2015

Identification of TMEM2 as a SOX4 Transcriptional Target Involved in Breast Cancer Metastasis

Hyeseung Lee

IDENTIFICATION OF TMEM2 AS A SOX4 TRANSCRIPTIONAL TARGET INVOLVED IN BREAST CANCER METASTASIS

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Hyeseung Lee

June 2015

IDENTIFICATION OF TMEM2 AS A SOX4 TRANSCRIPTIONAL TARGET INVOLVED IN BREAST CANCER METASTASIS

Hyeseung Lee, Ph.D.

The Rockefeller University 2015

The transcription factor SRY (sex determining region Y)-box 4 (SOX4) regulates embryonic development and has been shown to drive the progression and metastasis of multiple solid cancer types. A common transcriptional target of SOX4 that mediates its effects in multiple cancer types has yet to be identified. Through a systematic molecular and genomic approach, we identify the transmembrane protein 2 (TMEM2) gene as a direct transcriptional target of SOX4.

TMEM2 is transcriptionally activated by SOX4 in both breast cancer and lung adenocarcinoma cells and mediates pro-invasive and pro-migratory effects of SOX4. TMEM2, like SOX4, is sufficient to promote breast cancer metastatic colonization and its expression in high-risk primary breast tumors associates with metastatic relapse likelihood. Interestingly, independent studies wherein SOX4 or TMEM2 have been genetically inactivated have resulted in similar cardiac developmental defects, suggesting that TMEM2, in addition to mediating the pathologic effects of SOX4 on cancer progression, may also mediate its developmental effects.

ACKNOWLEDGEMENTS

During my graduate study at the Rockefeller University, I have imagined writing the acknowledgements of my thesis to overcome difficult times. This precious moment would not have been possible without the heartfelt support from many people who I have luckily met here.

Most of all, I would like to thank my mentor Sohail Tavazoie for his unwavering faith towards me. I have learned from him what to ask, how to approach, and how to critically view things as a scientist. Not only has he taught me these essential things as a thesis advisor, he has also shown his passion for science, by which I could not be more inspired. Every encounter with him gave me motivation to keep myself on the road of science. I also truly appreciate his patience and constant encouragement during my most difficult times. It is a great honor for me to be one of his early students and I will do my best to make him proud of me in future. I have a dream that someday in the future, he and I will meet in some conferences and discuss about our scientific work. I appreciate Sohail for enabling me to hold onto this dream.

I thank my thesis committee members, Sandy Simon and Daniel Mucida for their valuable advice and encouragement over the years, and Swarnali Acharyya for serving as the external examiner on my thesis committee.

I would next like to thank all the past and present members of the Tavazoie laboratory. First of all, I especially thank Claudio Alarcón, who was my mentor for my rotation project in this lab. He would then become an excellent teacher to me as well as a great research partner. It was a great pleasure for me to work with him. From working with him on our fantastic projects, I gained a broad knowledge of science, ranging from basic bench work to insights on how to design and analyze experiments. I also thank Hani Goodarzi for his expertise in analyzing bioinformatics-related data for all projects that I have been involved in as well as for all computer-related issues. I deeply thank my close lab mates and friends Jia Min Loo, Lisa Fish and Zander Nguyen for their time discussing my project and wonderfully editing my writing. I thank Emily Mandel for her help in administrative-related works and making lab life smoother. I am grateful to Yuehyi Gloria Wu, Raissa Tanqueco, Ethan Weinberg, Hoang Chuong Bui Nguyen, Helen Tian and Fung Ying Man for making the lab a very enjoyable place to work. I also appreciate the knowledge, reagents, and assistance from Nils Halberg, Caitlin Sengelaub, Xuhang Liu, Doowon Huh, Lisa Noble, Bernardo Tavora, Jason Ross, and Paul Furlow.

I appreciate the Rockefeller University community, especially Sidney Strickland, Emily Harms, Kristen Cullen, Marta Delgado, Cristian Rosario and Stephanie Fernandez in the Dean's Office for providing me with their support in the unique environment of this graduate program. My time at Rockefeller will be one of most memorable moments in my life.

I also appreciate my great undergraduate academic advisors Jaesang Kim, Sanghyuk Lee, and Chongmok Lee for their continuous interest and encouragement far away from Korea.

I would like to dedicate this thesis to my mother, Jungsoon Hwang, and to my father, Mooseong Lee. With their true love, I was able to exceed my potential. I deeply thank you for everything as always. I could not have my life here without mental support from grandmothers, Aerisook Im who has raised me up with her endless love and Soodeok Jang who has prayed for me all the time; my beloved sister, Yeonseung Lee, and my brother, Seungho Lee, for always acting as a confidant for me. I thank my uncles and aunts for their good wishes for me. Lastly I thank God for his blessings and for being with me at all times. I sincerely pray that God will hold me in my life-long pursuit of science for the benefit of humanity.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

PART I: BREAST CANCER

Breast cancer is the most common cancer in women and the second most common cancer across the entire population of the United States (The National Cancer Institutes Surveillance, Epidemiology, and End Results (SEER) database of 2015) (Fig. 1.1). It is estimated that 12 % of women in the United States will develop invasive breast cancer during their lifetime, which is estimated to be responsible for 3% of all female patients' death from cancer. Most patients die from metastatic disease (Gupta and Massagué, 2006), which is the spread of cancer cells from a primary tumor to distal organs and the growth of metastatic colonies in these organs.

	Common Types of Cancer	Estimated New Cases 2015	Estimated Deaths 2015
1.	**Breast Cancer (Female)**	**231,840**	**40,290**
2.	Lung and Bronchus Cancer	221,200	158,040
3.	Prostate Cancer	220,800	27,540
4.	Colon and Rectum Cancer	132,700	49,700
5.	Bladder Cancer	74,000	16,000
6.	Melanoma of the Skin	73,870	9,940
7.	Non-Hodgkin Lymphoma	71,850	19,790
8.	Thyroid Cancer	62,450	1,950
9.	Kidney and Renal Pelvis Cancer	61,560	14,080
10.	Endometrial Cancer	54,870	10,170

Figure 1.1 Breast cancer as the most common cancer in women

The National Cancer Institutes SEER (Surveillance, Epidemiology, and End Results) database of 2015 estimates 231,840 new cases of breast cancer and 40,290 patients who is predicted to die of this disease.

Most cancers by incidence such as skin, colon, prostate, lung, or breast cancer originate from epithelial tissues (Blanpain, 2013). Breast cancer arises from the mammary tissue, which consists of lobules (milk-producing glands), ducts (tubes that bring milk from the lobules to the nipple), and stroma (connective fatty tissue that surrounds lobules and vessels) (Fig. 1.2). Breast cancer is heterogeneous and can be classified into biologically meaningful and clinically meaningful subgroups (Weigelt et al., 2010). For clinical classification, breast cancer can be broadly classified into two types: *in situ* non-invasive cancer and invasive breast cancer. Non-invasive *in situ* breast cancer is comprised of ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS), which account for about 83% and 12% of all *in situ* cases respectively (The National Cancer Institutes SEER database from 2004 to 2010). Most *in situ* non-invasive breast cancers typically recur within 5 to 10 years. In the clinic, invasive breast cancers are staged based on the TNM staging system (T for primary tumors remaining within the breast, N for tumors infiltrating to the regional lymph nodes, M for presence or absence of distal metastases) according to the *AJCC* (American Joint Committee on Cancer) Cancer Staging Manual and Handbook. Cancers are further assigned a stage from 0 to IV (stage 0 reflecting *in situ* lesions, while stages I to III referring to invasive cancers in nearby lymph nodes and/or tissues; stage IV refers to metastatic cancer). The SEER Summary Stage system classifies patients into three groups; local stage that corresponds to stage I and partially stage II in the TNM staging system, regional stage that corresponds to

stage II or III, and distal stage that corresponds to stage IIIc and IV. The 5-year survival of patients diagnosed with localized or regional tumors (stage I-III) is 88% whereas only 22% of patients who develop tumors in distant organs survive greater than 5 years (The National Cancer Institutes SEER database of 2015) (Fig. 1.3).

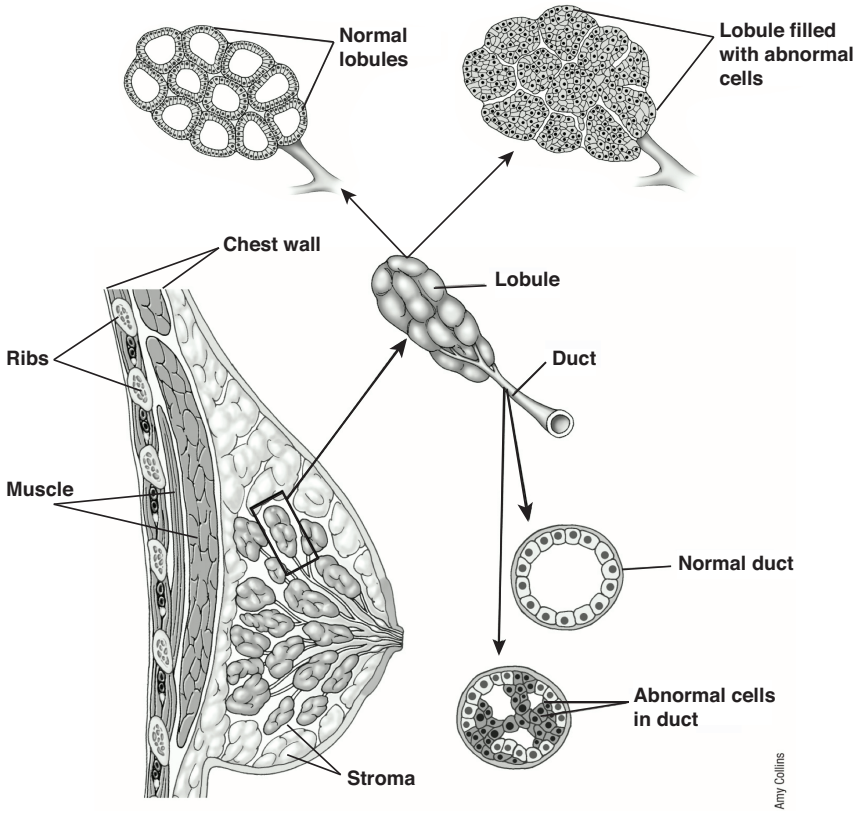


Figure 1.2 Breast tissue depicting normal or abnormal ducts and lobules

Figure was adapted from Breast cancer information provided by American cancer society and was further modified. A woman's normal breast mainly comprises lobules and ducts. The spaces between the lobules and ducts are filled with fibrous tissue and fat stroma. Most breast cancers originate from the abnormal cells lining the ducts or the lobules.

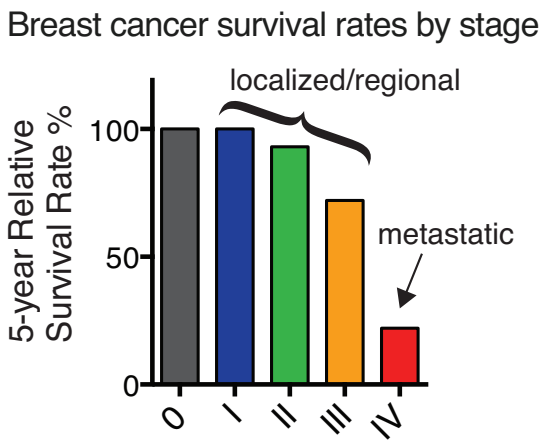


Figure 1.3 Breast cancer survival rates by stage

The rates above come from the National Cancer Institutes SEER database of 2015. The 5-year observed survival rate means the percentage of patients who live greater than or equal to 5 years after being diagnosed with breast cancer. Localized cancer refers to the cancer that is confined to primary site and regional cancer refers to the cancer that is spread to regional lymph nodes. Metastatic cancer refers to the cancer spread to distal organs. (stage 0 *in situ* lesions; stages I to III invasive cancers in nearby lymph nodes and/or tissues; stage IV metastatic cancer)

In addition to the TNM staging system, during laboratory examination of breast cancer tissue, pathologists preferentially use a tumor grading system. Tumor grade indicates how abnormal the tumor tissue look under a microscope, which is an indicator of how quickly a tumor is likely to grow and spread (*AJCC* Cancer Staging Manual. 7th ed. New York, NY: Springer; 2010). Grade 1 corresponds to well-differentiated tumors displaying morphology similar to normal

cells. Grade 2 refers to moderately-differentiated tumors. Grade 3 refers to poorly differentiated tumors showing enhanced growth and invasiveness. Generally, higher tumor grade correlates with higher stage, and indicates a poorer prognosis.

Breast cancers can be grouped into molecular subtypes which is associated with clinical outcomes. The major molecular subtypes of breast cancer are basal-like, luminal A, luminal B, HER2 positive/ER negative, and normal breast-like (Carey et al., 2006). It is possible that normal breast-like breast cancer could be falsely diagnosed as an artifact of sample presentation (contamination by high content of normal tissue) (Fig. 1.4). Currently there are two widely accepted hypothetical models explaining the origins of cancer; the cancer stem cell hypothesis and the clonal evolution model (Polyak, 2007). In the cancer stem cell theory, self-renewing, stem-like cancer cells exist and only this cancer stem cell population can drive cancer progression by accumulating genetic changes that are required for this process. In the clonal evolution theory, cross-talk among tumor-initiating cells and tumor stromal cells drives cancer progression and all tumor cells are capable of self-renewing division. The two models do not have to be mutually exclusive and a model encompassing clonal evolution of cancer stem cells is possible.

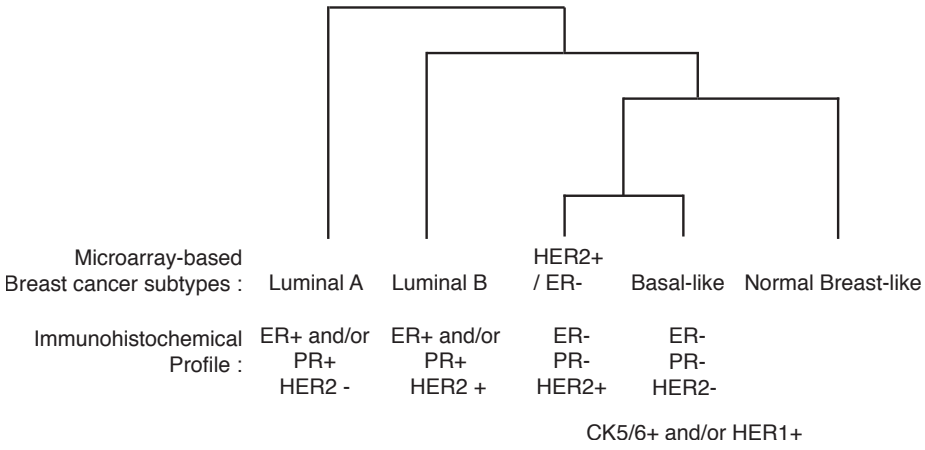


Figure 1.4 Microarray-based breast cancer subtypes with immunohistochemical profile

The hierarchical clustering analysis of intrinsic gene list with immunohistochemical classification schema is shown. The gene expression data for estrogen receptor (ER), human epidermal growth factor receptor-2 (HER2), CK5 (cytokeratin), and HER1 were obtained from the 115 patients' tumor samples from the Carolina Breast Cancer Study (ascertained between May 1993 and December 1996). This figure was adapted from (Carey et al., 2006) and reproduced.

Current therapies for breast cancer include surgery, radiation therapy, and systemic therapy including chemotherapy, hormone therapy, and targeted therapy (American Cancer Society. Breast Cancer Facts & Figures 2013-2014. Atlanta: American Cancer Society, Inc. 2013). In recent years, major accomplishments in drug development have resulted in the approval of drugs targeting hormone receptors highly expressed in breast cancer cells such as estrogen receptor (ER)

and progesterone receptor (PR) or targeting a growth-promoting protein, the erb-b2 receptor tyrosine kinase 2 (ERBB2, also known as HER2). Tamoxifen inhibits binding of estrogen to estrogen receptors. Trastuzumab (Herceptin), pertuzumab, ado-trastuzumb emtansine (adcyla, also known as TDM-1), and Lapatinib (Tykerb) function as inhibitors of HER2, resulting in reduced tumor size or repressed tumor growth. Treatment of ER positive, PR positive, or HER2 positive breast cancer with these targeted therapies has contributed to improved survival. However, about 14 to 20% of all breast cancers are classified as Triple Negative Breast Cancer (TNBC) and do not overexpress HER2 or ER and PR. TNBC exhibits the worst outcome of the all breast cancer subtypes because these tumors tend to grow aggressively and are not susceptible to hormone therapy and anti-HER2 therapies (Koboldt et al., 2012). Considering that the majority of breast cancers with mutations in BRCA1 and BRCA2 have been classified as TNBC, the potential benefit of Poly (ADP-ribose) polymerase (PARP) inhibitors such as olaparib have been investigated as agents against BRCA1/BRCA2 mutant TNBC (Turner and Reis-Filho, 2006; Atchley et al., 2008). However, the biology of TNBC is poorly understood and targeted therapies for TNBC have not been approved.

PART II: METASTASIS

Cancer is the second most common cause of death in the US, accounting for nearly 1 of every 4 deaths (American Cancer Society. Cancer Facts & Figures 2014. Atlanta: American Cancer Society; 2014). The main cause of death in most cancer patients is metastasis, the spread of cancer cells from a primary tumor to distal organs (Batistatou et al., 2009). Therefore, understanding the biology of metastasis and developing therapies targeting this lethal process of cancer progression is crucial to reducing patient deaths. Metastasis has been viewed as a collective process consisting of sequential steps. After the transformation of normal cells to malignant cells, metastasis follows sequential steps such as the formation of blood vessels feeding nutrients to cancer cells and providing a route to distant organs, intravasation (invasion of cancer cells into the circulation), survival in circulation, extravasation (invasion of cancer cells out of vessel and into distal organ parenchyma), metastatic colony formation by proliferating cancer cells in end organs, growth, and re-initiation of these processes at secondary sites for metastasis to additional organs (Fig. 1.5). Given the complexities of the metastatic cascade, it is unlikely that a single gene would be sufficient to complete all steps in the metastatic cascade.

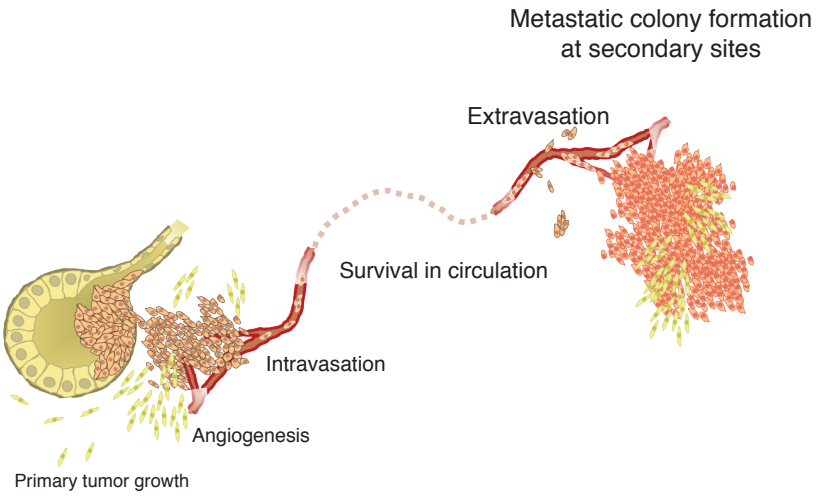


Figure 1.5 The sequential process of metastasis

After the transformation of normal cells to malignant cells, metastatic progression follows sequential steps such as angiogenesis, intravasation, survival in circulation, extravasation, colony formation in secondary organs, and re-initiation of these processes at secondary sites for metastasis to additional organs.

To understand this complex process of metastasis, I.J. Fidler and colleagues pioneered *in vivo* selection, a technique whereby sub-populations of tumor cells were selected for enhanced metastatic potential from the original non-metastatic tumor cell population (Fidler, 1973). It has been shown that primary tumors are genomically unstable with mutations in oncogenes and tumor suppressors and also exhibit epigenetic alterations. It also has been observed that the cell population in the primary tumor is heterogeneous and its subpopulations that endure selection pressures are capable of metastasis (Talmadge and Fidler, 1982; Nicolson and Dulski, 1986; Fidler, 2003). By systematically comparing the gene expression profiles of the poorly metastatic

parental breast cancer cells with highly metastatic sublines, we can identify the genes that are responsible for governing metastatic progression.

From studies in various types of cancers, it has been suggested that tumor cells have preferences for certain sites of metastasis (Chambers et al., 2002). In breast cancer, liver, brain, bone, and lung have been reported as common secondary sites of breast cancer metastasis (Paget, 1989; Yoneda et al., 2001; Kang et al., 2003; Minn et al., 2005) whereas prostate cancer usually disseminates to bone, and colorectal cancer to liver (Rinker-Schaeffer et al., 2000; Kuo et al., 1995). This organ specificity hypothesis has been called the “seed and soil theory” of cancer spread — the site of secondary tumors is dependent on the origin of the tumor cell (“seed”) and the target organ (“soil”) — originally outlined by Stephen Paget in 1889. In support of this hypothesis, it has been observed that endothelial cells in vessels in different organs express different adhesion proteins that could affect patterns of metastatic tumor spread (Nicolson, 1988; Pauli et al., 1990). Chemokines are also known to have a role in guiding organ specific metastasis. High levels of CXCL12, a soluble ligand for chemokine receptor CXCR4, in lung would guide breast cancer cells expressing high levels of CXCR4 through chemokine-mediated signal pathways, which would result in metastatic colonization by breast cancer cells in lung (Mukherjee and Zhao, 2013). Similarly, melanoma cells express elevated levels of CCR10 receptor which would strongly bind to its ligand CCL27 that is highly expressed in skin, thereby contributing to melanoma metastasis to distant skin (Monteagudo et al., 2011).

MicroRNAs (miRNAs) are 19 to 24 nucleotides-long single-stranded RNA molecules that regulate gene expression in a sequence-specific manner. In the nucleus, RNA polymerase II transcribes a much longer pri-miRNA, which forms a hairpin loop structure. Then, Drosha cleaves the large pri-miRNA to generate the pre-miRNA that is then translocated to the cytoplasm by exportin 5-mediated export. In the cytoplasm, Dicer processes it into a short RNA duplex. This duplex is unwound upon loading into the RISC complex (RNA-induced silencing complex, comprising proteins such as Dicer and members of the Argonaute family) and finally the single-stranded mature miRNA is generated. Consequently, the mature miRNA is able to bind mRNAs through its seed sequence that are partially complementary to its target mRNA sequences. Thus target mRNAs could be degraded and translation could be suppressed through miRNA binding, which ultimately leads to a reduction in protein expression levels (Nicoloso et al., 2009). Considering that deregulated miRNAs could initiate alterations in expression of many downstream target genes, miRNAs are attractive candidates for clinical biomarker (Bartel, 2004; Zhang et al., 2007; Ventura and Jacks, 2009).

Recently, *in vivo* selected cell lines have been used to identify molecular factors and key regulatory pathways that mediate metastasis in various cancer types (Minn et al., 2005; Gupta et al., 2005; Acharyya et al., 2012; Pencheva et al., 2012; Png et al., 2012; Loo et al., 2015). In breast cancer, molecular analysis of independently derived *in vivo* selected cell populations led to the identification of endogenous human microRNAs, miR-335 and miR-126, (miRNAs) that

suppress breast cancer metastasis (Tavazoie et al., 2008). These miRNAs are silenced in metastatic human breast cancer cell lines and restoring their expression in these cells dramatically suppresses lung and bone metastasis by breast cancer cells in mice. Supporting a clinical role for these miRNAs, reduced expression of miR-335 or miR-126 in human primary tumors is significantly associated with poor metastasis-free survival.

Recent studies revealed that the activities of miRNAs demonstrate tissue specificity and can regulate different pathways during metastasis in different types of tissues. MiR-206 and miR-1 has been shown to negatively regulate angiogenesis during zebrafish development by targeting Vascular endothelial growth factor A (VegfA) in muscle (Stahlhut et al., 2012). Gastric cancer cells having downregulated expression of miR-206 displayed enhanced cells growth and colony forming ability (Zhang et al., 2013b). Recently it also has been reported that miR-206 loss also promotes gastric cancer metastasis by activating paired box gene 3 (PAX3) and MET proto-oncogene (MET) pathways (Zhang et al., 2015). In breast cancer, lower expression levels of miR-206 are correlated with larger tumor size whereas overexpression of miR-206 resulted in reduced cell proliferation by targeting cyclinD2 (Zhou et al., 2013). Molecular mechanism of miR-206 has been further studied in triple-negative breast cancer cells; miR-206 repressed migration ability of triple-negative breast cancer cells by targeting actin-binding protein coronin 1C (CORO1C), resulting in changes in actin skeleton and cell morphology (Wang et al., 2014a).

Our laboratory has discovered the role of miR-126 in breast cancer metastasis (Png et al., 2012). By integrative transcriptomic profiling of highly lung metastatic compared to nonmetastatic breast cancer cells in the context of miR-126 overexpression and knockdown, Insulin-like growth factor binding protein 2 (IGFBP2), phosphatidylinositol transfer protein (PITPNC1), and c-Mer tyrosine kinase (MERTK), were identified as miR-126 target genes that promote endothelial cell recruitment; IGFBP2 recruits endothelial cells to cancer cells through IGF1/IGF1R pathways. PITPNC1 promotes secretion of IGFBP2 by cancer cells. MERTK in cancer cells competes with MERTK in endothelial cells for binding to its ligand, growth arrest-specific 6 (GAS6), thereby resulting in promoting IGFBP2 independent endothelial cell recruitment. MiR-126 and miR-126* (a miRNA pair derived from a same precursor) has been also reported to suppress breast cancer metastasis by inhibiting mesenchymal stem cells and inflammatory monocytes (Zhang et al., 2013c). MiR-126 functions as metastasis suppressor in other types of cancer including hepatocellular carcinoma by targeting LRP6 and PIK3R2 to repress angiogenesis (Chen et al., 2013; Du et al., 2014), colorectal cancer by repressing angiogenesis through inhibiting VEGF and by suppressing cell invasion via inactivation of RhoA/ROCK signaling pathway (Zhang et al., 2013d; Li et al., 2013)

Similar to miR-206 and miR-126, miR-335 has also been reported as a metastasis suppressor in various types of cancer. MiR-335 suppresses ovarian cancer by repressing expression of B-cell CLL/lymphoma 2 like 2 (Bcl-w) and its effecter molecule matrix metalloproteinase-2 (MMP-2) (Cao et al., 2013). In osteosarcoma, downregulated miR-335 has been shown to be correlated with lymph node positive metastasis and it inhibits the cell migration and invasion by repressing Rho-associated coiled-coil containing protein kinase 1 (ROCK1) (Wang et al., 2013). It is known that miR-335 negatively regulates octamer-binding transcription factor 4 (OCT4) expression to suppress pancreatic cancer progression (Gao et al., 2014). Loss of miR-335 expression in breast cancer led to the enhanced expression of tenascin C (TNC) and the transcription factor SRY (sex determining region Y)-box 4 (SOX4), yielding enhanced metastatic progression through increased cell invasion and migration (Tavazoie et al., 2008). Along with the discovery of miR-335's downstream target genes in breast cancer, our laboratory has also found that miR-335 expression is regulated in human breast cancer cells by genetic deletion and epigenetic silencing of the locus (Png et al., 2011). MiR-335 is located on human chromosome 7q32 domain, within the second intron of a paternally expressed gene, Mesoderm-specific transcript (MEST). Firstly, the expression of miR-335 and MEST have been found to be positively correlated in a number of cell lines. Additionally, the promoter region of MEST is highly methylated in metastatic breast cancer cells compared to poorly metastatic cells. Taken together, it is suggested that the low expression of

miR-335 in highly metastatic cells results from epigenetic silencing through hypermethylation of the promoter of MEST. Secondly, the miR-335 locus is deleted in highly metastatic derivatives compared to poorly metastatic parental cells. Thirdly, in the matched primary/metastasis clinical breast cancer patients samples, miR-335 expression in primary samples were higher than its expression in metastasis samples in which the reduction in miR-335 copy number were observed.

Identification of master metastasis-regulating molecules like miRNAs is important because not only do these molecules provide clues about the downstream pathways involved in metastasis but they also could be attractive targets for more effective drug development. In line with this, global transcriptomic analysis following recent advances in sequencing technologies and computational biology has identified several transcription factors as regulators of metastasis. For example, a basic leucine zipper transcription factor, BACH1, has been identified as a key regulator of breast cancer bone metastasis by regulating MMP1 and CXCR4, both of which are known to promote metastasis by various types of tumors (Liang et al., 2012). A basic helix-loop-helix leucine-zipper transcription factor, C-Myc, has also been shown to function as oncogenic transcription factor which regulates invasion and migration by activating genes in the epithelial-mesenchymal transition (EMT) pathway like osteopontin and RhoA (Wolfer and Ramaswamy, 2011). Transforming growth factor β (TGFβ) mediated activation of mutant p53 pathway is also known to promote invasion and

metastasis by blocking p63, another transcription factor that suppresses metastasis (Clohessy and Pandolfi, 2009). Discovery of transcription factors functioning as positive or negative regulators of metastasis has raised interest in the development of drugs targeting such transcription factors and their downstream effectors (Ell and Kang, 2013).

PART III: ROLE OF SOX4 IN METASTASIS

To date, 20 Sry-related HMG box (Sox) genes have been identified in vertebrate species, and are subdivided into 8 groups (A-H) based on sequence similarity of the HMG domain, full-length protein structure, and gene organization data (Bowles et al., 2000). Sox transcription factors are widely expressed in various tissues during early development. Since it was discovered as a master regulator of testis differentiation and expressed at the onset of gender determination (Eggers et al., 2014), Sox genes have been mostly studied in developmental biology. The group C of Sox transcription factors in most vertebrates consists of SOX4, SOX11, and SOX12. SoxC members have been shown to exhibit overlapping expression patterns and biological properties (Dy et al., 2008). They are highly co-expressed in neuronal and mesenchymal tissues in the developing mouse but also display independent expression patterns in various tissues. Sox11-deficient mice die at birth due to congenital cyanosis caused by heart defects (Sock et al., 2004). Sox11 depletion in sensory neurons results in reduction in neuronal survival and an arrest of axonal outgrowth (Lin et al., 2010b). Sox12-deficient mice develop normally due to compensation of its loss by nonreciprocal redundant functions of SOX4 and Sox11 (Hoser et al., 2008).

The human SOX4 gene was firstly characterized and mapped in 1993 (Farr et al., 1993) and its role as a transcription factor was first studied in mouse lymphocyte development (Vandewetering et al., 1993). It is known that the WWCAAG motif, a known functional T cell specific enhancer motif, is bound by DNA-binding HMG box proteins, T cell specific transcription factor 1 (TCF1) and lymphoid enhancer binding factor LEF1. It was observed that another HMG domain containing transcription factor family, the Sox transcription factors, were expressed in T and pre-B lymphocyte lines and in the murine thymus and could possibly bind the WWCAAG motif. The search for Sox genes that are responsible for transcriptional transactivation through the T cell specific enhancer motif revealed that SOX4 functions as a transcriptional activator in B lymphocyte differentiation. Hematopoietic cells with SOX4 deletion remained at the pro-B cell stage and the expression of SOX4 has been correlated with poor clinical outcome in B-cell leukemia. More recent studies have supported SOX4's oncogenic role in leukemia in cooperation with other proteins such as SPI1 (ETS transcription factor family involved in hematopoietic development), CREB (a nuclear transcription factor critical for hemopoiesis), PML-RARα (a fusion product serving as the initiating event in acute promyelocytic leukemia), FRA-2 (an AP-1 transcription factor family member), and PI3K/AKT signaling molecules (Aue et al., 2011; Sandoval et al., 2012; Omidvar et al., 2013; Zhang et al., 2013a; Higuchi et al., 2013; Ramezani-Rad et al., 2013).

In addition to the role of SOX4 in B cell differentiation, genetic inactivation of SOX4 in mouse revealed its critical role during heart development. Mammalian heart development requires coordinated regulation of cell differentiation at the correct time and migration into the correct place (Fig. 1.6 A). It begins with the formation of two endocardial tubes, which become a single primitive heart tube that eventually loops and septates into 4 chambers and arterial trunks. The truncus arteriosus should be further separated to become the two main arteries, the aorta and the pulmonary artery. The failure of the division of truncus arteriosus results in a phenotype called common arterial trunk with lack of ventricles (Buckingham et al., 2005) (Fig. 1.6 B). SOX4 null mice die from failure in circulation caused by severely impaired heart development at embryonic day (ED) 14 (Schilham et al., 1996). The homozygous SOX4 deficient embryos were not distinguishable externally until the onset of valvular morphogenesis in the arterial pole of the heart, which happens between ED 12 and ED 13 (Ya et al., 1998). The outflow tract malfunction eventually results in edema which leads to embryonic lethality at ED 14. Histological evaluation of the SOX4 knockout embryo further demonstrated the common arterial trunk, which could result from impaired growth of the aorticopulmonary septum and the endocardial ridges of the outflow tract. Hypoplastic endocardial cushions and semilunar valves led incomplete ventricular septation and caused arterial blood back flow.

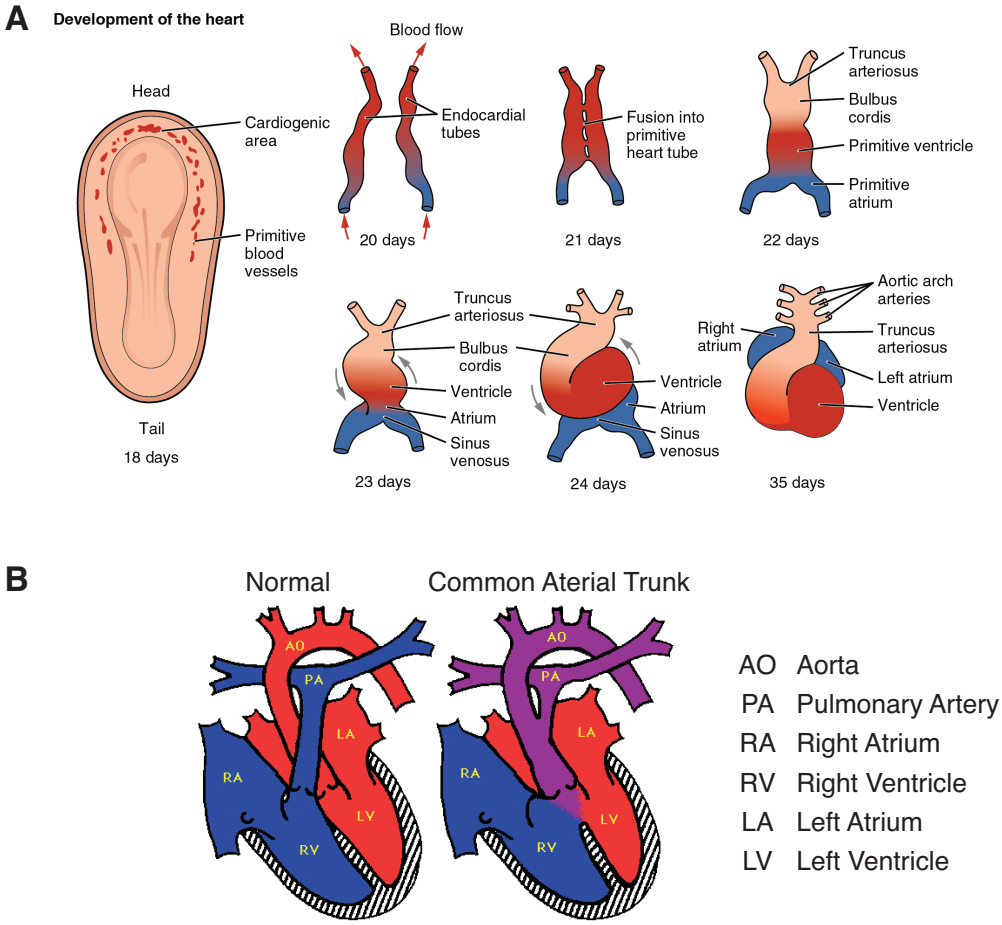


Figure 1.6 Heart development

(A) The embryological development of the human heart during the first 8 weeks is delineated. Heart originates from a pair of strands called cardiogeneic cords and endocardial tubes. The fusion of the strands then occurs and it starts differentiating into the truncus arteriosus, bulbus cordis, primitive ventricle, primitive atrium, and sinus venosus. This primitive heart forms an S shape, which determines the location of the chambers and major vessels of the adult heart. Illustration from Anatomy & Physiology, Connexions Web site. http://cnx.org/content/col11496/1.6/ (B) Common atrial trunk (Truncus arteriosus). Illustration from http://pixgood.com/persistent-truncus-arteriosus.html

In addition to its role in heart development, SOX4 is also known to regulate osteogenesis and insulin producing pancreatic cell development. SOX4 heterozygous mice show defects in bone formation from an early age resulting from suppressed proliferation, differentiation, and mineralization of osteoblasts (Nissen-Meyer et al., 2007). SOX4 null pancreatic explants displayed reduced numbers of insulin producing beta cells (Maria E Wilson, 2005).

Interestingly, as discussed earlier, the identification of SOX4 as a target of miR-335 shed light on the biological role of SOX4 in cancer. Silencing of miR-335 in breast cancer cells results in the increased expression of tenascin C (TNC) and SOX4, enhanced cell invasion, migration, and metastatic progression. One of the miR-335 regulated genes, TNC, has been shown to support the metastasis-initiating capacity of breast cancer cells by enhancing the expression of musashi homolog 1 (MSI1) and leucine-rich repeat-containing G protein-coupuled receptor 5 (LGR5), a positive regulator of NOTCH signaling and a target gene of the WNT pathway, respectively (Oskarsson et al., 2011). The other miR-335 regulated gene, SOX4, has been shown to play important roles in cancer progression and metastasis. Subsequent to its discovery as a mediator of breast cancer metastasis, more recent studies have shown that SOX4 expression is elevated in other cancer types, including leukemia (Zhang et al., 2013a; Ramezani-Rad et al., 2013), glioblastoma (Ikushima et al., 2009; Lin et al., 2010a), medulloblastoma (Lee et al., 2002; de Bont et al., 2008), hepatocellular carcinoma (HCC) (Liao et al., 2008), and prostate cancer (Liu et al., 2006; Scharer

et al., 2009; Lai et al., 2011).

Despite the above evidence supporting the oncogenic properties of SOX4, the SOX4 regulatory program appears to be cell type and context specific (Vervoort et al., 2013b). SOX4 regulatory networks found in prostate cancer include transducin-like enhance of split (TLE-1), BCL2 binding component 3 (BBC3/PUMA), and epidermal growth factor receptor (EGFR) (Liu et al., 2006; Scharer et al., 2009; Lai et al., 2011). In HCC, two genes in the category of axon guidance, neurophilin-1 (NRP1) and semaphorin 3C (SEMA3C), were identified as SOX4 target genes that promote migration and intrahepatic metastasis of HCC cell (Liao et al., 2008). In glioblastoma, TGF-β-mediated SOX4 induction promotes the expression of Sox2 to maintain the stemness of glioma-initiating cells (Ikushima et al., 2009; Lin et al., 2010a). In leukemia, SOX4 has been shown to transcriptionally activate promoters of multiple components within the PI3K/AKT and MAPK signaling pathways (Zhang et al., 2013a; Ramezani-Rad et al., 2013). However, the question remains whether there exists SOX4 target genes that drive a common metastasis-related SOX4-dependent phenotypes such as invasion and migration in multiple cell types. In this study, we identify and characterize TMEM2 as a metastasis-promoting gene that is transcriptionally regulated by SOX4 in metastatic breast cancer cells and mediates pro-migratory and pro-invasive effect downstream of this transcription factor.

CHAPTER 2

SYSTEMATIC ANALYSIS OF SOX4-REGULATED TARGET GENES IN BREAST CANCER

PART I: TRANSCRIPTOMIC PROFILING

To identify potential SOX4 target genes, we employed ChIP-Seq and transcriptomic profiling approaches. Given that SOX4 is a known transcriptional activator, and its expression is upregulated in highly metastatic LM2 cells compared to poorly metastatic MDA-MB-231 cells, we searched for genes for which expression levels were decreased following RNA interference-mediated SOX4 depletion, and that exhibited increased expression in the metastatic LM2 cells, which endogenously express higher levels of SOX4 relative to their parental line. By comparing microarray expression profiles of LM2 cells transfected with either an siRNA targeting SOX4 or a control siRNA, we identified possible SOX4-regulated genes which expression was diminished in cells depleted of SOX4 relative to control cells. Biological replicates were used for this analysis. The cells were transfected using independent siRNAs, and RNA was extracted 96 hours after siRNA transfection using the miRVana kit according to the manufacturer's instructions. RNA samples were processed by the Genomics core facility at Rockefeller University using Illumina BeadChip arrays (HumanHT-12 V4 BeadChip). RAW signals from each samples were median-normalized and the ratio between SOX4 knockdown and controls was calculated by averaging the normalized signal from the biological replicates. Of 31,424 genes tested, 2,225 genes were down-regulated in the SOX4 knockdown cells compared to the control cells (genes that were expressed lower than SOX4). The relative expression levels of these 2,225 genes were determined from a microarray experiment

comparing MDA-MB-231 cells and LM2 cells. Of the 2,225 genes downregulated upon SOX4 depletion, 1,961 genes were upregulated in the LM2 relative to the MDA-MB-231. There were 1,961 genes whose expression fold enrichment in LM2 relative to the MDA-MB-231 was equal or higher than the SOX4 fold enrichment.

PART II: PREDICTION OF POTENTIAL SOX4-REGULATED DIRECT TARGET GENES BY CHIP-SEQ

Summary

To determine which of these genes might be direct transcriptional targets of SOX4, we performed chromatin immunoprecipitation of SOX4, followed by high throughput sequencing (ChIP-Seq) in LM2 cells. Considering that the C-terminal 33 amino acids of SOX4 constitute the transactivation domain (Dy et al., 2008), we engineered LM2 cells to stably express exogenous N-terminal FLAG-tagged SOX4 to prevent interfere of the FLAG peptide with its transactivation activity. Additionally, we produced control cells, expressing an empty vector and used anti-FLAG antibodies for immunoprecipitation. Sequencing peaks from empty-vector expressing cells were used as the negative control to remove background reads caused by potential non-specific binding of the FLAG antibody to DNA. Analysis of peaks from SOX4 binding events in the LM2 cells revealed that the majority of SOX4-DNA interactions were in intergenic regions, followed by distal regions, introns, and promoter regions (Fig. 2.1 A). To identify SOX4 binding motifs in LM2 cells, the ChIP-Seq data was analyzed by FIRE (Finding Informative

Regulatory Elements), a motif discovery and characterization program based on mutual information (Elemento et al., 2007). FIRE analysis revealed that the DNA sequence SRAAATATCY was statistically over-represented in the ChIP-Seq peaks identified in the SOX4 overexpressing LM2 cells (Fig. 2.1 B). This A/T-rich motif is consistent with known SOX4 binding sites obtained from global analysis of the sequence-specific binding of human transcription factors using high-throughput SELEX and ChIP sequencing (Jolma et al., 2013).

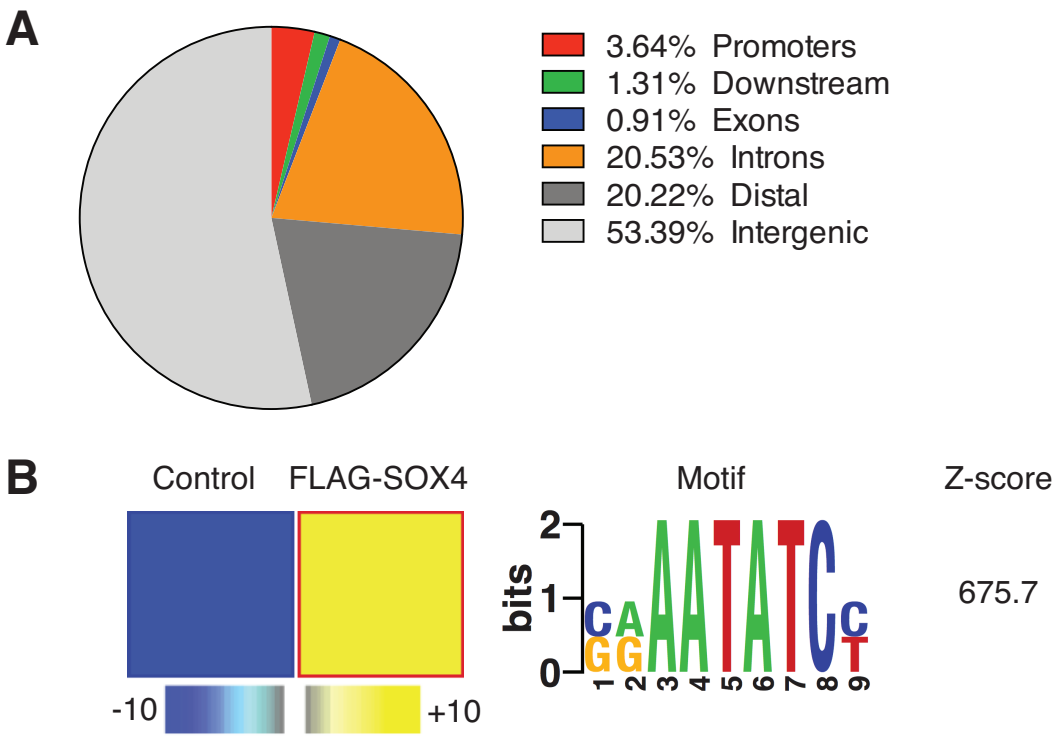


Figure 2.1 ChIP-Seq analysis of SOX4 in LM2 cells

(A) Pie chart depicting the distribution of clustering of enriched DNA sequences obtained from N-terminal FLAG tagged SOX4 expressing LM2 cells. (B) The enriched SOX4 binding motifs from the true peak sequences identified by FIRE analysis.

Although microarray technology has been widely used to screen target genes of transcription factors, it produces a large number of false positives. Also, it is impossible to distinguish a direct target from an indirect target by this approach. In the case of a direct target, a transcription factor binds directly to either the promoter or enhancer regions of the target to activate transcription. For

an indirect target, a transcription factor is still functionally responsible for regulating the downstream gene, but does not directly bind to the regulatory region of the gene. Instead, the transcription factor directly regulates the expression of an upstream regulatory protein that then directly controls the expression of the gene (Weinmann, 2004). To circumvent this, ChIP-Seq can be used to identify both direct targets and the binding motifs (Qin et al., 2011). SOX transcription factor family members all bind a similar DNA motif WWCAAW (Vandewetering et al., 1993) and various sequences for SOX4 binding sites have been reported in different types of cells. In human mammary epithelial cells, SOX4 was shown to bind an intronic region of CDH2, with the binding site sequence of either CCTTTGTT or AACAATGA (Vervoort et al., 2013a). In T helper type 2 cells, SOX4 binding sites are located near GATA-3 binding sites, and the sequence of the binding site is TGATTGTT (Kuwahara et al., 2012). In prostate cancer cells SOX4 was shown to bind the motif AACAAAGG (Scharer et al., 2009). Therefore we decided to use both SOX4 ChIP-Seq data and microarray gene expression data to discover direct targets of SOX4 and to identify SOX4 binding motifs.

Total cell lysate from LM2 cells that were transduced with either a control empty vector or a FLAG-SOX4 vector were immunoprecipitated with anti-FLAG beads and then immunoblotted with an anti-FLAG mouse monoclonal antibody (Fig. 2.2 A) or an anti-SOX4 rabbit polyclonal antibody (Fig. 2.2 B).

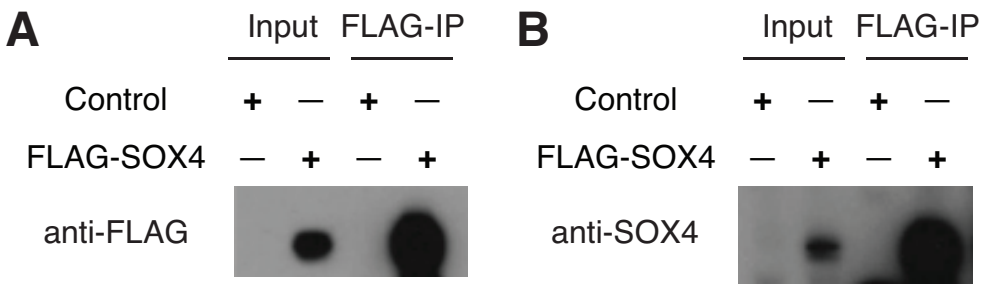


Figure 2.2 Immunoprecipitation of FLAG-SOX4 for ChIP-Seq

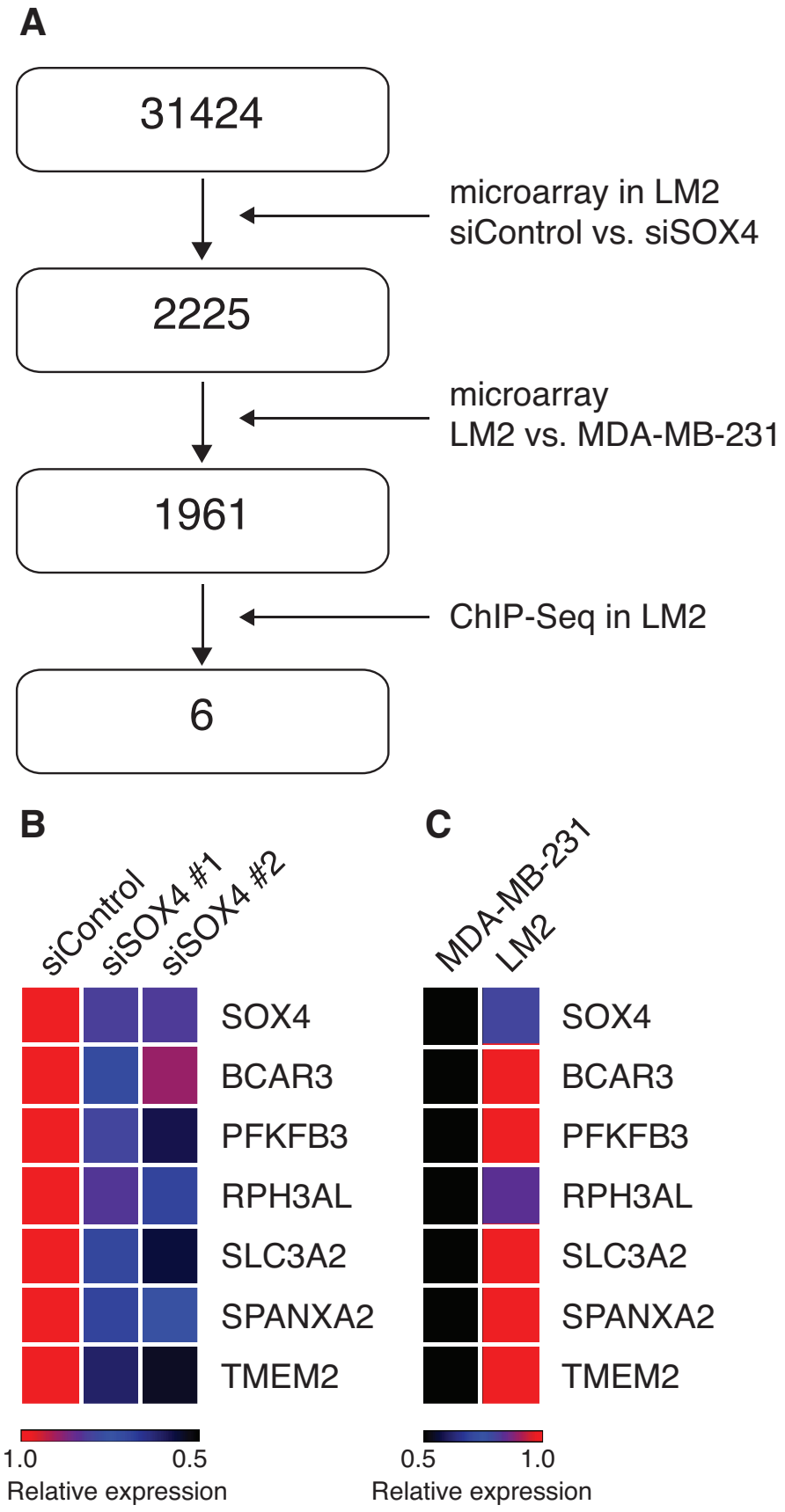
(A-B) Nuclear lysate from LM2 cells over-expressing either FLAG-SOX4 or control empty vector were immunoprecipitated with anti-FLAG beads, separated by SDS-PAGE and visualized by immunoblot with anti-FLAG antibody (A) and anti-SOX4 antibody (B).

The ChIP-Seq data was analyzed by FIRE to find the SOX4 consensus binding motifs in this system. As a control, we used randomly generated sequences of identical length while keeping the di-nucleotide frequencies constant. Frequencies of each motif were compared to the frequency of random sequences. The likelihood of SOX4 binding to the sequences obtained from the ChIP-Seq experiment versus random sequences were measured using mutual information and the Z score was calculated from the mutual information obtained from 10,000 random shufflings of the motifs. FIRE analysis revealed that the DNA sequence SRAAATATCY was statistically over-represented in the SOX4 ChIP-Seq data. This analysis of the ChIP-Seq data revealed that 426 genes were significantly bound by SOX4. The 1,961 genes that were downregulated in the context of SOX4 knockdown as well as upregulated in LM2 compared to

MDA-MB-231 cells, were then overlapped with the 426 genes that were bound by SOX4. From this analysis, we identified 6 genes that were putative direct SOX4 target genes (Fig. 2.3 A-C).

Figure 2.3 A systematic strategy for identification of direct SOX4 target genes that regulate breast cancer metastasis

(A) Potential functional SOX4 target genes were identified by the overlap between genes whose expression were depleted in the context of SOX4 knockdown by two independent siRNAs versus control siRNA (cut-off 1.2 fold-decrease), upregulated in metastatic breast cancer cell population LM2 relative to parental MDA-MB-231 (cut-off 1.3 fold-increase), and genes whose promoter regions were identified to be bound by SOX4 through ChIP-Seq ($P < 1\times10^{-5}$). (B-C) Heat map depicting the mRNAs of putative SOX4 target genes. Red represents higher expression and blue lower expression levels. (B) Relative expression levels of the 6 genes from microarray data from LM2 cells expressing control siRNA or two independent siRNAs targeting SOX4. (C) Relative expression levels of the 6 genes from microarray data comparing LM2 cells vs. MDA-MB-231 cells.


A
31424
microarray in LM2
siControl vs. siSOX4
2225
microarray
LM2 vs. MDA-MB-231
1961
ChIP-Seq in LM2
6
B
siControl
siSOX4 #1
siSOX4 #2
SOX4
BCAR3
PFKFB3
RPH3AL
SLC3A2
SPANXA2
TMEM2
1.0
0.5
Relative expression
C
MDA-MB-231
LM2
SOX4
BCAR3
PFKFB3
RPH3AL
SLC3A2
SPANXA2
TMEM2
0.5
1.0
Relative expression

CHAPTER 3

VALIDATION OF MICROARRAY ANALYSIS

PART I: QRT-PCR VALIDATION OF SOX4-REGULATED TARGET GENES

To validate our gene expression profiling analyses, we performed quantitative real-time polymerase chain reaction (qRT-PCR) for these genes. By qRT-PCR, we observed that the mRNA levels of RPH3AL, SLC3A2, and TMEM2, three of the six potential targets of SOX4, were upregulated in the highly metastatic LM2 cells which endogenously overexpress SOX4 relative to the poorly metastatic MDA-MB-231 parental cell line (Fig. 3.1 D, E, G). Depletion of SOX4 by two independent siRNAs targeting SOX4 resulted in significant reduction of the RNA levels of RPH3AL, SLC3A2, and TMEM2 relative to the control siRNA transfected cells (Fig. 3.1 K, L, N)

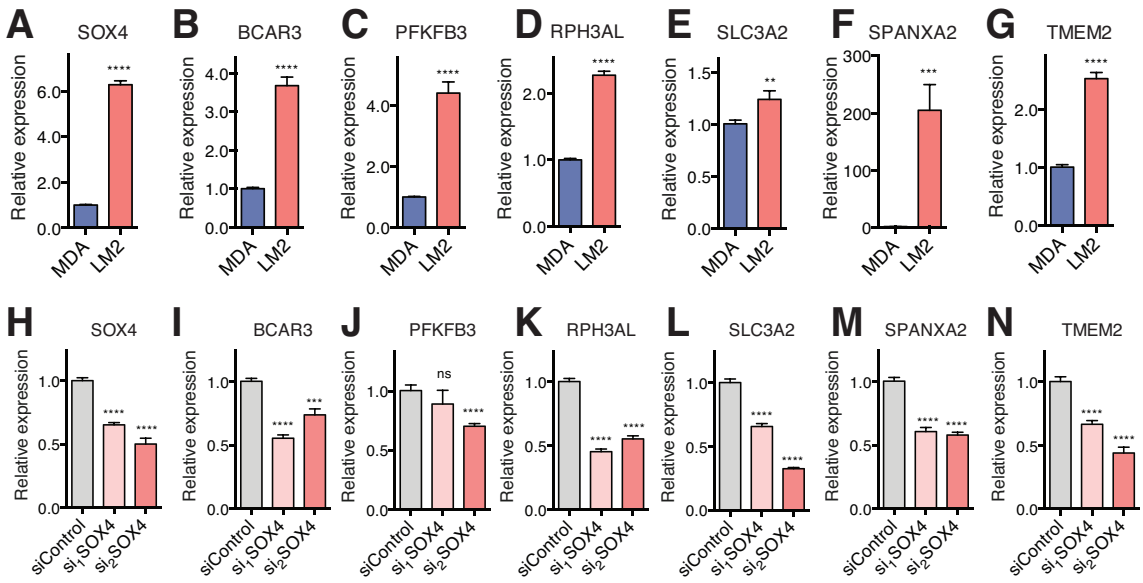


Figure 3.1 QRT-PCR validation of the 5 putative SOX4 regulated target genes in MDA-MB-231/LM2 cell lines

(A-G) qRT-PCR measurements of mRNA expression of 6 target genes in the parental MDA-MB-231 cells as well as LM2 cells (n=3). (H-N) qRT-PCR measurements of mRNA expression of 6 target genes in LM2 cells transfected with control siRNA or two siRNAs targeting SOX4 (n=3).

The relative expression level of these three genes and SOX4 was further assessed in the CN34 cell line, an independent malignant breast cancer cell population, and its highly metastatic lung derivative, the LM1A subline. All three genes were upregulated in the LM1A subline which over expresses endogenous SOX4 relative to its non-metastatic CN34 parental line (Fig. 3.2 D, E, G). Knockdown of SOX4 in LM1A cells also decreased mRNA levels of the three genes (Fig. 3.2 K, L, N), suggesting that these genes could be SOX4-regulated general promoters of breast cancer metastasis.

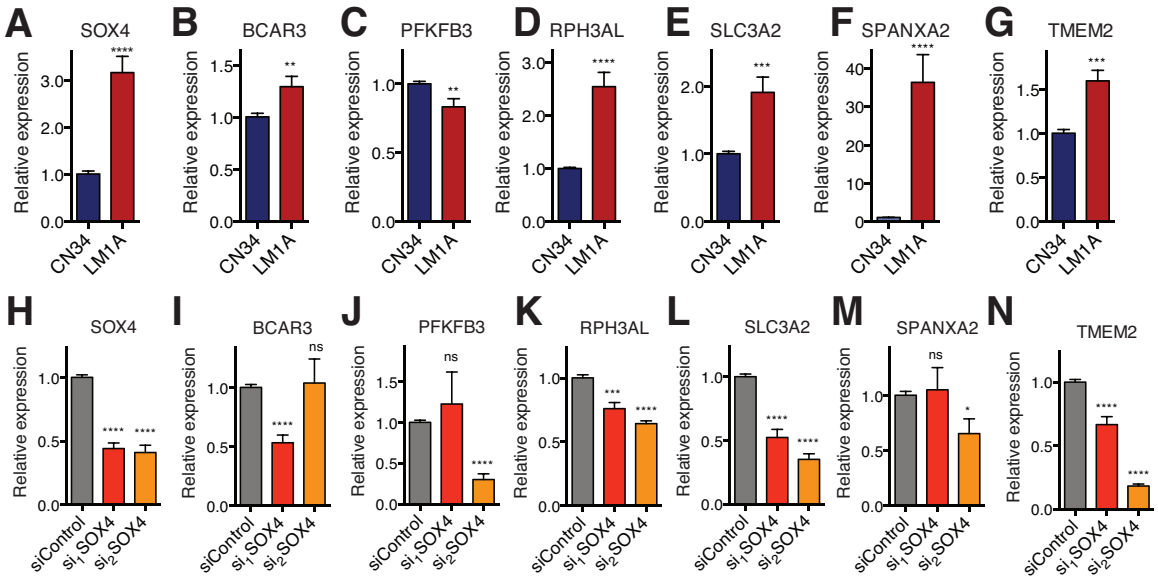


Figure 3.2 QRT-PCR validation of the 3 putative SOX4 regulated target genes in CN34/LM1A cell lines

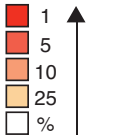
(A-G) Relative expression of the 6 genes in the CN34 parental cells as well as LM1A cells (n=3). (H-N) Relative expression of the 6 genes in LM1A cells expressing control siRNA or two siRNAs against SOX4 (n=3).

PART II: EVALUATION OF CLINICAL ASSOCIATION OF SOX4 AND TARGET GENES

We then asked if SOX4 and its putative target genes, RPH3AL, SLC3A2, and TMEM2, might be clinically relevant in breast cancer progression. To test if there is any correlation between the expression of these genes and SOX4 in breast cancer, we next investigated the expression levels of these genes in transcriptomic profiling datasets comprised of non-cancerous and cancerous breast tissue. Oncomine Cancer Microarray database showed that SLC3A2 and TMEM2 were highly ranked by overexpression in cancerous tissues when SOX4 was overexpressed (Fig. 3.3 A). Consistent with this, these two genes showed a statistically significant positive correlation with SOX4 in the Expression Project for Oncology (ExpO) microarray database (Fig. 3.3 B).

A

Gene	Median Rank	P-value	Tissue Type 1	2	3	4	5	6	7	8
SOX4	2090.0	1.79E-11								
SLC3A2	5156.5	0.001								
TMEM2	5291.5	0.031								
RPH3AL	14201.5	0.630								



1 Intraductal cribriform breast adenocarcinoma vs. Normal
2 Invasive breast carcinoma vs. Normal
3 Invasive ductal and lobular carcinoma vs. Normal
4 Invasive ductal breast carcinoma vs. Normal
5 Invasive lobular breast carcinoma vs. Normal
6 Male breast carcinoma vs. Normal
7 Mixed lobular and ductal breast carcinoma vs. Normal
8 Mucinous breast carcinoma vs. Normal

B

	ExpO data set (243 tumors)		
	Pearson correlation coefficient	p-value	critical p-value < 0.007
SLC3A2	0.212176086	0.0009	*
TMEM2	0.212599112	0.0009	*

Figure 3.3 Clinical correlation with SOX4 and the 3 putative target genes in breast cancer

(A) mRNA expression levels of SOX4 and its 3 putative target genes in either breast carcinoma or normal tissue obtained from Oncomine Cancer Microarray database. The rank for a gene is the median rank for that gene across each of the analyses. The p-value for a gene is its p-value for the median-ranked analysis. (B) The expression correlation of SOX4 with SLC3A2 or TMEM2 analyzed in the ExpO breast cancer database (www.intgen.org), which consists of a total of 243 tumors from 4 patients in stage 0, 38 patients in stage 1, 129 patients in stage 2, 67 patients in stage 3, and 5 patients in stage 4. Bonferroni correction was used to adjust confidence intervals (critical p value < 0.007).

Since SOX4 is a transcriptional activator, and a validated promoter of cancer progression, we focused on SLC3A2 and TMEM2, the genes that were positively correlated with SOX4 expression. Next, we wondered whether SOX4 was sufficient to increase mRNA levels of both genes. MDA-MB-231 cells expressing stable FLAG-SOX4 showed increased mRNA levels of SLC3A2 and TMEM2 relative to the control empty vector expressing cells (Fig. 3.4 A-C). As an independent method, we generated doxycycline inducible FLAG-SOX4 expressing MDA-MB-231 and CN34 cell lines. Interestingly in both cell lines, TMEM2 mRNA levels were increased 24 hours post doxycycline treatment whereas SLC3A2 mRNA levels were increased after 36 hours post doxycycline treatment (Fig. 3.4 D-I). The degree of increase was also different. The induction of TMEM2 mRNA levels (Fig. 3.4 F, I) was stronger than that of SLC3A2 (Fig. 3.4 E, H) by doxycycline inducible FLAG-SOX4 expression in both cell lines. These results suggest there may be differential sensitivity of target genes to SOX4. Additionally, SLC3A2 might have an indirect component to its up regulation that could explain the time lag after SOX4 induction.

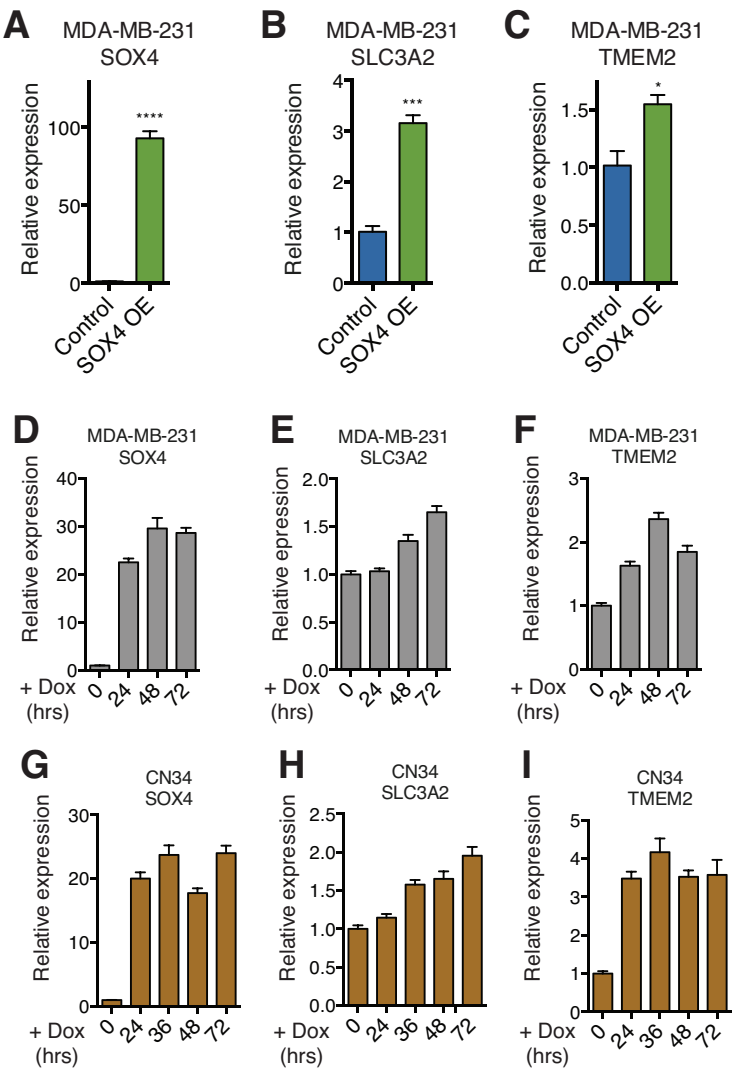


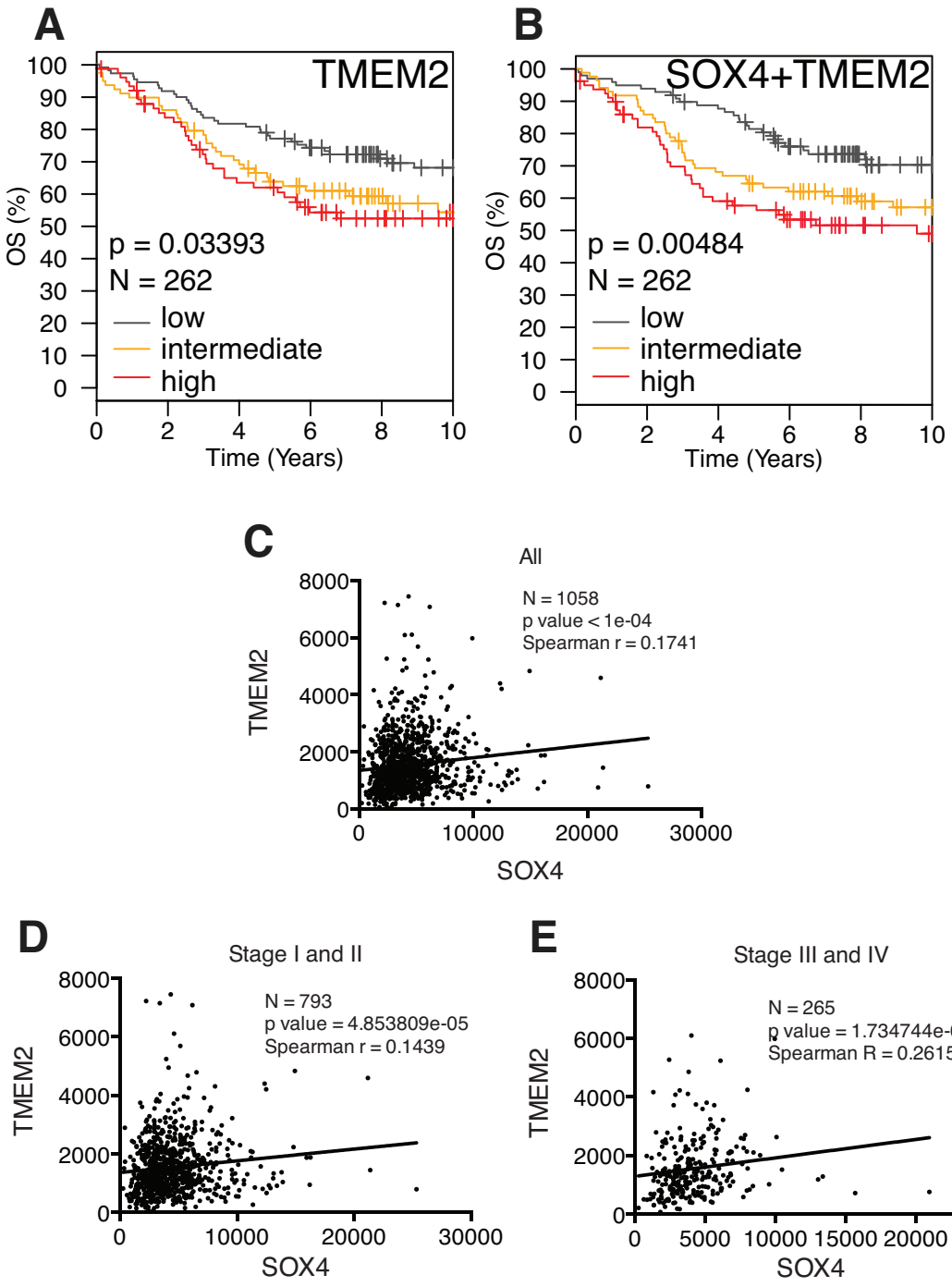
Figure 3.4 Induction of SLC3A2 and TMEM2 transcript levels by SOX4 overexpression in breast cancer cell lines

(A-C) Assessment of mRNA levels for SOX4, SLC3A2 and TMEM2 in the control or FLAG-SOX4 overexpressing MDA-MB-231 cells (n=2). (D-F) Assessment of mRNA levels for SOX4, SLC3A2 and TMEM2 in Doxycycline inducible FLAG-SOX4 overexpressing MDA-MB-231 cells (n=2). (G-I) Assessment of mRNA levels for SOX4, SLC3A2 and TMEM2 in Doxycycline inducible FLAG-SOX4 overexpressing CN34 cells (n=2).

To evaluate prognostic value of SLC3A2 and TMEM2, we examined the transcript levels of SOX4 and the target genes in a cohort of tissues of grade 3 from breast cancer patients. Patients with grade 3 breast tumors with higher SOX4 expression levels trended towards having shorter overall survival times relative to patients with grade 3 breast tumors with lower SOX4 expression. Interestingly, patients with higher levels of TMEM2 had significantly shorter overall survival times relative to patients with lower levels of TMEM2 (Fig. 3.5 A). Moreover, high expression of both SOX4 and TMEM2 most significantly correlated with poor overall survival times (Fig. 3.5 B). Consistent with this, TMEM2 was positively correlated with SOX4 in The Cancer Genome Atlas (TCGA) data set for invasive breast carcinoma (Fig. 3.5 C). In addition, patients assigned to higher stage (stage III or IV) showed 2.5 fold stronger positive correlation relative to lower stage (stage I or II) (Fig. 3.5 D-E). SLC3A2 did not show correlation in this patient data set.

Figure 3.5 Clinical relevance of TMEM2 in patients with aggressive breast tumors

(A-B) Kaplan-Meier overall survival curves demonstrating patients diagnosed with grade 3 tumors (N=262) expressing higher levels of TMEM2 (A) and both SOX4 and TMEM2 (B) to have poorer survival than those with grade 3 tumors expressing lower levels. p=0.03393 and 0.00484 respectively. Each color represents patients whose tumors' expression levels for each gene were higher (red) or lower (grey) than the median (yellow). (C-E) Clinical correlation of TMEM2 and SOX4 expression in TCGA breast cancer patient data set. SOX4 expression level is displayed on x-axis and TMEM2 on y-axis. Each data points represents individual breast cancer patients assigned from stage I to IV (C), stage I to II (D), or stage III to IV (E). TCGA Breast Invasive Carcinoma data were obtained through the cBioPortal for Cancer Genomics (www.cbioportal.org) and UCSC cancer browser (genome-cancer.ucsc.edu).


A
TMEM2
OS (%)
p = 0.03393
N = 262
low
intermediate
high
Time (Years)
B
SOX4+TMEM2
OS (%)
p = 0.00484
N = 262
low
intermediate
high
Time (Years)
C
All
N = 1058
p value < 1e-04
Spearman r = 0.1741
TMEM2
SOX4
D
Stage I and II
N = 793
p value = 4.853809e-05
Spearman r = 0.1439
TMEM2
SOX4
E
Stage III and IV
N = 265
p value = 1.734744e-05
Spearman R = 0.2615
TMEM2
SOX4

Clinical relevance of TMEM2 in breast cancer patients with grade 3 tumors and sensitive induction of mRNA levels by SOX4 overexpression suggest TMEM2 as a functional direct target of SOX4. We next wondered if mRNA levels of TMEM2 are also dependent on SOX4 in lung cancer, a cancer in which SOX4 has been reported to be overexpressed (Medina et al., 2009; Castillo et al., 2012). We depleted SOX4 in the A549 and H522 cell lines, derived from non-small cell lung cancer (NSCLC), which accounts for 85-90% of lung cancers (Haghgoo et al., 2015). TMEM2 transcripts were significantly downregulated upon SOX4 depletion in both cell lines, which is consistent with our observations in breast cancer cell lines (Fig. 3.6 A-D). Given that the A549 and H522 NSCLC cell lines tested in our study showed the expression of TMEM2 is dependent on SOX4 by RNAi, we wondered if TMEM2 is also clinically relevant in NSCLC. We used online survival analysis software (Gyorffy et al., 2010) to assess the prognostic value of TMEM2. Interestingly, high expression of TMEM2 was associated with shorter overall survival times in patients with lung adenocarcinoma, a NSCLC subtype that accounts for approximately 40% of NSCLC (American Cancer Society. Cancer Facts & Figures 2015. Atlanta: American Cancer Society; 2015.) (Fig. 3.6 E).

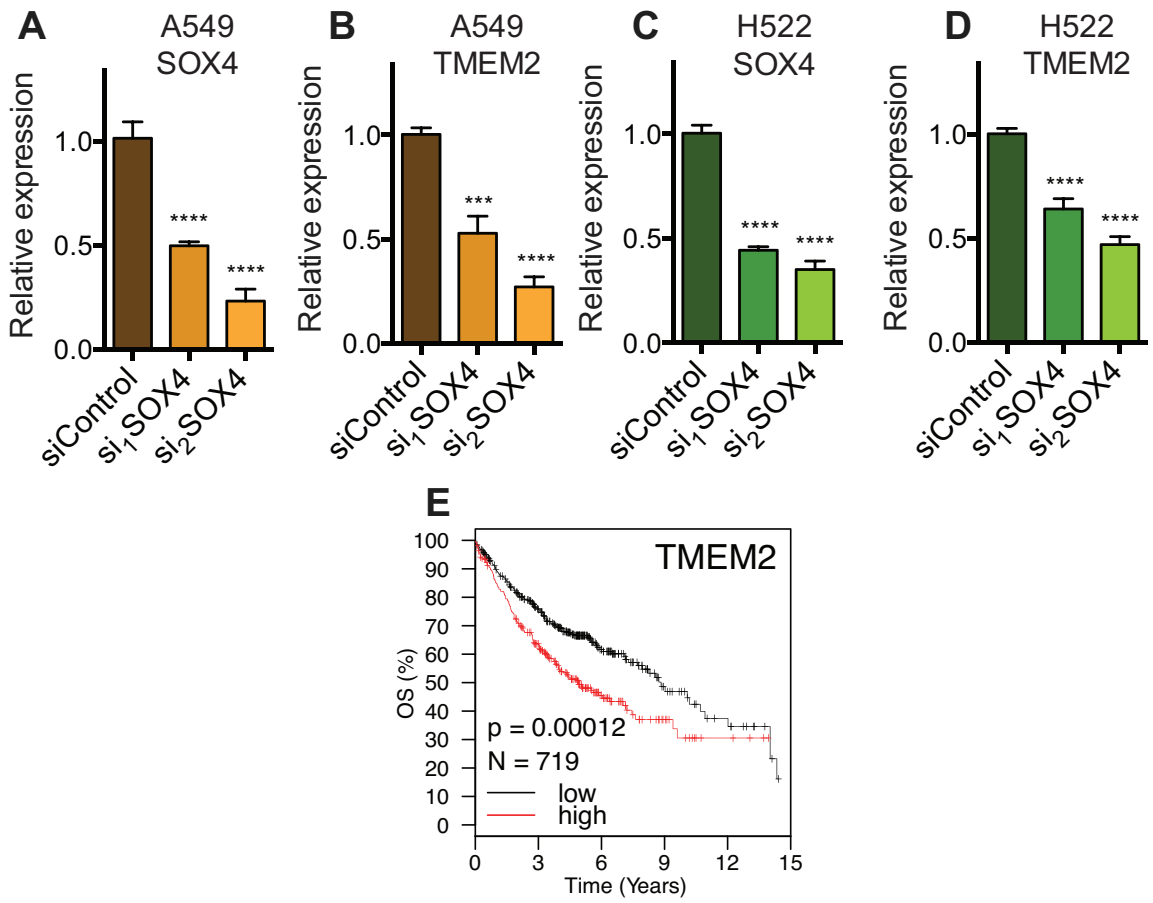


Figure 3.6 Clinical relevance of TMEM2 in patients with lung adenocarcinoma

(A-B) Relative expression levels of SOX4 (A) and TMEM2 (B) in lung adenocarcinoma A549 cells with or without SOX4 knockdown (n=2). (C-D) SOX4 (C) and TMEM2 (D) expression levels in SOX4 knockdown cells versus control in lung adenocarcinoma H522 cells (n=2). (E) Kaplan-Meier curves depicting overall survival time of patients having NSCLC adenocarcinoma histologic types (n=719) as a function of TMEM2 expression levels in tumors. The hazard ratio with 95% confidence intervals and log-rank P value for the two patient cohorts were calculated by Kaplan-Meier plotter (www.kmplot.com).

Given the lack of treatment for late stage breast cancer patients, the synergistic outcome of SOX4/TMEM2 in predicting survival of patients with advanced stage tumors is attractive. Additionally, as an extracellular transmembrane protein, TMEM2 could have advantages as a target for drug development since major accomplishments in drug discovery have resulted in the development of drugs targeting membrane-bound proteins like Tamoxifen and Herceptin. Therefore understanding the cell biological mechanism by which TMEM2 regulates metastatic progression could provide better therapeutic intervention for patients suffering from the most aggressive tumors.

PART III: TRANSCRIPTIONAL REGULATION OF TMEM2 BY SOX4

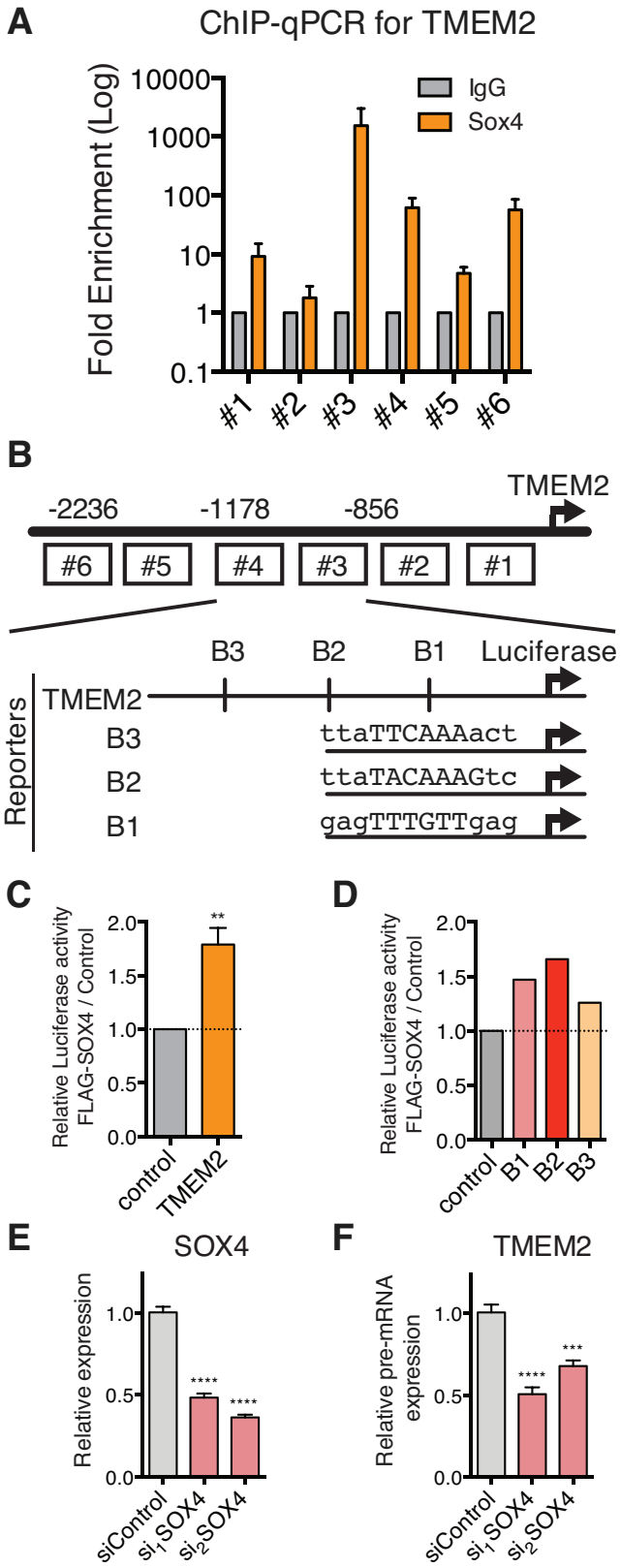
We next investigated whether TMEM2 is transcriptionally regulated by SOX4. The binding of SOX4 in the TMEM2 promoter was tested by CHIP-qPCR (Fig. 3.7 A). QPCR detection of distinct genomic regions of TMEM2 in immunoprecipitated DNA pulled down by a SOX4 antibody in metastatic LM2 cells revealed that a region spanning from -856 bp to -1178 bp was highly enriched relative to IgG control immunoprecipitated DNA (Fig. 3.7 B). We next asked which DNA sequences were sufficient to promote SOX4-dependent transcriptional induction of TMEM2. To do this, we used luciferase promoter assays. A reporter construct consisting of a luciferase reporter under the regulation of TMEM2 promoter regions were generated. Exogenously overexpressed SOX4 increased luciferase activity relative to control, consistent with a SOX4-dependent increase in transcription from the TMEM2 promoter (Fig. 3.7 C). In order to further identify the SOX4 binding motifs in this region, we searched for occurrences of the SOX4 binding motif in this region, cloned each of the three motifs found, and tested them in luciferase promoter reporter assays. The SOX4 high mobility group (HMG) box, a protein domain that mediates DNA binding, has been shown to preferentially bind the AACAAAG motif in electrophoretic mobility shift assays (EMSA) (Vandewetering et al., 1993) and TATTGTC and CATTGTG in promoter reporter constructs (Bergsland et al., 2006; Dy et al., 2008). In the context of SOX4 overexpression, TMEM2 promoter regions containing SOX4 motifs drove luciferase expression (Fig. 3.7 D). Among the three putative SOX4 binding sites,

the highest luciferase activity was driven by the B2 motif (TACAAAG), which is the closest motif to the known SOX4 binding motif (AACAAAG).

Given that pre-mRNAs are transient intermediates and their steady-state amount at any given time reflects endogenous transcription prior to nuclear processing (Lipson and Baserga, 1989; Luco and Misteli, 2011; de Almeida and Carmo-Fonseca, 2012; Montes et al., 2012; Shukla and Oberdoerffer, 2012), we analyzed the levels of intron-containing TMEM2 pre-mRNA precursors in control and SOX4 depleted cells by qRT-PCR. SOX4 knockdown caused a significant reduction of TMEM2 pre-mRNA levels relative to control cells (Fig. 3.7 E-F). These collective findings support a model whereby SOX4 transcriptionally regulates TMEM2 expression.

Figure 3.7 Identification of TMEM2 as a SOX4-regulated direct target gene

(A) ChIP-qPCR assessment with primers detecting different regions of TMEM2 promoter. Sheared DNA from LM2 cells fixed with 1% paraformaldehyde was immunoprecipitated with anti-SOX4 antibody or control IgG. (B) Schematic depicting promoter region of TMEM2 amplified by each qPCR primers used in (A). TMEM2 promoter region detected by qPCR primer #3 and #4 have three putative SOX4 binding motifs marked by B1, B2, and B3. (C) Functional validation of the ChIP-qPCR by luciferase reporter assay. TMEM2 promoter region containing all three putative SOX4 binding motifs were cloned upstream of a luciferase reporter. A relative luciferase activity of this region was increased by SOX4 overexpression relative to the control (n=3). (D) Identification of functional SOX4 binding motif in TMEM2 promoter region tested in (C) by luciferase reporter assays. Putative SOX4 binding motifs, B1, B2, and B3, were individually cloned upstream of a luciferase reporter and its effect on luciferase expression by SOX4 overexpression was measured. (E-F) pre-mRNA levels of TMEM2 in the LM2 cells expressing either siRNAs targeting SOX4 or a control siRNA were assessed by qRT-PCR with primers detecting intronic regions close to 3'UTR of TMEM2 (n=2).


A
ChIP-qPCR for TMEM2
Fold Enrichment (Log)
10000
1000
100
10
1
0.1
IgG
Sox4
#1
#2
#3
#4
#5
#6
B
TMEM2
-2236
-1178
-856
#6
#5
#4
#3
#2
#1
B3
B2
B1
Luciferase
Reporters
TMEM2
B3
ttaTTCAAAact
B2
ttaTACAAAGtc
B1
gagTTTGTTgag
C
Relative Luciferase activity
FLAG-SOX4 / Control
2.0
1.5
1.0
0.5
0.0
**
control
TMEM2
D
Relative Luciferase activity
FLAG-SOX4 / Control
2.0
1.5
1.0
0.5
0.0
control
B1
B2
B3
E
SOX4
Relative expression
1.0
0.5
0.0
****
****
siControl
si1SOX4
si2SOX4
F
TMEM2
Relative pre-mRNA expression
1.0
0.5
0.0
****
***
siControl
si1SOX4
si2SOX4

PART IV: *IN VITRO* AND *IN VIVO* FUNCTIONAL VALIDATION OF SOX4-REGULATED DIRECT TARGET GENE, TMEM2

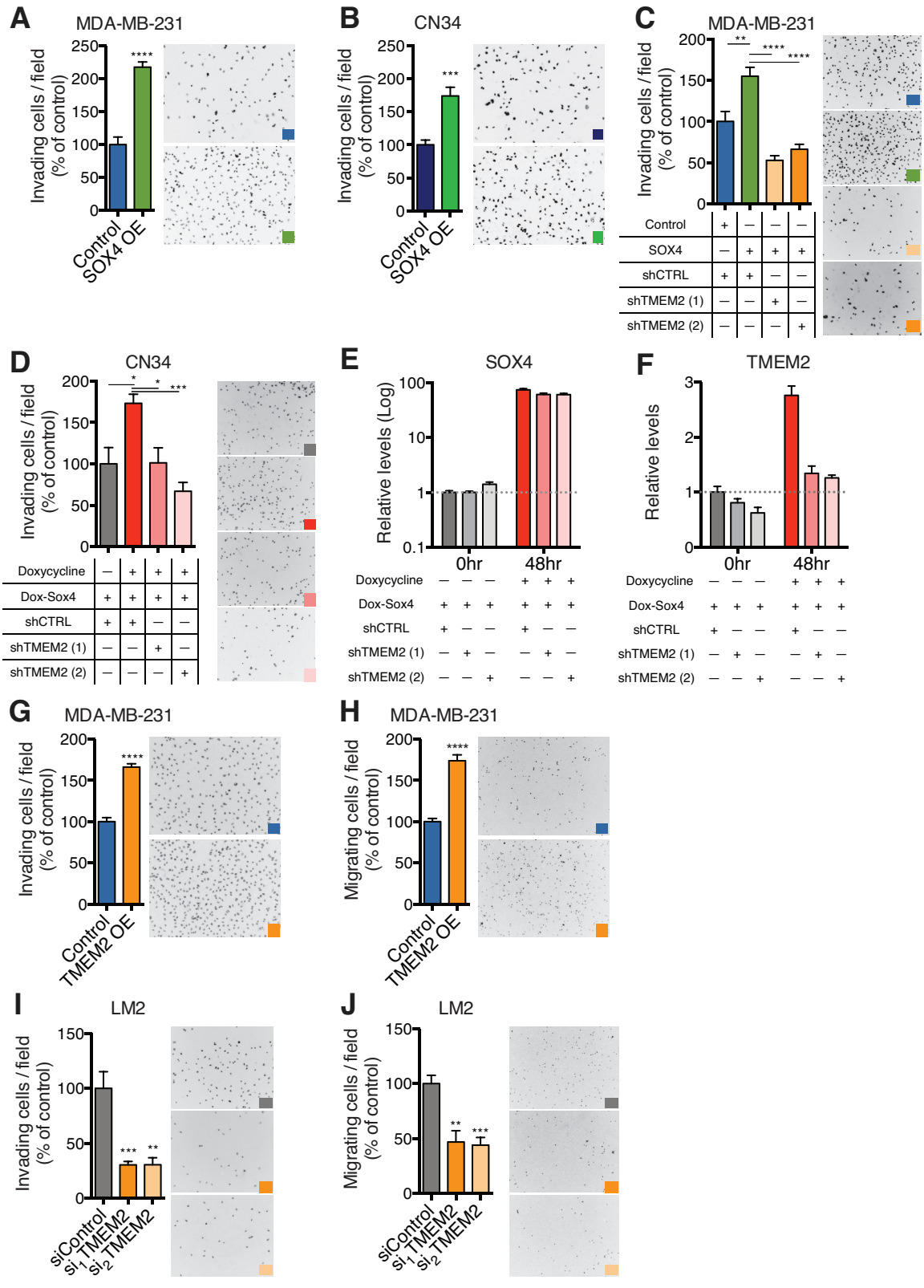
TMEM2-mediated effects of SOX4 on cell invasion and migration

Highly metastatic LM2 cells express higher levels of SOX4 relative to MDA-MB-231 cells, and the capacity of LM2 cells to invade through matrigel and to metastasize to the lungs *in vivo* was reduced upon SOX4 knockdown (Tavazoie et al., 2008). To determine if increased SOX4 expression is sufficient to enhance invasive capacity, we overexpressed SOX4 in two independent breast cancer cell lines, MDA-MB-231 and CN34, both of which express lower levels of SOX4 relative to their highly metastatic sub-lines, and performed matrigel invasion assays. Exogenous SOX4 overexpression increased matrigel invasion capacity by more than 50% in both cell lines (Fig. 3.8 A-B). In SOX4 overexpressing MDA-MB-231 cells, TMEM2 was transcriptionally induced by 1.5 fold relative to control cells expressing empty vector (Fig. 3.4 A, C). We next investigated whether the enhanced cell invasion caused by SOX4 overexpressing MDA-MB-231 cells was mediated by TMEM2. Indeed, knockdown of TMEM2 in SOX4 overexpressing cells significantly reduced (50%) the enhanced capacity of cells to invade through matrigel (Fig. 3.8 C). Consistent with this, transient overexpression of SOX4 in CN34 cells increased cellular matrigel invasion, an effect that was abrogated upon TMEM2 depletion (Fig. 3.8 D-F). This suggests that the increased invasion conferred by SOX4 overexpression is dependent, at least in part, on its downstream target gene, TMEM2. Similarly, significantly

greater invasion and migration capacity was observed in MDA-MB-231 cells overexpressing TMEM2 relative to control cells (Fig. 3.8 G-H), and the converse effect was observed in LM2 cells upon TMEM2 depletion (Fig. 3.8 I-J). These findings are consistent with a model whereby the pro-invasive effects of SOX4 are mediated, at least in part, through the transcriptional induction of TMEM2.

Figure 3.8 TMEM2 mediates the effects of SOX4 on cell invasion and migration

(A-B) Matrigel invasion by 5×10^4 MDA-MB-231 cells expressing either a control vector or FLAG-SOX4 (A, n=6) and by 1×10^5 CN34 cells expressing either control or FLAG-SOX4 (B, n=5). Images of DAPI stained cells were obtained at 10× magnifications and quantified by Image J. (C) Matrigel invasion by MDA-MB-231 cells after knockdown of TMEM2 in the context of constitutive expression of FLAG-SOX4 (n=4-5). (D) Matrigel invasion by knockdown of TMEM2 in the context of doxycycline inducible FLAG-SOX4 overexpression in CN34 cells (n=3-4). (E-F) Expression of SOX4 (E) and TMEM2 (F) in CN34 cells expressing doxycycline inducible FLAG-SOX4 48hours after doxycycline induction (n=2). (G) Enhanced matrigel invasion by MDA-MB-231 cells overexpressing TMEM2 compared to the control cells (n=4). (H) Increased trans-well migration by MDA-MB-231 cells overexpressing TMEM2 compared to the control cells (n=4). (I-J) Reduced matrigel invasion (I) and trans-well migration (J) by knockdown of TMEM2 in LM2 cells (n=4).


A
MDA-MB-231
Invading cells / field (% of control)
Control
SOX4 OE
B
CN34
C
MDA-MB-231
Control
SOX4
shCTRL
shTMEM2 (1)
shTMEM2 (2)
D
CN34
Doxycycline
Dox-Sox4
E
SOX4
Relative levels (Log)
0hr
48hr
F
TMEM2
Relative levels
G
MDA-MB-231
TMEM2 OE
H
MDA-MB-231
Migrating cells / field (% of control)
I
LM2
siControl
si1 TMEM2
si2 TMEM2
J
LM2

Promotion of in vivo metastasis by TMEM2

Based on our molecular and *in vitro* findings, we predicted that TMEM2 could have an impact on metastasis, similar to the one we observed by SOX4 (Tavazoie et al., 2008). The role of TMEM2 in metastatic colonization was tested by carrying out loss-of-function and gain-of-function studies. Depleting TMEM2 in highly metastatic LM2 cells strongly reduced metastatic colonization (Fig. 3.9 A-B) whereas TMEM2 over-expression significantly enhanced the metastatic capacity (Fig. 3.9 C-D). These *in vivo* metastatic phenotypes mirror the effects of SOX4 on metastasis and suggest that SOX4 mediates its effects on metastasis, at least in part, through transcriptional induction of TMEM2.

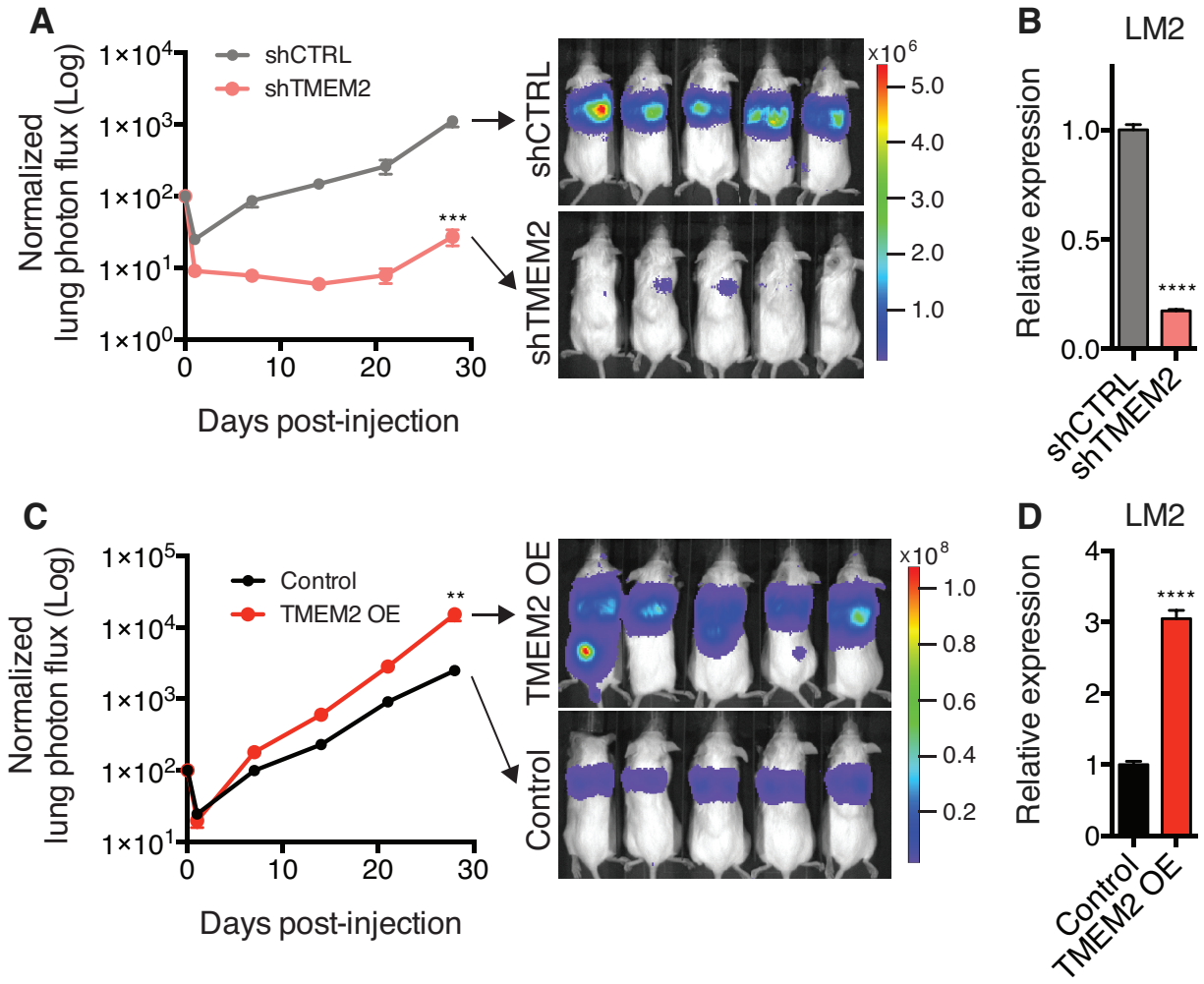


Figure 3.9 TMEM2 promotes *in vivo* metastatic lung colonization

(A) Bioluminescence imaging plot of lung metastasis by 4 × 10^4 LM2 cells expressing control shRNA or shRNA targeting TMEM2 (n=5). NOD-SCID gamma female mice were imaged every seven days and bioluminescence signals were normalized to the signal on day 0. (B) Knockdown efficiency of TMEM2 measured by qRT-PCR in LM2 cells injected into the tail-vein of mice. (C-D) Bioluminescence imaging plot of lung metastasis by LM2 cells expressing control or TMEM2 (C, n=5) and mRNA levels of TMEM2 measured by qRT-PCR in LM2 cells injected into the tail-vein of NOD-SCID gamma female mice (D).

CHAPTER 4

DISCUSSION

Summary

By using an unbiased, genomic and molecular approach, we have identified TMEM2 as a direct transcriptional target of SOX4 in multiple cancer types. We find that SOX4 directly binds to the TMEM2 promoter and transcriptionally activates it in a sequence-specific manner. Molecular studies reveal that SOX4 regulates endogenous TMEM2 pre-mRNA levels—consistent with a transcriptional regulatory role. TMEM2 promotes cancer cell migration, invasion, and metastatic colonization—phenocopying SOX4-dependent effects. Moreover, we find that TMEM2—a gene that has not been previously implicated in cancer progression—positively associates with breast cancer and lung cancer clinical progression.

The role of SOX4 in cancer progression

SOX4 was originally identified as a transcriptional activator in B and T lymphocytes (Schilham et al., 1996). Consistent with its expression in lymphocytes during development, SOX4 deficiency was found to result in abnormal thymocyte differentiation (Schilham et al., 1997). Further characterization showed that SOX4 belongs to the HMG box superfamily of DNA-binding proteins and contains a single HMG domain that allows it to bind DNA in a sequence specific manner (Vandewetering et al., 1993; Bergsland et al., 2006; Dy et al., 2008; Jolma et al., 2013). In addition to its role in lymphocytes, genetic studies in knockout mice revealed that SOX4 is required for normal

cardiac development. Loss of SOX4 activity was found to result in impaired maturation of endocardial ridges into semilunar valves as well as the ventricular septum outlet (Schilham et al., 1996; Ya et al., 1998; Buckingham et al., 2005). This phenotype mirrored what is seen in humans as the 'common arterial trunk' phenotype. SOX4 was subsequently found to be over-expressed in various cancers (Frierson et al., 2002; Sinner et al., 2007; Huang et al., 2012) and amplified in lung cancer (Medina et al., 2009).

The identification of metastasis suppressor miRNAs in breast cancer led to the identification of a cancer progression role for SOX4—as a driver of cancer cell invasion and migration. SOX4 was found to be targeted by the metastasis suppressor miRNA miR-335. As a result of miR-335 silencing in metastatic breast cancer, SOX4 expression is enhanced (Tavazoie et al., 2008). The enhancement of invasive and migratory phenotypes by SOX4 resulted in augmented metastatic lung colonization capacity of breast cancer cells. A number of additional studies have revealed that SOX4 can promote cancer progression in a number of additional cancer types such as prostate cancer (Liu et al., 2006; Scharer et al., 2009; Lai et al., 2011), hepatocellular carcinoma (Liao et al., 2008), glioblastoma (Ikushima et al., 2009; Lin et al., 2010a), and leukemia (Zhang et al., 2013a; Ramezani-Rad et al., 2013). Although SOX4 has been found to promote cancer cell migration and invasion in multiple cancer types, direct target(s) of SOX4 that mediate such effects in multiple cancer types have yet to be identified (Vervoort et al., 2013b).

TMEM2

TMEM2 is a poorly described transmembrane protein. It was identified to be associated with autosomal recessive non-syndromic hearing loss. However, no genetic mutations in TMEM2 have been identified that associate with hearing loss. A type II transmembrane protein, TMEM2 is composed of a N-terminal transmembrane domain, a G8 domain and a GG domain. Both G8 and GG domains are within extracellular sequences. The G8 domain is named for its eight conserved glycine residues, which may be important for correct folding of the β-strand/helix structure (He et al., 2006; Guo et al., 2006). The GG domain consists of two conserved glycines. To date, no biochemical functions have been identified for these domains (Guo et al., 2006) (Fig. 4.1 A). The structure of TMEM2 is highly conserved in vertebrates.

Recent studies have revealed a role for TMEM2 in heart development of zebrafish (Smith et al., 2011; Totong et al., 2011). Two independent groups have identified two zebrafish mutants, named wickham (wkm) and frozen ventricle (frv). Wkm displayed a specific loss of cardiac looping at later developmental stages and frv exhibited aberrant atrioventricular canal differentiation. Both mutations occur in the TMEM2 locus. Wild type TMEM2 has 1378 amino acids whereas the frv mutant expresses a truncated (Fig. 4.1 B-C). TMEM2 that does not have amino acids from 876 to the end results in a shortened extracellular domain. The above study suggests that a part of the missing extracellular domain is important for its function during cardiac development. These two groups have also shown

that TMEM2 can function in the endocardium to suppress atrioventricular differentiation within the ventricle by inhibiting bone morphogenetic protein 4 (Bmp4). During cardiac development, BMP2 and BMP4 are involved in the lengthening of the outflow tract, proper positioning of the outflow vessels, and septation of the atria, ventricle and atrioventricular canal (Goldman et al., 2009). The phenotype resulting from loss of BMP2 and BMP4 signaling resembles the common arterial trunk phenotype of the SOX4 null mouse. In the myocardium, TMEM2 promotes the migration of myocardial and endocardial cells. It is interesting that TMEM2 functions in heart development in part by regulating cell migration, which resembles the role of SOX4 in both mouse heart development and breast cancer metastasis.

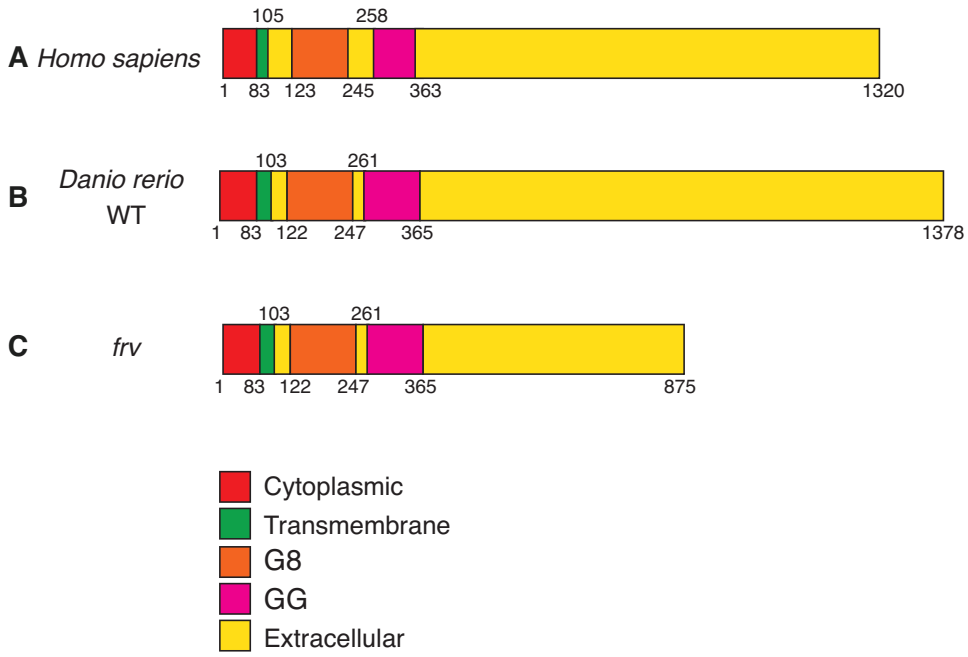


Figure 4.1 Graphical view of domain structure of TMEM2

(A) *Homo sapiens* TMEM2 (B-C) *Danio rerio* TMEM2 of wild type (B) and *frv* (C)

Given the above evidences that all three genes have roles in heart development, and that SOX4 is implicated in cancer metastasis, it is possible that the SOX4, BMP2/4 and TMEM2 axis could represent a common pathway that is important in both heart development and cancer progression. Interestingly, in the context of cancer, recent studies have revealed that SOX4 and Bmp4 have important roles in glioblastoma; Ikushima et al. showed that TGFβ activates SOX4, which then leads to elevated SOX2 expression, which is responsible for maintaining the glioma-initiating stem cell population (Ikushima et al., 2009). Piccirillo et al. showed that another member of the TGFβ family, Bmp4, reduced the glioma-initiating stem cell population (Piccirillo et al., 2006). In breast cancer, SOX4 was previously shown to directly regulate an EMT regulator, EZH2, in a TGFβ-dependent manner (Ikushima et al., 2009) and Bmp4 has been reported to inhibit metastasis by blocking myeloid-derived suppressor cell activity (Cao et al., 2014). Based on the above evidences, SOX4, TMEM2, and members of the BMP family could play dual roles in cardiac development and metastasis.

From the results presented in the current thesis, TMEM2 could possibly regulate cell migration and invasion by engaging with BMP signaling pathways. It would be interesting to investigate if TMEM2 could interact with either the extracellular signaling ligands BMP or BMP-specific transmembrane receptors expressed in breast cancer cells in a positive feedback regulation of the pathway. Alternatively and equally interesting, would be to determine the molecular mechanism by which TMEM2 induces cell migration and invasion phenotypes.

This could be performed through co-immunoprecipitation experiments and mass spectrometry to identify interacting partners of TMEM2.

While we have identified TMEM2 as one transcriptional target of SOX4 (Fig. 4.2), our findings do not exclude the likely possibility that there exist additional direct transcriptional targets that mediate the SOX4-dependent pro-invasive and pro-migratory effects. One open question is what are the co-factors required for the transcriptional activation of TMEM2? Are those co-factors predictors of metastasis and what is their tissue specificity? Additionally, it will be important to study the transcriptional landscape of SOX4 during development in specific tissues such as the heart and lymphocytes in order to establish the breadth of its transcriptional program during development and try to understand the mechanisms by which cancer cells appropriate those developmental programs to metastasize.

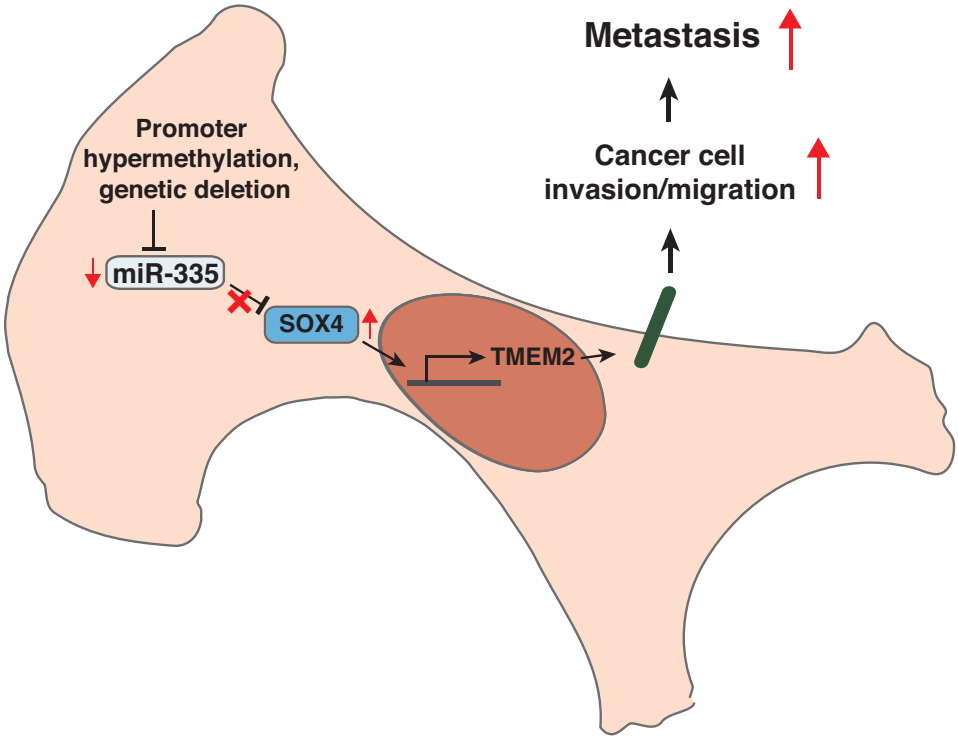


Figure 4.2 A SOX4 transcriptional network that regulates breast cancer metastasis through transcriptional activation of TMEM2

TMEM2 was identified as a gene that regulates myocardial and endocardial morphogenesis and its genetic inactivation was found to result in abnormal cardiac development. While the relationship between SOX4 and TMEM2 was previously unknown, the regulation of TMEM2 by SOX4 in cancer cells, the similarity of the SOX4 and TMEM2 phenotypes in cancer, and the common phenotypes resulting from the genetic inactivation of these genes, suggest that SOX4 and TMEM2 may form a pathway that governs normal cardiovascular development. Future studies are needed to determine if SOX4 regulates TMEM2 expression in the developing heart and, if so, whether enforced TMEM2 expression could rescue in part the SOX4 loss-of-function phenotype.

CHAPTER 5

N^6-METHYLADENOSINE MARKS PRIMARY MICRORNAS FOR PROCESSING

Claudio R. Alarcón, Hyeseung Lee, Hani Goodarzi, Nils Halberg and
Sohail F. Tavazoie, *Nature*, 2015

PART I: DISCOVERY OF THE ROLE OF METHYLTRANSFERASE-LIKE 3 (METTL3) IN MICRORNA PROCESSING

Summary

The first step in the biogenesis of miRNAs is the processing of primary miRNAs (pri-miRNAs) by the microprocessor complex, composed of the RNA binding protein DGCR8 and the ribonuclease type III DROSHA (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). This initial event requires the recognition of the junction between the stem and the flanking single-stranded RNA of the pri-miRNA hairpin by DGCR8 followed by recruitment of DROSHA, which cleaves the RNA duplex to yield the pre-miRNA product (Han et al., 2006). While the mechanisms underlying pri-miRNA processing have been elucidated, the mechanism by which DGCR8 recognizes and binds pri-miRNAs as opposed to other secondary structures present in transcripts is not understood.

We find that methyltransferase like 3 (METTL3) methylates pri-miRNAs, marking them for recognition and processing by DGCR8. Consistent with this, METTL3 depletion reduced the binding of DGCR8 to pri-miRNAs and resulted in the global reduction of mature miRNAs and concomitant accumulation of unprocessed pri-miRNAs. *In vitro* processing reactions confirmed the sufficiency of the m6A mark in promoting pri-miRNA processing.

Finally, gain-of-function experiments revealed that METTL3 is sufficient to enhance miRNA maturation in a global and non-cell-type-specific manner. Our findings reveal that the m6A mark acts as a key post-transcriptional modification that drives the initiation of miRNA biogenesis.

Identification of the METTL3 recognition motif, GGAC, in pri-miRNA sequences

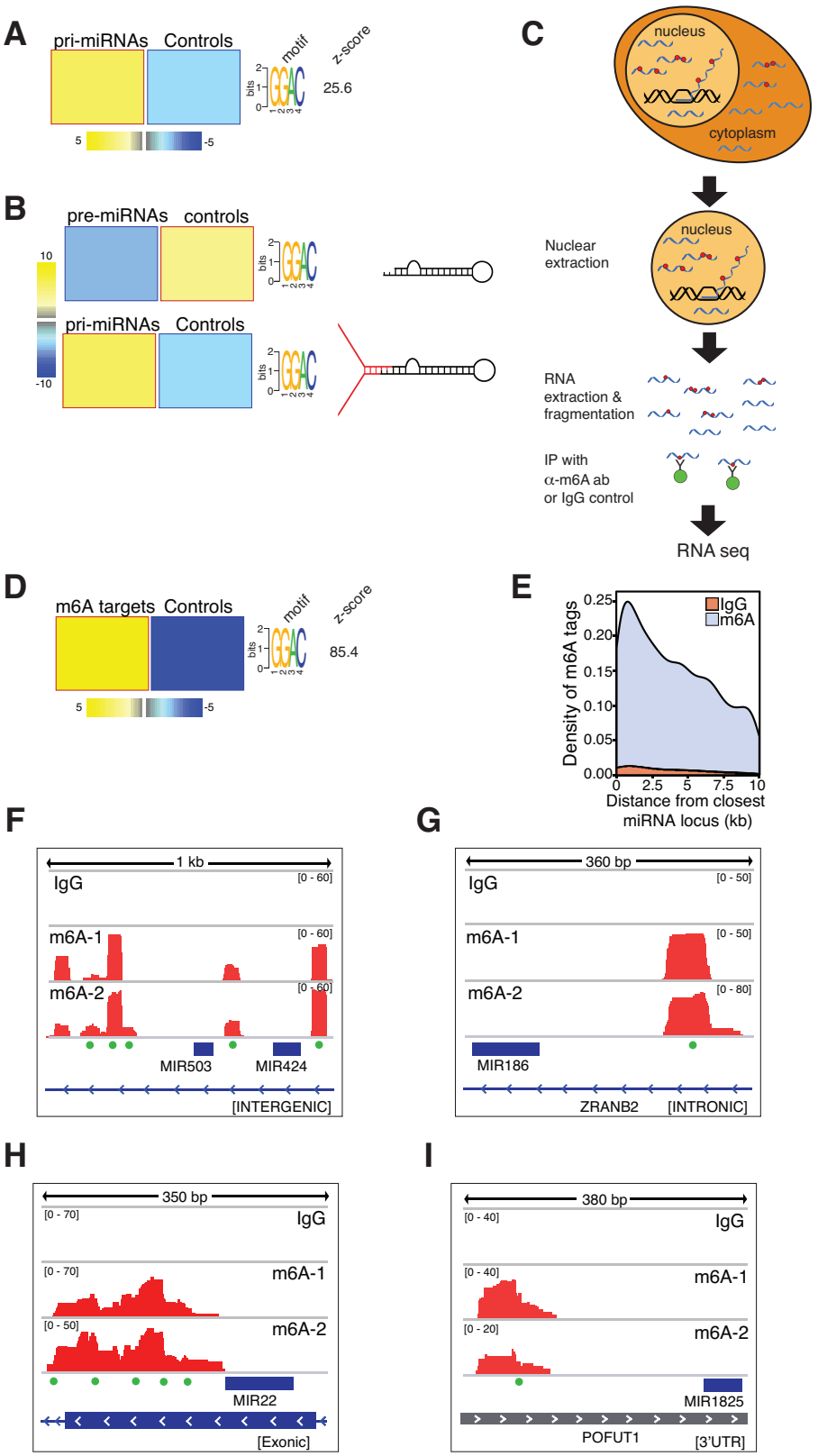
In our search for post-transcriptional modifications that regulate miRNA processing, we conducted a systematic search for sequence motifs that are over-represented in miRNA-containing regions using the FIRE algorithm, a well-established computational framework for the unbiased identification of cis-regulatory elements present in primary sequences (Elemento et al., 2007). We observed the over-representation of the GGAC motif in pri-miRNA sequences relative to shuffled sequences (Fig. 5.1 A). This motif is consistent with a previously established recognition sequence ([GAU][GAU]AC) for the RNA methyltransferase enzyme METTL3 (Wei and Moss, 1977; Dominissini et al., 2012; Meyer et al., 2012). In contrast to pri-miRNA sequences, this element was not enriched in pre-miRNA sequences, and was actually depleted relative to shuffled sequences (Fig. 5.1 B). METTL3 is the catalytic subunit of a multi-component enzyme that methylates RNA, thereby adding the N^6-methyladenosine (m6A) mark to eukaryotic RNAs (Bokar et al., 1997; Sibbritt et al., 2013). The biological function of the m6A modification in eukaryotes has just recently begun to be investigated, and includes RNA splicing (Dominissini et

al., 2012), mRNA stability (Wang et al., 2014b; c), and circadian clock speed control (Fustin et al., 2013).

To determine if the over-representation of the m6A methylation motif in pri-miRNA sequences signifies increased m6A methylated sequences, we conducted m6A-seq (Dominissini et al., 2012), by immunoprecipitating nuclear RNA from the MDA-MB-231 breast cancer cell line with an anti-m6A antibody followed by RNA seq (Fig. 5.1 C). An unbiased search for cis-regulatory elements from m6A-seq revealed a significant enrichment of the METTL3 motif relative to shuffled sequences (Fig. 5.1 D). Furthermore, when we analyzed the density of the peaks in the vicinity of miRNA loci, we found a substantial increase in the density of peaks proximal to pre-miRNA sequences, corresponding to pri-miRNA regions (Fig. 5.1 E). We next inspected individual clusters of reads using the Integrative Genomics Viewer (IGV) software (Thorvaldsdottir et al., 2013) and found numerous cases in which there were significant peaks in locations that correspond to pri-miRNAs. These clusters were located in both intergenic and intragenic pri-miRNA sites that contained canonical METTL3 motifs (Fig. 5.1 F). Thus, these results reveal that the m6A modification is enriched within pri-miRNA sequences.

Figure 5.1 m6A mark is present in pri-miRNAs regions

(A) Motif discovery analysis in pri-miRNA sequences using FIRE, reveals over-representation of the METTL3 motif, where yellow represents over-representation and blue under-representation. The magnitude of the over/under-representation is reflected in the linear scale heat-map shown at the bottom. The z-score is specified on the right. (B) Unbiased search for cis-regulatory elements using the FIRE algorithm. FIRE motif discovery analysis of pre-miRNAs and pri-miRNA sequences as well as random sequences of the same length reveals over-representation of the METTL3 motif in pri-miRNAs sequences but not in pre-miRNAs. Yellow represents over-representation, and blue under-representation. The magnitude of the over/under-representation is represented by the linear scale heat-map on the left. A schematic representation of a pre and a pri-miRNA is shown on the right. (C) Schematic representation of the m6A-seq protocol. (D) FIRE analysis of the m6A peaks compared to controls sequences of the same length. The over-represented motif and their z-score are depicted on the right as in (A). (E) Density plot of the abundance of the m6A marks and their proximity to miRNA within transcripts. Peaks obtained from the IgG immunoprecipitation were used as controls. (F-I) IGV tracks displaying examples of sequencing read clusters from two m6A-seq replicates are shown next to the pre-miRNA genomic loci. The green dots at the bottom of the tracks depict the position of METTL3 consensus motifs. The upper most tracks at the top represent the reads from the IgG immunoprecipitated control sample.

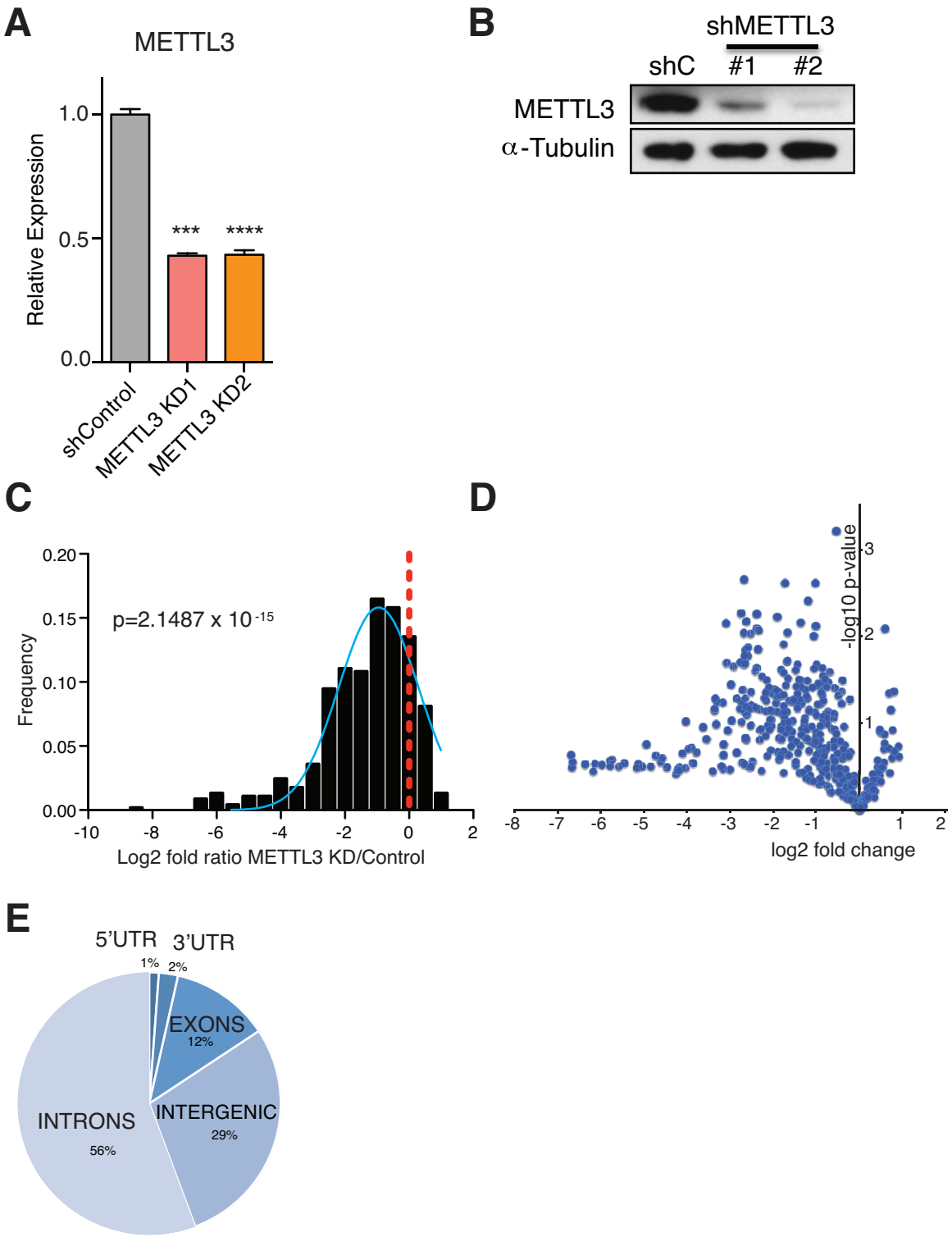

A
pri-miRNAs
Controls
motif
z-score
25.6
B
pre-miRNAs
controls
pri-miRNAs
Controls
C
nucleus
cytoplasm
Nuclear extraction
RNA extraction & fragmentation
IP with α-m6A ab or IgG control
RNA seq
D
m6A targets
Controls
85.4
E
IgG
m6A
Density of m6A tags
Distance from closest miRNA locus (kb)
F
1 kb
IgG
m6A-1
m6A-2
MIR503
MIR424
[INTERGENIC]
G
360 bp
MIR186
ZRANB2
[INTRONIC]
H
350 bp
MIR22
[Exonic]
I
380 bp
MIR1825
POFUT1
[3'UTR]

Role of METTL3 in pri-miRNA processing

To determine if METTL3 plays a role in miRNA processing, we conducted genome-wide miRNA expression profiling of MDA-MB-231 cells expressing a control shRNA as well as MDA-MB-231 cells expressing two independent shRNAs targeting METTL3 (Fig. 5.2 A-B). METTL3 depletion using independent shRNAs led to a global downregulation of mature miRNAs ($p=2.15 \times 10^{-15}$; Fig. 5.2 C-D). Remarkably, most miRNAs (~70%) were downregulated by at least 30%. The miRNAs affected by METTL3 depletion were diverse in terms of their genomic locations, with the most abundant being intronic (56%), followed by intergenic (29%), exonic (12%) and finally those located in untranslated regions (UTRs; 3%), indicating that the effect of METTL3 depletion on miRNA processing was not specific to miRNAs originating from specific locations within or outside of genes (Fig. 5.2 E)

Figure 5.2 METTL3 modulates the expression levels of miRNAs

(A-B) qRT-PCR (A) and western blot (B) quantifications of METTL3 depletion upon transduction with two independent shRNAs against METTL3 (METTL3 KD1 and METTL3 KD2) in MDA-MB-231 cells. Samples were normalized to GAPDH. The data from biological triplicates is shown. Bar graphs represent a linear scale and error bars represent s.d. ***, p-value <0.0005. (C) Histogram depicting fold change (log2) in miRNA expression, as obtained by genome-wide miRNA expression profiling, of MDA-MB-231 cells expressing two independent shRNAs targeting METTL3 compared to two samples expressing control shRNAs. The ratio of the average of the two independent shRNAs over the average of the two controls is shown. The p-value of the analysis is indicated. (D) A volcano plot representation of the microarray of miRNAs shown in (C), where the y-axis represents the -log10 of the p-value, and the x-axis represents the fold change (log2) between the expression levels of the miRNA from the METTL3 depletion (average of two independent shRNAs) versus the average of two control samples. (E) Pie-chart representation of the genomic locations of miRNAs downregulated upon METTL3 depletion.


A
METTL3
Relative Expression
1.0
0.5
0.0
***
****
shControl
METTL3 KD1
METTL3 KD2
B
shMETTL3
shC
#1
#2
METTL3
α-Tubulin
C
Frequency
0.20
0.15
0.10
0.05
0.00
p=2.1487 x 10-15
-10
-8
-6
-4
-2
0
2
Log2 fold ratio METTL3 KD/Control
D
-log10 p-value
3
2
1
-8
-7
-6
-5
-4
-3
-2
-1
1
2
log2 fold change
E
5'UTR
1%
3'UTR
2%
EXONS
12%
INTERGENIC
29%
INTRONS
56%

We next validated specific miRNAs by qRT-PCR (Fig. 5.3). Importantly, the effect of METTL3 depletion on miRNA processing was not restricted to a particular cell line and was also observed in HeLa cells and primary human umbilical vein endothelial cells (Fig. 5.4). These experiments demonstrate that METTL3 is required for the basal expression of the vast majority of miRNAs in cancerous and non-cancerous cells.

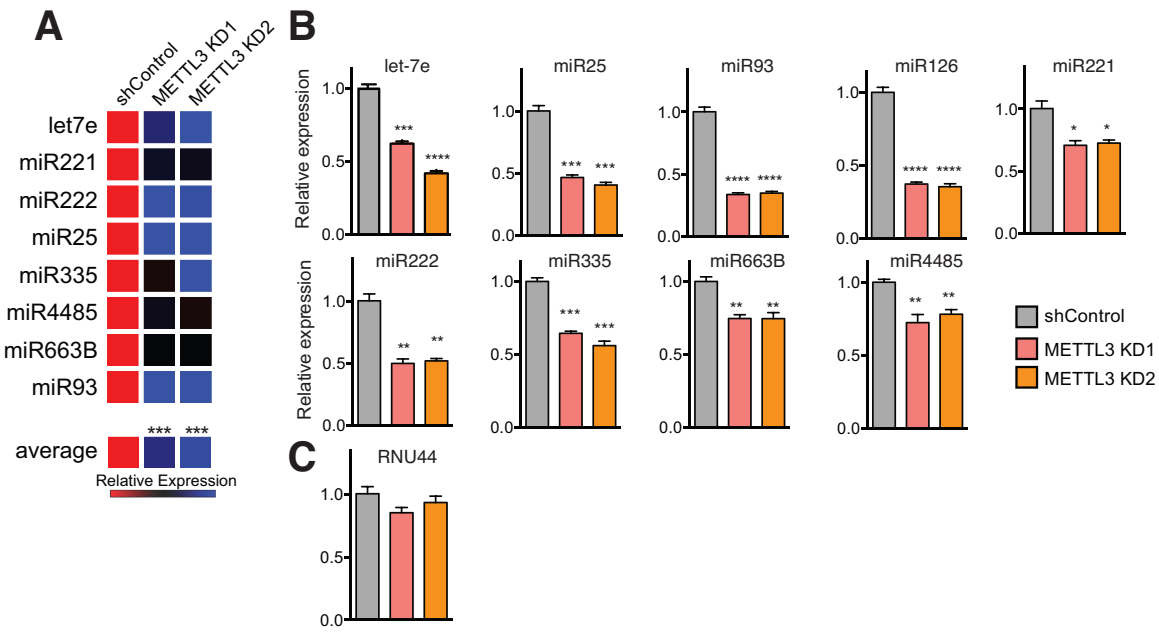
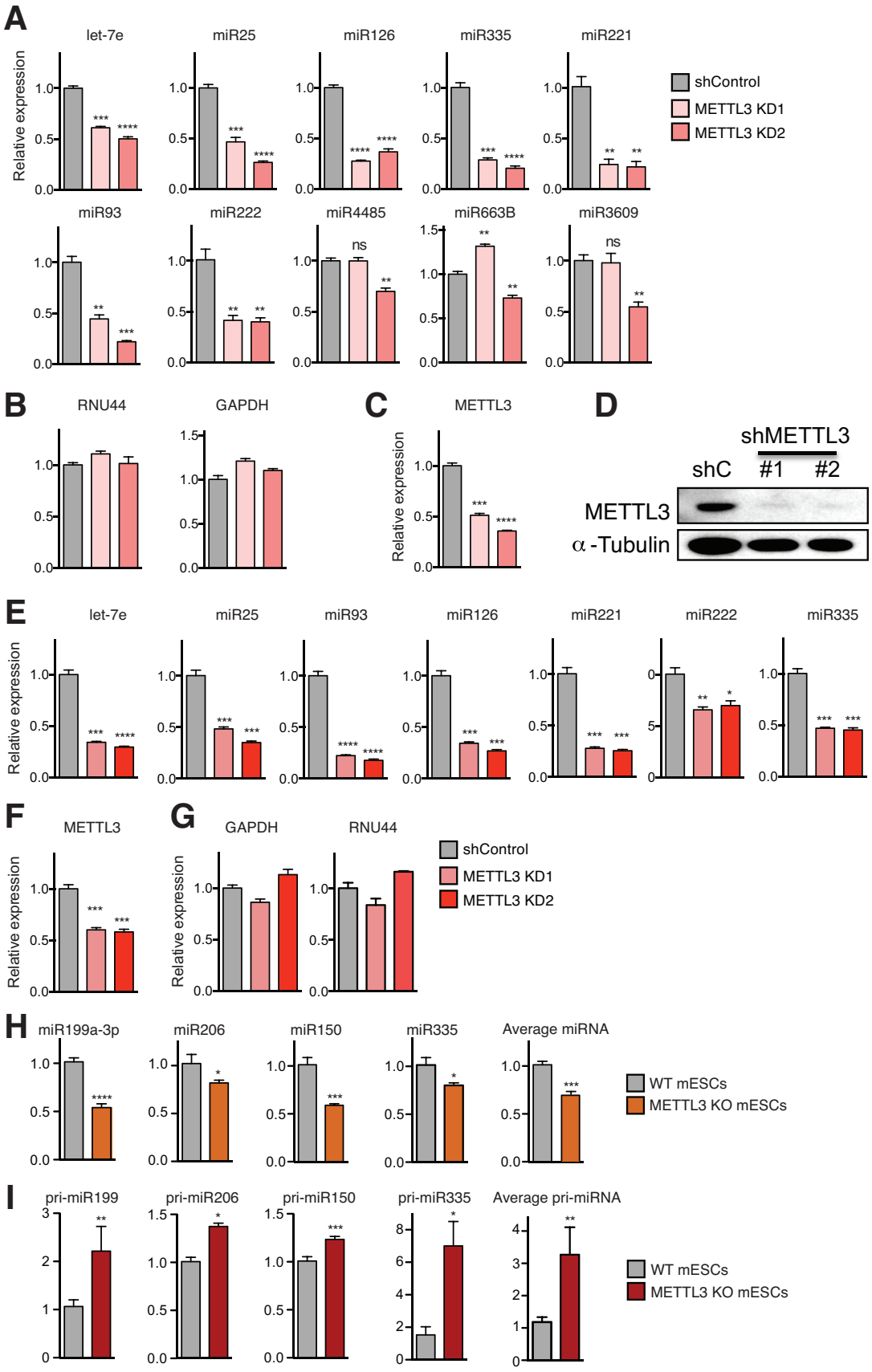


Figure 5.3 Mature miRNAs are downregulated upon METTL3 depletion in MDA-MB-231 cells

(A) Heat-map representation of the quantification by qRT-PCR of eight representative mature miRNAs that were affected by METTL3 depletion. Red represents increased expression while blue represents reduced expression. At the bottom, a heat map depicts their aggregate expression upon METTL3 modulation. (B) Quantification of representative miRNAs that were affected by METTL3 depletion in MDA-MB-231 cells as measured by qRT-PCR. Samples were normalized to SNORD44 (RNU44). (C) An example of a small RNA that did not display expression level changes upon METTL3 knockdown (SNORD44, small nucleolar RNA) normalized to 18S. All the experiments were done in biological replicates. Bar graphs represent a linear scale and error bars represent s.d. **** $P < 1\times10^{-4}$, *** $P < 5\times10^{-4}$, ** $P < 1\times10^{-3}$, * $P < 5\times10^{-2}$.

Figure 5.4 Mature miRNAs are downregulated upon METTL3 depletion in multiple mammalian cell lines

(A) qRT-PCR quantification of examples of miRNAs that were modulated by METTL3 depletion in HeLa cells. Samples were normalized to RNU44. (B) Expression levels of genes used for normalization. All experiments were done in biological replicates. (C-D) qRT-PCR (C) and western blot (D) quantifications of METTL3 depletion upon transduction with two independent shRNAs targeting METTL3. (E) Expression levels of representative miRNAs that were affected by METTL3 depletion in HUVEC cells, as measured by qRT-PCR. Normalization of the samples was obtained using RNU44 as endogenous control. (F) qRT-PCR quantification of METTL3 depletion upon transduction with two independent shRNAs against METTL3. (G) Quantification of the expression levels of control genes. (H-I) Examples of miRNAs affected in mouse embryonic stem cells in which Mettl3 has been targeted using CRISPR, whose expression levels were measured by qRT-PCR. All experiments were done in biological replicates. Bar graphs represent a linear scale and error bars represent s.d. **** $P < 1\times10^{-4}$, *** $P < 5\times10^{-4}$, ** $P < 1\times10^{-3}$, * $P < 5\times10^{-2}$.


A
let-7e
miR25
miR126
miR335
miR221
miR93
miR222
miR4485
miR663B
miR3609
Relative expression
shControl
METTL3 KD1
METTL3 KD2
B
RNU44
GAPDH
C
METTL3
Relative expression
D
shMETTL3
shC
#1
#2
METTL3
α-Tubulin
E
let-7e
miR25
miR93
miR126
miR221
miR222
miR335
Relative expression
F
METTL3
Relative expression
G
GAPDH
RNU44
Relative expression
shControl
METTL3 KD1
METTL3 KD2
H
miR199a-3p
miR206
miR150
miR335
Average miRNA
WT mESCs
METTL3 KO mESCs
I
pri-miR199
pri-miR206
pri-miR150
pri-miR335
Average pri-miRNA
WT mESCs
METTL3 KO mESCs

To test if METTL3 expression is sufficient to promote miRNA processing, we stably overexpressed METTL3 in MDA-MB-231 cells. METTL3 overexpression was sufficient to significantly increase the expression levels of mature miRNAs (Fig. 5.5). We then quantified the expression levels of pri-miRNAs to determine if METTL3 depletion could impact pri-miRNA levels. As expected, we observed a clear and significant up-regulation in the levels of pri-miRNAs upon METTL3 depletion (Fig. 5.6 A-C). These effects required enzymatically active METTL3 since they were abrogated upon mutation of the predicted catalytic residues (Bujnicki et al., 2002) (Fig. 5.6 D-E). Neither METTL3 depletion nor its overexpression altered the expression, sub-cellular localization, or activity of the Microprocessor—consistent with a direct role of METTL3 in pri-miRNA processing (Fig. 5.7). Our findings reveal that pri-miRNAs are marked by the METTL3 m6A modification and that METTL3 expression is required for the appropriate processing of most pri-miRNAs to mature miRNAs.

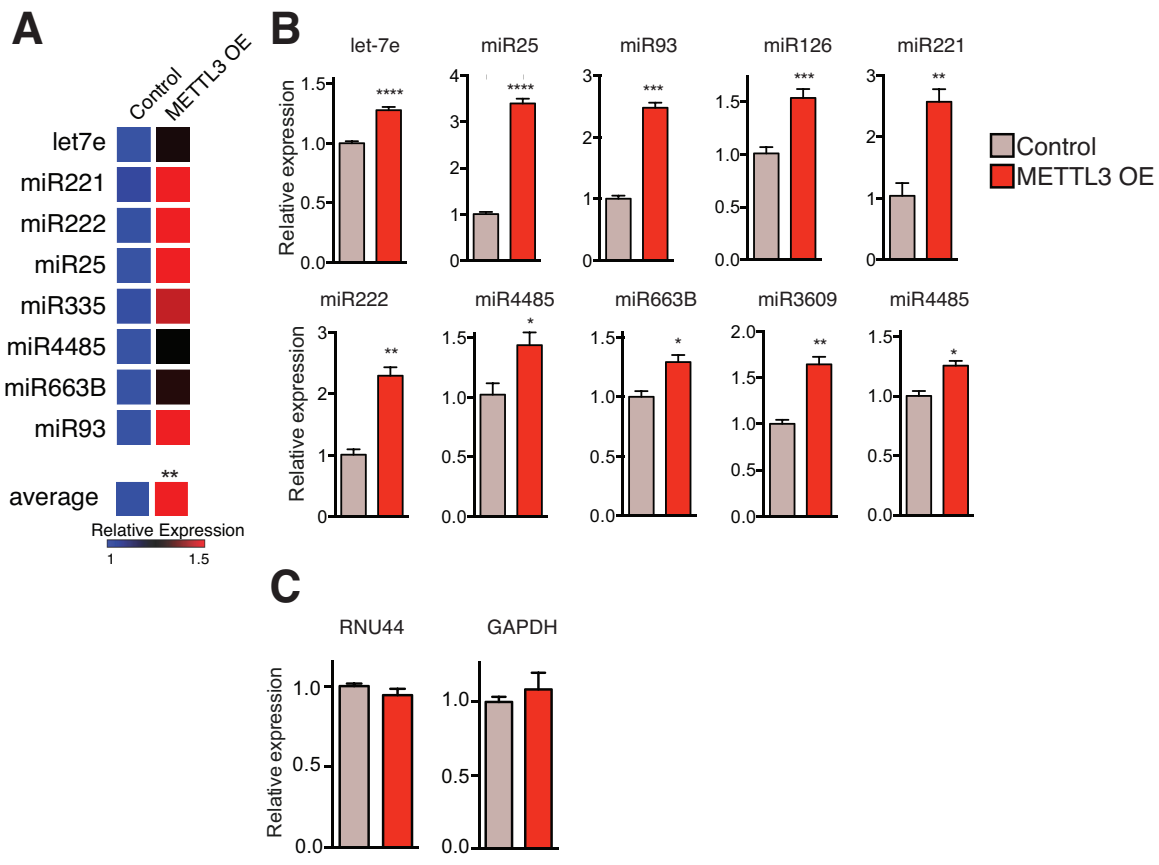
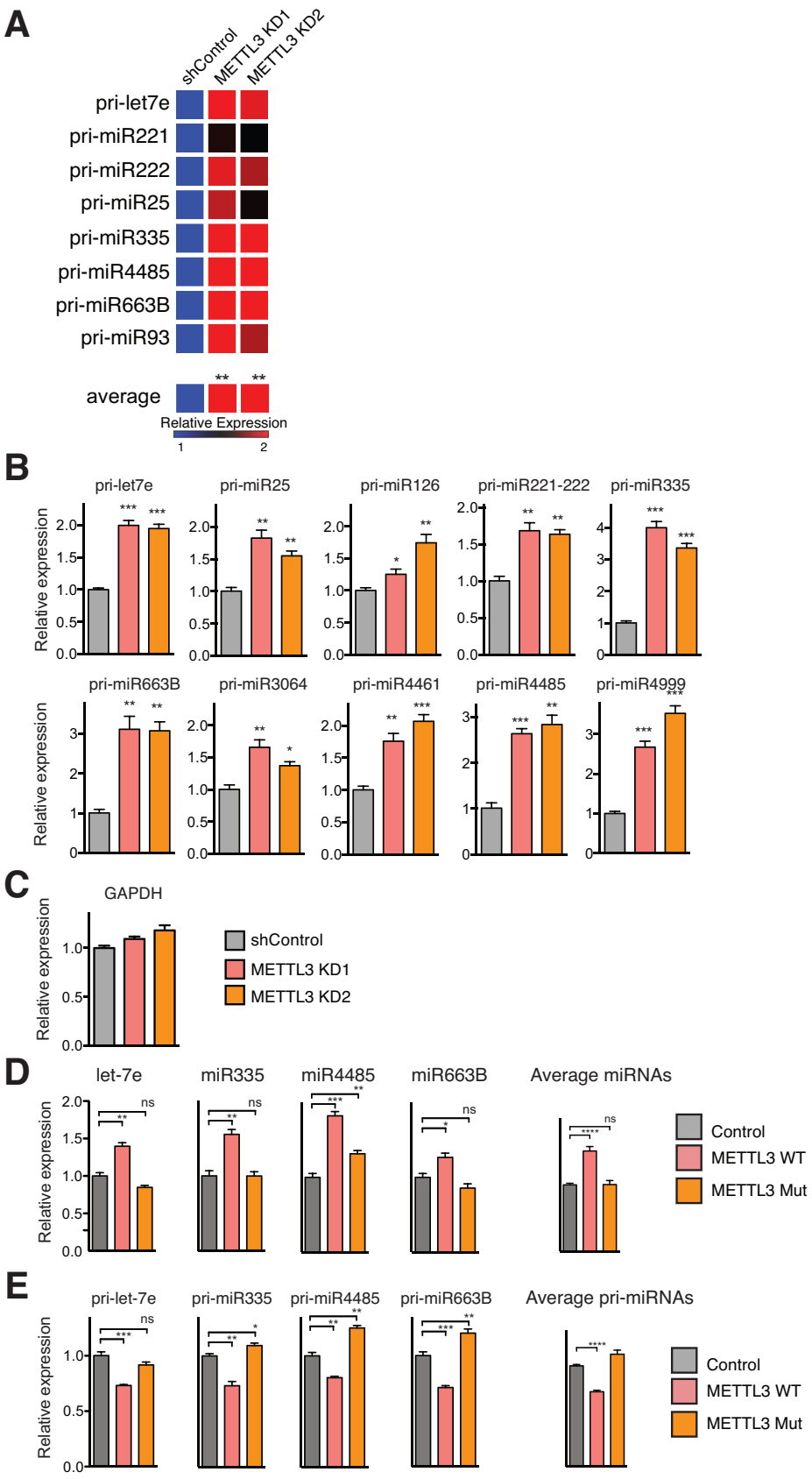


Figure 5.5 Mature miRNAs are upregulated by METTL3 over-expression in MDA-MB-231 cells

(A) Heat-map representation of the quantification of mature miRNAs from (Figure 5.3 A) by qRT-PCR upon METTL3 over-expression (B) qRT-PCR quantification of representative miRNAs that were modulated upon METTL3 over-expression (METTL3 OE) in MDA-MB-231 cells. Samples were normalized to RNU44. (C) qRT-PCR quantification of control RNU44 and GAPDH genes normalized to 18S. All experiments were done in biological replicates. Bar graphs represent a linear scale and error bars represent s.d. **** $P < 1\times10^{-4}$, *** $P < 5\times10^{-4}$, ** $P < 1\times10^{-3}$, * $P < 5\times10^{-2}$.

Figure 5.6 Quantification of pri-miRNAs levels upon depletion and catalytic inactivation of METTL3 in MDA-MB-231 cells

(A) Heat-map representing the quantification of pri-miRNA forms of miRNAs from Fig. 5.3 (A) and Fig. 5.5 (A) by qRT-PCR upon METTL3 depletion. (B) qRT-PCR quantification of representative pri-miRNAs that were impacted by METTL3 depletion using two independent hairpins in MDA-MB-231 cells. Expression levels were normalized to GAPDH. (C) qRT-PCR quantification of GAPDH, endogenous control. All experiments were done in biological replicates. (D-E) Quantification of mature (D) and pri-miRNAs (E) when MDA-MB-231 cells were stably transduced with either wildtype or a catalytic mutant METTL3. Mature miRNA expression was normalized by RNU44 and pri-miRNAs by GAPDH expression levels. The last bar graph shows the average of all the individual miRNA tested. The experiments were done in biological replicates. Bar graphs represent a linear scale and error bars represent s.d. **** $P < 1\times10^{-4}$, *** $P < 5\times10^{-4}$, ** $P < 1\times10^{-3}$, * $P < 5\times10^{-2}$; ns= not significant.


A
shControl
METTL3 KD1
METTL3 KD2
pri-let7e
pri-miR221
pri-miR222
pri-miR25
pri-miR335
pri-miR4485
pri-miR663B
pri-miR93
average
Relative Expression
B
pri-let7e
pri-miR25
pri-miR126
pri-miR221-222
pri-miR335
pri-miR663B
pri-miR3064
pri-miR4461
pri-miR4485
pri-miR4999
Relative expression
C
GAPDH
shControl
METTL3 KD1
METTL3 KD2
D
let-7e
miR335
miR4485
miR663B
Average miRNAs
Control
METTL3 WT
METTL3 Mut
E
pri-let-7e
pri-miR335
pri-miR4485
pri-miR663B
Average pri-miRNAs
Control
METTL3 WT
METTL3 Mut

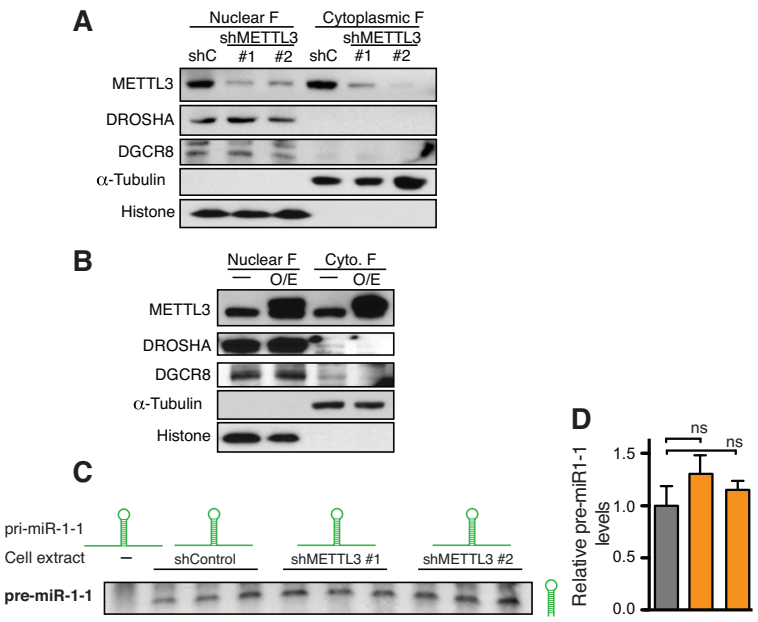


Figure 5.7 Expression and localization of the Microprocessor under depletion and overexpression of METTL3

(A) Western blot analysis of METTL3, DROSHA and DGCR8 obtained from nuclear and cytoplasmic fractions of cells transduced with two independent shRNAs against METTL3 (shMETTL3 #1 and #2) or with an shRNA control (shC). Tubulin and Histone 3 were used as loading controls as well as controls for the efficiency of the fractionation. (B) Same as (A), but in this case lysate from cells overexpressing METTL3 were compared to wildtype control cells. (C) *In vitro* pri-miRNA processing reactions. Whole cell extracts of control cells or cells depleted of METTL3 with 2 independent shRNAs were used to process *in vitro* transcribed pri-miR-1-1 to produce pre-miR-1-1 *in vitro*. Pre-miR-1-1 levels were then analyzed by Northern blot. (D) Hybridization intensities of (C) were quantified, normalized by their inputs and shown in a bar graph format. Bar graphs represent a linear scale and error bars represent s.d.

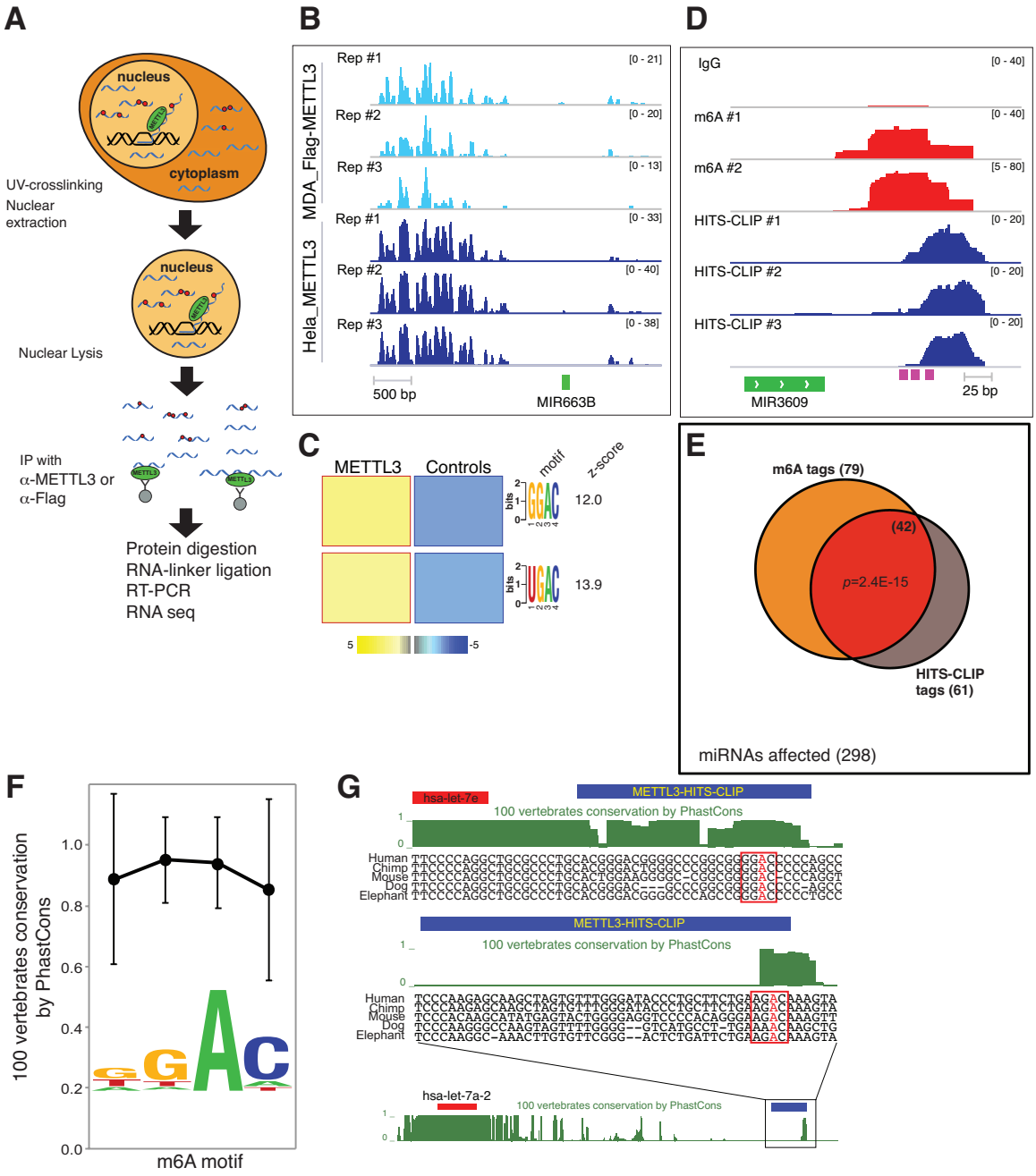
METTL3 marks pri-miRNAs

In order to directly implicate METTL3 in the m6A methylation of pri-miRNAs, we conducted high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) (Chi et al., 2009) experiments using an antibody against endogenous METTL3 in HeLa cells, and an anti-Flag antibody in MDA-MB-231 cells overexpressing Flag-epitope tagged METTL3 (Fig. 5.8 A-B). We next performed an unbiased search for cis-regulatory elements in the METTL3 HITS-CLIP reads to further validate that the sequences obtained from HITS-CLIP contained the METTL3 signature. Consistent with our findings from m6A-seq, METTL3-HITS-CLIP sequences displayed a significant enrichment of the METTL3 motif (Fig. 5.8 C). Despite the transient nature of the interaction between METTL3 and its targets, we observed a significant overlap in the miRNAs affected by METTL3 depletion that shared both the METTL3 footprint by HITS-CLIP and the m6A mark by m6A-seq ($P = 2.4\times10^{-15}$; Fig. 5.8 D-E).

In order to investigate the evolutionary conservation of the METTL3 motif across different species, we used the PhastCons software (Pollard et al., 2010) to analyze 30 miRNAs that are conserved among 100 vertebrates (targetscan.org) and also contain m6A and/or METTL3 tags. This analysis showed a high degree of conservation for the METTL3 motif among vertebrates (Fig. 5.8 F). These highly conserved regions were located next to the pre-miRNA regions but also hundreds of base pairs away from mature miRNA loci, as shown in the examples aligned using UCSC Genome Browser (Kent et al., 2002) (Fig. 5.8 G).

Figure 5.8 METTL3 targets pri-miRNAs for m6A methylation

(A) Schematic representation of the HITS-CLIP protocol used. (B) IGV tracks displaying sequencing read clusters from two independent HITS-CLIP experiments, one surveying endogenous METTL3 binding sites in HeLa cells (dark blue) and the other surveying Flag-METTL3 binding sites in MDA-MB-231 cells (light blue). The experiments were done in biological triplicates; the reads next to the miR-663b locus are shown as an example. (C) FIRE analysis provides motif of METTL3 HITS-CLIP binding sites. The color scale of the linear scale heat map is the same as in Fig. 5.1 A. (D) An example of sequencing clusters obtained from METTL3 HITS-CLIP (blue) and m6A-seq (red). m6A seq was done in duplicate using IgG as control. METTL3 HITS-CLIP was done in triplicate. The purple boxes at the bottom of the tracks represent conserved METTL3 motifs. (E) Venn diagram representation of the overlap of the miRNAs affected by METTL3 depletion and bearing the m6A and/or the METTL3 HITS-CLIP tags within 1kb from any particular miRNA locus. The overlap of miRNAs containing both m6A and METTL3 HITS-CLIP tags is depicted in red, $P = 2.4\times10^{-15}$. (F) Average conservation of the METTL3 motif among vertebrates using the PhastCons software. Error bars represent s.d. (G) Two examples of pri-miRNA genomic regions containing METTL3 clusters obtained from HITS-CLIP are shown. Partial pre-miRNA (filled red boxes) sequences are shown in the first example and METTL3 HITS-CLIP clusters are depicted in a blue box. The conserved METTL3 motif is framed in red with the putative methylated adenosine in red.


A
nucleus
cytoplasm
UV-crosslinking
Nuclear extraction
nucleus
Nuclear Lysis
IP with
α-METTL3 or
α-Flag
Protein digestion
RNA-linker ligation
RT-PCR
RNA seq
B
MDA_Flag-METTL3
Rep #1
[0 - 21]
Rep #2
[0 - 20]
Rep #3
[0 - 13]
Hela_METTL3
Rep #1
[0 - 33]
Rep #2
[0 - 40]
Rep #3
[0 - 38]
500 bp
MIR663B
C
METTL3
Controls
motif
z-score
12.0
13.9
5
-5
D
IgG
[0 - 40]
m6A #1
[0 - 40]
m6A #2
[5 - 80]
HITS-CLIP #1
[0 - 20]
HITS-CLIP #2
[0 - 20]
HITS-CLIP #3
[0 - 20]
MIR3609
25 bp
E
m6A tags (79)
(42)
p=2.4E-15
HITS-CLIP
tags (61)
miRNAs affected (298)
F
100 vertebrates conservation by PhastCons
m6A motif
G
hsa-let-7e
METTL3-HITS-CLIP
100 vertebrates conservation by PhastCons
Human
Chimp
Mouse
Dog
Elephant
METTL3-HITS-CLIP
100 vertebrates conservation by PhastCons
Human
Chimp
Mouse
Dog
Elephant
hsa-let-7a-2
100 vertebrates conservation by PhastCons

Role of the m6A modification by METTL3 in pri-miRNA processing

To demonstrate a direct role of the m6A modification in pri-miRNA processing we performed *in vitro* processing reactions using whole cell extracts from HEK293T cells transfected with DGCR8 and DROSHA (Lee and Kim, 2007). In this gain-of-function experiment, the extracts were used to process *in vitro* transcribed pri-miRNAs containing modified N^6-methyladenosine or unmodified bases. Consistent with our model, methylated pri-let-7e was more efficiently processed by the microprocessor to produce pre-let7e relative to its un-methylated counterpart as detected by northern blot. For a normalization control, we included in each reaction the un-methylated pri-miR-1-1 (Fig. 5.9 A-C). These *in vitro* experiments suggest that m6A marks in pri-miRNAs are required for efficient processing of pri-miRNAs *in vitro*. Additionally, we performed a loss-of-function experiment by testing the impact on miRNA processing of mutating select adenosines present in potential METTL3 motifs located in a pri-miRNA. To do this, we modified a previously described reporter construct (Auyeung et al., 2013). In one reporter, we introduced a wild type version of pri-let-7e, while in the other, the adenosines of the METTL3 motifs present in the pri-let-7e region (i.e. 5 out of 18 adenosines present outside the pre-miRNA sequence) were mutated (Fig. 5.9 D). Each reporter contained an internal control miRNA, pri-miR-1-1, which is co-transcribed and used for let-7e expression normalization. This control miRNA was also mutated in both vectors to make its processing independent of methylation. These reporters were transfected into

HEK293T cells and RNA was extracted to analyze the expression levels of mature let-7e and mature miR-1-1. Consistent with our findings, mutation of METTL3 motifs in pri-let-7e significantly reduced its processing to the mature form (Fig. 5.9 E).

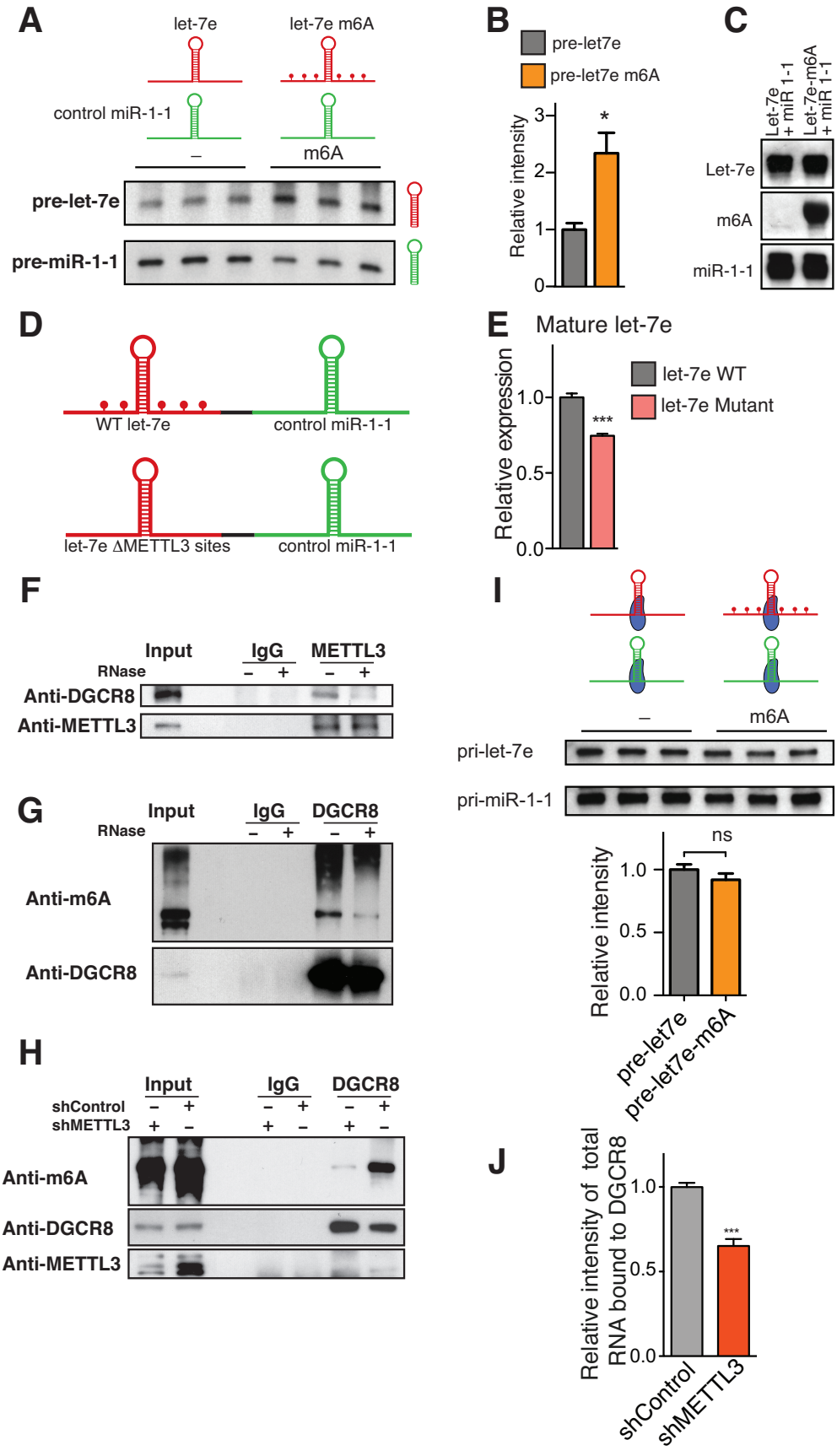
Since the m6A mark is added to the pri-miRNA, and we observed an accumulation of unprocessed pri-miRNAs upon METTL3 depletion, we questioned whether METTL3 is required for the engagement of the pri-miRNA by the microprocessor. We first tested whether METTL3 and DGCR8 are part of a complex. Immunoprecipitation of METTL3 from nuclear lysates co-precipitated DGCR8; however this interaction was mediated through RNA, since RNase treatment disrupted this association (Fig. 5.9 F). We next sought to determine whether DGCR8 interacts with methylated RNA *in vivo*. To do this, we UV-crosslinked cells, isolated nuclear fractions and immunoprecipitated endogenous DGCR8. After immunoprecipitation of DGCR8 and SDS-polyacrylamide gel electrophoresis (SDS-PAGE), m6A immunoblotting revealed that DGCR8 indeed interacts with methylated RNA (Fig. 5.9 G). Based on our observations, we predicted that a reduction in METTL3 levels should decrease the levels of methylated RNA bound by DGCR8. Indeed, depleting METTL3 significantly reduced the m6A methylated RNA bound by DGCR8 (Fig. 5.9 H). However, the binding of DGCR8 to methylated RNA is unlikely to be direct, as we detected no preference for DGCR8 binding to methylated RNA over unmethylated RNA *in vitro* (Fig. 5.9 I).

Since mRNAs tend to form secondary structures including short hairpins that resemble pri-miRNAs, a potential basis of pri-miRNA methylation might be to confer specificity for, and facilitate the recognition of, pri-miRNA structures by DGCR8. Based on this hypothesis, we would expect that a reduction in the levels of methylated pri-miRNAs would reduce the total amount of RNA recognized and bound by DGCR8. To test this, we immunoprecipitated DGCR8 from control and METTL3 depleted cells and radio-labeled the total RNA bound to DGCR8. Consistent with our model, METTL3 depletion significantly reduced the amount of total RNA (Fig. 5.9 J) as well as pri-miRNAs (Fig. 5.10) bound by DGCR8. These findings suggest that the m6A methylation mark allows for the effective recognition of pri-miRNAs by DGCR8 and their subsequent processing to pre-and mature miRNAs.

Figure 5.9 m6A methylation of pri-miRNAs is required for normal processing by DGCR8

(A) *In vitro* pri-miRNA processing reactions. Pri-let-7e, was transcribed with either a modified N^6-methyladenosine (depicted in the schematic as red dots) or with unmodified bases, was co-incubated in the cell extract with pri-miR-1-1 and the production of pre-let-7e as well as pre-miR-1-1 was analyzed by northern blot. (B) Hybridization intensities were quantified and shown in the bar graph. Bars represent the average normalized intensity of three biological replicates, * $P < 5\times10^{-2}$. (C) Input for experiment shown in (A). (D-E) *In vitro* assays. (D) Schematic representation of the reporters used to study the role of the METTL3 on the processing of pri-miRNAs. Represented in red is the pri-let-7e sequence and in green, the control pri-miR-1-1. The top reporter contains a wildtype sequence of the pri-let-7e and the potential sites of methylations are depicted as red dots. The reporter on the bottom contains a mutant version of pri-let-7e in which the 5 putative adenines of the METTL3 motif were mutated. (E) HEK293T cells were transfected with the reporters depicted in (D), RNA was extracted, and mature miRNA expression quantified by qRT-PCR. The bar graph represents the relative expression levels of mature let-7e normalized to mature miR-1-1. (F) Co-immunoprecipitation of the METTL3-interacting protein DGCR8. Western blot using the indicated antibodies with IgG used as control for the IP.

(G) Immunoprecipitation of endogenous DGCR8. Western blot using an antibody against m6A methylated RNA with IgG used as control for the IP. (H) Similar as (G), but control cells or cells depleted of METTL3 were compared. (I) *In vitro* binding assays using immunopurified DGCR8. Samples containing *in vitro* transcribed pri-let-7e with N^6-methyladenosine or unmodified adenosines were incubated with magnetic beads bound-DGCR8, washed, eluted and analyzed by Northern blot. All reactions contained unmodified pri-miR-1-1 as endogenous control. The upper panel shows pri-let-7e and the lower panel pri-miR-1-1. On the right side the bar graph depict the average intensity of the pri-let-7e normalized by pri-miR-1-1 levels. The experiment was done in biological triplicate. Bar graphs represent a linear scale and error bars represent s.d., *** $P < 5\times10^{-4}$, ns= not statistically significant. (J) Similar immunoprecipitation as (G), however, after immunoprecipitation of DGCR8, a radiolabelled RNA linker was ligated to the RNA bound by DGCR8. Bars represent the average normalized intensity of three biological replicates, *** $P < 5\times10^{-4}$.


A
let-7e
let-7e m6A
control miR-1-1
–
m6A
pre-let-7e
pre-miR-1-1
B
pre-let7e
pre-let7e m6A
Relative intensity
3
2
1
0
*
C
Let-7e + miR 1-1
Let-7e-m6A + miR 1-1
Let-7e
m6A
miR-1-1
D
WT let-7e
control miR-1-1
let-7e ΔMETTL3 sites
control miR-1-1
E
Mature let-7e
let-7e WT
let-7e Mutant
Relative expression
1.0
0.5
0.0
***
F
Input
IgG
METTL3
RNase
– + – +
Anti-DGCR8
Anti-METTL3
G
Input
IgG
DGCR8
RNase
– + – +
Anti-m6A
Anti-DGCR8
H
Input
IgG
DGCR8
shControl – + – + – +
shMETTL3 + – + – + –
Anti-m6A
Anti-DGCR8
Anti-METTL3
I
–
m6A
pri-let-7e
pri-miR-1-1
ns
Relative intensity
1.0
0.5
0.0
pre-let7e
pre-let7e-m6A
J
Relative intensity of total RNA bound to DGCR8
1.0
0.5
0.0
***
shControl
shMETTL3

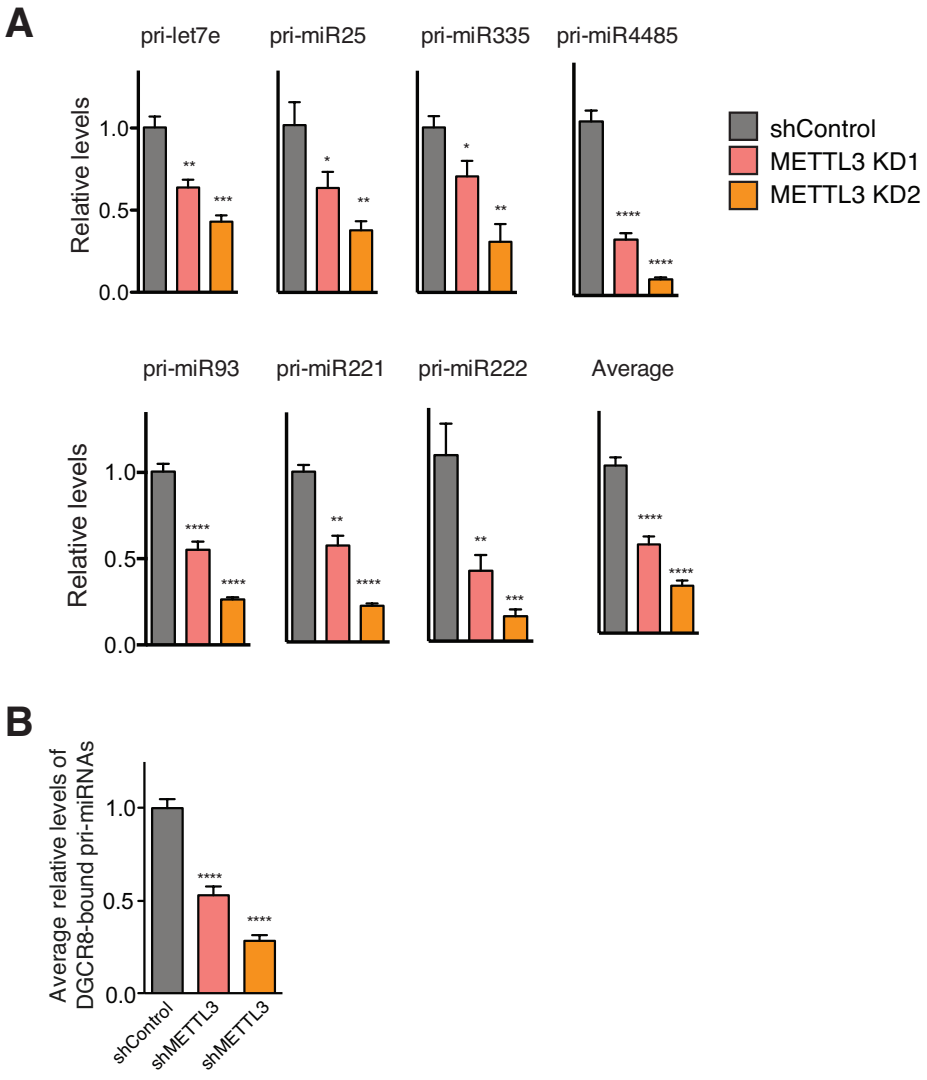


Figure 5.10 Effect of METTL3 depletion on endogenous pri-miRNAs binding to DGCR8

(A) Immunoprecipitation of endogenous DGCR8 crosslinked to RNA of control cells or cells depleted of METTL3 using two independent shRNAs. After the immunoprecipitation the RNA was extracted and a panel of pri-miRNAs was quantified by qRT-PCR. (B) The average quantification of (A). Bar graphs represent a linear scale and error bars represent s.d., **** $P < 1\times10^{-4}$, *** $P < 5\times10^{-4}$, ** $P < 1\times10^{-3}$, * $P < 5\times10^{-2}$.

PART II: DISCUSSION

By using an unbiased approach, we have identified m6A as a novel regulator of miRNA processing. Our findings reveal that METTL3 methylates primary inter- and intragenic miRNAs. The m6A modification facilitates the recognition of pri-miRNA sequences and marks an initiation event in miRNA biogenesis (Fig. 5.11). We propose that the m6A mark on transcripts at pri-miRNA sequences initiates a global co-transcriptional program comprising the engagement and processing of primary miRNAs by the microprocessor machinery. The m6A mark thus plays an important role in the nucleus—allowing the microprocessor complex to recognize its intended substrates as opposed to potential hairpin structures not containing miRNA products. Additionally, altered METTL3 expression in various human malignancies may significantly contribute to the aberrant expression of miRNAs seen cancer.

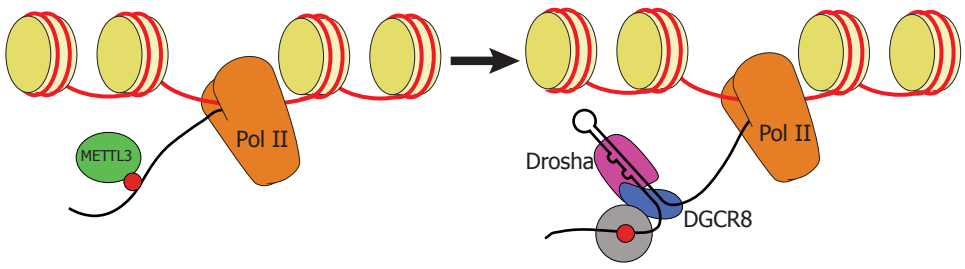


Figure 5.11 Model of METTL3 regulation of miRNA biogenesis

The molecules represented in the schematic are: histones (yellow), RNA Pol II (orange), METTL3 (green), m6A (red), DGCR8 (blue), DROSHA (pink) and a putative unknown reader of the m6A mark (grey).

MATERIALS AND METHODS

Cell Culture

The MDA-MB-231 line, its metastatic derivatives LM2 cells, CN34 parental cells, and its metastatic derivatives LM1A cells have been cultured as described previously (Minn et al., 2005). Briefly, all these four breast cancer cell lines, HEK293T, and HeLa cells were propagated *in vitro* with Dulbecco's modified Eagles medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin, 2mM L-Glutamine, 1mM sodium pyruvate, and 2.5 ug/mL fungizone. The lung cancer cell line H522 was purchased from ATCC and the A549 cell line was a kind gift from H. Choi, a graduate student in Dr. V. MITTAL laboratory in Weill Cornell Medical College. Lung cancer cell lines were cultured in RPMI media with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin, 2mM L-Glutamine, 1mM sodium pyruvate, 2.5 ug/mL fungizone, 1mM HEPES, and 0.15% Sodium bicarbonate. HUVECs were cultured with Endothelial Cell Growth Media (CC-3162, Lonza). Experiments with HUVEC were conducted with cells from passage 2 to 5.

Generation of lentivirus, retrovirus, knockdown and over-expressing cells

Generation of lentivirus-mediated knockdown and retroviral-overexpressing cells were performed as described previously (Minn et al., 2005; Tavazoie et al., 2008). For overexpression studies, human cDNA of SOX4 with N-terminal FLAG tag was cloned into pBabe-puro retroviral expression vector. 10 ug of DNA were transfected into H29 packaging cell line in a 10cm plate by Lipofectamine 2000 (ratio of ug DNA: uL Lipofectamine 2000 is 1:4). Virus-containing supernatant at

48hr and 72hr post-transfection were harvested, filtered through a 0.45 μm syringe filter, and added to the 70% confluent breast cancer cells in a 10cm plate with polybrene (8 ug/mL). After over-night incubation, cells were recovered with fresh growth media for 1 day and cultured with growth media containing puromycin (1 ug/mL) for 2-3 days. Human cDNA of TMEM2 with C-terminal V5 tag was obtained from CCSB Lentiviral Expression Library (OHS6087) (Yang et al., 2011). The C-terminal frame shift mutation present in the ORFeome library was fixed and cloned into pLX304 lentiviral expression vector by Gateway cloning. For overexpression of METTL3, human cDNA of METTL3 with C-terminal FLAG tag was cloned into pBabe-Puro retroviral expression vector. Refer to list of cloning primers in Appendix.

For knockdown studies, commercially available shRNA glycerol stocks were purchased (Sigma). Two shRNAs against SOX4 that showed best knockdown efficiency (TRCN0000018216, TRCN0000018217), two shRNAs against TMEM2 (TRCN0000148114, TRCN0000147636), two shRNAs targeting METTL3 (TRCN0000034715, TRCN0000034717) and control shRNA (SHC002) were used for experiments. 2.5 ug of pLKO lentiviral plasmid with lentiviral packaging plasmids (5 ug of gag/pol and 2.5 ug of env) was co-transfected into HEK293T cells in a 10cm plate by Lipofectamine 2000. Lentivirus-containing supernatant were harvested and processed in the same way as described above.

Generation of inducible over-expressing cells

N-terminal FLAG tagged SOX4 was pcr amplified and sub-cloned into pEN_Tmcs Entry vector (Addgene Plasmid #25751) which has TRE promoter followed by multiple cloning sites. Then this entry vector was digested with BamHI and EcoRV to be ligated with pSLIK-Neo lentiviral vector (Addgene plasmid #25735) that was digested with AleI and XhoI. Since EcoRV and AleI generate blunt ends, the digested pSLIK-Neo lentiviral vector was CIP-treated and then both insert and the vector were blunt-ended by Klenow followed by ligation. Sequence verified construct was used to generate doxycycline inducible SOX4 expression cell lines.

siRNA-mediated mRNA knockdown

Cells were seeded the day before siRNA transfection. siRNAs targeting either SOX4 (Thermo Scientific oligo ID ROSJN-00001, ROSJN-00003) or TMEM2 (Integrated DNA Technologies (IDT) siRNA identity HSC.RNAI.N013390.12.10, HSC.RNAI.N013390.12.2) and control siRNA (Thermo Scientific, siGENOME Control siRNA #1) were transfected with Lipofectamine reagent according to the manufacturer's protocol. Forty-eight hours after transfection, the cells were collected and applied to either trans-well invasion assay or trans-well migration assay.

Transcriptomic profiling (microarray)

1 ug of total RNA extracted from two independent RNAi-mediated knockdown of SOX4 and two control cells was labeled and hybridized on Illumina BeadChip array (GPL10558_HumanHT-12_V4_0_R2_15002873_B) by the Genomics core facility at Rockefeller University. Heat maps depicting relative expression levels of genes from microarray data was created with matrix2png (Pavlidis and Noble, 2003). For miRNA expression profiling, total RNA extracted from two independent stable knockdown of METTL3 and two control cells was labeled and hybridized on miRNA microarrays by LC sciences. The arrays have probes for all miRNAs of human (1872 precursors, 2578 mature) available in the miRBase database (Release 20). Of all the probes assayed, those corresponding to 438 miRNAs revealed a signal above background in at least two of the MDA-MB-231 cell lines.

Analysis of mRNA expression

Total RNA from cancer cells was extracted and purified using the mirVana (Applied Biosystems) or Total RNA Purification Kit (Norgen Biotek). mRNA expression quantification was performed as described previously (Png et al., 2012). Refer to list of qRT-PCR primers in Appendix.

Chromatin immunoprecipitation sequencing (ChIP-Seq)

2.2×10^7 cancer cells either expressing FLAG tagged SOX4 or control empty vector were harvested, washed with ice-cold PBS twice and flash-freeze in liquid nitrogen. The pellets were fixed with 1% paraformaldehyde (PFA) in PBS for 10 min at room temperature with rotation. Glycine was added to a final concentration of 0.125 M to quench unreacted PFA and incubate for 5 min at room temperature. The fixed cells were lysed in ice-cold LB1 buffer containing 50 mM Hepes-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Triton x-100, and protease inhibitors (Thermo Scientific, 78443). The pellet was washed with ice-cold LB2 buffer containing 10 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1mM EDTA, 0.5 mM EGTA and protease inhibitors. The pellet was digested with micrococcal nuclease (MNase, Worthington) and further lysed with LB3 buffer containing 10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1 % Na-Deoxycholate, 0.5 % N-lauroylsarconsine, 1 % Triton x-100, and protease inhibitors. The pellet was sonicated 60 times for 30 sec with intermediate incubation of 30 sec (Bioruptor, Diagenode). The lysate was incubated at 4°C for overnight with anti-FLAG M2 Magnetic Beads (M8823, Sigma). The immunoprecipitations were washed twice with the wash buffer containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 1% Triton x-100 followed by last wash with TE buffer. Then eluted for 30 min at 65°C with the elution buffer containing 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS. Crosslink was reversed for overnight at 65°C with shacking at a speed of 300 rpm (Eppendorf

Thermomixer 5350). RNA and protein digestions were performed sequentially by incubation with RNAse (Roche, 11-119-915-001) for 1 hour at 37°C and then with Proteinase K (Roche, 03-115-828-001) for 2 hours at 55°C with shacking at 300 rpm. The DNA was recovered and purified by QIAquick PCR Purification Kit (Qiagen, 28104) and further processed for high-throughput sequencing with NEXTflex ChIP-Seq Kit (5143-01) according to the manufacturer's protocol.

Western Blotting

Protein expression of FLAG tagged SOX4 was detected using anti-FLAG conjugated with HRP (A8592 Sigma) or SOX4 antibody generated from rabbits immunized with a synthetic peptide between 352 and 365 amino acids from the C-terminal region of human SOX4 (CGR SPA DHR GYA SLR).

Clinical correlation analysis

The Cancer Genome Atlas (TCGA) Data: Expression levels of SOX4 and TMEM2 in TCGA Breast Invasive Carcinoma were obtained from cBioPortal for Cancer Genomics (www.cbioportal.org) and UCSC cancer browser (genome-cancer.ucsc.edu). Among 1098 patients, 1073 patients had stage information ranging from Stage I to Stage IV. Nonparametric Spearman correlation with one-tailed p-value are shown in figure.

Expression Project for Oncology (ExpO) Data: Total of 236 breast cancer samples (4 of stage 0, 31 of stage 1, 129 of stage 2, 67 of stage 3, 5 of stage 4) were obtained from the ExpO microarray database (GSE2109) (http://www.intgen.org/research-services/biobanking-experience/expo/). Expression levels of SOX4 and the putative target genes were analyzed.

Gene expression-based Outcome for Breast cancer Online (GOBO) Data: Assessment of gene expression levels and association with overall survival outcome for TMEM2 and SOX4 in breast cancer subgroups (stage 0, I, II, III, IV) were done by GOBO (http://co.bmc.lu.se/gobo/gobo.pl).

Luciferase reporter assay

The SOX4 bound regions of TMEM2 obtained from ChIP-qPCR were cloned into the pGL3-Promoter vector (Promega, E1761) that contains the Firefly luciferase gene downstream from a multiple cloning site into which the promoter of interest is inserted. These vectors (25 ng) were co-transfected with the Renilla luciferase

vector (2.5 ng) into HEK293T cells either transiently overexpressing FLAG tagged SOX4 (22.5 ng) or control empty plasmid (22.5 ng). After 30 hours post transfection, the cells were processed with Dual-Luciferase Reporter Assay System (Promega, E1910) according to the manufacturer's protocol. Luciferase activities were measured in a luminometer and the Firefly luciferase activity was normalized to the Renilla luciferase activity.

Trans-well invasion assay

The trans-well invasion assays were performed as described previously (Tavazoie et al., 2008). Cells were incubated overnight in the culture media with low concentrations of FBS (0.2%). The next day, the cells were trypsinized, re-suspended in the low serum media, and seeded at 5×10^4 cells per well into growth factor reduced matrigel invasion chambers (8 μm pore size, BD Biosciences). After 22hours, the chambers were washed with PBS twice and the cells on the apical side of each inserts were scraped off. The invaded cells fixed with 4% paraformaldehyde were stained with DAPI and quantified with ImageJ.

Trans-well migration assay

Trans-well migration assays were performed as described previously (Tavazoie et al., 2008) with minor modifications: cells were incubated with low concentrations of FBS (0.2%) for overnight and then seeded at 1×10^5 cells per cell-culture insert made of Track-etched polyethylene terephthalate (PET) membranes with pores (3 μm pore size, BD Biosciences). After 12hours, the migrated cells were stained with DAPI and quantified with ImageJ.

Animal Studies

All animal work has conducted in accordance with a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at The Rockefeller University. NOD scid gamma female mice (The Jackson Laboratory) age-matched between six and eight weeks were used for tail-vein injection.

Statistical Data Analysis

All data are represented as mean ± standard error of mean (SEM) with p-values: **** $P < 1\times10^{-4}$, *** $P < 5\times10^{-4}$, ** $P < 1\times10^{-3}$, * $P < 5\times10^{-2}$. One-tailed Student's t test was used and $P < 1\times10^{-5}$ was considered to be statistically significant.

qRT-PCR for miRNAs

Mature miRNAs were quantified by either Taqman microRNA assays (Applied Biosystems) or poly-A tailing of total RNA followed by reverse-transcriptase (RT) reaction with T7 oligo dT (NEB). Quantitative miRNA expression data were acquired and analyzed using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Pri-miRNAs were measured using the TaqMan Pri-miRNA Assay. RNU44, GusB, GAPDH and 18S were used as endogenous controls.

m6A Immunoprecipitation and RNA seq

3×10^7 cells/sample were lysed using LB1 buffer (50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% Triton x-100, and protease inhibitors). The nuclear fraction was then lysed with M-PER buffer and diluted 10-fold in

dilution buffer before the immunoprecipitation. Rabbit a-m6A antibody (Synaptic Systems) and rabbit IgG control bound to protein A Dynabeads were used for the immunoprecipitations. The immunoprecipitated RNA was eluted with N^6-methyladenosine (Sigma-Aldrich), ethanol precipitated and resuspended in water. RNA was barcoded using ScriptSeq V2 kit (Epicentre) and sent for sequencing.

m6A analysis

Paired reads were mapped to the human genome (build hg19) using bowtie2 with "sensitive-local" option. Consistent paired reads were converted into genomic intervals (bed format). ChIPseeqer was used to detect statistically significant peaks that were present in the m6A co-precipitation samples and not the IgG control. The following parameters were used: -t 10 -fold_t 2 -min len 20. FIRE was used for non-discovery analysis, and ChIPseeqer was used for shuffled sequences.

HITS-CLIP

We performed HITS-CLIP as previously described (Zhang and Darnell, 2011) with some modifications. After UV crosslinking (400mJ/cm2 of 254nm UV), we isolated nuclear fractions from $3x10^7$ cells/sample using LB1 buffer. The nuclear fraction was then lysed in M-PER buffer (Thermo Scientific) and diluted 10 fold in a buffer containing 50 mM Tris-HCl pH 7.4 and 100mM NaCl, before the immunoprecipitation. For the endogenous immunoprecipitation we used 3ug of a-METTL3 rabbit antibody (Bethyl) or rabbit IgG as control bound to protein A

dynabeads (Invitrogen). For Flag immunoprecipitation we used cells overexpressing Flag-METTL3 or cells without a Flag-tagged protein as the control and used anti-Flag magnetic beads (Sigma). After the immunoprecipitation a ^{32}P-labeled linker was ligated to the RNA on the beads. The samples were resolved using SDS-PAGE, transferred to nitrocellulose membrane and exposed overnight. The films were used to guide the cut of the membrane to extract the radiolabeled RNA. RNA was extracted in 200ul of PK buffer containing proteinase K (4mg/ml), then adding 200ul of PK buffer containing 7M urea and lastly 400 ul of acid phenol:chloroform, each step 20 min at 37°C at 1,000 rpm shaking (Eppendorf Thermomixer 5350). The aqueous solution was then precipitated and resuspended in RNase-free water as described in the *in vitro* processing section. An RNA linker was ligated to the purified RNA at 16°C overnight. The RNA was once again extracted using phenol:chloroform and precipitated as previously described. This new RNA was reverse transcribed and PCR enriched. These PCR amplicons were resolved in a TBE-Urea poly-acrylamide gel and DNA between 90 and 140 bp was extracted. This DNA was further enriched with the addition of Solexa fusion primers and the product was resolved in a 2% metaphor agarose/TBE gel, from which and a DNA between 150-170 bp was extracted. The samples were submitted for high throughput sequencing using an Illumina HiSeq 2000 instrument with 50bp single-end runs.

HITS-CLIP analysis

The reads were trimmed and clipped using cutadapt (v1.2.1) and were subsequently mapped to the human genome (build hg19) using bowtie. CIMS package was used to detect CLIP clusters with cross-linking induced mutations. FIRE was used for non-discovery analysis, and ChIPseeqer for shuffled sequences.

Density plots analysis

The distance between peaks located in introns to closest miRNAs were measured using closestBed. In R, ggplot was used to generate the density plots.

Co-Immunoprecipitations

Cells were lysed with LB1 and the nuclear fraction was then lysed with M-PER buffer (Thermo Scientific) and diluted 10-fold in dilution buffer. Immunoprecipitations were performed with a-METTL3 rabbit antibody (Bethyl) or a-DGCR8 rabbit antibody (Abcam), previously bound to magnetic Dynabeads Protein A (Life Technologies), in the presence with either RNase A or RNase inhibitor (Promega). The immunoprecipitations were washed twice with high-salt buffer (50 mM Tris-HCl pH 7.4, 300 mM NaCl) followed by two times more wash with low-salt buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl). The following antibodies were used for western blot analysis a-METTL3 (mouse, Novus Biologicals), a-DGCR8 (mouse, Abcam), a-m6A (mouse, Synaptic Systems).

For the *in vitro* binding assay once DGCR8 was immunopurified and still bound to the magnetic beads it was incubated with 1ug of pri-let-7e and 200ng of pri-miR-1-1, in the presence of RNAse inhibitors for 20 min at room temperature, washed twice and eluted with 170 ul of RNA elution buffer (0.3 M sodium acetate, pH 5.5 and 2% SDS). RNA was then extracted with 200ul of acid phenol:chloroform, precipitated and resuspended in RNase-free as described in the *in vitro* processing section. An aliquot of the elution was then used for northern blot analysis.

For the radiolabeled experiments the cells were UV-crosslinked, the nuclear fraction was lysed with M-PER buffer and diluted 10-fold in dilution buffer. The nuclear extracts were immunoprecipitated with a-DGCR8 antibody or IgG as control. A RNA oligo was labeled with ^{32}P-gamma-ATP (Perkin Elmer) and ligated to the co-immunoprecipitated RNA as in the HITS-CLIP protocol. The samples (triplicates) were loaded in a SDS-PAGE, transferred, and the membrane was exposed overnight or longer if needed. After the image of the radiolabeled RNA was obtained, the same membrane was immunoblotted for DGCR8 for normalization. Quantitation was achieved using ImageJ.

For the endogenous DGCR8-pri-miRNA binding experiments, similar to the radiolabeled experiments, the cells were UV-crosslinked and the nuclear fraction was lysed with M-PER buffer and diluted in dilution buffer. The nuclear extracts were immunoprecipitated with an anti-DGCR8 antibody or IgG as a control. The RNA was extracted by treatment with 200 ul of proteinase K (4mg/ml) in PK buffer

(100mM Tris-Cl pH 7.5, 50mM NaCl and 10mM EDTA) and incubation at 37°C for 20 min with shaking. Then, 200ul 7M urea in PK buffer was added and incubated 20 min more at 37°C with shaking. After that 400ul acid phenol:chloroform (Ambion) was added to the solution, vortexed and centrifuged at maximum speed for 5 min.The aqueous phase was precipitated with 50ul 3M NaOAc pH 5.2, 0.75ul glycoblue and 1mL of 1:1 ethanol:isopropanol at -20°C at least 30 min. The mix was then centrifuged at 4°C at maximum speed for 15 min, and the pellet washed with 80% ethanol and air-dried. The RNA was resuspended in 20ul of RNase-free water and 8 ul were used to produce cDNA by ramdom hexamers (SuperScript III First-Strand Synthesis System, Life Technologies) and qRT-PCR was carried out using specific miRNA primers.

Northern Blots

Non-radioactive northern blots were performed as previously described (Kim et al., 2010). RNA was extracted with miRVana or Norgen Biotek RNA extraction kits. DIG-labeled LNA probes against let-7e and miR-1-1 were obtained from EXIQON. Pre-stained marker (DynaMarker) was obtained from BioDynamics Laboratory Inc and Low range ssRNA Ladder from NEB. Mini-protean 10% and 15% pre-cast TBE-Urea gels were purchased from Bio-RAD and TBE buffer from Invitrogen. Samples were prepared using 2x Gel loading buffer (Ambion) and denatured 5min at 95C. Gels were at 200V in 1X TBE buffer and then transfered to a nylon membrane at 80V for 1 hour in 0.5X TBE buffer. Membranes were

either chemically crosslinked with EDC (0.753g of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide in 24ml of 125mM solution of 1-methylimidazole pH of 8.0) for mature and pre-miRNAS (2 hours at 65C) or UV crosslinking for pri-miRNAs (240 mJoules). The membrane was then pretreated with hybridization buffer (DIG Easy Hyb Granules, Roche) at 37°C for at least 30 min in hybridization oven. The probe was denatured at 95°C for 1 min and added to the hybridization buffer (37°C overnight). The membrane was washed twice with Low Stringent Buffer at 37°C for 15 min, twice with High Stringent Buffer at 37°C for 5 min. and once with Washing Buffer at 37°C for 10 min. Then the membrane was incubated in Blocking Buffer (DIG Wash and block buffer set, Roche) for 3 hours at room temperature and the anti-Digoxigenin-AP antibody (Roche) was added (1:15,000 dilution) and incubated at room temperature for an extra 30 min. The membrane was washed in DIG Washing buffer four times for 15 min each at room temp. Finally the membrane was incubated in development buffer for 5 min. and developed with CDP-Star (Roche) for 5 min and exposed to X-ray films.

In vitro pri-miRNA processing

Pri-miRNAs were *in vitro* transcribed using the T7 based MEGAshortscript kit (Life Technologies) according to the manufacture's indications. In cases were m6A modifications were added *in vitro,* N^6-methyl-ATP (tri-link) was used in a ratio 4:1 to ATP in the *in vitro* transcription reaction. Pri-miRNA processing reactions were performed as previously described (Lee and Kim, 2007). In brief, 293 cells were co-transfected plasmids carrying DROSHA and DGCR8 (obtained from Addgene). 48 hours later, cells were lysed in lysis buffer (20 mM Tris-HCl pH 8.0, 100 mM KCl, and 0.2 mM EDTA) and sonicated for 1 min with 5 sec pulses at 30% amplitude. The whole cell was obtained after centrifugetion at 13,523 rcf at 4°C for 15 min. In the case of testing the effect of METTL3 depletion on pri-miRNA processing, cells were not transfected and the endogenous activity was measured. The *in vitro* processing reaction contained 3 ul of 64 mM MgCl2 solution, 1 ug of Let-7e pri-miRNA (wildtype or m6A methylated), 50ng of control pri-miR-1-1, 0.75 ul of RNase inhibitor, 15 ul of HEK293T whole extract and RNase-free water to a final volume of 30 ul. The reactions were incubated at 37°C for 90 min. Then 170 ul of RNA elution buffer (0.3 M sodium acetate, pH 5.5 and 2% SDS) was added to the reaction to dissociate the RNA from the proteins and terminate the reaction. RNA was then extracted with 200 ul of acid phenol:chloroform (Ambion), vortexed and centrifuged for 5 min at room temperature. After extraction the RNA was precipitated at -20°C by adding 20 ul of sodium acetate (3M), 1 ul of glycoblue (Life Technologies), and 800 ul of a mix

isopropanol:ethanol (1:1). The mix was then centrifuged at 4°C at maximum speed for 15 min, and the pellet washed with 80% ethanol and air-dried. The RNA was resuspended in 30ul of RNase-free water and 5ul were used for Northern Blot analysis.

Reporter assays

Analysis of pri-miRNA processing using ectopic reporter constructs was performed as previously described (Auyeung et al., 2013). We used the pri-miRNA reporter construct developed by Bartel and collegues (Auyeung et al., 2013) and replaced the miRNA control pri-miR-1-1 with a mutant version in which the adenosines of the potential METTL3 motifs were mutated. Then we placed the query miRNA, either wildtype pri-let-7e or its altered version in which the A's of the putative METTL3 motifs were mutated to T's, upstream of the pri-miR-1-1 control. We then used these two constructs to transfect HEK293T cells using Lipofectamine 2000. 48 hours later the RNA was extracted and qRT-PCR was performed to test the production of mature let-7e and mature miR-1-1.

APPENDIX – LIST OF QRT-PCR AND CLONING PRIMERS

QRT-PCR primers

QRT-PCR primers			
Gene	FWR primer	REV primer	Amplicon size (bp)
18S	gtaacccgttgaaccccatt	ccatccaatcggtagtagcg	151
GAPDH	agccacatcgctcagacac	gcccaatacgaccaaatcc	66
SOX4	gaaggtgaagtccggcaac	accgaccttgtctcccttct	82
BCAR3	ctgcgcctggacataattg	gctcggtcctccaaagtg	92
PFKFB3	ttggcgtccccacaaaagt	agttgtaggagctgtactgctt	75
RPH3AL	ctggagcagcagagaatcg	ccgttccccatcacattc	71
SLC3A2	gtgctgggtccaattcacaag	caccccggtagttgggagta	166
SPANXA2	cgaggccaacgagatgat	tggtcgaggactcagatgttt	92
TMEM2	ttcagggtggtcagaacaagt	aagcctctaattgggaacgtc	94
TMEM2_pre-mRNA	gtcctcctccttgttgtgga	aaaatcggacttgaccaggc	207
ChIP-qPCR primers			
GAPDH_promoter	tactagcggttttacgggcg	tcgaacaggaggagcagagagcga	166
TMEM2_promoter_#3	ctggtactgcgcctgtca	atatttattccggcctgcgag	300
TMEM2_promoter_#4	agctcttctggtcccttgga	agtcctgcagctgtcactag	321
TMEM2_promoter_#6	ctagtgacagctgcaggact	gtagattgcgcgacgtctg	323
TMEM2_promoter_#7	agtcctgtaaccgcgacac	ctttcctccctgcctatccc	349
TMEM2_promoter_#12	ctgtgctgggccaactttat	cgcctggcctaattgctatt	257
TMEM2_promoter_#13	ttcccaaggccacacagtaa	caccttccttagcctcccaa	131

Cloning primers

N-terminal FLAG tagged SOX4	
EcoRI-FLAG-huSox4	gggggaattcatggattacaaggatgacgacgataaggtgcagcaaaccaacaatgc
huSox4-SalI	gggggtcgactcagtaggtgaaaaccaggt
C-terminal V5 tagged TMEM2	
Mutagenic primers to fix TMEM2 C-terminal frame-shift point mutation present in the ORFeome library	
t3928g_antisense	gatgagctggtttggctaatcccagaggtactagt
t3928g_sense	actagtacctctgggattagccaaaccagctcatc
Gateway cloning to pDONR221 vector	
12attB1_TMEM2	aaaaagcaggcttcaccatgtatgccactgattccag
12attB2_TMEM2_C-term	agaaagctgggtgcaaatgtgcttttgaagctt
Universal attB1_adapter	ggggacaagtttgtacaaaaaagcaggct
Universal attB2_adapter	ggggaccactttgtacaagaaagctgggt
Clone TMEM2 promoter region into pGL3 Promoter Luciferase vector	
MluI_-1kb_F	gatacgcgtacacctttggaccatgaaaaca
XhoI_-1kb_R	gatgctcgagatccgggccgtcgtcatc
Clone individual putative SOX4 motifs in the promtoer region of TMEM2 into pGL3 Promoter Lucifease vector	
MluI_-1kb_#3_F	gatacgcgtttttgtcccttccatcttgtt
XhoI_-1kb_#3_R	gatgctcgagggaggaaaggtttttaaaataagg
MluI_-1kb_#2_F	gatacgcgtccaagatttggaaacagtaatttctta
XhoI_-1kb_#2_R	gatgctcgagcctatttcaagggtattagtttagga
MluI_-1kb_#1_F	gatacgcgtgtcgcggttacaggactgag
XhoI_-1kb_#1_R	gatgctcgagcagggttgttcattcctcct

BIBLIOGRAPHY

Acharyya, S., T. Oskarsson, S. Vanharanta, S. Malladi, J. Kim, P.G. Morris, K. Manova-Todorova, M. Leversha, N. Hogg, V.E. Seshan, L. Norton, E. Brogi, and J. Massagué. 2012. A CXCL1 Paracrine Network Links Cancer Chemoresistance and Metastasis. *Cell*. 150:165-178. doi:10.1016/j.cell.2012.04.042.

Alarcón, C.R., H. Lee, H. Goodarzi, N. Halberg, and S.F. Tavazoie. 2015. N6-methyladenosine marks primary microRNAs for processing. *Nature*. 1–16. doi:10.1038/nature14281.

Atchley, D.P., C.T. Albarracin, A. Lopez, V. Valero, C.I. Amos, A.M. Gonzalez-Angulo, G.N. Hortobagyi, and B.K. Arun. 2008. Clinical and Pathologic Characteristics of Patients With BRCA-Positive and BRCA-Negative Breast Cancer. *Journal of Clinical Oncology*. 26:4282-4288. doi:10.1200/JCO.2008.16.6231.

Aue, G., Y. Du, S.M. Cleveland, S.B. Smith, U.P. Davé, D. Liu, M.A. Weniger, J.Y. Metais, N.A. Jenkins, N.G. Copeland, and C.E. Dunbar. 2011. Sox4 cooperates with PU.1 haploinsufficiency in murine myeloid leukemia. *Blood*. 118:4674-4681. doi:10.1182/blood-2011-04-351528.

Auyeung, V.C., I. Ulitsky, S.E. McGeary, and D.P. Bartel. 2013. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*. 152:844-858. doi:10.1016/j.cell.2013.01.031.

Bartel, D.P. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*. 116:281-297. doi:10.1016/S0092-8674(04)00045-5.

Batistatou, A., A. Charalabopoulos, and K. Charalabopoulos. 2009. Molecular basis of metastasis. *N. Engl. J. Med.* 360:1679-author reply 1679-80. doi:10.1056/NEJMc090143.

Bergsland, M., M. Werme, M. Malewicz, T. Perlmann, and J. Muhr. 2006. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes & development*. 20:3475-3486. doi:10.1101/gad.403406.

Blanpain, C. 2013. Tracing the cellular origin of cancer. 1-9. doi:10.1038/ncb2657.

Bokar, J.A., M.E. Shambaugh, D. Polayes, A.G. Matera, and F.M. Rottman. 1997. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N^6-adenosine)-methyltransferase. *RNA (New York, N.Y.)*. 3:1233-1247.

Bowles, J., G. Schepers, and P. Koopman. 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Developmental biology*. 227:239-255. doi:10.1006/dbio.2000.9883.

Buckingham, M., S. Meilhac, and S. Zaffran. 2005. Building the mammalian heart from two sources of myocardial cells. *Nature reviews. Genetics*. 6:826-837. doi:10.1038/nrg1710.

Bujnicki, J.M., M. Feder, M. Radlinska, and R.M. Blumenthal. 2002. Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. *J. Mol. Evol.* 55:431-444. doi:10.1007/s00239-002-2339-8.

Cao, J., J. Cai, D. Huang, Q. Han, Q. Yang, T. Li, H. Ding, and Z. Wang. 2013. miR-335 represents an invasion suppressor gene in ovarian cancer by targeting Bcl-w. *Oncology reports*. 30:701-706. doi:10.3892/or.2013.2482.

Cao, Y., C.Y. Slaney, B.N. Bidwell, B.S. Parker, C.N. Johnstone, J. Rautela, B.L. Eckhardt, and R.L. Anderson. 2014. BMP4 inhibits breast cancer metastasis by blocking myeloid-derived suppressor cell activity. *Cancer research*. 74:5091-5102. doi:10.1158/0008-5472.CAN-13-3171.

Carey, L.A., C.M. Perou, C.A. Livasy, L.G. Dressler, D. Cowan, K. Conway, G. Karaca, M.A. Troester, C.K. Tse, S. Edmiston, S.L. Deming, J. Geradts, M.C.U. Cheang, T.O. Nielsen, P.G. Moorman, H.S. Earp, and R.C. Millikan. 2006. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 295:2492-2502. doi:10.1001/jama.295.21.2492.

Chambers, A.F., A.C. Groom, and I.C. MacDonald. 2002. Metastasis: Dissemination and growth of cancer cells in metastatic sites. *Nature reviews. Cancer*. 2:563-572. doi:10.1038/nrc865.

Chen, H., R. Miao, J. Fan, Z. Han, J. Wu, G. Qiu, H. Tang, and Z. Peng. 2013. Decreased expression of miR-126 correlates with metastatic recurrence of hepatocellular carcinoma. *Clin. Exp. Metastasis*. 30:651-658. doi:10.1007/s10585-013-9569-6.

Chi, S.W., J.B. Zang, A. Mele, and R.B. Darnell. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. 460:479-486. doi:10.1038/nature08170.

Clohessy, J.G., and P.P. Pandolfi. 2009. β-tting on p63 as a Metastatic Suppressor. *Cell*. 137:28-30. doi:10.1016/j.cell.2009.03.028.

de Almeida, S.F., and M. Carmo-Fonseca. 2012. Design principles of interconnections between chromatin and pre-mRNA splicing. *Trends in*

Biochemical Sciences. 37:248-253. doi:10.1016/j.tibs.2012.02.002.

de Bont, J.M., J.M. Kros, M.M.C.J. Passier, R.E. Reddingius, P.A.E. Sillevis Smitt, T.M. Luider, M.L. den Boer, and R. Pieters. 2008. Differential expression and prognostic significance of SOX genes in pediatric medulloblastoma and ependymoma identified by microarray analysis. *Neuro-oncology*. 10:648-660. doi:10.1215/15228517-2008-032.

Denli, A.M., B.B.J. Tops, R.H.A. Plasterk, R.F. Ketting, and G.J. Hannon. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature*. 432:231-235. doi:10.1038/nature03049.

Dominissini, D., S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, and G. Rechavi. 2012. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 485:201-206. doi:10.1038/nature11112.

Du, C., Z. Lv, L. Cao, C. Ding, O.-A.K. Gyabaah, H. Xie, L. Zhou, J. Wu, and S. Zheng. 2014. MiR-126-3p suppresses tumor metastasis and angiogenesis of hepatocellular carcinoma by targeting LRP6 and PIK3R2. *Journal of translational medicine*. 12:259. doi:10.1186/s12967-014-0259-1.

Dy, P., A. Penzo-Méndez, H. Wang, C.E. Pedraza, W.B. Macklin, and V. Lefebvre. 2008. The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties. *Nucleic acids research*. 36:3101-3117. doi:10.1093/nar/gkn162.

Eggers, S., T. Ohnesorg, and A. Sinclair. 2014. Genetic regulation of mammalian gonad development. *Nat Rev Endocrinol*. 10:673-683. doi:10.1038/nrendo.2014.163.

Elemento, O., N. Slonim, and S. Tavazoie. 2007. A universal framework for regulatory element discovery across all genomes and data types. *Molecular cell*. 28:337-350. doi:10.1016/j.molcel.2007.09.027.

Ell, B., and Y. Kang. 2013. Transcriptional control of cancer metastasis. *Trends in cell biology*. 23:603-611. doi:10.1016/j.tcb.2013.06.001.

Farr, C.J., D.J. Easty, J. Ragoussis, J. Collignon, R. Lovell-Badge, and P.N. Goodfellow. 1993. Characterization and mapping of the human SOX4 gene. *Mammalian Genome*. 4:577-584. doi:10.1007/BF00361388.

Fidler, I.J. 1973. Selection of successive tumour lines for metastasis. *Nature: New biology*. 242:148-149.

Fidler, I.J. 2003. Timeline - The pathogenesis of cancer metastasis: the “seed and soil” hypothesis revisited. *Nature reviews. Cancer*. 3:453-458. doi:10.1038/nrc1098.

Frierson, H.F., A.K. El-Naggar, J.B. Welsh, L.M. Sapinoso, A.I. Su, J. Cheng, T. Saku, C.A. Moskaluk, and G.M. Hampton. 2002. Large scale molecular analysis identifies genes with altered expression in salivary adenoid cystic carcinoma. *The American journal of pathology*. 161:1315-1323. doi:10.1016/S0002-9440(10)64408-2.

Fustin, J.M., M. Doi, Y. Yamaguchi, H. Hida, S. Nishimura, M. Yoshida, T. Isagawa, M.S. Morioka, H. Kakeya, I. Manabe, and H. Okamura. 2013. RNA-methylation-dependent RNA processing controls the speed of the circadian clock. *Cell*. 155:793-806. doi:10.1016/j.cell.2013.10.026.

Gao, L., Y. Yang, H. Xu, R. Liu, D. Li, H. Hong, M. Qin, and Y. Wang. 2014. MiR-335 functions as a tumor suppressor in pancreatic cancer by targeting OCT4. *Tumour Biol.* 35:8309-8318. doi:10.1007/s13277-014-2092-9.

Goldman, D.C., N. Donley, and J.L. Christian. 2009. Genetic interaction between Bmp2 and Bmp4 reveals shared functions during multiple aspects of mouse organogenesis. *Mechanisms of Development*. 126:117-127. doi:10.1016/j.mod.2008.11.008.

Gregory, R.I., K.-P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature*. 432:235-240. doi:10.1038/nature03120.

Guo, J., H. Cheng, S. Zhao, and L. Yu. 2006. GG: A domain involved in phage LTF apparatus and implicated in human MEB and non-syndromic hearing loss diseases. *FEBS Letters*. 580:581-584. doi:10.1016/j.febslet.2005.12.076.

Gupta, G.P., A.J. Minn, Y. Kang, P.M. Siegel, I. Serganova, C. Cordon-Cardo, A.B. Olshen, W.L. Gerald, and J. Massague. 2005. Identifying site-specific metastasis genes and functions. *Cold Spring Harbor symposia on quantitative biology*. 70:149-158. doi:10.1101/sqb.2005.70.018.

Gupta, G.P., and J. Massagué. 2006. Cancer metastasis: building a framework. *Cell*. 127:679-695. doi:10.1016/j.cell.2006.11.001.

Haghgoo, S.M., A. Allameh, E. Mortaz, J. Garssen, G. Folkerts, P.J. Barnes, and I.M. Adcock. 2015. Pharmacogenomics and targeted therapy of cancer: Focusing on non-small cell lung cancer. *Eur. J. Pharmacol.* 754:82-91. doi:10.1016/j.ejphar.2015.02.029.

Han, J., Y. Lee, K.-H. Yeom, J.-W. Nam, I. Heo, J.-K. Rhee, S.Y. Sohn, Y. Cho,

B.-T. Zhang, and V.N. Kim. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*. 125:887-901. doi:10.1016/j.cell.2006.03.043.

Han, J., Y. Lee, K.-H. Yeom, Y.-K. Kim, H. Jin, and V.N. Kim. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & development*. 18:3016-3027. doi:10.1101/gad.1262504.

He, Q.Y., X.H. Liu, Q. Li, D.J. Studholme, X.W. Li, and S.P. Liang. 2006. G8: a novel domain associated with polycystic kidney disease and non-syndromic hearing loss. *Bioinformatics (Oxford, England)*. 22:2189-2191. doi:10.1093/bioinformatics/btl123.

Higuchi, T., T. Nakayama, T. Arao, K. Nishio, and O. Yoshie. 2013. SOX4 is a direct target gene of FRA-2 and induces expression of HDAC8 in adult T-cell leukemia/lymphoma. *Blood*. 121:3640-3649. doi:10.1182/blood-2012-07-441022.

Hoser, M., M.R. Potzner, J.M.C. Koch, M.R. Bösl, M. Wegner, and E. Sock. 2008. Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors. *Molecular and cellular biology*. 28:4675-4687. doi:10.1128/MCB.00338-08.

Huang, H.Y., Y.Y. Cheng, W.C. Liao, Y.W. Tien, C.H. Yang, S.M. Hsu, and P.H. Huang. 2012. SOX4 transcriptionally regulates multiple SEMA3/plexin family members and promotes tumor growth in pancreatic cancer. *PloS one*. 7:e48637. doi:10.1371/journal.pone.0048637.

Ikushima, H., T. Todo, Y. Ino, M. Takahashi, K. Miyazawa, and K. Miyazono. 2009. Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell stem cell*. 5:504-514. doi:10.1016/j.stem.2009.08.018.

Jolma, A., J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R. Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale. 2013. DNA-binding specificities of human transcription factors. *Cell*. 152:327-339. doi:10.1016/j.cell.2012.12.009.

Kang, Y., P.M. Siegel, W. Shu, M. Drobnjak, S.M. Kakonen, C. Cordón-Cardo, T.A. Guise, and J. Massagué. 2003. A multigenic program mediating breast cancer metastasis to bone. *Cancer cell*. 3:537-549. doi:10.1016/S1535-6108(03)00132-6.

Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler. 2002. The human genome browser at UCSC. *Genome research*.

12:996-1006. doi:10.1101/gr.229102.

Kim, S.W., Z. Li, P.S. Moore, A.P. Monaghan, Y. Chang, M. Nichols, and B. John. 2010. A sensitive non-radioactive northern blot method to detect small RNAs. *Nucleic acids research*. 38:e98-e98. doi:10.1093/nar/gkp1235.

Koboldt, D.C., R.S. Fulton, M.D. McLellan, H. Schmidt, J. Kalicki-Veizer, J.F. McMichael, L.L. Fulton, D.J. Dooling, L. Ding, E.R. Mardis, R.K. Wilson, A. Ally, M. Balasundaram, Y.S.N. Butterfield, R. Carlsen, C. Carter, A. Chu, E. Chuah, H.-J.E. Chun, R.J.N. Coope, N. Dhalla, R. Guin, C. Hirst, M. Hirst, R.A. Holt, D. Lee, H.I. Li, M. Mayo, R.A. Moore, A.J. Mungall, E. Pleasance, A. Gordon Robertson, J.E. Schein, A. Shafiei, P. Sipahimalani, J.R. Slobodan, D. Stoll, A. Tam, N. Thiessen, R.J. Varhol, N. Wye, T. Zeng, Y. Zhao, I. Birol, S.J.M. Jones, M.A. Marra, A.D. Cherniack, G. Saksena, R.C. Onofrio, N.H. Pho, S.L. Carter, S.E. Schumacher, B. Tabak, B. Hernandez, J. Gentry, H. Nguyen, A. Crenshaw, K. Ardlie, R. Beroukhim, W. Winckler, G. Getz, S.B. Gabriel, M. Meyerson, L. Chin, P.J. Park, R. Kucherlapati, K.A. Hoadley, J. Todd Auman, C. Fan, Y.J. Turman, Y. Shi, L. Li, M.D. Topal, X. He, H.-H. Chao, A. Prat, G.O. Silva, M.D. Iglesia, W. Zhao, J. Usary, J.S. Berg, M. Adams, J. Booker, J. Wu, A. Gulabani, T. Bodenheimer, A.P. Hoyle, J.V. Simons, M.G. Soloway, L.E. Mose, S.R. Jefferys, S. Balu, J.S. Parker, D. Neil Hayes, C.M. Perou, S. Malik, S. Mahurkar, H. Shen, et al. 2012. Comprehensive molecular portraits of human breast tumours. *Nature*. 490:61-70. doi:10.1038/nature11412.

Kuo, T.H., T. Kubota, M. Watanabe, T. Furukawa, T. Teramoto, K. Ishibiki, M. Kitajima, A.R. Moossa, S. Penman, and R.M. Hoffman. 1995. Liver colonization competence governs colon cancer metastasis. *Proceedings of the National Academy of Sciences of the United States of America*. 92:12085-12089.

Kuwahara, M., M. Yamashita, K. Shinoda, S. Tofukuji, A. Onodera, R. Shinnakasu, S. Motohashi, H. Hosokawa, D. Tumes, C. Iwamura, V. Lefebvre, and T. Nakayama. 2012. The transcription factor Sox4 is a downstream target of signaling by the cytokine TGF-[beta] and suppresses TH2 differentiation. *Nat. Immunol.* 13:778-786. doi:10.1038/ni.2362.

Lai, Y.H., J. Cheng, D. Cheng, M.E. Feasel, K.D. Beste, J. Peng, A. Nusrat, and C.S. Moreno. 2011. SOX4 interacts with plakoglobin in a Wnt3a-dependent manner in prostate cancer cells. *BMC cell biology*. 12:50. doi:10.1186/1471-2121-12-50.

Landthaler, M., A. Yalcin, and T. Tuschl. 2004. The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr. Biol.* 14:2162-2167. doi:10.1016/j.cub.2004.11.001.

Lee, C.J., V.J. Appleby, A.T. Orme, W.I. Chan, and P.J. Scotting. 2002. Differential expression of SOX4 and SOX11 in medulloblastoma. *Journal of neuro-oncology*. 57:201-214.

Lee, Y., and V.N. Kim. 2007. In Vitro and In Vivo Assays for the Activity of Drosha Complex. *In* MicroRNA Methods. Elsevier. 87-106.

Li, N., A. Tang, S. Huang, Z. Li, X. Li, S. Shen, J. Ma, and X. Wang. 2013. MiR-126 suppresses colon cancer cell proliferation and invasion via inhibiting RhoA/ROCK signaling pathway. *Mol. Cell. Biochem.* 380:107-119. doi:10.1007/s11010-013-1664-0.

Liang, Y., H. Wu, R. Lei, R.A. Chong, Y. Wei, X. Lu, I. Tagkopoulos, S.-Y. Kung, Q. Yang, G. Hu, and Y. Kang. 2012. Transcriptional network analysis identifies BACH1 as a master regulator of breast cancer bone metastasis. *The Journal of biological chemistry*. 287:33533-33544. doi:10.1074/jbc.M112.392332.

Liao, Y.L., Y.M. Sun, G.Y. Chau, Y.P. Chau, T.C. Lai, J.L. Wang, J.T. Horng, M. Hsiao, and A.P. Tsou. 2008. Identification of SOX4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of SOX4 potentiates metastasis in hepatocellular carcinoma. *Oncogene*. 27:5578-5589. doi:10.1038/onc.2008.168.

Lin, B., A. Madan, J.-G. Yoon, X. Fang, X. Yan, T.-K. Kim, D. Hwang, L. Hood, and G. Foltz. 2010a. Massively parallel signature sequencing and bioinformatics analysis identifies up-regulation of TGFBI and SOX4 in human glioblastoma. *PloS one*. 5:e10210. doi:10.1371/journal.pone.0010210.

Lin, L., V.M. Lee, Y. Wang, J.S. Lin, E. Sock, M. Wegner, and L. Lei. 2010b. Sox11 regulates survival and axonal growth of embryonic sensory neurons. *Dev. Dyn.* 240:52-64. doi:10.1002/dvdy.22489.

Lipson, K.E., and R. Baserga. 1989. Transcriptional activity of the human thymidine kinase gene determined by a method using the polymerase chain reaction and an intron-specific probe. *Proceedings of the National Academy of Sciences of the United States of America*. 86:9774-9777.

Liu, P., S. Ramachandran, M. Ali Seyed, C.D. Scharer, N. Laycock, W.B. Dalton, H. Williams, S. Karanam, M.W. Datta, D.L. Jaye, and C.S. Moreno. 2006. Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells. *Cancer research*. 66:4011-4019. doi:10.1158/0008-5472.CAN-05-3055.

Loo, J.M., A. Scherl, A. Nguyen, F.Y. Man, E. Weinberg, Z. Zeng, L. Saltz, P.B.

Paty, and S.F. Tavazoie. 2015. Extracellular Metabolic Energetics Can Promote Cancer Progression. *Cell*. 1-14. doi:10.1016/j.cell.2014.12.018.

Luco, R.F., and T. Misteli. 2011. More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. *Current Opinion in Genetics & Development*. 21:366-372. doi:10.1016/j.gde.2011.03.004.

Maria E Wilson, K.Y.Y.A.K.J.L.Y.K.F.C.L.J.W.C.M.V.E.H.C.C.M.S.G. 2005. The HMG Box Transcription Factor Sox4 Contributes to the Development of the Endocrine Pancreas. *DIABETES*. 54.

Medina, P.P., S.D. Castillo, S. Blanco, M. Sanz-Garcia, C. Largo, S. Alvarez, J. Yokota, A. Gonzalez-Neira, J. Benitez, H.C. Clevers, J.C. Cigudosa, P.A. Lazo, and M. Sanchez-Cespedes. 2009. The SRY-HMG box gene, SOX4, is a target of gene amplification at chromosome 6p in lung cancer. *Human molecular genetics*. 18:1343-1352. doi:10.1093/hmg/ddp034.

Meyer, K.D., Y. Saletore, P. Zumbo, O. Elemento, C.E. Mason, and S.R. Jaffrey. 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 149:1635-1646. doi:10.1016/j.cell.2012.05.003.

Minn, A.J., G.P. Gupta, P.M. Siegel, P.D. Bos, W. Shu, D.D. Giri, A. Viale, A.B. Olshen, W.L. Gerald, and J. Massague. 2005. Genes that mediate breast cancer metastasis to lung. *Nature*. 436:518-524. doi:10.1038/nature03799.

Monteagudo, C., A. Pellín-Carcelén, J.M. Martín, and D. Ramos. 2011. Role of Chemokines in Melanoma Progression. *Actas Dermo-Sifiliográficas (English Edition)*. 102:498-504. doi:10.1016/j.adengl.2011.03.004.

Montes, M., S. Becerra, M. Sánchez-Álvarez, and C. Suñé. 2012. Functional coupling of transcription and splicing. *Gene*. 501:104-117. doi:10.1016/j.gene.2012.04.006.

Mukherjee, D., and J. Zhao. 2013. The Role of chemokine receptor CXCR4 in breast cancer metastasis. *Am J Cancer Res*. 3:46-57.

Nicoloso, M.S., R. Spizzo, M. Shimizu, S. Rossi, and G.A. Calin. 2009. MicroRNAs--the micro steering wheel of tumour metastases. *Nature reviews. Cancer*. 9:293-302. doi:10.1038/nrc2619.

Nicolson, G.L. 1988. Organ specificity of tumor metastasis: role of preferential adhesion, invasion and growth of malignant cells at specific secondary sites. *Cancer metastasis reviews*. 7:143-188. doi:10.1007/BF00046483.

Nicolson, G.L., and K.M. Dulski. 1986. Organ specificity of metastatic tumor colonization is related to organ-selective growth properties of malignant cells. *International journal of cancer. Journal international du cancer*. 38:289-294.

Nissen-Meyer, L.S., R. Jemtland, V.T. Gautvik, M.E. Pedersen, R. Paro, D. Fortunati, D.D. Pierroz, V.A. Stadelmann, S. Reppe, F.P. Reinholt, A. Del Fattore, N. Rucci, A. Teti, S. Ferrari, and K.M. Gautvik. 2007. Osteopenia, decreased bone formation and impaired osteoblast development in Sox4 heterozygous mice. *Journal of cell science*. 120:2785-2795. doi:10.1242/jcs.003855.

Omidvar, N., M.L. Maunakea, L. Jones, S. Sevcikova, Bin Yin, K.L. Himmel, T.R. Tennant, M.M. Le Beau, D.A. Largaespada, and S.C. Kogan. 2013. PML-RARα co-operates with Sox4 in acute myeloid leukemia development in mice. *Haematologica*. 98:424-427. doi:10.3324/haematol.2011.057067.

Oskarsson, T., S. Acharyya, X.H.-F. Zhang, S. Vanharanta, S.F. Tavazoie, P.G. Morris, R.J. Downey, K. Manova-Todorova, E. Brogi, and J. Massagué. 2011. Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nature medicine*. 17:867-874. doi:10.1038/nm.2379.

Paget, S. 1989. The distribution of secondary growths in cancer of the breast. 1889. 8. 4 pp.

Pauli, B.U., H.G. Augustin-Voss, M.E. El-Sabban, R.C. Johnson, and D.A. Hammer. 1990. Organ-preference of metastasis. *Cancer metastasis reviews*. 9:175-189. doi:10.1007/BF00046359.

Pavlidis, P., and W.S. Noble. 2003. Matrix2png: a utility for visualizing matrix data. *Bioinformatics (Oxford, England)*. 19:295-296.

Pencheva, N., H. Tran, C. Buss, D. Huh, M. Drobnjak, K. Busam, and S.F. Tavazoie. 2012. Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. *Cell*. 151:1068-1082. doi:10.1016/j.cell.2012.10.028.

Piccirillo, S.G.M., B.A. Reynolds, N. Zanetti, G. Lamorte, E. Binda, G. Broggi, H. Brem, A. Olivi, F. Dimeco, and A.L. Vescovi. 2006. Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature*. 444:761-765. doi:10.1038/nature05349.

Png, K.J., M. Yoshida, X.H.-F. Zhang, W. Shu, H. Lee, A. Rimner, T.A. Chan, E. Comen, V.P. Andrade, S.W. Kim, T.A. King, C.A. Hudis, L. Norton, J. Hicks, J. Massagué, and S.F. Tavazoie. 2011. MicroRNA-335 inhibits tumor reinitiation and is silenced through genetic and epigenetic mechanisms in human breast cancer. *Genes & development*. 25:226-231. doi:10.1101/gad.1974211.

Png, K.J., N. Halberg, M. Yoshida, and S.F. Tavazoie. 2012. A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. *Nature*. 481:190-194. doi:10.1038/nature10661.

Pollard, K.S., M.J. Hubisz, K.R. Rosenbloom, and A. Siepel. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research*. 20:110-121. doi:10.1101/gr.097857.109.

Polyak, K. 2007. Breast cancer: origins and evolution. *J. Clin. Invest.* 117:3155-3163. doi:10.1172/JCI33295.

Qin, J., M. Li, P. Wang, M. Zhang, and J. Wang. 2011. ChIP-Array: combinatory analysis of ChIP-seq/chip and microarray gene expression data to discover direct/indirect targets of a transcription factor. *Nucleic acids research*.

Ramezani-Rad, P., H. Geng, C. Hurtz, L.N. Chan, Z. Chen, H. Jumaa, A. Melnick, E. Paietta, W.L. Carroll, C.L. Willman, V. Lefebvre, and M. Muschen. 2013. SOX4 enables oncogenic survival signals in acute lymphoblastic leukemia. *Blood*. 121:148-155. doi:10.1182/blood-2012-05-428938.

Rinker-Schaeffer, C.W., D.R. Welch, and M. Sokoloff. 2000. Defining the biologic role of genes that regulate prostate cancer metastasis. *Curr Opin Urol*. 10:397-401.

Sandoval, S., C. Kraus, E.-C. Cho, M. Cho, J. Bies, E. Manara, B. Accordi, E.M. Landaw, L. Wolff, M. Pigazzi, and K.M. Sakamoto. 2012. Sox4 cooperates with CREB in myeloid transformation. *Blood*. 120:155-165. doi:10.1182/blood-2011-05-357418.

Scharer, C.D., C.D. McCabe, M. Ali Seyed, M.F. Berger, M.L. Bulyk, and C.S. Moreno. 2009. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. *Cancer research*. 69:709-717. doi:10.1158/0008-5472.CAN-08-3415.

Schilham, M.W., M.A. Oosterwegel, P. Moerer, J. Ya, P.A. de Boer, M. van de Wetering, S. Verbeek, W.H. Lamers, A.M. Kruisbeek, A. Cumano, and H. Clevers. 1996. Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature*. 380:711-714. doi:10.1038/380711a0.

Schilham, M.W., P. Moerer, A. Cumano, and H.C. Clevers. 1997. Sox-4 facilitates thymocyte differentiation. *European Journal of Immunology*. 27:1292-1295. doi:10.1002/eji.1830270534.

Shukla, S., and S. Oberdoerffer. 2012. Co-transcriptional regulation of alternative pre-mRNA splicing. *BBA - Gene Regulatory Mechanisms*. 1819:673-683.

doi:10.1016/j.bbagrm.2012.01.014.

Sibbritt, T., H.R. Patel, and T. Preiss. 2013. Mapping and significance of the mRNA methylome. *Wiley interdisciplinary reviews. RNA*. 4:397-422. doi:10.1002/wrna.1166.

Sinner, D., J.J. Kordich, J.R. Spence, R. Opoka, S. Rankin, S.-C.J. Lin, D. Jonatan, A.M. Zorn, and J.M. Wells. 2007. Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells. *Molecular and cellular biology*. 27:7802-7815. doi:10.1128/MCB.02179-06.

Smith, K.A., A.K. Lagendijk, A.D. Courtney, H. Chen, S. Paterson, B.M. Hogan, C. Wicking, and J. Bakkers. 2011. Transmembrane protein 2 (Tmem2) is required to regionally restrict atrioventricular canal boundary and endocardial cushion development. *Development*. 138:4193-4198. doi:10.1242/dev.065375.

Sock, E., S.D. Rettig, J. Enderich, M.R. Bösl, E.R. Tamm, and M. Wegner. 2004. Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. *Molecular and cellular biology*. 24:6635-6644. doi:10.1128/MCB.24.15.6635-6644.2004.

Stahlhut, C., Y. Suárez, J. Lu, Y. Mishima, and A.J. Giraldez. 2012. miR-1 and miR-206 regulate angiogenesis by modulating VegfA expression in zebrafish. *Development*. 139:4356-4364. doi:10.1242/dev.083774.

Talmadge, J.E., and I.J. Fidler. 1982. Cancer Metastasis Is Selective or Random Depending on the Parent Tumor Population. *Nature*. 297:593-594. doi:10.1038/297593a0.

Tavazoie, S.F., C. Alarcón, T. Oskarsson, D. Padua, Q. Wang, P.D. Bos, W.L. Gerald, and J. Massagué. 2008. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature*. 451:147-152. doi:10.1038/nature06487.

Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics*. 14:178-192. doi:10.1093/bib/bbs017.

Totong, R., T. Schell, F. Lescroart, L. Ryckebusch, Y.F. Lin, T. Zygmunt, L. Herwig, A. Krudewig, D. Gershoony, H.G. Belting, M. Affolter, J. Torres-Vazquez, and D. Yelon. 2011. The novel transmembrane protein Tmem2 is essential for coordination of myocardial and endocardial morphogenesis. *Development*. 138:4199-4205. doi:10.1242/dev.064261.

Turner, N.C., and J.S. Reis-Filho. 2006. Basal-like breast cancer and the BRCA1 phenotype. *Oncogene*. 25:5846-5853. doi:10.1038/sj.onc.1209876.

Vandewetering, M., M. Oosterwegel, K. Vannorren, and H. Clevers. 1993. Sox-4, an Sry-Like Hmg Box Protein, Is a Transcriptional Activator in Lymphocytes. *EMBO J.* 12:3847-3854.

Ventura, A., and T. Jacks. 2009. MicroRNAs and Cancer: Short RNAs Go a Long Way. *Cell*. 136:586-591. doi:10.1016/j.cell.2009.02.005.

Vervoort, S.J., A.R. Lourenco, R. van Boxtel, and P.J. Coffer. 2013a. SOX4 mediates TGF-beta-induced expression of mesenchymal markers during mammary cell epithelial to mesenchymal transition. *PloS one*. 8:e53238. doi:10.1371/journal.pone.0053238.

Vervoort, S.J., R. van Boxtel, and P.J. Coffer. 2013b. The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? *Oncogene*. 32:3397-3409. doi:10.1038/onc.2012.506.

Wang, J., E. Tsouko, P. Jonsson, J. Bergh, J. Hartman, E. Aydogdu, and C. Williams. 2014a. miR-206 inhibits cell migration through direct targeting of the actin-binding protein coronin 1C in triple-negative breast cancer. *Mol Oncol*. 8:1690-1702. doi:10.1016/j.molonc.2014.07.006.

Wang, X., Z. Lu, A. Gomez, G.C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan, and C. He. 2014b. N^6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*. 505:117-120. doi:10.1038/nature12730.

Wang, Y., W. Zhao, and Q. Fu. 2013. miR-335 suppresses migration and invasion by targeting ROCK1 in osteosarcoma cells. *Mol. Cell. Biochem.* 384:105-111. doi:10.1007/s11010-013-1786-4.

Wang, Y., Y. Li, J.I. Toth, M.D. Petroski, Z. Zhang, and J.C. Zhao. 2014c. N^6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature cell biology*. 16:191-198. doi:10.1038/ncb2902.

Wei, C.M., and B. Moss. 1977. Nucleotide sequences at the N^6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry*. 16:1672-1676.

Weigelt, B., F.C. Geyer, and J.S. Reis-Filho. 2010. Histological types of breast cancer: How special are they? *Mol Oncol*. 4:192-208. doi:10.1016/j.molonc.2010.04.004.

Weinmann, A.S. 2004. Novel ChIP-based strategies to uncover transcription factor target genes in the immune system. *Nat. Rev. Immunol.* 4:381-386.

doi:10.1038/nri1353.

Wolfer, A., and S. Ramaswamy. 2011. MYC and metastasis. *Cancer research*. 71:2034-2037. doi:10.1158/0008-5472.CAN-10-3776.

Ya, J., M.W. Schilham, P.A.J. de Boer, A.F.M. Moorman, H. Clevers, and W.H. Lamers. 1998. Sox4-Deficiency Syndrome in Mice Is an Animal Model for Common Trunk. *Circ. Res.* 83:986-994. doi:10.1161/01.res.83.10.986.

Yang, X., J.S. Boehm, K. Salehi-Ashtiani, T. Hao, Y. Shen, R. Lubonja, S.R. Thomas, O. Alkan, T. Bhimdi, T.M. Green, C.M. Johannessen, S.J. Silver, C. Nguyen, R.R. Murray, H. Hieronymus, D. Balcha, C. Fan, C. Lin, L. Ghamsari, M. Vidal, W.C. Hahn, D.E. Hill, and D.E. Root. 2011. A public genome-scale lentiviral expression library of human ORFs. *Nature methods*. 8:659-661. doi:10.1038/nmeth.1638.

Yoneda, T., P.J. Williams, T. Hiraga, M. Niewolna, and R. Nishimura. 2001. A bone-seeking clone exhibits different biological properties from the MDA-MB-231 parental human breast cancer cells and a brain-seeking clone in vivo and in vitro. *J. Bone Miner. Res.* 16:1486-1495. doi:10.1359/jbmr.2001.16.8.1486.

Zhang, B., X. Pan, G.P. Cobb, and T.A. Anderson. 2007. microRNAs as oncogenes and tumor suppressors. *Developmental biology*. 302:1-12. doi:10.1016/j.ydbio.2006.08.028.

Zhang, C., and R.B. Darnell. 2011. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature biotechnology*. 29:607-614. doi:10.1038/nbt.1873.

Zhang, H., M. Alberich-Jorda, G. Amabile, H. Yang, P.B. Staber, A. Diruscio, R.S. Welner, A. Ebralidze, J. Zhang, E. Levantini, V. Lefebvre, P.J. Valk, R. Delwel, M. Hoogenkamp, C. Nerlov, J. Cammenga, B. Saez, D.T. Scadden, C. Bonifer, M. Ye, and D.G. Tenen. 2013a. Sox4 is a key oncogenic target in C/EBPalpha mutant acute myeloid leukemia. *Cancer cell*. 24:575-588. doi:10.1016/j.ccr.2013.09.018.

Zhang, L., L. Xia, L. Zhao, Z. Chen, X. Shang, J. Xin, M. Liu, X. Guo, K. Wu, Y. Pan, and D. Fan. 2015. Activation of PAX3-MET pathways due to miR-206 loss promotes gastric cancer metastasis. *Carcinogenesis*. 36:390-399. doi:10.1093/carcin/bgv009.

Zhang, L., X. Liu, H. Jin, X. Guo, L. Xia, Z. Chen, M. Bai, J. Liu, X. Shang, K. Wu, Y. Pan, and D. Fan. 2013b. MiR-206 inhibits gastric cancer proliferation in part by repressing CyclinD2. *Cancer Letters*. 332:94-101. doi:10.1016/j.canlet.2013.01.023.

Zhang, Y., P. Yang, T. Sun, D. Li, X. Xu, Y. Rui, C. Li, M. Chong, T. Ibrahim, L. Mercatali, D. Amadori, X. Lu, D. Xie, Q.-J. Li, and X.-F. Wang. 2013c. miR-126 and miR-126* repress recruitment of mesenchymal stem cells and inflammatory monocytes to inhibit breast cancer metastasis. *Nature cell biology*. 15:284-294. doi:10.1038/ncb2690.

Zhang, Y., X. Wang, B. Xu, B. Wang, Z. Wang, Y. Liang, J. Zhou, J. Hu, and B. Jiang. 2013d. Epigenetic silencing of miR-126 contributes to tumor invasion and angiogenesis in colorectal cancer. *Oncology reports*. 30:1976-1984. doi:10.3892/or.2013.2633.

Zhou, J., Y. Tian, J. Li, B. Lu, M. Sun, Y. Zou, R. Kong, Y. Luo, Y. Shi, K. Wang, and G. Ji. 2013. miR-206 is down-regulated in breast cancer and inhibits cell proliferation through the up-regulation of cyclinD2. *Biochemical and biophysical research communications*. 433:207-212. doi:10.1016/j.bbrc.2013.02.084.