September 2015

Function of Long Noncoding RNAs in Breast Cancer

Edward J. Richards
University of South Florida, erichards@mail.usf.edu

Function of Long Noncoding RNAs in Breast Cancer

by

Edward Richards

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Cell Biology, Microbiology, and Molecular Biology
College of Arts and Sciences
University of South Florida

Major Professor: Jin Q. Cheng, M.D., Ph.D.
Ed Seto, Ph.D.
Jerry Wu, Ph.D.
Domenico Coppola, M.D.
Conor Lynch, Ph.D.

Date of Approval:
June 15, 2015

Keywords: TGFβ, epithelial to mesenchymal transition, metastasis

DEDICATION

This dissertation is written with dedication to my family and friends.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Breast cancer is a disease that will be diagnosed in about 1 in 10 women throughout their lifetime. The majority of breast cancers are originated from the epithelial cells of the mammary ducts, and this occurrence can be due to several factors including hereditary and acquired mutation. There are several major breast cancer subtypes, including estrogen receptor-α (ERα)-positive, HER2-enriched and triple-negative (TNBC). Patients diagnosed with ER+ tumors are generally treated with estrogen blockers (e.g., tamoxifen, letrozole and fulvestrant). Patients with HER2+ tumors are commonly administered with drugs that block HER2 signaling (e.g., trastuzumab) or inhibit HER2's tyrosine kinase activity (e.g., lapatinib). For patients with TNBC, chemotherapies such as taxanes and anthracyclines are standard of care therapies.

However, for each breast cancer subtype, a significant number of patients develop resistance to these therapies and eventually die from metastasis, a process which accounts for ~90% of breast cancer mortality. Currently, metastatic breast cancer is incurable, and the short median survival of 3 years for patients with metastatic breast cancer has not significantly changed in over 20 years. Therefore identification of new molecules that are involved in breast cancer metastasis and development of more

precisely targeted therapeutic strategies are urgently needed to improve the clinical outcome for this disease.

The transforming growth factor pathway beta (TGFβ) pathway has been show to play a key role in metastasis through induction of epithelial-mesenchymal transition (EMT), cell migration and invasion. Over more than a decade, this pathway has been studied across several cancers and it is now better established that it has context-dependent tumor suppressive and oncogenic qualities. In the early stages of breast cancer, TGFβ pathway is a suppressor of benign and early stage tumor growth. However, as disease progresses and corresponding levels of TGFβ ligands become elevated, a "switch" will take place and promote oncogenic phenotypes like EMT and cancer cell stemness which drive metastasis.

Long noncoding RNAs (lncRNAs) are an emerging subclass of RNA molecules in cancer biology. LncRNAs are >200nt and can influence target gene expression locally in "*cis*", or along a distant chromosome in "*trans*", through various mechanisms and interactions with other biological molecules. The contribution of TGFβ-regulated lncRNAs to associated phenotypes like EMT and cancer cell stemness has not been very well studied. The aim of this doctoral dissertation is to address the functional and mechanistic roles of lncRNAs in these processes. Using a well-established TGFβ-induced EMT model (e.g., mouse mammary epithelial cell NMuMG treated with TGFβ), we have identified 3 conserved lncRNAs (lncRNA-HIT, WDFY3-AS2 and TIL) that are significantly upregulated upon TGFβ-induced EMT. They all mediate TGFβ-induced EMT, cell migration and invasion. Overexpression of these lncRNAs is frequently detected during the breast cancer progression and is associated with high grade and

late stage of breast cancer as well as metastatic lesion. We have also demonstrated that lncRNA-HIT positively regulates HOXA13 through "*cis*" mechanism and that WDFY3-AS2 induces WDFY3 and STAT3 expression at mRNA level by direct interaction with hnRNP-R. Interestingly, TIL stimulates C-MYC protein but not mRNA expression by promoting Akt phosphorylation of NF90 leading to its translation from the nucleus to cytosol where NF90 binds to C-MYC mRNA and enhances C-MYC translation. Importantly, we have shown that knockdown of lncRNA-HIT and WDFY3-AS2 significantly reduces breast cancer growth and lung metastasis in orthotopic breast cancer model. These findings indicate that these TGFβ-induced lncRNAs play critical role in EMT, metastasis, and are relevant in human patient tumors. Therefore, it is important to consider utilizing these molecules for clinical applications like diagnosis, monitoring recurrence, predicting a response to therapy, and even as a direct target for therapeutic intervention.

CHAPTER 1

Background

Breast Cancer. In 2014, there will be an estimated 230,000 new cases of breast cancer diagnoses and 40,000 deaths resulting from the disease [1]. Breast cancer can arise in various anatomic locations within the breast. Commonly, cancer will arise in the ducts towards to front of the breast, and less frequently in the lobules that locate closer to the chest wall and the other areas in between. Upon diagnosis breast cancer staging will occur, which accounts for size of primary mass, degree of invasiveness, and presence of metastasis in lymph nodes or other parts of the body. Patients diagnosed in early stages (e.g., stage I or stage II) and who present no local dissemination of cells in lymph node have a very good prognosis with 5-year survival greater than 90%. However, if diagnosed at later stages, 5-year survival drops significantly to 72% at stage III and 22% presenting at stage IV. For this reason, early diagnosis and prevention are critical for improving patient survival.

Several factors have been linked to increasing an individuals risk for developing breast cancer throughout their lifetime including genetic predisposition, atypical menstruation patterns, obesity, high fat diet, and lacking exercise. There is no way to predict who will develop breast cancer, however it is important to be aware of familial history, have regular mammogram screening, and maintain healthy lifestyle to best

prevent disease occurrence. There are 3 standard therapies used to treat breast cancer, surgery, radiation, and chemotherapy [2]. For most early stage cancers, surgical lumpectomy is sufficient treatment which removes a small mass and some normal surrounding tissue to clear borders of any residual cancer cells. If the disease is more advanced, surgery can become more invasive and favor partial or modified radical mastectomy where larger portions or complete breast is removed. External or internal radiation treatments (high energy x-rays) are usually given post-surgery to locally clear the area of any remaining cancer cells. Several chemotherapeutic drugs are now approved for treatment of breast cancer in form of alkylating agents, anthracyclines, platinum drugs and taxanes [2]. The chemotherapeutic treatment chosen will also likely take into account pathological characteristics of tumor as more targeted therapies become available.

Types of Breast Lesions and Carcinomas. Breast cancers originate in the epithelial compartment once increased cellular proliferation and abnormal cellular architecture becomes evident. There is typical progression from normal to pre-cancerous lesion to carcinoma that initially presents itself in the form of a hyperplasia. The term of hyperplasia refers increased number of proliferative cells, and is accompanied by an increase in cell size referred to as hypertrophy. A hyperplasia is a benign lesion, but can develop into a neoplasia or pre-cancerous lesion once genetic abnormalities persist and the cell becomes unresponsive to cues that halt cell cycle and resist apoptosis. The aggressiveness of the lesion can be evaluated by the tumor grade. These features are clinically and pathologically relevant as to how a pre-cancerous

lesion is diagnosed and provide information as to rate of recurrence and how aggressive the carcinoma may become [3].

Ductal carcinoma *in situ* (DCIS) is the most common form of breast tumor associated with development of invasive breast cancer. DCIS is considered stage 0 breast cancer, as this refers to the fact that the mass has yet to become invasive and push past the boundaries of the basement membrane remaining confined to the duct [4]. Depending on how the mass is formed and patterned DCIS can be subdivided into papillary, with finger-like projections, crib-form in which there are gaps between cells, or solid in which gaps are completely filled inside and blocking the duct [5]. Lobular carcinoma *in situ* (LCIS) anatomically presents itself in the lobular compartment of the breast. The lobular component of the breast is the milk producing area located beyond the ducts and closer to the chest wall. LCIS, like DCIS, is confined to the lobules and has not invaded or spread to any surrounding tissues. LCIS is less frequently diagnosed however, as it is hard to view via standard mammogram and no symptoms are present during this stage of pre-cancerous lesion.

Invasive ductal carcinoma (IDC) is the most common type of breast cancer encompassing about 80% of all diagnoses. Advanced DCIS will form IDC once the mass has spread beyond the boundaries of the basement membrane and has infiltrated the surrounding tissue of the breast. Early detection of IDC is critical in that once the tumor protrudes through the duct wall metastasis to the lymph node or other parts of the body can rapidly occur. The majority of cases of women at diagnosis are post-menopausal and close above the age of 55. Invasive lobular carcinomas (ILCs) are much less commonly diagnosed and account for about 10% of all new breast cancer

cases. ILC results once LCIS has spread into the surrounding tissues and areas of the breast surrounding the lobules. It is common for women to be diagnosed with ILC at a later age than IDC and it has also been suggested that hormone replacement therapy can trigger development of this disease.

Hereditary Breast Cancer. Hereditary breast cancer refers to inherited mutations in particular genes that put women at higher risk for developing breast cancer throughout their lifetime and at an earlier age. The most common type of hereditary breast cancer is inheriting mutations in a single copy of either BRCA1 or BRCA2 genes [6]. The mutations in BRCA1 and BRCA2 are referred to as single nucleotide polymorphisms (SNPs), which are naturally occurring variations within the DNA of the genome that differs by one nucleotide within a population. SNPs that locate within the coding region or open reading frame of a protein have the ability to alter amino acid sequence and protein secondary structure which can alter normal gene function and protein stability. BRCA1 and BRCA2 are referred to as tumor suppressor genes, responsible for DNA damage repair and maintaining chromosomal stability and integrity [7]. Patients who inherit a mutation within one of these two genes will have deficient DNA damage repair response and accumulate genetic mutations at a higher rate than those with normal BRCA status. These associated mutations account for the majority of hereditary breast cancer, and about 10% of all new case of diagnosed breast cancers.

There are several other genes that have also been implicated in familial and hereditary breast cancer, however no other gene mutations have showed comparable penetrance rate as that of BRCA1 and BRCA2. Many of the other genes that have

shown susceptibility to breast cancer development are also involved in the DNA damage repair pathway, which include ATM, CHEK2, PALB, RAD50, and MRE11 [8-12].

These genes all play critical roles in DNA damage repair, and can form complex with each other to elicit a robust functional response. Therefore, it is possible that a mutation rendering one of these genes non-functional can result in faulty repair due to dominant-negative effect. Other classical tumor suppressor genes have also been implicated as contributing to hereditary breast cancer, including p53 and PTEN genes [13, 14]. P53 is the most commonly mutated gene in all of human cancer (~50%), and is critical for cell cycle arrest and initiation of apoptosis. PTEN tumor suppressor is the major negative regulator of PI3K/AKT signaling pathway. Loss of PTEN expression and function can occur through various mechanisms, resulting in increased cell survival and proliferation.

Molecular Pathological Subtypes of Breast Cancer. Luminal A and luminal B are the most common types of breast cancer accounting for 40% and 20% of all diagnosed cases, respectively [15]. Luminal breast cancers are named such in that they originate from the epithelial layer of the luminal lining in the mammary duct. Luminal A breast cancer is defined molecularly as estrogen receptor-positive (ER+) and/or progesterone receptor PR-positive (PR+), HER2/neu receptor-negative (HER2-), and a low percentage of p53 mutation (<15%). Luminal A has the most positive prognosis of all subtypes of the breast cancers, a low recurrence rate, and targeted hormone therapies based on ER-positive status [16, 17].

Luminal B cancers are classified same as luminal A in ER+ and/or PR+, however are also HER2+ and have a greater Ki-67 index which is a marker of cellular proliferation. The prognosis for luminal B tumors is poorer than those diagnosed for luminal A. Several factors such as up-regulation of the HER2 receptor, frequent p53 mutations, poorly differentiated tumor, and larger tumor size all contribute to poorer survival rates. Although, given ER+ and HER2+ status, targeted and hormonal therapies do exist for luminal B cancers [18].

HER2/neu subtype refers to a breast cancer in which this is the only the HER2 receptor up-regulated, and not ER or PR. HER2/neu subtype accounts for the least amount of cases diagnosed amongst all the subtypes at only about 10-15% of all cases. HER2/neu tumors tend to be diagnosed at younger age, have more frequent recurrence, higher rates of metastasis, and poorer prognosis primarily due to accompanying large percentage of p53 mutations [19]. A monoclonal antibody targeted therapy called trastuzumab (Herceptin) is commonly used to treat HER2+ tumors [20].

Triple-negative breast cancer (TNBC) subtypes are negative for all three receptors (e.g., ER, PR and HER2) and have by far the poorest prognosis of all breast cancers. The 5-year survival for early stage diagnosis of triple negative breast cancer is only 77% [21]. TNBC is more basal-like as compared to other subtypes having features reminiscent of the cells surrounding the mammary duct and not in the lumen. These tumors tend to be very aggressive with high rates of metastasis, and due to the negative status of all receptors there are limited treatment options outside of surgery, radiation, and standard chemotherapy [22].

Recent advances in technology like whole genome sequencing, microarray, RNA sequencing, and proteomic analysis have been valuable tools in continually redefining the molecular subtypes of breast cancer [23, 24]. Although the technology is still in its youth, these approaches, and more importantly, the combination of these approaches will likely stratify these established subtypes further or give rise to novel subtypes of breast cancer. This is an important area of research as the sorting and understanding of all this new information is critical for personalized medicine and the development of new targeted therapies.

Human genome and long noncoding RNAs (lncRNAs)

Genomic landscape. The central dogma of molecular biology has defined the linear progression of biological identity and function as "DNA makes RNA, makes protein." Translation to protein has traditionally been the requirement for a particular DNA sequence to be assigned the title of "gene". The completion of the Human Genome Project significantly expanded our insight into absolute genomic landscape, and has allowed for scientists to map and organize ~3 billion base pairs to a more precise number of protein-coding genes, often cited as 20,000 to 25,000 [25]. RNA has customarily been reduced to the role of an intermediate, conveniently modified and processed in a highly regulatory manner, with a sole purpose to relay the appropriate language for building a functional protein. It is interesting to note however, that when a more global view of the human genome reveals that only ~1.5% accounts for protein coding exons, increasing to ~2% if including introns and untranslated regions (UTRs)

[26]. The vast majority of the remaining uncharacterized genome aside from housekeeping RNAs was thought not to be expressed and coined "junk DNA," and this has served as a suitable designation for some time.

In recent years, however, modern technologies such as tiling arrays, massive parallel sequencing, and RNA-sequencing have demonstrated that the human transcriptome is much larger than previously thought, giving rise to a paradigm shift in molecular biology known as the "RNA Hypothesis" [27-29]. Current estimates now consider that ~70% of the human genome is transcribed into RNA, a discovery that provides justification to reconsider what constitutes the functional genome, and place greater emphasis on what is defined a "gene" [30]. To date, several classes of non-coding RNAs (ncRNAs) have been categorized and well-defined, demonstrating the diversity and dynamic role of RNA in cellular function (Table 1) [31-33]. A recent example of this shift is the extensive study and documentation of microRNAs (miRNAs), a class of small RNAs that illustrate the sophistication RNA potentiates in exerting direct function in biological processes [34]. The most recent advance has led to the discovery and rapid characterization of another novel group of ncRNA, termed long noncoding RNA (lncRNA).

Characteristics of lncRNA. The current and general parameters for defining lncRNAs are as follows; 1). polyadenylation at 3' UTR in a majority of lncRNAs, 2). epigenetic markers including trimethylation at lysine-4 of H3 (H3K4me3) at the promoter and trimethylation at lysine-36 of H3 (H3K36me3) along transcribed region, 3). active transcription via RNA polymerase II (RNA pol. II), 4). regulation by established

8

transcription factors, 5). processing via canonical exonic splice sites, 6). >200 bp long

without open-reading frame and 7). tissue-specific expression (Fig 1A) [35, 36]. There

Table 1. Species of RNA in the human genome.

Species of RNAs in human genome	
Housekeeping RNAs	Transfer RNAs (tRNAs)
	Ribosomal RNAs (rRNAs)
	Small nuclear RNAs (snRNAs)
	Small nucleolar RNAs (snoRNAs)
Small RNAs (<200nts)	Micro RNAs (miRNAs)
	Short interfering RNAs (siRNAs)
	Tiny transcription initiation RNAs
	Repeat associated small interfering RNAs
	Promoter associated RNAs
	Retrotransposon derived RNAs
	Splice site RNAs
	Anti-sense termini associated RNAs
Large RNAs (>200nts)	Long noncoding RNAs (lncRNAs)
	Pseudogenes
	Transcribed ultra-conserved regions
	Enhancer RNAs (eRNAs)
	Repeat associated long ncRNAs

are exceptions however, and examples of non-polyadenylated lncRNAs transcribed via

RNA polymerase III have been reported [37, 38]. The majority of these features emulate

characteristics of messenger RNAs (mRNAs), emphasizing the notion that lncRNAs are

under precise transcriptional control and processed in the same manner as functional

protein-coding transcripts. The directionality of lncRNA is usually defined in reference to the nearest protein coding gene, and can exhibit orientations in antisense, bidirectional, sense, intronic and intergenic (Fig 1B).
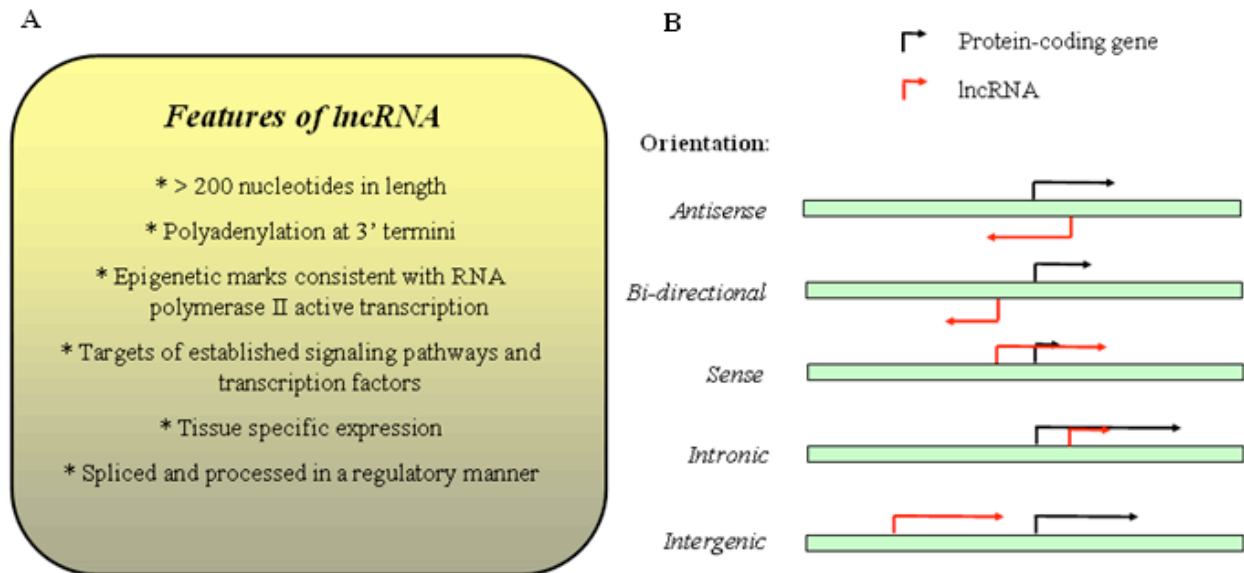
Figure 1. The characteristic features and orientation of lncRNAs. (A) There are several defining features of lncRNAs which distinguishes these transcripts as a class. (B) LncRNA orientation is defined in relation to the transcriptional start site (TSS) of the nearest protein coding gene. LncRNA can be transcribed antisense, bi-directional, sense, intronic, or intergenic orientations. Reproduced from 'Edward J. Richards and Jin Q. Cheng, 2014, Long noncoding RNAs and their expanding roles in cancer, Current Topics in Biochemical Research, Vol. 16, 81 – 95'.

Conservation of lncRNA. The conservation of lncRNAs has been the topic of extensive debate and study. The significance of linking these attributes to lncRNA would provide strong evidence for functionality suggesting these transcripts were retained through selective pressures. By and large, however, lncRNAs exhibit poor sequence conservation across species in comparison to protein coding genes and microRNAs [39, 40]. One study reported that in starting with a pool of tissue specific human lncRNAs, 80% contained an ortholog transcript in chimpanzee, and this decreased to only 39% in

mice [41]. Thus, given the fact that the protein coding genome is almost entirely conserved across higher organisms, the noncoding genome likely is a key contributor to species identity and complexity. From an evolutionary perspective, it has been suggested that lncRNAs are rapidly evolving regulatory elements of high plasticity responsible for continually mobilizing mammalian evolution [40]. This concept is very striking given the short evolutionary distance between humans and chimpanzee, and lncRNAs account for about half the genetic differences between the two genomes [42].

It is also important to remember that selecting for a protein coding gene and lncRNA needs to be considered independent and mutually exclusive processes. Conservation of a functional protein coding gene requires high sequence preservation to ensure proper amino acid coding and secondary protein folding. Any disruption in sequence can drastically alter the biological function of a protein, and likely this contributes to the high conservation of proteins across several species. On the contrary, lncRNAs do not require significant sequence retention along the entire body of the transcript to retain appropriate architecture and secondary structure necessary for function. LncRNA function can occur through a protein binding site, a miRNA seed site, mRNA recognition sequence, or DNA binding element, all of which require relatively short stretches of sequence for utility. A prime example is the X-inactive specific transcript (XIST), a 17-kb transcript in human in which conservation is only retained in exon 1, and this conserved region is critical for proper function [43]. Also, despite a lack of lncRNA sequence conservation in and of itself, the promoters of lncRNAs demonstrate very high sequence conservation and are capable of binding established transcription factors [44].

Mechanisms of lncRNA Function

   Epigenetic and chromatin modification. LncRNAs can regulate gene expression within a few megabases along the same chromosome or from a distance at a different chromosome. This regulation is referred to as acting in "cis" and in "trans", respectively (Figure 2). The XIST is a very well studied lncRNA that functions in "cis". In developing mammalian females, X-chromosome inactivation is a critical compensatory mechanism to ensure gene dosage equivalency between males and females along the X-chromosome. XIST mediates this process through its exclusive expression on the inactive X-allele via the X-chromosome inactivation center (Xic) [43, 45]. The first exon of XIST contains a double hairpin motif capable of binding polycomb recessive complex 2 (PRC2), and recruitment of this histone methyltransferase complex results in tri-methylation of lysine 27 at H3 (H3K27me3) and transcriptional silencing along the inactive X-chromosome [46]. Interestingly, the active X-chromosome represses XIST expression through interaction with an antisense lncRNA known as TSIX, yet an exact mechanism has yet to be determined for this modulation [47].

   One of the better characterized lncRNAs that functions in epigenetic and chromatin modification in "trans" is HOX antisense intergenic ncRNA (HOTAIR). HOTAIR is located on chromosome 12 and is oriented antisense to the HoxC cluster, and transcriptionally represses the HoxD locus on chromosome 2 [48]. There are two functional domains of HOTAIR, a PRC2-binding domain at the 5' end which interacts with subunit Suz12, and an LSD1/CoREST1-binding domain at the 3' end. It is

suspected that these two complexes use HOTAIR as a modular scaffold for collaboration and recruitment to target gene locations [49]. The precise mechanisms for
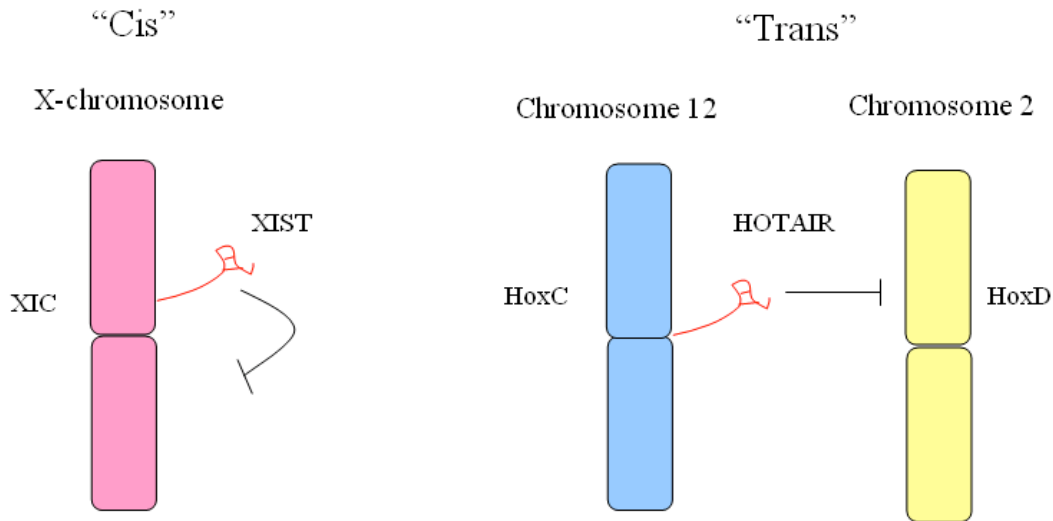
Figure 2. LncRNA functions in "cis" or in "trans". LncRNAs that function locally on the chromosome of origin are referred to as functioning in "cis". This is the case of XIST, which is transcribed at X-chromosome inactivation center (XIC) and is responsible for X-chromosome inactivation. LncRNAs can also regulate target gene expression on distant chromosomes in "trans", such as HOTAIR which is transcribed from HoxC gene cluster on chromosome 12, yet represses target HOXD expression on distant chromosome 2. Reproduced from 'Edward J. Richards and Jin Q. Cheng, 2014, Long noncoding RNAs and their expanding roles in cancer, Current Topics in Biochemical Research, Vol. 16, 81 – 95'.

how "trans" functioning lncRNAs target specific genes on distant chromosomes are still not well understood. There have been several additional mechanisms reported for lncRNA function (Table 2).

Table 2. The biological mechanisms of lncRNA function.

Mechanisms of lncRNA function		
Function	**Specific lncRNA mechanism**	**Reference**
Epigenetic and chromatin modification	**XIST**: Exclusively expressed on the inactive X chromosome, recruits PRC2 to X-chromosome inactivation center (XIC) for gene silencing	43
Sponge for miRNAs	**H19**: Contains 4 Let-7 binding sites throughout transcript body to competitively bind and down-regulate Let-7 in Ago2 dependent manner.	51
Messenger RNA processing and protein translation	**ZEB2-AS**: Transcribed antisense to ZEB2 mRNA, binds ZEB2 transcript and prevents splicing of 5' intron preventing protein translation.	57
Enhancers and transcriptional activation	**CCAT2**: Transcribed in enhancer region ~334 kb upstream of C-MYC transcriptional start site, recruits TCF7L2 to C-MYC promoter.	61
Decoy for protein binding	**PANDA**: Induced in DNA damage response, binds and sequesters transcription factor NF-YA to prevent induction of pro-apoptotic gene targets	63

Reproduced from 'Edward J. Richards and Jin Q. Cheng, 2014, Long noncoding RNAs and their expanding roles in cancer, Current Topics in Biochemical Research, Vol. 16, 81 – 95'.

Sponge for miRNAs. It is becoming more apparent that some lncRNAs contain multiple miRNA response elements (MREs), which layers ncRNA regulation through competition. This theory is known as the competing endogenous RNA (ceRNA)

hypothesis [50]. LncRNAs can interact with miRNAs through these repeat MREs along the length of the transcript to influence global gene expression. For example, within the body of lncRNA H19 resides 4 canonical and noncanonical MREs for Let-7 [51]. It was demonstrated that enforced expression of H19 results in more Let-7 binding at these MREs in an Ago2-dependent manner. Through competitive binding and sequestering of Let-7, downstream targets of Let-7 such as DICER and HMGA2 increased their expression [51].

Linc-ROR, a regulator of reprogramming, can also act as a sponge and shares MREs with several miRNAs that target the core transcription factors in maintaining embryonic stem cells. [52, 53]. Expression patterns of linc-ROR are consistent and associate with undifferentiated status. Linc-ROR contains MREs for miR-145, miR-181, and miR-99-3p, which target Oct4, Nanog, and Sox2. Enforced expression of linc-ROR competitively binds these miRNAs relieving target repression, enhancing protein expression of these core transcription factors. Loss of linc-ROR leads to decreased expression of Oct4, Nanog, and Sox2 and movement towards the differentiated state [53].

Messenger RNA Processing and Protein Translation. LncRNAs can physically interact with mRNAs to influence stability of the transcript or translation in dynamic ways [54, 55]. This relationship is especially apparent in the case of protein coding genes which contain overlapping natural antisense transcripts (NATs). Uchl1 protein gene overlaps with antisense lncRNA Uchl1-AS at the 5' region and also contains an inverted SINE2B repeat element downstream [56]. In the stress response or

using rapamycin treatment to induce stress, Uchl1-AS will rapidly shuttle from the nucleus to the cytosol where it binds to and stabilizes the Uchl1 mRNA transcript along the 5' UTR region leading to increased translation. The detailed underlying mechanism is not very well understood, but the SINE2B element is critical for enhanced translation during this process, indicating a possible novel function for transposable elements within the genome [56].

In another example, Snail1-induced epithelial to mesenchymal transition (EMT) does not activate Zeb2 mRNA transcription, but rather promotes Zeb2 translation through splicing of a large intron at the 5' UTR region. This intron contains an internal ribosome entry site (IRES) necessary for the expression of Zeb2. Maintenance of 5'-UTR Zeb2 intron is dependent on the expression of Zeb2-AS1 that overlaps the 5' splice site in the intron. In the non-EMT state this intron is spliced out, which does not affect mRNA level yet does prevent protein translation. In the case of Snail1-induced EMT, Zeb2-AS1 is transcribed and its complimentary base pairs with the 5' UTR of Zeb2 prevents splicing of the 5' IRES sequence allowing for Zeb2 translation [57].

Enhancers and transcriptional activation. Enhancers are regulatory DNA elements that activate gene transcription in "cis" through chromosomal looping and recruitment of cofactors [58, 59]. It is now emerging that these regulatory regions frequently contain lncRNAs which have target gene specificity and promote cofactor recruitment. In one study, enhancer-associated ncRNA-a7 was characterized as being able to specifically regulate SNAI1 transcription, yet had no relationship to RNF114, UBE2VI, TMEM189, or CEBPB which reside within the same 100 kb radius [60]. It was

also shown that SNAI1 migration phenotype was driven through ncRNA-a7 providing strong evidence that lncRNAs play a key regulatory role in gene expression and downstream biological processes [60].

Another lncRNA CCAT2, which locates ~334 kb away from C-MYC transcriptional start site, was shown to be over-expressed in colorectal cancers and promote chromosomal instability, invasion, and metastasis [61]. CCAT2 binds to transcription factor TCF7L2 for recruitment to the enhancer region and results in activation of C-MYC transcription through chromosomal looping and interaction at the promoter. Interestingly, CCAT2 has a single nucleotide polymorphism (SNP) that associates with increased risk of developing colon cancer. This study suggests that this SNP is functional through the RNA product, and enhances TCF7L2 recruitment to the C-MYC promoter prompting a paradigm shift that SNPs are typically functional through DNA elements [61].

Decoy for protein binding. LncRNAs can also function as decoys through binding protein targets and thereby preventing normal function and localization of the protein. For example, in response to DNA damage a subset of lncRNAs will be induced in a p53-dependent manner including 2 bidirectional transcripts upstream of the p21 promoter, lincRNA-p21 and PANDA [62, 63]. During this response PANDA will bind to the transcription factor NF-YA and prevent its localization at promoters of pro-apoptotic target genes. Thus, PANDA allows for cell cycle arrest and DNA repair rather than programmed cell death in the p53 response [63]. Another example can be seen in extra-coding CEBPA lncRNA (termed ecCEBPA), a ~4.5kb transcript encompassing the

entire ~2.6 kb portion of the CEBPA gene in the sense direction [64]. When ecCEBPA is expressed it will competitively bind and exclude DNMT1 for the occupying chromatin at the CEBPA promoter, thus preventing DNA methylation and activating transcription of the CEBPA [64].

LncRNAs in Cancer and Malignancy

LncRNAs regulate genes that participate in a wide range of biological processes, and for this reason they can act as oncogenes and tumor suppressors. Aberrant expression of lncRNA can be a deciding point in cancer etiology, promote tumor onset and disease progression, as well as drive metastasis. LncRNAs are also a novel and untapped resource for pathological and clinical utility for use as novel biomarkers. Although a number of lncRNAs have been associated with cancer in literature, there is a clear need for further identification and characterization of more candidates. The following section will discuss the function and clinical significance of specific lncRNAs reported in human cancer to date.

HOTAIR. HOTAIR, as discussed above, locates on chromosome 12 and is responsible for "trans" silencing the HoxD genes of chromosome 2 through recruitment of PRC2 complex [48]. HOTAIR expression has been shown to be up-regulated in both primary and metastatic breast cancers, in some cases as high 2,000-fold over-expression. Furthermore, the elevated expression correlates with increased risk of metastasis, as well as poor metastasis-free and overall survival [65]. *In vitro* studies

demonstrate that knockdown of HOTAIR can inhibit invasive capacity. While *in vivo*, enforced expression of HOTAIR results in a greater capacity for breast cancer cells to colonize the lung following tail vein injection [65].

In non-small cell lung cancer (NSCLC), HOTAIR has been shown to have elevated expression levels when comparing 42 lung carcinomas to matched normal tissues [66]. Elevated HOTAIR in NSCLC patients also associates with late stage tumor, positive lymph node metastasis, and poor overall survival. This study also demonstrates that HOTAIR can control cell migration, invasion, and metastasis *in vitro* and *in vivo*, and shows that knockdown of HOTAIR leads to apoptosis. Depletion of HOTAIR resulted in significant loss of expression of known metastasis promoting genes MMP2 and MMP9, and increased expression of HOXA5, which has been shown to be important in inhibiting lung cancer metastasis [66].

A recent study shows that in renal carcinoma cells HOTAIR is inversely correlated in expression with miR-141 [67]. This study also shows that HOTAIR plays a critical role in promoting migration, invasion, and survival in renal carcinoma. Interestingly, enforced expression of miR-141 results in significant loss of HOTAIR expression, reduced invasive phenotypes, and increased apoptosis. Further investigation revealed that HOTAIR is a *bona fide* miR-141 target in an AGO2 dependent manner. This study illustrates the dynamic relationship between ncRNA species regulating each other, and their potential role in kidney cancer [67].

XIST. In one study the contribution of XIST to hematological cancers using a conditional knockout mouse model was evaluated. The majority of the females in which

XIST is conditionally removed from cells in the blood compartment results in aggressive and rapid onset of multi-lineage hyperproliferations, myelodysplasia, and myeloproliferative neoplasms [68]. These results suggest a potent tumor suppressor role for XIST in hematopoietic stem cells and potentially other tissues. Deletion of XIST resulted in X chromosome reactivation and aberrant gene expression in several major tumor suppressive pathways contributing to the rapid onset of disease [68].

A second study used microarray technology to compare global changes in RNA expression in primary ovarian tumor to recurrent ovarian tumors after chemotherapy [69]. Interestingly, in recurrent ovarian cancer XIST was the most significantly downregulated gene in comparison to primary tumor [69]. From a therapeutic standpoint, this suggests that XIST can potentially play a role in sensitizing patients to taxol and cisplatin based chemotherapies. XIST can also serve as a marker for chemoresistant ovarian cancer.

XIST has also been linked to the BRCA1 tumor suppressor gene in patients with breast or/and ovarian cancers. Initial reports indicated that breast cancers with deficient BRCA1 expression or function have defects in X chromosome inactivation, and enforced expression of BRCA1 gene can initiate XIST expression [70]. This view was later challenged using a BRCA1 conditional knockout mouse model which demonstrated that depleting BRCA1 had no affect on X inactivation and gene dosing, as well as showed no evidence of BRCA1 and XIST colocalization [71]. The relationship between BRCA1 and XIST in breast cancer will require future investigation to be better established.

MEG3. Maternally expressed gene 3 (MEG3), a paternally imprinted gene comprised of 10 exons, is expressed in many normal tissues [72]. There are 12 different MEG3 gene transcripts, generated by alternative splicing. They contain the common exons 1-3 and exons 8-10, but each uses one or more exons 4-7 in a different combination in the middle. All MEG3 RNA isoforms contain three distinct secondary folding motifs M1, M2, and M3 that are critical for activation of p53. MEG3 isoform expression patterns are tissue and cell type specific. However, each isoform stimulates p53-mediated transactivation and suppresses tumor cell growth. MEG3 retains the ability to activate p53 in spite of replacing regions unrelated to M1, M2, or M3, thus provides evidence that a proper folding structure of the MEG3 RNA molecule is critical for its biological functions and that only short stretches of sequence conservation are vital to maintain function [73]. In addition to transcriptionally activating p53, MEG3 can also enhance binding of p53 at target gene promoters including GDF15 [74]. MEG3 expression is dramatically lost in several human cancers [73]. Reintroduction of MEG3 into NSCLC cells was able to slow colony formation capacity and cell proliferation [73].

In panel of hepatocellular carcinoma (HCC) cells MEG3 is frequently downregulated, and enforced expression of MEG3 was able to slow cell proliferation, inhibit anchorage independent growth, and promote apoptosis [75]. This same study also demonstrated that maternal allele of MEG3 is hypermethylated in a DNMT1 and DNMT3b dependent manner. Interestingly, miR-29a expression is also commonly lost in HCC and DMNT1 and DNMT3b are direct targets of miR-29a. Loss of miR-29a results in aberrantly increased DMNT1/3b expression in HCC and provides a novel mechanism for MEG3 downregulation due to hypermethylation at the promoter [75].

ANRIL. Antisense noncoding RNA in the INK4 locus (ANRIL) was initially identified as being deleted in the germ-line deletion of the 9p21 gene cluster [76]. This gene cluster contains several important tumor suppressor genes including, p14ARF, p15INK4B, and p16INK4A that are deleted along with ANRIL across ~ 403 kb region and results in cancer predispositions [77-79]. ANRIL locates in the antisense direction and overlaps with the p14ARF promoter and 2 exons of the p15INK4B gene body [76]. Although initial reports identified this transcript as associated with deletion, later reports suggest that ANRIL over-expression can also drive cancers [80, 81].

Elevated expression of ANRIL was detected in prostate cancer and inhibiting ANRIL expression resulted in increased INK4A and INK4B gene products, denoting a transcriptional repression in "cis". In this same study it was shown that ANRIL can bind to CBX7 component of the PRC1 repressive complex and epigenetically silence the associated tumor suppressive genes on 9p21 gene locus [80]. ANRIL is also elevated human gastric cancers and associates with poor overall survival. This same study demonstrated that ANRIL can silence oncogenic miRNAs, including miR-99a and miR-449a, through binding to the components of PRC2 transcriptional repression complex [81].

Emerging lncRNAs in cancer. The 8q24 chromosomal amplification associates with increased risk for several cancers. Linked within this region resides a gene desert containing no protein coding genes in addition to the C-MYC oncoprotein [82-85]. Several lncRNAs (CCAT2, Carlo-5, and PVT1) have been identified within this region

and regulate C-MYC expression through a variety of mechanisms including transcriptional activation, enhancer-associated function, and protein stability [61, 86, 87].

Another interesting lncRNA, FAL1, has recently been identified as locating in a focally amplified chromosome 1q21 region across several epithelial cancers including breast and ovarian [88]. FAL1 binds to PRC1 core subunit BMI1 to transcriptionally repress tumor suppressor gene p21. Interestingly, FAL1 is a predictor of poor overall survival in ovarian cancer patient and elevated FAL1 closely associates with chromosome 1q21 DNA copy number gain [88].

TGFβ-induced lncRNA-ATB is also emerging as an important driver in EMT and metastasis in liver cancer. LncRNA-ATB is a potent driver of the EMT by acting as miRNA sponge for the miR200 family [89]. This family is the master regulator of ZEB1/2 expression through 3' UTR mediated repression in an Ago2 dependent manner. Also, lncRNA-ATB promotes cell proliferation and distant colonization through directly binding and stabilizing IL-11 mRNA resulting in activation of phospho-STAT3 [89].

TGFβ Signaling Pathway

TGFβ Signaling Pathway and Ligands. The TGFβ signaling pathway is dynamic and involves in a variety of normal cellular programs such as cell differentiation, proliferation, inflammation, tissue morphogenesis, regeneration, and maintenance (Figure 3). It has become clear that TGFβ pathway is dependent on cellular context. For example, depending on cell type and tissue of origin the TGFβ pathway can elicit a radically different, and sometimes opposite functional response [90]. Significant

research efforts have been put forth in defining the related cytokines, receptors, and intracellular signaling molecules in order to establish the pathway, which made it possible to study its functional role in a given cellular context. We now have a better understanding of the pleiotropic effects elicited by the TGFβ pathway.

The TGFβ related ligands and cytokines consist of 2 subfamilies (including 42 ligands and over 300 target genes) as categorized by sequence homology and downstream effectors in the pathway [91]. The first family includes the TGFβ's, Activin, and Nodal ligands, while the second family consists of the BMPs (bone morphogenetic proteins) and MISs (muellerian inhibiting substances) [92]. Both families work in a variety of developmental cellular functions such as embryogenesis, osteogenesis, and hormone regulation. [93-95] The TGFβ monomer contains a conserved "cysteine knot" which consists of several beta strands held together by three tight disulfide bridges. These monomers typically take on the active dimerized form that is held via hydrophobic interactions and disulfide bridges. [96] Ligand traps are a group of proteins that regulate the TGFβ ligands and cytokines access to receptors. These traps can promote or inhibit the ligand/receptor interaction and provides a mechanism of specificity. [97]

TGFβ Signaling Pathway Receptors. The receptors associated with TGFβ pathway are serine/threonine kinases and are referred to as type I (7 receptors) or type II (5 receptors). These receptor kinases contain an extracellular N-terminal ligand binding domain, a transmembrane domain, and a C-terminal kinase signaling domain. Type I receptors are 53-KD proteins that contain a conserved SGSGSG motif, called

GS domain, located N-terminally and in close proximity to the kinase domain. This domain on type I receptor is directly phosphorylated by type II receptor resulting in activation of dimeric complex, and therefore a heterodimer is critical for activation. [98] There are 2 distinctive ligand-receptor interactions which are distinguished by the 2 superfamilies of ligands mentioned above. These distinctions refer to each subfamilies affinity to TGFβ receptor types, for example TGFβ's, Activins, and Nodal ligands preferentially bind to type II receptor and do not interact with isolated type I receptors, while BMPs prefer to interact and have high affinity for type I receptors. [99]

TGFβ Signaling Pathway SMAD Proteins. The intracellular effector and regulatory molecules of the TGFβ pathway are the SMAD family proteins. Three classes of SMAD proteins exist in humans, receptor regulated SMADs or "R-SMADs" (includes SMAD1, 2, 3, 5, 8/9), co-mediator SMAD or "co-SMAD" (SMAD4), and inhibitory SMAD or "I-SMADs" (SMAD6, 7) (Figure 4). R-SMADs are the direct substrates of receptor mediated phosphorylation by type I receptors. SMAD2 and SMAD3 are activated in response to TGFβ ligand subfamily, while SMAD1, SMAD5, and SMAD8 respond to BMP subfamily of ligands. [92]

R-SMADs are typically about 400-500 amino acids, contain a conserved N-terminal MH1 (MAD-homology 1) domain (40-95% conservation across R-SMADs), C-terminal MH2 (MAD-homology 2) domain (38-98% conservation), and have 2 conserved structural domain to provide secondary structure. [100] The MH1 domain functions as the DNA recognition and binding element while MH2 domain facilitates homo- or hetero-
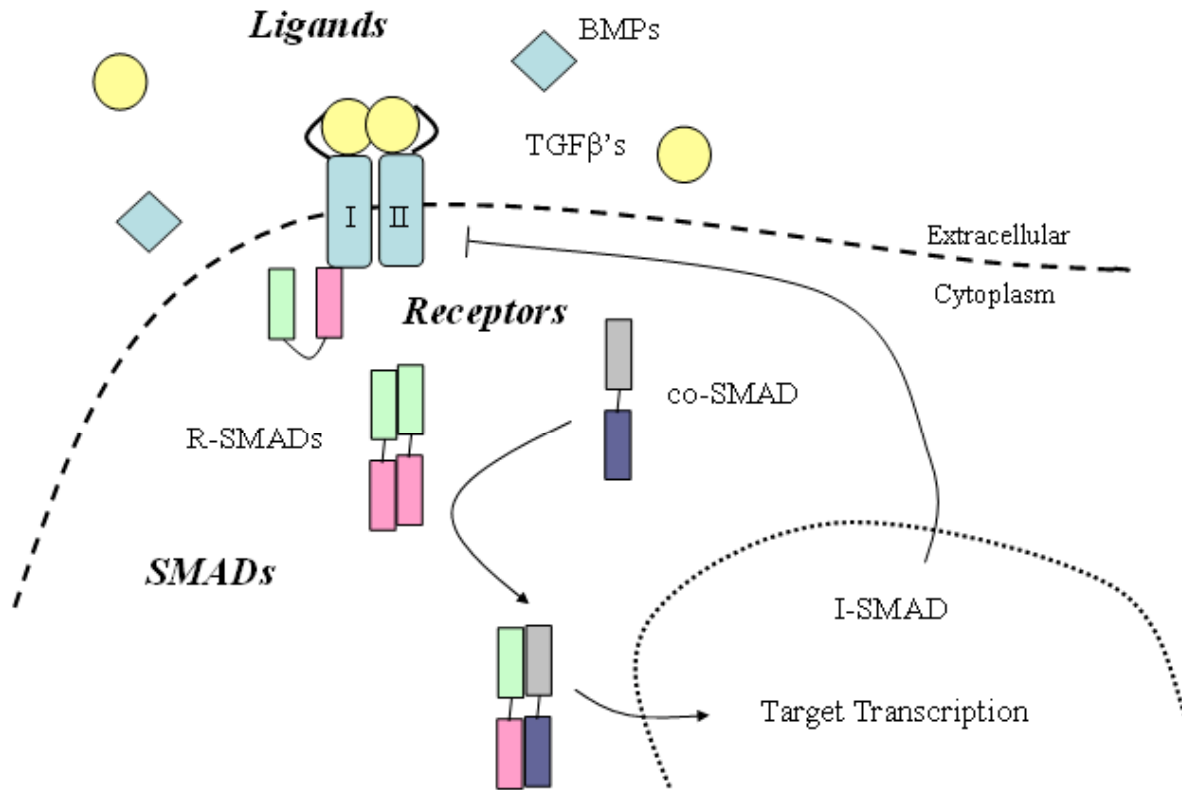
Figure 3. The TGFβ signaling pathway and associated molecules. The TGFβ pathway consists of ligands, receptors, and intracellular signaling molecules called SMADs to influence target gene expression. The TGFβ (transforming growth factor beta) and BMP (bone morphogenic protein) ligands bind type I and II TGFβ receptors to initiate R-SMAD (regulatory) and co-SMAD (cooperative) signaling cascade. Inhibitory SMADs (I-SMADs) can directly bind to and inhibit the receptors through a negative feedback loop.

oligermization to promote dimer activation. The R-SMADs also contain an SXS motif at the C-terminal region and these serine residues are critical for activation [101]. Two members of the Hecht-family of E3 ubiquitin ligases, SMURF1 and SMURF2, can target and negatively regulate the R-SMADs and receptors [102, 103]. Subsequent R-SMAD activation via type I or type II TGFβ receptor, homo-oligomerization will first occur, and

26

downstream with co-SMAD4 as a hetero-dimer or -trimer. The active heterocomplex of R-SMAD/co-SMAD4 can then translocate to the nucleus and activate transcription of numerous target genes. Target gene transcription and functional outcome occurs through the dynamic interactions with the MH1 and MH2 domains as well as the recruitment of other transcriptional mediators.

Negative regulation of the TGFβ pathway is a critical feedback process and tightly regulated mechanism to haul over-active signaling. Two members of the Hecht-family of E3 ubiquitin ligases, SMURF1 and SMURF2, can target and negatively regulate the R-SMADs and type I and II receptors [103]. Specifically SMAD1 and SMAD5 is a substrate for SMURF1 ubiquitin ligase activity. [104] SMURF2, which contains significant homology to SMURF1, can preferentially inhibit SMAD1 and SMAD2 [105, 106]. Inhibitory SMAD7 plays a critical role as co-regulator in this process.

Once TGFβ pathway is activated, SMAD7 will translocate from the nucleus to the cytosol to locate at the plasma membrane through an unknown mechanism. As a result of this translocation, SMAD7 can form complexes with both SMURF1 and SMURF2 and become tethered to the membrane to directly bind type I receptors, including those involved in the late endosomal compartments. Binding of SMAD7 to type I receptor will competitively inhibit phosphorylation of R-SMADs [106]. Also, recruitment of SMURFs to the membrane will promote E3 ubiquitinase activity to directly facilitate degradation of the receptors. This negative feedback look is perpetuated and balanced given that during TGFβ activation SMAD7 is itself ubiquitated and degraded, yet is also a transcriptionally activated target gene [108]. This ensures proper SMAD7 turnover and dosage during the response.

SMAD Family Proteins
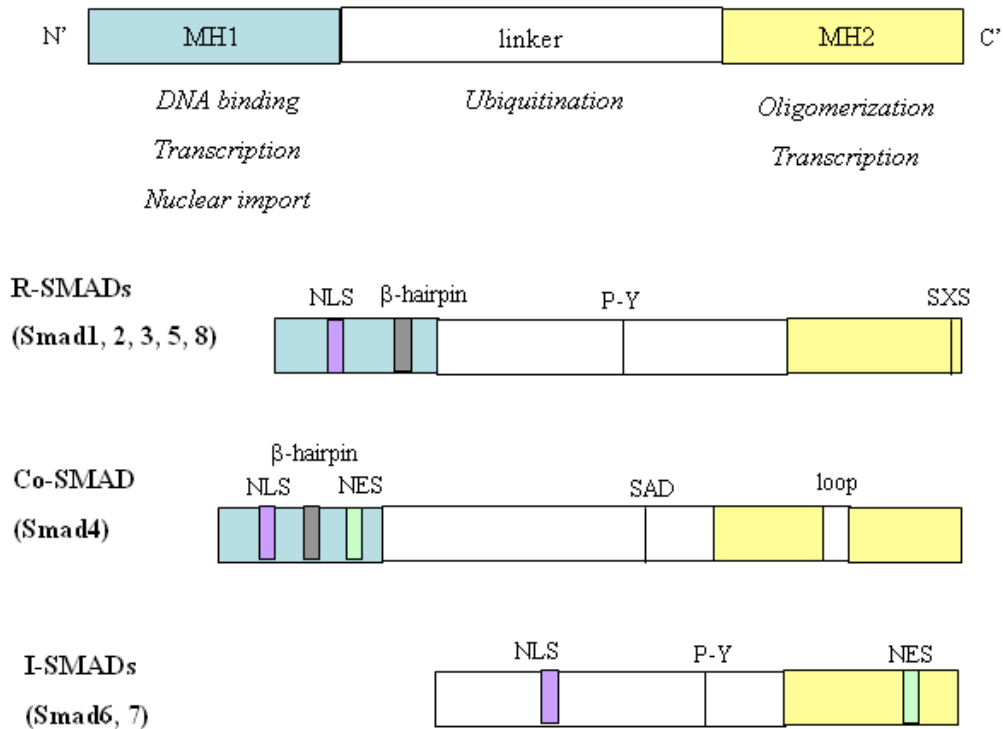
Figure 4. The SMAD family of proteins. The SMAD family is comprised of R-SMADS (regulatory), Co-SMAD (cooperative), and I-SMAD3 (inhibitory) and function through varying regions N-terminal MH1 domain (MAD homology-1), middle linker region, and C-terminal MH2 domain. Abbreviations: NLS (nuclear localization signal), P-Y (proline-tyrosine), SXS (serine-X-serine), NES (nuclear export signal), SAD (SMAD activation domain). Figure adapted from Moustakas et. al. [107].

TGFβ-associated phenotypes and genes

Epithelial to Mesenchymal Transition (EMT). The epithelial to mesenchymal transition (EMT) is a highly organized process facilitated through distinct biological pathways that shift a polarized epithelial cell to a mesenchymal phenotype.

Mesenchymal phenotypes associate with increased motility and invasion, production of extracellular matrix components (fibroblast-like), and resistance to apoptosis [109, 110]. EMT exemplifies the inherent ability of epithelial cell plasticity and how dynamic interactions from extracellular components can influence terminally differentiated tissue to promote cancer phenotypes. Ultimately, this process is the driving force in which a cancer cells can degrade the basement membrane and gain access to the blood stream and initiate metastasis (Figure 5).

It is important to note that the process of EMT is initial breaching of the basement membrane and escape from the primary site. Downstream, other process such as resistance to anoikis (anchorage-independent cell death) and mesenchymal to epithelial transition (MET) must still take place for a cell to successfully colonize and form a secondary carcinoma. The process of EMT is considered a fundamental hallmark of cancer, and further understanding this process is critical given the consequences that once cells leave the primary site acquire these aggressive characteristics to enable metastasis [111].

Changes in Gene Expression Associated with EMT. There are several fundamental gene changes associated with conversion of an epithelial cell to more mesenchymal cell. Down-regulation of cell to cell adhesion and tight junction molecules like e-cadherin (CHD1) and ZO-1 which stabilize and immobilize the epithelium is typically characteristic to early EMT. This down-regulation is also true for several other more specific cytoskeletal and adherent molecules like claudins, occludins, desmoplakins, and cytokeratins [112]. At the same time as there is a loss of adherent
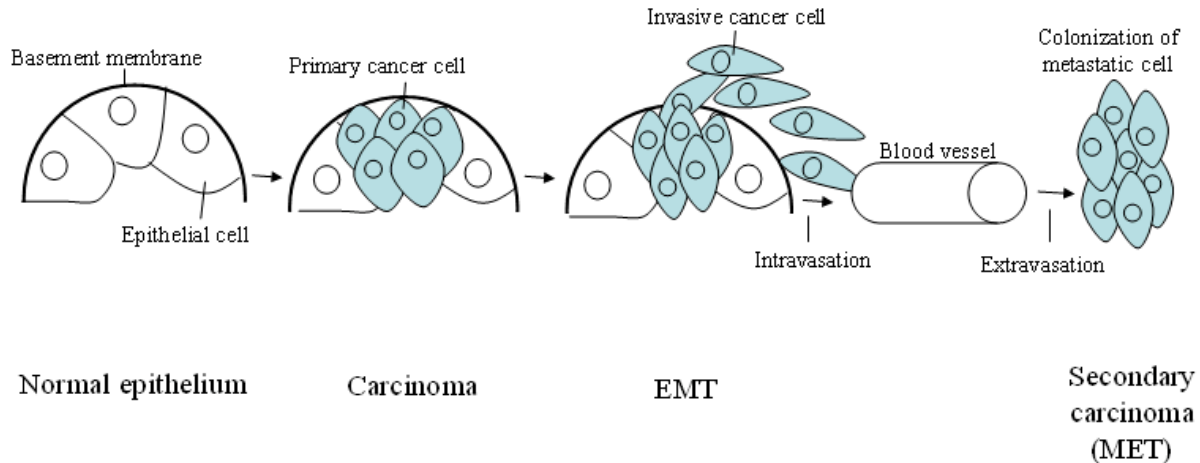
Figure 5. The epithelial to mesenchymal transition (EMT) in breast cancer. In breast cancer progression, the normal epithelial cell layer within the breast duct will become hyperplastic and eventually lead to an uncontrollable proliferative mass known as carcinoma. The primary carcinoma will fill the duct while at the edge of the tumor some cells will break through the basement membrane. These cells are considered to have undergone the epithelial to mesenchymal transition (EMT). These invasive cells can gain access to blood vessels and travel to distant organs in the body to colonize and form secondary carcinomas. This process is referred to as mesenchymal to epithelial transition (MET) and metastasis. Adapted from Kalluri and Weinberg [110].

molecules, there is an up-regulation of another set of molecules responsible for motion and invasive potential. There is a "cadherin switch" from E-cadherin to N-cadherin expression which decreases cell to cell contact. The up-regulation of N-cadherin not only affects adhesion qualities on cell surface, but also shifts signaling pathways inside the cell to promote migration and invasion. These pathways and interactions include β-catenin, receptor tyrosine kinases (RTKs), platelet derived growth factor (PDGF), fibroblast growth factor receptors (FGFRs), and SRC family kinases [113-115]. In addition, several other genes such as vimentin, fibronectin, collagen I/IV, lamin I, and matrix metalloproteinases MMP2/MMP9 are up-regulated enhancing cell's ability to break through the basement membrane and ECM components (Figure 6) [104].

There are also several established transcription factors that drive EMT including ZEB1/2, SNAI1, and basic helix loop helix transcription factors (bHLH). ZEB1/2 can transcriptionally repress or activate several target genes to promote EMT in cancer through it's interaction with DNA regulatory sequences called E-boxes [116]. ZEB1 can interact with c-terminal-binding protein (CTBP) or switch sucrose/non fermentable protein (SWI/SNF) protein BRG1 to transcriptionally repress e-cadherin [104]. SNAI1 (SNAIL) and SNAI2 (SLUG) are known to drive mesenchymal phenotype during embyrogenesis and development and can become reactivated in cancer [117].

Similar to ZEB1/2 function, SNAI1/2 can bind to E-boxes through their zinc finger motifs to shut down E-cadherin transcription. Once bound to E-cadherin promoter, SNAIL can recruit polycomb recessive complex 2 (PRC2) which contains several subunits histone methyltransferase, histone demethylases, histone deacetylases, and co-repressors. Together the subunits of PRC2 work through modifying histones and chromatin to coordinate transcriptional repression [104]. The bHLH transcription factors that play a major role in contributing to EMT are TWIST1 and TWIST2. TWIST1/2 can repress E-cadherin expression independently from SNAIL and also transcriptional activate several mesenchymal genes such as N-cadherin [116]. All three of these transcription factors can be up-regulated via several mechanisms associated with tumor proliferation and tumor microenvironment, including but not limited to elevated levels of TGFβ, hypoxia inducible factor 1α (HIF1α), WNT signaling pathway, and RAS-MAPK signaling pathway [118-121].

Epithelial to Mesenchymal Transition (EMT)

Markers:
E-cadherin
ZO-1
Cytokeratins

TGF-β

Markers:
Vimentin
Fibronectin
Zeb1/2

Phenotypically:
Cell polarity
Tight junctions
Stationary

Phenotypically:
Fibroblast-like morphology
Few cell adhesions
Motile and invasive
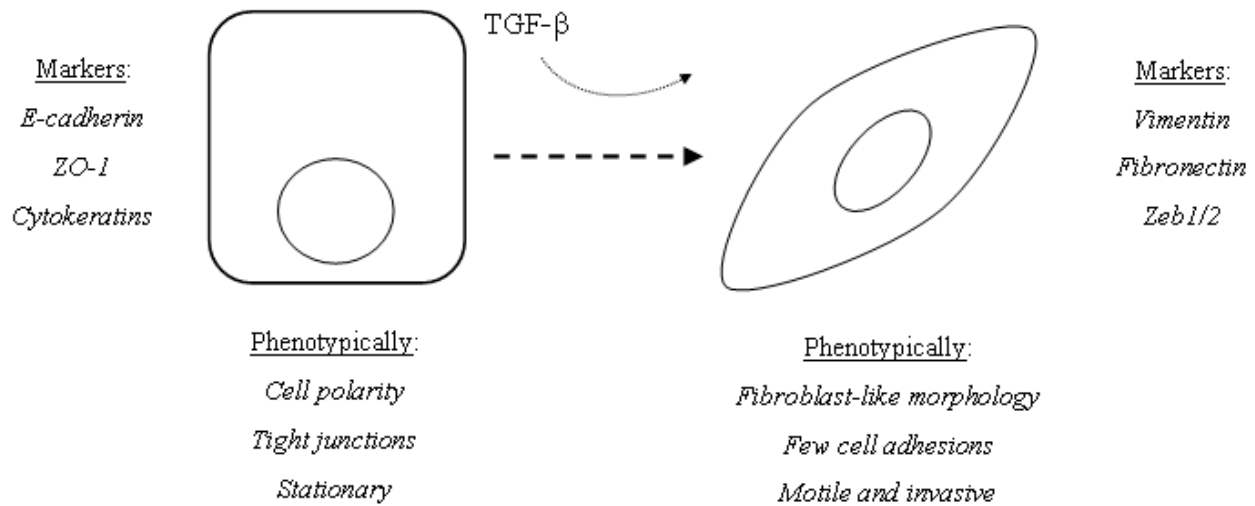
Figure 6. EMT-associated markers and phenotypes. Throughout the process of EMT gene signatures and cell phenotype will change from epithelial status to a more mesenchymal state. Characteristic expression of epithelial markers such as e-cadherin, ZO-1, and specific cytokeratins will be down-regulated, and the TGFβ-induced transition with trigger a switch towards mesenchymal markers such as vimentin, fibronectin, and Zeb1/2. The epithelial cells will phenotypically appear more spindle-like, contain fewer tight junctions, and become more motile and invasive.

Cancer Stem Cells (CSCs). Historically, clinicians have observed in that giving a single agent therapy to a patient a subset of cells in the primary tumor are sensitive to the treatment while others remain resistant. It is now established that a primary tumor is not composed of a monoclonal mass of cells with equal cellular make-up and constituents, but rather tumor heterogeneity exists within the mass [122]. In fact a primary tumor is not only composed of genetically and phenotypically distinct carcinoma cells, but is also composed of endothelial, stromal, immune, and hematopoietic cells

that function as a dynamic system. Terminally, well-differentiated epithelial cells are typically thought of as being functional, healthy adult tissue. Therefore, the term "cancer stem cell" in part refers to cancer's ability to revert from this well-differentiated state back to a more "stem-like" state. A loose definition for CSCs is a population of cancer cells that can be isolated with a capacity for long-term repopulation and self-renewal of the original tumor. [123, 124]

Functionally, CSCs can initiate tumor formation in xenograft model, self-renew through serial passaging of spheres, and spawn daughter cells which can proliferate and also form new tumor, although this tumor is may be different from parental [104, 125]. The origin of CSCs remains up to debate and several theories have surfaced. One known as the hierarchal model proposes that a small subset of CSCs exists within the tumor population, and these cells obtain the characteristics for tumor propagation as well as self renewal of more CSCs. In contrast the stochastic model assumes that all cells within the tumor have the ability for tumor initiation and these qualities are intrinsically determined by several factors and influenced by surrounding environment (Figure 7) [124].

Breast Cancer Stem Cells (BCSCs). BCSCs originate in the breast and obtain the properties described above. Initial reports of BCSCs utilized cell surface markers CD44 and CD24 status to distinguish the stem cell population. BCSCs are typically defined as population of cells that is CD44+/CD24-. In one study, total breast cancer cells from human patient tumor were sorted for CD44+/CD24- population or cells without this surface receptor status. In eight out of nine patients as few as 100 cells

isolated with CD44+/CD24- were able to establish a xenograft tumor in mice, while tens of thousands of cells were required if there was alternate cell surface status [126]. The classification of BCSCs based solely on CD44 and CD24 status remains up to debate, as several reports claim that elevated CD24 associates with poorer patient outcome [127, 128]

Another cellular marker that is widely used to distinguish BCSCs as well as other CSCs is alcohol dehydrogenase isoform 1, ALDH1. This enzyme normally catalyzes oxidation of intracellular aldehydes in the process of normal stem cell differentiation [129]. Interestingly, the population of ALHD1+ cells sorted from total CD44+/CD24- population is only about 1%. However, if CD44+/CD24- population is also then sorted for ALDH1+, these cells obtain ~5-fold greater tumor initiating capacity as compared to those sorted for CD44+/CD24- alone [130]. In addition to this, it has been reported that patients with elevated levels of ALDH1 in tumor have overall poorer survival [131].
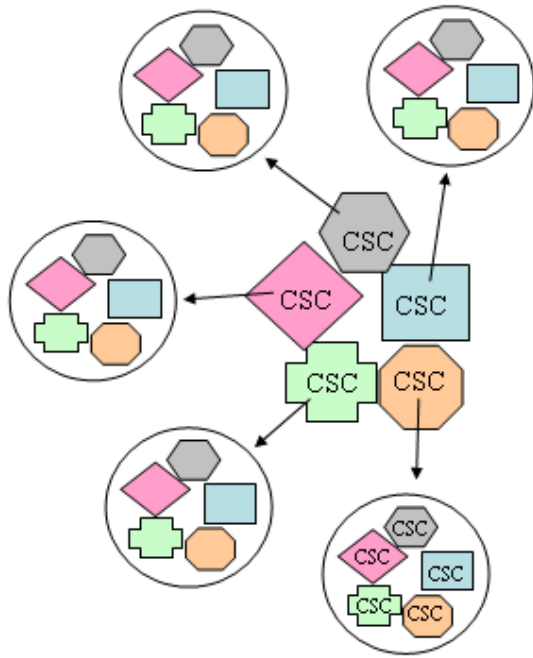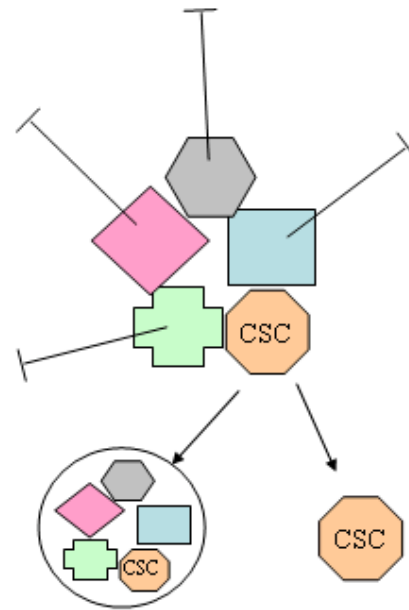
Figure 7. The "Stochastic" and "Hierarchal" models of cancer stems cells. (CSCs). There are 2 theories that exist for CSCs referred to as stochastic and hierarchical models. The stochastic model suggests that cancer cells are dynamic, and through gene expression changes any cell can acquire properties necessary to initiate de novo tumor formation. The hierarchal model suggests that the CSCs are more static and fixed, and only a few distinct cells obtain the potential to be tumor initiating.

CHAPTER 2

Identification of lncRNAs regulated by TFGβ in mouse mammary epithelia and strategy to identify human orthologs

Introduction

It has been well documented that metastasis is the most common cause of death from human cancer. TGFβ has been shown to play a pivotal role in cell migration, invasion, EMT and metastasis. A number of protein coding genes and miRNAs have been shown to be regulated by TGFβ and mediate TGFβ action and phenotypes. However, lncRNAs regulated by TGFβ and their role in EMT and tumor metastasis has not been well documented.

The miR-200 family (miR-200a, miR-200b, miR-200c, miR-141, and miR-429) is now well established as a key negative regulator of EMT through targeting ZEB1 and ZEB2, considered to be key drivers in the process [132, 133]. Interestingly, ZEB1 can also negatively feedback on miR-200 family through direct binding to the promoter regions and transcriptionally repress expression through the recruitment of co-factors. [134] Although loss of miR-200 family results in increased EMT and intravasation from primary site, in a not well understood process re-expression of miR-200 has been shown to be critical for MET and colonization at distant site. For example, there is

higher miR-200 expression in the highly metastatic mouse cell line 4T1, which is capable of colonizing and forming macrometastases, compared to 4T07 which can form micrometastases but not colonization. Ectopic expression of miR-200 in 4T07 cell line enhanced MET and formation of macrometastasis at distant site, primarily through targeting Sec23a [135, 136]. Other miRNAs such as MYC induced miR-9, TGFβ induced miR-155, and miR-221/222 cluster have been shown to be potent promoters of EMT through direct targeting of E-cadherin, RHOA, and TRPS1 [137-139]. In 2008, the first connection between lncRNA, Zeb2-AS, as a mediator of EMT was elegantly described (mentioned above). Recently, a TGFβ-activated lncRNA, lncRNA-ATB, has been shown to play an important role in TGFβ-mediated EMT by acting as a sponge for miR-200 family. LncRNA-ATB is significantly up-regulated in response to TGFβ and contains 3 bonafide miRNA binding sites for miR-200. These sites can act as a miRNA sponge to bind and sequester miR200s, and provide a novel mechanism for ZEB1/2 up-regulation and initiate EMT [89].

Noncoding RNAs can have a significant influence in the biological processes associated with cancer stems cells such as pluripotency, self renewal, and differentiation. In several cancers, including pancreatic, prostate, and gastric, the p53-induced miR-34 family (miR-34a, miR-34b, and miR-34c) has demonstrated to suppress tumor formation and tumor initiating cells [140-142]. miR-34 directly targets several key tumor promoting genes which likely all contribute to inhibiting this process, such as CD44, C-MYC, BCL-2, MET, CDK4/6, and Cyclin E2 [140, 141, 143-145]. Let-7 and miR-200c have also been shown to be down-regulated in CSCs compared to differentiated state and non-CSC tumor cells [146, 147]. The functional role of lncRNAs

contributing to CSCs has not been very well studied and established. However, there are certainly candidate lncRNA genes that can contribute to tumor initiating population given there implications in other studies. For example, lncRNA H19 acts a miRNA sponge for Let7-family and inhibit Let-7 function leading to upregualtion of Let-7 target genes. [51] Given that let-7 loss has previously been implicated in driving CSCs, H19 likely contributes to this process.

It is well established that the NMuMG cell line undergoes EMT and takes upon more fibroblastic features if treated with TGFβ for 24 hours. We have isolated total RNA from NMuMG cells treated with vehicle (DMSO) or TGFβ for 24 hours. The RNA was reverse transcribed, end-labeled with digoxigenin, and hybridized to NCode Mouse Non-coding RNA Microarray (Invitrogen). This microarray contains high-density oligo based probes designed to interrogate 10,802 unique putative mouse lncRNAs, as well as 25,179 mRNAs from the RefSeq database.

Although the majority of lncRNAs are not conserved from mouse to human, we hypothesized that a small fraction of them are conserved and have a strong functionality. Retention of sequence through evolutionary pressures has been very important for preservation of structure and function in proteins, and it is possible this same occurrence has taken place among lncRNAs. Furthermore, of the significantly induced lncRNAs by TGFβ in mouse, we implemented a system based on 4 criteria to predict a strong potential for ortholog lncRNA in human. Using this strategy, we were able to successfully predict 10 human ortholog lncRNAs and detect expression of 9. Three of these transcripts were also regulated by TGFβ in human primary mammary epithelial cell line HMEC. We further investigated these transcripts.

Results

Profile of lncRNA expression in TGFβ-induced EMT in NMuMG cells. Using the widely employed NMuMG mammary epithelial cells as a model to study TGFβ-induced EMT, we have previously shown that TGFβ up-regulates ncRNA BIC, which processes into miR-155 to contribute to the process of EMT [138]. Therefore, we hypothesized that TGFβ could similarly regulate lncRNAs that control EMT-associated phenotypes. To address this, we treated NMuMG cells with vehicle or TGFβ (5 ng/ml) for 24 hours to induce EMT (Fig 8A). Total RNA was labeled with digoxigenin and hybridized to NCode Mouse Noncoding RNA Microarray. This is a high density oligo array that contains lncRNAs and protein coding genes in order to study the dynamic expression of a large subset of lncRNAs and associated protein coding genes within the same biological sample. This array contains 2 replicates for each gene per sample to ensure signal is consistent and reliable. We have plotted relative intensity signals of replicate spots for lncRNAs in control cell and TGFβ-induced EMT to assess the consistency of signal for lncRNA within given sample. Linear regression analysis shows significance for both vehicle and TGFβ treated replicates indicating that signal within our samples is consistent (Fig 8B-C). The microarray contains replicates for 10,802 lncRNAs in which signal was normalized, averaged, and plotted to assess overall changes in expression in vehicle versus TGFβ induced EMT (Fig 8D).

We determined that lncRNA expression was significantly changed if average signal of replicates changed ≥2-fold from vehicle to TGFβ treated samples (Fig 9A).
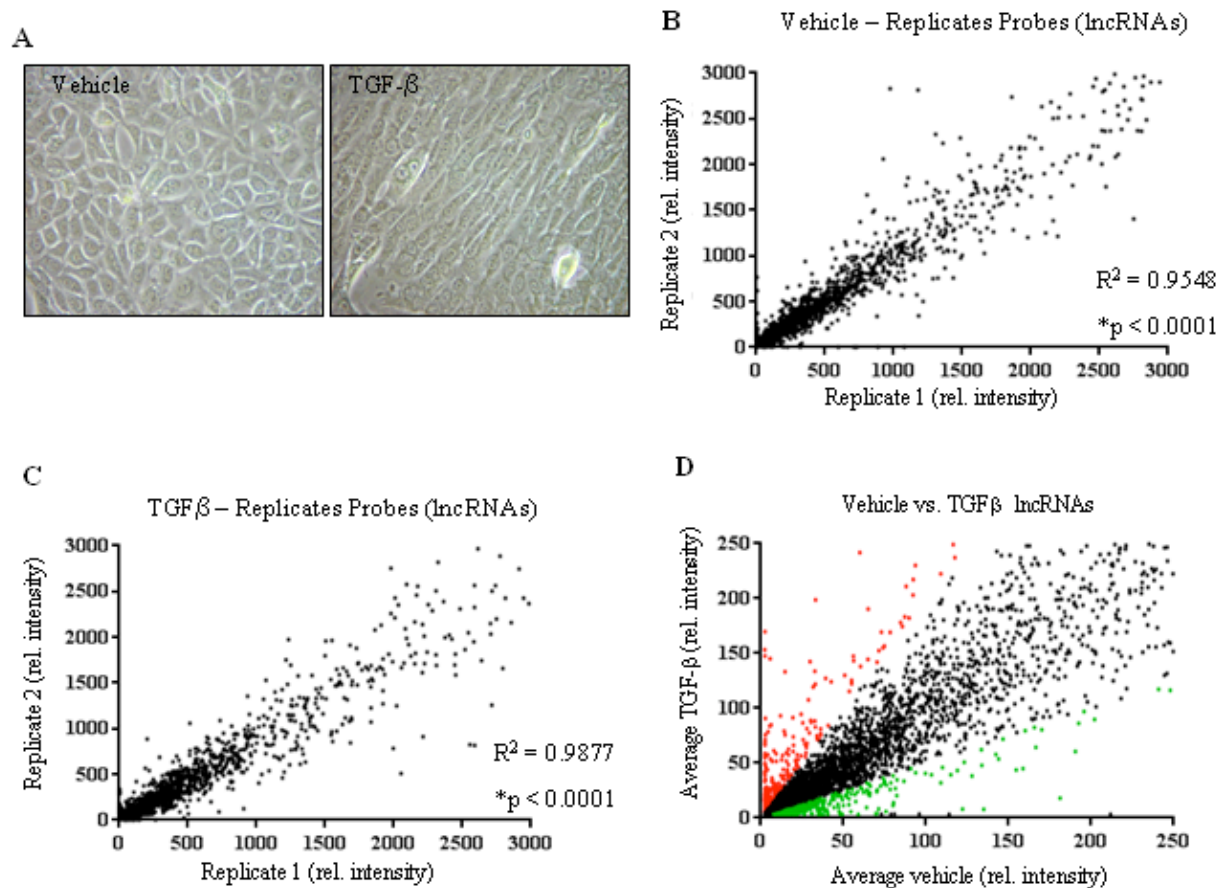
Figure 8. LncRNA expression profile of TGFβ-induced EMT in NMuMG cells. (A) TGFβ induces cell morphological change of EMT in NMuMG cells. The cells were treated with vehicle (left) or 5 ng/ml TGFβ (right) for 24 hours and then photographed. (B - C) Linear regression analysis of lncRNAs in 2 replicates after vehicle (B) and TGFβ (C) treatment of NMuMG cells for 24 hours. (D) Scatter plot analysis comparing all lncRNA changes vehicle vs. TGFβ-treated NMuMG. Note: red and green dots represent up-regulated and down-regulated lncRNAs, respectively.

Using ≥2-fold change as criteria, 633 lncRNAs were down-regulated and 680 lncRNAs were up-regulated following TGFβ-induced EMT (Fig 9B). Furthermore, we confirmed the expression of several top changed lncRNAs by real-time PCR (Fig 9C-D). Tables 3

and 4 summarize top 15 up-regulated and 15 down-regulated lncRNAs. (Tables 3 and 4).
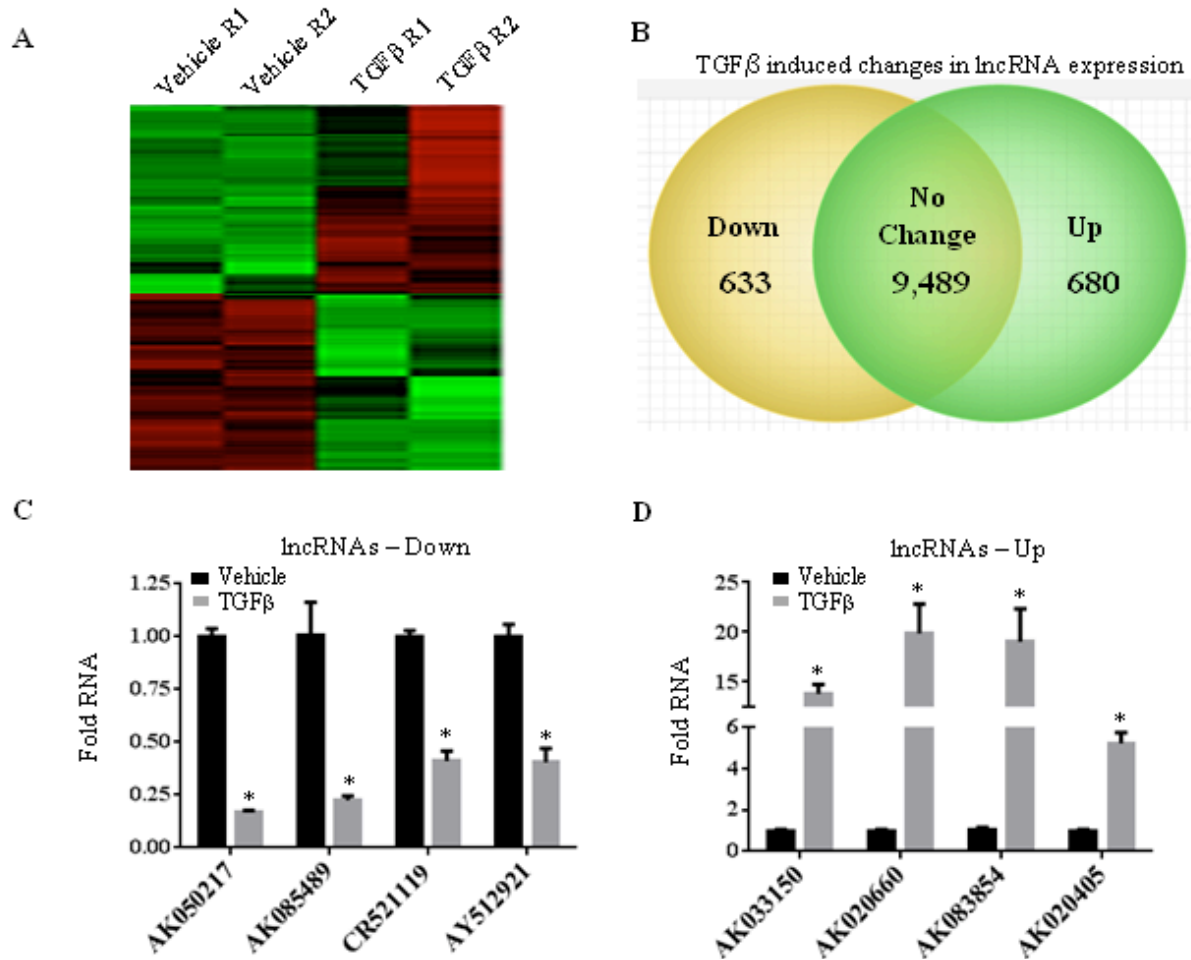
Figure 9. Verification of TGFβ-regulated lncRNAs. (A) Heatmap and (B) diagram representations of dysregulated lncRNAs between vehicle- and TGFβ-treated NMuMG cells. (C) Real-time PCR analysis of representative TGFβ-upregulated and (D) downregulated lncRNAs.

In 25,179 mRNAs represented on the chip, signal for each probe was averaged and plotted to assess overall changes in EMT response (Fig 10A). Genes were considered significantly changed if average fold change was greater than 2-fold. In total,

1,096 genes increased and 2,220 decreased expression upon TGFβ-induced EMT (Fig 10B). Further RT-qPCR validation showed significant changes of several top scoring and established typical TGFβ target genes including Col1a1, Ctgf, Itga5, Serpine1, Kif12, Rbp4, ect., indicating that the lncRNA and mRNA data generated in this study are dependable (Fig 10 C-D).

Strategy used to discover conserved human ortholog lncRNAs. In total, 680 lncRNAs were up-regulated 2-fold or greater. The ultimate goal of this project is to identify lncRNAs important in driving human cancer development and/or metastasis. Therefore from the data generated using the mouse lncRNA microarray, we developed a strategy to predict mouse lncRNAs with a strong potential to have a conserved human ortholog. To do this, we designed a workflow to which we subjected significantly TGFβ-induced mouse lncRNAs (Figure 11). These mouse lncRNAs had to fit 4 criteria or be thrown out. If all 4 criteria were met, we considered these transcripts strong candidates to be conserved in human and potentially across other species. In this study, we focused on transcripts induced and not repressed (Figure 11), as these genes could be knocked down for functional analysis in both mouse and human. These 4 criteria are:

1. **Mouse lncRNA was induced by TGFβ >2-fold in NMuMG by follow up RT-qPCR validation**. This criterion served to validate the array and eliminate false positive data points which may have been introduced during statistical analysis. This also confirmed that the mouse lncRNA is regulated by an established signaling pathway (TGFβ pathway) and associated transcription factors. Due to

the high conservation of lncRNA promoter elements, lncRNAs are more likely to be functional if they are regulated by established transcription factors which are tightly regulated in the cell.

2. **Mouse lncRNA has significant sequence identity to human**. Mouse lncRNA sequence was used as query and BLAST analysis (blast.ncbi.nlm.nih.gov/ Blast.cgi) was performed against human genome (all assemblies) with a "semi-discontiguous" cutoff. The BLAST was against the genome because using this strategy we were likely going to identify novel regions within the human DNA that contained sequence similarity, however did not contain an annotated gene. Currently, there are only about 3,991 lncRNAs annotated in NCBI database [148]. Also, a "semi-discontiguous" cutoff was used because it was unlikely there would be exact sequence identity across the entire length of the gene, but rather many short or some long stretches of high complementarity. In relation to the query, we refer to the length of the stretch as "coverage", and the amount of sequence similarity as "identity".

3. **The BLAST query must locate at syntenic loci in human genome**. Strong sequence coverage and identity that also locates at syntenic loci in human genomes provide strong evidence that this lncRNA has been retained through selective pressures of evolution. It is also important to check if this same sequence identity locates at syntenic loci across species, this will further support the argument for evolutionary conservation. If the BLAST sequence does not map to syntenic loci, this could be due to redundancies and repeat elements within the genome that may be transposable elements. Although these regions

(i.e. Alu repeats and SINEs) may be functional in their own right, this study focuses to characterize lncRNA genes and not transposable elements.

4. **Mouse lncRNA and identified genomic location in human can not overlap with a protein coding gene**. This is to eliminate doubt that this lncRNA is not actually a conserved region within a protein coding gene; i.e. exonic, intronic, and untranslated regions (UTRs). Proteins retain high sequence similarity across species. It is important that the mouse and putative human ortholog are free standing and intergenic, to eliminate any doubt that lncRNA is associated with a protein coding transcript. There are examples of functional antisense and sense lncRNAs that reside within and overlap with protein coding mRNAs, however these were not considered in this study.

Test for expression of predicted human ortholog transcript. We have selected the top 200 TGFβ-induced mouse lncRNAs identified in microarray to be further analyzed using the 4 criteria described above. Using this strategy only 10 of the 200 mouse lncRNAs met all 4 criteria. Using the largest stretch of known conserved sequence region in human, real-time PCR (RT-qPCR) primers were designed within this region to detect presence of a transcript in human cells and its regulation by TGFβ. To test this, we treated the primary human mammary epithelial cell, HMEC, with vehicle or TGFβ for 24 hours and performed RT-qPCR. The aim of this strategy would be to predict orthologs that are co-regulated by TGFβ in mouse and human, and provide an avenue to demonstrate primers designed were on target. We can further validate this through use of siRNAs.
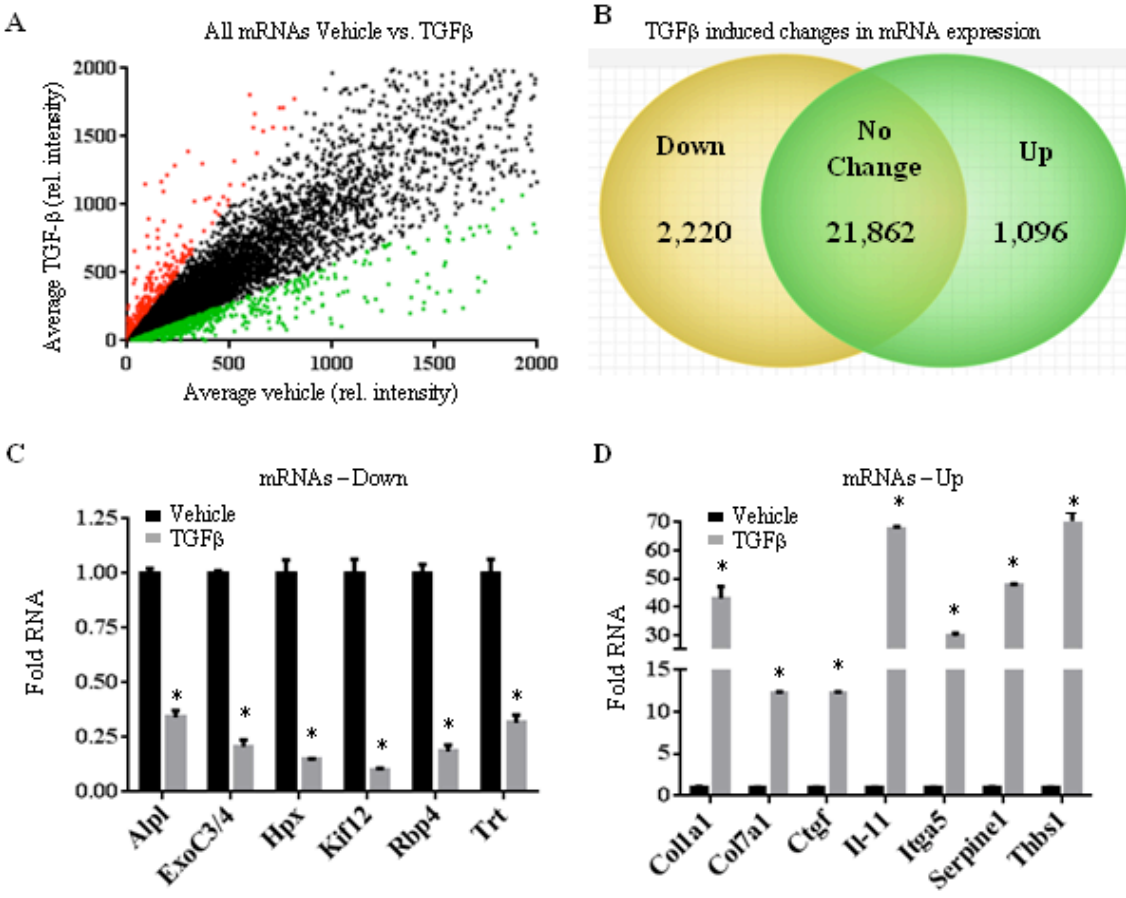
Figure 10. Dysregulated mRNAs in TGFβ-induced EMT. (A) Scatter plot analysis comparing all mRNA changes vehicle vs. TGFβ-treated NMuMG cells (red and green dots represent up-regulated and down-regulated mRNAs, respectively). (B) Diagram summarizing number genes changed. (C & D) Real-time PCR analysis of representative TGFβ down-regulated (C) and up-regulated (D) mRNAs. Asterisks represent $p < 0.05$.

The stretches of similarity ("coverage") typically ranged from ~100nt to 1kb, and therefore designing RT-qPCR primers for a 100-150 base amplicon was not a problem. It is also important to note that all samples were DNase treated prior to reverse transcription step to eliminate genomic contamination and signal during PCR step.

Interestingly, we observed that 3 of the 10 predicted human lncRNA orthologs were also induced greater than 2-fold by TGFβ in HMEC (Fig 12A).

4 Criteria used to identify mouse lncRNAs with strong potential for human ortholog

1. Up-regulated by TGF-β in mouse
Known mouse transcript is validated as being up-regulated at least 2-fold by TGF-β (induced via established transcription factors)

2. Sequence conservation from mouse to human
BLAST (discontiguous megablast) mouse transcript to human genome, demonstrates significant sequence identity and coverage is > 10% of total input transcript

3. Conserved sequence maps to syntenic loci in human genome
BLAST hit maps to snytenic loci in both mouse and human genome

4. No overlap with protein coding transcript
Both mouse and human transcripts can not overlap in sense with coding or intronic sequence of protein coding gene (avoid protein coding associated transcripts).

Hypothesis:
If all criteria are met, there is a strong case that a functional conserved lncRNA in human resides within identified genomic location

of top **200** mouse lncRNAs induced >2-fold by TGF-β in mouse NMuMG,

10 candidates met all 4 criteria,

3 were induced >2-fold by TGFβ in human HMEC

1 resided in genomic region of interest

Figure 11. Schematic representation and workflow for identifying conserved human orthologs. Each significantly up-regulated lncRNA in microarray was subjected to 4 criteria. If all criteria were met, RTqPCR primers were designed for human lncRNA to check changes in expression in TGFβ-treated HMEC cells.

The 3 mouse lncRNAs that have a human ortholog also regulated by TGFβ are AK578, BB049, and AK060. Interestingly, BB049 maps to annotated human lncRNA

WDFY3-AS2. In addition, although human ortholog of AK562 was not induced greater than 2-fold by TGFβ in HMEC, we were still interested in this transcript given its genomic location within the 5-prime distal region of the HOXA locus. Several conserved and functional lncRNAs have been shown to reside in HOX gene clusters, and therefore we did not want to rule out the possibility that a functional lncRNA locates at this location [48, 149, 150]. We next checked several established breast cancer cell lines for expression of each of the 4 conserved lncRNAs. Interestingly, all 4 human orthologs were over expressed in human breast cancer cell lines by RT-qPCR in comparison to HMEC control, and each had a unique expression pattern (Fig 12B).

Human AK578 is over-expressed in human breast cancer patient. To provide a proof of concept for how we settled on these 4 candidates, we will retrospectively discuss AK578. Given the fact that if mouse AK578 is (1) regulated by TGFβ, (2) shares significant sequence identity with human, (3) maps to syntenic loci, and (4) this region does not overlap with any protein coding genes, we hypothesized that there was a strong chance a human ortholog existed.

Mouse AK578 is ~2.8 kb transcript that locates on mouse chromosome 13 and resides in an intergenic region with no overlap with protein coding genes. AK578 was BLAST-ed to the human genome on "semi-discontiguous" cutoff, and this showed that human genome contains a conserved region at syntenic loci in chromosome 5q14 (Fig 13A). This region of the genome is conserved in human with 68% "coverage" and 85% "identity" to mouse AK578. This genomic region, to varying degrees, is conserved across several other species, including chimpanzee, rat, bull, and even zebrafish (Fig

13B), suggesting that this region has been selected through evolutionary pressures and contains a functional element.
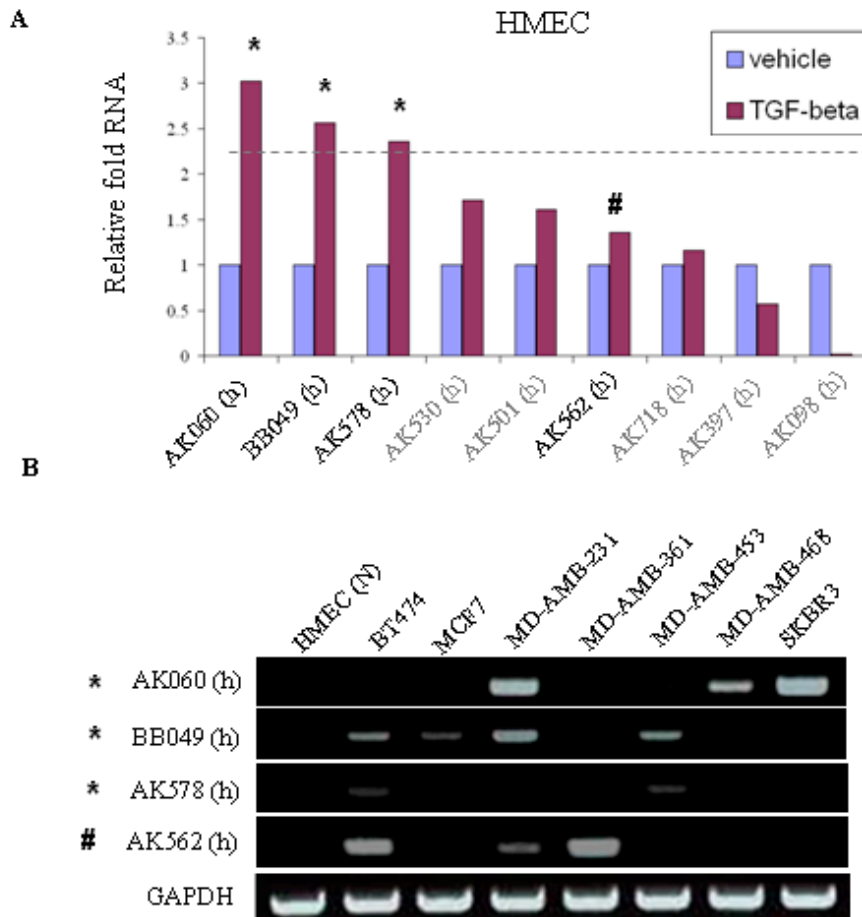
Figure 12. Ortholog lncRNAs are regulated by TGFβ and overexpressed in breast cancer cell lines. (A) Basal level expression was detected for all 10 predicted orthologs, including 3 were regulated by TFGβ in HMEC (*). A fourth ortholog remained of interest due to the fact that it's genomic location in the very well conserved HOXA locus (#). (B) All 4 human orthologs show elevated expression across a panel of established breast cancer cell lines in comparison to normal cell HMEC.

RT-qPCR primers were designed for mouse and human AK578. NMuMG and HMEC cells were treated with TGFβ (5 ng/ml) for 24 hours and subsequently assessed

for AK578 expression. As indicated in microarray analysis, mouse AK578 was validated as being induced by TGFβ >2.5-fold. Interestingly, we were able to detect ~3-fold induction of a human ortholog when using qPCR primers designed within conserved human region (Fig 13C). These data suggest that a conserved human transcript regulated by TGFβ does exist and our strategy was successful. We also obtained human patient breast normal and tumor RNAs from Moffitt Cancer Center Tissue Core to assess expression of human AK578. We performed semi-quantitative qPCR (semi-qPCR) and observed significant elevated expression of AK578 in patient breast tumor (T) compared to normal control (N) RNA samples (Fig 13D). This result suggests that AK578 expression is relevant in human and may functionally contribute to breast cancer development and metastasis.

The following chapters will discuss functional characterization and mechanisms of the other 3 mouse and human ortholog candidates AK562, BB049, and AK060.

Methods

LncRNA Microarray. NCode™ Mouse Noncoding Microarray chip from Life Technologies (Carlsbad, CA, USA), which contains 10,802 lncRNAs and 25,178 protein-coding genes, was used to interrogate lncRNA and mRNA changes in vehicle- versus TGFβ-treated (5 ng/ml) NMuMG cells after 24 hours. TGFβ was purchased from R&D Systems (Minneapolis, MN, USA). Total RNA was end labeled, and hybridized to the array. Hybridization and analysis were performed in MicroArray Core at H. Lee Moffitt Cancer Center.

RNA isolation and RT- and semi- qPCR polymerase chain reaction. Total
RNAs were isolated with Trizol reagent following manufacturer's protocol (Life
Technologies and then subjected to RT reaction using High Capacity cDNA Reverse
Transcription Kit (Life Technologies, Carlsbad, CA, USA). The RT product was used for
subsequent qPCR. The qPCR was performed with SYBR Green 2x Master Mix (Life
Technologies Carlsbad, CA, USA) on ABI HT9600 from Applied Biosystems (Foster City,
CA, USA) and the data were collected and analyzed using ABI SDS version 2.3. ΔCTs
were normalized to GAPDH, and $\Delta\Delta$CT analysis was performed to calculate relative
RNA expression.

A

Mouse
Ch13

Human
Ch5

AK578

BLAST

B

Conservation AK578 across species

Species	Identity	Coverage	Chromosome
Mus musculus	100	100	13
Homo sapiens	85	68	5
Ratus norvergicus	91	94	2
Pan troglodytes	82	62	5
Bos taurus	81	69	14
Danio Rerio	87	6	10

C

Human Breast Patient RNA

AK578 (h)
GAPDH
N

AK578 (h)
GAPDH
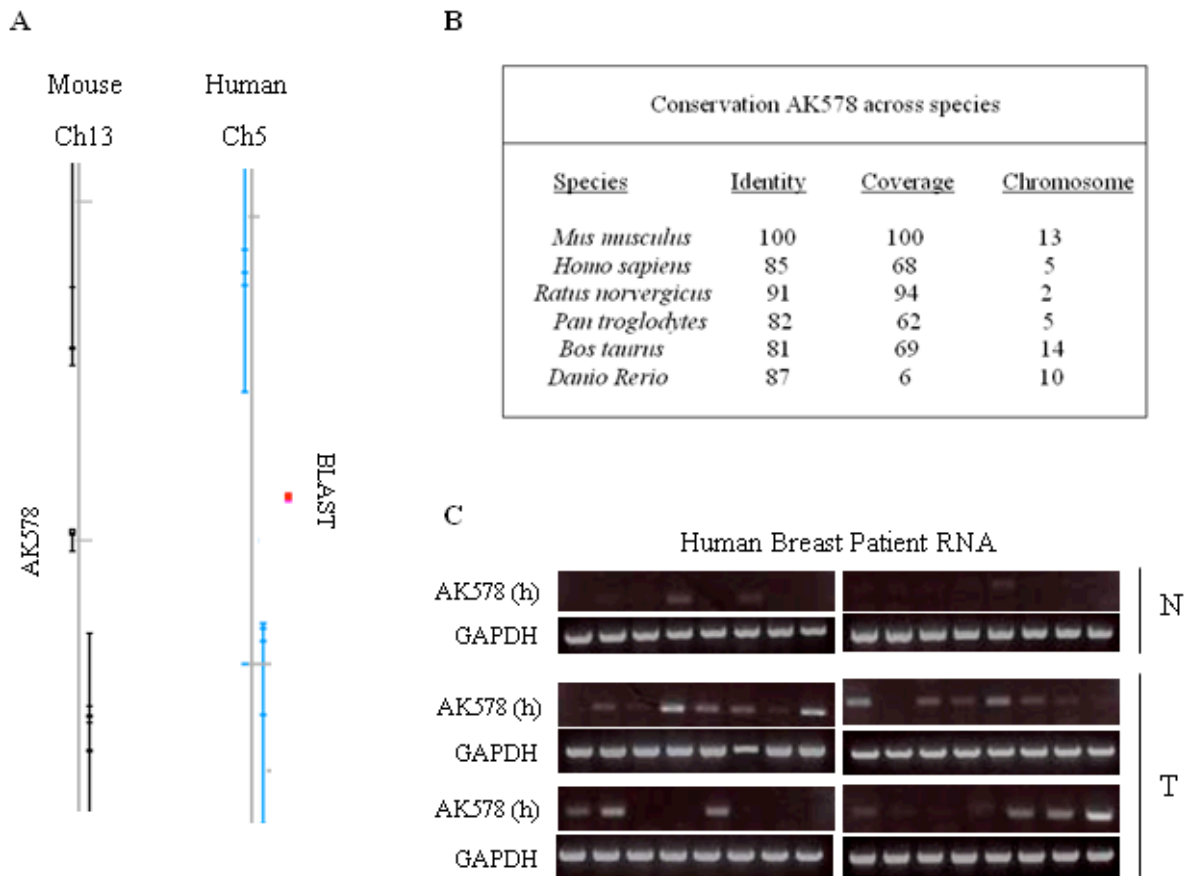
AK578 (h)
GAPDH
T

Figure 13. Mouse AK578 has human ortholog transcript and is up-regulated in patient breast tumor. (A) Mouse AK578 and AK578 (h) locate at syntenic loci are intergenic sharing no overlap with protein coding gene. (B) Mouse AK578 is very well conserved across several species in sequence identity and coverage including zebrafish (*Danio rerio*). (C) AK578 (h) has elevated expression in 16 human breast tumor RNAs examined in comparison to 8 normal breast controls.

Table 3. Top up-regulated lncRNAs following TGFβ-induced EMT in NMuMG.

Gene ID	Vehicle (ave. relative intensity)	TGF-β (ave. relative intensity)	Average Fold Change	Chromosome	Length (nt)
AK171786	3.19	23,171.79	7257.60	18	3,372
AK043331	8.67	8287.23	954.83	9	2,462
AK160425	2.48	169.80	68.27	5	2,013
AK085133	2.28	153.48	67.25	3	2,662
AK014868	2.46	147.81	60.06	X	1,979
AK039984	2.67	90.99	34.03	11	986
AK020562	2.42	68.74	28.38	6	1,245
AK005664	2.61	71.50	27.32	2	652
AK143298	2.49	66.65	26.69	12	3,389
AK035982	5.68	145.23	25.56	5	1,388
AK015200	2.39	58.93	24.59	5	737
AK039295	2.45	55.29	22.53	6	1,385
AK020168	2.67	57.04	21.34	3	1,196
AK019733	2.62	55.21	21.00	10	355
AK016032	2.46	50.17	20.32	7	1,118

Table 4. Top down-regulated lncRNAs in TGFβ-induced EMT in NMuMG.

Gene ID	Vehicle (ave. relative intensity)	TGF-β (ave. relative intensity)	Average Fold Change	Chromosome	Length (nt)
AK032278	38,274.10	3.45	0.0001	1	4,151
AK082403	380.93	3.45	0.009	5	1,925
AK076494	251.28	3.45	0.013	18	2,377
AK016398	211.75	3.45	0.016	14	1,401
AK083030	313.03	6.04	0.019	10	1,958
AK084380	113.95	3.45	0.030	16	4,202
AK036704	95.84	3.55	0.037	10	2,401
AK087232	78.73	3.45	0.043	4	3,346
AK020718	80.89	3.59	0.044	X	644
AK080421	73.04	3.45	0.047	2	1,108
AK039030	69.51	3.99	0.057	1	3,090
AK019328	134.91	7.95	0.059	16	205
AK007385	58.94	3.50	0.059	3	839
AK141015	70.40	4.22	0.059	X	2,752
AK015950	122.10	8.16	0.066	1	483

CHAPTER 3

LncRNA-HIT mediates TGFβ-induced EMT in mammary epithelia and associates with metastasis and poorer overall patient survival

Introduction

The Homeobox genes (Hox genes) are well-conserved across several species and encode for a cluster of transcription factors important in development [151]. In mouse, human, and other vertabrates, the Hox genes are organized into 4 groups; Hoxa, Hoxb, Hoxc, and Hoxd. These genes are spatially and temporally expressed during embryogenesis and development to ensure proper segmentation and bilateral symmetry across the body axis [152]. For this reason, the expression of the Hox genes is very tightly regulated by several factors, as disruption can result in major developmental complications [153]. It has now become apparent that in malignancies such as cancer, Hox genes can become dysregulated and drive tumorigenesis and metastasis. It has also been suggested that Hox gene expression is partly governed by locally residing lncRNAs that can act in "*cis*" or in "*trans*" [48, 149].

In this Chapter, we identified mouse AK562 to be an uncharacterized lncRNA which is located in Hoxa gene cluster and significantly up-regulated in the lncRNA microarray. AK562 resides downstream in the sense direction but not overlapping with

Hoxa13 that was significantly Induced by TGFβ, and thus was named lncRNA-HIT. Depletion of lncRNA-HIT inhibits TGFβ induced migration, invasion, and EMT in both NMuMG and 4T1 cells. LncRNA-HIT is well conserved in sequence and chromosomal location from mouse to human. We were able to detect lncRNA-HIT in human and observed that increased expression directly correlates with breast cancer progression. Furthermore, we showed that human lncRNA-HIT can drive breast cancer by positively regulating HOXA13. The following discusses mouse and human ortholog, lncRNA-HIT.

Results

LncRNA-HIT mediates TGFβ-induced invasion, migration, and EMT in NMuMG cell. LncRNA-HIT is one of the top TGFβ up-regulated lncRNAs, and was of particular interest to us given its genomic location i.e., within the homeobox domain A (HOXA) gene cluster. LncRNA-HIT resides in the sense orientation to the HOXA protein genes and downstream of HOXA13. There is no overlap between these transcripts and a gap of ~1.4 kb between the HOXA13 3-prime termini and lncRNA-HIT 5-prime transcriptional start site (Fig 14A**).** Previous studies have shown that the HOX gene clusters (HOXA, HOXB, HOXC, and HOXD) are very well conserved across species and several lncRNAs have been shown to be functionally important within these regions [48, 149, 150].

Consistent with the microarray finding, RT-qPCR analysis showed that lncRNA-HIT expression was induced by TGFβ (Fig 14B). LncRNAs have been shown to play functional roles in both the nuclear and cytoplasmic compartments [56, 154-156].

Therefore, we performed cellular fractionation after 24 hours of treatment with TGFβ to assess the subcellular localization of lncRNA-HIT. We observed that lncRNA-HIT expression was induced and localized to the nucleus upon TGFβ treatment (Fig 14C), suggesting that lncRNA-HIT functions in the nucleus.
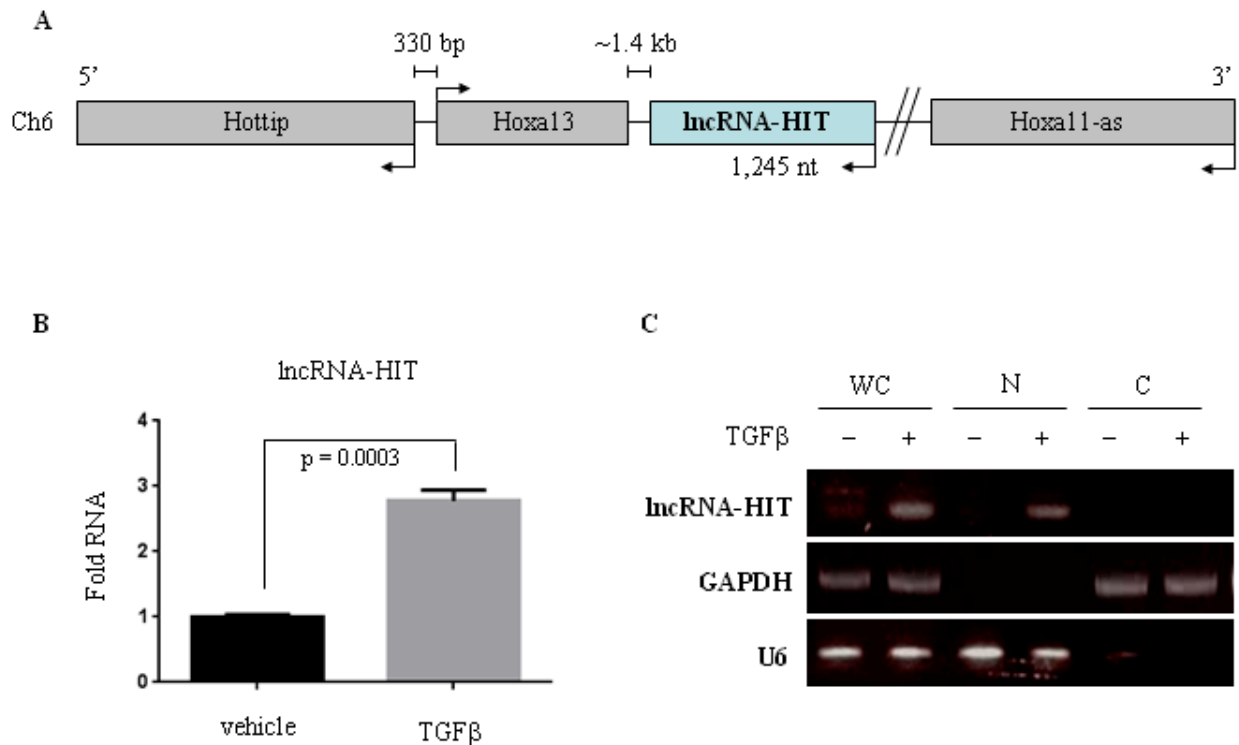
Figure 14. LncRNA-HIT locates in Hoxa gene cluster and is induced by TGFβ. (A) Diagram shows lncRNA-HIT location. (B) RT-qPCR revealed that lncRNA-HIT is induced in NMuMG cells after TGFβ treatment for 24 hours. (C) lncRNA-HIT is predominantly localized in the nucleus. Following treatment with vehicle or TGFβ for 24 hours, NMuMG cells were fractionated and then subjected to semi-quantitative PCR (top), and semi-quantitative PCR with control primers (medium and bottom).

LncRNA-HIT mediates TGFβ-induced motility and EMT. We further designed 2 siRNAs against lncRNA-HIT (si-HIT-1 and si-HIT-2) to experimentally address its

functional significance in TGFβ-induced migration, invasion and EMT. NMuMG cells were transfected with scrambled control siRNA (si-CTL), si-HIT-1, or si-HIT-2 for 48 hours. Cells were then treated with TGFβ (5 ng/ml) or vehicle control for 24 hours. Following confirmation of knockdown of lncRNA-HIT (Fig 15A), we performed two-chamber migration and invasion assays and found that depletion of lncRNA-HIT significantly inhibited TGFβ-induced migration and invasion compared to si-CTL (Fig 15B-C).

TGFβ-induced EMT in NMuMG causes the cell to move from an epithelial to a more mesenchymal and fibroblast-like gene signature. A hallmark of EMT is the loss of E-cadherin and increase of Vimentin expression. To test whether depletion of lncRNA-HIT can abrogate TGFβ-induced EMT, we transfected the cells for 48 h with si-CTL, si-HIT-1, and si-HIT-2 and then treated cells with and without TGFβ for 24 h to induce EMT. Immunoblotting analysis revealed that TGFβ treatment reduced E-cadherin and increased Vimentin in si-CTL cells. However, knockdown of lncRNA-HIT largely overrode TGFβ action in E-cadherin and Vimentin expression (Fig 15D). This result was recapitulated using immunofluorescence staining to visualize expression of E-cadherin and Vimentin (Fig 15E). We noted that TGFβ-disrupted tight junctions were largely restored by knockdown of lncRNA-HIT.

Ectopic expression lncRNA-HIT increases cell motility and EMT. Moreover, we examined the effects of overexpression of lncRNA-HIT alone on EMT in NMuMG cells. Following transfection of lncRNA-HIT into NMuMG cells (Fig 16A), we observed a significant increase in both migration and invasion (Fig 16B-C). Ectopic
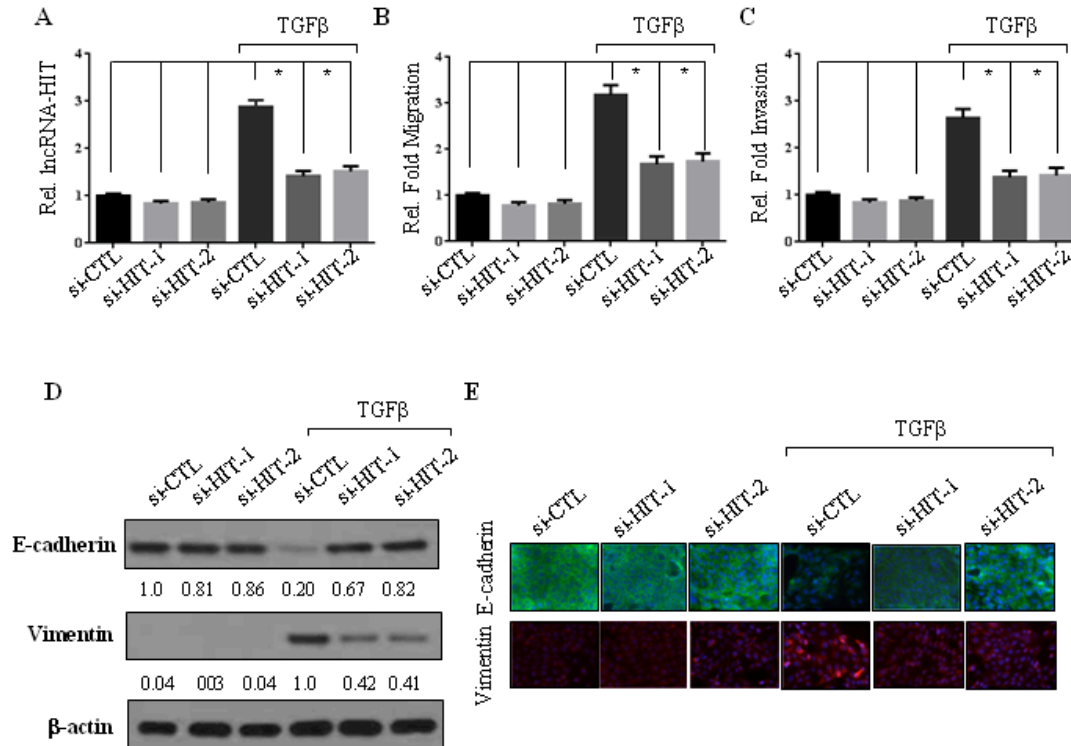
Figure 15. LncRNA-HIT mediates TGFβ-induced cell migration, invasion and EMT.
(A) NMuMG cells were transfected with indicated siRNAs. After treatment with vehicle or TGFβ for 48 hours, RNAs were isolated and analyzed for lncRNA-HIT expression using real-time PCR. (B-C) Depletion of lncRNA-HIT reduces TGFβ-triggered cell migration and invasion. LncRNA-HIT knockdown and control NMuMG cells were seeded in Boyden chamber without (B) and with (C) matrigel to assess changes in migration and invasion. (D-E) Knockdown of lncRNA-HIT abrogates a hallmark change of EMT induced by TGFβ. NMuMG cells were transfected with indicated siRNAs. After 48 hours of transfection, the cells were left untreated or treated with TGFβ for 16 hours. Expression of E-cadherin and Vimentin were assessed by Western blot (D). Immunofluorescence staining was also performed with antibodies against E-cadherin and Vimentin (E). Note: TGFβ-induced tight-junction dissolution was largely rescued by depletion of lncRNA-HIT. Asterisks represent p<0.05.

expression of lncRNA-HIT was also able to disrupt tight junction as indicated by E-cadherin immunofluorescence staining (Fig 16D). Collectively, these data indicate that lncRNA-HIT plays a pivotal role in TGFβ-induced cell migration, invasion, and EMT.
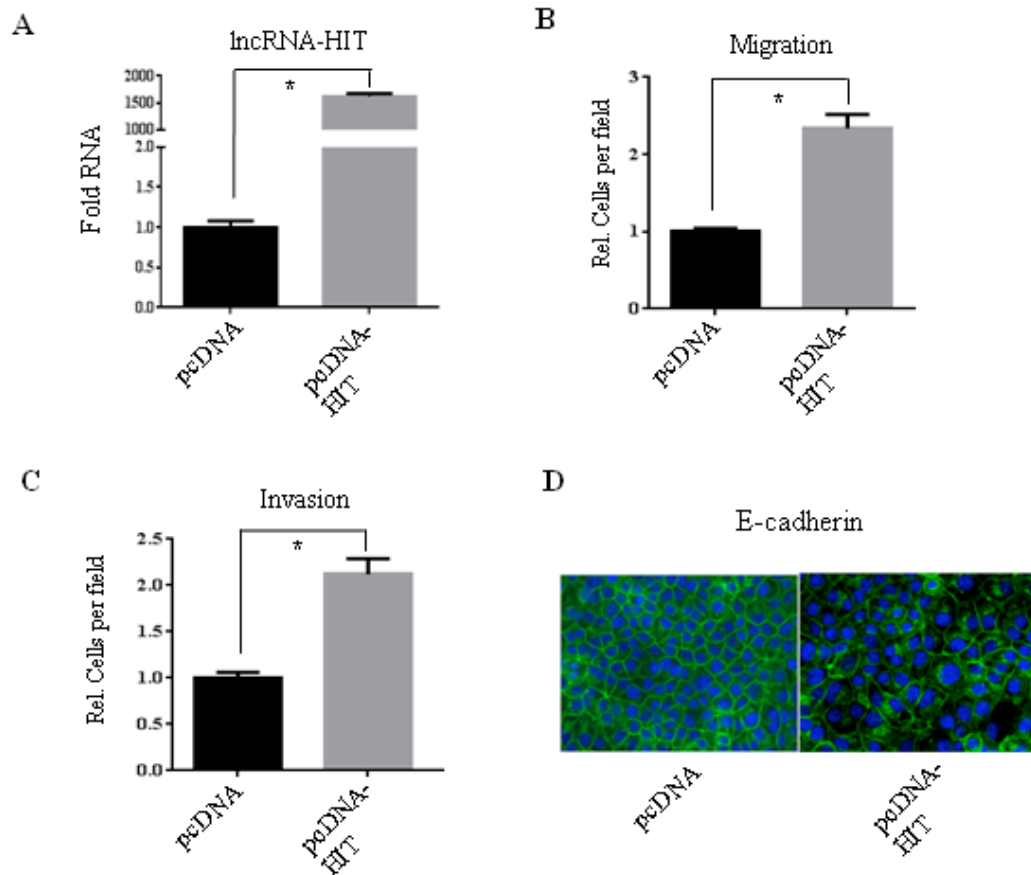
Figure 16. Over-expression of lncRNA-HIT promotes migration, invasion and disrupts tight junction. (A) LncRNA-HIT was ectopically expressed in NMuMG cells for 48 hours. Cells were then seeded in the upper Boyden Chamber without (B) and with (C) matrigel to determine changes in migration and invasion. (D) Presence of tight junction following overexpression of lncRNA-HIT was assessed using E-cadherin antibody for immunofluorescence staining.

LncRNA-HIT is up-regulated in 4T1 cells and its depletion inhibits cell migration, invasion, lung metastasis and tumor growth. We further examined lncRNA-HIT expression in four well-characterized mouse mammary tumor cell lines (67NR, 168FARN, 4TO7 and 4T1) derived from a single spontaneously arising mammary tumor in a BALB/c mouse. Although each of these tumor cell lines is able to form primary tumors, they have different metastatic properties. 67NR cells form primary

tumors readily, but tumor cells do not intravasate. 168FARN cells can be detected in lymph nodes but rarely in other tissues, suggesting that they can enter the vasculature, but extravasate inefficiently. 4TO7 cells can disseminate from primary mammary tumors into the lungs but don't form visible lung nodules. Moreover, disseminated 4TO7 cells in the lungs rapidly disappear when the primary tumor is removed, suggesting that they are unable to colonize in distant sites. 4T1 cells are fully metastatic and form macroscopic lung nodules from primary mammary tumors. Interestingly, lncRNA-HIT was significantly elevated in the 4T1 cell and expressed at low levels in other 3 cell lines (Fig 17A), suggesting that lncRNA-HIT may be involved in breast cancer metastasis. The importance of lncRNA-HIT in 4T1 cell migration and invasion was further investigated. After cells were transfected with si-HIT-1, and si-HIT-2 as well as si-CTL for 48 h (Fig 17B), two-chamber migration and invasion assays were performed as described above. After 16 hours, we observed a significant reduction in both migration and invasion of 4T1 cells in which lncRNA-HIT was depleted (Fig 17C-D). However, we did not observe that TGFβ induced lncRNA-HIT in 4T1 cells and that knockdown of lncRNA-HIT had no effect on TGFβ-induced 4T1 cell migration and invasion (data not shown).

We next examined the effect of lncRNA-HIT on anchorage independent growth in soft agar following depletion of lncRNA-HIT. Compared to si-CTL treated 4T1 cells, si-HIT-1 and -2 transfected cells resuspended in soft agar and cultured for 2 weeks showed markedly reduced colony formation capacity (Fig 17E). Notably, the orthotopic breast cancer model revealed that knockdown of lncRNA-HIT dramatically inhibited number of metastatic lung nodules, breast tumor volume, and tumor weight (Fig 17F-H).

Identification of E-cadherin as a major target of lncRNA-HIT. To identify the genes regulated by lncRNA-HIT, we performed Affymetrix gene expression analysis after 48 h ectopic expression of lncRNA-HIT in NMuMG (Fig 18A). One of the significant deregulated genes that are associated with EMT was E-cadherin. Accumulating studies have shown that loss of E-cadherin is not only a hallmark of EMT but also a key driver of EMT and metastasis [157-160]. Thus, we further examined if lncRNA-HIT inhibits E-cadherin transcription by performing RT-qPCR and using E-cadherin luciferase promoter assay. Following ectopic expression of lncRNA-HIT, we observed a significant loss of E-cadherin mRNA and promoter activity (Fig 18B-C). Furthermore, this correlated with a loss of E-cadherin protein after 72 h overexpression of lncRNA-HIT (Fig 5D), however we did not see changes of other EMT-associated genes ZEB1 or Snail at this time point. Further, the effects of lncRNA-HIT induced EMT, migration, and invasion were rescued through introduction of ectopic E-cadherin (Fig 18E-G).

LncRNA-HIT is conserved across several species. In "BLAST-ing" mouse lncRNA-HIT to the human genome, we found a significant conservation at syntenic locus in human chromosome 7p15 located downstream of HOXA13 in sense direction (Fig 19A), suggesting that lncRNA-HIT was retained through evolutionary pressures. The conservation from mouse to human genomic sequence is 79% identity across 99% coverage. Interestingly, lncRNA-HIT is well conserved in several other higher organisms at the DNA level. (Fig 19B). Furthermore, to check for evidence and expression at the RNA level in human, we "BLAST-ed" the human genomic DNA sequence to the

expressed sequence tag bank (ESTs). This result showed several reported transcripts residing within this region and thus lncRNA-HIT expression at the RNA level (Fig 19C).
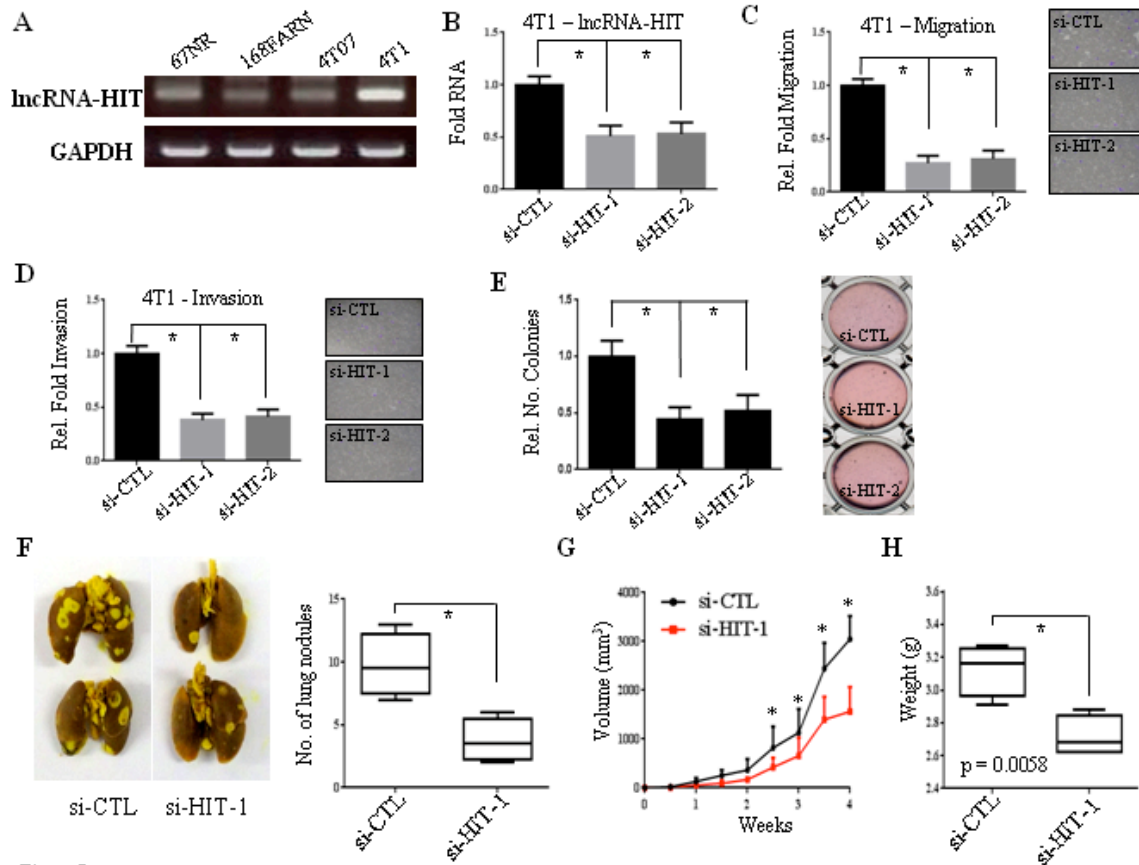
Figure 17. LncRNA-HIT is elevated in metastatic 4T1 cells and its knockdown results in significant reduction of cell migration, invasion, and metastasis. (A) Expression of lncRNA-HIT was evaluated in 4 isogenic mouse breast cancer cell lines 67NR, 168FARN, 4T07 and 4T1 by semi-quantitative RT-PCR. (B) Knockdown of lncRNA-HIT with 2 siRNAs in 4T1 cells. Cells were transfected with indicated siRNAs and then analyzed lncRNA-HIT expression by RT-qPCR. (C and D) Depletion of lncRNA-HIT dramatically reduces 4T1 cell migration and invasion. Following treatment with indicated siRNAs, 4T1 cells were assayed for cell migration and invasion. (E) Soft agar colony formation assay was performed after 4T1 cells were transfected with si-CTL, si-HIT-1, or si-HIT-2. Colony forming capacity was dramatically decreased after 2 weeks in lncRNA-HIT depleted cells. (F) Orthotopic breast cancer model. Control and lncRNA-HIT knockdown 4T1 cells (4 x 10^6) were injected to mammary fat pad of female nude mice and tumor growth was monitored for 4 weeks. Number of metastatic lung nodules (F), tumor volume (G), and tumor weight (H) were assessed at completion of experiment. Asterisks represent p<0.05.
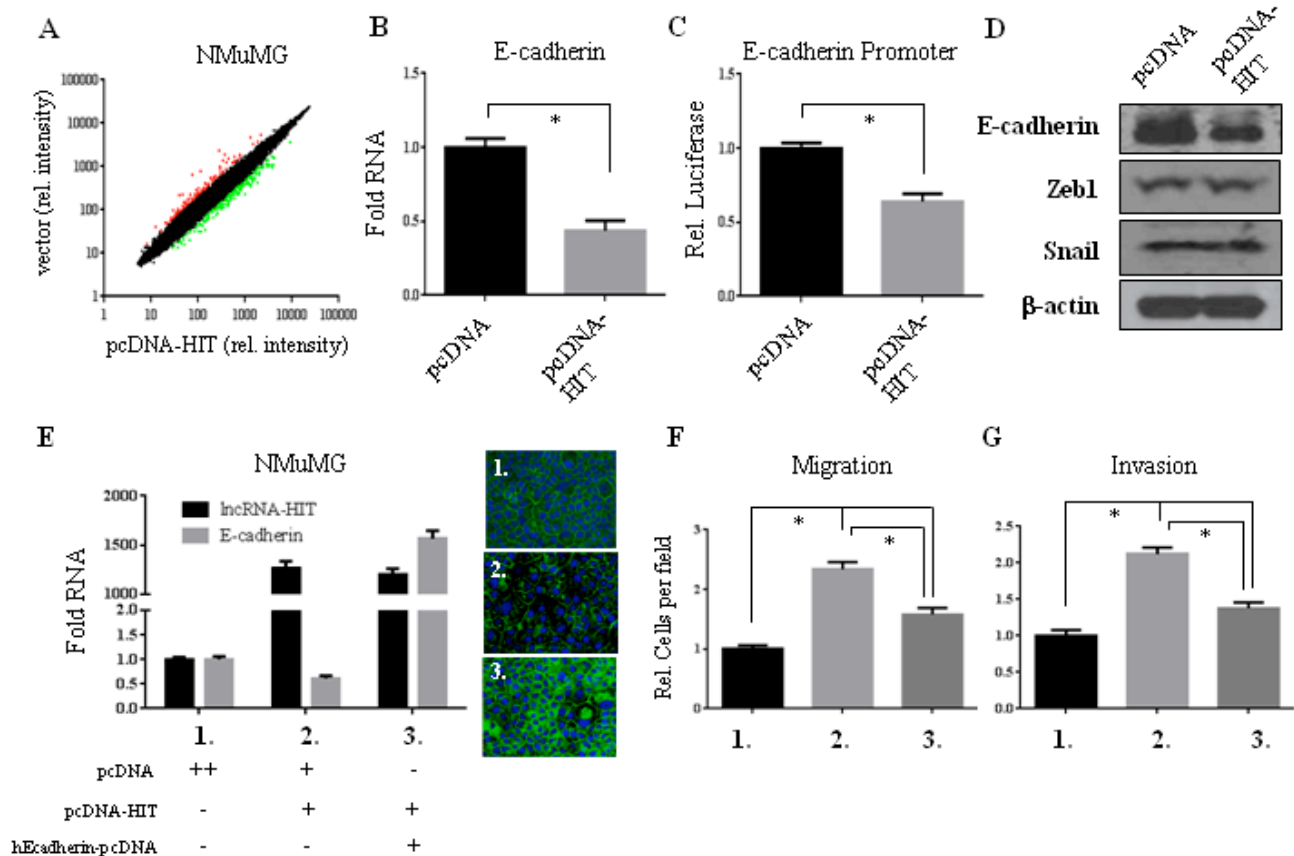
Figure 18. E-cadherin is a target of lncRNA-HIT. (A) Affymetrix Gene Expression Analysis was performed on NMuMG cells transfected with vector and lncRNA-HIT. (B) RT-qPCR was performed to confirm a loss of E-cadherin expression. (C) PGL3-E-cadherin promoter was co-transfected with vector or lncRNA-HIT into NMuMG cell for 48 hours and assayed for luciferase activity. (D) After 72 h overexpression of lncRNA-HIT, expression of E-cadherin, ZEB1 and Snail were examined by immunoblot with indicated antibodies. (E-G) After 72 hours transfection, ectopic expression of E-cadherin was able to rescue lncRNA-HIT effects on loss of tight junction (E), migration (F) and invasion (G). Asterisks represent p<0.05

LncRNA-HIT is elevated in human breast cancer cell lines and knockdown reduces migration and invasion. Since we observed that lncRNA-HIT is conserved in human genome and several ESTs reside in this location, we designed PCR primers and siRNAs within this region to test for expression and knockdown. We performed semi-
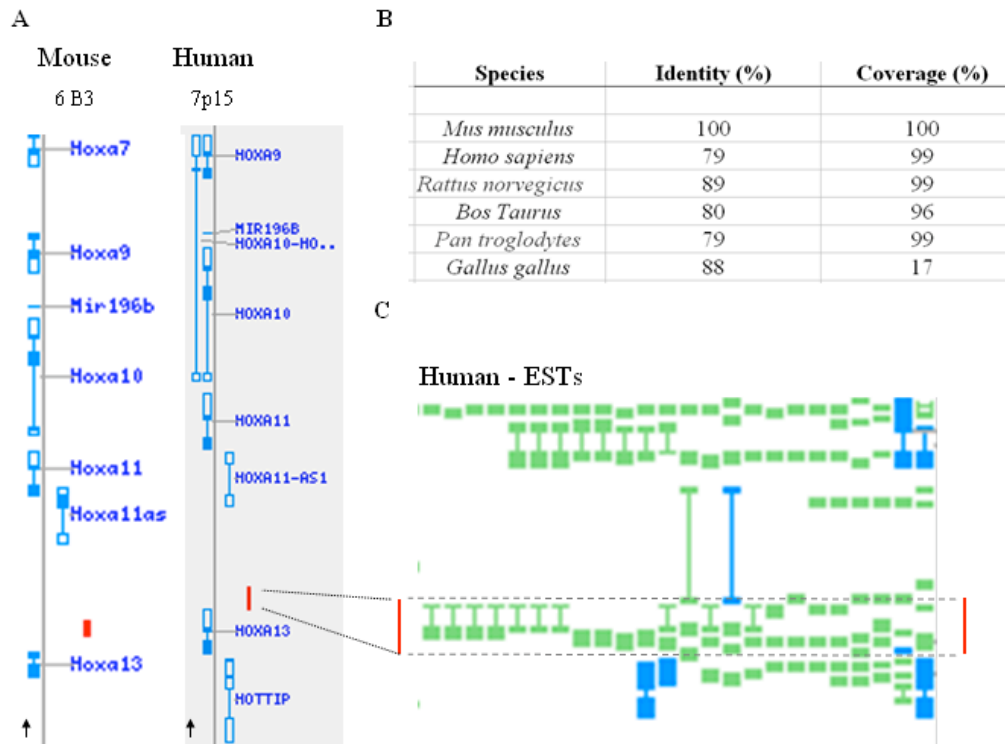
Figure 19. LncRNA-HIT is conserved in across several species and human ESTs have been reported. (A) Mouse lncRNA-HIT was "BLAST-ed" to human genomic DNA, and results show significant similarity and coverage at snytenic loci in human chromosome 7p15. (B) LncRNA-HIT is well conserved in several species at genomic DNA level. (C) In "BLAST-ing" human genomic sequence corresponding to lncRNA-HIT to the EST database, several RNA transcripts have been reported in this location.

quantitative PCR (semi-qPCR) in a panel of established breast cancer cell lines and observed up-regulation of lncRNA-HIT in several including MD-AMB-231, HS578T, BT474, MD-AMB-361, and T47D (Fig 20A). In the highly metastatic MD-AMB-231 we were able to knockdown lncRNA-HIT to 0.32-fold and 0.39-fold expression respectively using si-hHIT-1 and si-hHIT-2 in comparison to control siRNA (si-CTL) (Fig 20B). After knockdown of lncRNA-HIT for 48 h, cells were seeded in upper Boyden Chambers and assayed for changes in migration and invasion (matrigel coated). Both si-hHIT-1 and -2 resulted in significantly reduced migration and invasion after 16 h (Fig 20C-D).

Depletion of lncRNA-HIT decreases primary tumor growth and metastasis

in vivo. Given that we were able to reduce 4T1 cell tumor growth and lung metastasis (Figure 17) and human breast cancer cell migration and invasion *in vitro* setting (Figure 20), we next investigated whether knockdown of lncRNA-HIT could reduce cancer phenotypes *in vivo* in human breast cancer cells. To do this, we employed a mouse orthotopic xenograft model using stable expressing luciferase MD-AMB-231 in
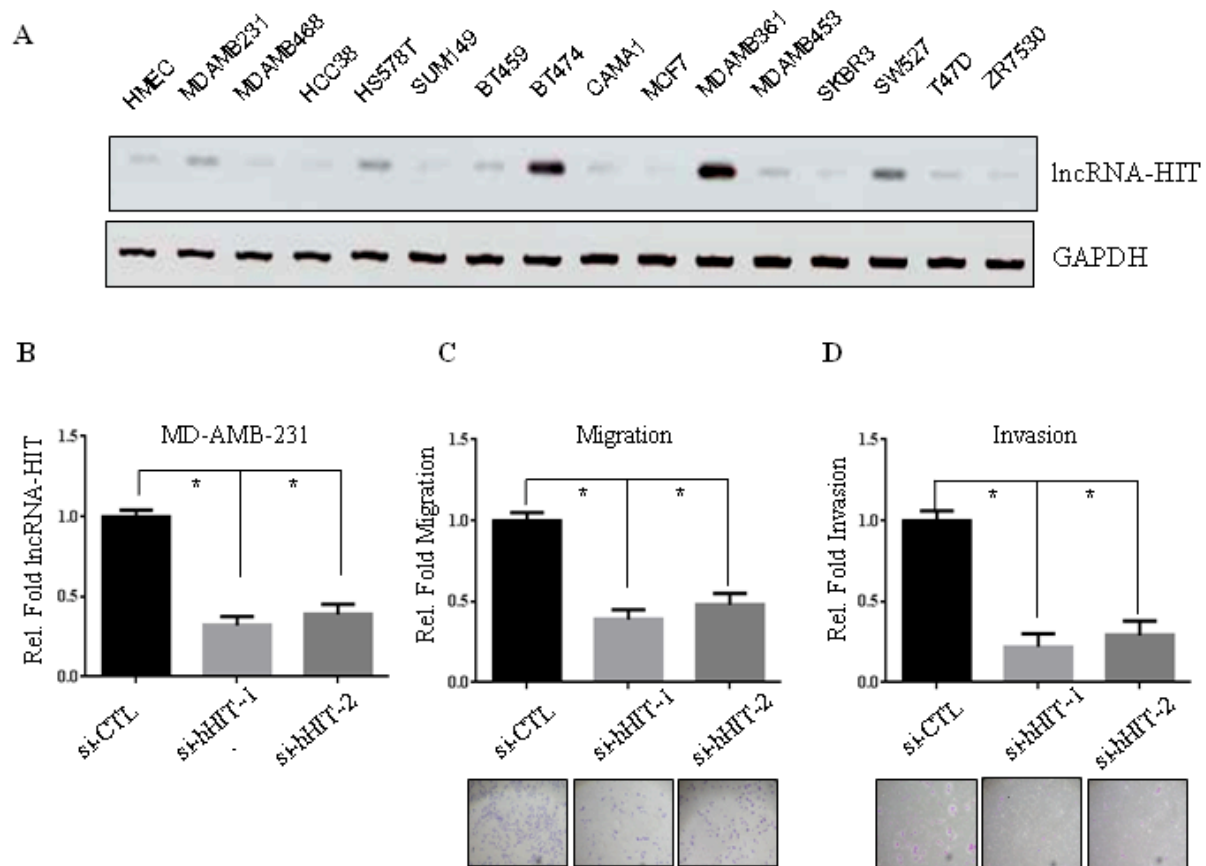
Figure 20. Human lncRNA-HIT is over-expressed in several breast cancer cells lines and depletion of lncRNA-HIT results in reduced migration and invasion. (A) Semi-qPCR was performed in a panel of breast cancer cell lines for expression of lncRNA-HIT. (B) Knockdown of lncRNA-HIT was confirmed by RT-qPCR in MD-AMB-231. After 48 h post-knockdown, cells were assayed for migration (C) and invasion (D).

NOD/Scid gamma (NSG) mice. NSG mice are advantageous in orthotopic cancer models due to germline SCID mutation in the Prkdc gene and an additional targeted mutation in the IL2RG gene (IL2 receptor) [161, 162]. The Prkdc mutation results in mutation of DNA-PK gene product that can no longer function in DNA repair process. This results in B and T lymphocytes being able to successfully undergo V-D-J recombination and therefore can not express function B and T receptors. The IL2 receptor is critical for natural killer cell (NKT) function and this mutation will disrupt the ability of IL2R to recognize the IL2 ligand. Therefore, the NSG mice have no adaptive immunity and a severely impaired innate immunity. These deficiencies allow for reproducible xenografts of human cancer cells, as well as more reliable metastasis models including breast cancer [163].

MD-AMB-231-Luciferase cells were transfected with si-CTL or si-HIT-1 for 48 hours and then washed in PBS and trypsinized. Cells were then spun at low speed and resuspended in PBS at a concentration of 4×10^6 cells per 100 ul. Four NSG mice from each group received injection in the lower mammary fat pad. Tumors were established and tracked over the course of 7 weeks. Similar to mouse 4T1 cells, knockdown of lncRNA-HIT in MDA-MB-231 had significantly slowed tumor growth and volume (mm^3) as measured by standard caliper measurements and as quantified by luciferase activity using Xenogen Bioluminescent Imaging System and Software (Fig 21A-C). At the endpoint week 7, there was also significantly more decreased luciferase activity in the lung in lncRNA-HIT knockdown cell compared to si-CTL (Fig 21D-E). Further, hematoxylin and eosin (H & E) staining revealed less lung nodules in si-HIT group (Fig 21F).

LncRNA-HIT regulates 5-prime distal HOXA genes including HOXA13. A previous report has shown that HOTTIP, a distal 5-prime residing lncRNA, can positively regulate locally residing genes like HOXA13 and HOXA11 through recruitment of the WDR5/MLL1 chromatin remodeling complex [149]. Since lncRNA-HIT also resides in this region we hypothesized that it also may regulate genes within HOXA cluster. To test this, we transfected MD-AMB-231 cell line with si-CTL and si-HIT-1 for
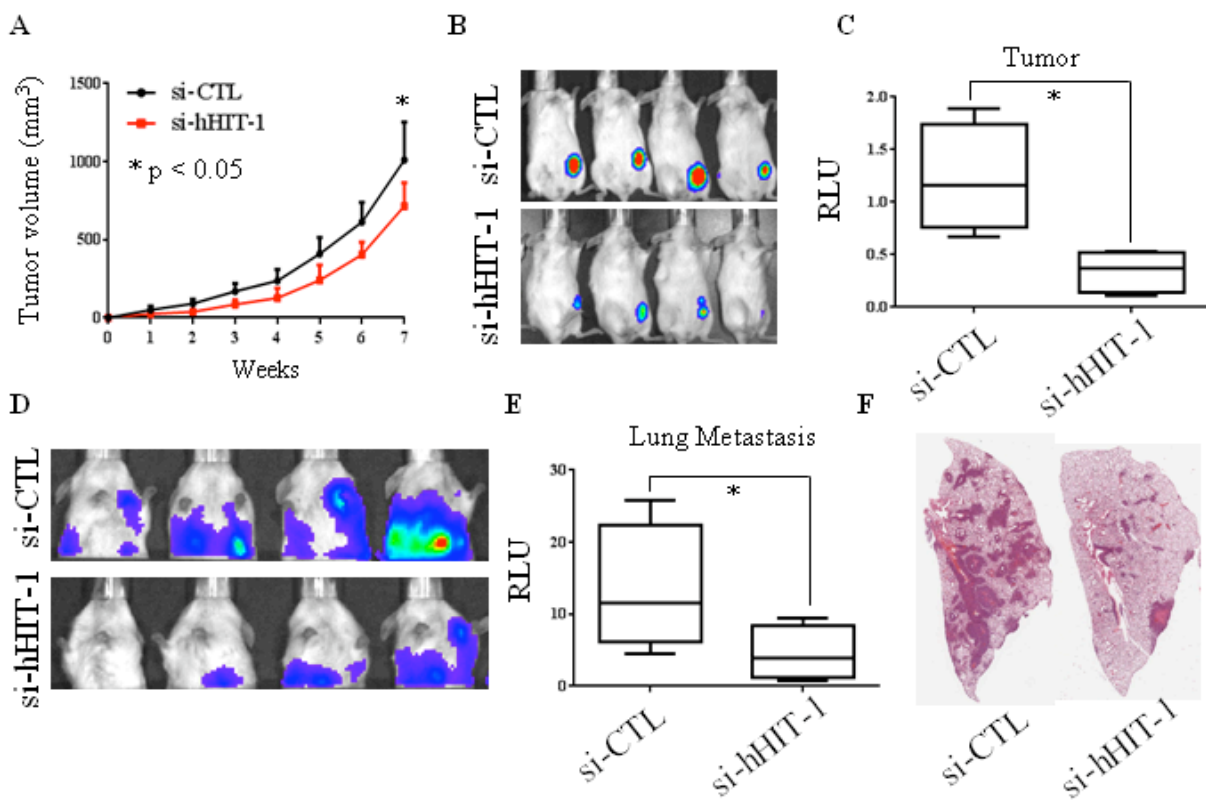
Figure 21. NSG mouse orthotopic model of control versus depleted lncRNA-HIT. (A-C) MD-AMB-231-Luciferase cells were transfected with control (si-CTL) or si-HIT-1 for 48 h, washed, trypsinized, and 4 x 10^6 cells were injected into lower fat pad at day 0. Tumor volume (mm^3) was tracked and quantified via standard caliper measurements (A) and bioluminescent imaging (B & C). Image is representative of primary tumor at week 3 post injection. (D-E) Just prior to sacrificing, lungs were imaged and quantified for luciferase activity. (F) After sacrificing lungs were removed, paraffin embedded, and H & E stained to visualize metastasis. Asterisks represents p<0.05.

48 h, isolated total RNA, and performed RT-qPCR for expression of genes in this cluster. Interestingly we found a similar pattern in that depletion of lncRNA-HIT resulted in decreased RNA expression of HOTTIP, HOXA13, and HOXA11 (Fig 22A). However we observed no significant change in HOXA10, HOXA9, and HOXA7. This suggests lncRNA-HIT is "cis" acting and can positively regulate genes in close proximity. The loss of HOXA13 was also evident at the protein level (Fig 22B).
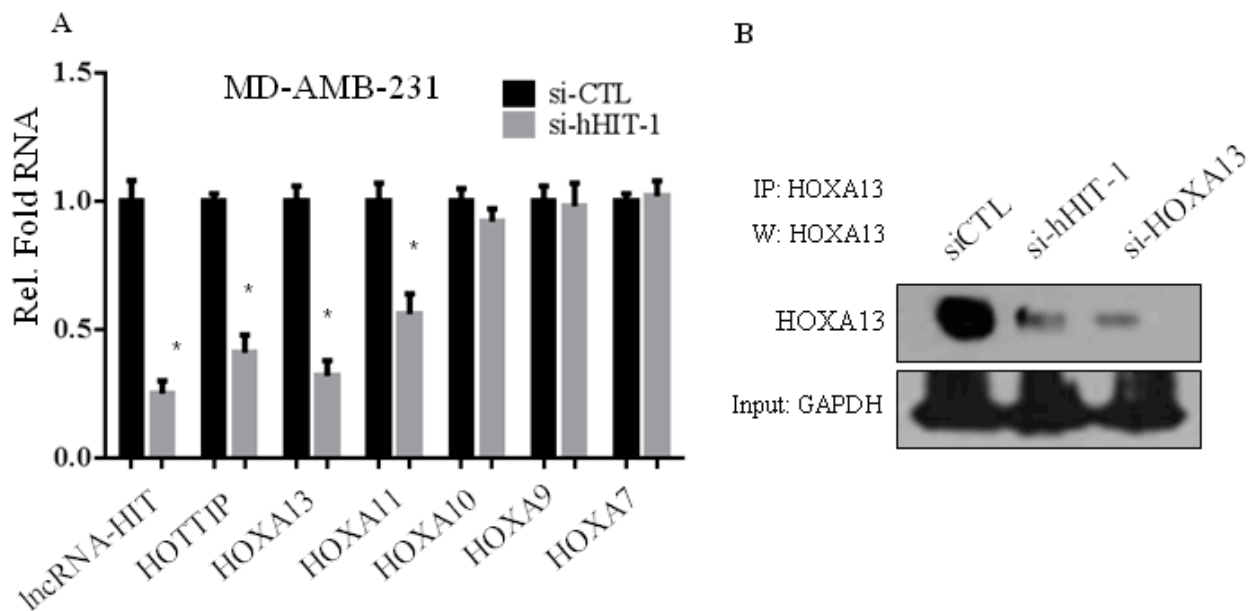
Figure 22 LncRNA-HIT knockdown results in decreased mRNA expression neighboring genes. (A) MD-AMB-231 was transfected with si-CTL or si-hHIT-1 for 48 h, total RNA was isolated. RT-qPCR analysis was performed on a panel of genes within HOXA locus. (B) Similarly transfected cell were lysed and assayed for HOXA13 expression. Asterisks represents p<0.05.

HOXA13 is important in driving breast cancer migration and invasion. It has previously been reported that HOXA13 can be cancer promoting and indicator of poorer

patient survival in esophageal squamous carcinoma and hepatocellular carcinoma [164, 165]. Since lncRNA-HIT regulates HOXA13 mRNA and protein, we aimed to assess if phenotypic changes are associated with loss of HOXA13 gene product. We transfected MD-AMB-231 cells with siRNA against HOXA13 (si-HOXA13). We observed almost complete loss of mRNA as measured by semi-qPCR after 48 h knockdown (Figure 23A). MD-AMB-231 is a triple-negative highly metastatic cell line that phenotypically looks very mesenchymal-like. Interestingly, MD-AMB-231 cells transfected with si-HOXA13 seem to dramatically revert a more rounded, epithelial-like cell shape and state (Figure 23B). The significance of this phenotypic change at the molecular level warrants further research. Loss of HOXA13 also resulted in significantly impaired migration and invasion (Figure 23C and 23D).

LncRNA-HIT is elevated in patient invasive ductal carcinoma, high grade tumor, positive lymph node, and associates with poorer patient survival. We asked if the expression of lncRNA-HIT associated with cancer invasiveness in human primary breast carcinoma. A total of 89 breast cancer specimens (15 noninvasive and 74 invasive breast carcinomas) and 4 normal and 9 hyperplasia breast tissue samples were examined for the expression of lncRNA-HIT. Locked nucleic acid *in situ* hybridization (LNA-ISH) analyses revealed high levels of lncRNA-HIT in 29 of 74 invasive tumors but in only 2 of 15 noninvasive cancer tissues (Fig 24A). The level of expression of lncRNA-HIT in normal breast tissue was low and gradually increased to invasive carcinoma, suggesting lncRNA-HIT may play a role in breast tumor progression (Fig 24B). These data further support the findings demonstrating the

involvement of lncRNA-HIT in EMT and invasion as we previously observed in NMuMG

and 4T1 cells, and suggest that conserved human lncRNA-HIT could play a pivotal role

in breast cancer metastasis.

Figure 23. HOXA13 depletion changes cell morphology and results in decreased migration and invasion. (A) MD-AMB-231 cells were transfected with si-CTL or si-HOXA13 for 48 hours and mRNA changes were measured by semi-qPCR. (B) Depleting HOXA13 results in a change in MD-AMB-231 cell morphology from a mesenchymal-like to a more epithelial-like state. Loss of HOXA13 also resulted in decreased cell migration (C) and invasion (D). Asterisks represents $p<0.05$.

We performed additional TMA-ISH to evaluate if there is clinical or pathological

relevance in patients with higher expression of lncRNA-HIT. Indeed, elevated levels of

lncRNA-HIT correlated with poorly differentiated late-grade breast tumor (p = 0.0036)

70
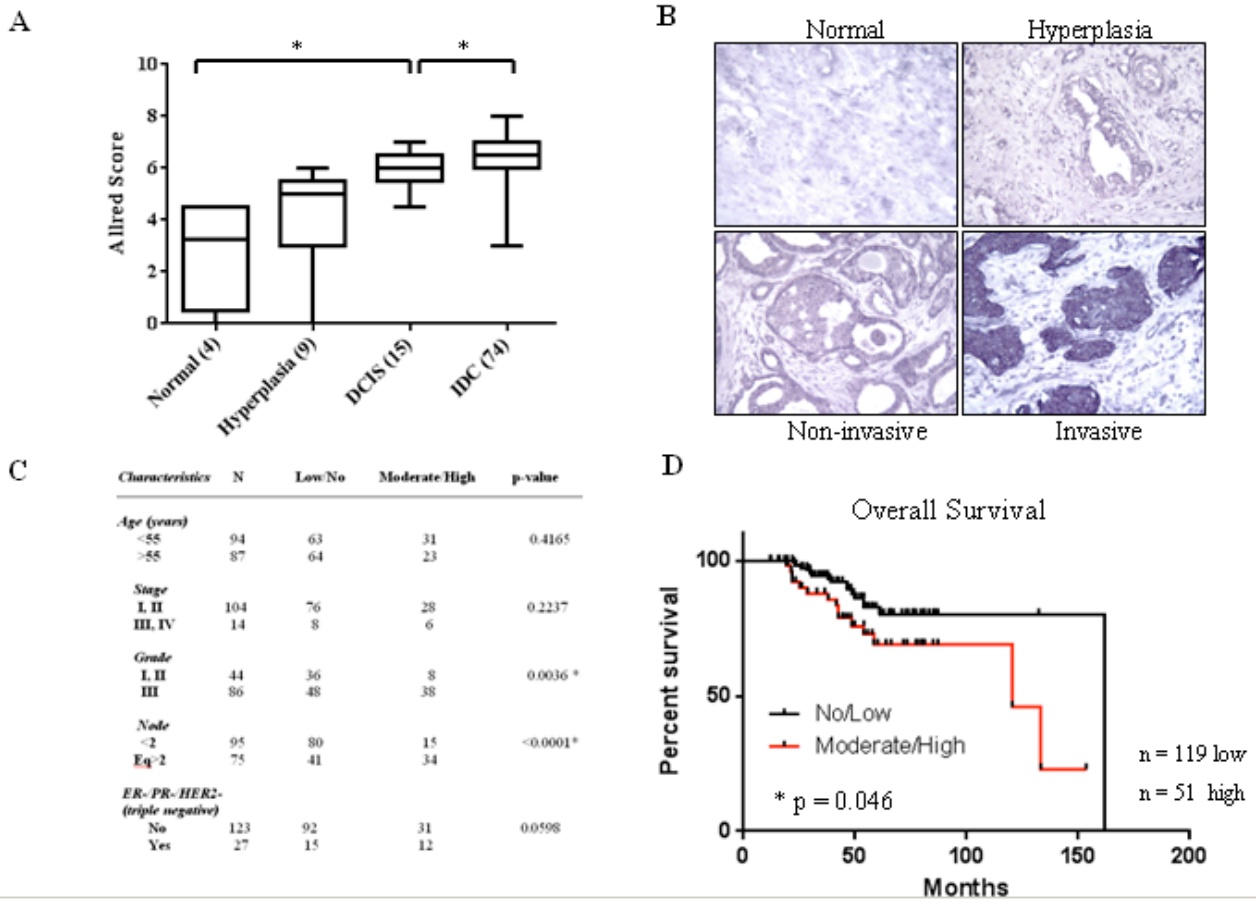
Figure 24. Elevated levels of lncRNA-HIT associates with invasive breast cancer in human, late grade tumor, positive lymph node, and poorer overall patient survival. (A) Locked nucleic acid *in situ* hybridization (LNA-ISH). LNA-oligonucleotide of lncRNA-HIT was labeled with digoxigenin-ddUTP using the Dig-5'-end labeling kit and hybridized to a human breast cancer TMA. Signal was measured using Allred Scale. (B) Representative photomicrographs of sections of a normal, hyperplasia, non-invasive and invasive breast tumor tissues are shown. (C and D) Summary of clinical and pathological characteristics associated with elevated lncRNA-HIT expression. (D) Kaplan-meier curve was plotted to assess overall patient survival in IDC patients with high and low expression of lncRNA-HIT. Asterisks represent p<0.05.

and positive lymph node (p < 0.0001) (Fig 24C). It is interesting to note that although not significant, high levels of lncRNA-HIT closely associate with triple-negative breast cancer subtype as MD-AMB-231 cell line. Of importance, elevated lncRNA-HIT

associates with poorer overall patient survival (p = 0.046) (Fig 24D). This result strongly suggests that lncRNA-HIT may play a functional role in patient mortality by driving EMT-associated phenotypes and metastasis in human cancer.

Discussion

Accumulating studies have demonstrated that TGFβ pathway plays a critical role in breast cancer metastasis as the roles of several protein coding genes and miRNAs have been described in this process [166-169]. In this study, we report lncRNA expression signature of TGFβ-induced EMT in mouse mammary gland epithelial (NMuMG) cells. Over 600 lncRNAs were significantly up-regulated or down-regulated during EMT, respectively. Further, we showed that lncRNA-HIT, one of top up-regulated lncRNA, plays an important role in TGFβ-induced EMT, cell migration, and invasion. Depletion of lncRNA-HIT can reverse the process of EMT associated gene expression of E-cadherin and vimentin. Furthermore, lncRNA-HIT expression is significantly elevated in the highly metastatic cell lines 4T1 and MDA-MB-231. Depletion of lncRNA-HIT results in significant reduction of both migration and invasion as well as tumor growth and lung metastasis. LncRNA-HIT is well conserved in sequence and genomic location from mouse to human. Human lncRNA-HIT expression is associated with more invasive tumor and breast cancer progression. These findings are important for several reasons. First, this is the first study to provide lncRNA expression signature of TGFβ-induced EMT in mouse mammary gland epithelial cells. Second, this study established a critical role of a previously uncharacterized lncRNA,

lncRNA-HIT, in TGFβ-induced EMT. Finally, conserved lncRNA-HIT was shown to be an important lncRNA in mouse and human breast cancer progression.

LncRNA-HIT resides in the 5 prime distal HOXA gene cluster in between but not overlapping with HOXA13 and HOXA11-AS. This region of the genome is highly conserved across several species and several functional lncRNAs have demonstrated to be important within the HOXA clusters, including neighboring HOTTIP and HOXA11-AS [149, 170, 171]. HOTTIP has been shown to positively regulate the protein coding genes in close proximity, notable HOXA13, through the recruitment of WDR5, a component of the histone methyltransferase protein MLL complex that promotes transcriptional activation [149]. It has also recently been reported that elevated levels of human HOTTIP and HOXA13 expression associates with disease progression and poorer survival in hepatocellular carcinoma. [165]. However, lncRNAs HOTTIP and HOXA11-AS were not induced by TGFβ in our microarray analysis, suggesting lncRNA-HIT has a unique role in TGFβ-induced EMT.

While microarray analysis revealed expression changes of a number of typical TGFβ-regulated protein-coding genes upon TGFβ-induced EMT, the array didn't show altered expression of lncRNA-HIT neighbor genes including HOXA13 and HOXA11. Further investigation is warranted to determine the exact mechanism and target genes of lncRNA-HIT, as well as its use as a therapeutic target for breast cancer metastasis intervention

Methods

Plasmids and siRNAs. A 1,178nt insert of lncRNA-HIT was amplified and TA-cloned into pGEM-T-easy vector from Promega (Madison, WI, USA) using NMuMG genomic DNA template. Plasmid containing insert was sequenced, insert was then isolated by EcoRI restriction digest, and cloned into pcDNA3.1 (pcDNA-HIT). SiRNAs were designed using IDT siRNA Design Software and 2 best siRNAs against lncRNA-HIT (si-HIT-1 and si-HIT-2) were determined empirically and used for experimentation.

Affymetrix Gene Expression Array. GeneChip Human Genome U133 Plus 2.0 Array was purchased from Affymetrix (Santa Clara, CA, USA). After 48 hours transfection of pcDNA-HIT in NMuMG, total RNA was isolation from samples and hybridized to array according to manufacturer's protocol. Hybridization and analysis was performed in MicroArray Core at H. Lee Moffitt Cancer Center.

Invasion and migration assays. NMuMG and 4T1 cell lines were transfected with control siRNA (si-CTL), si-HIT-1 or -2. After transfection for 48 hours, NMuMG cells were treated with TGFβ (5ng/ml) or vehicle control for 24 hours and then seeded into the upper chamber of Boyden Chambers coated without (migration) and with (invasion) Matrigel. Top chambers contained serum-free media while lower chambers had media containing 10% fetal bovine serum. After 16 hours, invasion and migration were evaluated and quantified by taking mean and standard deviations of 4 non-biased image fields.

Cell Lines, treatment, and tumor specimens. Cancer cell lines were obtained from ATCC (Manassas, VA, USA) and grown according to recommended culture conditions. Primary NMuMG and HMEC cell lines were maintained in Dulbecco's modified Eagle medium plus 10% fetal bovine serum and MEGM medium from Lonza Inc. (Hopkinton, MA, USA) plus manufacturer's supplements, respectively. Cells were treated with TGFβ at a concentration of 5 ng/ml for times indicated. Cell transfections were performed with Lipofectamine 2000 (Life Technologies) according to manufacturer's protocol. Frozen primary normal and tumor breast tissues were procured anonymously from patients who underwent surgery at H. Lee Moffitt Cancer Center, and approved by the institutional review board and in accord with an assurance filed with and approved by the US Department of health and human services. Each tumor contained at least 80% tumor cell as confirmed by microscopic examination. Tissues were snap-frozen to avoid RNA degradation and RNA was isolated with Trizol. Breast cancer tissue microarray (TMA) was obtained from Moffitt Cancer Center Tissue Core and approved under institutional review board.

Immunofluorescence, immunoblotting, and antibodies. The cells were grown to 60% confluence on cover slips and transfected with si-CTL, si-HIT-1 or -2 for 48 hours. Subsequently, TGFβ (5 ng/ml) was added to siRNA transfected cells for times indicated in Figure Legends. Briefly, cells were washed with PBS, fixed with 10% formalin containing methanol, and permeabilized with 1% NP-40 in PBS. Cells were blocked in 10% normal goat serum for 1 hour, and 1:200 dilution of primary antibodies

were incubated at 4°C overnight. Cover slips were washed in PBS 3 times and then appropriate secondary antibodies were added at 1:500 dilution. Cover slips were washed in TBS 3 times, counterstained with DAPI, and fixed for visualization.

Western blot was performed as previously described [172]. The band intensity of Western blots was quantified using ImageJ Software and represented as ratio of target gene/β-actin. Antibody against E-cadherin was purchased from BD Transduction Labs (San Jose, CA, USA), Lamin A/C from Santa Cruz Biotechnology (Santa Cruz, CA, USA), and antibodies for Vimentin and GAPDH from Cell Signaling Technologies (Danvers, MA). Alexa Fluor secondary anti-mouse and anti-rabbit antibodies for immunofluorescence were purchased from Life Technologies and HRP linked antibody for Western blot analysis was purchased from Sigma (St. Louis, MO, USA).

Soft agar colony formation assay. 4T1 cells were transfected with si-CTL, si-HIT-1 or -2 for 48 hours then washed, trypsinized, and resuspended in RPMI. A bottom layer of 0.6% agar in RPMI was poured, and top layer of 0.3% agar in RPMI was placed on top containing 1000 cells per 12-well and allowed to grow for 2 weeks. Four replicates were plated for each condition and quantitation is represented as mean and standard deviations of colonies counted in 4 non-biased fields.

Luciferase reporter assay. E-cadherin promoter luciferase plasmid was co-transfected with vector or pcDNA-HIT into NMuMG for 48 hours. Cells were then washed in PBS, lysed, and then firefly substrate was added to detect expression of

luciferase activity. The reaction was stopped using renilla substrate as a control and performed using Promega Dual Luciferase Assay Kit (Madison, WI USA).

Orthotopic mouse breast cancer model. Stable luciferase expressing MD-AMB-231 cell line was transfected with si-CTL or si-AS2-1 for 48 hours. Cells were washed in PBS, trypsinzed, centrifuged, and resuspended in PBS at a concentration of 4×10^6 cells per 100ul. Six-week old NOD Scid gamma (NSG) mice were purchased from Jackson Laboratories (Bar Harbor, ME, USA) and each mouse received 100ul injection into lower mammary fat pad. Mice tumor volume and luminescence signal were documented by injecting D-Luciferin K+ Salt Bioluminescent Substrate and imaged using Xenogen IVIS Imaging System 200 (Waltham, MA, USA).

Locked nucleic acid *in situ* hybridization of formalin-fixed, paraffin-embedded tissue microarray (LNA-TMA-ISH). LncRNA-HIT locked nucleic acid (LNA) probe was prepared by 5 prime end labeling with digoxigenin-ddUTP terminal transferase using the DIG 5 Prime End Labeling Kit from Roche (Indianapolis, IN, USA). Following deparaffinization and proteinase K digestion, breast tumor TMAs were prehybridized for 1 h and then hybridized with 10 nmol/L LNA lncRNA-HIT probe in a hybridization buffer (Roche) for 12 h. After three consecutive washes in 4× SSC/50% formamide, 2× SSC, and 0.1× SSC, sections were treated with a blocking buffer (Roche) for 1 h and incubated with anti-DIG-AP Fab fragments (Roche) for 12 h. Following wash for three times in 1× maleic acid and 0.3% Tween 20 buffer, reactions were detected in a detection solution [100 mmol/L Tris-HCl (pH 9.5) and 100 mmol/L

NaCl] in the presence of nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate

from Promega (Madison, WI, USA) and then visualized under a microscope.

CHAPTER 4

WDFY3-AS2 promotes breast cancer EMT, metastasis, and associates with poorer patient survival and triple-negative breast cancer

Introduction

In the array, we also found mouse lncRNA, BB049, to be up-regulated in response to TGFβ treatment. This transcript shares sequence and maps to human ortholog, WDFY3-AS2. Interestingly, WDFY3-AS2 expression is also induced by TGFβ in the primary human mammary epithelial cell line HMEC. WDFY3-AS2 expression is elevated in human breast tumor and metastasis. Clinical and pathological analysis revealed that high WDFY3-AS2 associates with poorer patient survival, high grade tumor, positive lymph node metastasis, and triple-negative breast cancer subtype. Knockdown and over-expression experiments demonstrate that WDFY3-AS2 can regulate TGFβ induced EMT, migration, invasion, and proliferation. In vivo, loss of WDFY3-AS2 resulted in slowed primary tumor growth and decreased lung metastasis in mouse orthotopic xenograft model. Lastly, we have shown that WDFY3-AS2 functions in "*cis*" to regulate WDFY3 and in "*trans*" to regulate STAT3 through protein binding partner hnRNP-R.

The WDFY3 gene product, also known as ALFY, is important in selective macroautophagy, which can promote clearing of cargo by stimuli other than starvation. WDFY3 can interact with the core autophagy promoting machinery, like p62, ATG5, ATG12, and LC3 [173]. Interestingly, TGFβ has been shown to selectively activate autophagy in human cancers, and in triple-negative breast cancers this can be an oncogenic force [174, 175]. STAT3 is a well established oncogenic transcription factor that is activated or elevated expression in many tumor types [176]. Hyperactive signaling through Janus kinase receptor (JAK) or over-expression of total STAT3 promotes G1/S transition to drive cell cycle progression and cell proliferation [177]. Notably, in triple-negative breast cancers STAT3 can promote autophagic-addiction and selective sensitivity to drugs targeting autophagy bafilomycin and chloroquine [178]. Given that WDFY3-AS2 regulates both WDFY3 and STAT3 through TGFβ pathway, this may provide a novel mechanism for cooperation between pathways in breast cancers. This chapter will discuss role of WDFY3-AS2 in further detail.

Results

BB049 and human ortholog WDFY3-AS2 are regulated by TGFβ in mammary epithelia. Mouse lncRNA BB049, a 659 base transcript on mouse chromosome 5, was another transcript which we predicted to have a human ortholog. BLAST analysis showed that BB049 is not only conserved in human (78% identity, 26% coverage), but across several species with >75% identity (Figure 25A). Importantly, BB049 maps to annotated human lncRNA WDFY3-AS2, providing independent

validation that our strategy for discovering human orthologs is effective (Figure 25B). Given both BB049 and WDFY3-AS2 are already in NCBI ref seq database, we designed primers for RT-qPCR based on these sequences. We confirmed that TGFβ treatment significantly induces both mouse and human transcripts greater than 2-fold in NMuMG and HMEC, respectively. (Fig 25C-D).
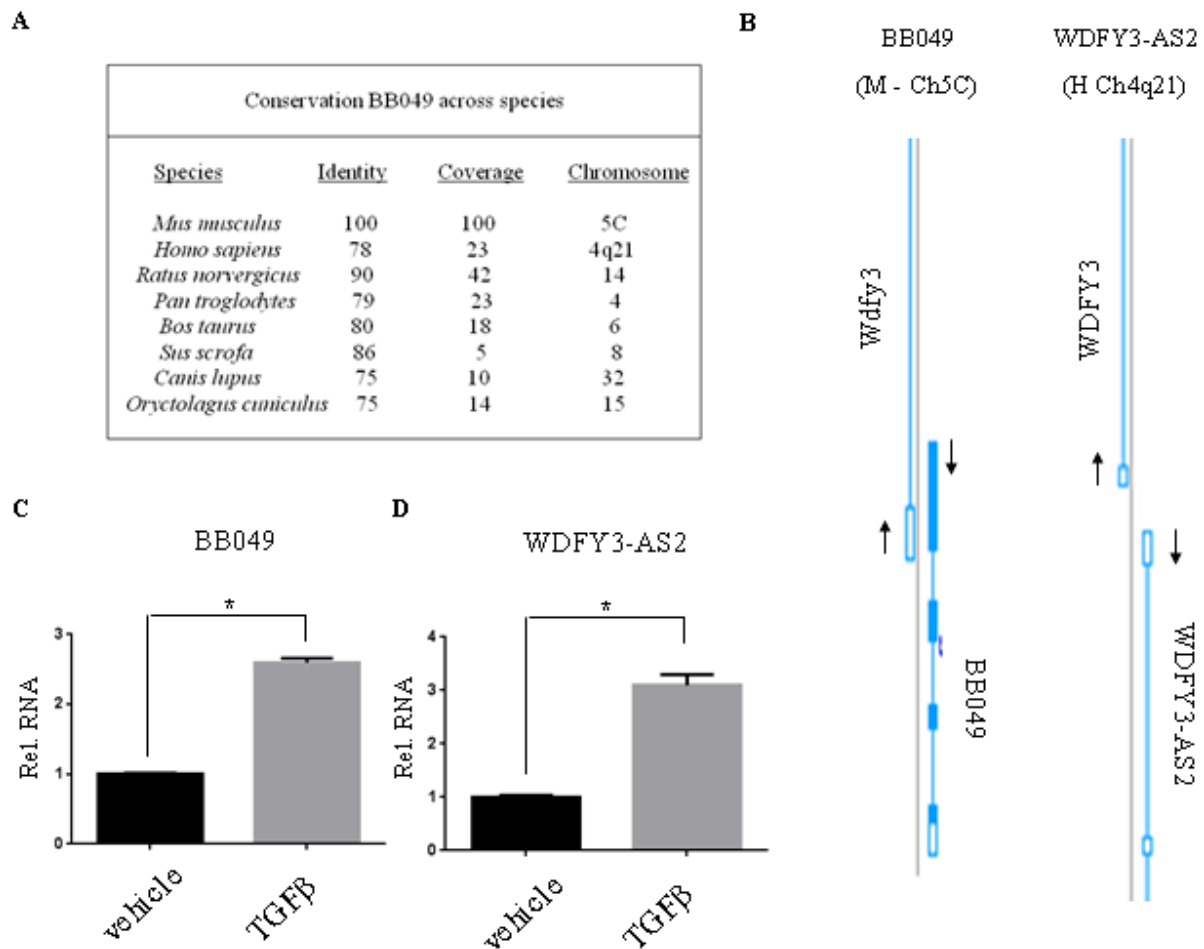
Figure 25. BB049 and human ortholog WDFY3-AS2 are regulated by TFGβ in mammary epithelia. (A) Using BLAST analysis, BB049 is well conserved across several species. (B) BB049 maps to human ortholog lncRNA at chromosome 4q21. (C) NMuMG and (D) HMEC were treated with 5ng/ml TGFβ for 24 h and RT-qPCR was performed to check expression of BB049 and WDFY3-AS2. Asterisks represents p<0.05.

WDFY3-AS2 is elevated in human cancer and associates with lymph node metastasis, high grade tumor, triple-negative subtype, and poor patient survival. Since WDFY3-AS2 is elevated upon TGFβ-induced EMT in HMEC, we evaluated its expression level in human breast cancer. We obtained fresh frozen patient RNA from 16 normal breast epithelia and 32 primary breast tumor samples from H. Lee Moffitt Cancer Center Tissue Core. Semi-qPCR analysis showed that expression of WDFY3-AS2 is elevated in breast tumors as compared to normal controls, suggesting that this transcript may be functionally relevant in patient (Figure 26A). We next performed TMA-ISH that was blindly scored by a pathologist based on 2 criteria, percentage of positive cells and staining intensity. The breast TMA is composed of cores of normal epithelia, DCIS, IDC, and metastasis. We observed significant elevated expression of WDFY3-AS2 in DCIS, IDC and metastatic lymph node as compared to normal (Figure 26B). High levels of WDFY3-AS2 were associated with high grade tumor, positive lymph node metastasis, and triple-negative subtype (Figure 26C). Furthermore, high expression of WDFY3-AS2 correlated with overall poorer patient survival (Figure 26D). Collectively, these data suggest that WDFY3-AS2 is involved in breast cancer initiation and progression.

WDFY3-AS2 regulates TGFβ-induced EMT in HMEC. HMEC cells treated with TGFβ will become more spindle-shaped and elongated, and protein expression levels of E-cadherin will decrease, while Vimentin protein expression will increase. Since WDFY3-AS2 is induced in response to TGFβ, we hypothesized that WDFY3-AS2 may functionally contribute to TGFβ-induced EMT. To test this, we first designed 2 siRNAs
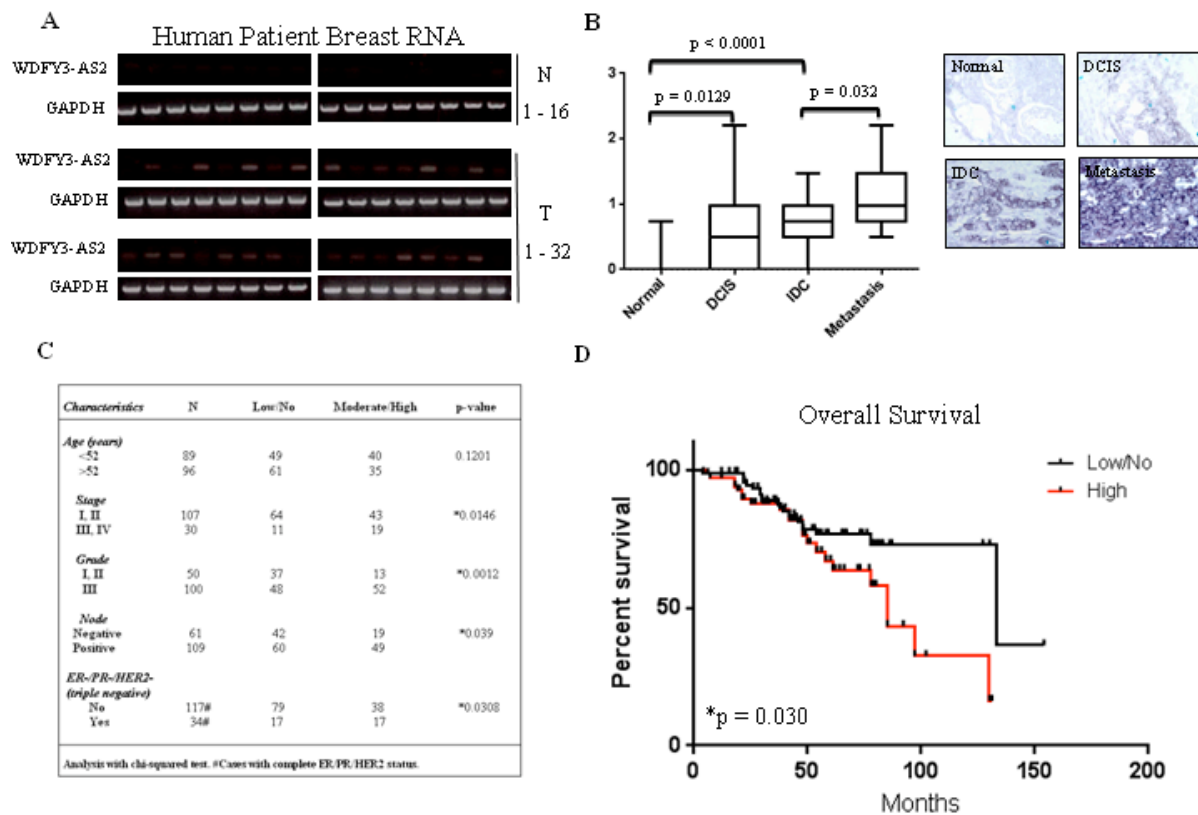
Figure 26. WDFY3-AS2 is over-expressed in breast cancer and is associated with poor patient survival and triple-negative breast cancer. (A) Semi-RT-qPRC analysis of WDFY3-AS2 expression in 16 normal breast epithelia and 32 primary human breast tumors. (B) TMA-ISH was performed on breast cancer TMA to assess expression of WDFY3-AS2 at various stages of breast lesions. (C) Relation of expression of WDFY-AS2 with clinical parameters. (D) Overall patient survival was analyzed and represented as Kaplan-Meier Curve.

against WDFY3-AS2 (si-AS2-1 and si-AS2-2) and validated knockdown in MD-AMB-231 cell line which expresses high level of WDFY3-AS2. We next transfected HMEC cell using control siRNA (si-CTL) and best scoring si-AS2-1 for 48 h, and then treated cells with TGFβ to induce EMT for 24 h. Si-AS2-1 was able to almost completely eliminate the induction of WDFY3-AS2 in response to TGFβ treatment (Figure 27A). Compared to untreated control (UNT), TGFβ treated samples showed significant loss of E-cadherin

and gain of Vimentin protein expression in si-CTL treated cells. However, both of these responses were partially abolished in cells depleted WDFY3-AS2 by si-AS2-1 (Figure 27B). This result was also reaffirmed using immunofluorescence staining to visualize E-cadherin and Vimentin expression on the cell surface (Figure 27C).
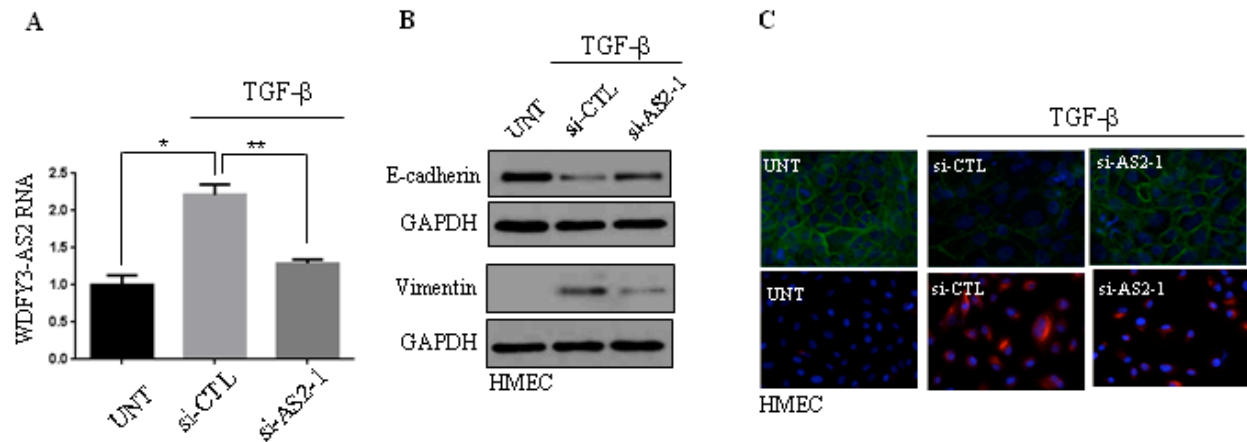
Figure 27. Knockdown of WDFY3-AS2 inhibits TGFβ induced EMT. (A) HMEC was left untreated (UNT) or transfected with control siRNA (si-CTL) or siRNA against WDFY3-AS2 (si-AS2-1) for 48 h, and then treated with TGFβ for 24 h. Total RNA was isolated and examined for WDFY3-AS2 expression by RT-qPCR. Similarly treated cells were assessed for characteristic EMT markers, E-cadherin and Vimentin, by Western blot (B) and immunofluorescence staining (C).

WDFY3-AS2 regulates migration, invasion, proliferation, and cell cycle. TGFβ treatment increases cell migration and invasion in HMEC cell line similar to the NMuMG cell line. To test WDFY3-AS2's contributed to this, we pre-treated HMEC cells for 24 hours with TGFβ (5 ng/ml) and then transfected si-CTL, si-AS-1, or si-AS2-2 for 48 hours while still in the presence of TGFβ. After 72 h of this treatment, we noticed a

significant loss of WDFY3-AS2 expression using both siRNAs (Figure 28A). We then

assayed for cell migration and invasion using Boyden chambers. In both cases, TGFβ-

induced migration and invasion were significantly reduced upon depletion of WDFY3-

AS2 in HMEC compared to si-CTL control (Figure 28B and 28C).
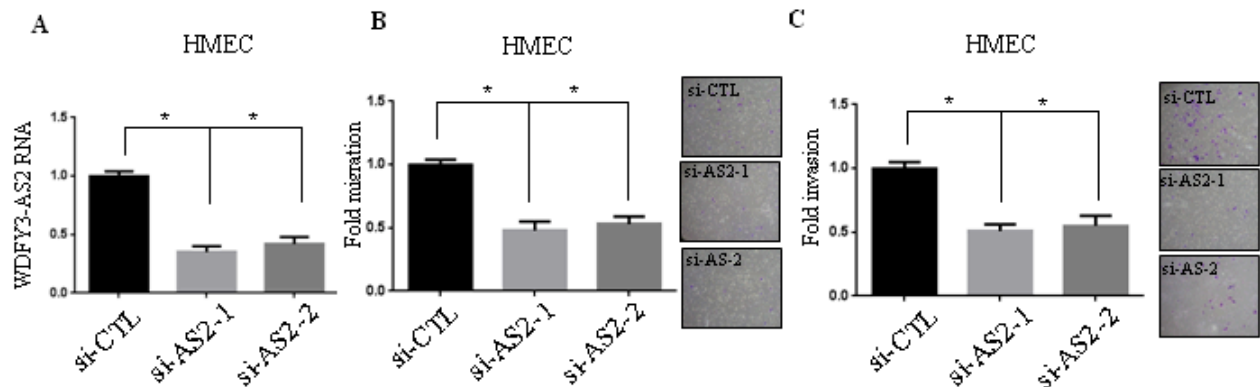
Figure 28. Knockdown of WDFY3-AS2 inhibits TGFβ-induced migration and invasion. (A) HMEC cells were pre-treated for 24 hours with TGFβ (5ng/ml) to induced migration and invasion, and then transfected for 48 hours with si-CTL or siRNAs against WDFY3-AS2. At 72 hours RNA was isolated and examined for WDFY3-AS2 expression. Once knockdown was confirmed, cells were subjected Boyden 2-Chamber Assay without matrigel (B) and with matrigel (C) to assess cell migration and invasion, respectively.

We performed similar assays using highly metastatic MD-AMB-231 human cell

line. In these experiments, we knocked down WDFY3-AS2 for 48 hours, then placed

cells in top of Boyden chambers in serum free medium, and initiated cell motility using

10% serum containing media as chemo-attractant in bottom chamber. Again, after 16

hours we observed that cell migration and invasion were significantly reduced following

knockdown of WDFY3-AS2 (Figure 29A-B). MD-AMB-231 cell line was also assayed

for changes in proliferation after WDFY3-AS2 knockdown by EdU incorporation. Cells

were seeded in chamber slides at 70% confluence and transfected with si-CTL, si-AS2-

1, and si-AS2-2 for 48 hours. The mean number of positive proliferating cells was calculated by quantifying 4 non-biased fields. Cells treated siRNAs against WDFY3-AS2 showed significant reduction of percentage of cells with EdU incorporation (Figure 29C). To further investigate contribution of WDFY3-AS2 to cell cycle progression in MD-AMB-231, after 48 hours knockdown we fixed cells in ethanol and stained nuclear DNA with propidium iodide (PI). In cells depleted WDY3-AS2, we observed a significant accumulation of cells in G1-phase, and a decrease in S-phase and G2-phase populations (Figure 29D). These data suggest that WDFY3-AS2 regulates cell proliferation, DNA synthesis and G1-S transition.

Next we wanted to determine if over-expression of ectopic WDFY3-AS2 could also have an effect on invasion, proliferation, and cell cycle. We cloned a 759nt cDNA into pcDNA3.1 expression vector (pcDNA-AS2). Ectopic expression of pcDNA-AS2 in low metastatic MCF7 cells caused an increase in invasion and EdU incorporation (Figure 30A-B). We also performed PI staining and cell cycle analysis on MCF7 cells over-expressing WDFY3-AS2 for 48 hours. Ectopic expression resulted in an increase of both S-phase and G2-phase cells, and a decrease in G1-population (Figure 30C).

Knockdown of WDFY3-AS2 inhibits primary tumor growth and metastasis *in vivo*. We next investigated whether WDFY3-AS2 knockdown resulted in reduced motility and proliferation *in vivo*. To test this, we employed an orthotopic mouse xenograft model by injecting 4×10^6 MD-AMB-231-Luciferase cells which had been treated with si-CTL and si-AS2-1, into the lower mammary fat pad. Using the NSG mice,

primary tumors were detected and established by the end of week 2. The size of the

primary tumor and the presence of lung metastasis were measured each week via
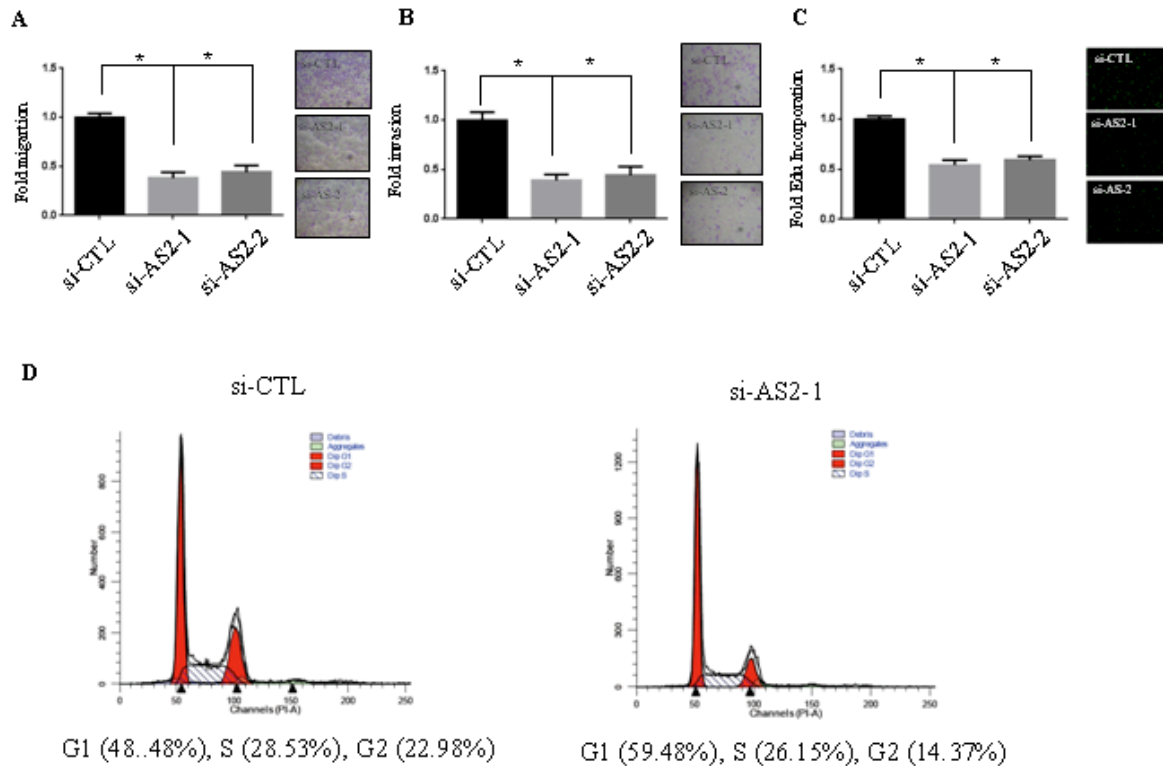
Figure 29. Knockdown of WDFY3-AS2 in MD-AMB-231 reduces migration, invasion, proliferation, and G1 accumulation. (A and B) WDFY3-AS2 was knocked down using 2 siRNAs in the highly metastatic cell MD-AMB-231 for 48 hours. Migration was subsequently assessed by seeding cells in Boyden 2 Chamber without (A) and with (B) matrigel. (C) Cell proliferation was assessed by incubation of cell with thymidine nucleoside analogue 5-ethynyl-2′-deoxyuridine (EdU) for one hour. (D) Cell cycle analysis was performed using propidium iodine (PI) staining of nuclei and followed flow cytometry. Asterisks denotes $p < 0.05$.

standard caliper measurement and bioluminescent imaging. The growth and size of the

primary tumor were significantly less in the mice injected with cells depleted WDFY3-

AS2 at week 5 and throughout endpoint at week 7 (Figure 31A-C). WDFY3-AS2

knockdown group also had significantly less tumor weight compared to control (Figure

31D). At week 7 endpoint, both groups of mice were injected with luciferase substrate and lungs were assessed for photon flux signal. There was a dramatic and significantly
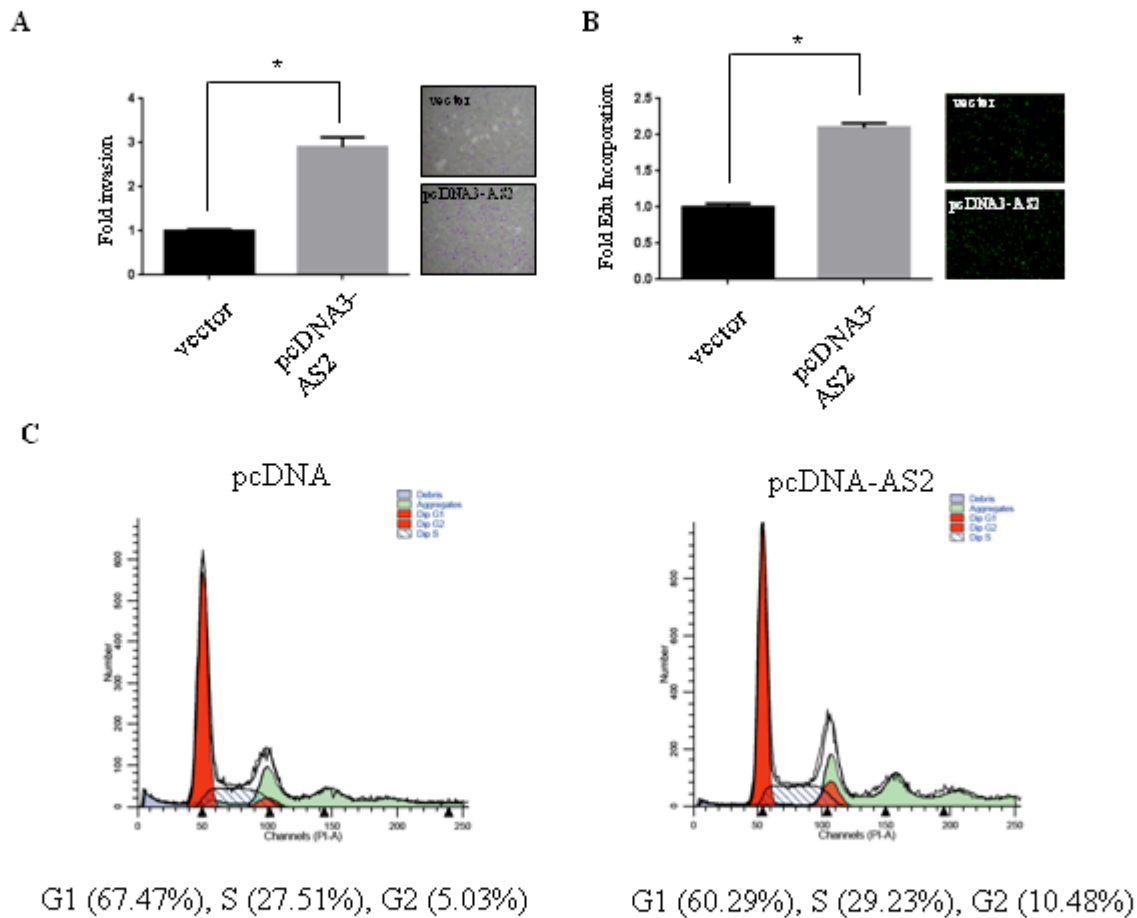
Figure 30. Ectopic expression of WDFY3-AS2 increases invasion, proliferation, and G2-S population. (A and B) pcDNA-AS2 was transfected into low metastatic MCF7 cells for 48 hours. Invasion (A) and proliferation (B) were assayed using Boyden chamber and EdU incorporation. (C) Cells were stained with PI to examine change in cell cycle. Asterisks represents $p < 0.05$.

more robust signal of luciferase in lungs of control mice compared to mice injected with the cells depleted WDFY3-AS2 (Figure 31E). Lungs from each group were paraffin embedded, sectioned, and H&E stained. The number of lung metastasis nodules for each group was counted across 3 different sections per lung and averaged. There was

a significant decrease in number of lung nodules present in WDFY3-AS2-knockdown mice as compared to control mice (Figure 31F).

WDFY3-AS2 regulates protein coding gene WDFY3 and STAT3 via 'cis' 'trans' mechanism. There are several examples of antisense lncRNAs regulating closest neighboring protein coding gene in "cis" and regulating distant genes in "trans" [56, 80, 149, 179]. In order to determine which genes are possible targets of WDFY3-AS2, we performed Affymetrix Chip Gene Expression analysis on si-CTL or si-AS2-1 transfected MD-AMB-231 cells (Figure 32A). We observed 2 genes, WDFY3 and STAT3 to be significantly down-regulated in response to WDFY3-AS2 knockdown. To validate this finding, we designed real-time PCR primers for WDFY3 and STAT3 and evaluated RNA expression for both after WDFY3-AS2 knockdown compared to control. Consistent with the array data, we observed a significant decrease in both WDFY3 and STAT3 mRNA levels after depletion of WDFY3-AS2 by semi-qPCR (Figure 32B). Furthermore, we also observed a loss in total WDFY3, phospho-STAT3 and total STAT3 at 72 hours knockdown of WDFY3-AS2 in MD-AMB-231 cells (Figure 32C).

Interestingly, if we first deplete cells of endogenous WDFY3-AS2 by siRNA treatment, ectopic expression of pcDNA-WDFY3-AS2 construct was able to completely rescue mRNA level of WDFY3 and partially rescue STAT3 (Figure 32D and 32E). In MCF7 cells, ectopic expression of pcDNA-WDFY3-AS2 increased phosph-STAT3 and total STAT3 protein, however this effect was not seen with WDFY3 protein level (Figure 32F).

WDFY3-AS2 and WDFY3 are co-expressed and elevated WDFY3 is also a predictor of poorer patient survival. Since WDY3-AS2 positively regulates WDFY3, we next assessed the significance of WDFY3 protein expression in human breast tumors and then correlated expression of WDFY3 with WDY3-AS2.
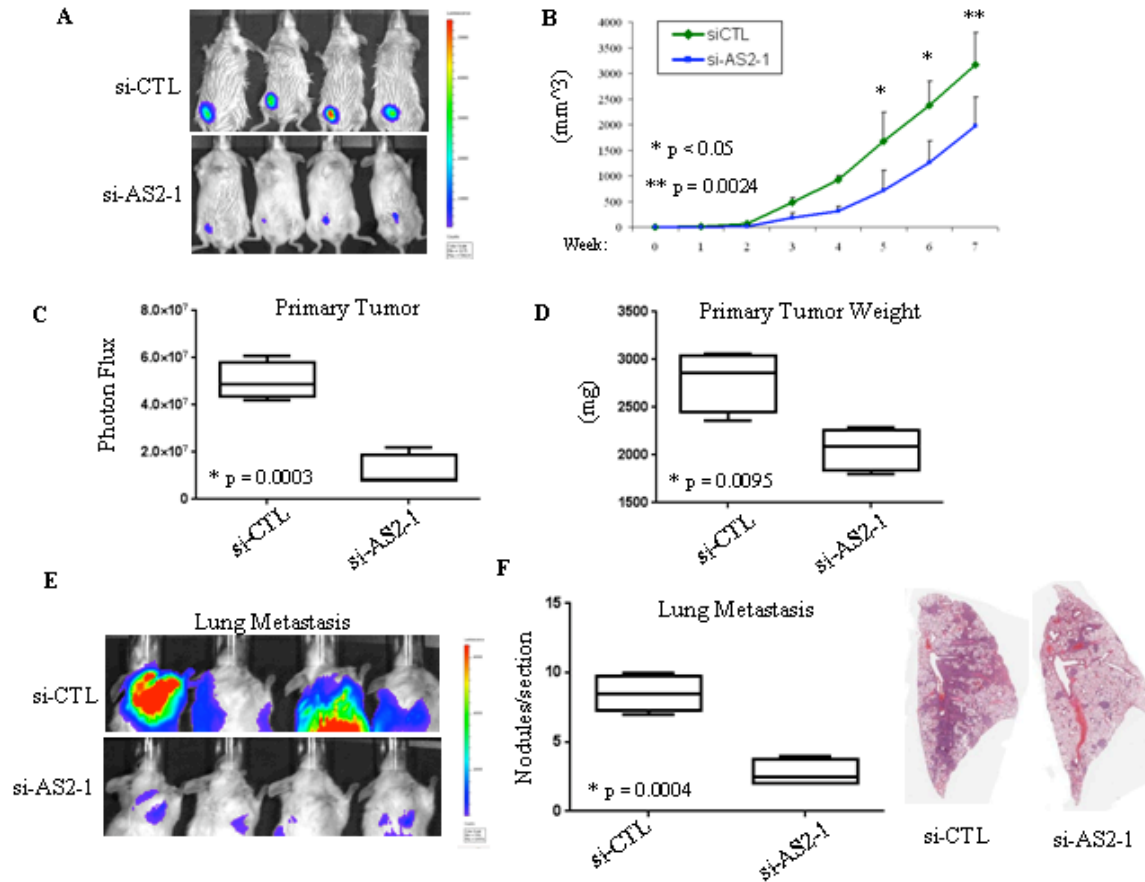
Figure 31. Knockdown of WDFY3-AS2 significantly inhibits tumor growth and metastasis. (**A** and **B**) Stable expressing luciferase MD-AMB-231 cells were transfected with si-CTL or si-AS2-1 for 48 hours. Cells were washed in PBS, trypsinized, resuspended in PBS, and 4 x 10^6 cells were injected into lower mammary fat pad per mouse. Primary tumor growth was tracked by bioluminescent imaging (A) and caliper measurement (B). (**C**) Quantification of primary tumor size using photon flux measurement at week 5. (**D**) Tumors were removed from each mice and weighed. Average mean and deviation are plotted. (**E**) At endpoint week 7, lungs metastasis were imaged using bioluminescent. (**F**) Lungs from mice were paraffin embedded, sectioned, H&E stained, and number of metastasis were counted and averaged.

Immunohistochemical staining the same breast cancer TMAs was performed with antibody against WDFY3. In a total of 135 specimens (normal, DCIS, and IDC), we observed 43 cases (31.8 %) of low/low expression (WDFY3-AS2/WDFY3), 40 cases (29.6 %) high/high expression, 22 cases (16.2 %) high/low expression, and 30 cases (22.2 %) low/high expression. Plotting data in a 2 x 2 contingency table statistical
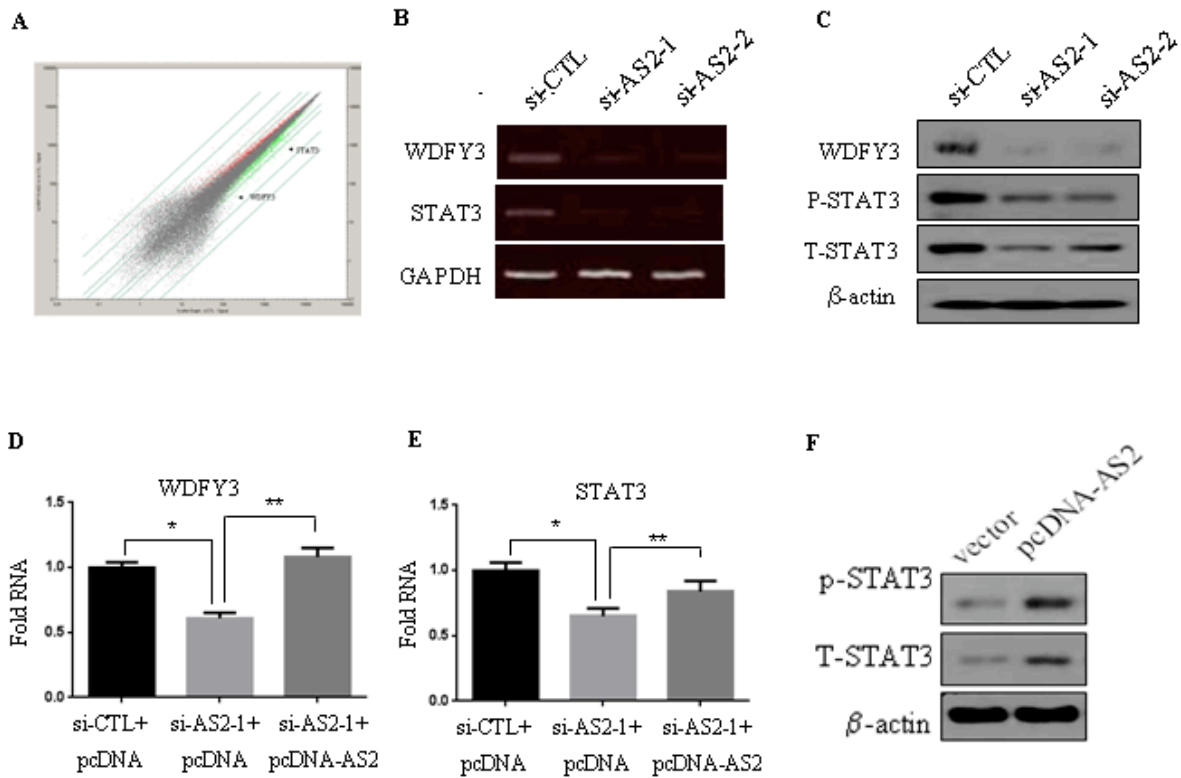
Figure 32. WDFY3-AS2 regulates protein coding gene WDFY3 and STAT3. (**A**) MD-AMB-231 cells were treated with si-CTL or si-AS2-1 for 48 hours and Affymetrix Gene Expression Analysis was performed. WDFY3 and STAT3 mRNA were shown to be significantly down-regulated in response to WFY3-AS2 knockdown. (**B**) Semi qRT-PCR analysis of WDFY3 and STAT3 expression. (**C**) Western blot protein analysis of WDFY3, phosho-STAT3, and total STAT3 protein in MD-AMB-231 cells treated with si-AS2-1, si-AS2-2 and si-CTL. (**D** and **E**) WDFY3 and STAT3 mRNAs were rescued with introduction of pcDNA-WDFY3-AS2 after knockdown of endogenous WDFY3-AS2. (**F**) Over-expression of pcDNA-WDFY3-AS2 construct induces phospho-STAT3 and total STAT3 protein levels.

analysis demonstrated a significantly positive correlation between expression WDFY3-AS2 (RNA) and WDFY3 (protein) in patients with breast cancer (Figure 33A-B).

Furthermore, we also observed co-expression of WDFY3-AS2 and WDFY3 (mRNA) in 15 established breast cancer cell lines as measured by RT-qPCR and plotted by linear regression and R-squared analysis (data not shown). Collectively, these data suggest that WDFY3-AS2 positively regulates WDFY3 expression *in vivo*. Since we have shown that WDFY3-AS2 is strongly associated with triple-negative subtype and high grade tumor (Figure 26), we used Kaplan Meier Plotter Database for independent validation and found that elevated WDFY3 is also a predictor of poorer overall survival. In patients stratified for ER-negative, basal subtype, or high grade breast cancers, elevated WDFY3 mRNA is a significant predictor of poorer patient survival in all 3 cases (Figure 33C-E). Interestingly, elevated WDFY3 mRNA is of no, or reverse, prognostic value in patients with ER-positive, luminal A subtype, or low grade tumor (data not shown).

Since WDFY3 and STAT3 are the targets of WDFY3-AS2, we next examined their expression in orthotopic mouse breast tumors derived from WDFY3-AS2 knockdown and si-CTL transfected MDA-MB-231 cells. Immunoblot analysis revealed decreased expression of both WDFY3 and STAT3 in WDFY3-AS2 depleted tumors (Figure 34A). Furthermore, immunohistochemical staining and *in situ* hybridization of paraffin embedded sections of xenograft tumor and lungs showed significantly reduction of STAT3 and WDFY3 expression in WDFY3-AS2 knockdown breast tumor and lung metastases. In addition, Ki-67 proliferation marker was also significantly reduced in si-AS2-1 treated group (Figure 34B).

WDFY3-AS2 binds to hnRNP-R to regulate WDFY3 and STAT3. In attempt to identify a protein binding partner for WDFY3-AS2, RNA immunoprecipitation (RIP) assay was performed. "Sense" and "Antisense" orientations of WDFY3-AS2 RNA were *in vitro* transcribed in the presence of biotin-UTP and mixed with 1 mg of MD-AMB-231 nuclear extract. Ribonucleoprotein (RNP) complexes were immunoprecipitated using agarose-streptavidin. The bound protein was run on an SDS page gel and silver stained for analysis (Figure 35A). We observed a unique protein band in the "sense" lane just below the 72-kDa protein marker. This band was extracted and submitted for protein identification via mass spectrometry. Analysis revealed a significant enrichment of hnRNP-R protein as compared to "antisense" gel band extracted at same molecular weight. To confirm this interaction, we transfected hnRNP-R-FLAG construct into HEK293T cells for 72 hours and performed RIP assay using biotinylated "sense" and "antisense" RNA probes. After immunoprecipitation using streptavidin-agarose, immunoblotting for FLAG showed a significant enrichment in sense binding to hnRNP-R versus antisense probe (Figure 35B). We performed similar experiments using 1 mg nuclear extract isolated from MDA-MB-231 cell line and immunoblotting against endogenous hnRNP-R (data not shown).

Next we performed the reverse experiment to assess binding of endogenous WDFY3-AS2 to hnRNP-R. We transfected either vector or hnRNPR-FLAG into MDA-MB-231 cells for 72 hours followed by immunoprecipitation using FLAG antibody. Total RNA was isolated from the immunoprecipitates. Semi-qPCR and real-time PCR was performed to examine endogenous WDFY3-AS2 binding to hnRNP-R. We observed
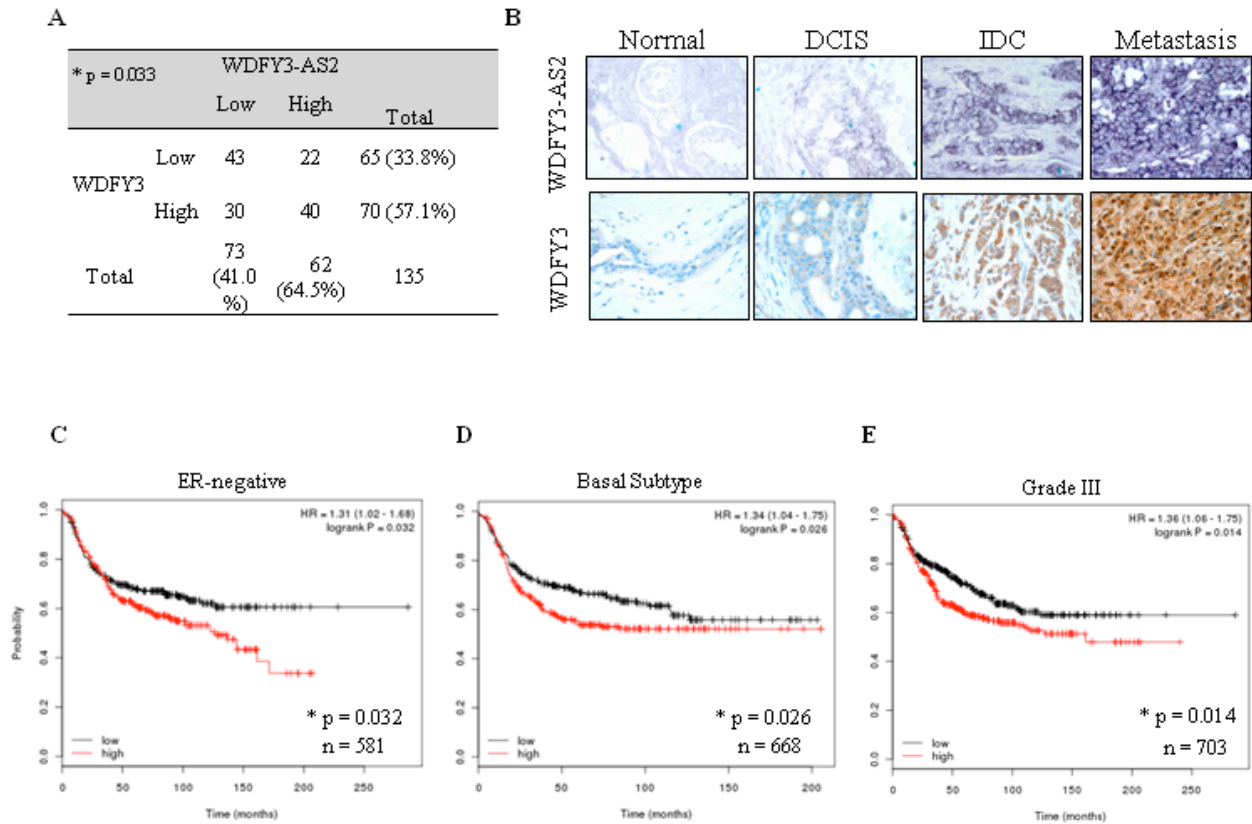
Figure 33. WDFY3-AS2 and WDFY3 are co-expressed in patient and elevated WDFY3 predicts poorer patient overall survival. (**A**) WDFY3-AS2 and WDFY3 were stained using LNA-probe and antibody, respectively. They are significantly co-expressed. (**B**) Representative tissue cores demonstrating increased co-expression of both WDFY3-AS2 and WDFY3. (**C** - **D**) In stratifying patients, WDFY3, like WDFY3-AS2, is also a predictor of poorer overall survival in ER-negative (C) basal subtype (D) and grade III tumors (E).

about an 8-fold increase in bound WDFY3-AS2 transfected with hnRNP-R-FLAG as compared to that transfected with vector (Figure 35C). Endogenous hnRNP-R in MDA-MB-231 and MCF7 cell lines was also evaluated for bound endogenous WDY3-AS2 by comparing immunoprecipitations of anti-hnRNPR and anti-GFP. In both cell lines we found greater than 2-fold increase in bound WDFY3-AS2 compared to GFP control (Figure 35D).
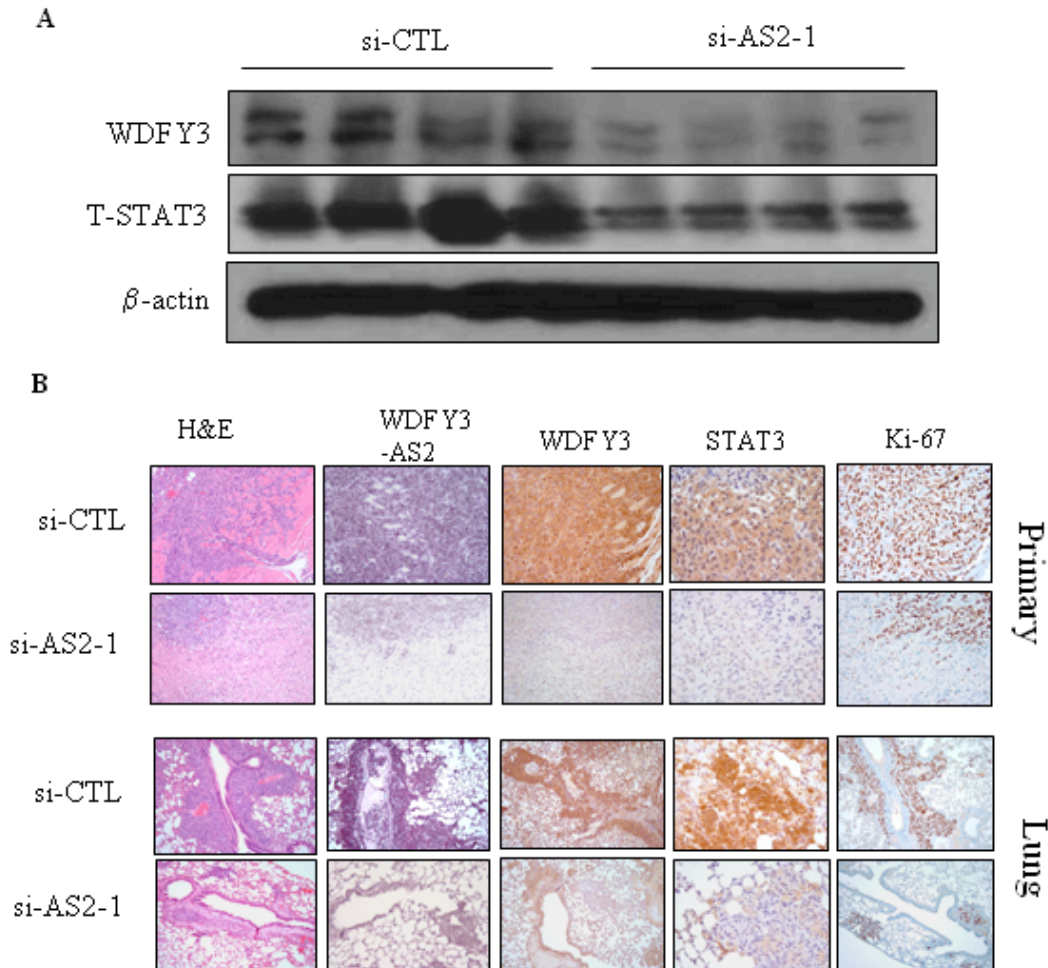
Figure 34. Reduced expression of WDFY3 and STAT3 in WDFY3-AS2 knockdown orthotopic mouse breast tumor and lung metastatic lesions. (**A**) Depletion of WDFY3-AS2 in MDA-MB-231 xenograft of primary tumor results in decreased expression of both WDFY3 and STAT3 protein. (**B**) Primary tumor and lungs were removed, paraffin embedded, and stained with H&E, LNA-WDFY3-AS2 RNA probe, and antibodies against WDFY3 and STAT3, as well as Ki-67.

Previously we have shown that knockdown of endogenous WDFY3-AS2 followed by introduction of ectopic WDFY3-AS2 can rescue both WDFY3 and STAT3 RNA levels. To assess whether hnRNP-R functionally contributes to this process, we performed the same rescue experiments and included cells deficient in hnRNP-R. We noticed that in

cells depletion of hnRNP-R, we were unable to restore WDFY3 and STAT3 mRNA (Figure 35E). This denoted that WDFY3-AS2 promotes WDFY3 and STAT3 expression in an hnRNP-R dependent manner.

Discussion

We identified a long non-coding mouse transcript BB179049 to be significantly up-regulated in response to TGFβ, which was mapped to syntenic loci and shared sequence with human ortholog WDFY3-AS2. We have shown that WDFY3-AS2 is also induced by TFGβ in human HMEC cells, and this regulation in mammary epithelia is conserved from mouse to human. We observed WDFY3-AS2 to be elevated in several breast cancer cell lines and primary breast tumors compared to normal controls. Furthermore, WDFY3-AS2 expression directly increases with progression of disease, and is most significantly overexpressed in metastatic lesions. WDFY3-AS2 also has clinical and pathological significance in that elevated expression of WDFY3-AS2 is a predictor of poorer overall survival, as well as associates with last stage and high grade tumors as well as triple-negative breast cancer subtype. Depletion of WDFY3-AS2 slowed the onset of TGFβ mediated EMT program and also inhibited migration, invasion, and proliferation. These results were recapitulated in a mouse orthotopic xenograft model in that mice deficient in WDFY3-AS2 showed decreased primary tumor growth and lung metastasis.

WDFY3-AS2 is a natural antisense transcript (NAT) located at the 5-prime region but not overlapping with protein coding gene WDFY3. Several examples of NATs are

emerging in literature and it is becoming more apparent that a large subset of protein coding genes may be regulated by these lncRNAs [180-183]. Here we have shown that knockdown of WDFY3-AS2 transcript results in loss of WDFY3 transcript and protein.
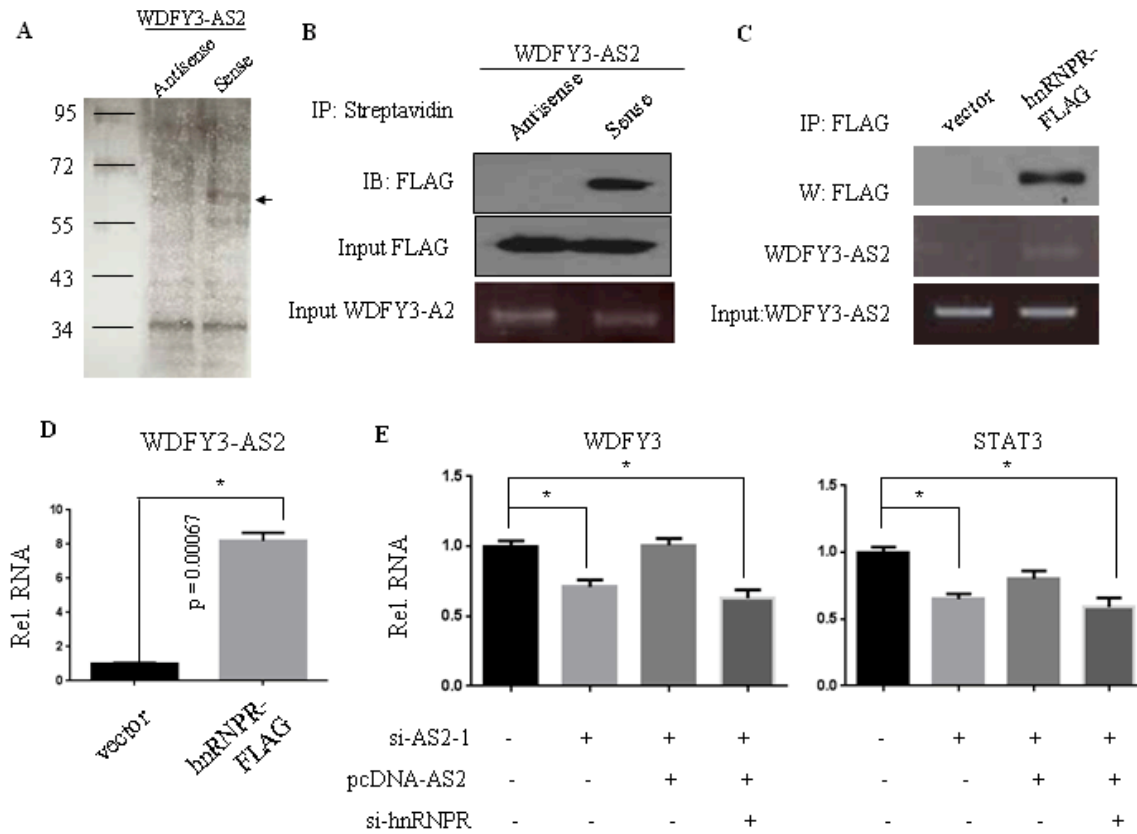
Figure 35. WDFY3-AS2 binds to hnRNP-R. (**A**) RNA-Immunoprecipitation (RIP) was performed using biotinylated "sense" and "antisense" probes and 1mg MDA-MB-231 nuclear extract. Total bound RNA-protein complexes was run on SDS-page gel and silver stained. (**B**) HEK293T cells were transfected with ectopic hnRNP-R-FLAG. After 48 h of incubation, RIP was performed. Subsequent Western blotting was done using anti-FLAG antibody. (**C** and **D**) MDA-MB-231 cells were transfected with hnRNP-R-FLAG. IP was performed after 72 h, total RNA was isolated from IPs and checked for endogenous WDFY3-AS2 expression by semi-qPCR (C) and real-time PCR (D). (**E**) Endogenous hnRNP-R was knocked down using siRNA and then co-transfected with pcDNA-AS2. WDFY3 and STAT3 mRNAs were assessed by RT-qPCR.

The WDFY3 gene encodes for a protein that enhances selective macroautophagy through interaction with p62 and the autophagy effector complex consisting of Atg5, Atg12, Atg16, and LC3 [173]. Another report has demonstrated that autophagy promotes cell invasion and EMT in hepatocellular carcinoma using TGFβ pathway as a key regulatory element [184]. Studies have also shown that potent inhibitors of autophagy can have therapeutic benefit in specifically in triple-negative breast cancer subtype including those driven by STAT3 [175, 178, 185]. For this reason, we found it interesting that elevated WDFY3 associates with poorer patient overall survival in these ER-negative and basal subtype, but not ER-positive or luminal subtype of breast cancer.

We have also identified STAT3 as a downstream target of WDFY3-AS2. STAT3 is a well-established oncogene across several human malignancies [186-188]. The JAK/STAT signaling pathway has also been shown to be critical for EMT, invasion, and metastasis [189, 190]. One report has demonstrated that there is therapeutic benefit using metformin to target STAT3 specifically in triple-negative breast cancer [191]. Our data suggest that WDFY3-AS2 is a novel route for TGFβ mediated activation of STAT3 to enhance both proliferation and EMT phenotypes in breast cancer.

We have identified hnRNP-R to be a protein binding partner of WDFY3-AS2, and this interaction functions to regulate WDFY3 expression. It has recently been reported that hnRNP-R can facilitate transcription through its interaction with scaffolding components Mediator, TBP, and TFIIH, which can then recruit RNA polymerase II [192]. Since WDFY3-AS2 lies in the antisense orientation and upstream to the transcriptional start site of WDFY3, this may provide a novel mechanism by which lncRNA can recruit

RNA polymerase II to the promoter to activate target gene transcription. However, further experimentation is required to solidify this claim.

Taken together, WDFY3-AS2 is a novel target for therapy in triple-negative breast cancer. Since we have shown that WDFY3-AS2 is critical for expression of both WDFY3 and STAT3, inhibition of WDFY3-AS2 could suppress associated phenotypes facilitated through the function of these two genes. WDFY3-AS2 can also serve as a novel biomarker for diagnostic and prognostic utility in the clinic. Future studies in this area of research are warranted.

Methods

Invasion and migration assays. Briefly, MD-AMB-231 cell were transfected with si-CTL, si-AS2-1, and si-AS2-2 for 48 hours. Cells were washed in PBS, trypsinized, and seeded in upper chambers without (migration) and with matrigel (invasion). After 16 hours top chambers were cleared with cotton, and bottom chamber was fixed in crystal violet plus 10% methanol. Average mean and standard deviations were calculated from 4 non-biased fields.

Cell Lines, treatment, and tumor specimens. Cells were obtained from ATCC and grown according to recommended conditions. MD-AMB-231 and MCF7 cell lines were maintained in RPMI plus 10% fetal bovine serum. Cell transfections were performed using Lipofectamine 2000. Frozen RNAs were procured anonymously from

patients at H. Lee Moffitt Cancer Center, and approved by the IRB. Tissues were snap-frozen to avoid degradation and isolated with Trizol.

Locked nucleic acid in situ hybridization of formalin-fixed, paraffin-embedded tissue microarray (LNA-TMA-ISH). WDFY3-AS2 LNA probe was prepared by end labeling with digoxigenin-ddUTP terminal transferase. Following deparaffinization and antigen retrieval, TMAs were hybridized with 10 nmol/L LNA WDFY3-AS2 probe in a hybridization buffer for 12 hours. After washing 4× SSC/50% formamide, 2× SSC, and 0.1× SSC, sections were blocked for 1 hour and incubated with anti-DIG-AP Fab fragments. Following washing in 1× maleic acid and 0.3% Tween 20 buffer, reactions were processed [100 mmol/L Tris-HCl (pH 9.5) and 100 mmol/L NaCl] in the presence of nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate and then visualized.

Mouse orthotopic breast cancer model. MD-AMB-231-Luciferase cell line was transfected with si-CTL or si-AS2-1 for 48 hours. Cells were washed, trypsinzed, and resuspended in PBS at a concentration of 4 x 10^6 cells per 100ul. Six-week old NSG mice received 100ul injection into lower fat pad. Tumor volume and luminescence signal were documented and imaged using Xenogen IVIS Imaging System 200 .

Affymetrix Gene Expression Array. As described in Chapter 3. Briefly, MDA-MB-231 cells were transfected with si-CTL or si-AS2-1, total RNA was isolated and

hybridized to the chip. Analysis was performed by H. Lee Moffitt Cancer Center Molecular Genomics Core.

Immunoblotting Immunofluorescence, and antibodies. Western blotting was performed using standard procedures. Antibodies against WDFY3 and total STAT3 was purchased from Santa Cruz Biotechnologies (Santa Cruz, CA, USA), and phosho-STAT3 was from Cell Signaling Technologies (Danvers, MA). For immunofluorescence staining, HMEC cells were seeded onto glass cover slips and then transfected with siRNAs for 48 hours. Media was then supplemented with TGFβ for times indicated in figure legend. Cover slips were then washed in PBS and then fixed in 10% cold methanol. Cover slips were then permeablized using 1% NP-40 and blocked in 10% normal goat serum. Primary antibody incubation was performed overnight at 4 degrees and secondary incubation was at room temperature for 1 hour. Cells were then washed in TBS, counterstained with DAPI, and fixed with mounting media for visualization.

Plasmids and siRNAs. A 758nt insert of WDFY3-AS2 was amplified and TA-cloned into pGEM-T-easy vector from Promega (Madison, WI, USA) using T47D cDNA library as template. Plasmid containing insert was sequenced, insert was then isolated by EcoRI restriction digest, and cloned into pcDNA3.1 (pcDNA-AS2). SiRNAs were designed using IDT siRNA Design Software. Two siRNAs against human WDFY3-AS2 (si-AS2-1 and si-AS2-2) were determined empirically and used for experimentation.

RNA immunoprecipitation (RIP) assay. WDFY3-AS2 sense and antisense DNA templates were generated using standard PCR and incorporating a T7 RNA polymerase II recognition site through the forward primer, respectively. PCR fragments were purified and 1ug template was used for *in vitro* transcription reaction using MAXIscript In Vitro Transcription Kit (Life Technologies) in the presence of biotin labeled UTP Roche (Branchburg, NJ, USA). Protocol used was previously described. [49] Briefly, 3ug *in vitro* transcribed RNA was incubated with 1mg nuclear extract, immunoprecipitated with 60ul streptavidin beads (Life Technologies), and run on a SDS-page gel. Gel was silver stained, and unique bands in sense lane were submitted for protein identification via mass spectrometry.

CHAPTER 5

LncRNA TIL involved in breast cancer EMT, metastasis, and cancer cell

stemness through regulation of NF90/C-MYC

Introduction

Another mouse lncRNA AK060 was also found to be significantly up-regulated by TGFβ in the microarray analysis and predicted to have a human ortholog. We were in fact able to detect a conserved lncRNA also regulated by TGFβ in human HEMC cells and was also elevated in human breast cancer. We have decided to refer to the human lncRNA as TIL, because this is TGFβ Induced LncRNA. TIL was of particular interest given its proximity to the C-MYC gene (e.g., residing ~10MB upstream of C-MYC) in 8q23-24 region. Recently, several reports have suggested that locally residing lncRNAs can regulate C-MYC expression by various mechanisms [61, 86, 87]. Therefore this led to the hypothesis that TIL may also regulate C-MYC expression.

Knockdown and over-expression studies demonstrated that TIL regulated C-MYC protein expression, however, had no effect on C-MYC mRNA levels. These results suggest TIL regulation of C-MYC at post-transcriptional level. We identified NF90 as a binding partner for TIL. NF90 is a dsRBP capable of regulating target gene expression through interaction with AU-rich elements (AREs) within the 3-prime UTR [193]. We

next found that C-MYC messenger contains 4 AREs and also physically associates with NF90, and this interaction likely stabilizes the mRNA. Furthermore, it has been reported that AKT phosphorylation of NF90 at Ser647 is critical for NF90 shuttling from nuclear to cytosol [194]. We have shown that TIL enhances phospho-NF90 level and promotes NF90 export from the nucleus. Taken together, we hypothesize that TIL enhances nuclear export of NF90 which increases C-MYC mRNA stability in cytosol and C-MYC protein product. This chapter aims to discuss these mechanisms in further detail.

Results

TGFβ-regulated AK060 and human ortholog TIL. AK060 is a mouse lncRNA predicted to have a human ortholog. In "BLAST-ing" mouse AK060 to the human genome, we observed significant sequence conservation at intergenic region on chromosome 8. AK060 is conserved across several species to varying degrees, and in human exhibits 88% identity across 44% coverage (Figure 36A).

This result was interesting to us given the close proximity to the C-MYC gene locus (~10MB; Figure 36B). There have been previous reports of lncRNAs residing within this genomic region that regulate C-MYC [61, 86, 195]. Next we validated regulation by TGFβ of AK060 and human ortholog in NMuMG and HMEC cell lines, respectively (Figure 36C and 36D). In human, this conserved transcript is induced greater than 3-fold in response to 24 hours TGFβ treatment versus vehicle control. We thus named the human transcript TIL (TGFβ Induced LncRNA).

Figure 36. AK060/TIL is a conserved TGFβ-upregulated lncRNA. (**A**) "BLAST" analysis of mouse AK060 transcript showed significant sequence identity and coverage at syntenic loci 8q23 in human. (**B**) Diagram shows that TIL is ~10 MB away from C-MYC. (**C**) NMuMG and (**D**) HMEC were treated with TGFβ (5ng/ml) for 24 hours and expression of AK060 and TIL were measured by RT-qPCR.

TIL is elevated in human breast cancer and associates with late stage tumor and triple-negative breast cancer. We next evaluated TIL expression in a panel of human breast cancer cell lines. In comparison to HMEC normal cell, several cancer lines show significant over-expression of TIL, notably HS578T, MDA-MB-231, SKBR3, and CAMA1 (Figure 37A). Furthermore, we obtained 16 normal and 32 primary

105

breast tumor RNAs from H. Lee Moffitt Tissue Core. In the normal breast, semi-qPCR

showed almost no expression of TIL. However, TIL is frequently overexpressed in tumor

RNA samples (Figure 37B). Furthermore, we also obtained 8 pairs of adjacent normal

and breast tumor RNAs. In 6 of the 8 pairs we observed overexpression of TIL (Figure

37C).
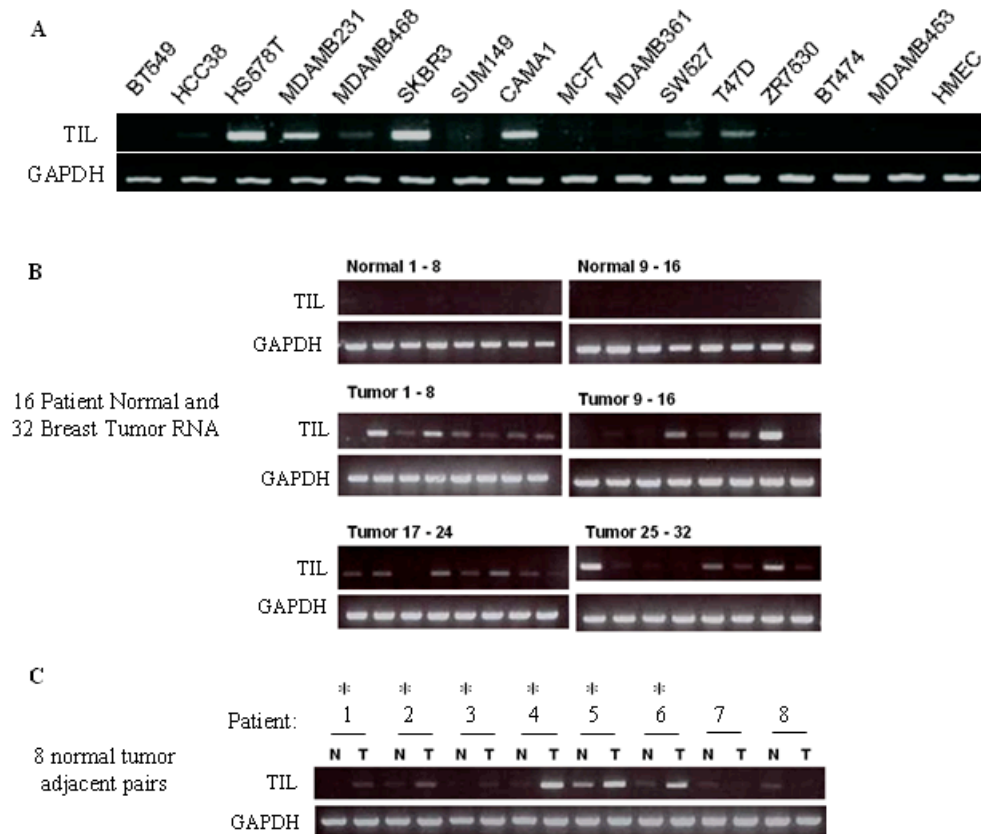
Figure 37. TIL is elevated in breast cancer cell lines and primary tumors. (**A**) Semi-qPCR for TIL expression was performed in a panel of e breast cancer cell lines. (**B** and **C**) Indicated normal and breast tumors were subjected into semi-qPCR analysis for expression of TIL.

Next, we performed *in situ* hybridization analysis on breast TMAs which contain

cores of normal epithelia, DCIS, IDC, and lymph node metastasis. A pattern emerged in

that TIL was shown to have progressively higher expression as disease became more

advanced. For example, expression in DCIS was higher than normal, and expression in IDC was higher than DCIS. The most significant overexpression was observed in the metastasis (Figure 38A). Representative images were taken for cores in each group (Figure 38B). Based on the data available, we assessed if there was any significant association between elevated TIL and patient clinical and pathological characteristics. Our analysis revealed that high expression of TIL correlates with late stage tumor, positive lymph node metastasis, and triple-negative breast cancer (Figure 38C). There was however no significant correlation with overall survival as shown by Kaplan-Meier Plot (Figure 38D). This result may be due to a non-stratified population and the limited number of breast cancer examined. For example, if we were to separate the population into late-stage versus early-stage, or triple-negative versus ER-positive populations, we may see significant survival data. In this study, however, there are very few patients with late-stage and triple-negative breast cancer to make this conclusion (data not shown). The prognostic value of TIL warrants further investigation.

TIL regulates migration, invasion, and has elevated expression in BCSCs. Expression of TIL is high in the triple-negative and highly invasive MDA-MB-231 breast cancer cell line. Therefore, we used this cell line to confirm knockdown of TIL expression using 2 siRNAs (si-TIL-1 and si-TIL-2) compared to si-CTL (Figure 39A). Once knockdown was confirmed, we assayed MDA-MB-231 for phenotypic changes in migration and invasion. After 48 hours of TIL depletion, cells for each group were seeded in upper chambers and assayed for migration and invasion. After 16 h, top
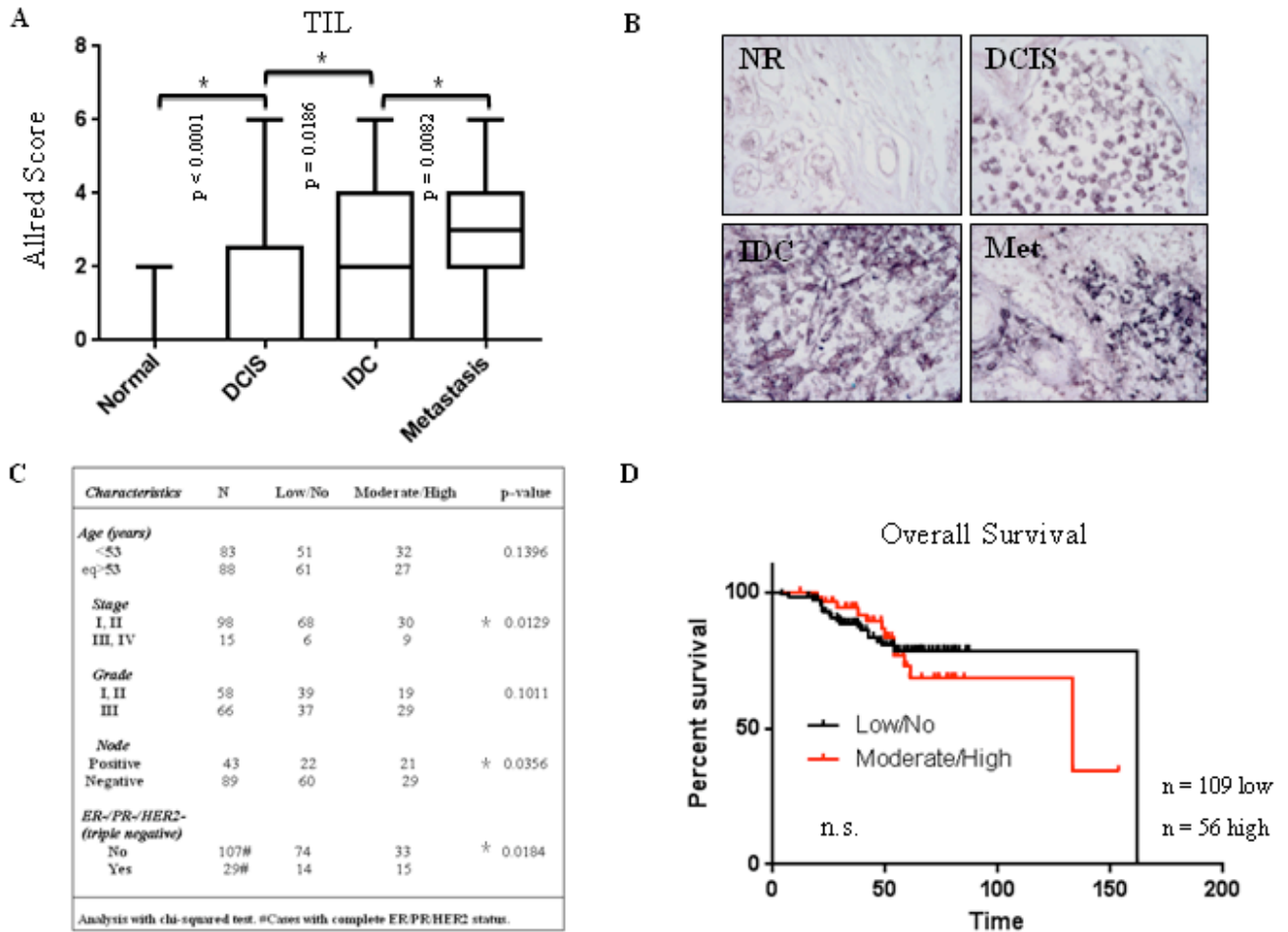
Figure 38. TIL is progressively up-regulated with advanced disease and associates with late stage tumor, positive lymph node, and triple-negative breast cancer. (**A**) TMA-ISH was performed using LNA-probe and scored. (**B**) Representative images of patients in each group are shown. (**C**) Correlation of clinical and pathological data with TIL expression. (**D**) Kaplan-Meier curve representing overall survival.

chambers were cleared and number of migratory cells were quantified by 4 non-biased fields. Loss of TIL resulted in significant reduction of both migration and invasion in comparison to control (Figure 39B and 39C).

Next, a cDNA clone of TIL was generated using SKBR3 RNA as a template. Primers were designed based on the large human exon conserved in human from

mouse AK060. We called this region "Exon 3" and hypothesized this to contain functional domain. As predicted we amplified a 657nt insert in which we cloned into the pcDNA3.1+ expression vector. Next, we over-expressed TIL in MCF7 cells (Figure 39D). Ectopic introduction was able to induce invasion capacity >3-fold compared to vector control (Figure 39E).

To address if TIL is involved in regulation of breast cancer stem cell (BCSC) property, we treated the HMEC cells with TGFβ (5 ng/ml) for 9 days. HMEC cells underwent EMT in about 24 hours, and beyond this continually moved towards a stem-like status. Using this method, it is possible to generate a significant population of CD44+/CD24- cells in culture, indicative of BCSC population. Next, the cells were sorted by flow for CD44-/CD24+ and CD44+/CD24- populations and RNA was isolated. RT-qPCR analysis showed that there was ~13-fold higher expression of TIL in the BCSC population compared to non-BCSCs (Figure 39F).

TIL post-transcriptionally upregulates C-MYC. TIL locates at chromosome 8q23, about 10MB away from C-MYC gene. Due to close proximity, and ability of lncRNAs to target neighboring gene expression in "*cis*", we directly addressed whether TIL regulates C-MYC. We first silenced TIL expression in MDA-MB-231 cell using si-TIL-1 (Figure 40A). As shown in Figure 39A, this siRNA gives ~80% knockdown. Next, we checked whether loss of TIL expression resulted in changes in C-MYC mRNA or protein levels. Interestingly, we observed no change in C-MYC mRNA after TIL knockdown, however we did detect significant drop in C-MYC protein expression (Figure 40B and 40C). Furthermore, we introduced TIL into MCF7 cells and found that

ectopic expression of TIL also had no effect on C-MYC mRNA, but dramatically induced total C-MYC protein levels (Figure 40D-F).
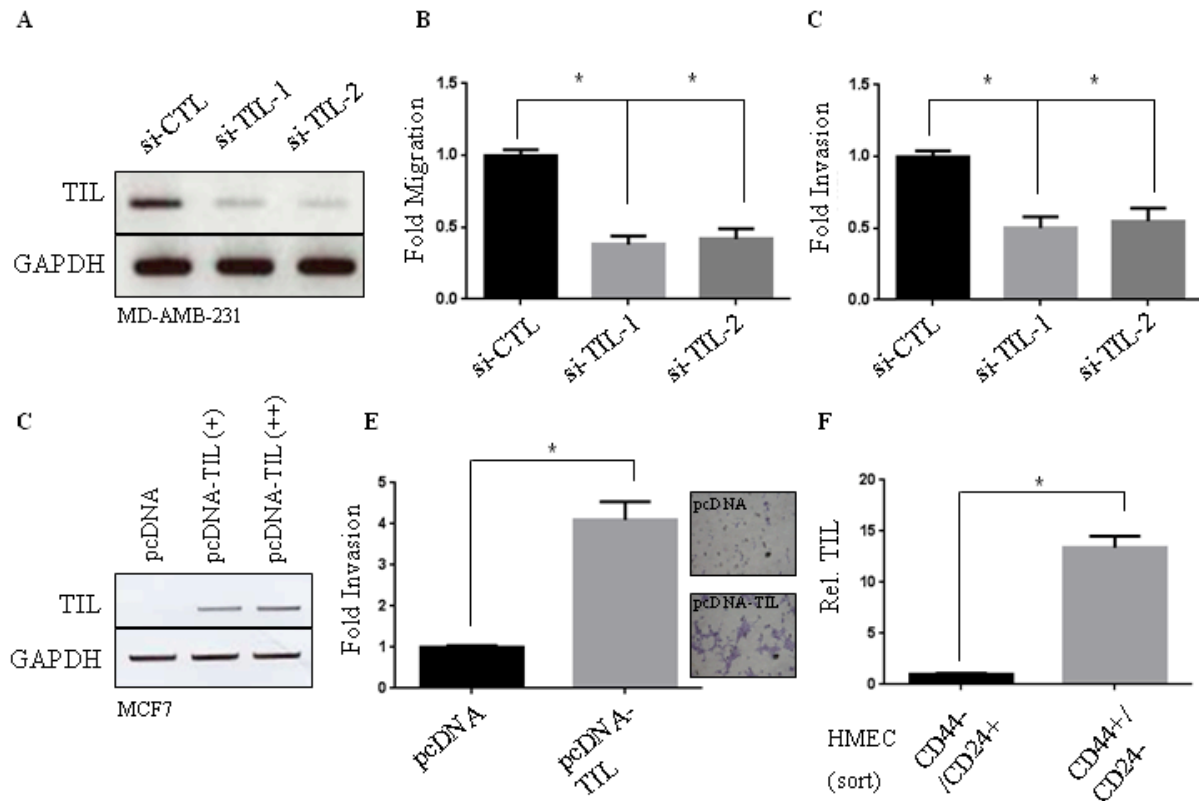
Figure 39. TIL regulates migration, invasion, and has higher expression in BCSCs.
(**A**) MDA-MB-231 cells were transfected with si-CTL, si-TIL-1, and si-TIL-2 for 48 hours and semi-qPCR was performed. (**B** and **C**) After 48 hours transfection, cells were seeded in upper chamber for 16 h and assayed for migration (B) and invasion(C). (**D**) pcDNA-TIL was transfected into MCF7 and measured by semi-qPCR. (**E**) After 48 hours transfection, MCF7 cells were seeded in upper chamber and assayed for invasion. (**F**) HMEC cells were treated with TGFβ for 9 days and sorted by flow for CD44-/CD24+ and CD44+/CD24- populations. RNA was isolated and examined for TIL expression by RT-qPCR. Asterisks represent p<0.05.

It has also been well documented that in normal breast and early onset of breast cancer, TGFβ treatment will result in down-regulation of C-MYC mRNA and protein expression [196]. In this case such as in HMEC, TGFβ pathway will act as a tumor suppressor. However, in breast cancers where the "switch" has occurred, TGFβ

treatment will induce C-MYC mRNA and eventual protein expression to trigger oncogenic phenotypes. This is the case with MDA-MB-231 line, in which 24 h TGFβ treatment (5ng/ml) induced C-MYC mRNA ~3-fold (Figure 40G). MDA-MB-231 cells were transfected with si-CTL and si-TIL for 48 hours and left untreated or treated with TGFβ for 24 hours. In the absence of TIL, TGFβ-induced C-MYC mRNA levels remain unaffected after 24 hours (Figure 40G). Interestingly, C-MYC protein expression is dramatically inhibited as shown by time course of TGFβ treatment (Figure 40H). This result suggests that TIL post-transcriptionally regulates C-MYC expression.

TIL binds to NF90. To elucidate the mechanism by which TIL post-transcriptionally regulates C-MYC, we aimed to identify a protein binding partner. To do this we used the non-biased RIP assay described in Chapter 4. Briefly, *in vitro* transcribed biotinylated "Sense" and "Antisense" TIL RNAs were incubated with 1mg MDA-MB-231 protein extract. RNA-protein complexes were immunoprecipitated using.

Streptavidin-agarose, washed, boiled, and run on SDS-page gel. We observed a very strong unique band at ~90 kDa in the "Sense" lane that was not visible in the "Antisense" lane (Figure 41A). This band was excised and subjected to mass spectrometry protein identification. The data suggested a strong enrichment for the NF90 protein in "Sense" versus "Antisense" samples. To further investigate the interaction of TIL with NF90, we performed similar RIP experiment as described above with minimal modification. After running SDS-page gel, total bound proteins were transferred to nitrocellulose membrane. Western blotting analysis was carried out with

an antibody against NF90. We observed significant expression of NF90 in "sense" lane

and virtually no expression in the "antisense" lane (Figure 41B).
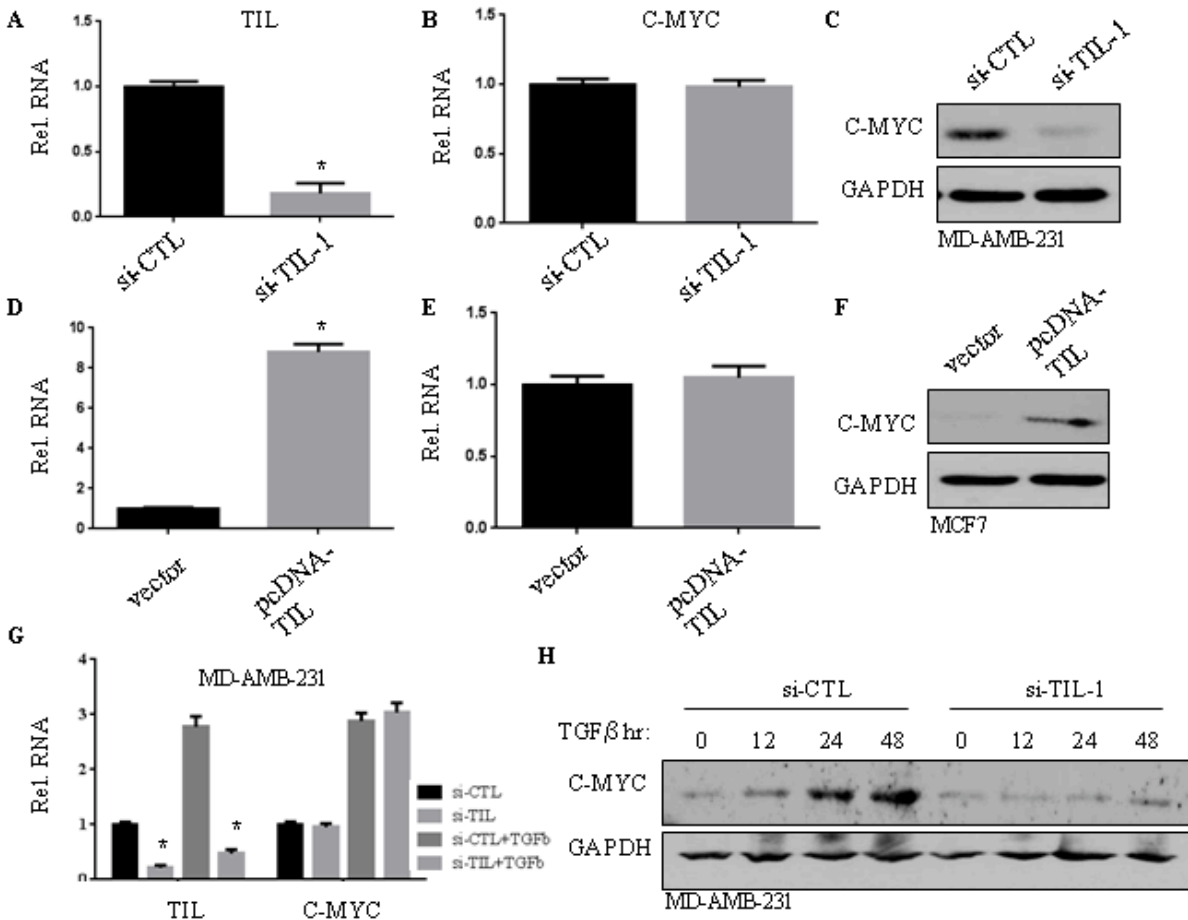
Figure 40. TIL regulates C-MYC at protein but not mRNA level. (**A**) MDA-MB-231 cells were transfected with si-CTL and si-TIL-1 and subsequently subjected into real-time PCR analysis of TIL expression. (**B** and **C**) After 48 hours of knockdown of TIL, C-MYC mRNA and protein expression was measured by RT-qPCR (B) and Western blot analysis (C). (**D** - **F**) MCF7 cells were transfected with pcDNA-TIL and vector control. Following 48 hours of incubation, RT-qPCR was performed to examine the expression of transfected plasmids (D) and C–MYC mRNA (E). C-MYC protein levels were assessed by western blotting (F). (**G**) MDA-MB-231 cells were transfected with si-CTL and si-TIL-1 for 48 h. Samples were left untreated or treated with TGFβ (5 ng/ml) for 24 hours. Total RNA was isolated and expression of TIL and C-MYC mRNA was assessed by RT-qPCR. (**H**) Cells were similarly transfected as described in panel G. Following treatment with TGFβ for indicated time, immunoblot analysis was performed with anti-C-MYC antibody. GAPDH was used as a loading control.

Next, we performed the reverse experiment to see if NF90 bound endogenous TIL *in vivo*. To do this, we lysed MDA-MB-231 cells, and then incubated with anti-GFP or anti-NF90 antibodies. Immunoprecipitation was performed using protein A/G beads and total bound protein-RNA complexes were resuspended in Trizol buffer. Total RNA was then isolated and semi-qPCR was performed. We observed a significant enrichment of endogenous TIL in the anti-NF90 samples as compared to anti-GFP (Figure 41C). These results suggest that endogenous NF90 binds to endogenous TIL *in vivo*.

C-MYC binds to NF90. Since TIL induces C-MYC protein level and binds to NF90, we next investigated if NF90 mediates TIL effect on C-MYC. NF90 is known to bind AU-rich elements (AREs) within the 3'UTR of mRNAs to influence target gene expression. NF90 can be bi-faced and multifunctional in that it can promote mRNA stability and degradation, as well as facilitate nuclear export of transcripts to the cytosol [197, 198]. We therefore analyzed the 3'UTR of C-MYC for presence of AREs using "AREsite" software [199] and found four distinct AUUUA motifs within 3'UTR of C-MYC (Figure 42A).

Furthermore, we cloned C-MYC 3'UTR (e.g., 474nt) into pcDNA3 vector for *in vitro* transcription. Next, we generated biotinylated "sense" and "antisense" RNA probes and performed similar RIP experiment with MDA-MB-231 extract as described above to assess whether the 3'UTR of C-MYC binds to NF90. Indeed, after isolating RNA-protein
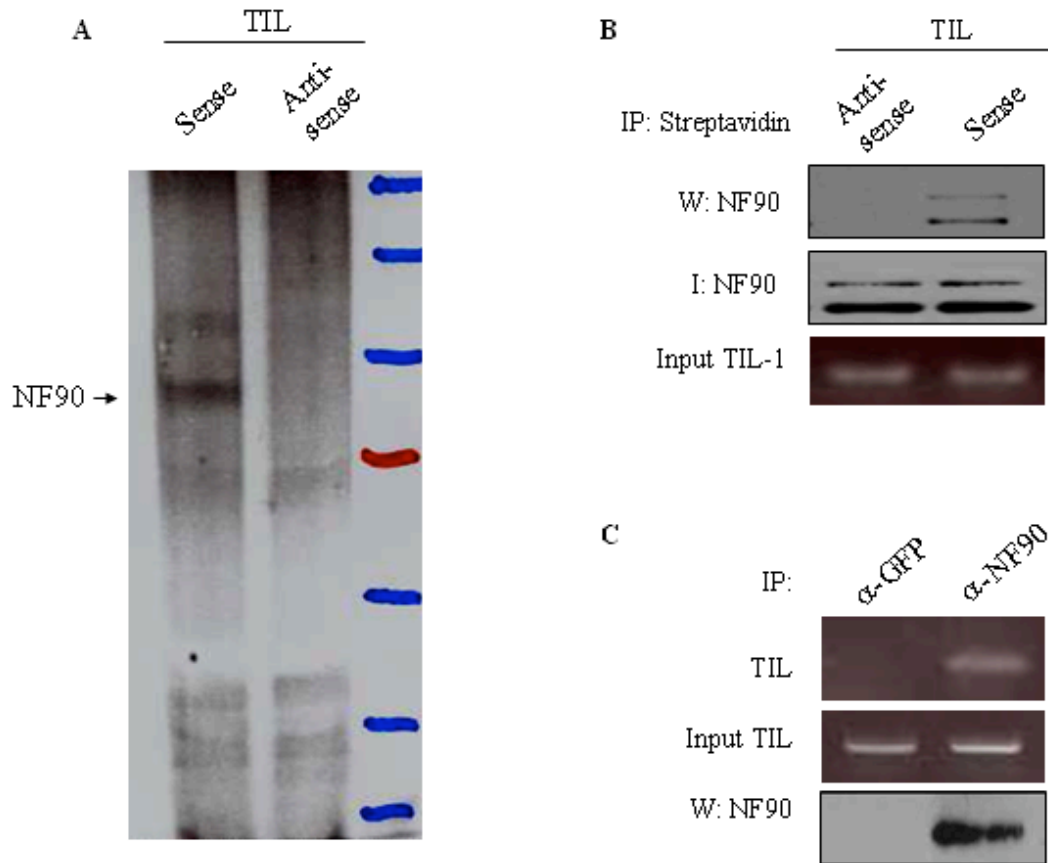
Figure 41. TIL binds to NF90. (**A**) RIP was performed by incubation of 3ug biotinylated "sense" and "antisense" RNAs of TIL with 1mg MDA-MB-231 extract in the presence of, Streptavidin-agarose. Following wash, total bound protein was run on SDS-page gel and silver stained. Unique bands were subjected to protein identification by mass spectrometry. (**B**) Similar RIP was performed and total bound protein was transferred to nitrocellulose for Western blot analysis with NF90 antibody. (**C**) Immunoprecipitates were prepared with antibodies against GFP and NF90 and then were used to isolate RNA. Endogenous TIL expression was evaluated with semi-qPCR.

complexes by Streptavidin-agarose, Western blot analysis showed a strong enrichment for NF90 in "sense" but not "antisense" (Figure 42B). Next, we performed IP on endogenous NF90 and confirmed interaction with C-MYC mRNA transcript *in vivo* by semi-qPCR (Figure 42C). Taken together, these data suggest that the C-MYC 3'UTR contains AREs which interact with NF90.

TGFβ induces TIL-dependent shuttling NF90/C-MYC to cytosol. We next investigated the significance of TGFβ stimulation in relation to the TIL/NF90/C-MYC interaction. Since NF90 can stabilize transcripts in the cytosol to enhance protein expression and since NF90 binds to C-MYC mRNA, We hypothesized that TGFβ and TIL may facilitate this process. First, we treated the MDA-MB-231 with TGFβ to examine for change of NF90 localization. Indeed, TGFβ significantly induced shuttling of NF90 from nuclear to cytosol following the treatment for 24 hours (Figure 43A). However, if cells are first depleted TIL using siRNA, NF90 shuttling in response to 24 hours TGFβ treatment is almost completely abolished (Figure 43B). In the low-TIL expressing BT474 cell line, ectopic expression of TIL can enforce NF90 shuttling from nuclear to cytosol, however has no effect on total NF90 protein levels (Figure 43C). This result suggests that TIL/NF90 interaction is important for NF90 shuttling from nuclear to cytosol.

Since C-MYC mRNA also binds to NF90, we tested whether TIL can also drive C-MYC mRNA to cytosol. We ectopically expressed TIL in BT474 and observed that TIL promoted C-MYC mRNA from nuclear to cytosol. Presumably this shuttling is through C-MYC mRNA being bound to NF90 (Figure 43D). We have shown that ectopic TIL results in increased C-MYC protein, likely the result of increased C-MYC mRNA. To further demonstrate that ectopic TIL increase of C-MYC protein is through C-MYC 3'UTR interaction with NF90, we cloned the 3'UTR of C-MYC into luciferase reporter construct (pMIR-Luciferase-3'-UTR-C-MYC). Cells were first transfected with reporter alone or 3'UTR containing reporter. After 24 hours, cells were co-transfected with

A

3'UTR C-MYC NF90 contains several AU-rich elements (AREs) in 3'UTR

```
1876  GGAAAAGTAAGGAAAACGATTCCTTCTAACAGAAATGTCCTGAGCAATCACCTATGAACTTGTTTCAAATGCATGATCAA
1956  ATGCAACCTCACAACCTTGGCTGAGTCTTGAGACTGAAAGATTTAGCCATAATGTAAACTGCCTCAAATTGGACTTTGGG
2036  CATAAAAGAACTTTTTTATGCTTACCATCTTTTTTTTTTTCTTTAACAGATTTGTATTTAAGAATTGTTTTTAAAAAATTT
2116  TAAGATTTACACAATGTTTCTCTGTAAATATTGCCATTAAATGTAAATAACTTTAATAAAACGTTTATAGCAGTTACACA
2196  GAATTTCAATCCTAGTATATAGTACCTAGTATTATAGGTACTATAAACCCTAATTTTTTTTTATTTAAGTACATTTTGCTT
2276  TTTAAAGTTGATTTTTTTCTATTGTTTTTAGAAAAAATAAAATAACTGGCAAATATATCATTGAGCCAAA
```

B   **C**
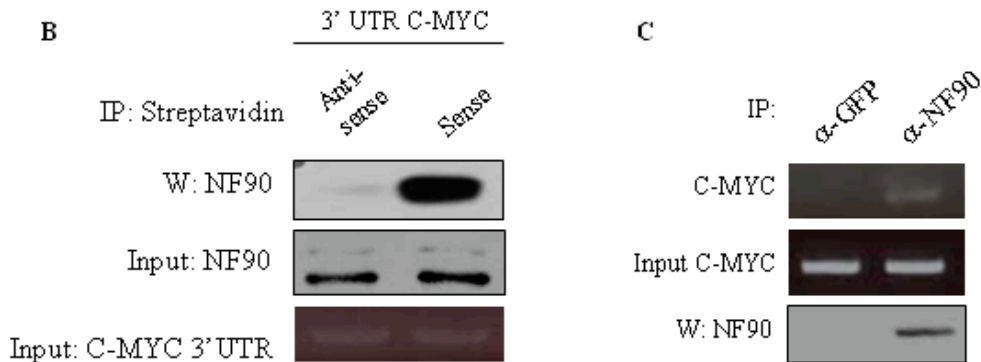
Figure 42. NF90 binds to C-MYC 3'UTR. (**A**) AREsite software identified four AUUUA motifs within the 3'UTR of C-MYC. (**B**) Three microgram biotinylated "sense" and "antisense" C-MYC 3'UTR RNA probes were incubated with 1mg MDA-MB-231 extract. Western blotting was performed for NF90 interaction with 3'UTR of C-MYC. (**C**) Anti-GFP and anti-NF90 were incubated with MDA-MB-231 extract and then immunoprecipitated in the presence of protein A/G. Total RNA was isolated and semi-qPCR was performed for C-MYC mRNA.

pcDNA or pcDNA-TIL. In cells co-transfected with 3'UTR containing reporter and TIL (V2+TIL), bioluminescence activity was significantly increased compared to controls. (Figure 43E). Taken together, these results suggest that TIL forces C-MYC mRNA to the cytosol, and this shuttling is dependent on the 3'UTR of C-MYC.

TIL enhances AKT phosphorylation of NF90 at Ser647. A previous report has demonstrated that phosphorylation of NF90 at Ser647 by AKT is essential for NF90

shuttling out of the nucleus to cytosol for binding and stabilization of IL-2 [194, 198]. We

therefore hypothesized that TIL may be involved in this phosphorylation event. To test

this, we transfected MCF7 with vector and TIL, and then NF90 was immunoprecipitated.

Since no antibody exists for NF90-p-Ser647, we probed the Western blot using
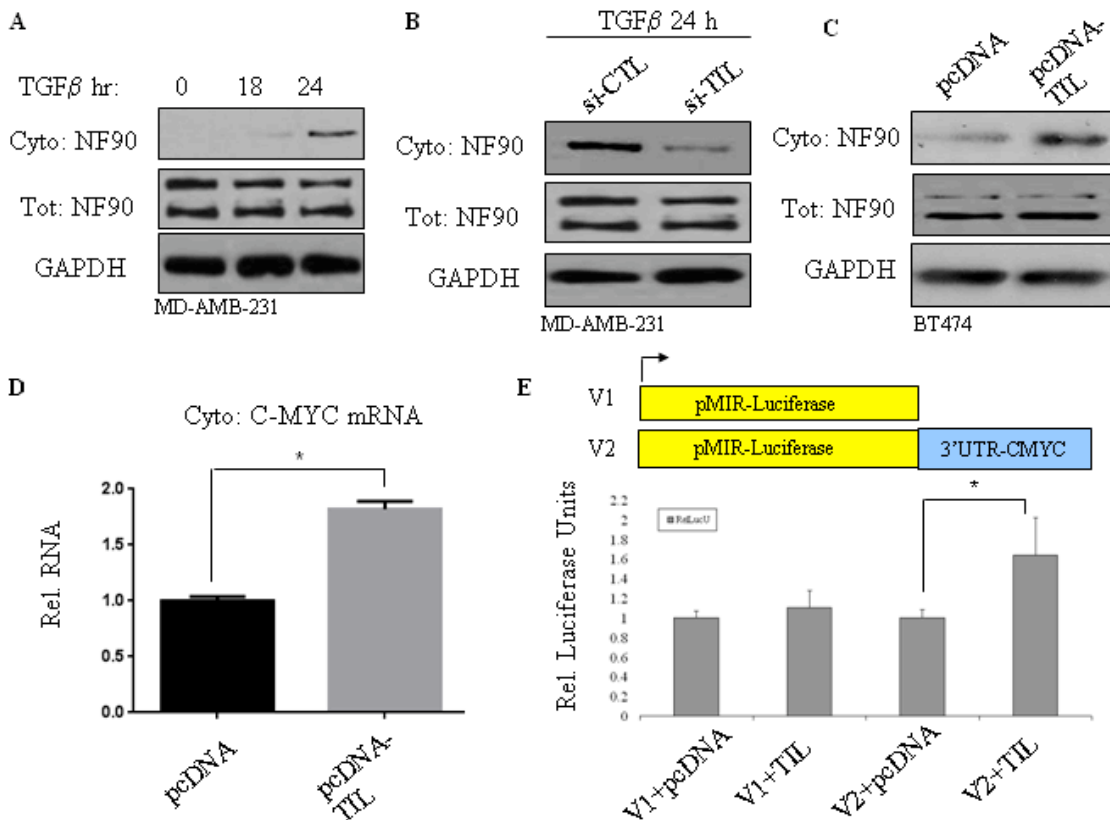
Figure 43. TGFβ enhances TIL-dependent NF90 and C-MYC mRNA shuttling to cytosol. (**A**) MDA-MB-231 cells were treated with TGFβ (5 ng/ml) and cells were fractioned to check for localization of NF90. (**B**) Cells were transfected with si-CTL or si-TIL for 48 hours and then treated with TGFβ for 24 hours. Cells were fractioned for Western blot analysis with NF90 antibody. (**C** and **D**) BT474 cells were transfected with ectopic TIL and fractioned for Western blot for NF90 (C) and RT-qPCR for C-MYC mRNA (D). (**E**) HEK293 cells were transfected with V1 (pMIR-Luciferase) or V2 (pMIR-Luciferase-3'UTR-C-MYC) for 24 hours. Cells were co-transfected with pcDNA or pcDNA-TIL for 48 hours. At total 72 hours time point, cells were lysed and luciferase activity was measured.

phospho-AKT-substrate antibody, which recognizes the RXRXXS/T motif of typical AKT phosphorylation. We observed a significant increase in signal in cells transfected with TIL at band consistent with size NF90. However, after stripping and reprobing the blot we observed no change in total NF90 levels (Figure 44A). It is important to note that Ser647 is the only residue of NF90 being phosphorylated by AKT. Therefore any signal observed using phospho-AKT-substrate antibody is indicative of p-Ser647. The same result was observed by co-transfecting NF90-GFP with TIL (Figure 44B).

We further determine if loss of TIL results in a drop in p-Ser647 signal. Previously it has been shown that TGFβ-induced EMT in mammary epithelia is through downstream signaling of the AKT pathway. This report demonstrated that TGFβ treatment coincided with activation of p-SMAD3/1 and p-AKT-473 [200]. Therefore, we knocked down TIL for 48 hours in MDA-MB-231, and then treated cells for 30 minutes with TGFβ to check for changes in p-Ser-647 of NF90. We observed a substantial increase of p-Ser647 in si-CTL transfected cells treated TGFβ compared to untreated, as predicted. However, if TIL was depleted prior to TGFβ treatment, NF90 p-Ser647 levels were completely lost to levels even below untreated cells (Figure 44C). These results suggest that TIL is involved in enhancing the phosphorylation of NF90 at Ser647 by AKT.

Discussion

We propose a model by which TIL regulates C-MYC expression through the RNA-binding protein NF90 (Figure 45). TIL and C-MYC locate ~10MB apart in the 8q23-

24 locus, a region of the genome that has been implicated in several cancers. Elevated expression of C-MYC has also been reported to be a driver in several malignancies including breast cancer. We have demonstrated that TGFβ induces TIL and C-MYC expression at RNA level. We have referred to this arm of TIL-pathway as the "transcriptional" pathway, and likely TIL and C-MYC are together up-regulated by canonical TGFβ signaling and SMAD-dependent target gene expression. Therefore TGFβ treatment will result in elevated levels of both TIL and C-MYC.
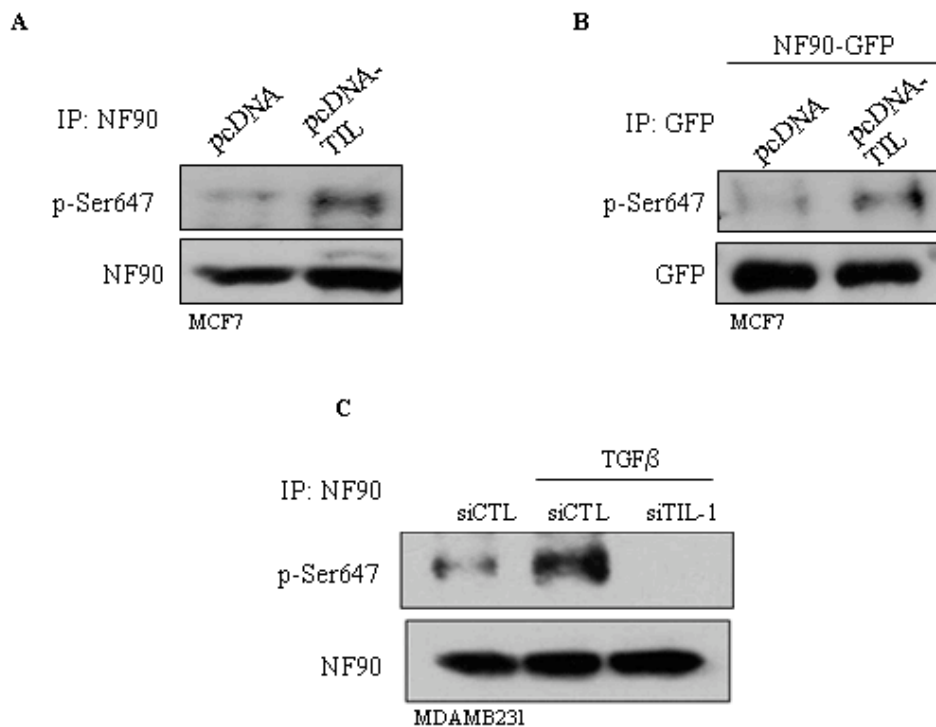
Figure 44. TIL promotes phosphorylation of NF90 at Ser-647. (A) MCF7 cells were transfected with vector or TIL for 48 hours. NF90 was immunoprecipitated and the immunoprecipitates were then immunoblotted with AKT-substrate antibody. (**B**) MCF7 cells were co-transfected with NF90-GFP and either vector or TIL for 48 h. Immunoprecipitation was performed using anti-GFP and then immunobloted with AKT-substrate antibody. (**C**) MDA-MB-231 cells were transfected with si-CTL or si-TIL for 48 hours. Cells were then treated with TGFβ (5 ng/ml) for 30 m. Immunoprecipitation and Western blot were done as described in panel A**.**

The second arm of the pathway we have referred to as the "translational", named such since it acts to regulate C-MYC post-transcriptionally. This pathway requires TGFβ activation of AKT, a known regulator in translational process. We have shown that TIL binds to NF90, and that this binding is critical for enhancing AKT phosphorylation at Ser647. This phosphorylation site is the critical event which triggers NF90 translocation to the cytosol. In the absence of TIL, NF90 phosphorylation and shuttling from nucleus to cytosol are dramatically inhibited. Enforced expression of TIL can increase p-NF90-Ser647 and shuttling NF90 to cytosol. Once in the cytosol, NF90 also binds to AREs of C-MYC-3'UTR, which promotes C-MYC mRNA stability and increases C-MYC translation.

A major question remains as to how TIL enhances phosphorylation of NF90. One explanation is that TIL binds to NF90 first, and this binding results in protein conformational change to open up Ser647 availability for AKT. Another possibility is that AKT phosphorylation happens first, and this site is stabilized and protected from phosphatase activity subsequently by TIL binding. Yet a third possible scenario is that TIL acts as a molecular scaffold that can also bind to AKT, and serve to direct the successful interaction between the two proteins. The C-MYC mRNA also serves as a player in this complex, and similar questions remain as to the order of events of TIL and C-MYC mRNA binding to NF90. NF90 contains 2 RBDs, so it is possible that TIL and C-MYC mRNA binding occurs independently, or lncRNA or mRNA bind first to induce a protein conformational change to load the second RNA product. The axis by which TIL/C-MYC/AKT/NF90 work together is a very dynamic process, and one that warrants future investigation.
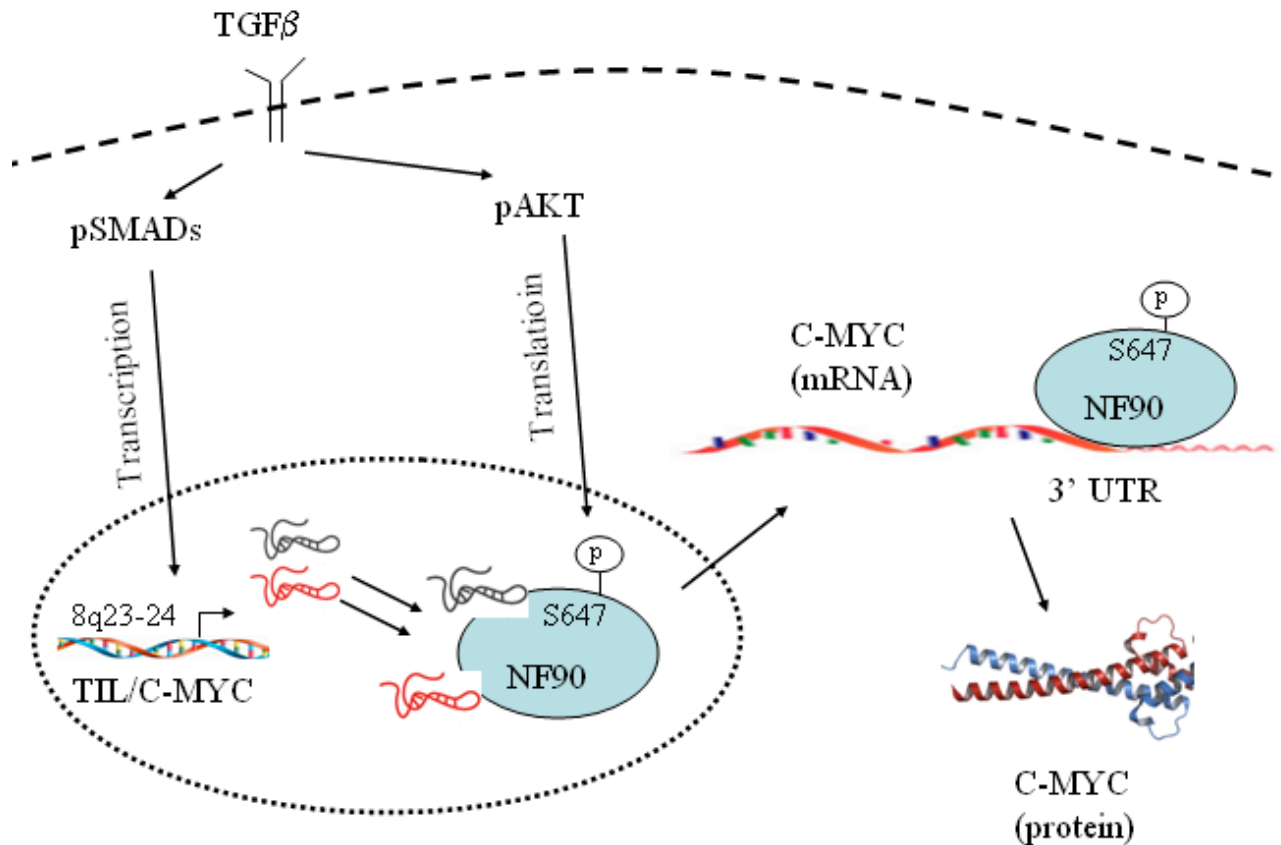
Figure 45. A model proposed by which TIL regulates C-MYC through NF90. TGFβ can activate target gene transcription through SMAD signaling. TIL and C-MYC are both transcriptionally activated. TGFβ can also activate the AKT pathway which can influence translation regulatory machinery in the cell. TIL binding to NF90 enhances AKT phosphorylation of NF90-Ser647 and C-MYC mRNA shuttling to the cytosol. Once in the cytosol, NF90 can bind to ARE-containing elements in the 3'UTR of C-MYC to stabilize transcript and increase protein expression.

Methods

RNA immunoprecipitation (RIP) assay. Briefly, 3ug biotinylated Sense and

Antisense in vitro transcribed TIL was incubated with 1mg HEK293 extract in RIP Buffer,

respectively. Ribonucleoprotein complexes were immunoprecipitated using Streptavidin-agarose (Life Technologies). IPs were washed in RIP Buffer and total isolated protein was run on SDS-PAGE gel and silver stained. Unique bands were submitted for Mass Spectrometry protein identification. Previously described in Chapter 4.

Locked nucleic acid in situ hybridization of formalin-fixed, paraffin-embedded tissue microarray (LNA-TMA-ISH). TIL LNA probe was end labeled with digoxigenin-ddUTP terminal transferase. After deparaffinization and antigen retrieval, hybridizaton with 10 nmol/L LNA TIL probe for 12 hours. After washing sections were blocked for 1 h and incubated with anti-DIG-AP Fab. Following washing TMAs were processed and then visualized.

Immunofluorescence, immunoblotting, and antibodies. As described in Chapters 3 and 4. Antibodies against total C-MYC was from Santa Cruz Biotechnologies (Santa Cruz, CA, USA), NF90 from Bethyl Laboratories (Montgomery, TX USA), phospho-AKT substrate (RXXS/T) from Cell Signaling Technologies (Danvers, MA USA).

Plasmids and siRNAs. A 656nt cDNA insert of TIL (Exon3) was amplified and TA-cloned into pGEM-T-easy vector Plasmid containing insert was sequenced, insert was then isolated by EcoRI restriction digest, and cloned into pcDNA3.1 (pcDNA-TIL). C-MYC 3-prime UTR cDNA was amplified using Green Taq PCR Master Mix (Promega) and cloned into pMIR-REPORT luciferase reporter vector (Life Technologies). SiRNAs

were designed using IDT siRNA Design Software. Two best siRNAs against human TIL were determined empirically and used for experimentation.

BCSCs sorting by flow for RT-qPCR. HMEC cells were treated with TGFβ (5 ng/ml) for 9 days. Cells were then washed in PBS, trypsinzed, and incubated with antibodies for CD24-FITC and CD44-PE from BD Biosciences (San Jose, CA USA). Cells were then sorted by flow cytometry for FITC-positive cells, indicative of CD24+/CD44- population, and PE-positive cells indicative of CD44+/CD24- population. Total RNA was isolated from both populations via Trizol and TIL expression was evaluated by RT-qPCR.

Dual reporter luciferase assay. HEK293 cells were transfected with pMIR-Luciferase (firefly) containing constructs and renilla luciferase control plasmid were co-transfected. After 24 hours, pcDNA or TIL was then transfected for 48 hours. At 72 hours time point, cells were lysed, and signal for firefly and renilla luciferase was measured using Dual-Luciferase Reporter Assay System from Promega. Results are representative of triplicate ratios firefly/renilla luciferase signals.

CHAPTER 6:

Conclusions and Implications

Accumulating evidence indicates that the majority of the mouse and human genomes are transcribed into ncRNAs. Several classes of functional ncRNAs are emerging within the cell, including lncRNAs. The functional contribution of lncRNAs to tumorigenesis and cancer progression remains elusive.

The central focus of this dissertation work is to characterize the biological mechanisms of lncRNAs in the context TGFβ-mediated phenotypes in metastatic breast cancer, such as EMT and cancer cell stemness. In the initial stages, the primary mouse mammary cell line NMuMG was treated with TGFβ to induce EMT. Global changes in lncRNA gene expression were profiled by microarray and significant changes were validated by RT-qPCR. Our analysis revealed that a total of ~600 lncRNAs were induced or repressed in this response, respectively. LncRNAs are poorly conserved across species, however there are examples of well conserved lncRNAs that function in cancer [33, 48]. Therefore, we developed a strategy to predict mouse lncRNAs that had a strong potential for a human ortholog.

The motivation behind this strategy was that the end goal was to characterize important lncRNAs in human cancer. Also, proteins are prime examples of genomic regions that are very well conserved and contain function gene products. If lncRNAs

demonstrated similar features, it was likely that this evolutionary selection was due to the fact that a functional gene product resided in this genomic location. Therefore, our hypothesis suggested that ortholog lncRNAs are more likely to be functional due to selective evolutionary pressures. Strategically, conserved lncRNAs will allow us to carry out knockout and transgenic animal studies in order to define physiological and pathological function of these lncRNAs *in vivo*.

The strategy to predict mouse lncRNAs with a strong potential for a human ortholog took into account 4 parameters; significant sequence identity and coverage of transcript, identity mapped to syntenic loci in both genomes, regulated by established transcriptional factor and pathway (TGFβ), and corresponding transcripts were intergenic and free standing. Next, we subjected the top 100 mouse lncRNAs induced by TGFβ in the microarray analysis to these parameters. Ten mouse lncRNAs emerged as potentially having a human ortholog. Follow-up experimentation showed that 9 were detected at the RNA level in human (denoting ortholog transcript), 3 were also under regulation of TGFβ in human, and 1 resided in a genomic location of interest.

This study focused on characterizing 3 of conserved lncRNAs; lncRNA-HIT, WDFY3-AS2, and TIL. LncRNA-HIT is a 5-prime distal Hoxa residing transcript. It is becoming well established that Hox gene clusters are lncRNA-governed, and several important examples have surfaced in literature detailing this regulation [149, 150]. Knockdown experiments using siRNA showed that mouse lncRNA-Hit is functional in driving TGFβ-induced EMT, migration, and invasion. Overexpression of lncRNA-Hit disrupted tight junction in NMuMG and this correlated with a drop in E-cadherin expression. Furthermore, human lncRNA-HIT is significantly elevated in breast cancer,

and this elevated expression associates with invasive disease, high grade tumor, and metastatic breast cancer (i.e., lymph node metastasis). LncRNA-HIT positively regulates HOXA13 mRNA and protein expression in human cell lines. Elevated expression of lncRNA-HIT could drive high levels HOXA13 in patient tumor to drive oncogenic and metastatic processes.

WDFY3-AS2 lies in the antisense orientation to protein coding gene WDFY3 on chromosome 4q21, and also drives TGFβ-mediated phenotypes such as cell motility and EMT. High levels of WDFY3-AS2 in patients are associated with late stage tumor, triple-negative subtype, and poorer overall patient survival. Our experimentation has shown that WDFY3-AS2 positively regulates WDFY3 and STAT3 through protein binding partner hnRNPR. WDFY3 is a critical regulator of selective macroautophagy by serving as an adapter for recruitment of the core autophagic machinery. Interestingly, it has been suggested that several triple-negative breast cancers are driven by STAT3 addicted autophagy. WDFY3-AS2 could be a bridge to link these 2 molecules in TGFβ-induced autophagy potentiating a signaling loop that drives cancer phenotypes.

TGFβ-regulated TIL is an 8q23-24 residing lncRNA that also promotes EMT and cancer cell stemness. In breast cancer patients, elevated levels of TIL associate with late stage tumor, metastasis, and triple-negative subtype. There have been several examples of functional lncRNAs in the 8q23-24 region that can regulate C-MYC expression. Therefore, we examined if TIL also has similar function toward C-MYC. Our experiments demonstrated that TIL can post-transcriptionally regulate C-MYC through NF90. Interestingly, TIL enhances phosphorylation of NF90 by AKT resulting in

increased shuttling of NF90 to the cytosol. In cytosol, NF90 can bind to and stabilize C-MYC mRNA and increase C-MYC protein expression.

The 3 candidates mechanistically function very differently from each other illustrating the dynamic roles of lncRNA can play within a cell. Clearly disruption and deregulation of lncRNA expression can be important in driving tumorigenesis and metastatic processes.

The translational impact of lncRNAs is an emerging aspect that has yet to be extensively explored. In future, there is a great potential for using lncRNA as biomarkers for diagnostics, monitoring recurrence, and predicting response to therapy. A major advantage of using RNA as tool and readout is the fact that it can be detected using PCR based technologies. This means that only a small amount of material is needed to be informative, and this material can be acquired through non-invasive means like blood and urine. If a panel of lncRNAs could be identified as specific to a particular cancer, this could provide a novel avenue to non-invasive preventative screening and diagnosis. This same principle could be applied to monitoring tumor recurrence and metastasis after surgery and chemotherapy.

LncRNAs could also serve as important biomarkers in personalized and precision medicine. For example, it has been reported that triple-negative breast cancers driven by STAT3-addicted autophagy are responsive to drugs like chloroquine and bafilomycin. In this case, the status of WDFY3-AS2 expression in the primary tumor would be a powerful readout for prediction and repurposing these FDA approved drugs. Similarly, TIL can be a marker of MYC driven malignancy and these patients could be benefit from newer drugs like JQ-1.

Of course with the identification of novel oncogenic drivers of cancer, the question is always asked if these molecules can serve as direct targets for therapy. There is a major research effort underway to begin targeting the "un-druggable", which currently includes the nucleic acids and proteins like transcription factors. This is an exciting time to expand the therapeutic toolbox, however I do believe directly targeting lncRNAs in cancer may still take a lot of efforts to prove efficacious. Traditionally, delivering nucleic acid based therapies, like small antisense oligos (ASOs), has been difficult due to limitations in methods of delivery and stability of the molecules once systemically administered. In principle, however, drugs designed against RNAs could be more specific than those against protein given the exactness of Watson-Crick base pairing. Future extensive investigation is required for better understanding the basic underlying biology of lncRNAs and their role in human malignancy such as metastatic breast cancer with aiming to develop biomarkers and effective novel targeted therapy.

REFERENCES CITED

1. Siegel, R., et al., *Cancer statistics, 2014.* CA: a cancer journal for clinicians, 2014. **64**(1): p. 9-29.

2. Matsen, C.B. and L.A. Neumayer, *Breast cancer: a review for the general surgeon.* JAMA surgery, 2013. **148**(10): p. 971-9.

3. Bombonati, A. and D.C. Sgroi, *The molecular pathology of breast cancer progression.* The Journal of pathology, 2011. **223**(2): p. 307-17.

4. Allred, D.C., *Ductal carcinoma in situ: terminology, classification, and natural history.* Journal of the National Cancer Institute. Monographs, 2010. **2010**(41): p. 134-8.

5. Virnig, B.A., et al., *Ductal carcinoma in situ of the breast: a systematic review of incidence, treatment, and outcomes.* Journal of the National Cancer Institute, 2010. **102**(3): p. 170-8.

6. Ford, D., et al., *Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium.* American journal of human genetics, 1998. **62**(3): p. 676-89.

7. Roy, R., J. Chun, and S.N. Powell, *BRCA1 and BRCA2: different roles in a common pathway of genome protection.* Nature reviews. Cancer, 2012. **12**(1): p. 68-78.

8. Ahmed, M. and N. Rahman, *ATM and breast cancer susceptibility.* Oncogene, 2006. **25**(43): p. 5906-11.

9. Nevanlinna, H. and J. Bartek, *The CHEK2 gene and inherited breast cancer susceptibility.* Oncogene, 2006. **25**(43): p. 5912-9.

10. Antoniou, A.C., et al., *Breast-cancer risk in families with mutations in PALB2.* The New England journal of medicine, 2014. **371**(6): p. 497-506.

11. Heikkinen, K., et al., *RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability.* Carcinogenesis, 2006. **27**(8): p. 1593-9.

12. Hsu, H.M., et al., *Breast cancer risk is associated with the genes encoding the DNA double-strand break repair Mre11/Rad50/Nbs1 complex.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2007. **16**(10): p. 2024-32.

13. Malkin, D., et al., *Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms.* Science, 1990. **250**(4985): p. 1233-8.

14. Li, J., et al., *PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer.* Science, 1997. **275**(5308): p. 1943-7.

15. Ignatiadis, M. and C. Sotiriou, *Luminal breast cancer: from biology to treatment.* Nature reviews. Clinical oncology, 2013. **10**(9): p. 494-506.

16. Weigelt, B. and J.S. Reis-Filho, *Histological and molecular types of breast cancer: is there a unifying taxonomy?* Nature reviews. Clinical oncology, 2009. **6**(12): p. 718-30.

17. Fisher, B., et al., *A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors.* The New England journal of medicine, 1989. **320**(8): p. 479-84.

18. Tran, B. and P.L. Bedard, *Luminal-B breast cancer and novel therapeutic targets.* Breast cancer research : BCR, 2011. **13**(6): p. 221.

19. Yarden, Y., *Biology of HER2 and its importance in breast cancer.* Oncology, 2001. **61 Suppl 2**: p. 1-13.

20. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2.* The New England journal of medicine, 2001. **344**(11): p. 783-92.

21. Dent, R., et al., *Triple-negative breast cancer: clinical features and patterns of recurrence.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2007. **13**(15 Pt 1): p. 4429-34.

22. Rakha, E.A. and I.O. Ellis, *Triple-negative/basal-like breast cancer: review.* Pathology, 2009. **41**(1): p. 40-7.

23. Neve, R.M., et al., *A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes.* Cancer cell, 2006. **10**(6): p. 515-27.

24. *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.

25. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

26. Alexander, R.P., et al., *Annotating non-coding regions of the genome.* Nature reviews. Genetics, 2010. **11**(8): p. 559-71.

27. Mockler, T.C., et al., *Applications of DNA tiling arrays for whole-genome analysis.* Genomics, 2005. **85**(1): p. 1-15.

28. Morin, R., et al., *Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing.* BioTechniques, 2008. **45**(1): p. 81-94.

29. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nature biotechnology, 2010. **28**(5): p. 511-5.

30. Birney, E., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.

31. Kapranov, P., et al., *RNA maps reveal new RNA classes and a possible function for pervasive transcription.* Science, 2007. **316**(5830): p. 1484-8.

32. Gibb, E.A., C.J. Brown, and W.L. Lam, *The functional role of long non-coding RNA in human carcinomas.* Molecular cancer, 2011. **10**: p. 38.

33. Prensner, J.R. and A.M. Chinnaiyan, *The emergence of lncRNAs in cancer biology.* Cancer discovery, 2011. **1**(5): p. 391-407.

34. Farh, K.K., et al., *The widespread impact of mammalian MicroRNAs on mRNA repression and evolution.* Science, 2005. **310**(5755): p. 1817-21.

35. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions.* Nature reviews. Genetics, 2009. **10**(3): p. 155-9.

36. Rinn, J.L. and H.Y. Chang, *Genome regulation by long noncoding RNAs.* Annual review of biochemistry, 2012. **81**: p. 145-66.

37. Cheng, J., et al., *Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution.* Science, 2005. **308**(5725): p. 1149-54.

38. Oler, A.J., et al., *Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors.* Nature structural & molecular biology, 2010. **17**(5): p. 620-8.

39. Johnsson, P., et al., *Evolutionary conservation of long non-coding RNAs; sequence, structure, function.* Biochimica et biophysica acta, 2014. **1840**(3): p. 1063-71.

40. Pang, K.C., M.C. Frith, and J.S. Mattick, *Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function.* Trends in genetics : TIG, 2006. **22**(1): p. 1-5.

41. Washietl, S., M. Kellis, and M. Garber, *Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals.* Genome research, 2014. **24**(4): p. 616-28.

42. Khaitovich, P., et al., *Evolution of primate gene expression.* Nature reviews. Genetics, 2006. **7**(9): p. 693-702.

43. Brown, C.J., et al., *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus.* Cell, 1992. **71**(3): p. 527-42.

44. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome research, 2012. **22**(9): p. 1775-89.

45. Clemson, C.M., et al., *XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure.* The Journal of cell biology, 1996. **132**(3): p. 259-75.

46. Zhao, J., et al., *Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome.* Science, 2008. **322**(5902): p. 750-6.

47. Lee, J.T., *Regulation of X-chromosome counting by Tsix and Xite sequences.* Science, 2005. **309**(5735): p. 768-71.

48. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.

49. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes.* Science, 2010. **329**(5992): p. 689-93.

50. Salmena, L., et al., *A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?* Cell, 2011. **146**(3): p. 353-8.

51. Kallen, A.N., et al., *The imprinted H19 lncRNA antagonizes let-7 microRNAs.* Molecular cell, 2013. **52**(1): p. 101-12.

52. Loewer, S., et al., *Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells.* Nature genetics, 2010. **42**(12): p. 1113-7.

53. Wang, Y., et al., *Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal.* Developmental cell, 2013. **25**(1): p. 69-80.

54. Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR.* Nature, 2013. **493**(7431): p. 231-5.

55. Gong, C. and L.E. Maquat, *lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements.* Nature, 2011. **470**(7333): p. 284-8.

56. Carrieri, C., et al., *Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat.* Nature, 2012. **491**(7424): p. 454-7.

57. Beltran, M., et al., *A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition.* Genes & development, 2008. **22**(6): p. 756-69.

58. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nature reviews. Genetics, 2014. **15**(4): p. 272-86.

59. Ong, C.T. and V.G. Corces, *Enhancer function: new insights into the regulation of tissue-specific gene expression.* Nature reviews. Genetics, 2011. **12**(4): p. 283-93.

60. Orom, U.A., et al., *Long noncoding RNAs with enhancer-like function in human cells.* Cell, 2010. **143**(1): p. 46-58.

61. Ling, H., et al., *CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer.* Genome research, 2013. **23**(9): p. 1446-61.

62. Huarte, M., et al., *A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.* Cell, 2010. **142**(3): p. 409-19.

63. Hung, T., et al., *Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters.* Nature genetics, 2011. **43**(7): p. 621-9.

64. Di Ruscio, A., et al., *DNMT1-interacting RNAs block gene-specific DNA methylation.* Nature, 2013. **503**(7476): p. 371-6.

65. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.* Nature, 2010. **464**(7291): p. 1071-6.

66. Liu, X.H., et al., *The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer.* BMC cancer, 2013. **13**: p. 464.

67. Chiyomaru, T., et al., *Long non-coding RNA HOTAIR is targeted and regulated by miR-141 in human cancer cells.* The Journal of biological chemistry, 2014. **289**(18): p. 12550-65.

68. Yildirim, E., et al., *Xist RNA is a potent suppressor of hematologic cancer in mice.* Cell, 2013. **152**(4): p. 727-42.

69. Huang, K.C., et al., *Relationship of XIST expression and responses of ovarian cancer to chemotherapy.* Molecular cancer therapeutics, 2002. **1**(10): p. 769-76.

70. Ganesan, S., et al., *BRCA1 supports XIST RNA concentration on the inactive X chromosome.* Cell, 2002. **111**(3): p. 393-405.

71. Xiao, C., et al., *The XIST noncoding RNA functions independently of BRCA1 in X inactivation.* Cell, 2007. **128**(5): p. 977-89.

72. Miyoshi, N., et al., *Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q.* Genes to cells : devoted to molecular & cellular mechanisms, 2000. **5**(3): p. 211-20.

73.   Zhang, X., et al., *Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions.* Endocrinology, 2010. **151**(3): p. 939-47.

74.   Zhou, Y., et al., *Activation of p53 by MEG3 non-coding RNA.* The Journal of biological chemistry, 2007. **282**(34): p. 24731-42.

75.   Braconi, C., et al., *microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer.* Oncogene, 2011. **30**(47): p. 4750-6.

76.   Pasmant, E., et al., *Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF.* Cancer research, 2007. **67**(8): p. 3963-9.

77.   Sasaki, S., et al., *Molecular processes of chromosome 9p21 deletions in human cancers.* Oncogene, 2003. **22**(24): p. 3792-8.

78.   Cunnington, M.S., et al., *Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression.* PLoS genetics, 2010. **6**(4): p. e1000899.

79.   Williamson, M.P., et al., *p16 (CDKN2) is a major deletion target at 9p21 in bladder cancer.* Human molecular genetics, 1995. **4**(9): p. 1569-77.

80.   Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.* Molecular cell, 2010. **38**(5): p. 662-74.

81.   Zhang, E.B., et al., *Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a.* Oncotarget, 2014. **5**(8): p. 2276-92.

82.   Witte, J.S., *Multiple prostate cancer risk variants on 8q24.* Nature genetics, 2007. **39**(5): p. 579-80.

83.   Ahmadiyeh, N., et al., *8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(21): p. 9742-6.

84.   Wasserman, N.F., I. Aneas, and M.A. Nobrega, *An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer.* Genome research, 2010. **20**(9): p. 1191-7.

85.   Jia, L., et al., *Functional enhancers at the gene-poor 8q24 cancer-linked locus.* PLoS genetics, 2009. **5**(8): p. e1000597.

86.   Kim, T., et al., *Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(11): p. 4173-8.

87.   Tseng, Y.Y., et al., *PVT1 dependence in cancer with MYC copy-number increase.* Nature, 2014. **512**(7512): p. 82-6.

88.   Hu, X., et al., *A Functional Genomic Approach Identifies FAL1 as an Oncogenic Long Noncoding RNA that Associates with BMI1 and Represses p21 Expression in Cancer.* Cancer cell, 2014. **26**(3): p. 344-57.

89.   Yuan, J.H., et al., *A long noncoding RNA activated by TGF-beta promotes the invasion-metastasis cascade in hepatocellular carcinoma.* Cancer cell, 2014. **25**(5): p. 666-81.

90. Massague, J., *TGFbeta signalling in context.* Nature reviews. Molecular cell biology, 2012. **13**(10): p. 616-30.

91. Vilar, J.M., R. Jansen, and C. Sander, *Signal processing in the TGF-beta superfamily ligand-receptor network.* PLoS computational biology, 2006. **2**(1): p. e3.

92. Shi, Y. and J. Massague, *Mechanisms of TGF-beta signaling from cell membrane to the nucleus.* Cell, 2003. **113**(6): p. 685-700.

93. Wu, M.Y. and C.S. Hill, *Tgf-beta superfamily signaling in embryonic development and homeostasis.* Developmental cell, 2009. **16**(3): p. 329-43.

94. Oka, K., et al., *The role of TGF-beta signaling in regulating chondrogenesis and osteogenesis during mandibular development.* Developmental biology, 2007. **303**(1): p. 391-404.

95. Ewan, K.B., et al., *Latent transforming growth factor-beta activation in mammary gland: regulation by ovarian hormones affects ductal and alveolar proliferation.* The American journal of pathology, 2002. **160**(6): p. 2081-93.

96. Sun, P.D. and D.R. Davies, *The cystine-knot growth-factor superfamily.* Annual review of biophysics and biomolecular structure, 1995. **24**: p. 269-91.

97. Hinck, A.P. and M.D. O'Connor-McCourt, *Structures of TGF-beta receptor complexes: implications for function and therapeutic intervention using ligand traps.* Current pharmaceutical biotechnology, 2011. **12**(12): p. 2081-98.

98. Massague, J., *TGF-beta signal transduction.* Annual review of biochemistry, 1998. **67**: p. 753-91.

99. Kirsch, T., W. Sebald, and M.K. Dreyer, *Crystal structure of the BMP-2-BRIA ectodomain complex.* Nature structural biology, 2000. **7**(6): p. 492-6.

100. Heldin, C.H., K. Miyazono, and P. ten Dijke, *TGF-beta signalling from cell membrane to nucleus through SMAD proteins.* Nature, 1997. **390**(6659): p. 465-71.

101. Kretzschmar, M. and J. Massague, *SMADs: mediators and regulators of TGF-beta signaling.* Current opinion in genetics & development, 1998. **8**(1): p. 103-11.

102. Zhang, Y., et al., *Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(3): p. 974-9.

103. Ebisawa, T., et al., *Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation.* The Journal of biological chemistry, 2001. **276**(16): p. 12477-80.

104. !!! INVALID CITATION !!!

105. Lin, X., M. Liang, and X.H. Feng, *Smurf2 is a ubiquitin E3 ligase mediating proteasome-dependent degradation of Smad2 in transforming growth factor-beta signaling.* The Journal of biological chemistry, 2000. **275**(47): p. 36818-22.

106. Kavsak, P., et al., *Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation.* Molecular cell, 2000. **6**(6): p. 1365-75.

107. Moustakas, A., S. Souchelnytskyi, and C.H. Heldin, *Smad regulation in TGF-beta signal transduction.* Journal of cell science, 2001. **114**(Pt 24): p. 4359-69.

108. Suzuki, C., et al., *Smurf1 regulates the inhibitory activity of Smad7 by targeting Smad7 to the plasma membrane.* The Journal of biological chemistry, 2002. **277**(42): p. 39919-25.

109. Kalluri, R. and E.G. Neilson, *Epithelial-mesenchymal transition and its implications for fibrosis.* The Journal of clinical investigation, 2003. **112**(12): p. 1776-84.

110. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* The Journal of clinical investigation, 2009. **119**(6): p. 1420-8.

111. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.

112. Huang, R.Y., P. Guilford, and J.P. Thiery, *Early events in cell adhesion and polarity during epithelial-mesenchymal transition.* Journal of cell science, 2012. **125**(Pt 19): p. 4417-22.

113. Yilmaz, M. and G. Christofori, *EMT, the cytoskeleton, and cancer cell invasion.* Cancer metastasis reviews, 2009. **28**(1-2): p. 15-33.

114. Wheelock, M.J., et al., *Cadherin switching.* Journal of cell science, 2008. **121**(Pt 6): p. 727-35.

115. Hansen, S.M., V. Berezin, and E. Bock, *Signaling mechanisms of neurite outgrowth induced by the cell adhesion molecules NCAM and N-cadherin.* Cellular and molecular life sciences : CMLS, 2008. **65**(23): p. 3809-21.

116. Peinado, H., D. Olmeda, and A. Cano, *Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype?* Nature reviews. Cancer, 2007. **7**(6): p. 415-28.

117. Barrallo-Gimeno, A. and M.A. Nieto, *The Snail genes as inducers of cell movement and survival: implications in development and cancer.* Development, 2005. **132**(14): p. 3151-61.

118. Vincent, T., et al., *A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition.* Nature cell biology, 2009. **11**(8): p. 943-50.

119. Yang, M.H., et al., *Direct regulation of TWIST by HIF-1alpha promotes metastasis.* Nature cell biology, 2008. **10**(3): p. 295-305.

120. Yook, J.I., et al., *A Wnt-Axin2-GSK3beta cascade regulates Snail1 activity in breast cancer cells.* Nature cell biology, 2006. **8**(12): p. 1398-406.

121. Xu, J., S. Lamouille, and R. Derynck, *TGF-beta-induced epithelial to mesenchymal transition.* Cell research, 2009. **19**(2): p. 156-72.

122. Polyak, K., *Heterogeneity in breast cancer.* The Journal of clinical investigation, 2011. **121**(10): p. 3786-8.

123. Clarke, M.F., et al., *Cancer stem cells--perspectives on current status and future directions: AACR Workshop on cancer stem cells.* Cancer research, 2006. **66**(19): p. 9339-44.

124. Nguyen, L.V., et al., *Cancer stem cells: an evolving concept.* Nature reviews. Cancer, 2012. **12**(2): p. 133-43.

125. Kreso, A. and J.E. Dick, *Evolution of the cancer stem cell model.* Cell stem cell, 2014. **14**(3): p. 275-91.

126. Al-Hajj, M., et al., *Prospective identification of tumorigenic breast cancer cells.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(7): p. 3983-8.

127. Mylona, E., et al., *The clinicopathologic and prognostic significance of CD44+/CD24(-/low) and CD44-/CD24+ tumor cells in invasive breast carcinomas.* Human pathology, 2008. **39**(7): p. 1096-102.

128. Shipitsin, M., et al., *Molecular definition of breast tumor heterogeneity.* Cancer cell, 2007. **11**(3): p. 259-73.

129. Chute, J.P., et al., *Inhibition of aldehyde dehydrogenase and retinoid signaling induces the expansion of human hematopoietic stem cells.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(31): p. 11707-12.

130. Ginestier, C., et al., *ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome.* Cell stem cell, 2007. **1**(5): p. 555-67.

131. Croker, A.K. and A.L. Allan, *Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44(+) human breast cancer cells.* Breast cancer research and treatment, 2012. **133**(1): p. 75-87.

132. Park, S.M., et al., *The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2.* Genes & development, 2008. **22**(7): p. 894-907.

133. Korpal, M., et al., *The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2.* The Journal of biological chemistry, 2008. **283**(22): p. 14910-4.

134. Brabletz, S. and T. Brabletz, *The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer?* EMBO reports, 2010. **11**(9): p. 670-7.

135. Dykxhoorn, D.M., et al., *miR-200 enhances mouse breast cancer cell colonization to form distant metastases.* PloS one, 2009. **4**(9): p. e7181.

136. Korpal, M., et al., *Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization.* Nature medicine, 2011. **17**(9): p. 1101-8.

137. Ma, L., et al., *miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis.* Nature cell biology, 2010. **12**(3): p. 247-56.

138. Kong, W., et al., *MicroRNA-155 is regulated by the transforming growth factor beta/Smad pathway and contributes to epithelial cell plasticity by targeting RhoA.* Molecular and cellular biology, 2008. **28**(22): p. 6773-84.

139. Stinson, S., et al., *TRPS1 targeting by miR-221/222 promotes the epithelial-to-mesenchymal transition in breast cancer.* Science signaling, 2011. **4**(177): p. ra41.

140. Ji, Q., et al., *MicroRNA miR-34 inhibits human pancreatic cancer tumor-initiating cells.* PloS one, 2009. **4**(8): p. e6816.

141. Liu, C., et al., *The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44.* Nature medicine, 2011. **17**(2): p. 211-5.

142. Ji, Q., et al., *Restoration of tumor suppressor miR-34 inhibits human p53-mutant gastric cancer tumorspheres.* BMC cancer, 2008. **8**: p. 266.

143. Christoffersen, N.R., et al., *p53-independent upregulation of miR-34a during oncogene-induced senescence represses MYC.* Cell death and differentiation, 2010. **17**(2): p. 236-45.

144. Cheng, C.Y., et al., *miR-34 cooperates with p53 in suppression of prostate cancer by joint regulation of stem cell compartment.* Cell reports, 2014. **6**(6): p. 1000-7.

145. He, L., et al., *A microRNA component of the p53 tumour suppressor network.* Nature, 2007. **447**(7148): p. 1130-4.

146. Yu, F., et al., *let-7 regulates self renewal and tumorigenicity of breast cancer cells.* Cell, 2007. **131**(6): p. 1109-23.

147. Shimono, Y., et al., *Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells.* Cell, 2009. **138**(3): p. 592-603.

148. Xing, Z., et al., *lncRNA directs cooperative epigenetic regulation downstream of chemokine signals.* Cell, 2014. **159**(5): p. 1110-25.

149. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.

150. Bertani, S., et al., *The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin.* Molecular cell, 2011. **43**(6): p. 1040-6.

151. Santini, S., J.L. Boore, and A. Meyer, *Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters.* Genome research, 2003. **13**(6A): p. 1111-22.

152. Mallo, M., D.M. Wellik, and J. Deschamps, *Hox genes and regional patterning of the vertebrate body plan.* Developmental biology, 2010. **344**(1): p. 7-15.

153. Mallo, M. and C.R. Alonso, *The regulation of Hox gene expression during animal development.* Development, 2013. **140**(19): p. 3951-63.

154. Yang, F., et al., *Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis.* Molecular cell, 2013. **49**(6): p. 1083-96.

155. Hacisuleyman, E., et al., *Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre.* Nature structural & molecular biology, 2014. **21**(2): p. 198-206.

156. Tripathi, V., et al., *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation.* Molecular cell, 2010. **39**(6): p. 925-38.

157. Onder, T.T., et al., *Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways.* Cancer research, 2008. **68**(10): p. 3645-54.

158. Perl, A.K., et al., *A causal role for E-cadherin in the transition from adenoma to carcinoma.* Nature, 1998. **392**(6672): p. 190-3.

159. Vleminckx, K., et al., *Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role.* Cell, 1991. **66**(1): p. 107-19.

160. Frixen, U.H., et al., *E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells.* The Journal of cell biology, 1991. **113**(1): p. 173-85.

161. Shultz, L.D., et al., *Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice.* Journal of immunology, 1995. **154**(1): p. 180-91.

162. Shultz, L.D., et al., *Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells.* Journal of immunology, 2005. **174**(10): p. 6477-89.

163. Iorns, E., et al., *A new mouse model for the study of human breast cancer metastasis.* PloS one, 2012. **7**(10): p. e47995.

164. Gu, Z.D., et al., *HOXA13 promotes cancer cell growth and predicts poor survival of patients with esophageal squamous cell carcinoma.* Cancer research, 2009. **69**(12): p. 4969-73.

165. Quagliata, L., et al., *Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients.* Hepatology, 2014. **59**(3): p. 911-23.

166. Kang, Y., et al., *Breast cancer bone metastasis mediated by the Smad tumor suppressor pathway.* Proc Natl Acad Sci U S A, 2005. **102**(39): p. 13909-14.

167. Nam, J.S., et al., *Bone sialoprotein mediates the tumor cell-targeted prometastatic activity of transforming growth factor beta in a mouse model of breast cancer.* Cancer Res, 2006. **66**(12): p. 6327-35.

168. Roberts, A.B. and L.M. Wakefield, *The two faces of transforming growth factor beta in carcinogenesis.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 8621-3.

169. Xie, W., et al., *Alterations of Smad signaling in human breast carcinoma are associated with poor outcome: a tissue microarray study.* Cancer Res, 2002. **62**(2): p. 497-505.

170. Chau, Y.M., S. Pando, and H.S. Taylor, *HOXA11 silencing and endogenous HOXA11 antisense ribonucleic acid in the uterine endometrium.* The Journal of clinical endocrinology and metabolism, 2002. **87**(6): p. 2674-80.

171. Manzanares, M., et al., *Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head.* Nature, 2000. **408**(6814): p. 854-7.

172. Yang, H., et al., *MicroRNA expression profiling in human ovarian cancer: miR-214 induces cell survival and cisplatin resistance by targeting PTEN.* Cancer research, 2008. **68**(2): p. 425-33.

173. Filimonenko, M., et al., *The selective macroautophagic degradation of aggregated proteins requires the PI3P-binding protein Alfy.* Molecular cell, 2010. **38**(2): p. 265-79.

174. Kiyono, K., et al., *Autophagy is activated by TGF-beta and potentiates TGF-beta-mediated growth inhibition in human hepatocellular carcinoma cells.* Cancer research, 2009. **69**(23): p. 8844-52.

175. Thomas, S., et al., *Preferential killing of triple-negative breast cancer cells in vitro and in vivo when pharmacological aggravators of endoplasmic reticulum stress are combined with autophagy inhibitors.* Cancer letters, 2012. **325**(1): p. 63-71.

176. Yu, H., et al., *Revisiting STAT3 signalling in cancer: new and unexpected biological functions.* Nature reviews. Cancer, 2014. **14**(11): p. 736-46.

177. Fukada, T., et al., *STAT3 orchestrates contradictory signals in cytokine-induced G1 to S cell-cycle transition.* The EMBO journal, 1998. **17**(22): p. 6670-7.

178. Maycotte, P., et al., *STAT3-Mediated Autophagy Dependence Identifies Subtypes of Breast Cancer Where Autophagy Inhibition Can Be Efficacious.* Cancer research, 2014. **74**(9): p. 2579-90.

179. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Molecular cell, 2011. **43**(6): p. 904-14.

180. Yelin, R., et al., *Widespread occurrence of antisense transcription in the human genome.* Nature biotechnology, 2003. **21**(4): p. 379-86.

181. Yu, W., et al., *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.* Nature, 2008. **451**(7175): p. 202-6.

182. Takayama, K., et al., *Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer.* The EMBO journal, 2013. **32**(12): p. 1665-80.

183. Katayama, S., et al., *Antisense transcription in the mammalian transcriptome.* Science, 2005. **309**(5740): p. 1564-6.

184. Li, J., et al., *Autophagy promotes hepatocellular carcinoma cell invasion through activation of epithelial-mesenchymal transition.* Carcinogenesis, 2013. **34**(6): p. 1343-51.

185. Singha, P.K., et al., *Manumycin A inhibits triple-negative breast cancer growth through LC3-mediated cytoplasmic vacuolation death.* Cell death & disease, 2013. **4**: p. e457.

186. Yu, H., D. Pardoll, and R. Jove, *STATs in cancer inflammation and immunity: a leading role for STAT3.* Nature reviews. Cancer, 2009. **9**(11): p. 798-809.

187. Grivennikov, S., et al., *IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer.* Cancer cell, 2009. **15**(2): p. 103-13.

188. Gao, S.P., et al., *Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas.* The Journal of clinical investigation, 2007. **117**(12): p. 3846-56.

189. Xiong, H., et al., *Roles of STAT3 and ZEB1 proteins in E-cadherin down-regulation and human colorectal cancer epithelial-mesenchymal transition.* The Journal of biological chemistry, 2012. **287**(8): p. 5819-32.

190. Rokavec, M., et al., *IL-6R/STAT3/miR-34a feedback loop promotes EMT-mediated colorectal cancer invasion and metastasis.* The Journal of clinical investigation, 2014. **124**(4): p. 1853-67.

191. Deng, X.S., et al., *Metformin targets Stat3 to inhibit cell growth and induce apoptosis in triple-negative breast cancers.* Cell cycle, 2012. **11**(2): p. 367-76.

192. Fukuda, A., et al., *Heterogeneous nuclear ribonucleoprotein R cooperates with mediator to facilitate transcription reinitiation on the c-Fos gene.* PloS one, 2013. **8**(8): p. e72496.

193. Masuda, K., et al., *NF90 in posttranscriptional gene regulation and microRNA biogenesis.* International journal of molecular sciences, 2013. **14**(8): p. 17111-21.

194. Pei, Y., et al., *Nuclear export of NF90 to stabilize IL-2 mRNA is mediated by AKT-dependent phosphorylation at Ser647 in response to CD28 costimulation.* Journal of immunology, 2008. **180**(1): p. 222-9.

195. Hung, C.L., et al., *A long noncoding RNA connects c-Myc to tumor metabolism.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(52): p. 18697-702.

196. Chen, C.R., Y. Kang, and J. Massague, *Defective repression of c-myc in breast cancer cells: A loss at the core of the transforming growth factor beta growth arrest program.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(3): p. 992-9.

197. Kuwano, Y., et al., *NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif.* Nucleic acids research, 2010. **38**(1): p. 225-38.

198. Shim, J., et al., *Nuclear export of NF90 is required for interleukin-2 mRNA stabilization.* Molecular cell, 2002. **10**(6): p. 1331-44.

199. Gruber, A.R., et al., *AREsite: a database for the comprehensive investigation of AU-rich elements.* Nucleic acids research, 2011. **39**(Database issue): p. D66-9.

200. Lamouille, S. and R. Derynck, *Cell size and invasion in TGF-beta-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway.* The Journal of cell biology, 2007. **178**(3): p. 437-51.

APPENDIX