Summer 2024

Implications of 5-methylcytosine (m5C) and N4-actylcytidine (ac4C) in epithelial to mesenchymal transition (EMT) in human breast cancer

Manuel A. Pazos II
mpazos@albany.edu

Implications of 5-methylcytosine (m5C) and N4-actylcytidine (ac4C) in epithelial to mesenchymal transition (EMT) in human breast cancer

by

Manuel Antonio Pazos II

A Thesis

Submitted to the University at Albany, State University of New York

In Partial Fulfillment of

the Requirements for the Degree of

Master of Science

School of Public Health

Department of Biomedical Sciences

Summer 2024

ABSTRACT

Breast cancer is a complex disease that continues to haunt women and their loved ones around the world. It is currently the leading cause of cancer-related mortality among women. While most breast cancer patients can catch the disease before progression, others are less fortunate. The prognosis of those whose disease has advanced to metastatic or invasive breast cancer (IBC) is grim. Metastasis occurs when cancer cells undergo epithelial to mesenchymal transition (EMT), migrate to distant body tissues, then form new tumors. It is imperative to better understand the process of EMT to develop superior ways to detect breast cancer before becoming metastatic. Recently, post-transcriptional modifications have been implicated in the regulation of EMT in different cancer types. Previous studies lead us to believe that N4-actylcytidine may be involved in the process of EMT in human breast cancer. N4-actylcytidine (ac4C) is produced by the enzyme N-acetyltransferase 10 (NAT10). To establish an understanding of ac4C's role, we conducted several experiments while altering the expression of NAT10 in Human Mammary Epithelial (HMLE) cells. In this study, we show that the overexpression of NAT10 in HMLE cells induces certain characteristics of the mesenchymal phenotype. We also confirm a proposed method of ac4C-seq by confirming one N4-acetylcytidine site (C-1842) on the 18S ribosome of HMLE cells. These results help us to better understand the role of ac4C in the process of EMT in human breast cancer which may help to pave the way for better preventative interventions and treatments for patients with metastatic breast cancer.

INTRODUCTION

Overview and rationale

Breast cancer is the leading cause of cancer-related mortality among women (DeSantis et al., 2015). Survival rates are dependent on many factors, including time of detection, subtype, morphology, etc. Regardless of the subtype, effective treatments for patients whose disease has been detected early and remains localized within the mammary glands/ducts involve surgical interventions, such as mastectomy (removal of breast) or breast-conserving surgery (removal of tumor only). For instance, standard treatment for early-stage breast cancers, such as ductal carcinoma in situ (DCIS), involves breast-conserving surgery followed by radiation therapy (Maughan et al., 2010). Contrarily, if the disease has advanced to later stages and/or has become metastatic, treatment methods become limited to chemotherapy and radiation therapy (Moo et al., 2018). Both of which can be debilitating to the patient.

Prior to becoming metastatic, or malignant, tumorous tissue is made up of cells that exert the epithelial phenotype. Epithelial cells are characterized by having cell-cell interactions that are held together by extracellular components, such as E-cadherin and B-catenin, at intercellular junctions (Gennari, 2011). In living systems, epithelial cells compose the linings of organs and other structures (Blanpain et al., 2007). In contrast, malignant tumors contain cells that display the mesenchymal phenotype. These cells are not obligated to maintain cell-cell interactions; therefore, they are able to freely travel throughout the body (Kalluri et al., 2009).

In a process called epithelial to mesenchymal transition (EMT), cells within cancerous tumors branch from the original tumor site and spread to another site within the body. Cells that have undergone EMT (EMT-positive cells) display the mesenchymal phenotype which gives them

the ability to travel throughout the body via blood and lymphatic vessels. While EMT is associated with cancerous metastasis, it plays a crucial role in normal physiology.

There are three distinct types of EMT. The first, type I, occurs during early embryonic development. During this process, some cell types must acquire migratory properties (a hallmark of the mesenchymal phenotype) to travel to different areas of the embryo (Potenta et al., 2008). Another important role of EMT, type II, is the process of wound healing. This process involves three steps: inflammation, proliferation, and maturation. During these processes, surrounding tissue is preserved by phagocytosis of infected cells, angiogenesis occurs to rebuild lost blood vessels within the wound, and new extra cellular matrix (ECM) is deposited. Terminating this process is the re-epithelialization of this new tissue (Barriere et al., 2015). The process of EMT is sometimes referred to as a "double-edged sword." While types I and II fulfill imperative biological functions, this process can go awry.

The occurrence of type III EMT yields the other end of the figurative "double-edged sword" – the production of malignant cancer cells. The process of cancerous tissue undergoing EMT and becoming malignant is known as metastasis. Migratory EMT-positive cells circulate in the body until they reach distant tissue and revert to their epithelial phenotype, by the reverse process mesenchymal to epithelial transition (MET), to form tumors in distant body tissues. This process is also responsible for relapse (Luo et al., 2015). It is important to understand the role that EMT plays in cancer as 90 % of cancer-related mortality results from metastasis (Chaffer et al., 2011). Currently, it is unknown what drives cancer cells to undergo EMT and become metastatic; however, there is a growing body of evidence that suggests that post-transcriptional modifications are involved.

Increasing attention has been given to post-transcriptional modifications due to their potential implications in the progression of cancer. N6-methyladenosine (m6A), for example, is known to be the most abundant in eukaryotes (Machnicka et al, 2013) and has been shown to be upregulated in several cancer types. It also has been implicated in EMT-like characteristics such as invasiveness and increased proliferation (Dong and Cui, 2020).

Understanding that m6A and other post-transcriptional modifications play essential roles in EMT in breast cancer, we are interested in studying the significance of others. For instance, little is known about the roles of 5-methylcytosine (m5C) and N4-acetylcytidine (ac4C) in human breast cancer. N4-acetylcytidine (ac4C) has been shown to increase the efficiency of translation and drive certain hallmarks of epithelial to mesenchymal transition (EMT) in HeLa cells (Arango et al., 2018). N4-acetylcytidine is produced by the N-acetyltransferase 10 (NAT10) (Ito et al., 2014). By knocking-out this gene, researchers were able to reverse phenotypes associated with EMT in HeLa cells (cervical cancer) (Arango et al., 2018). The experiments conducted in this thesis help us move closer to understanding the roles that the post-transcriptional modifications m5C and ac4C play in EMT in human breast cancer and aim to uncover the potential link between the two.

Several experiments, including immunoprecipitations, were performed in EMT model cell lines to get a better understanding of the role of m5C in EMT. To better understand the role of ac4C in EMT, we perform NAT10 knockdowns and develop a NAT10 overexpression cell line using Human Mammary Epithelial (HMLE) cells. Using this cell line, we conduct a series of experiments whose outcomes we expect to observe in cells that have undergone EMT. Recognizing the significance of post-transcriptional modifications, such as m5C and ac4C, can allow us to better understand their roles in the process of epithelial to mesenchymal transition in

breast cancer, paving the way for the development of targeted treatments for patients whose disease has progressed to metastatic breast cancer. Ideally, we will have the capability to prevent metastasis.

Forms of breast cancer

Breast cancer is a complex disease and the most frequently diagnosed cancer type among women in almost every region of the world (Ferlay et al., 2019). According to the American Cancer Society, 12.8 % (approximately 1 in 8) of women will develop breast cancer in their lifetime and 3 % will die from it. It is estimated that 20 – 30 % of localized breast cancers reoccur and eventually become metastatic. It is estimated that up to 90 % of cancer-related mortality is caused by metastatic cancers (Chaffer et al., 2011). Unmistakably, metastatic breast cancer can be devastating.

There are several subtypes of breast cancer HER2+, luminal A, luminal B, and triple negative breast cancer (TNBC). The latter of which is the most difficult to treat and has the worse prognosis of all. To define these subtypes, it is important to understand that the terms positive and negative are relative to normal, and therefore, refer to overexpression and decreased expression respectively.

Patients whose cancer displays the HER2+ subtype exhibit an increase in the human epidermal growth factor receptor 2 (HER2). It is estimated that 20 – 30 % of patients' tumors are HER2+. Therapies for this subtype include blocking these receptors which prevents the cells from obtaining sufficient epidermal growth factor (EGF) signal. A widely used example of a HER2 blocker is the monoclonal antibody trastuzumab (Mitri et al., 2012). This essentially starves the cancer cells as they require more resources to thrive.

Tumors of the luminal A subtype are estrogen receptor (ER) and progesterone receptor (PR) positive. Patients with this subtype respond very well to hormone therapies and have the best prognosis. The luminal A and B subtypes are both estrogen receptor (ER) positive (Gao and Swain, 2018).

Finally, patients whose cancer displays the TNBC subtype lack HER2, ER, and PR which renders hormone or HER2-targeted therapy ineffective (Anders and Carey, 2008). Consequentially, this subtype is the most difficult to treat and has the worse prognosis of all the subtypes. Treatment for this subtype is limited to radiation therapy and chemotherapy (Moo et al., 2018). Due to the low survival rate of women whose cancer has become metastatic; we are interested in elucidating the mechanisms that drive epithelial to mesenchymal transition (EMT) in breast cancer.

Epithelial to Mesenchymal Transition (EMT)

In order to understand EMT, it is necessary to define and understand the differences between epithelial and mesenchymal cells. Epithelial cells are polarized cells that can communicate intercellularly through adherens junctions, tight junctions, and desmosomes (Blanpain et al., 2007). This allows the cells to function as a sheet forming epithelium. Contrarily, mesenchymal cells are those that have lost their cell polarity and ability to form intercellular connections. As a result, these cells display an increased capacity for migration and invasiveness. Mesenchymal cells also display an elevated resistance to apoptosis and increase in the production of extra cellular matrix (ECM) components (Kalluri et al., 2009). Hence, epithelial to mesenchymal transition (EMT) is the process by which epithelial cells undergo biochemical changes that drive cells to exert the mesenchymal phenotype.

EMT is imperative for embryonic development. During development, several rounds of EMT and the reverse process, mesenchymal to epithelial transition (MET), occur before final cell differentiation. These rounds (primary, secondary, and tertiary EMT) are responsible for the creation of complex cellular structures such as internal organs (Thiery et al., 2009). In order to create some of these complex structures in distant parts of the embryo, it is necessary for some cells to acquire migratory properties- a hallmark of the mesenchymal phenotype.

Another important role of EMT is the process of wound healing. This process is a three-step process: inflammation, proliferation, and maturation. During these processes, tissue is preserved by phagocytosis of infected cells, angiogenesis occurs to rebuild lost blood vessels, deposition of new extra cellular matrix (ECM), then re-epithelialization (Barriere et al., 2015).

While EMT plays crucial roles in embryonic development and other important biological processes, it also plays a role in diseases such as cancer. When cancer cells undergo EMT, the cancer becomes metastatic. It is important to understand the role that EMT plays in cancer as 90 % of cancer-related mortality is the result of metastasis (Chaffer et al., 2011). Currently, it is unknown what drives cancer cells to undergo EMT and become metastatic; however, there is a growing body of evidence that suggests that RNA modifications are involved.

Post-Transcriptional Modifications

RNA modifications are chemical changes in ribonucleic acid molecules that occur post-transcriptionally. Additions include, but are not limited to methylation, acetylation, and the addition of amino acids. One example of a post-transcriptional modification is m6A. This modification emerges from a methylation at the N6 position of adenosine. N6-methyladenosine (m6A) is the most abundant RNA modification in humans (Zhao et al., 2017). Since the discovery of the first modified nucleoside – pseudouridine (Ψ) – in 1957 (Grosjean, 2015), many other RNA

modifications have been discovered and implicated in various diseases. Currently, over 170 post-transcriptional modifications have been identified (Dong and Cui, 2020).

Mutations in nearly half of the enzymes responsible for producing post-transcriptional modifications have been associated with human diseases (Jonkhout et al., 2017). This implies that RNA modifications play a significant role in many diseases. For instance, a mutation in the methyltransferase NSun2 gene, which produces 5-methylcytosine (m5C), is associated with microcephaly in mice and humans. When NSun2 is knocked out in mice, neurological abnormalities occur as a result of the lack of m5C in tRNA (Blanco et al., 2014).

RNA Modifications in Cancer and EMT

N6-methyladenosine (m6A) and 5-methylcytosine (m5C) have been implicated in tumorigenesis and increase in cancer cell proliferation. For example, m6A and m5C expression have been implicated in the progression and aggressiveness of glioblastoma (Dong and Cui, 2020). N6-methyladenosine (m6A) has also been found in other types of cancer, such as colorectal cancer. Overexpression of the METTL3; the methyltransferase responsible for producing m6A, has been found to be negatively correlated with survival rates in patients with colorectal cancer (Peng et al., 2019). While m6A and m5C are well known to play a role in several cancer types, we suspect that other post-transcriptional modifications are involved.

Previous mass spectrometric data from our lab shows an abundance of N4-acetylcytosine (ac4C) in a mesenchymal cancer cell line when compared to an epithelial cell line. Compared to MDA-MB-231 cell line; derived from a 51-year-old Caucasian female, we see an increase in the relative abundance of ac4C in the SUM159 cell line; a triple negative breast cancer cell line that displays the mesenchymal phenotype. This data leads us to believe that ac4C is implicated in epithelial to mesenchymal transition. Since we are particularly interested in EMT in breast cancer,

we look closer at this modification. To study the role of ac4C in EMT, we used Human Mammary Epithelial (HMLE) cell lines; a model for cells that have undergone EMT.

HMLE cell line

Experiments were conducted using the Human Mammary Epithelial (HMLE) cell line. The HMLE cell lines that were used had been transfected with the pWZL plasmid harboring GFP along with one of the following EMT-inducing genes: Snai1, Twist1, and TGF-β1. The emerging cell lines: HMLE-pWZL-Snail, HMLE-pWZL-Twist, HMLE-pWZL-TGF-β display overexpression of their respective EMT-inducing gene resulting in EMT-positive model (mesenchymal) cell lines. All HMLE cell lines were cultured as described in Mani et al., 2008.

According to the information displayed in **table 1**, Snai1 (Snail) is a transcription factor that directly downregulates the CDH1 (E-cadherin) (M.A. Nieto, 2002), Twist1 (Twist) is a master regulator of morphogenesis associated with activation of mesenchymal markers (Yang et al., 2004), and TGF-β1 (TGF-β) activates Snai1, Twist1, and other EMT-promoting transcription factors such as Snai2 (Slug), Zeb1, Zeb2, ect (Hao et al., 2019). E-cadherin is a key component of adherens junctions, which are crucial for cell-cell adhesion in epithelial tissues. Loss of E-cadherin results in the detachment of individual cell from the epithelial layer (Hugo et al., 2011).

Overexpression of Snai1, Twist1, and TGF-β1 causes HMLE cells to display the mesenchymal phenotype. Compared to the control HMLE cell line, pWZL-GFP, the EMT-inducing gene overexpression cell lines show decreased levels of E-cadherin, and increased expression of vimentin (Mani et al., 2008), which are markers of EMT-positive cells. Since overexpression of these genes has been confirmed to transform the cells from the epithelial to the mesenchymal phenotype, we used these cell lines as models for cells that have undergone EMT.

METHODS:

Generation of NAT10-overexpressing cell line (HMLE-pLenti-NAT10)

HMLE cells were transduced with pLenti-NAT10 to generate a stable overexpression cell line (**Figure 4**) that we designated "HMLE-pLenti-NAT10." To generate this cell line, human embryonic kidney cells (HEK293T) were transfected with pLenti-NAT10 (2 µg), pDelta8.2 (2 µg), and pVSVG (200 ng) using X-tremeGENE™HP from Millipore Sigma (Burlington, MA) in a 3:1 ratio according to manufacturer instructions. Following overnight incubation, the transfected HEK293T cells' media - Dulbecco's Modified Eagle Medium (DMEM; ThermoFisher Scientific) supplemented with 10 % FBS (Fetal bovine serum) - was replaced with that of its target cell line - Human Mammary Epithelial Cell Basal Medium supplemented with Mammary Epithelial Growth Supplement (Gibco™) - "HMLE media." Virus-infected HEK293T cells were allowed to populate the replaced media with viral particles for 24 hours. Virus-containing HMLE media was then removed and filtered through a 0.45 µm syringe filter. To the filtered media, polybrene (1:100; ThermoFisher Scientific) was added to enhance efficiency of transduction. The target cell line (HMLE) was then incubated with virus-containing media overnight. Transduced cells were selected via incubation with 1 µg/mL puromycin yielding HMLE-pLenti-NAT10. This protocol was performed using the empty pLenti vector to serve as our control to produce HMLE-pLenti.

Cell growth, harvest, and RNA extraction

Cell lines were grown in 15 cm dishes with HMLE media and incubated at 37 °C and 5 % CO_2 until plates reached approximately 80 % confluency. When ready for harvest, cells were washed twice with PBS then collected from the plates using a cell scraper. Cells were resuspended in cold PBS, then spun down into cell pellets. Pellets were then flash frozen by ethanol-dry ice bath and then stored at -80 °C until later use.

Total RNA was extracted from the flash-frozen cell pellets using Omega Bio-Tek's "E.Z.N.A.® Total RNA Kit I." Since we are only interested in analyzing the transcripts, and to maximize our immunoprecipitation efficiency, ribosomal RNA (rRNA) was depleted from the samples using Invitrogen's "RiboMinus™ Human/Mouse Transcriptome Isolation Kit." The rRNA-depleted RNA was then concentrated using the RiboMinus™ Concentration Module."

m5C RNA Immunoprecipitation (m5C-RIP)

To each sample of rRNA-depleted RNA, 5 µg of anti-m5C [5-methylcytosine (5-mC) Antibody, Clone 33D3 – Diagenode] monoclonal antibody along with an RNase inhibitor were added to m5C-RIP buffer. This solution was incubated overnight at 4 °C. As a control, the same amount of rRNA-depleted RNA from the same cell lines was incubated with 5 µg IgG monoclonal antibody under the same conditions. After incubation of RNA and antibody, 30 µl of prewashed Thermo Scientific™ Pierce™ Protein G Magnetic Beads were added to each sample. Once the beads were added, the samples were allowed to rotate at 4 °C for 2 hours to capture the m5C-bound antibodies.

To elute the methylated RNA sample from the beads, 500 µl of Qiazol (from Qiagen's RNeasy Plus Universal Mini Kit) were added to each sample m5C-selected RNA-bound beads. After incubating this mixture at room temperature for 10 minutes, 100 µl of chloroform were added to separate unwanted products from the desired water-soluble products. The mixture was centrifuged for 15 minutes at 4 °C. After centrifugation, the aqueous layer was removed and transferred to another tube. As per the RNeasy Plus Universal Mini Kit (Qiagen) protocol, 1.5 volumes of ethanol were added to the RNA solution and RNA isolation was performed following steps 6 – 14 of the protocol. The eluent was passed twice through the membrane to maximize yield. The final volume of each sample was 30 µl.

Sample preparation for qPCR analysis

From all samples, including our controls, 10 µl were removed for reverse transcription (RT). The RT reaction was carried out using the MultiScribe Reverse Transcription Kit (Invitrogen™) which consists of 10X buffer, random hexamers, dNTPs, and reverse transcriptase. All components were added to each 10 µl sample for a final volume of 20 µl, then placed into a thermocycler to complete the reaction. Upon completion of RT, samples were diluted accordingly using RNase-free water. Diluted RT products were stored at 4 °C until ready for analysis. Samples were prepared for qPCR analysis using the "PowerUp™ SYBR® Green Master Mix," mixed with the gene-specific primer. Differences in methylated transcripts between HMLE pWZL-GFP and -Snail cell lines were noted.

RNA immunoprecipitation (RIP) buffer

To prepare the RIP buffers for m5C and ac4C containing RNA, a solution containing 10 mM Tris-HCl, 150 mM NaCl, and 0.1 % Ipegal, was brought to a pH of 7.4. To create the buffer, 2 ml of 1M Tris-HCl was added to 30 ml of 1 M NaCl. This solution was adjusted to a pH of 7.4. To the pH 7.4 solution, 20 ml of 1 % Ipegal were added along with enough water to reach 200 ml. The complete buffer was then sterile filtered and stored at 4 °C.

qPCR analysis

2 µg of total RNA (from all cell lines used) were reverse transcribed. Samples were diluted to 50 ng/µl. 5 µl of each cDNA sample were added to a 384-well plate. To each sample, 5 µl of PowerUp™ SYBR® Green Master Mix green power up master mix and 1 µl of primer mixture (5 µM forward primer and 5 µM reverse primer) (**Table S2**) were added. Plates were sealed, briefly centrifuged, then read using QuantStudio 12 K Flex (Thermo Fisher Scientific).

CTG proliferation assay

To determine the difference in proliferation rates between the HMLE pLenti and HMLE pLenti-Nat10 cell lines, a CellTiter-Glo® (CTG) Luminescent Cell Viability Assay was performed. The assay uses a reagent which reacts with ATP to produce luminescence. The magnitude of luminescence is directly proportional to the amount of live cells. To perform this assay, two thousand cells per well were plated into 5 wells of 96-well plates. Cells were allowed to incubate at 37 °C and plates were set up to be analyzed at the following time-points: 24 hours, 48 hours, 72 hours, and 96 hours. For each time-point, plates were removed from incubator and allowed to cool to room temperature for 15 minutes. Once cool, 50 µl of the CTG reagent were to each well to be read. The plates were then allowed to incubate at room temperature for 10 minutes in darkness. After incubation, luminescence was analyzed using a plate reader.

Mammosphere assay

To visualize the mesenchymal phenotype caused by the overexpression of Nat10, a mammosphere assay was performed using the HMLE pLenti-NAT10 cell line and comparing it to the HMLE pLenti (control) cell line. For each cell line, ten thousand cells per well were plated into three wells of a six-well plate. Cells were plated in media containing: 20 ng/ml Fibroblast Growth Factor – Basic (bFGF), 10 ng/ml Epidermal Growth Factor (EGF), 4 µg/ml heparin sulfate, and 0.5 % methylcellulose in MEGM media (LONZA, product # CC-3150). Cells were allowed to incubate at 37 °C for 10 days. After ten days, plates were removed from incubator and pictures were taken using EVOS FL Auto Imaging System (Thermo Fisher Scientific). Mammospheres were individually counted, and the average number of mammospheres for all three wells of each cell line were recorded.

Future directions: N4-acetylcytosine sequencing (ac4C-seq)

Total RNA from HMLE pWZL-Twist was used as a template for human 18S rRNA. To confirm the acetylation sites, total RNA was treated with 100 mM sodium cyanoborohydride (Sigma-Aldrich) (with 100 mM HCl) to reduce N4-acetylcytidine (ac4C) to N4-acetyl-tetrahydrocytidine. Reduction of ac4C leads to a nucleotide misincorporation (from C to T) upon reverse transcription (RT) of the chemically manipulated RNA. After treatment with sodium cyanoborohydride, primer-specific reverse transcription was performed using primers (**Table S1**) that were designed to capture the two expected ac4C sites on the 18S ribosomal subunit – C-1337 and C-1842 (Sharma et al., 2015). The cDNA obtained from RT was then amplified by PCR. The amplicons associated with each ac4C site, located at positions C-1337 and C-1842 of the human 18S rRNA subunit, were predicted to be 175 and 214 base pairs respectively. The size of each amplicon was confirmed by gel electrophoresis. Bands were excised from gel and gel purified. To determine the presence or absence of acetylation at C-1337 and C-1842, samples were prepared for sanger sequencing.

Once we confirmed the validity of this method for differentiating between canonical cytidine (C) and N4-acetylcytidine (ac4C), we performed the reduction method introduced by Sas-Chen et al. (2020). Sodium cyanoborohydride ($NaCNBH_3$) treatments were performed on all poly-A-selected RNA obtained from all three of our EMT-positive HMLE model cell lines. Treated and untreated RNA samples were stored at -80 °C to be sequenced in the future.

5-METHYLCYTOSINE (m5C) EXPERIMENTS

According to bisulfite conversion data, displayed in **figure S1**, the following transcripts: GNB4, EIF2AK2, PRR11, EEF1Δ, and the lncRNA, OPI5-AS1 show an increase in methylation in all three EMT model cell lines when compared to the control cell line. The greatest difference in methylation was noted in the HMLE pWZL-Snail cell line. Consequently, this cell line was used to perform 5-methylcytosine-specific RNA Immunoprecipitation (m5C-RIP) using Diagenode's anti-mouse "5-methylcytosine (5-mC) Antibody – Clone 33D3."

We confirm hypermethylation in GNB4, EIF2AK2, PRR11, EEF1Δ, and the lncRNA, OPI5-AS1 by immunoprecipitation. Immunoprecipitation was performed on rRNA-reduced HMLE pWZL-Snail RNA samples described in the "methods" section. As presented in **figure 1**, relative the control cell line, the selected transcripts show higher amounts of m5C in the EMT-positive cell line HMLE pWZL-Snail.

When performing qPCR on our m5C-RIP product to verify the presence of predicted hypomethylated genes (**Table S2**) results were inconsistent. Several immunoprecipitation conditions were attempted; however, each condition yielded contradictory results. As displayed in **figure S1**, most of the genes were found to be hypomethylated in the EMT model cell lines. We tested over 20 different genes where hypomethylation was expected based on bisulfite conversion data; however, all gave conflicting results. This could be attributed to there being fewer 5-methylcytosine sites, limiting the anti-m5C antibody's ability to bind to targets.

N4-ACETYLCYTOSINE (ac4C) EXPERIMENTS

GO Analysis and gene selection

Previously obtained mass spec data from our lab (**Figure S2**) shows that ac4C is greatly upregulated in certain cancer cell-lines that exert the mesenchymal phenotype. This preliminary data prompted us to study the role that ac4C plays in EMT. In preparation for investigating ac4C's role in EMT, gene ontology (GO) analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) was performed on acetylated genes found in HeLa cells by Arango et al (2018). From this analysis, we obtained seven genes that are putatively differentially acetylated as well as involved in EMT (**Table 3**).

NAT10 knockdown

Using siRNA (small interfering RNA), N-acetyltransferase 10 (NAT10) knockdown was performed on HMLE pWZL-Twist to determine the effect that it has on the differentially acetylated genes. As illustrated by **figure 2**, NAT10 was successfully knocked down decreasing its expression to approximately 20 – 25 % of normal. This cell line was used since previous mass spec data (**Figure S3**) suggests that HMLE pWZL-Twist shows a high level of ac4C.

Since NAT10 was successfully knocked down, we sought to determine the effect that the knockdown has on the differentially acetylated genes that we found by gene ontology analysis. As with NAT10 expression, TGF-β2, NOTCH1, BMP2, CUL7, FGFR2, LOXL2, and GSK3β expression was analyzed by SYBR Green qPCR. As presented in **figures 3a and 3b**, we observed a difference in expression between siRNAs, which may be indicative of non-specific binding. For instance, siNAT10 #1 (**Figure 3a**) appears to decrease the expression of BMP2 while siNAT10 #2 (**Figure 3b**) appears to increase its expression. Neglecting the possibility of non-specific binding; however, **Figure 3a** shows that using siNAT10 #,1 NOTCH1 and TGF-β2 seem to have had the

most dramatic differences in expression resulting from NAT10 knockdown. Expression of both genes decreases. This result, however, is not corroborated in **figure 3b**, where expression of only TGF-β2 decreases.

NAT10 overexpression

In order to explore the effect of increased acetylation in the HMLE cell line, we produced an HMLE cell line overexpressing N-acetyltransferase 10 (NAT10); the acetyltransferase responsible for producing ac4C (**Figure 4**). As described in the "methods" section, HMLE cells were transduced using lentiviral tranduction with pLenti harboring NAT10 resulting in our NAT10 overexpression cell line – HMLE pLenti-NAT10.

Since we obtained results following the knockdown of NAT10 in our EMT-positive model HMLE cell line (HMLE- pWZL-Twist), we were interested in contrasting these results with our NAT10-overexpressing cell line (HMLE pLenti-NAT10) and analyzing the difference. Expression of the differentially acetylated genes were compared between HMLE pLenti and HMLE pLenti-NAT10 by qPCR analysis. For the majority of the genes, we noted little difference in gene expression between the NAT10 overexpression (**Figure 5**) cell line and NAT10 knockdown (**Figures 3a and 3b**).

Proliferation Assay

To compare the difference in proliferation rate between HMLE pLenti and HMLE pLenti-NAT10, a CellTiter-Glo® (CTG) Luminescent Cell Viability Assay was conducted. Cells were allowed to grow for 96 hours. **Figure 6** shows two replicates of a CTG assay that was conducted to compare the growth of HMLE pLenti-NAT10 to that of the empty vector control cell line (HMLE pLenti). Although only two replicates had been completed, after the 72-hour time point,

we were able to see escalated proliferation in the HMLE pLenti-NAT10 cell line relative to the HMLE pLenti cell line.

Mammosphere formation

HMLE pLenti-NAT10 and HMLE pLenti (control) were plated in a 6-well ultra-low attachment (ULA) dish containing 2 ml of media. Each plate was able to accommodate three technical replicates for each cell line. Plates were incubated at incubated for 10 days. After this incubation period, plates were imaged and mammospheres from each well were manually counted using ImageJ software. This experiment was repeated three times yielding three biological replicates.

Across the three replicates (**Figure 7**) the average number of mammospheres formed by the HMLE pLenti-NAT10 cell line (mean: 160.7, 207.7, and 165 respectively) were higher than those formed by the control cell line (mean: 97.7, 122, and 172.7 respectively). Furthermore, mammospheres formed by the HMLE pLenti-NAT10 cell line (**Figure 8b**) appeared slightly larger and were generally more abundant than those in control cell line (**Figure 8a**).

This data suggests that overexpression of NAT10 in HMLE cells increases the number of mammospheres which suggests an increase in the cells' abilities to evade anoikis - apoptosis that is induced by inadequate cell-cell interactions (Frisch and Screaton, 2001). In conclusion, these results suggest that NAT10 overexpression in HMLE cells may encourage epithelial to mesenchymal transition by increasing N4-acetylcytossine (ac4C) levels.

ac4C-seq

Our initial strategy for isolating ac4C-containing RNA for experiments involving ac4C was to perform RNA immunoprecipitations (RIP) on mRNA using Abcam's anti-rabbit "Anti-N4-acetylcytidine (ac4C) antibody [EPRNCI-184-128].". We intended to perform ac4C-RIP on

rRNA-reduced or poly-A-selected RNA samples from our EMT model cell lines to isolate ac4C-containing transcripts. The results would then be compared to those collected from our control cell line. These transcripts would later be used for ac4C-sequencing.

We performed immunoprecipitation using total RNA extracted from HMLE pWZL-Twist to determine the utility of the anti-ac4C antibody. After several failed attempts, we finally observed an approximate 3-fold increase in immunoprecipitated ac4C-containing RNA (**Figure S4**); however, this result was not replicable. As a result, we determined that isolating ac4C-containing transcripts using this antibody was not sufficient for our purposes; thus, attempts at isolating acetylated transcripts by immunoprecipitation were discontinued.

As an alternative to performing immunoprecipitation on total RNA to isolate ac4C-containing transcripts, transcripts will be isolated by poly-A selection using the NEBNext® Poly(A) mRNA Magnetic Isolation Module. The poly-A-selected RNA will be treated with sodium cyanoborohydride as per the methods described by Sas-Chen et al. (2020). Treatment of ac4C-containing RNA using the reducing agent sodium cyanoborohydride ($NaCNBH_3$) under acidic conditions reduces N4-acetylcytidine (ac4C) to N4-acetyl-tetrahydrocytidine. Reduction results in a nucleotide misincorporation upon reverse transcription of $NaCNBH_3$-treated RNA. The method alluded to above was tested for efficacy using 18S rRNA from HMLE pWZL-Twist as described thoroughly in the "methods" and "discussion" sections.

After reverse transcription of $NaCNBH_3$-treated samples, amplification, followed by next generation sequencing (NGS) will be performed on cDNA of the poly-A-selected EMT model cell lines. reverse transcription products. cDNA from untreated RNA will be compared to that of treated RNA to determine the presence or absence of ac4C. Thus, the presence of ac4C can be detected. Before conducting the sodium cyanoborohydride treatments on poly-A-selected RNA;

however, it was necessary to confirm whether the treatment could successfully reduce ac4C which could be verified by Sanger sequencing.

Sodium cyanoborohydride treatment of acetylated RNA

The methods described by Sas-Chen et al. (2020) were applied to total RNA which was used as a template for 18S rRNA. As previously stated, there are two confirmed ac4C sites on the 18S ribosomal subunit; C-1337 and C-1842, (Sharma et al., 2015) and reduction of ac4C leads to a nucleotide misincorporation by the reverse transcriptase enzyme which causes the read-out to change from C to T (Sas-Chen et al., 2020). To reduce N4-acetylcytidine (ac4C) to N4-acetyl-tetrahydrocytidine, 100 mM sodium cyanoborohydride (in 100 mM HCl) was added to 1 µg of total RNA from HMLE pWZL-Twist.

Using the primers listed on **table S1** referenced in the "methods" section, we were able to amplify both expected ac4C sites which were confirmed by Sanger sequencing performed at the Center for Functional Genomics using the ABI 3730XL. Although both ac4C sites were successfully amplified, only the ac4C site at C-1842 was observed to have been successfully reduced by sodium cyanoborohydride. As expected, based on findings from Sas-Chen et al. (2020), we detected acetylated C-1842 on the central "C" within the "CCG" motif. As a result of reduction by sodium cyanoborohydride, the "CCG" motif was read as "CTG" by Sanger sequencing, confirming the presence of ac4C (**Figure 9).**

DISCUSSION

The results introduced in these experiments provide us with preliminary evidence of a change in RNA profile upon epithelial to mesenchymal transition. We see that 5-methylcytosine levels are upregulated in several transcripts in EMT model cell lines (**Figure 1**), and although not confirmed, are likely to be downregulated in other transcripts (**Figure S1**). As previously mentioned, we believe that there were inconsistent results for the predicted hypomethylated transcripts due to the lack of m5C sites on the transcripts.

We also show evidence that N4-acetylcytosine is involved in EMT. **Figure 4** confirms successful overexpression of NAT10 in HMLE cell line. Consequently, we postulate that our NAT10-overexpressing cell line (HMLE pLenti-NAT10) expresses increased levels of ac4C compared to the parental cell line (HMLE pLenti). To determine whether this presumed increase in ac4C translated to EMT-positive phenotypes, we conducted a CellTiter-Glo® (CTG) Luminescent Cell Viability Assay to determine this cell line's difference in proliferation rate. According to **figure 6**, it appears as though an increased production of ac4C in HMLE cells increases their proliferation rate. Although more replicates of the CTG assay need to be done to confirm, this provides preliminary evidence of ac4C's involvement in EMT in breast cancer.

To further support our hypothesis that ac4C is associated with EMT, we performed a mammosphere assay. The mammosphere assay is useful for determining a cell's ability to evade anoikis or essentially "death by isolation." When epithelial cells are isolated from surrounding cells, critical cell-cell interactions are non-existent. This signals apoptosis in the isolated epithelial cells. Contrarily, mesenchymal cells do not require these cell-cell interactions for survival. As a result, they are able to proliferate. In the context of this assay, the survival of isolated epithelial cells is represented by the formation of mammary epithelial cell aggregates called

"mammospheres." As illustrated by **figures 7, and 8b**, we observed more mammosphere formation in the NAT10 overexpression cell line (HMLE pLenti-NAT10) compared to the empty vector control cell line (HMLE pLenti). These results support our hypothesis as the ability to evade anoikis and increased proliferation rate are hallmarks of EMT.

To better understand which genes could be affected by increased ac4C resulting from NAT10 overexpression, we obtained a list of differentially acetylated genes presented by Arango et al. (2018). Using this list, differentially acetylated genes involved in EMT (**Table 3**) were uncovered by gene ontology (GO) analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID). When these genes were analyzed by qPCR in HMLE pLenti-NAT10, some notable changes were observed particularly in NOTCH1 and TGF-β2. According to the data displayed in **figure 5** (NAT10 overexpression), the expression of NOTCH1, CUL7, BMP2, and LOXL2 decrease. As displayed in **table 3**, NOTCH1 is involved in cell proliferation and metastasis, CUL7 promotes evasion of apoptosis by ubiquitination of caspases, BMP2 promotes cell invasion and migration, and LOXL2 increases the stability of Snai1. According to the sources of the information presented in **table 3**, the data displayed in **figure 5** is inconsistent with our hypothesis that NAT10 overexpression increases the exertion of EMT-positive phenotypes. Contrarily, while insignificant, there is a trend toward significance showing an increase in TGF-β2. According to the information regarding TGF-β2 by Hachim et al. (2018) displayed in **table 3**, TGF-β2 is implicated in matrix remodeling and evasion of immune surveillance; therefore, the trend toward significance of increased TGF-β2 resulting from NAT10 overexpression is consistent with our hypothesis.

Noting these differences resulting from NAT10 overexpression, we were interested in the effect that knocking-down NAT10 using siRNA would have on the genes displayed in **table 3**.

Underwhelmingly, the results obtained from each siRNA (siNAT10 #1 and siNAT10 #2) were significantly different (**Figures 3a and 3b**). For instance, as depicted in **figure 3a**, NOTCH1, TGF-β2, and BMP2 were significantly decreased resulting from NAT10 knockdown. Furthermore, there was a trend toward significant decrease in the expression of LOXL2 and GSK3β. In contrast, **figure 3b**, depicts only a significant decrease in expression of TGF-β2 and a significant increase in BMP2. Based on the data presented in **table 3**, the majority of the data obtained using siNAT10 #1 (**Figure 3a**) is consistent with our hypothesis. We expected that a decrease in NAT10 expression would result in a decrease in the expression of these EMT-associated differentially acetylated transcripts. While more research is required, the results obtained from these experiments provide further preliminary evidence of ac4C's involvement in EMT in breast cancer.

Assuming the data from **figure 3a** (using siNAT10#1) is correct, the decrease in expression of NOTCH1, TGF-β2, and BMP2 are consistent with our hypothesis. As inferred from the information presented in **table 3**, a decrease in expression of NOTCH1, TGF-β2, and BMP2 would reduce a cell's ability to proliferate and metastasize, evade the immune surveillance, and facilitate EMT respectively.

Since acetylation of all the genes listed in **table 3** are described to promote tumor progression, metastasis, or immune surveillance evasion, we anticipated that knocking down NAT10, would decrease the promotion of these genes. Conversely, we anticipated that the overexpression of NAT10 in the HMLE cell line would increase acetylation; therefore, promoting the expression of the genes listed in **table 3**. Unimpressively, we only observed this for TGF- β2. Expression of TGF-β2 decreased as result of NAT10 knockdown (**Figures 3a and 3b**), while it increased in the NAT10-overexpressing HMLE cell line (**Figure 5**).

As previously described, as we accomplished for m5C-containing transcripts, we intended to perform RNA immunoprecipitations (RIP) of ac4C-containing transcripts using rRNA-reduced or poly-A-selected RNA extracted from our EMT model cell lines. Using total RNA extracted from HMLE pWZL-Twist, we attempted several ac4C-RIP attempts; however, we were unsuccessful in replicating our best immunoprecipitation which was accomplished using a tRNA block. We observed a 3-fold increase in immunoprecipitated ac4C-containing RNA (**Figure S4**). Since we were unable to replicate this result, attempts at isolating ac4C-containing transcripts by ac4C-RIP were abandoned. Although we observed significant immunoprecipitation of ac4C-containing RNA, we determined that further attempts at using tRNA as a block should be abandoned as tRNA contains ac4C. The presence of ac4C in the tRNA block could have potentially produce false positive results.

In an effort to validate and sequence ac4C-containing transcripts in our EMT model cell lines, we implemented a method of determining ac4C sites proposed by Sas-Chen et al. (2020). This method involved chemical manipulation of poly-A-selected RNA extracted from HMLE pWZL-Snail, Twist, and TGF- β as well as the vector control, HMLE pWZL-GFP.

The method involved reducing N4-acetylated cytidine (ac4C) to N4-acetyl-tetrahydrocytidine using sodium cyanoborohydride ($NaCNBH_3$) under acidic conditions. After reduction, reverse transcription was performed on the reduced product. Reverse transcription of canonical cytidine (C) yields complimentary DNA (cDNA) containing guanidine (G) at the corresponding ac4C site. Contrarily, reverse transcription of N4-acetyl-tetrahydrocytidine (reduced ac4C) is read as uridine (U) by DNA polymerase as opposed to C. This yields cDNA containing the mis-incorporated base pair, adenosine (A) rather than G, at that site. After chemical manipulation, and reverse transcription, the cDNA template from the reduced ac4C-containing

RNA is amplified by PCR. The resulting amplified cDNA will yield thymine (T) in place of C as indicated in **figure 9b**.

To accomplish this amplification, primers were designed to amplify sections containing both ac4C-containing sites of the human 18S rRNA subunit (C1337 and C1842 with predicted sizes of 175 and 214 base pairs respectively) (**Table S1**). After amplification, gel electrophoresis of each amplicon was performed to isolate and validate the size of each ac4C-containing section. Bands were excised and gel purified. Finally, gel purified samples of each amplicon were Sanger sequenced to determine whether reverse transcription mis-incorporation of the ac4C sites was accomplished; thereby, confirming the utility of this method for discovery of ac4C-containing sites of RNA.

Future directions

After confirming the efficacy of this method of determining ac4C sites using 18S rRNA, it was performed on the poly-A-selected RNA extracted from our EMT model cell lines. The samples were stored at -80 °C for future analysis. Poly-A-selected RNA reduced by $NaCNBH_3$ treatment will then be amplified using random hexamer primers (oligonucleotides). The amplified product of which will be used for the discovery of new ac4C sites on several transcripts.

For future experiments, it may be worth performing ac4C-RIP using different antibodies under several conditions. While we were able to isolate or concentrate mRNA (transcripts) by performing Ribo-reduction or Poly-A selection, it would be ideal to isolate transcripts containing ac4C as we were able to accomplish with m5C-containing transcripts. Furthermore, as previously indicated, this study was mainly focused on experiments involving ac4C. As a result, future experiments would involve altering the expression of m5C by performing knockdowns and producing overexpressing cell lines of its “writers” NSUN2 and TRDMT1. As presented in this

study, alteration of NAT10, the "writer" of ac4C, allows us to get a functional understanding of the phenotypic effects of altering the expression of this post-transcriptional modification in the HMLE cell line.

In conclusion, this study provides a foundation of evidence for our hypothesis that the abundance of the post-transcriptional modifications m5C and ac4C are correlated with epithelial to mesenchymal transition in human breast cancer. While it is unknown whether changes in post-transcriptional modifications are responsible for driving EMT or result from EMT, there seems to be an association between the two. As a result, there is a need for more investigation into unveiling the roles that post-transcriptional modifications may play in EMT and metastasis. Furthermore, we validate a method of determining ac4C-containing RNA sites discovered by Sas Chen et al. (2020) by replicating it using 18S rRNA, which has been confirmed to contain two ac4C sites (C1337 and C1842). We hope that the results and trends presented in this study will provide encouragement to future investigators who hope to add to the growing body of knowledge in the determination of the functions of post-transcriptional modifications in cancer. These findings provide steps toward elucidating the causal factors contributing to metastatic breast cancer.

TABLES

Gene	Full gene name	Gene function / implications in cancer
Snai1	Snail	Zinc-finger transcription factor. Directly downregulates E-cadherin (CDH1). (M.A. Nieto, 2002)
Twist1	Twist	A master regulator of morphogenesis associated with activation of mesenchymal markers. (Yang et al., 2004)
TGF-β1	Transforming growth factor beta 1	Promotes EMT through the expression of E-cadherin transcriptional repressors (SNAIL, ZEB, and TWIST). (Hao et al., 2019)
CDH1	E-cadherin	Responsible for the formation of adherens junctions and desmosomes. (Neel et al., 2022)
EpCAM	Epithelial cellular adhesion molecule	Involved in cell signaling, proliferation, and differentiation. (Gires et al., 2020)
VIM	Vimentin	Major cytoskeletal component of mesenchymal cells. Marker of EMT+ or mesenchymal cells. (Battaglia et al., 2018)

Table 1. EMT-associated genes.

Gene	Full gene name	Gene function / implications in cancer
GNB4	Guanine nucleotide binding protein subunit-4	Promotes cancer cell infiltration by immune cells. (Liu et al., 2022)
EIF2AK2	Eukaryotic Translation Initiation Factor 2 Alpha Kinase 2	Implicated in the suppression of metastatic capabilities. (Kim et al., 2017)
PRR11	Proline rich 11	Involved in cell cycle regulation. (Han et al., 2022)
EEF1Δ	Eukaryotic elongation factor 1-delta	Modulates proliferation, migration, and invasion. (Klopfleisch et al., 2010)
OIP5-AS1	Opa-interacting protein 5-antisense RNA 1	Has been shown to promote breast cancer metastasis. (Meng et al., 2020)

Table 2. Hypermethylated transcripts determined by bisulfite conversion data obtained from Chris Mason's team (**Figure S1**).

Gene	Full gene name	Gene function / implications in cancer
TGF-β2	Transforming growth factor beta 2	Evasion of immune surveillance and matrix remodeling. (Hachim et al., 2018)
NOTCH1	Notch receptor 1	Cancer cell proliferation and metastasis. (Zhong et al., 2022)
BMP2	Bone morphogenetic protein 2	Member of the TGF-β superfamily. Facilitates EMT and metastasis in bone and breast cancer. (Huang et al., 2017)
CUL7	Cullin 7	Member of the cullin family of E3 ubiquitin ligases. Evasion of apoptosis by caspase ubiquitination. (Kong et al., 2019) (Qiu et al., 2018)
FGFR2	Fibroblast growth factor receptor 2	Suppresses BRCA1 and promotes tumor progression. (Lei et al., 2021)
LOXL2	Lysyl oxidase-like 2	Increases stability of Snai1. (Hollosi et al., 2009)
GSK3β	Glycogen synthase kinase 3 beta	Promotes cell invasion and migration via AKT-GSK3β-Snail pathway. (Zhang et al., 2017)

Table 3. Proposed differentially acetylated EMT-associated genes obtained from gene ontology (GO) analysis of "gene list" presented by Arango et al.

SUPPLEMENTARY TABLES

Target site	Forward primer sequence	Reverse primer sequence	Amplicon size (base pairs)
C-1337	5'-GGAGCCTGCGGCTTAATTTG-3'	5'-AGCATGCCAGAGTCTCGTTC-3'	175
C-1842	5'-GGGTCATAAGCTTGCGTTGAT-3'	5'-TAATGATCCTTCCGCAGGTT-3'	214

Table S1. Primers used to amplify the two ac4C sites of 18S rRNA.

Gene name	Forward primer	Reverse primer
GNB4	5'- TCCTATCCAAAGGCATCCACA-3'	5'- TGTTCAGTTGACCACGAGTGT-3'
EIF2AK2	5'-CTTCCATCTGACTCAGGTTTGC-3'	5'-ACTCCCTGCTTCTGACGGTA-3'
PRR11	5'-TCACTGAGAATAAAGCAGTTGAGT-3'	5'-TCCCACAAGAATGGTCAAGTCA-3'
EEF1Δ	5'-GAGGCCCTGCCACGATTAAA-3'	5'-AGGAAATGACCAGGACGGAG-3'
OIP5-AS1	5'-AAGCAGAGGGCTGTGCTAAA-3'	5'- AGGTGGTTTAGGGGGATGGT-3'
TGF-β2	5'-CTCCGAAAATGCCATCCCGC-3'	5'-GCTCAATCCGTTGTTCAGGC-3'
NOTCH1	5'-GGACGTCAGACTTGGCTCAG-3'	5'-ACATCTTGGGACGCATCTGG-3'
BMP2	5'-ACTCGAAATTCCCCGTGACC-3'	5'-CCACTTCCACCACGAATCCA-3'
CUL7	5'-TCCGCAGACATGCTCAATCA-3'	5'-CTTGAGGACCCCTCCTGGTA-3
FGFR2	5'-CGAGCAAAGTTTGGTGGAGG-3'	5'-TCGGATTTGGGGAACGAGAG-3'
LOXL2	5'-CCAGTGTGGTCTGCAGAGAG-3'	5'-CCTGTGCACTGGATCTCGTT-3'
GSK3β	5'-GCTTCAACCCCCACAAATGC-3'	5'-CCAAACGTGACCAGTGTTGC-3'

Table S2. SYBR Green primers used for qPCR analysis.

FIGURES

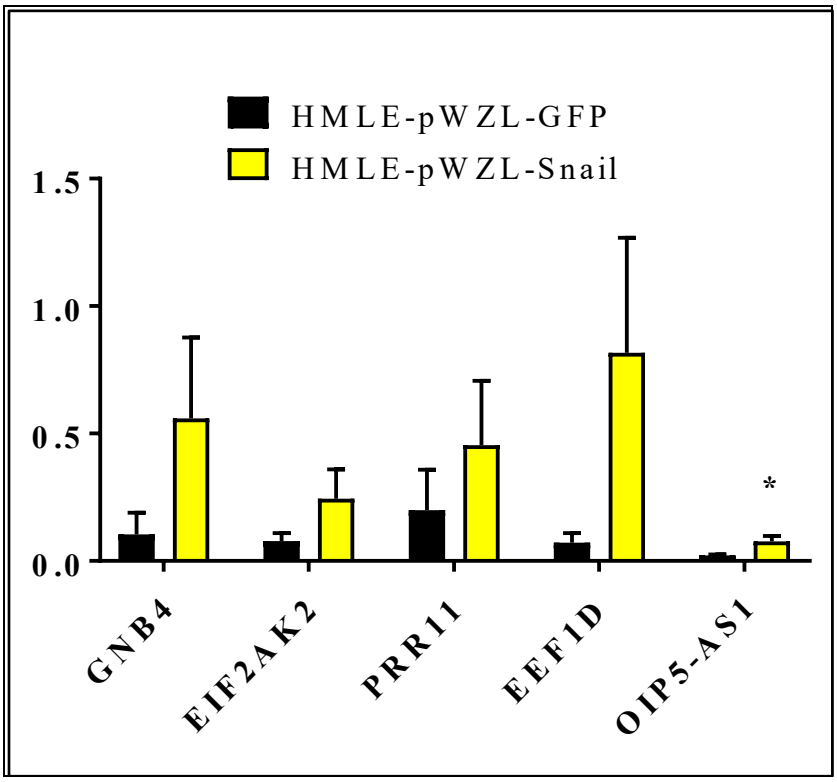

Figure 1. m5C-RIP in both control (HMLE-GFP) and EMT induced (HMLE-Snail) cell lines was performed to determine if target genes become hypermethylated following EMT. Three replicates. +/- standard deviation is graphed. * indicates $p<0.05$.

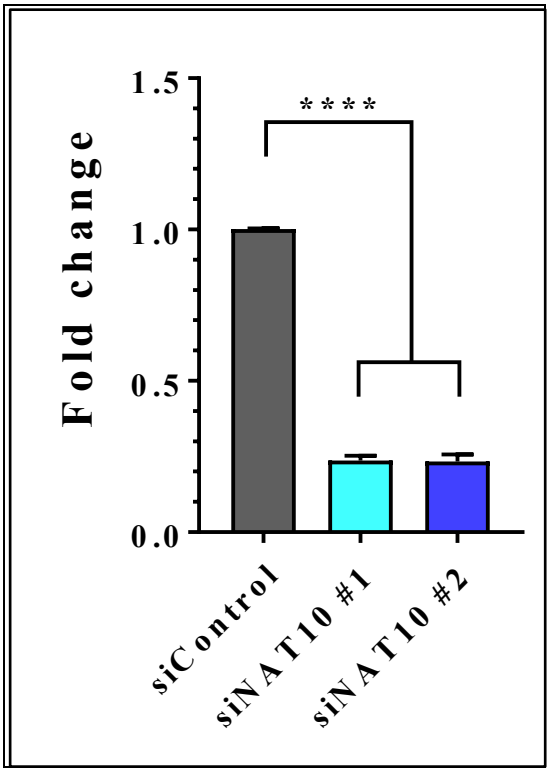

Figure 2. 50nM of the indicated siRNA were transfected into 2.0 x10^5 HMLE-Twist cells. Total RNA was extracted 72h later and assessed for NAT10 expression. **** indicates $p<0.0001$

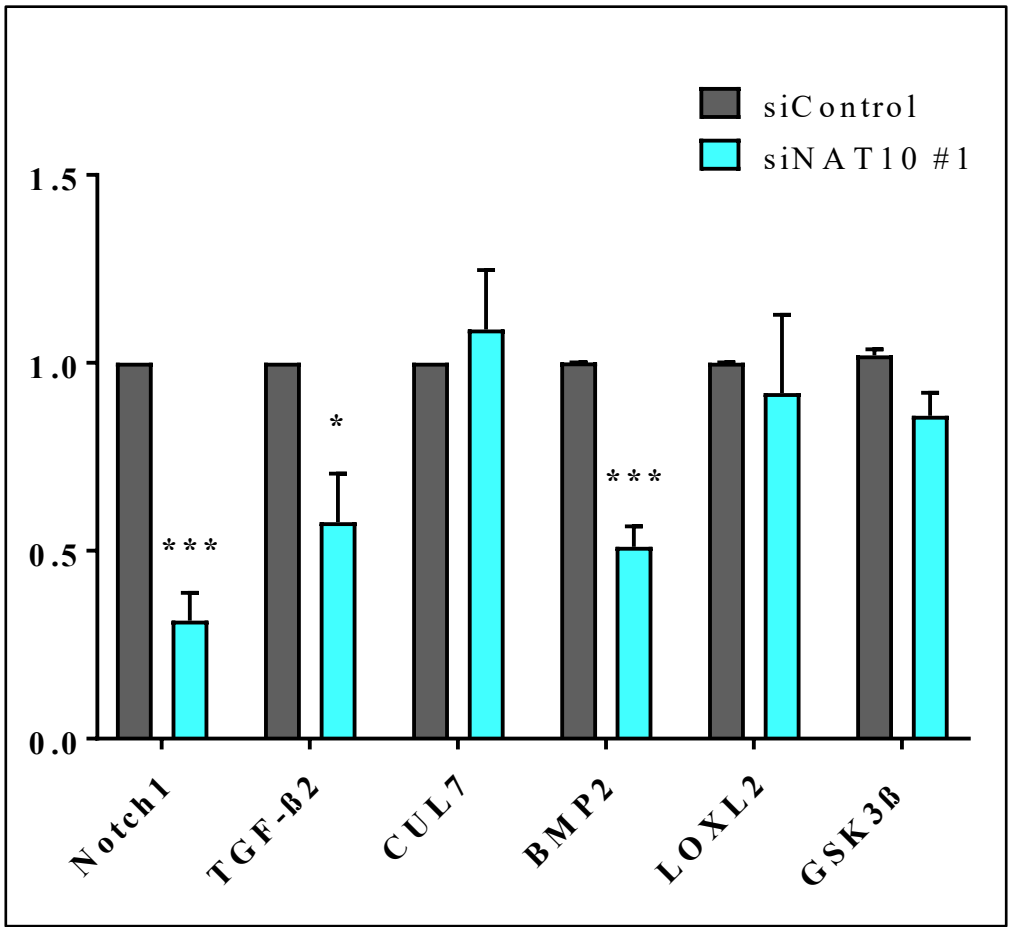

Figure 3a. Expression of NAT10 target genes following 72-hour siRNA knockdown. Three replicates. +/- standard deviation is graphed. * indicates $p<0.05$. *** indicates $p<0.001$.

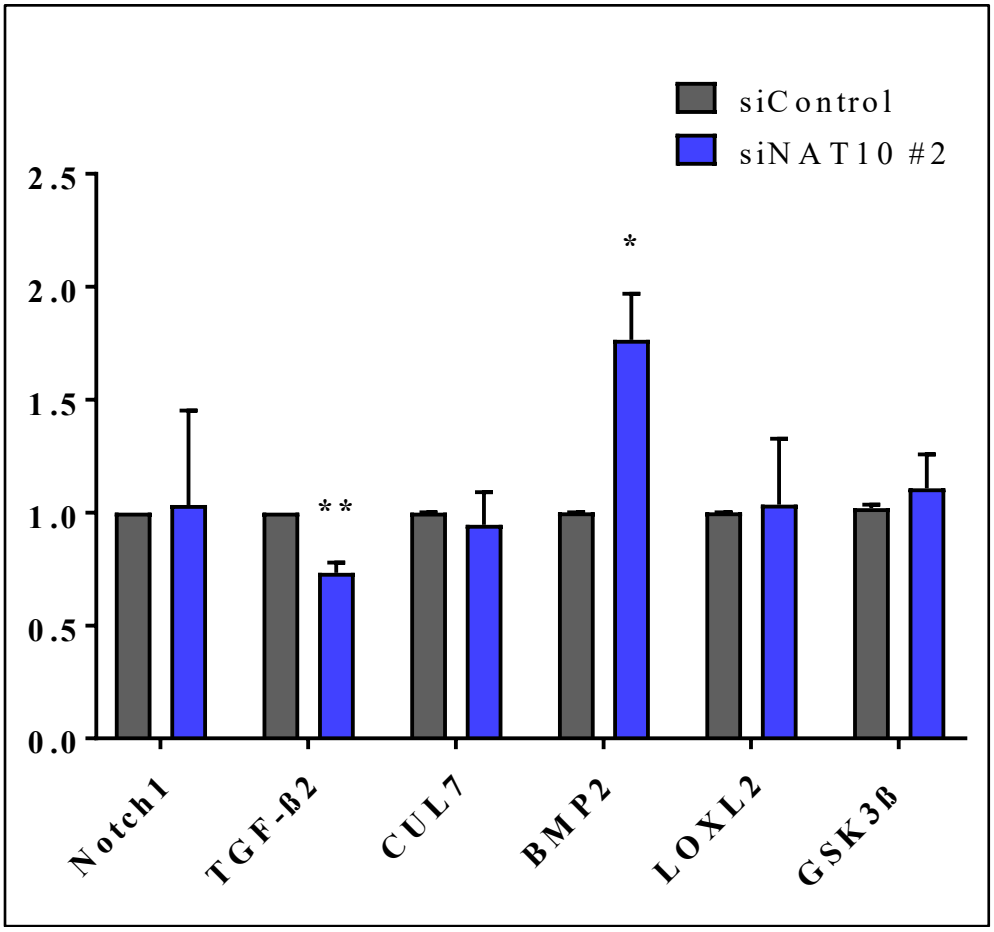

Figure 3b. Expression of NAT10 target genes following 72-hour siRNA knockdown. Three replicates. +/- standard deviation is graphed. * indicates $p<0.05$. ** indicates $p<0.01$.

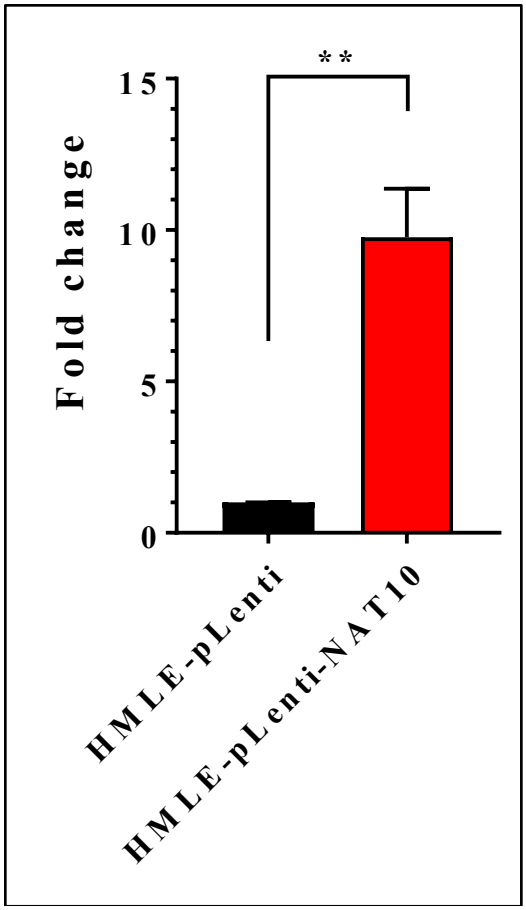

Figure 4. Overexpression of NAT10 by lentiviral transfection in HMLE. Three replicates. +/- standard deviation is graphed. ** indicates $p<0.01$.

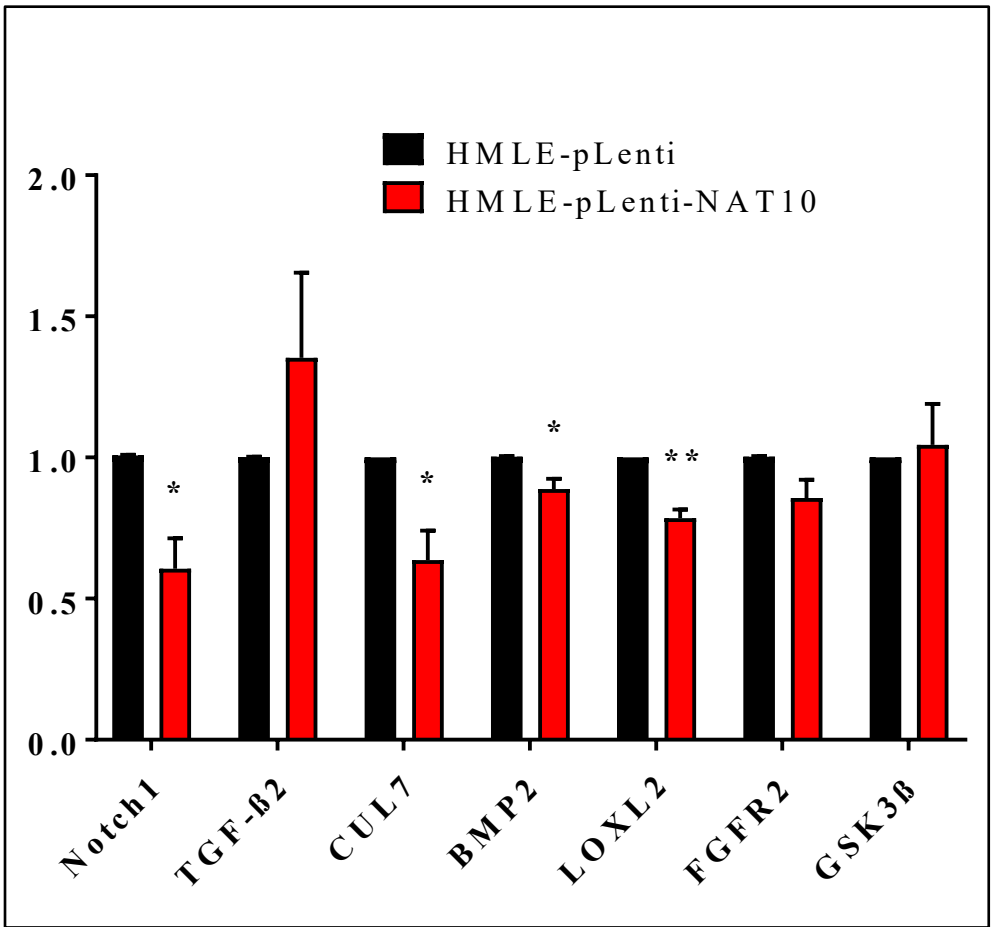

Figure 5. Expression of putative NAT10 target genes in either control (black) or NAT10 overexpressing (red) cells. Three replicates. +/- standard deviation is graphed. * indicates $p<0.05$. ** indicates $p<0.01$

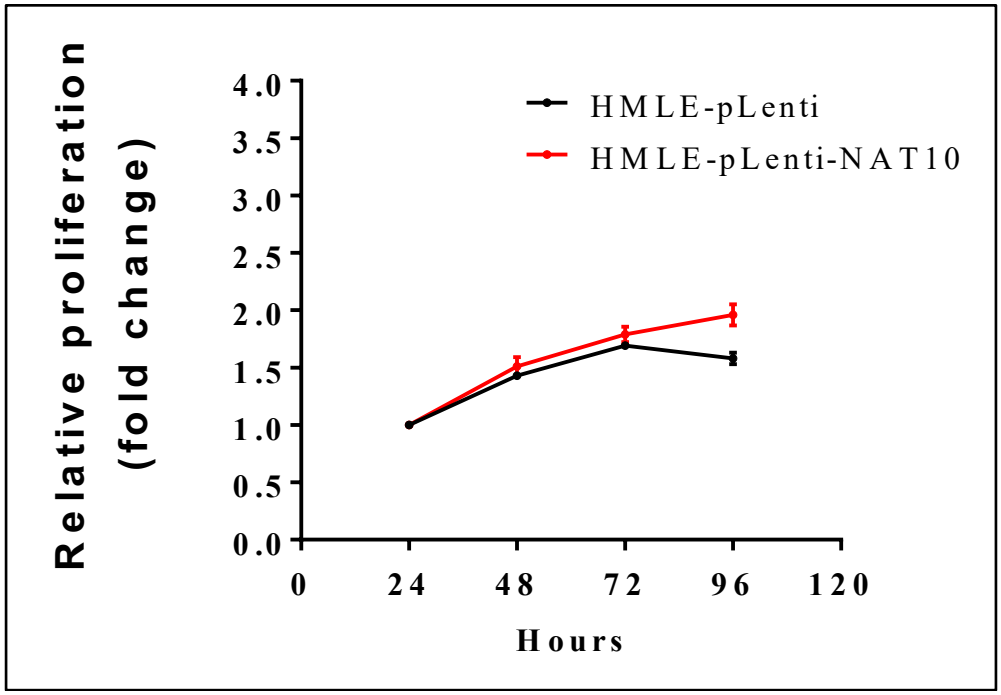

Figure 6. 5,000 cells of the indicated cell lines were seeded in 96-well plates on day 1. Cell proliferation was assessed by CellTiter-Glo® (CTG) Luminescent Cell Viability assay at indicated time points. Mean +/- range is graphed.

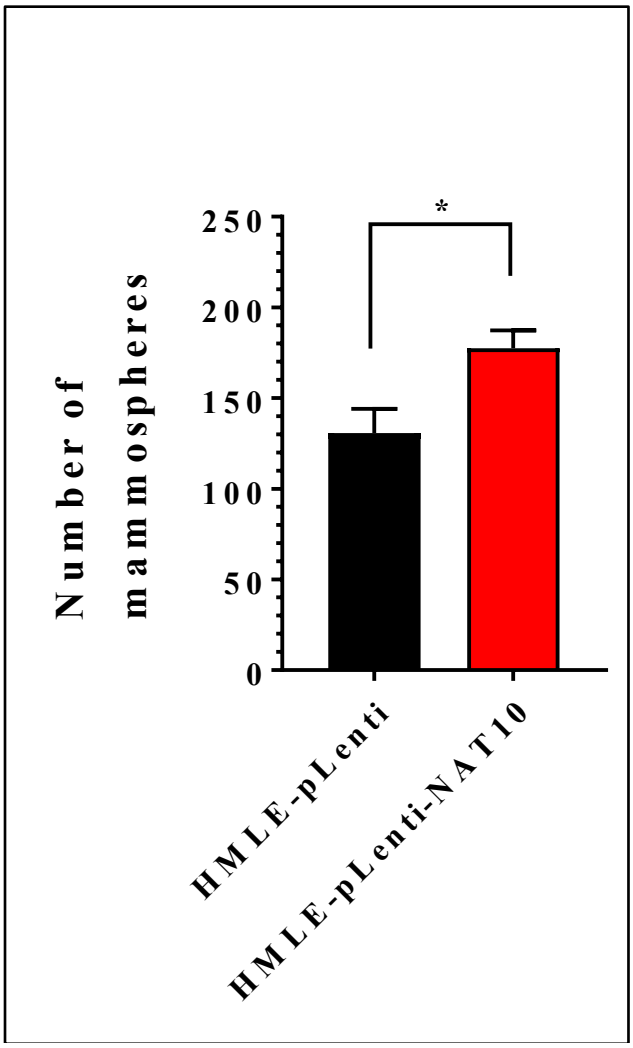

Figure 7. Mammosphere formation in control and NAT10 overexpressing HMLE cells. Three replicates. +/- standard deviation is graphed. * indicates $p<0.05$.

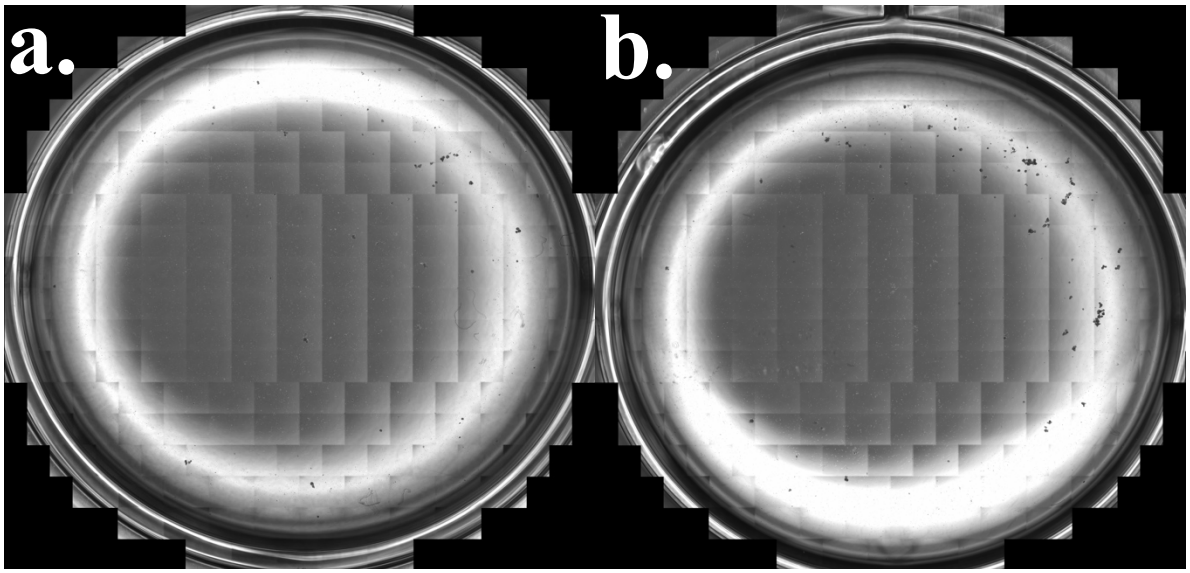

Figure 8. Representative composite photos obtained from EVOS FL Auto Imaging System (Thermo Fisher Scientific) displaying mammospheres. 50,000 cells/well plated on 6-well Ultra-Low Binding Culture Plate left undisturbed for 10 days in incubator at 37°C and 5% CO_2. **a.** HMLE pLenti, and **b.** HMLE pLenti-NAT10.

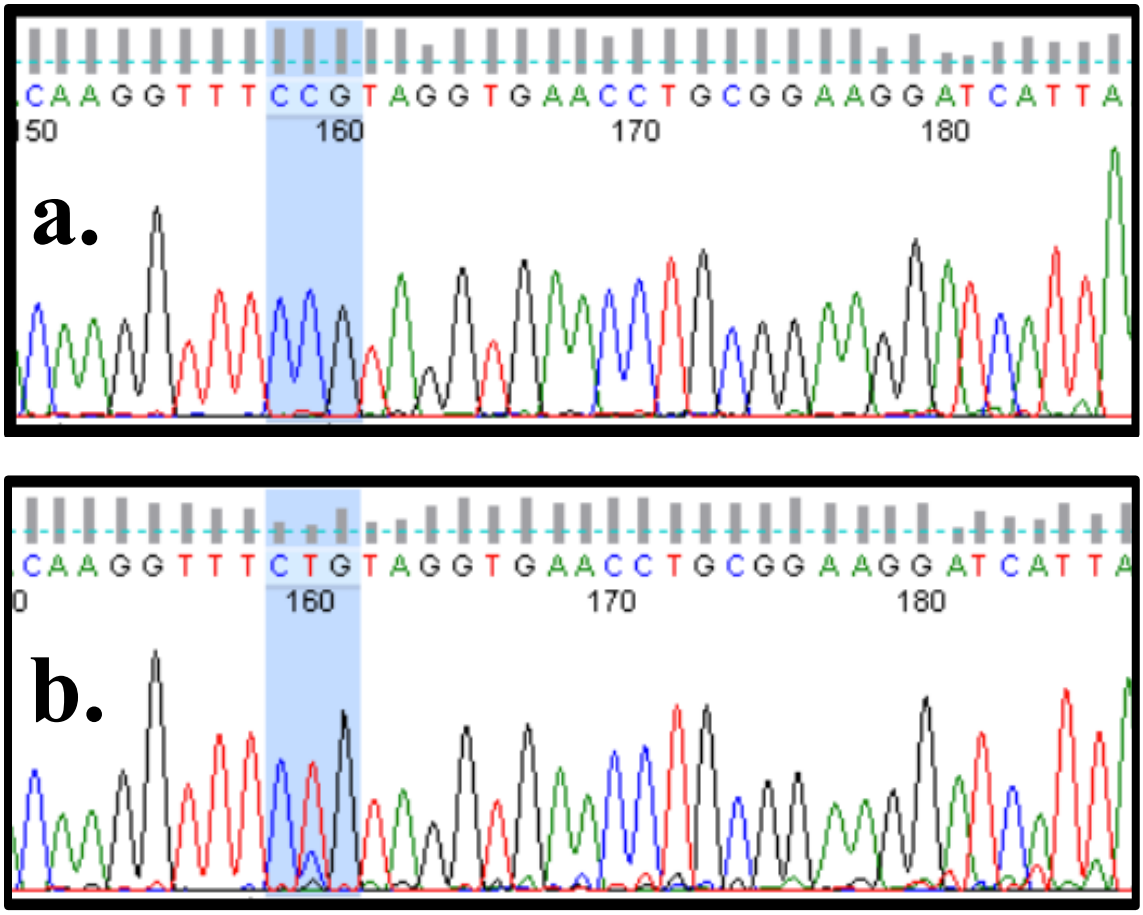

Figure 9. Sanger sequencing representative spectra highlighting the "CCG" motif at C-1842 of the reverse transcription product of 18S rRNA extracted from HMLE-Twist. **a.** Untreated. **b.** Sodium cyanoborohydride ($NaCNBH_3$)-treated. The change in read-out from "CCG" to "CTG" indicates successful reduction of ac4C by $NaCNBH_3$.

SUPPLEMENTARY FIGURES

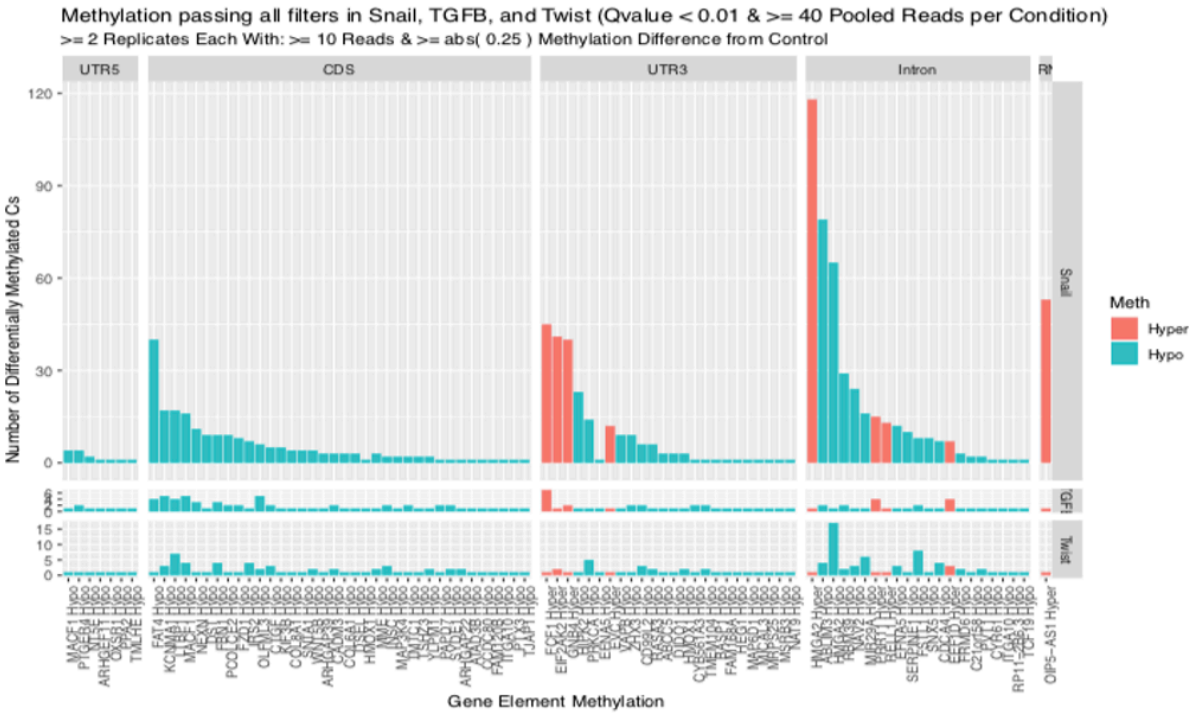

Figure S1. Bisulfite conversion data of differentially methylated Cs in EMT model cell lines (Unpublished data from Dr. Christopher Mason's team at Weill Cornell Medical College)

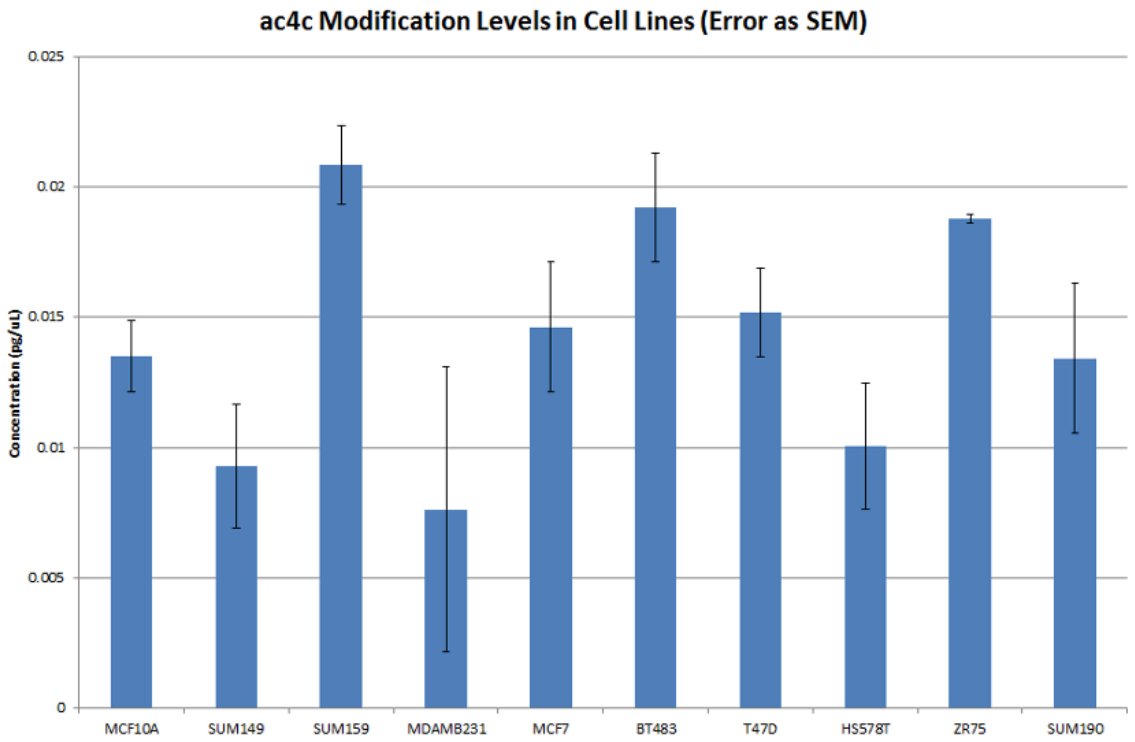

Figure S2. Relative N4-acetylcytosine levels in various breast cancer cell lines (epithelial and mesenchymal morphologies) obtained by mass spectrometry (Unpublished data from our lab)

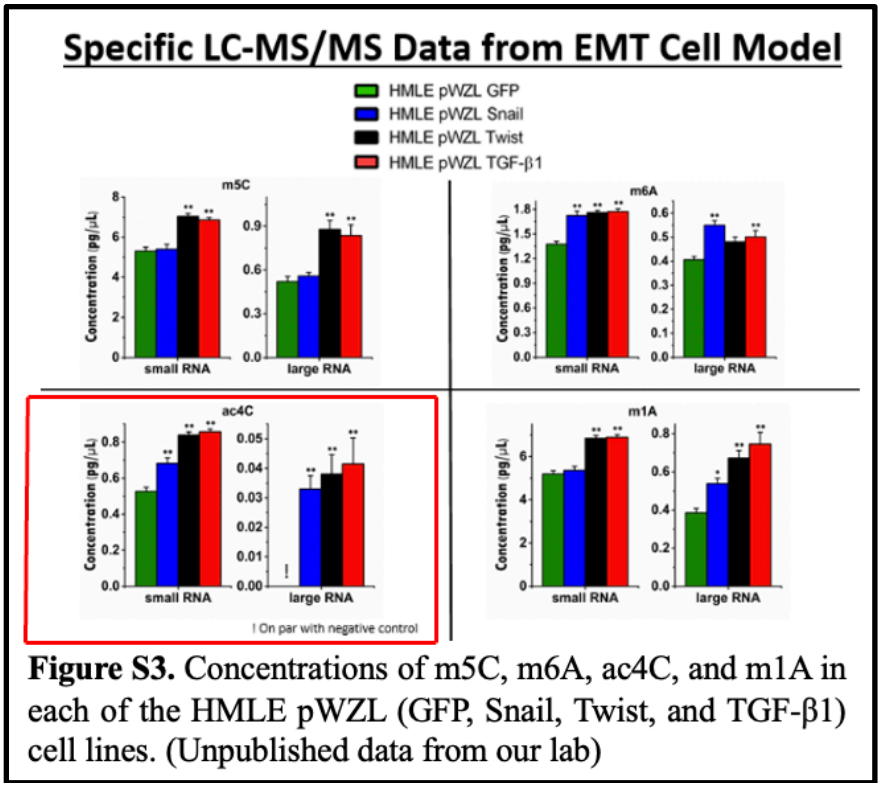

Figure S3. Concentrations of m5C, m6A, ac4C, and m1A in each of the HMLE pWZL (GFP, Snail, Twist, and TGF-β1) cell lines. (Unpublished data from our lab)

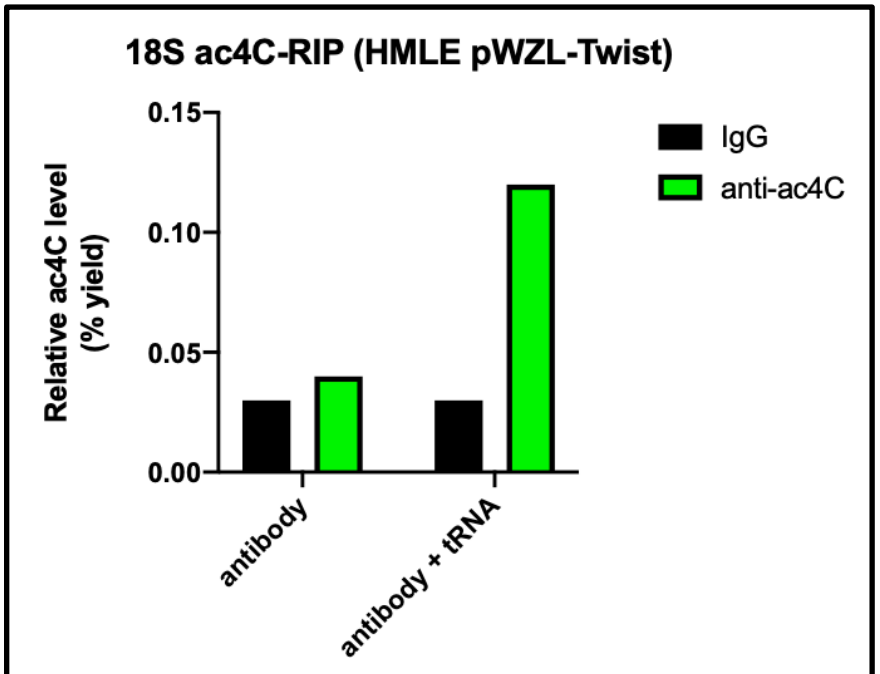

Figure S4. Best ac4C-RNA immunoprecipitation (ac4C-RIP) result. 18S rRNA samples were immunoprecipitated using anti-ac4C antibody (Supplied by Abcam). ac4C-RIP enrichment was compared to that of IgG control. Samples were blocked with tRNA. (n=1)

References

Anders, C., & Carey, L. A. (2008). Understanding and treating triple-negative breast cancer. *Oncology (Williston Park, N.Y.)*, *22*(11), 1233–1243.

Arango, D., Sturgill, D., Alhusaini, N., Dillman, A. A., Sweet, T. J., Hanson, G., Hosogane, M., Sinclair, W. R., Nanan, K. K., Mandler, M. D., Fox, S. D., Zengeya, T. T., Andresson, T., Meier, J. L., Coller, J., & Oberdoerffer, S. (2018). Acetylation of Cytidine in mRNA Promotes Translation Efficiency. *Cell*, *175*(7), 1872–1886.e24. https://doi.org/10.1016/j.cell.2018.10.030

Barriere, G., Fici, P., Gallerani, G., Fabbri, F., & Rigaud, M. (2015). Epithelial Mesenchymal Transition: a double-edged sword. *Clinical and translational medicine*, *4*, 14. https://doi.org/10.1186/s40169-015-0055-4

Battaglia, R. A., Delic, S., Herrmann, H., & Snider, N. T. (2018). Vimentin on the move: new developments in cell migration. *F1000Research*, *7*, F1000 Faculty Rev-1796. https://doi.org/10.12688/f1000research.15967.1

Blanco, S., Dietmann, S., Flores, J. V., Hussain, S., Kutter, C., Humphreys, P., Lukk, M., Lombard, P., Treps, L., Popis, M., Kellner, S., Hölter, S. M., Garrett, L., Wurst, W., Becker, L., Klopstock, T., Fuchs, H., Gailus-Durner, V., Hrabě de Angelis, M., Káradóttir, R. T., … Frye, M. (2014). Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *The EMBO journal*, *33*(18), 2020–2039. https://doi.org/10.15252/embj.201489282

Blanpain, C., Horsley, V., & Fuchs, E. (2007). Epithelial stem cells: turning over new leaves. *Cell*, *128*(3), 445–458. https://doi.org/10.1016/j.cell.2007.01.014

Chaffer, C. L., & Weinberg, R. A. (2011). A perspective on cancer cell metastasis. *Science (New York, N.Y.)*, *331*(6024), 1559–1564. https://doi.org/10.1126/science.1203543

DeSantis, C. E., Bray, F., Ferlay, J., Lortet-Tieulent, J., Anderson, B. O., & Jemal, A. (2015). International Variation in Female Breast Cancer Incidence and Mortality Rates. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, *24*(10), 1495–1506. https://doi.org/10.1158/1055-9965.EPI-15-0535

DeVaux, R. S., & Herschkowitz, J. I. (2018). Beyond DNA: the Role of Epigenetics in the Premalignant Progression of Breast Cancer. *Journal of mammary gland biology and neoplasia*, *23*(4), 223–235. https://doi.org/10.1007/s10911-018-9414-2

Dong, Z., & Cui, H. (2020). The Emerging Roles of RNA Modifications in Glioblastoma. *Cancers*, *12*(3), 736. https://doi.org/10.3390/cancers12030736

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International journal of cancer*, *144*(8), 1941–1953. https://doi.org/10.1002/ijc.31937

Frisch, S. M., & Screaton, R. A. (2001). Anoikis mechanisms. *Current opinion in cell biology*, *13*(5), 555–562. https://doi.org/10.1016/s0955-0674(00)00251-9

Gangnon, R. E., Stout, N. K., Alagoz, O., Hampton, J. M., Sprague, B. L., & Trentham-Dietz, A. (2018). Contribution of Breast Cancer to Overall Mortality for US Women. *Medical decision making : an international journal of the Society for Medical Decision Making*, *38*(1_suppl), 24S–31S. https://doi.org/10.1177/0272989X17717981

Gao, J. J., & Swain, S. M. (2018). Luminal A Breast Cancer and Molecular Assays: A Review. *The oncologist*, *23*(5), 556–565. https://doi.org/10.1634/theoncologist.2017-0535

Gennari, A., D'amico, M., & Corradengo, D. (2011). Extending the duration of first-line chemotherapy in metastatic breast cancer: a perspective review. *Therapeutic advances in medical oncology*, *3*(5), 229–232. https://doi.org/10.1177/1758834011413423

Gires, O., Pan, M., Schinke, H., Canis, M., & Baeuerle, P. A. (2020). Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years?. *Cancer metastasis reviews*, *39*(3), 969–987. https://doi.org/10.1007/s10555-020-09898-3

Grosjean H. (2015). RNA modification: the Golden Period 1995-2015. *RNA (New York, N.Y.)*, *21*(4), 625–626. https://doi.org/10.1261/rna.049866.115

Hachim, M. Y., Hachim, I. Y., Dai, M., Ali, S., & Lebrun, J. J. (2018). Differential expression of TGFβ isoforms in breast cancer highlights different roles during breast cancer progression. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine*, *40*(1), 1010428317748254. https://doi.org/10.1177/1010428317748254

Han, W., & Chen, L. (2022). PRR11 in Malignancies: Biological Activities and Targeted Therapies. *Biomolecules*, *12*(12), 1800. https://doi.org/10.3390/biom12121800

Hao, Y., Baker, D., & Ten Dijke, P. (2019). TGF-β-Mediated Epithelial-Mesenchymal Transition and Cancer Metastasis. *International journal of molecular sciences*, *20*(11), 2767. https://doi.org/10.3390/ijms20112767

Hollosi, P., Yakushiji, J. K., Fong, K. S., Csiszar, K., & Fong, S. F. (2009). Lysyl oxidase-like 2 promotes migration in noninvasive breast cancer cells but not in normal breast epithelial cells. *International journal of cancer*, *125*(2), 318–327. https://doi.org/10.1002/ijc.24308

Huang, P., Chen, A., He, W., Li, Z., Zhang, G., Liu, Z., Liu, G., Liu, X., He, S., Xiao, G., Huang, F., Stenvang, J., Brünner, N., Hong, A., & Wang, J. (2017). BMP-2 induces EMT and breast cancer stemness through Rb and CD44. *Cell death discovery*, *3*, 17039. https://doi.org/10.1038/cddiscovery.2017.39

Hugo, H. J., Kokkinos, M. I., Blick, T., Ackland, M. L., Thompson, E. W., & Newgreen, D. F. (2011). Defining the E-cadherin repressor interactome in epithelial-mesenchymal transition: the PMC42 model as a case study. *Cells, tissues, organs*, *193*(1-2), 23–40. https://doi.org/10.1159/000320174

Ito, S., Horikawa, S., Suzuki, T., Kawauchi, H., Tanaka, Y., Suzuki, T., & Suzuki, T. (2014). Human NAT10 is an ATP-dependent RNA acetyltransferase responsible for N4-acetylcytidine formation in 18 S ribosomal RNA (rRNA). *The Journal of biological chemistry*, *289*(52), 35724–35730. https://doi.org/10.1074/jbc.C114.602698

Jonkhout, N., Tran, J., Smith, M. A., Schonrock, N., Mattick, J. S., & Novoa, E. M. (2017). The RNA modification landscape in human disease. *RNA (New York, N.Y.)*, *23*(12), 1754–1769. https://doi.org/10.1261/rna.063503.117

Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation*, *119*(6), 1420–1428. https://doi.org/10.1172/JCI39104

Kim, T. H., & Cho, S. G. (2017). Kisspeptin inhibits cancer growth and metastasis via activation of EIF2AK2. *Molecular medicine reports*, *16*(5), 7585–7590. https://doi.org/10.3892/mmr.2017.7578

Klopfleisch, R., Klose, P., Weise, C., Bondzio, A., Multhaup, G., Einspanier, R., & Gruber, A. D. (2010). Proteome of metastatic canine mammary carcinomas: similarities to and differences from human breast cancer. *Journal of proteome research*, *9*(12), 6380–6391. https://doi.org/10.1021/pr100671c

Kong, Y., Wang, Z., Huang, M., Zhou, Z., Li, Y., Miao, H., Wan, X., Huang, J., Mao, X., & Chen, C. (2019). CUL7 promotes cancer cell survival through promoting Caspase-8 ubiquitination. *International journal of cancer*, *145*(5), 1371–1381. https://doi.org/10.1002/ijc.32239

Lei, J. H., Lee, M. H., Miao, K., Huang, Z., Yao, Z., Zhang, A., Xu, J., Zhao, M., Huang, Z., Zhang, X., Chen, S., Jiaying, N. G., Feng, Y., Xing, F., Chen, P., Sun, H., Chen, Q., Xiang, T., Chen, L., Xu, X., … Deng, C. X. (2021). Activation of FGFR2 Signaling Suppresses BRCA1 and Drives Triple-Negative Mammary Tumorigenesis That is Sensitive to Immunotherapy. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)*, *8*(21), e2100974. https://doi.org/10.1002/advs.202100974

Lin, X., Chai, G., Wu, Y., Li, J., Chen, F., Liu, J., Luo, G., Tauler, J., Du, J., Lin, S., He, C., & Wang, H. (2019). RNA m^6A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail. *Nature communications*, *10*(1), 2065. https://doi.org/10.1038/s41467-019-09865-9

Liu, B., Chen, L., Huang, H., Huang, H., Jin, H., & Fu, C. (2022). Prognostic and Immunological Value of GNB4 in Gastric Cancer by Analyzing TCGA Database. *Disease markers*, *2022*, 7803642. https://doi.org/10.1155/2022/7803642

Luo, M., Brooks, M., & Wicha, M. S. (2015). Epithelial-mesenchymal plasticity of breast cancer stem cells: implications for metastasis and therapeutic resistance. *Current pharmaceutical design*, *21*(10), 1301–1310. https://doi.org/10.2174/1381612821666141211120604

Machnicka, M. A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K. M., Helm, M., Bujnicki, J. M., & Grosjean, H. (2013). MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic acids research*, *41*(Database issue), D262–D267. https://doi.org/10.1093/nar/gks1007

Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J., & Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, *133*(4), 704–715. https://doi.org/10.1016/j.cell.2008.03.027

Maughan, K. L., Lutterbie, M. A., & Ham, P. S. (2010). Treatment of breast cancer. *American family physician*, *81*(11), 1339–1346.

Meng, L., Yue, X., Zhou, D., & Li, H. (2020). Long non coding RNA OIP5-AS1 promotes metastasis of breast cancer via miR-340-5p/ZEB2 axis. *Oncology reports*, *44*(4), 1662–1670. https://doi.org/10.3892/or.2020.7724

Mitri, Z., Constantine, T., & O'Regan, R. (2012). The HER2 Receptor in Breast Cancer: Pathophysiology, Clinical Use, and New Advances in Therapy. *Chemotherapy research and practice*, *2012*, 743193. https://doi.org/10.1155/2012/743193

Moo, T. A., Sanford, R., Dang, C., & Morrow, M. (2018). Overview of Breast Cancer Therapy. *PET clinics*, *13*(3), 339–354. https://doi.org/10.1016/j.cpet.2018.02.006

Neel, B. L., Nisler, C. R., Walujkar, S., Araya-Secchi, R., & Sotomayor, M. (2022). Collective mechanical responses of cadherin-based adhesive junctions as predicted by simulations. *Biophysical journal*, *121*(6), 991–1012. https://doi.org/10.1016/j.bpj.2022.02.008

Nieto M. A. (2002). The snail superfamily of zinc-finger transcription factors. *Nature reviews. Molecular cell biology*, *3*(3), 155–166. https://doi.org/10.1038/nrm757

Nisticò, P., Bissell, M. J., & Radisky, D. C. (2012). Epithelial-mesenchymal transition: general principles and pathological relevance with special emphasis on the role of matrix metalloproteinases. *Cold Spring Harbor perspectives in biology*, *4*(2), a011908. https://doi.org/10.1101/cshperspect.a011908

Peng, W., Li, J., Chen, R., Gu, Q., Yang, P., Qian, W., Ji, D., Wang, Q., Zhang, Z., Tang, J., & Sun, Y. (2019). Upregulated METTL3 promotes metastasis of colorectal Cancer via miR-1246/SPRED2/MAPK signaling

pathway. *Journal of experimental & clinical cancer research : CR*, *38*(1), 393. https://doi.org/10.1186/s13046-019-1408-4

Potenta, S., Zeisberg, E., & Kalluri, R. (2008). The role of endothelial-to-mesenchymal transition in cancer progression. *British journal of cancer*, *99*(9), 1375–1379. https://doi.org/10.1038/sj.bjc.6604662

Qiu, N., He, Y., Zhang, S., Hu, X., Chen, M., & Li, H. (2018). Cullin 7 is a predictor of poor prognosis in breast cancer patients and is involved in the proliferation and invasion of breast cancer cells by regulating the cell cycle and microtubule stability. *Oncology reports*, *39*(2), 603–610. https://doi.org/10.3892/or.2017.6106

Sas-Chen, A., Thomas, J. M., Matzov, D., Taoka, M., Nance, K. D., Nir, R., Bryson, K. M., Shachar, R., Liman, G., Burkhart, B. W., Gamage, S. T., Nobe, Y., Briney, C. A., Levy, M. J., Fuchs, R. T., Robb, G. B., Hartmann, J., Sharma, S., Lin, Q., Florens, L., … Schwartz, S. (2020). Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature*, *583*(7817), 638–643. https://doi.org/10.1038/s41586-020-2418-2

Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., & Sharma, K. K. (2010). Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research*, *1*(2), 109–126.

Sharma, S., Langhendries, J. L., Watzinger, P., Kötter, P., Entian, K. D., & Lafontaine, D. L. (2015). Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. *Nucleic acids research*, *43*(4), 2242–2258. https://doi.org/10.1093/nar/gkv075

Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., Hartwell, K., Onder, T. T., Gupta, P. B., Evans, K. W., Hollier, B. G., Ram, P. T., Lander, E. S., Rosen, J. M., Weinberg, R. A., & Mani, S. A. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(35), 15449–15454. https://doi.org/10.1073/pnas.1004900107

Thiery, J. P., Acloque, H., Huang, R. Y., & Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell*, *139*(5), 871–890. https://doi.org/10.1016/j.cell.2009.11.007

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., & Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell*, *117*(7), 927–939. https://doi.org/10.1016/j.cell.2004.06.006

Yue, B., Song, C., Yang, L., Cui, R., Cheng, X., Zhang, Z., & Zhao, G. (2019). METTL3-mediated N6-methyladenosine modification is critical for epithelial-mesenchymal transition and metastasis of gastric cancer. *Molecular cancer*, *18*(1), 142. https://doi.org/10.1186/s12943-019-1065-4

Zhang, X., Jiang, G., Sun, M., Zhou, H., Miao, Y., Liang, M., Wang, E., & Zhang, Y. (2017). Cytosolic THUMPD1 promotes breast cancer cells invasion and metastasis via the AKT-GSK3-Snail pathway. *Oncotarget*, *8*(8), 13357–13366. https://doi.org/10.18632/oncotarget.14528

Zhao, B. S., Roundtree, I. A., & He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nature reviews. Molecular cell biology*, *18*(1), 31–42. https://doi.org/10.1038/nrm.2016.132

Zhong, Y., Shen, S., Zhou, Y., Mao, F., Lin, Y., & Guan, J. (2022) NOTCH1 is a poor prognostic factor for breast cancer and is associated with breast cancer stem cells. *OncoTargets and Therapy*, 9(2016), 6865-6871. https://doi.org/10.2147/OTT.S109606

Zhu, L. J., & Altmann, S. W. (2005). mRNA and 18S-RNA coapplication-reverse transcription for quantitative gene expression analysis. *Analytical biochemistry*, *345*(1), 102–109. https://doi.org/10.1016/j.ab.2005.07.028