8-2016

¬¬DEFINE THE EPIGENETIC PROFILES AND SUBTYPE-SPECIFIC GENES OF BREAST CANCER

Wenqian Li

DEFINE THE EPIGENETIC PROFILES AND SUBTYPE-SPECIFIC GENES OF BREAST CANCER

By

Wenqian Li, Ph.D. Candidate

APPROVED:

_____

Sharon Y.R. Dent, Ph.D., Supervisory Professor

_____

Mark Bedford, Ph.D.

_____

Taiping Chen, Ph.D.

_____

Xiaobing Shi, Ph.D.

_____

Rick A. Finch, Ph.D.

APPROVED:

_____

Dean, The University of Texas

Graduate School of Biomedical Sciences at Houston

DEFINE THE EPIGENETIC PROFILES AND SUBTYPE-SPECIFIC GENES OF BREAST CANCER

A

Dissertation

Presented to the Faculty of

The University of Texas Health Science Center at Houston

And

The University of Texas M.D. Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the degree of

DOCTOR OF PHILOSOPHY

By

Wenqian Li, Ph.D. Candidate

Houston, Texas

August 2016

ACKNOWLEDGEMENTS

Ph.D. study is by no means an easy path for most people, at least for me. I definitely would not achieve the success without the support and help from so many people I met in this special five-year life journey.

First and foremost, I would like to thank my mentor, Dr. Sharon Dent, for everything. Four years ago, she kindly accepted me as her graduate student and has been a role model for me as a successful female scientist and an encouraging advisor. Sharon not only influences me with her passion and dedication to science, but also her way to keep the lab motivated, happy, and united. I learned from her in many ways, which I believe will benefit me in my future life and career.

I am very lucky to have a supportive committee, Dr. Mark Bedford, Dr. Taiping Chen, Dr. Xiaobing Shi, and Dr. Rick Finch. Every committee meeting helped shape my project better and more clearly. I appreciate a lot for their kindness throughout these four years.

I also feel fortunate to be in a lab with pleasant environment and willing-to-help members. Dr. Boyko Atanassov has always been my go to person from my rotation. He taught me many techniques and went through repeated discussion and interpretation of my experimental designs and results. Xianjiang Lan kindly taught me a number of biochemical experiments and exchanged ideas. Aimee Farria, who sit across my bench, is a great bench company and always cheers my days. Andrew Salinger is such a responsible lab manager to take care of my

routine orders and freezer emergencies. Xianghong Kuang, Li Wang, Dr. Andria Schibler, and Dr. Lia Koutelou, make me feel the lab is a home.

The love and support from my family is a huge motivation for me. My parents always respect my choice and trust my decision, as long as I live healthily and happily. My boyfriend, Chao Xie, always brings me positive attitude for work and life every day.

I am also grateful to Dr. Wei Li, Dr. Yuanxin Xi, Dr. Helen Piwnica-Worms, Dr. Shirong Cai, and Dr. Khandan Keyomarsi for collaborating in the breast cancer subtype project, and MB/NGS core in science park for their patience and consistent good attitude.

DEFINE THE EPIGENETIC PROFILES AND SUBTYPE-SPECIFIC GENES OF BREAST CANCER

Wenqian Li, Ph.D. Candidate

Supervisory Professor: Sharon Y.R. Dent, Ph.D.

Molecular profiling has identified 5 distinct subtypes of breast cancer, luminal A, luminal B, HER2-enriched, basal-like, and claudin-low breast cancer. These 5 subtypes correlate with hormone response, patient prognosis, and response to therapy. Although steady state gene expression patterns have been explored using expression microarrays, very little is known about the initial, disease-driving transcriptional changes in these cancers or epigenetic changes associated with the differential gene expression signatures. Defining these changes may provide new insights into the mechanisms by which these subtypes arise, as well as new avenues for breast cancer prevention, diagnosis, and treatment. Using Chromatin Immunoprecipitation sequencing and immunoblot technology, we investigated the genome wide profiling and global levels of eleven histone modifications localizing at promoter, gene body, and enhancer in distinct subtypes of breast cancer cell lines. With this knowledge, we will address how changes in chromatin modification landscape contribute to differential breast cancer subtypes, and identify the subtype-specific features and responses to gene regulation.

By examining the global levels of histone modifications in distinct subtypes, I found that H2Bub1 level is lower in claudin-low breast cancer cell lines compared with basal-like breast cancer cell lines. However, the mRNA levels of deubiquitinases and E3 ligases targeting H2Bub1 have no significant difference between basal-like and claudin-low subtypes. By analyzing the differential H2Bub1-enriched genes in claudin-low and basal-like subtypes, we found that miR200 family is highly enriched for H2Bub1 and has higher levels of expression in basal-like cells. miR200b overexpression decreases cell proliferation and colony formation specifically in claudin-low cells.

The main deubiquitinase for H2Bub1 is USP22 in the DUB module of the SAGA complex. USP22 is the catalytic subunit of the DUB module. Two adaptor proteins, ATXN7L3 and ENY2, are necessary for DUB activity of USP22 towards histone H2Bub1 and other substrates. ATXN7L3B shares 74% identity with the N-terminal region of ATXN7L3, but the functions of ATXN7L3B are not known. Here we report that ATXN7L3B interacts with ENY2 and USP22, but not other SAGA components. Even though ATXN7L3B localizes in the cytoplasm, ATXN7L3B overexpression increases H2Bub1 levels, while overexpression of ATXN7L3 decreases H2Bub1 levels. In vitro, ATXN7L3B competes with ATXN7L3 to bind ENY2, and in vivo, knockdown of ATXN7L3B leads to concomitant loss of ENY2. Unlike the ATXN7L3 DUB complex, the USP22-ATXN7L3B-ENY2 complex cannot deubiquitinate H2Bub1 efficiently in vitro. Moreover, ATXN7L3B knockdown inhibits migration of breast cancer cells in vitro

and limits expression of ER target genes. Collectively our studies suggest that ATXN7L3B may influence breast cancer cell behavior through effects on H2Bub1 levels and SAGA DUB activity.

From analyzing the histone modification ChIP-seq result, we found that specific histone modification patterns that are characteristic of five distinct breast cancer subtypes. We also identified subtype specific genes with unique chromatin signatures and expression patterns. For example, the AFAP1-AS1 gene locus is highly enriched in H3K9ac and H3K79me2 active transcription markers in TNBC cell lines. In addition, AFAP1-AS1 expression levels are higher in triple negative cells and tumors. Moreover, silencing of AFAP1-AS1 decreases cell proliferation and colony formation of TNBC cells.

Overall, these studies confirm and extend the importance of histone modifications to gene regulation in breast cancer and may provide new avenues for therapy development in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

SAGA: Spt-Ada-Gcn5-Acetyltransferase

ER: Estrogen Receptor

PR: Progesterone Receptor

HER2: Human Epidermal Growth Factor Receptor 2

TNBC: Triple Negative Breast Cancer

PHD: plant homeodomain

KMT: Lysine methyltransferase

HAT: Histone acetyltransferase

DUB: Deubiquitinase

EZH2: Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit

LSD1: Lysine-specific demethylase 1A

HDAC: Histone Deacetylase

CBP: CREB-binding protein

PCAF: P300/CBP-associated factor

TIP60: Tat Interacting Protein, 60kD

GCN5: General Control of Amino-Acid Synthesis 5-Like 2

USP22: Ubiquitin carboxyl-terminal hydrolase 22

ENY2: Enhancer of Yellow 2 Transcription Factor Homolog

TRF1: Telomeric repeat-binding factor 1

FBP1: far upstream element (FUSE)-binding protein 1

SIRT1: Sirtuin 1

RNF20/40: Ring Finger Protein 20/40

AFAP1: Actin Filament Associated Protein

AFAP1-AS1: AFAP1-Antisense RNA 1

EMT: Epithelial-Mesenchymal Transition

dNSAF: distributed normalized spectral abundance factor

Chapter 1: Introduction

1.1 Breast cancer clinical subtypes and therapies

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among females worldwide, accounting for 23% of the total cancer cases and 14% of the cancer deaths (1). In 2015, there were 2.8 million women affected by breast cancer (2). According to American Cancer Society, approximately 246,660 new cases of invasive breast cancer and 40,450 deaths are expected among US women in 2016 (www.cancer.org). Death rates from breast cancer have been declining since about 1989. These decreases are believed to result from earlier detection through better screening techniques and improved treatment.

One big challenge of breast cancer therapy is that breast cancer is a heterogenous disease, encompassing several subtypes, which have distinct clinicopathological features and clinical behaviors. In the clinic, breast cancer is classified based on the status of hormore receptors, estrogen and progesterone receptors (ER and PR), and HER2. The tumors of all breast cancer patients are routinely examined for these three markers using immunohistochemical staining. Patients with cancers that are ER and/or PR positive can be treated by hormone therapy. Estrogen promotes the growth of hormone receptor-positive cancers. The mechanisms of hormone therapy include lowering estrogen levels in the body (Aromatase inhibitors) and blocking estrogen from acting on breast cancer cells (Tamoxifen). Patients with Her2/neu positive breast cancers can be treated

with drugs that target the HER2/neu protein, most commonly monoclonal antibodies for HER2/neu protein, such as Trastuzumab and Pertuzumab. However, patients with cancers that are hormone receptor and HER2/neu negative, or triple negative, do not respond to hormone therapy or targeted therapies. Currently, there is no efficient targeted therapy for triple negative breast cancer, although radio-or chemo-therapy can be applied.

1.2 Breast cancer molecular subtypes

Phenotypic diversity of breast cancers defined by receptor status is useful when determining therapy plans but is also limiting since the heterogeneity of breast cancer goes beyond expression levels of ER, PR, and HER2. Diversity in gene expression patterns defined by cDNA microarrays provided better insights to the heterogeneity of breast cancer (3). 65 breast cancer specimens from 42 different patients were classified into normal-breast-like, luminal epithelial/ER+, HER2+, and basal-like subtypes according to the gene expression patterns of 456 human genes. Each subtype expresses a specific cohort of genes. In a further study, the luminal epithelial/ER+ subtype was divided into two subgroups, Luminal A and Luminal B, which express distinctive gene expression patterns and exhibit a significant difference in prognosis (4). Another study identified a new molecular subtype, claudin-low (5), which was classified into the basal-like subtype in a previous study. Claudin-low tumors express low levels of tight junction proteins claudin3, 4 and 7, and E-cadherin (5), exhibit a tumor initiating cell genomic

signature (6), and resemble mammary epithelial stem cell (7). Clinically, the majority of both basal-like and claudin-low tumors are ER, PR, and HER2 negative (triple negative), which cannot be treated by hormone therapy or targeted therapy.

Various breast cancer cell lines are useful models when defining the molecular mechanisms of breast cancer and developing novel potential treatment strategies. In order to make the studies in cell lines more accurate and efficient in predicting the mechanisms in human tumors, researchers also explored the expression profiles of a panel of breast cancer cell lines, and compared them to the expression patterns of breast cancer subtypes (7-9). About 50 breast cancer cell lines, based on their expression profile, are classified into comparable subtypes as in breast tumor tissues. Using corresponding breast cancer cell lines in certain subtype to study subtype-specific mechanism undoubtedly increases the reliability of mechanism-exploration and treatment-response.

1.3 Epigenetics: histone modification

Epigenetics refers to heritable events that alter the phenotype of a cell or organism without changing the DNA sequence or copy number, including DNA methylation, histone post translational modifications, and non-coding RNAs (10). These important epigenetics factors influence gene transcription states, mRNA splicing, DNA replication, DNA damage response, and other DNA-templated processes. Since a comprehensive DNA methylation profiling has been

investigated in human breast cancer samples (11), our study only focused on histone modification changes in breast cancer.

Nucleosomes are the fundamental units of chromatin in eukaryotes. The core particles of nucleosomes are composed of approximately 147 bp of DNA wrapping a histone octamer consisting of two of each histone, H2A, H2B, H3, and H4. Histone tails, protruding out from the nucleosome, are subject of a large variety of post-translational modifications, including acetylation, methylation, ubiquitination, phosphorylation among others (12). These histone modifications can affect interactions between nucleosomes, between histones and DNA, and between histones and trans-acting factors.

Different modifications have specific genomic locations. Most acetylation marks and some lysine methylation marks, such as H3K9ac, H3K23ac, H4K8ac, H4K16ac, H3K4 di- and tri- methylation (me2/3), and H3K9me2/3, are usually enriched around transcription start sites. H3K4me1 and H3K27ac are markers for active enhancers, while H3K27me3, H3K36me3, H3K79me3, and H2Bub1 are distributed within gene coding regions. In addition, different modifications are also associated with specific transcription states. All acetylation events, specific lysine methylation marks, and H2Bub1 predict active gene transcription, while H3K9me2/3 and H3K27me3 are usually associated with gene silencing. Histone modifiers, which can be called histone modification “writers” and “erasers”, catalyze the addition or removal of these modifications, affecting the states of

gene transcription, the processes of gene replication and repair, and overall chromatin states. Histone modification binding proteins, often referred as “readers”, contain specific domains that can recognize and bind to specific modifications. Proteins with chromodomains, Tudor domains, and PHD domains interact with specific lysine methylation events, while proteins with bromodomains can bind to several different lysine acetylation states.

1.4 Histone modification and cancer

Alterations in histone modifications, and their writers, erasers, and readers, have all been linked to both cancer development and cancer progression. Immunohistochemistry of H3K4me2 and H3K18ac in primary tissue specimens reveals that lower global levels of H3K4me2 and H3K18ac predict a higher risk of recurrence of prostate, lung, and kidney cancers (13, 14). Mutations or changes in expression of a number of writers, erasers, and readers are also associated with cancer. Gain of function and mutation of EZH2, a lysine methyltransferase (KMT) for H3K27, are associated with progression of distinctive cancer types (15). Given that EZH2 acts as a cancer driver, EZH2 inhibitors have been actively investigated in pre-clinical and clinical studies (15). Over expression of LSD1, a histone demethylase, has been observed in a broad spectrum of malignancies (16). Meta-analysis also revealed that high levels of LSD1 expression are correlated with poor overall survival in cancer patients (17). Changes in histone acetylation related proteins are also associated with cancer formation and development. Both histone acetyltransferases (HATs) and deacetylases

(HDACs), such as CBP/P300, PCAF, TIP60, and HDAC1/2, have been reported to promote cancer progression (18, 19). HAT inhibitors and HDAC inhibitors are also widely explored as possible cancer therapies (18, 19).

1.5 SAGA complex

GCN5 was the first HAT to be defined as important in transcription. GCN5 is part of SAGA, a multi-subunit complex that is functionally and structurally conserved from yeast to mammals. SAGA contains two enzymatic modules, the HAT and deubiquitinase (DUB) modules, and two other modules essential for its integrity and interactions with the transcriptional machinery, the TAF and SPT modules (20). The SAGA complex does not interact with DNA directly but can be recruited to chromatin by sequence-specific DNA-binding transcription factors, such as MYC, contributing to the active transcription of Myc target genes (21, 22). Gcn5 is the enzymatic protein in which the HAT domain resides. Recombinant yeast Gcn5 preferentially acetylates lysine 14 of H3 and lysine 8 and 16 of H4 of free histone substrates in vitro (23). However, yeast Gcn5 itself cannot acetylate histones in the form of nucleosomes (24). Only after incorporated into SAGA complex, yeast Gcn5 gains the ability to access and acetylate pre-assembled nucleosomes (25). On the other hand, mammalian GCN5 is able to acetylate histones incorporated into nucleosomes. Gcn5-null mouse embryos die at an early embryonic stage due to severe growth retardation (26, 27). Conditional neural progenitor cell-specific deletion of Gcn5 results in reduction of brain mass

and microcephaly (28). Moreover, Gcn5 depletion in SCA7 mice accelerates cerebellar and retinal degeneration (29).

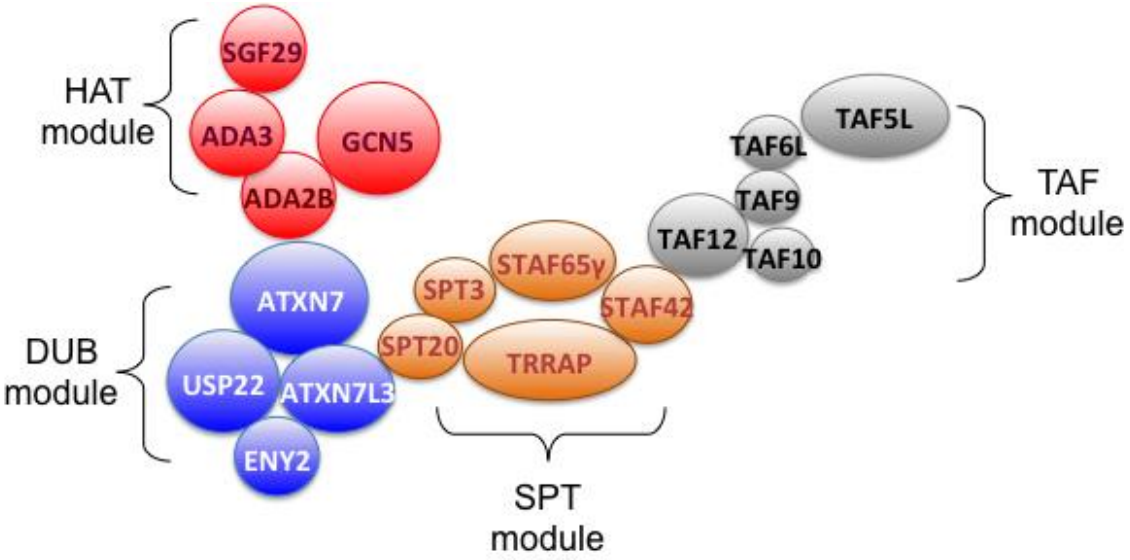


Figure 1. Schematics of SAGA complex.

SAGA complex is composed of four modules, HAT, DUB, SPT, and TAF modules. HAT and DUB modules possess acetyltransferase and deubiquitinase activities, respectively, with GCN5 and USP22 being the enzymes in each module. Adapted from (20).

1.6 DUB module

The DUB module consists of a catalytic subunit, USP22, and three other proteins, ATXN7, ATXN7L3 and ENY2, required for deubiquitinase activity (30-32). ATXN7, functions as a bridge to integrate the core DUB module into the greater SAGA complex (20, 33). USP22 alone has almost no activity until incorporated into SAGA complex (31, 32).

The monoubiquitination on histone H2A and H2B can be removed by the DUB module, and the ubiquitination status of these histones is associated with gene activation (H2Bub1) or silencing (H2Aub) (34). In mammals, monoubiquitination of H2B is catalyzed by the E2 conjugating enzymes HR6A/HR6B and the E3 ligase RNF20/RNF40 (35, 36). H2Bub1 was found to localize at transcribed regions of actively transcribed genes (37). H2B monoubiquitination is a highly dynamic process, with both ubiquitination and deubiquitination are required for optimal transcription (32, 38).

Functionally, USP22 is required during early embryogenesis in mice, since Usp22 null embryos die at 10.5 days and are significantly smaller in size (39). USP22 has also been widely indicated to play a role in cancer initiation and development. As early as 2005, USP22 was identified as a member of an 11 gene “death-from-cancer” signature for highly aggressive tumors (40). Elevated expression of USP22 has since been reported to be a prognostic factor for poor

survival in patients with colorectal cancer (41), breast cancer (42), pancreatic cancer (43, 44), cervical cancer (45), oral squamous cell carcinoma (46), hepatocellular carcinoma (47), and gastric carcinoma (48). Depletion of USP22 slows cell proliferation and cell cycle progression (30, 49). For molecular mechanisms, USP22 regulates the functions of transcription factors via acting as a co-activator, and deubiquitinate non-histone proteins to affect their stability or affinity. As a co-activator, USP22 can be recruited by the onco-protein Myc and is required for the transcription of Myc target genes (30). In addition, USP22 DUB activity is required for full transactivation of androgen receptor-dependent genes (31). However, USP22 is also required for castration-resistant AR expression and activity, thereby promoting castrate-resistant prostate cancer cell/tumor growth (50). By targeting non-histone proteins, USP22 also regulates the functions of proteins related to cancer. For example, USP22 regulates the expression of p21 by affecting the ubiquitination status of FBP1. Increased polyubiquitination of FBP1 resulted from ablation of USP22 does not decrease its protein levels, but compromises its ability to bind target loci and regulate downstream target genes (49). Additionally, depletion of USP22 increases ubiquitination of TRF1, a primary component of the telomeric shelterin complex, thereby leading to telomere dysfunction (51). USP22 can also stabilize Sirt1 by removing ubiquitins on Sirt1, thereby leading to increased deacetylation of p53 and suppression of p53-mediated functions (39). Moreover, USP22 expression is significantly positively correlated with BMI-1, c-Myc, and pAkt, and downregulation of USP22 up

regulation of BMI-1 mediated INK4a/ARF pathway but down regulation of Akt pathway (52).

1.7 ATXN7L3

ATXN7L3 contains ENY2-binding region in its N-terminus, ZnF-Sgf11 and SCA7 domains that reside in C-terminal region. The presence of ATXN7L3 is essential for the deubiquitinase activity of DUB module. No stable complex could be formed in the absence of ATXN7L3 (32). Moreover, the ZnF-Sgf11 domain of ATXN7L3 plays a pivotal role in the enzymatic activity, although is dispensable for the assembly, of the DUB module (32). The ZnF-Sgf11 domain of ATXN7L3 is essential for DUB activity towards H2Bub1 in vitro (32). Interestingly, others have reported that a truncated form of ATXN7L3 that only contains amino acids 3-76 interacts with USP22 and ENY2 in vitro, but the complex containing the truncated protein cannot remove monoubiquitination from H2Bub1 (32). However, the complex containing 3-76 truncated ATXN7L3 still exhibits activity since it could be bound by ubiquitin vinyl sulfone (Ub-VS), an irreversible and specific inhibitor of deubiqitinating enzymes that only binds to active deubiquitinase (32). The ZnF-Sgf11 domain is required for ATXN7L3 binding to nucleosomal DNA (53), and the crystal structure of the DUB module reveals an arginine cluster in the ZnF-Sgf11 domain directly interacts with ubiquitinated nucleosomes and H2A/H2B heterodimer (54).

ATXN7L3 can also bind to two other deubiquitinating enzymes, USP27x and USP51, to regulate the balance of the levels of H2Bub1 (55). These two novel DUB modules possess deubiquitinating activity but are independent of the SAGA complex. Furthermore, similar to USP22, depletion of USP27x or USP51 can also inhibit the growth of cancer cells and xenograft tumors.

1.8 Summary

As stated in Chapter 1.2, molecular profiling has identified 5 distinct subtypes of breast cancer, luminal A, luminal B, HER2-enriched, basal-like, and claudin-low breast cancer, which correlate with hormone response, patient prognosis, and response to therapy. Although use of expression microarray techniques has explored the steady state gene expression patterns among subtypes, very little is known about the initial, disease-driving transcriptional changes in these cancers or epigenetic changes associated with the different gene expression signatures. Using Chromatin Immunoprecipitation sequencing and immunoblot technology, we investigated the genome wide profiling and global levels of eleven histone modifications localizing at promoter, gene body, and enhancers in distinct subtypes of breast cancer cell lines. With this knowledge, we will address how changes in the chromatin modification landscape contribute to differential breast cancer subtypes, and identify the subtype-specific features and responses to gene regulation. Defining these changes may provide new insights into the mechanisms by which these subtypes arise, as well as new avenues for breast cancer prevention, diagnosis, and treatment.

Chapter 2: Materials and Methods

Antibodies

Antibodies used in this study include: anti-ATXN7L3B 75-97aa, 4331-1, Home Made; anti-ATXN7L3, Cat. # A301-800A, Bethyl Labs; anti-ENY2, Cat. # GTX128034, GeneTex; anti-USP22, 3933-1, Home Made; anti-GCN5, Cat. # 3305, Cell Signaling; anti-β-actin, Cat. # Sc-47778, Santa Cruz; anti-β-tubulin, Cat. # 2146S, Millipore; anti-H2Bub1, Cat. # 05-1312, Millipore; anti-H2B, Cat. # 07-371, Millipore; anti-Flag (M2), Cat. # F3165, Sigma; anti-HA, Cat. # 3724, Cell Signaling; anti-V5, Cat. # 46-0705, Invitrogen.

Cell culture

293T, MCF7, MDA-MB-231 cells were grown in high-glucose Dulbecco's Modified Eagle Medium (DMEM) (Thermo Scientific) supplemented with 10% Fetal Bovine Serum (FBS) (HyClone, Cat. # SH30910.03) and 1% penicillin-streptomycin (Thermo Scientific). MCF7T, HCC1937 cells were provided by Dr. Khandan Keyomarsi and were grown in alpha-Media (Minimum Essential Medium alpha (Corning, Cat. # 15-012-CV) supplemented with 10% FBS, 10mM HEPES, 1% non-essential amino acids, 2 mM glutamine, 1% sodium pyruvate, 1 ug/ml insulin, 1 ug/ml hydrocortisone, 12.5 ng/ml epidermal growth factor). Sf21 insect cells were grown in Sf-900^{TM} II SFM (Life Technologies, Cat. # 10902-096) supplemented with 1% penicillin-streptomycin. Cells were maintained in flasks (Corning, 250 mL Cat. # 431144 or 500 mL Cat. # 431145) on an orbital shaker at 26°C and 120 rpm. Sf21 cells were passaged every 3 days.

Protein Lysates

Adherent cells were trypsinized and neutralized in culture media, then pelleted by centrifugation at 200g for 5 min at room temperature (RT). Suspension Sf21 cells were harvested by centrifugation at 200g for 5 min. In both cases, cell pellets were then washed twice in ice-cold PBS containing protease inhibitors (Sigma protease inhibitor cocktail cat. #P8340). Washed cell pellets were lysed in Buffer C (20 mM Tris-HCl pH 7.9, 20% glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 0.2 mM EDTA, 1 mM Dithiothreitol (DTT), 1 mM Phenylmethylsulfonyl fluoride (PMSF), protease inhibitors) by pipetting vigorously several times, vortexed briefly and incubated on ice for 20 min. An equal volume of Buffer A (10 mM HEPES pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, 1 mM PMSF, protease inhibitors) was then added to the lysates. Cell lysates were sonicated for 5 min using Biorupter Twin (Diagenode, UCD-400) and then centrifuged at 12000 rpm for 10 min at 4°C. The cleared supernatants were taken as whole cell lysates (WCL).

Immunoprecipitation (IP)

After obtaining WCL, protein concentrations were measured by Bradford assays (Bio-Rad, Cat. # 500-0006). 0.8-1mg of total proteins were used for each IP. 30ul anti-FLAG M2 (Sigma cat. # A2220) beads were added to the lysates and incubated for 4 hours on a rocking platform at 4°C. After incubation, beads were centrifuged at 1000 rpm for 1 min, and washed in Wash buffer 150 (10 mM Tris-

HCl pH 7.9, 10% glycerol, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF, protease inhibitors) once and Wash buffer 350 (10 mM Tris-HCl pH 7.9, 350 mM NaCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF, protease inhibitors) once. In each wash, beads were incubated at 4°C for 5 min on a rocking platform and then centrifuged at 1000 rpm for 1 min. Beads with precipitated complexes were boiled in equal volume of 2x SDS sample buffer at 95°C for 10 min.

Immunoblots

20-40 µg of WCL or IP proteins were resolved on 4-12% NuPAGE gels (Life Technologies cat. # NW04122BOX). After electrophoresis, proteins were transferred to 0.2 µm nitrocellulose in Transfer buffer (25 mM Tris, 190 mM glycine, 10% Methanol) for 1 hour at constant 300 mA at 4°C. After blocked in 5% non-fat milk-TBST for 1 hour at RT, nitrocellulose membranes were incubated with primary antibodies overnight at 4°C. After three washes in TBST for 5 min, membranes were incubated with secondary-horseradish peroxidase (HRP) conjugated antibodies (GE Healthcare, Cat. # NA934V for Rabbit, NA931V for Mouse) for 45 min. After three washes in TBST for 5min, membranes were incubated with ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences, RPN2232) for 2 min and exposed.

Tandem IP sample preparation for MudPIT Analyzes

Sample preparation for MudPIT analyzes was performed as previously described in (Atanassov, 2016) 293T cells, stably expressing pINTO-N-FH-ATXN7L3, pINTO-N-FH-ATXN7L3B, or empty vector were grown in culture medium containing 100 µg/ml Zeocin ™ (Life Technologies cat. # R25001) in 150 mm plastic dishes at 37°C and 5% CO_2. Thirty approximately 80% confluent dishes were used for each purification.

Cells were harvested by trypsinization, washed twice in ice-cold PBS containing protease inhibitors, and pelleted by centrifugation for 5 min at 200g after each wash. Cell pellets were resuspended in 5 pellet volumes of Buffer C (20 mM Tris-HCl pH 7.9, 20% glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF, protease inhibitors) and lysates were incubated on ice for 30 min. Lysates were further homogenized using glass homogenizer with 15 strokes using a tight pestle. Homogenates were collected in fresh 50 mL falcon tubes and an equal volume of ice-cold Buffer A (10 mM HEPES pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, 1 mM PMSF, protease inhibitors) was added. To precipitate the crude cell debris, lysates were centrifuged for 15 min at 12,000g at 4°C. The supernatants were then moved to ultra centrifuge tubes, balanced, and centrifuged for 1.5 hours at 100,000g at 4°C using a Beckman 50.2Ti rotor. After this centrifugation the clear supernatants were moved to new Falcon tubes (without touching any pellets or clouds). After measuring the protein concentration a small fraction was saved as input.

Equal amounts of total protein for all samples were moved to 50 mL Falcon tubes and 500uL of anti-FLAG M2 beads, pre-washed in Wash buffer 150 (10 mM Tris-HCl pH 7.9, 10% glycerol, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF, protease inhibitors), were added to each tube. The bead-protein mixtures were incubated for 4 hours on a rocking platform at 4°C. After this incubation, the bead-protein mixtures were moved to chromatography columns (Bio-Rad Poly-Prep Cat. # 7311550) and solutions were allowed to flow through. The beads were then washed subsequently with 5 mL of Wash buffer 150, 5 mL of Wash buffer 350 (10 mM Tris-HCl pH 7.9, 350 mM NaCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF, protease inhibitors) and two more times with 5 mL of Wash buffer 150.

Precipitated complexes were eluted by incubating the beads in 2.5ml of Wash buffer 150 containing 150 µg/mL 3xFLAG peptide (Sigma Cat. # F3290) for 30 min on a rocking platform at 4°C. Elutes were collected in 15 mL Falcon tubes and the elution process was repeated one more time using 2.5ml of Wash buffer 150. The same sample elutes were pooled in one 15 mL Falcon tube. After the elution using 3XFLAG peptide, 100 uL of pre washed anti-HA beads (Roche Cat. # 11815016001) were added to the elutes and mixtures were incubated overnight on a rocking platform at 4°C. After the second IP using anti-HA beads, the beads were washed as after the first IP-1 time with Wash buffer 150, 1 time with Wash buffer 350 and 2 times with Wash buffer 150. Precipitated complexes were eluted with 500 uL of 100 mM glycine (pH 2.0) for 15 min on a rocking platform at 4°C.

The eluates were collected in 15 mL falcon tubes containing 50 uL of 1M Tris-HCl, pH 9.0, to neutralize the acidic eluates. The elution was repeated one more time and eluates collected in new 15 ml Falcon tubes containing 50 uL of 1M Tris-HCl, pH 9.0. The same sample eluates were pooled, and 10% of the pooled sample (100 uL) was saved as input, for use silver staining and immunoblotting following gel electrophoresis. Inputs were further concentrated down to ~25-30 uL volume using centricons (Amicon Ultra, Millipore Cat. # UFC 501024, 10K cut size). For MudPIT samples, the final eluates were treated with Benzoase (Sigma Cat. # E8263-5KU) 0.1U/tube on 37°C for 30 min. Eluates were then aliquoted into several tubes, volumes brought to 400 uL with 100 mM Tris-HCl pH 8.5 and precipitated immediately by adding 100 uL of ice-cold trichloroacetic acid (TCA) (Sigma, Cat. # T6399). Eluate-TCA mixtures were incubated at 4°C overnight. After this incubation, precipitated proteins were pelleted by centrifugation at 21,000g for 20 min at 4°C followed by 1 immediate centrifugation at 21,000g for 10 min at 4°C. Pellets were then washed twice with ice-cold acetone, air-dried in a fume hood and submitted for MudPIT analysis.

Transfection and viral infection

When 293T cells reached 60% confluence, pINTO-N-FH empty vector, pINTO-N-FH-ATXN7L3, or pINTO-N-FH-ATXN7L3B, were transfected using Lipofectamine® 2000 (Life Technologies cat. # 11668019) following manufacturer's instructions. Six to eight hours after transfection, the medium was changed to normal culture medium as described above.

Viral medium was produced using 293T cells co-transfected with lentiviral or retroviral vectors and psPAX.2 and pMD2.G (Addgene) for lentiviral production or pCL-Ampho (Addgene) for retroviral production. 48 hours post transfection, the culture medium containing the viral particles (viral medium) was collected, filtered through 0.45 µm filters and used for infection. For infection of cells, the viral medium was diluted 1:3 with the culture medium containing 8 ug/ml polybrene (Sigma, Cat. # H9268).

Stable cell lines generation

After transfection of 293T cells with pINTO-N-FH-ATXN7L3, pINTO-N-FH-ATXN7L3B or pINTO-N-FH empty vector, cells were subsequently selected in culture medium containing 300 µg/ml Zeocin for approximately 2 weeks. Then 300 cells were plated on 10cm dishes and allowed to form single colonies. After 2 weeks, single colonies were picked and grown in a 96 well plate, followed by expansion into 24 well plates and then 6 well plates. Expressed proteins were detected as FLAG signals in WCLs using immunoblot. Stable lines were maintained in culture medium containing 100µg/ml Zeocin.

To achieve stable cell lines expressing pGIPZ shRNAs, MCF7T cells were infected with viral medium containing pGIPZ, shATXN7L3, shATXN7L3B lentivirus. 48 hours later, cells were selected in culture medium containing 2 µg/mL Puromycin (Calbiochem cat. # 540222) for 48 hours. Stable lines were maintained in culture medium containing 2 ug/mL Puromycin.

Immunofluorescence

Immunofluorescence assay was performed as previously described. (56) Appropriate amounts of cells were seeded onto sterile coverslips placed in 6 well plates 24 hours before staining. Cells were rinsed with PBS once and fixed with 3% formaldehyde in PBS for 15 min at RT, followed by three rinses with PBS. Then cells were permeabilized with 0.5% Triton X-100 for 5 min at RT and rinsed again with PBS three times. After blocking with 10% FBS in PBS for 45 min at RT, cells were incubated with anti-Flag antibody for 1 hour at RT, followed by three rinses with PBS. Then cells were incubated with a secondary anti-mouse antibody for 1hour at RT and rinsed three times with PBS. 1 mL 100% ethanol was added onto the cells for 1 min and then removed. Slides were dried for 5min at RT, coverslips were removed from the 6 well plate and mounted onto a microslide using Vectashield mounting medium with DAPI (Vector Laboratories, Cat. # H-1200). Images were acquired using a laser scanning spectral confocal microscope (Leica STP6000).

Baculovirus expression system

Baculovirus transfection, infection, and maintenance were performed as previously described in (Atanassov, 2016). Sf21 cells were co-transfected with 3 ug of pBacPAK8.3 expression vectors, containing the indicated cDNAs, and 0.5ug of linearized baculovirus DNA (BestBac 1.0, Expression systems, Cat. # 91-001). Cells were transfected using Lipofectamine 2000 following manufacturer's instructions. 72 hours after the transfection, the medium

containing the recombinant viral particles was collected. This medium was further used (in 1:5 dilution) to infect new cells. For infection, exponentially growing Sf21 cells were seeded in 15 cm dishes ($2x10^7$ cells/dish) and viral medium was added after cells were attached to the dish. Cells were incubated in a humid chamber (plastic box with wet paper towels at the bottom) for 72 hours at 26°C. The viral medium was then collected, and stored in dark boxes at 4°C for several weeks, or at -80°C for longer periods. To increase the viral titer, several rounds of viral infection followed by medium collection were conducted. The expression level of each cDNA was monitored by western blot in the end of each round, using WCL from infected cells.

Protein purification from Baculovirus system

Large-scale protein purification from Baculovirus system was performed as previously described. (56) $4x10^7$ cells were infected and maintained in flasks in an orbital shaker for 3 days at 27°C and 120 rpm. Cells were harvested at 1500 rpm for 10 min and lysed in 3 ml Buffer C for protein purification (50 mM HEPES, pH 7.9, 20% glycerol, 300 mM NaCl, 5 mM $MgCl_2$, 0.1% NP-40, 1 mM DTT, 1 mM PMSF, protease inhibitors). The suspensions were then homogenized by bead beating 3 times for 30 sec each time on ice (Mini-Beadbeater, Biospec Products). The supernatant was recovered by centrifuging at 15,000 rpm for 10 min, and then further cleared by being centrifuged at 40,000 rpm for 1 hour at 4°C. The supernatant was then incubated with 80 ul of anti-Flag M2, anti-HA matrix (Roche, Cat. # 1181506001), or anti-V5 Agarose Affinity Gel (Sigma, Cat.

A7345) equilibrated with Wash buffer (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% Triton X-100, 300 mM NaCl, 1 mM DTT, 1 mM PMSF, protease inhibitor) overnight at 4°C. After being washed with Wash buffer 3 times for 10 min each time, bound proteins were eluted with 200 ug/ml FLAG peptide (Sigma, Cat. # F3290) at 4°C or 400 ug/ml HA peptide (Sigma, Roche, Cat. # 11666975001) at RT or 700 ug/ml V5 peptide (Sigma, Cat. # V7754) at RT for 2hr in Elution Base buffer (10 mM HEPES, pH 7.9, 100 mM NaCl, 1.5 mM $MgCl_2$, 0.05% Triton X-100, 0.2 mM EDTA, 10% glycerol, 0.5 mM DTT, 0.2 mM PMSF, protease inhibitor). The eluted complexes or recombinant proteins were further purified through a gel filtration Superdex 200 column (GE Healthcare).

In vitro deubiquitination assay

In vitro deubiquitination assay was performed as previously described (56). Total free histones were extracted from 293T cells using a histone purification mini kit (Active Motif, Cat. # 40026) per manufacturer's instructions. Free histones were incubated with purified recombinant USP22 or deubiquitinase complexes in reaction buffer (100 mM Tris-HCl, pH 8.0, 5% glycerol, 1 mM EDTA, 3 mM DTT) for 2 hours at 37°C. ¼ volume of 5X SDS loading buffer was added and the mixture was boiled to stop the reaction. Immunoblots were performed to assess ubiquitination levels of the histones.

In vitro competition assay

Individual recombinant proteins were purified from Sf21 cells. Equal amounts of V5-ATXN7L3 and FLAG-ENY2, and increasing amounts of V5-ATXN7L3B were added into 1 ml wash buffer (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% Triton X-100, 300 mM NaCl, 1 mM DTT, 1 mM PMSF, protease inhibitor). Reactions were incubated in 4°C overnight on a rocking platform. After adding anti-FLAG M2 beads, reactions were incubated in 4°C for another 1 hour. Then bead-protein complexes were washed three times with wash buffer and boiled in equal volume of 1X SDS loading buffer. The supernatants were analyzed by immunoblot.

Subcellular fractionation

Subcellular fractionations were done following published protocols. (57) Cells were re-suspended in 200 ul of Buffer A (10 mM HEPES, pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 0.1% Triton X-100, 1 mM DTT, 1 mM PMSF, protease inhibitor) to separate nuclei from a crude cytoplasmic fraction, which was further clarified by high-speed centrifugation to remove cell debris and insoluble aggregates. The supernatant was designated the cytoplasmic fraction and mixed with ¼ volume of 5X SDS loading buffer and boiled. Nuclei were washed with Buffer A and resuspended in 200ul of 2X SDS loading buffer, boiled for 10min, and sonicated for 15 sec at 25% amplitude using EpiShear Probe Sonicator (Active Motif, Model CL-18).

RNA extraction, Reverse transcription and quantitative real-time PCR (qRT-PCR)

Total RNA was isolated using an RNeasy Mini kit (Qiagen, Cat. # 74104), following the manufacturer's recommended procedure. 1 ug of RNA was reverse-transcribed to cDNA using a SuperScript VILO cDNA Synthesis kit (ThermoFisher Scientific, Cat. # 11754050) following the manufacturer's procedure. 10 ng of cDNA was used for one qRT-PCR reaction with three technical replicates.

Transwell migration assay

MCF7T cells were used for transwell migration assays. Transwell chambers with transparent PET membranes (Falcon, Cat. # 353097) were placed in 24 well plates with normal culture medium, alpha-Media. Then 80,000 cells suspended in alpha-MEM containing 0.1% BSA were placed on the upper layer of the chambers. After 24 hours of incubation at 37°C with 5% CO_2, the membranes at the bottom of chambers were fixed with methanol for 5 min and stained with 0.5% crystal violet for 10 min. Then the cells above the membranes were removed with cotton swabs. Membranes were dried and cells attached below the membranes were imaged using microscope (ZEISS, Axiovert 40 CFL).

Cell Proliferation assay

15,000 cells were seeded into each well of 12 well plate with three replicates for each sample and each day of counting. At indicated day, cell numbers were counted using hemocytometer.

Colony formation assay

Stable MCF7T pGIPZ, shATXN7L3, shATXN7L3B expressing cells were trypsinized thoroughly, neutralized with alpha-Media, and counted 3 times for each sample. 300 cells for each sample were put into single wells of a 6 well plate, with three replicates, and incubated at 37°C with 5% CO_2. 20~30 days later, cells were washed with PBS once and fixed with Fixation buffer (Acetic acid/methanol, 1:7 v/v) for 5 min. Then cells were stained with 0.5% crystal violet for 2 hours to overnight, washed, dried, and imaged.

Chromatin Immunopreciptation

Approximately 4 x 10^7 cells were crosslinked with 1% formaldehyde in the culture media for 10 min at RT, quenched with 125 mM glycine and washed twice with PBS. The cells were scraped into cold PBS containing protease inhibitors and pelleted at 2,500 rpm for 8 min. Cell pellets were lysed in cell lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP40, 1 mM DTT, 1 mM PMSF, protease inhibitor), incubated on ice for 10 min, span down at 5,000 rpm for 5 min at 4°C. Then the nuclear pellets were lysed in 200 ul nuclei lysis buffer (50 mM Tris-Cl pH 8.0, 10 mM EDTA, 1% SDS, 1 mM DTT, 1 mM PMSF, protease inhibitor). The nuclear lysates were transferred into TPX tubes (Diagenode, Cat. # M-

50050) and incubated on ice for 5 min. Chromatin were sonicated into 200-500 bp fragments using Bioruptor Twin. Sonicated products were span at 13,000 rpm for 10 min at 4°C. The supernatants were transferred to a new tube and diluted 10X with ChIP-dilution buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.01% SDS, 1 mM DTT, 1 mM PMSF, protease inhibitor), and pre-cleared with 30 ul of BSA-blocked Agarose G beads (Millipore, Cat. # 16-266) for 30 min at 4°C. Agarose G beads were removed by centrifugation at 2,000 rpm for 1 min. Supernatants were transferred to new tubes and 40 ul were saved as 2% input. 4 ug of antibodies were added into pre-cleared chromatin solutions and incubated overnight at 4°C on a rocking platform. Then 30 ul BSA-blocked Agarose G beads were added into ChIP reactions and incubated for 1 hour at 4°C. Beads-chromatin complexes were span down at 2,000 rpm for 1 min and washed consecutively with 1 ml of Low salt wash buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 1 mM DTT, 1 mM PMSF, protease inhibitor), High salt wash buffer (20 mM Tris-Cl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 1 mM DTT, 1 mM PMSF, protease inhibitor), LiCl wash buffer (20 mM Tris-HCl pH 8.0, 250 mM LiCl, 1mM EDTA, 1% NP40, 1% Na-deoxycholate, 1 mM DTT, 1 mM PMSF, protease inhibitor), and TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, protease inhibitor). For each wash, reactions were rotated for 5 min at 4°C and spined down at 1,000 rpm for 1min at 4°C. Immunoprecipitated chromatin were eluted from beads twice with 50 ul freshly prepared IP elution buffer (50 mM $NaHCO_3$, 1% SDS) and de-crosslinked by adding 8 ul of 4M NaCl,

followed by incubating at 65°C for overnight. For the 2% Input, 60 ul of IP elution buffer and 8 ul of 4M NaCl were added, followed by incubating at 65°C for overnight. 1 ul of 10 mg/ml RNase A were added to each sample and incubated for 1 hour at 37°C. Then DNAs were purified using Qiagen PCR purification kit following manufacturer's instructions.

Plasmids

Plasmid	Vector	Tag
pINTO-NF-H-ATXN7L3	pINTO-NF-H	N-Flag-HA
pINTO-NF-H-ATXN7L3B	pINTO-NF-H	N-Flag-HA
pBacPAK8.3-V5-ATXN7L3	pBacPAK8.3	N-V5
pBacPAK8.3-V5-ATXN7L3B	pBacPAK8.3	N-V5
pBacPAK8.3-Flag-ENY2	pBacPAK8.3	N-Flag
pBacPAK8.3-HA-USP22	pBacPAK8.3	N-HA
pBABE-Flag-ATXN7L3B	pBABE-hygro	N-Flag

Chapter 3: Define epigenetic profiles of breast cancer subtypes

3.1 Define genome-wide locations of histone modifications in breast cancer subtypes.

To explore the genetic distribution of histone modifications in distinct subtypes of breast cancer, the ChIP next generation sequencing (ChIP-Seq) technique was adopted. We selected 11 histone modifications for this study (Table 1), in which I personally performed the ChIP-Seq for acetylation and ubiquitination marks. They localize at different parts of genes, including promoter, gene body, and enhancer. They are associated with either active or repressed gene transcription state. The epigenome differences in distinct breast cancer subtypes were studied using cell lines that typify the five molecular subtypes, Luminal A, Luminal B, HER2-enriched, Basal-like, and Claudin-low (Table 2). Normal, immortal mammary cell lines were used as controls. We performed ChIP-seq assay for each histone modification across all 14 cell lines (Table 3). By analyzing the histone modification ChIP-seq data, we defined 13 chromatin states, which were used to define the subtype specific patterns. The results showed that we can classify breast cancer subtypes using distinct chromatin states.

Table 1. Selected histone modifications

Histone Modifications	State of gene transcription	Genetic localization
H3K9ac	Activation	Promoter
H3K23ac		
H4K8ac		
H3K4me3		
H3K9me3	Repression	
H3k27me3		
H2BK120ub	Activation	Gene Body
H3K36me3		
H3K79me2		
H3K27ac		Enhancer
H3K4me1		

Table 2. Selected cell lines of normal and five subtypes of breast cancer.

	Cell Line	ER	PR	HER2	Subtype
1	76NF2V	-	-	-	Normal
2	MCF-10A	-	-	-	
3	MCF-7	+	+	-	Luminal A
4	MCF-7T	+	+	-	
5	ZR75-1	+	-	-	
6	MDA-MB-361	+	-	+	Luminal B
7	UACC812	+	-	+	
8	SKBR3	-	-	+	HER-2
9	AU565	-	-	+	
10	HCC1954	-	-	+	
11	MDA-MB-231	-	-	-	Claudin Low
12	MDA-MB-436	-	-	-	
13	MDA-MB-468	-	-	-	Basal-like
14	HCC1937	-	-	-	

Table 3. ChIP-seq table.

			Promoter						Gene Body			Enhancer			
			Activation				Repression		Activation			Activation			
	Cell Line	Subtype	H3K9ac	H3K23ac	H4K8ac	H3K4me3	H3K9me3	H3K27me3	H2B-ub	H3K36me3	H3K79me2	H3K27ac	H3K4me1	H3	Input
1	76NF2V	Normal	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
2	MCF-10A		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
3	MCF-7	Luminal A	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
4	MCF-7T		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
5	ZR75-1		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
6	MDA-MB-361	Luminal B	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
7	UACC812		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
8	SKBR3	HER-2	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
9	AU565		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
10	HCC1954		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
11	MDA-MB-231	TNBC/Claudin Low	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
12	MDA-MB-436		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
13	MDA-MB-468	TNBC/Basal-like	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
14	HCC1937		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑

3.2 Investigate the global protein levels of histone modifications and SAGA components in breast cancer cell lines.

Since global levels of histone modifications are correlated with prognosis of prostate cancers (13, 14), we wondered whether global levels of histone modifications have specific patterns in distinct subtypes of breast cancer. By immunoblot assays, we found that H2Bub1 levels were lower in claudin-low cell lines compared with basal-like cell lines (Figure 2). However, for other modifications, there were no obvious patterns across breast cancer subtypes.

To examine the expression levels of GCN5 and USP22, which are two enzymatic proteins in SAGA complex, across 14 cell lines, we performed immunoblots. We found that the protein levels showed big variation across cell lines in each subtype and there were no clear patterns across subtypes (Figure 3).

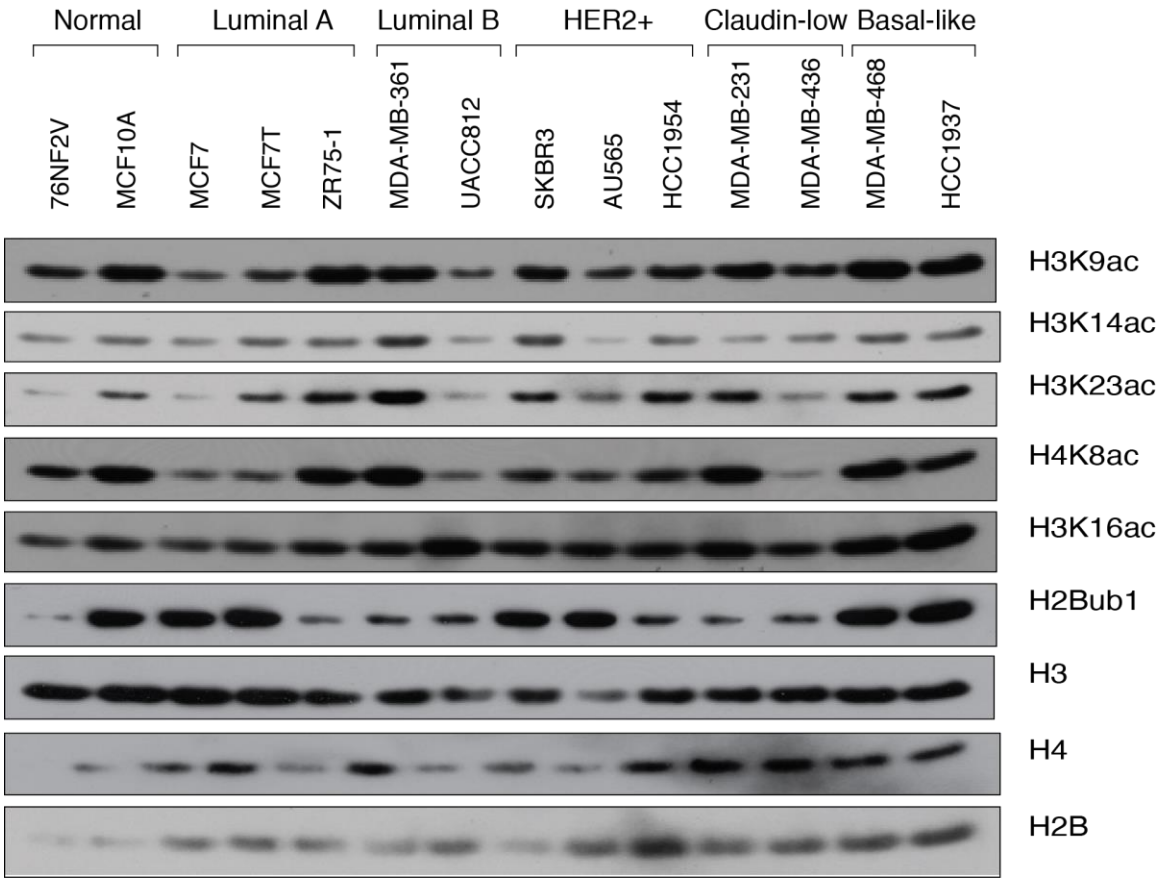


Figure 2. Histone modification levels across 14 cell lines.

Histones were extracted and separated in 4-12% gradient SDS PAGE. then probed by immunoblot with indicated antibodies.

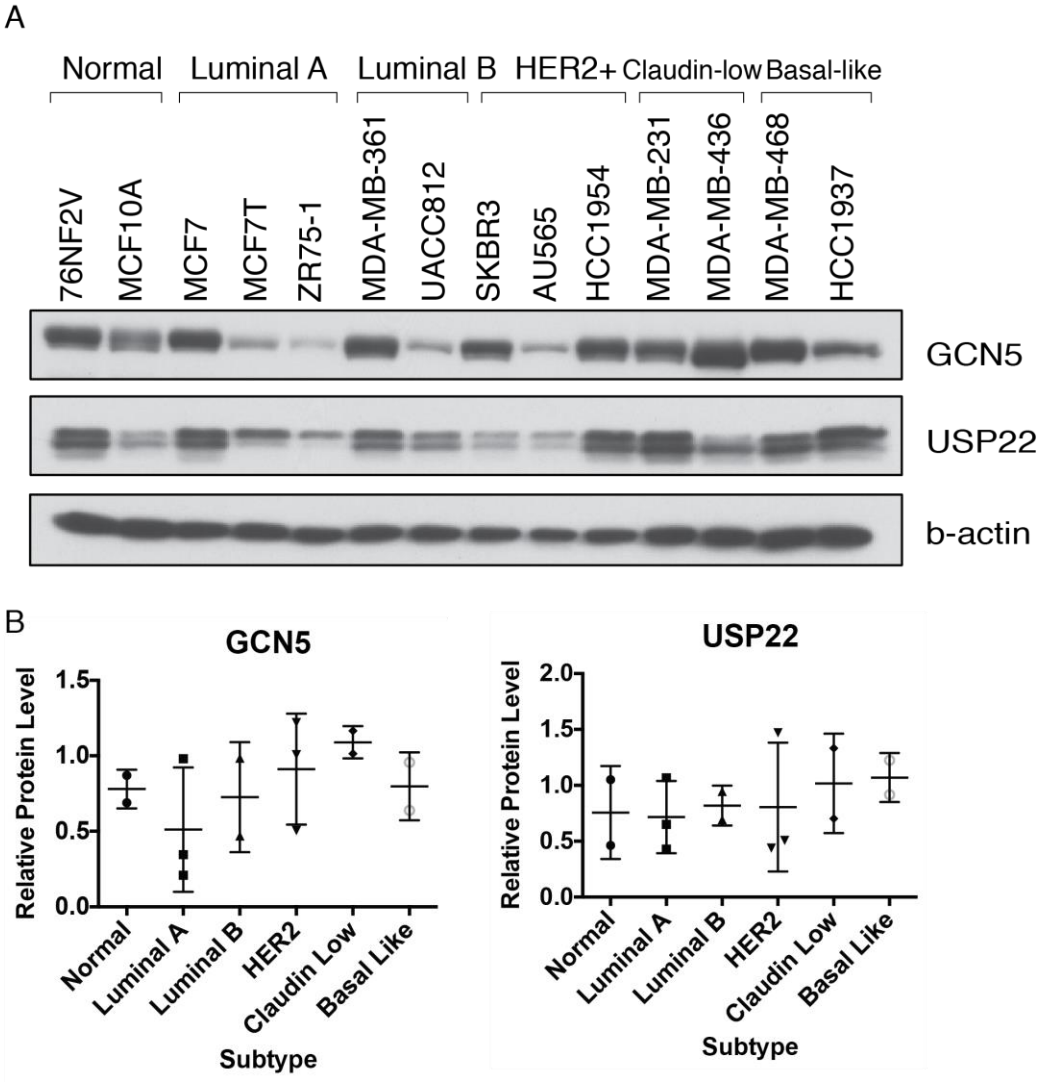


Figure 3. The protein levels of enzymatic proteins in SAGA complex.

A. Whole cell lysates from each cell line were resolved by electrophoresis. Proteins were probed by immunoblot with indicated antibodies. B. Protein signals in panel A were quantified using ImageJ and normalized to β-actin.

3.3 Conclusions

By examining the global levels of histone modifications in 14 cell lines, we found that H2Bub1 levels were lower in two claudin-low cell lines compared with two basal-like cell lines. Considering that both claudin-low and basal-like subtype are triple negative breast cancer, we are interested in characterizing the subtype-specific features in order to provide potential therapy target and detangle the heterogeneity of triple negative breast cancer. The more detailed investigation on lower H2Bub1 levels in claudin-low subtype will be discussed in Chapter 4. Then, a novel H2Bub1 regulator, ATXN7L3B, and its other functions will be discussed in Chapter 5. Finally, the findings discovered from the ChIP-Seq data will be described and discussed in Chapter 6.

Chapter 4: H2Bub is lower in claudin-low subtype than basal-like subtype

4.1 H2Bub1 level is lower in claudin-low than basal-like

By examining the protein levels of histone modifications in 14 mammary cell lines, we found that even though there were no clear patterns across all subtypes, in four TNBC cell lines, two claudin-low cell lines, MDA-MB-231 and MDA-MB-436, express lower levels of H2Bub1 compared with two basal-like cell lines, MDA-MB-468 and HCC1937 (Figure 4A). The H2Aub levels, however, were not different between two subtypes. Four triple-negative HIM breast cancer tumors, HIM2, HIM4, HIM5, and HIM6, were obtained from our collaborator, Dr. Helen Piwnica-Worms (UTMDACC). Two tumors, HIM2 and HIM5, according to their high expression in Vimentin and Twist, are classified as claudin-low tumors. Two other tumors, HIM4 and HIM6, expressing high levels of claudin-3 and E-cadherin, are classified as basal-like tumors (Figure 4B). Interestingly, H2Bub1 levels are significantly lower in two claudin-low tumors compared with two basal-like tumors, but H2Aub1 levels do not have this difference (Figure 4B). We then sought to determine the molecular basis of this difference by checking the mRNA levels and protein levels of several E3 ligases and deubiquitinases. However, no one mRNA or protein change for these factors was observed in these 4 cell lines and 4 tumors, so at present we cannot explain the basis for the difference in H2Bub1 levels (Figure 4C,D,E).

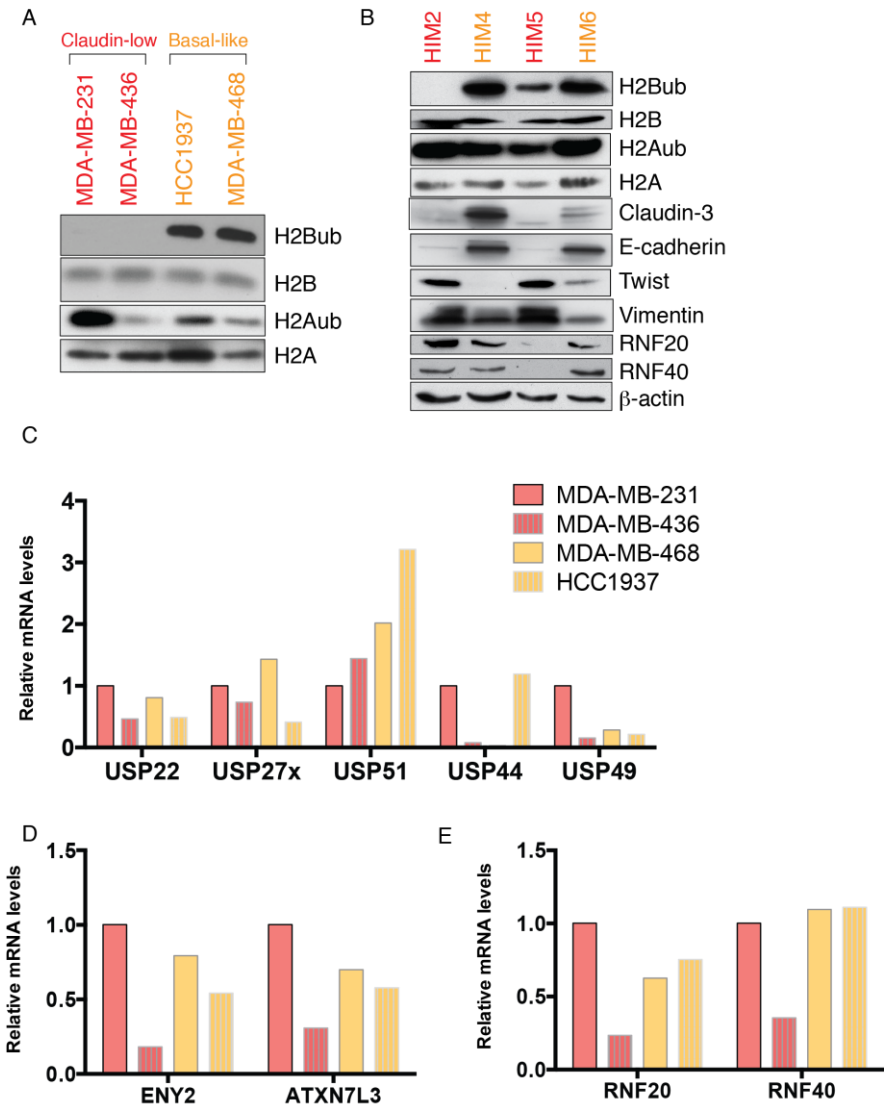


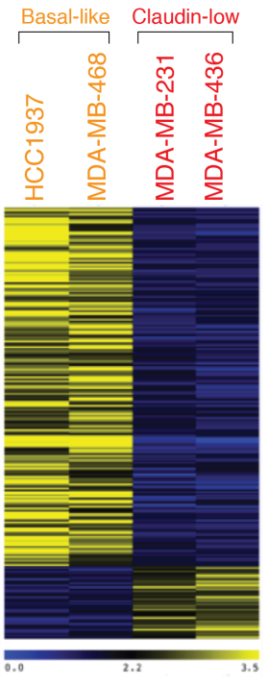
Figure 4. H2Bub1 level is lower in claudin-low cells than basal-like cells.

A. Histones from indicated four cell lines were extracted and separated in 4-12% gradient SDS-PAGE. Modified and non-modified histones were probed by immunoblots with indicated antibodies. B. Histones and whole cell lysates from indicated four HIM tumors were resolved using electrophoresis and probed by immunoblots with indicated antibodies. C.D.E. RNA were extracted from indicated four cell lines and reverse-transcribed to cDNAs, which were used for quantitative real-time PCR.

4.2 Differential H2Bub enriched genes in claudin-low vs basal-like subtype

Lower H2Bub1 levels have been reported in pluripotent stem cells (58). Also, claudin-low tumors express higher levels of genes correlated with stemness (59). Therefore, we hypothesized that the lower H2Bub1 levels in claudin-low breast cancer cells and tumors are correlated with the stemness of claudin-low tumors. We analyzed data from H2Bub1 ChIP in four triple negative cell lines followed by next generation sequencing to obtain the differential H2Bub1 enriched genes in claudin-low versus basal-like subtypes (Figure 5A, Appendix 1 and 2). Each subtype has a specific cohort of genes that are highly enriched for H2Bub1. Then we selected the genes with the biggest changes in H2Bub1 enrichment between two subtypes (Figure 5B, C). 187 genes are highly enriched for H2Bub1 in basal-like cells. The average of values in two basal-like cells and two claudin-low cells are compared and the genes with larger than 5-fold difference were selected (Figure 5B). 39 genes are highly enriched for H2Bub1 in claudin-low cells and the genes with larger than 3-fold difference were selected (Figure 5C). The functions of most of these genes are either correlated with tumor suppression and tumor promotion. To confirm the RNA levels of these genes, we performed real-time PCR to examine the mRNA and microRNA levels in these four cell lines (Figure 5D, E). The mRNA and microRNA levels of these genes are consistent with the H2Bub1 enrichment.

A



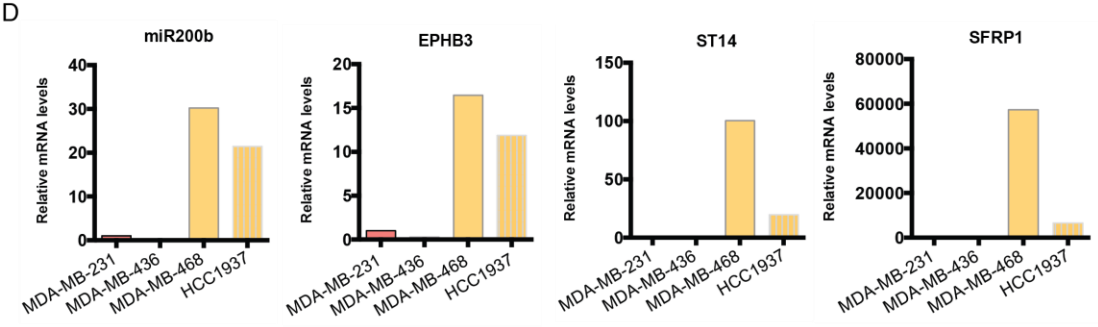
B

Gene	Claudin-Low		Basal-Like		Function
	MDA-MB-231	MDA-MB-436	MDA-MB-468	HCC1937	
MIR200B	0.46	0.20	18.79	5.92	inhibit EMT and suppress metastasis
MIR200A	0.72	0.12	15.05	5.17	
MIR429	0.77	0.29	13.74	3.36	
EPHB3	0.91	0.47	4.10	4.44	suppress metastasis
ST14	0.86	0.41	4.09	2.84	predict poor survival
SFRP1	0.49	0.48	2.61	4.01	inhibit canonical Wnt signaling, loss of function in cancer

C

Gene	Claudin-Low		Basal-Like		Function
	MDA-MB-231	MDA-MB-436	MDA-MB-468	HCC1937	
MIR645	3.49	2.78	0.49	1.50	unknown
FGFRL1	3.63	2.65	0.71	0.94	promote cancer cell proliferation
BIN1	2.45	2.69	0.73	0.94	cancer suppression

D



E

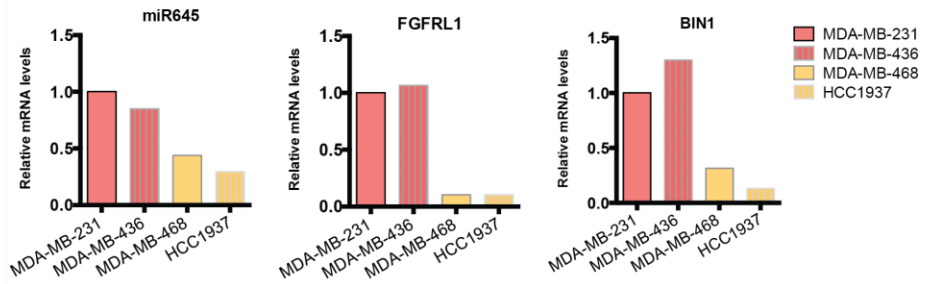


Figure 5. Differential H2Bub1-enriched genes in claudin-low versus basal-like cells.

A. Heatmap for differential H2Bub-enriched genes in two basal-like cell lines and two claudin-low cell lines. B. Six selected genes highly enriched for H2Bub in basal-like cells with the largest difference compared to claudin-low cells. C. Three selected genes highly enriched for H2Bub in claudin-low cells with largest difference compared to basal-like cells. D.E. RNA were extracted from indicated four cell lines and reverse-transcribed to cDNAs, which were used for quantitative real-time PCR to examine the expression levels of genes listed in panel B and C.

4.3 miR200b specifically repressed proliferation of claudin-low cells

Among these genes, the miR200 family has been reported to function in tumor suppression via inhibiting EMT by targeting the E-cadherin transcriptional repressors ZEB1 and ZEB2. In addition, miR200 family expression is also strongly decreased in breast cancers with EMT features. Compared with basal-like cells, claudin-low cells express lower levels of miR200a, miR200b, and miR429. To find out whether over expression of miR200 family members has specific effects on claudin-low cells, we overexpressed miR200b in four triple negative cell lines (Figure 6A). Interestingly, we found that overexpression of miR200b has a specific effects on claudin-low cells to inhibit cell proliferation (Figure 6B, C) and colony formation (Figure 6D, E).

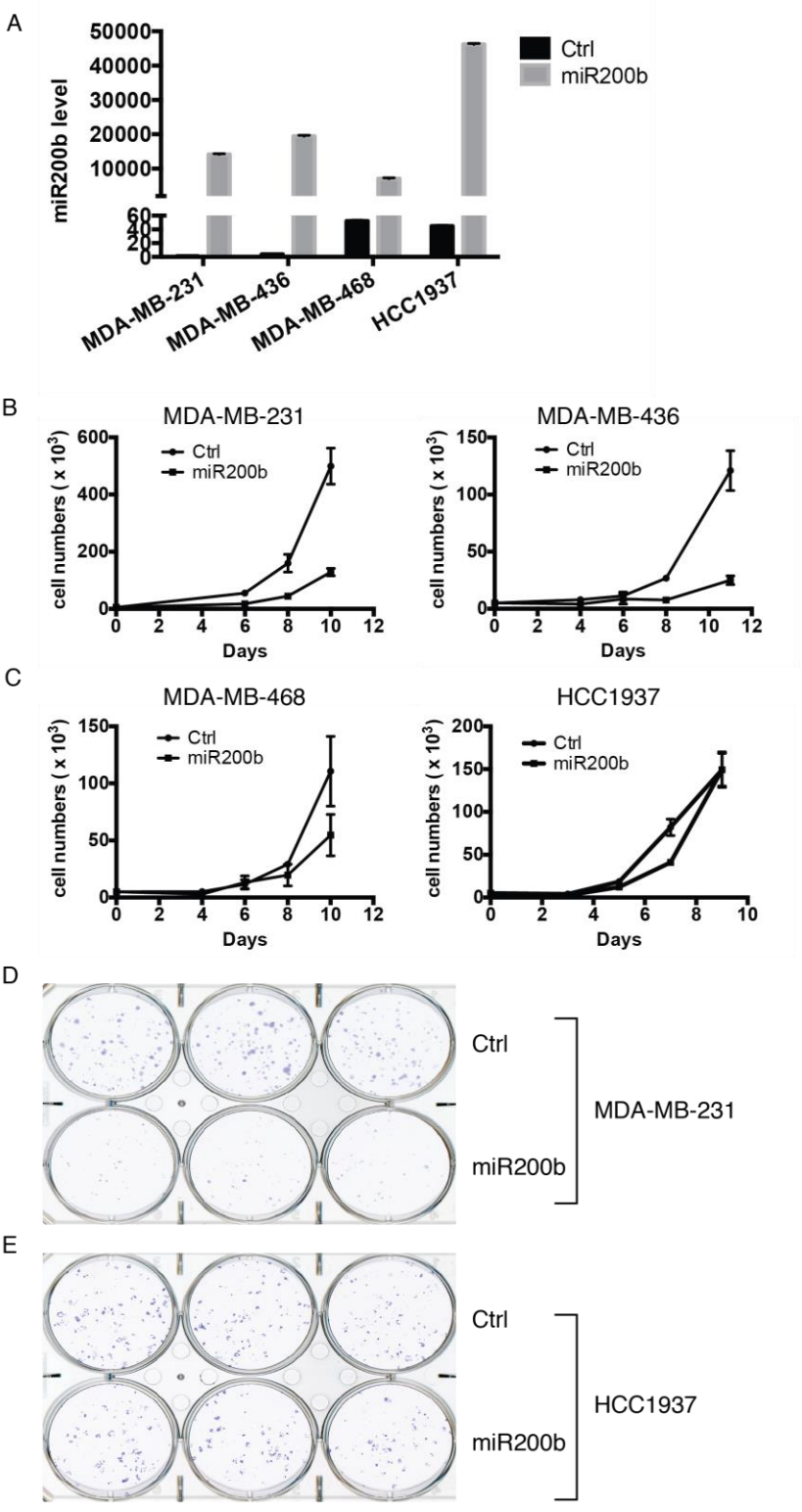


Figure 6. Overexpression of miR200b specifically inhibits cell proliferation and colony formation of claudin-low cells.

A. Control microRNA or miR200b were transfected into the indicated cell lines. After 48 hours, microRNAs were extracted and used for quantitative real-time PCR. B.C. Cells transfected with control microRNA or miR200b were seeded into 12 well plates. Cell numbers were counted at the indicated days using a hemocytometer. D.E. 300 cells transfected with control microRNA or miR200b were seeded into 6 well plates. Two weeks later, colonies were stained with crystal violet and imaged.

4.4 Conclusions

From examining the global levels of a number of histone modifications and histone-related proteins across subtypes, we found that H2Bub1 level is significantly lower in claudin-low subtype compared to basal-like subtype. Considering the facts that H2Bub1 is lower in pluripotent cells and that claudin-low subtype expresses higher levels of stem-cell like genes, it is an interesting finding that H2Bub1 level is also lower in claudin-low cells and tumors. It suggests that H2Bub1 is a critical and consistent marker for cells with stem cell features. It is possible that lower H2Bub1 level is a cause or a consequence of the stemness of claudin low subtype.

There are quite a number of DUBs and E3 ligases for H2Bub1, and even more proteins that regulate the enzymes in mammalian cells. The gain or loss of function of these proteins may result from gene copy number amplification, tranlocation, mutation, or deletion. Therefore, it is a rather difficult task to identify the reasons for lower H2Bub1 level in claudin-low subtypes. At least in our study, through examining the mRNA and protein levels of USP22, USP27x, USP51, USP44, USP49, RNF20, RNF40, ATXN7L3, and ENY2, we have not found the reasons for lower H2Bub1 level in claudin-low cells. We may be able to get insight of the causes by analyzing gene expression profiles of patients from large datasets. However, it is very possible that in each cell line or tumor, there are

different reasons causing lower H2Bub1 level, which makes finding the causes even more difficult.

Analyzing the differential H2Bub1 enriched genes in claudin-low versus basal-like subtypes may shed light on what role of H2Bub1 plays in two subtypes and why low H2Bub1 co-occurs with claudin-low subtype. The H2Bub1 enrichment on miR200 gene locus, including miR200a, miR200b, and miR429, is significantly lower in claudin-low cells compared to basal-like cells. The expression of miR200b is consistent with H2Bub1 enrichment, lower in claudin-low subtype. Interestingly, by overexpression miR200b in four triple negative cell lines, the cell proliferation and colony formation are inhibited only in claudin-low cells other than basal-like cells, indicating that miR200b can be a therapy target specifically for claudin-low subtype breast cancer.

Chapter 5: Cytoplasmic ATXN7L3B Interferes with Nuclear Functions of the SAGA Deubiquitinase Module

Parts of this text were taken from the following paper, which is under revision for publication in Molecular Cellular Biology.

Cytoplasmic ATXN7L3B Interferes with Nuclear Functions of the SAGA Deubiquitinase Module.

Wenqian Li,[a,b,c] Boyko S. Atanassov,[a,b] Xianjiang Lan,[a,b,c] Ryan D. Mohan,[d] Selene K. Swanson,[d] Aimee T. Farria,[a,b,c] Laurence Florens,[d] Michael P. Washburn,[d,e] Jerry L. Workman,[d] and Sharon Y.R. Dent[a,b]#

Department of Epigenetics and Molecular Carcinogenesis, The University of Texas M.D. Anderson Cancer Center, Smithville, Texas, 78957, USA[a]; Center for Cancer Epigenetics, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, 77030, USA[b]; Program in Epigenetics and Molecular Carcinogenesis, Graduate School of Biomedical Sciences, The University of Texas M.D. Anderson Cancer Center, Smithville, Texas, 78957, USA[c]; Stowers Institute for Medical Research, Kansas City, Missouri, USA[d]; Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA[e].

Driven by our interesting findings in H2Bub1 difference among breast cancer subtypes, we deeply explored the functions of the DUB module and its related proteins. At the end, we identified ATXN7L3B as an interesting regulator of H2Bub1 levels. In this chapter, we will discuss thoroughly the general molecular functions of ATXN7L3B and its functions specifically in breast cancer cells.

5.1 ATXN7L3B interacts with DUB module components but not SAGA.

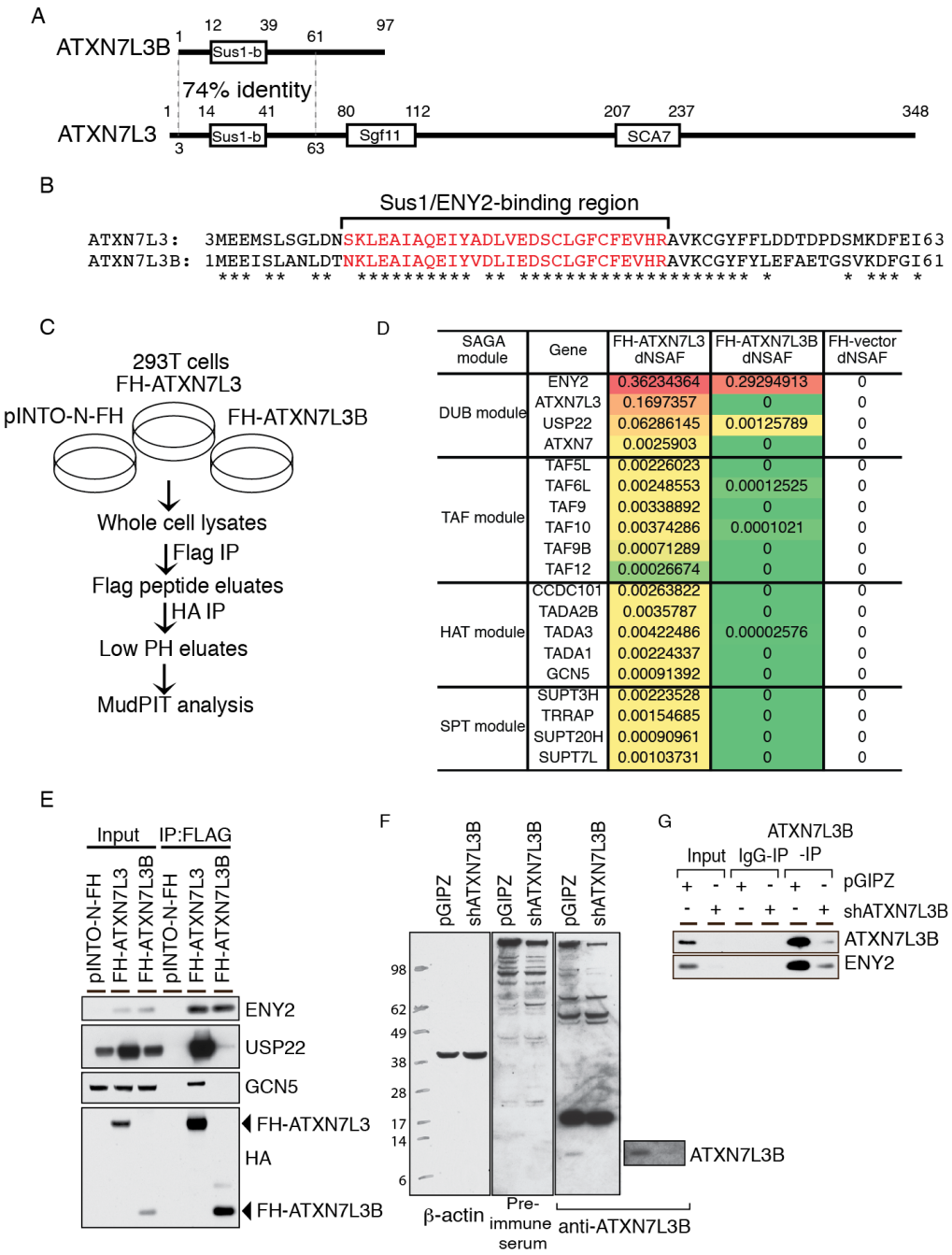
ATXN7L3B and ATXN7L3 share 74% identity within their N-terminal ~60 amino residues (Figure 7A), including a region that interacts with ENY2 (Figure 7B, in red) (32), called the Sus1-binding region for the yeast ENY2 ortholog (60). Interestingly, others have reported that a truncated form of ATXN7L3 that only contains amino acids 3-76 interacts with USP22 and ENY2 in vitro, but the complex containing the truncated protein cannot remove monoubiquitination from H2Bub1 (32). Since the native structure of ATXN7L3B resembles this artificial form of ATXN7L3, we reasoned that it might also interact with USP22 or other SAGA components and affect DUB activity in vivo.

We first determined whether ATXN7L3B is incorporated into the SAGA complex. Using tandem immunoprecipitation, we affinity purified ATXN7L3 or ATXN7L3B interacting proteins from 293T cells stably expressing Flag and HA-tagged ATXN7L3 (FH-ATXN7L3) or ATXN7L3B (FH-ATXN7L3B) (Figure 7C). Multidimensional Protein Identification Technology (MudPIT) indicated that FH-ATXN7L3 interacted with most known components of SAGA complex, with ENY2

and USP22 at the top, as expected (Figure 7D). ATXN7L3B is still capable of binding to ENY2 and USP22 of DUB module, but not ATXN7. However, even though the interaction between ATXN7L3B and ENY2 is comparable with that between ATXN7L3 and ENY2, the interaction between ATXN7L3B and USP22 is much weaker than that between ATXN7L3 and USP22 (~60 fold difference). For other modules, except for very low levels of TAF10, TAF6L, and TADA3, no other SAGA components were detected among FH-ATXN7L3B associated proteins (Figure 7D). Interactions between USP22, ENY2 and FH-ATXN7L3 or FH-ATXN7L3B were confirmed by co-immunoprecipitations followed by immunoblots (Figure 7E). These results indicate that ATXN7L3B does not associate with SAGA.

To investigate the functions of endogenous ATXN7L3B, we generated an antibody targeting the ATXN7L3B C-terminus 75-97 amino acids, which is not homologous with ATXN7L3. Whole cell lysates from shCtrl (pGIPZ) and shATXN7L3B cells were separated using SDS-PAGE gel, and blotted using pre-immune serum or our ATXN7L3B 4331-1 antibody. Even though the antibody recognizes multiple non-specific bands on the blot, it also specifically recognizes endogenous ATXN7L3B band around 11kDa that disappears after knocking down ATXN7L3B (Figure 7F). Unfortunately, a very strong non-specific band is detected at ~17kD, which is the same size as FH-ATXN7L3B. Therefore, FH-ATXN7L3B proteins were always detected using anti-Flag or anti-HA antibodies. Our ATXN7L3B specific antibody was also used to perform IP assays. Using this

antibody, we found that endogenous ATXN7L3B also binds strongly to endogenous ENY2 (Figure 7G). All of these data together, including the tandem IP, Flag-IP, and endogenous IP, demonstrate that ATXN7L3B strongly binds to ENY2.



SAGA module	Gene	FH-ATXN7L3 dNSAF	FH-ATXN7L3B dNSAF	FH-vector dNSAF
DUB module	ENY2	0.36234364	0.29294913	0
	ATXN7L3	0.1697357	0	0
	USP22	0.06286145	0.00125789	0
	ATXN7	0.0025903	0	0
TAF module	TAF5L	0.00226023	0	0
	TAF6L	0.00248553	0.00012525	0
	TAF9	0.00338892	0	0
	TAF10	0.00374286	0.0001021	0
	TAF9B	0.00071289	0	0
	TAF12	0.00026674	0	0
HAT module	CCDC101	0.00263822	0	0
	TADA2B	0.0035787	0	0
	TADA3	0.00422486	0.00002576	0
	TADA1	0.00224337	0	0
	GCN5	0.00091392	0	0
SPT module	SUPT3H	0.00223528	0	0
	TRRAP	0.00154685	0	0
	SUPT20H	0.00090961	0	0
	SUPT7L	0.00103731	0	0

Figure 7. ATXN7L3B interacts with DUB module components but not SAGA.

A. Comparison of the protein structures of ATXN7L3 and ATXN7L3B. The N-terminus share 74% of identity but ATXN7L3B lacks the Sgf11 and SCA7

domains that are present in ATXN7L3. B. Comparison of the protein sequences of ATXN7L3 and ATXN7L3B N-termini. Identical amino acid residues are indicated by asterisks. Red lettering indicates the Sus1/ENY2-binding region. C. Schematic of tandem Flag and HA affinity purification using 293T stable cells expressing pINTO-N-FH, FH-ATXN7L3, or FH-ATXN7L3B. D. dNSAF of SAGA components in different modules immunoprecipitated by FH-ATXN7L3 or FH-ATXN7L3B. E. Flag-IP from 293T cells stably expressing pINTO-N-FH vector, pINTO-N-FH-ATXN7L3, or pINTO-N-FH-ATXN7L3B. Bound proteins were resolved by electrophoresis and probed by immunoblot with indicated antibodies. F. Whole cell lysates from MCF7T cells stably expressing non-targeting shRNA (pGIPZ) or shATXN7L3B were resolved on SDS PAGE and probed by immunoblot with indicated antibodies. G.H. ATXN7L3-IP and ATXN7L3B-IP using MCF7T cells stably expressing non-targeting shRNA (pGIPZ), shATXN7L3 (G), or shATXN7L3B (H). Bound proteins were resolved by electrophoresis and probed by immunoblot with indicated antibodies.

5.2 ATXN7L3B mainly localizes to the cytoplasm whereas ATXN7L3 is nuclear.

SAGA is best characterized as a transcriptional co-activator functioning in the cell nucleus (31). To determine the subcellular distribution of ATXN7L3B, we performed immunofluorescence assays using 293T cells that stably express FH-ATXN7L3 or FH-ATXN7L3B. Staining with anti-Flag antibodies and DAPI revealed that ATXN7L3B is mainly localized outside of the nucleus, in the cytoplasm (Figure 8A). In contrast, as expected, ATXN7L3 mainly localized to the nucleus (Figure 8A). We confirmed the different distributions of these proteins by separating cells into cytoplasmic, soluble nuclear, and chromatin fractions, followed by immunoblots (Figure 8B). Whereas FH-ATXN7L3 localized in the chromatin fraction with histone H2B, FH-ATXN7L3B localized in the cytoplasmic fraction with β-tubulin. The distinct subcellular locations of ATXN7L3 and ATXN7L3B suggest that despite their similarities in sequence, these two proteins likely perform different functions in cells.

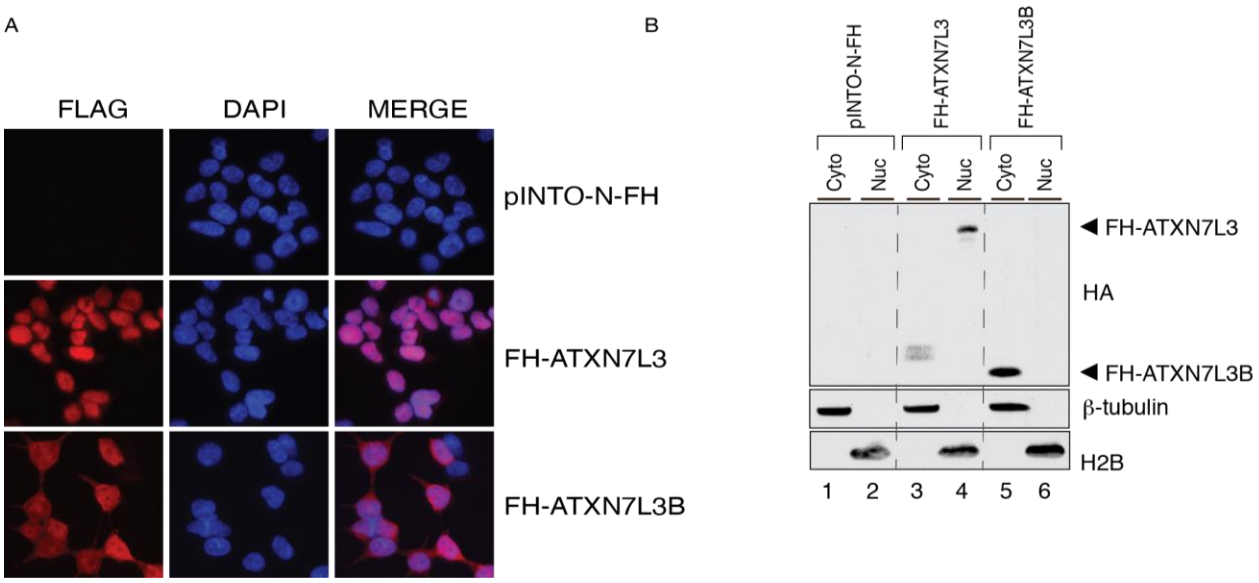


Figure 8. ATXN7L3B mainly localizes to the cytoplasm whereas ATXN7L3 is nuclear.

A. Immunofluorescence using Flag antibody (Red) on 293T cells stably expressing pINTO-N-FH vector, pINTO-N-FH-ATXN7L3, or pINTO-N-FH-ATXN7L3B. Nuclei were counterstained with DAPI (blue). B. Cytoplasmic and nuclear fractions were isolated from 293T cells stably expressing pINTO-N-FH vector, pINTO-N-FH-ATXN7L3, or pINTO-N-FH-ATXN7L3B. Fractions were resolved on SDS PAGE and blotted with the indicated antibodies. FH-ATXN7L3 and FH-ATXN7L3B proteins bands are indicated with arrows. H2B marks the nuclear fraction, while β-tubulin marks the cytoplasmic fraction.

5.3 ATXN7L3 and ATXN7L3B affect global levels and subcellular distributions of H2Bub1, ENY2, and USP22.

The best-characterized substrate of the SAGA DUB module is histone H2Bub1. Fluctuations in ENY2 protein levels affect DUB module activity and overall levels of H2Bub1 (31, 61, 62). Since ATXN7L3B interacts with ENY2, we determined how changes in ATXN7L3B levels affect H2Bub1. As expected, based on previously published data (62), over expression of ATXN7L3 greatly diminished H2Bub1, consistent with increased DUB activity. Surprisingly, over expression of ATXN7L3B had an opposite effect, dramatically increasing H2Bub1 levels (Figure 9A), consistent with loss of DUB activity. Examination of ENY2 and USP22 levels by immunoblot revealed that over expression of ATXN7L3 led to increased levels of both ENY2 and USP22 proteins (Figure 9A, lanes 1 and 2), consistent with the increased DUB activity and decreased H2Bub1 levels observed. Interestingly, ATXN7L3B over expression also led to increased of levels of ENY2 and USP22 (Figure 9A, lanes 1 and 3), although the extent of increase in USP22 levels was lower than that caused by ATXN7L3 overexpression. These results were puzzling, as previous work has shown that overexpression of USP22 leads to decreased H2Bub1 (62). Examination of the effects of over expression of ATXN7L3 or ATXN7L3B on the endogenous levels of each other provided a possible explanation. Overexpression of ATXN7L3B resulted in significantly decreased levels of endogenous ATXN7L3 (Figure 9A, lanes 1 and 3), consistent with the increased H2Bub1 levels observed.

Conversely, over expression of ATXN7L3 led to decreased levels of endogenous ATXN7L3B (Figure 9A, lanes 1 and 2). Unexpectedly, we also observed a significant decrease in endogenous ATXN7L3B after its ectopic protein overexpression (Figure 9A). This is possible due to negative self-regulation of ATXN7L3B protein, which need further investigation. Interestingly, these effects on protein levels upon overexpression of FH-ATXN7L3 or FH-ATXN7L3B were not reflected in changes in corresponding RNA levels, indicating these effects occur post-transcriptionally (Figure 9C, D).

Since ATXN7L3 and ATXN7L3B are located in different subcellular compartments (Figure 8), we isolated cytoplasmic, soluble nuclear, and chromatin fractions from FH-ATXN7L3 or FH-ATXN7L3B expressing cells in order to better understand effects on H2Bub1, ENY2, and USP22 levels. H2Bub1 levels in chromatin faithfully reflected levels observed in whole cell lysates, as expected (Figure 9B, compare lanes 2, 4, and 6). Interestingly, over expression of FH-ATXN7L3 led to increased levels of ENY2 and USP22 in both the cytoplasmic and chromatin fractions (Figure 9B, compare lanes 1 and 2 to lanes 3 and 4), whereas over expression of FH-ATXN7L3B led to increased levels of ENY2 and USP22 only in the cytoplasm, but decreased levels of ENY2 in nucleus (Figure 9B, compare lanes 1 and 2 to lanes 5 and 6). These results further explain why increased protein levels of global USP22 and ENY2 upon overexpression of FH-ATXN7L3B did not correlate to decreased levels of H2Bub1. Consistent with the reduction in global levels of endogenous ATXN7L3

observed upon overexpression of ATXN7L3B, decreased levels of ATXN7L3 were found in chromatin fraction (Figure 9B, compare lanes 1 and 2 to lanes 5 and 6), which is consistent with the increased H2Bub1 levels observed. Interestingly, these effects on protein levels upon overexpression of FH-ATXN7L3 or FH-ATXN7L3B were not reflected in changes in corresponding RNA levels, indicating these effects occur post-transcriptionally (Figure 9C, D).

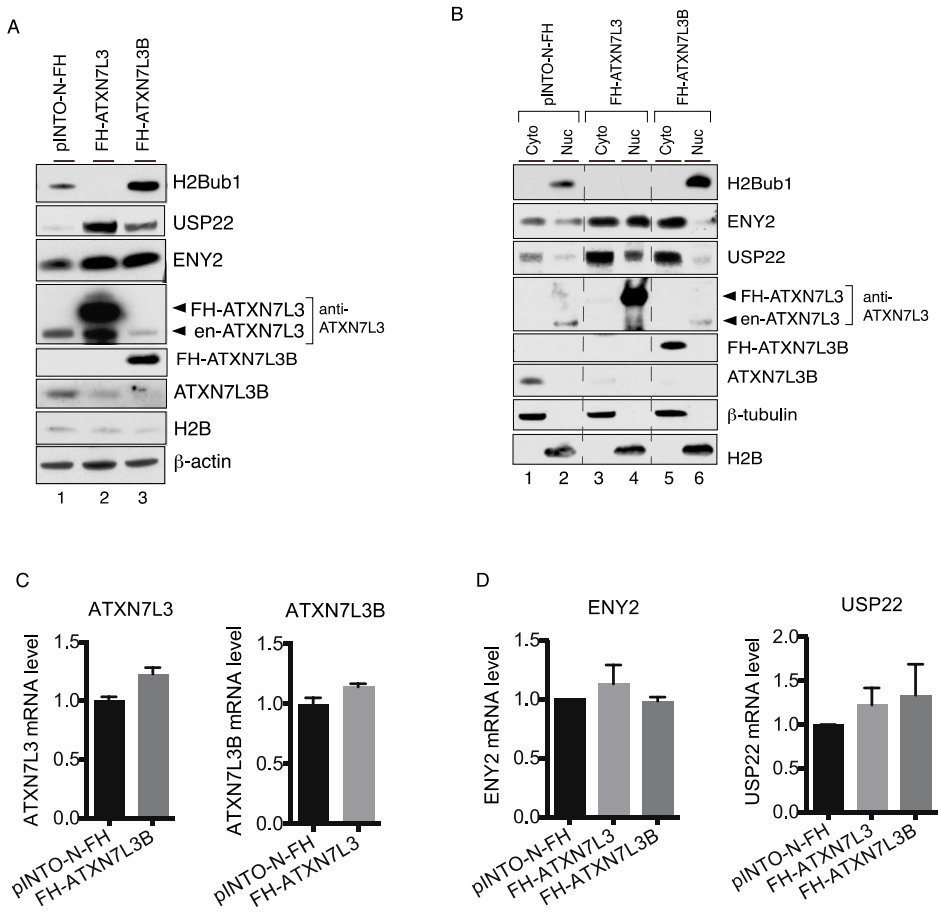


Figure 9. ATXN7L3 and ATXN7L3B affect global levels and subcellular distributions of H2Bub1, ENY2, and USP22.

A. Whole cell lysates from 293T cells stably expressing pINTO-N-FH vector, pINTO-N-FH-ATXN7L3, or pINTO-N-FH-ATXN7L3B, were resolved on SDS PAGE and probed by immunoblot with indicated antibodies. Arrows indicate endogenous (en-) and exogenous (FH-) ATXN7L3 proteins. B. Cytoplasmic, nuclear fractions were isolated from 293T cells stably expressing pINTO-N-FH vector, pINTO-N-FH-ATXN7L3, or pINTO-N-FH-ATXN7L3B. Fractions were resolved by electrophoresis and probed by immunoblot with indicated antibodies.

Arrows indicate endogenous (en-) and exogenous (FH-) ATXN7L3 proteins. C. D. Reciprocal effects of overexpression of FH-ATXN7L3 or FH-ATXN7L3B on protein levels occur post-transcription. RNA was extracted from 293T cells stably expressing pINTO-N-FH vector, pINTO-N-FH-ATXN7L3, or pINTO-N-FH-ATXN7L3B, and used as a template for cDNA generation and quantitative real-time PCR. The mRNA levels of ATXN7L3, ATXN7L3B, ENY2, and USP22 were examined.

5.4 ATXN7L3B competes with ATXN7L3 for ENY2 binding in vitro.

Since both ATXN7L3 and ATXN7L3B physically interact with ENY2, we determined whether ATXN7L3B can compete with ATXN7L3 for ENY2 binding in vitro. We purified recombinant Flag-ENY2, V5-ATXN7L3, and V5-ATXN7L3B separately from insect cells (Figure 10A), and added increasing amounts of V5-ATXN7L3B into reactions containing a constant amount of Flag-ENY2 and V5-ATXN7L3 (Figure 10B). Using Flag immunoprecipitations, we observed increasing amounts of ATXN7L3B, but decreasing amounts of ATXN7L3, associated with Flag-ENY2, indicating that ATXN7L3B is able to compete with ATXN7L3 for ENY2 binding. This result provides a plausible explanation for the decreased levels of ATXN7L3 protein level observed after overexpression of ATXN7L3B (Figure 9A, B), as previous studies have shown that ENY2 is necessary for stabilization of ATXN7L3 (62, 63). Overexpression of ATXN7L3B likely sequesters ENY2, resulting in less ENY2-ATXN7L3 association and destabilization of ATXN7L3.

To test whether USP22, ATXN7L3B, and ENY2 can form a stable complex in vitro, we co-infected baculovirus containing expressing vectors for HA-USP22, V5-ATXN7L3 or V5-ATXN7L3B, and Flag-ENY2, and used the lysates to perform HA pull down assay (Figure 10C). By analyzing the proteins that have been pulled down, we performed immunoblot assay (Figure 10D). As expected, USP22 binds to ATXN7L3 and ENY2. For ATXN7L3B, we found that HA-USP22

also binds to ATXN7L3B and ENY2 (Figure 10D). Then we wonder whether the ATXN7L3B DUB complex has deubiquitinase activity toward ubiquitinated histones, we performed in vitro deubiquitination assays using histone H2Bub1 as substrate. Consistent with previous results, a complex that contains USP22, ATXN7L3, and ENY2 deubiquitinated H2Bub1 completely after one hour (Figure 10D, lane 3). However, a complex containing ATXN7L3B instead of ATXN7L3 had only weak if any deubiquitinase activity towards H2Bub1, even though there was more ATXN7L3B complex than ATXN7L3 complex (Figure 10D, compare lane 6 and lane 3). These results indicate that ATXN7L3B containing complexes may function as DUBs towards non-histone proteins in vivo, again consistent with the cytoplasmic location of ATXN7L3B.

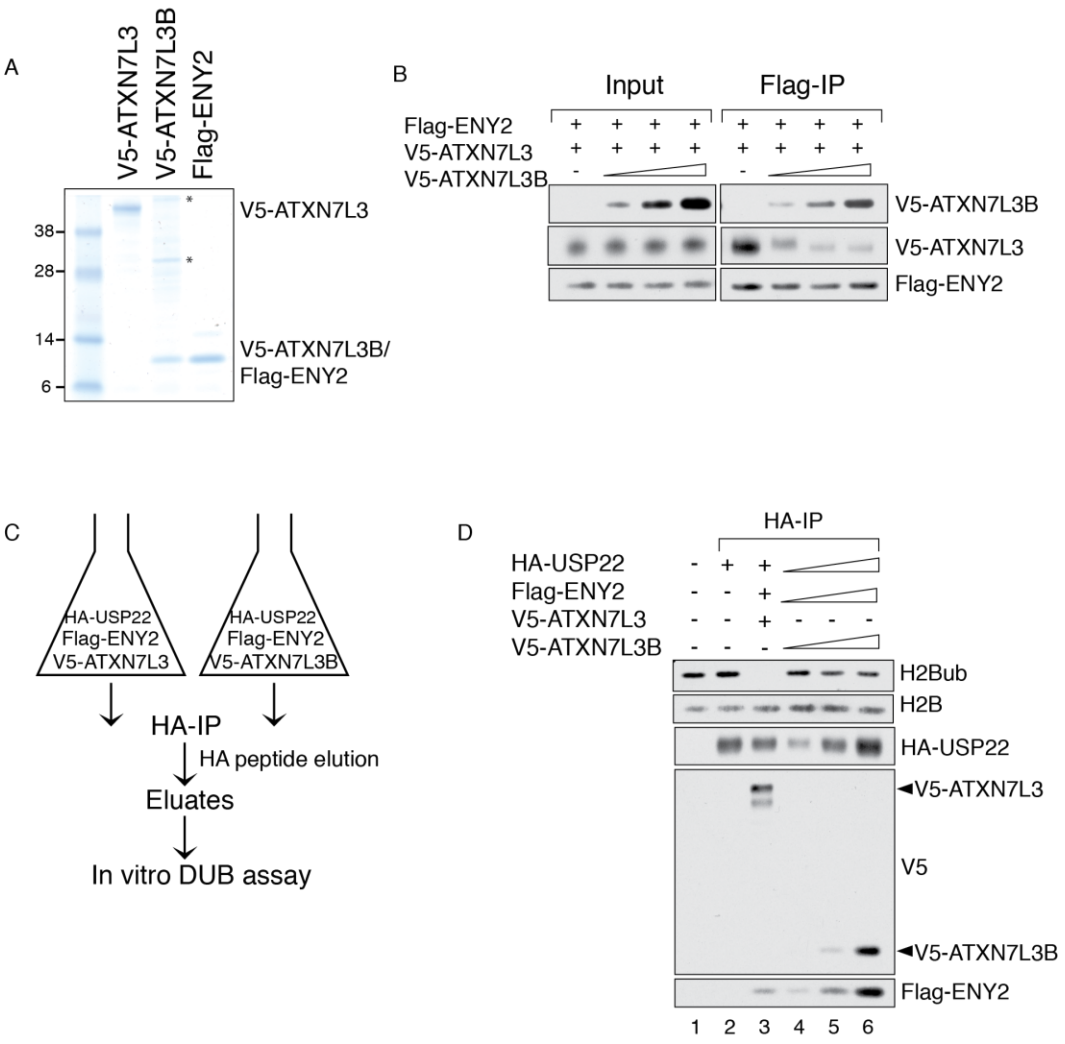


Figure 10. ATXN7L3B competes with ATXN7L3 for ENY2 binding in vitro.

A. Recombinant proteins were purified from Sf21 cells and resolved by electrophoresis, followed by colloidal blue staining. Asterisks indicate background protein bands. B. Indicated recombinant proteins were mixed in an in vitro competition reactions and the formed complexes were precipitated with anti-Flag resin. The resin bound proteins were eluted, resolved on SDS PAGE and detected by western blot with indicated antibodies. 5% of the in vitro reactions were used as input. C. Schematic of HA affinity purification using Sf21 cells co-

infected with HA-USP22, Flag-ENY2, and V5-ATXN7L3 or V5-ATXN7L3B. D. Indicated immunoprecipitated complexes from (C) were used in in vitro deubiquitination reactions in which purified histones served as substrates. Reactions were resolved by electrophoresis and detected by western blot with indicated antibodies.

5.5 ATXN7L3B regulates the protein levels of ENY2.

To complement the overexpression experiments, we next examined the effects of depletions of these proteins on levels of H2Bub1 and DUB module components. We used MCF7T cells for these experiments, as endogenous ATXN7L3B is easily detectable in these cells (Figure 11A, lane 1). Knockdown of ATXN7L3B slightly decreases H2Bub1 levels in MCF7T cells (Figure 11A, compare lanes 1 and 3). Interestingly, knockdown of ATXN7L3B caused a significant reduction in ENY2 levels, whereas knockdown of ATXN7L3 increased levels of ENY2 (Figure 11A). In addition, knockdown of ATXN7L3 or ATXN7L3B increases the protein levels of the other protein (Figure 11A). Knockdown of ATXN7L3 or ATXN7L3B does not significantly affect mRNA levels of ENY2, again indicating regulation of ENY2 by ATXN7L3 or ATXN7L3B occurs post-transcriptionally (Figure 11B). Knockdown of ATXN7L3B also does not alter mRNA levels of ATXN7L3, whereas knockdown of ATXN7L3 increases mRNA levels of ATXN7L3B (Figure 11B). Therefore, the induction of ATXN7L3B protein level might result from transcriptional regulation after ATXN7L3 knockdown. Additionally, the reduction of ENY2 upon knock down of ATXN7L3B was restored by expression of exogenous ATXN7L3B (Figure 11C), indicating the effect is not due to an off target effect of the knock down reagents.

We next determined how ATXN7L3B or ATXN7L3 depletion affects the subcellular distribution of ENY2. Knockdown of ATXN7L3B significantly

decreased levels of ENY2 in the cytoplasm without affecting chromatin associated ENY2 levels (Figure 11D, compare lanes 1 and 3 to lanes 7 and 9, Figure S2B). Conversely, depletion of ATXN7L3 led to an increase in the cytoplasmic ENY2 levels without affecting chromatin associated ENY2 (Figure 11D, compare lanes 1 and 4).

To examine whether ENY2 levels are regulated through proteasome or lysosome-dependent protein degradation, we treated control, ATXN7L3, or ATXN7L3B depleted MCF7T cells with MG132, a proteasome inhibitor, or Bafilomycin A1 (BafA1), a lysosome inhibitor. Neither MG132 or BafA1 treatment rescued reductions in ENY2 protein levels caused by ATXN7L3B knockdown (Figure 11E, F). These data indicate that ATXN7L3B likely regulates ENY2 levels at the level of RNA processing, RNA transport or protein translation.

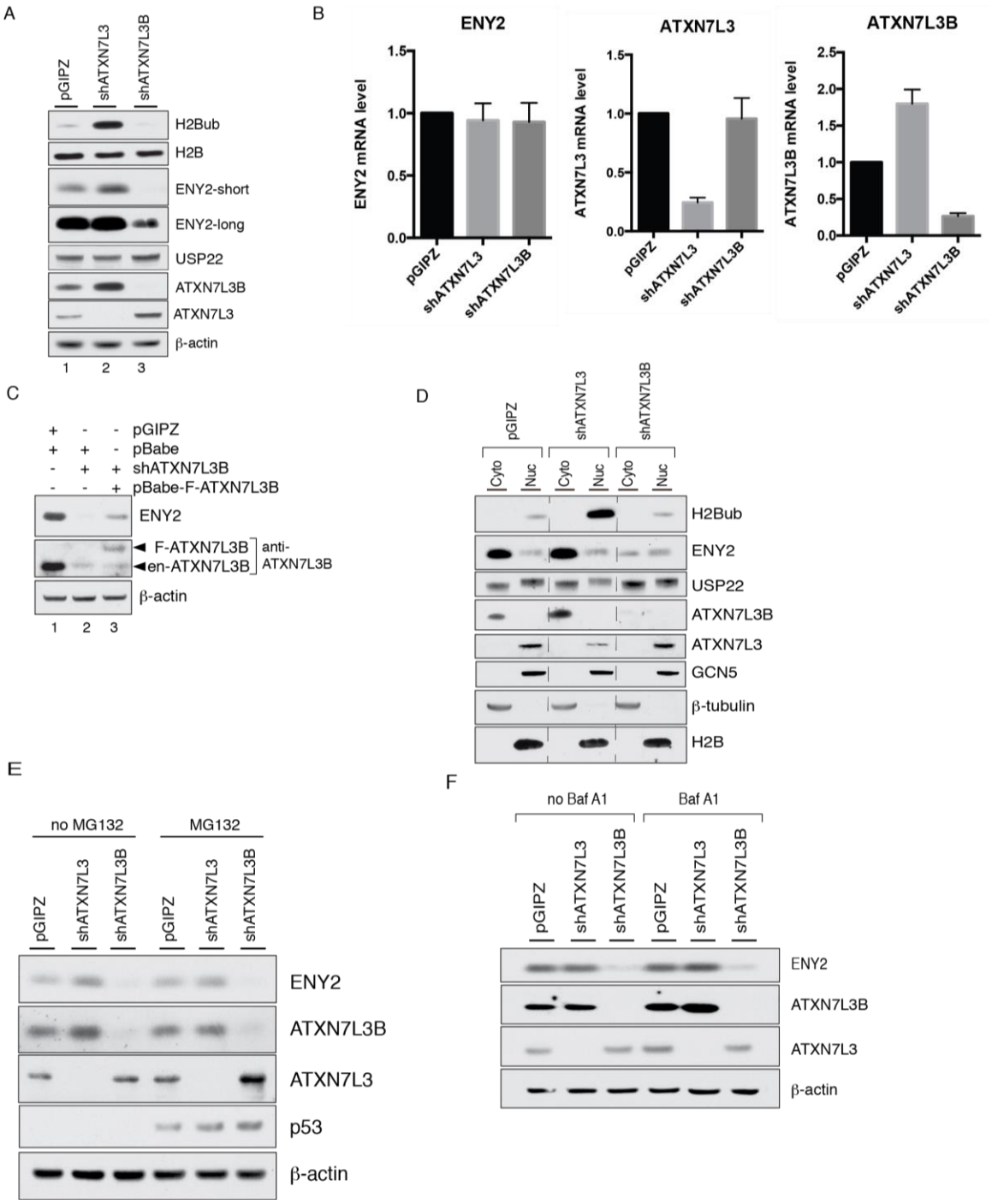


Figure 11. ATXN7L3B regulates the protein levels of ENY2.

A. Whole cell lysates from MCF7T cells stably expressing shATXN7L3, shATXN7L3B on non-targeting shRNA (pGIPZ), were resolved on SDS PAGE

and probed by immunoblot with indicated antibodies. B. RNA from ATXN7L3 depleted, ATXN7L3B depleted, or control (pGIPZ) MCF7T cells were isolated. The mRNA levels of ENY2, ATXN7L3, and ATXN7L3B were examined using quantitative real-time PCR. C. Ectopic expression of shRNA immune ATXN7L3B restore the ENY2 levels in ATXN7L3B depleted cells. Whole cell lysates from ATXN7L3B depleted MCF7T cells were used to monitor the ENY2 protein levels before (lane 2) and after (lane 3) ectopic expression of FLAG-ATXN7L3B. Arrows indicate endogenous (en-) and exogenous (F-) ATXN7L3B proteins. D. Cytoplasmic, nuclear fractions were isolated from the cell used in (A), resolved by electrophoresis and probed by immunoblot with indicated antibodies. E.F. ATXN7L3 depleted, ATXN7L3B depleted, or control (pGIPZ) MCF7T cells were treated with MG132 (20uM, 5 hours), or BafA1 (100nM, 2 hours). Lysates were extracted and resolved by electrophoresis and probed by immunoblot with indicated antibodies.

5.6 ATXN7L3B regulates migration of ER positive breast cancer cells.

Given that USP22 overexpression is correlated with several aggressive cancers (41, 42, 64), we examined ATXN7L3B protein levels in normal mammary, ER+, and HER2 positive (HER2+) breast cancer cell lines (Figure 12A). Interestingly, higher ATXN7L3B protein levels were observed only in the ER+ breast cancer cell lines (Figure 12A). mRNA levels of ATXN7L3B were also detected in these cell lines (Figure 12B). Even though not all ER+ cell lines express higher levels of ATXN7L3B mRNAs, MCF7T and MDA-MB-361 express significantly more ATXN7L3B mRNAs (Figure 12B). We depleted ATXN7L3B using an shRNA in MCF7T cells (Figure 12C, D) to determine whether ATXN7L3B loss affects ER function and cell migration or colony formation of ER+ breast cancer cell line. By examining the mRNA levels of three ER targets genes, GREB1, Trefoil Factor 1 (TFF1) (65) and Insulin-Like Growth Factor Binding Protein 4 (IGFBP4) (66), we found that ATXN7L3B depletion decreased expression of these genes (Figure 12E). Interestingly, ATXN7L3B depletion also decreased cell migration without affecting colony formation ability of MCF7T cells (Figure 12F, G).

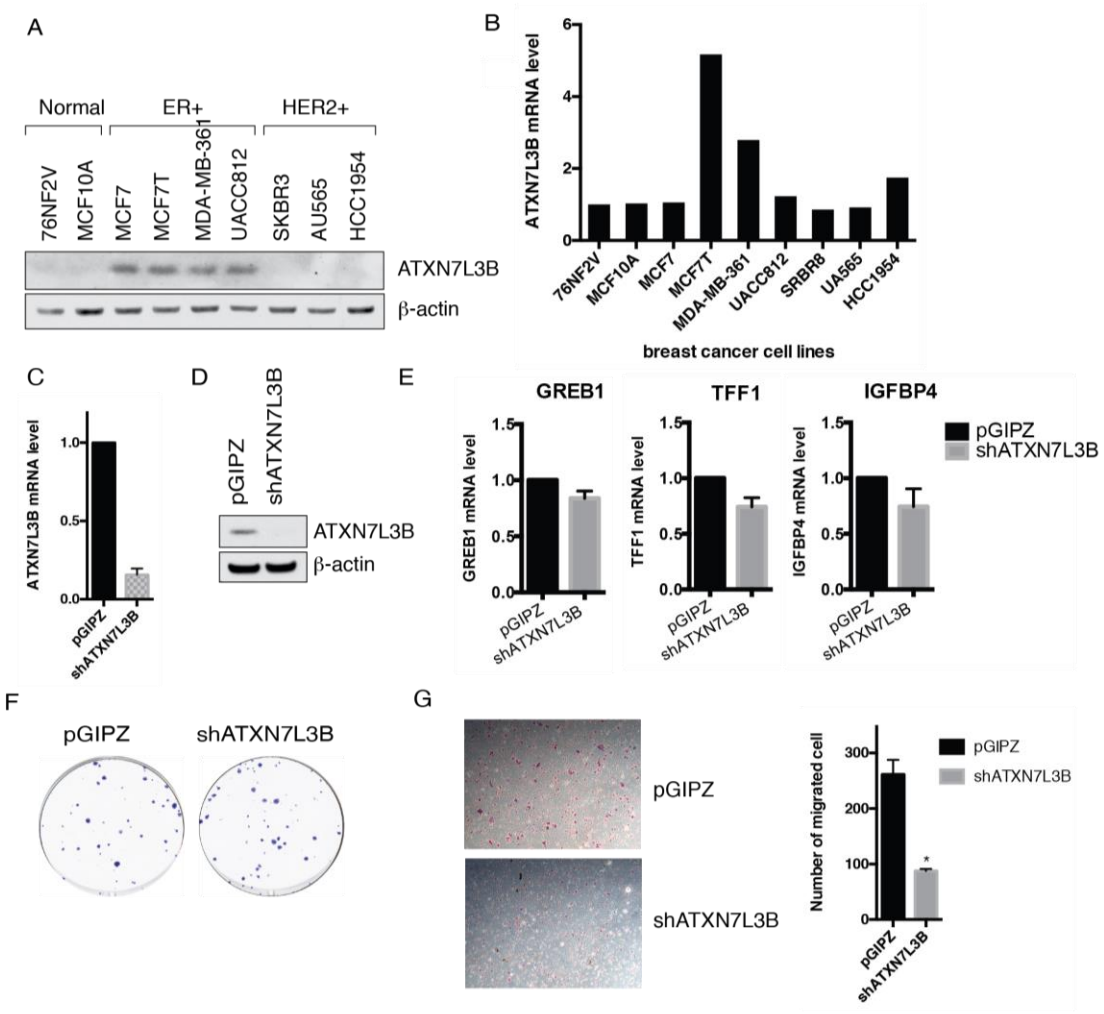


Figure 12. Depletion of ATXN7L3B inhibits migration of ER positive breast cancer cells.

A. Protein levels of ATXN7L3B in two normal mammary, four ER+, and three HER2+ cell lines were detected using immunoblot. B. mRNA levels of ATXN7L3B in same cells as (A) were detected using quantitative real-time PCR. RNA was extracted from the indicated lines and used as a template for cDNA generation and quantitative real-time PCR. C.D. Efficient silencing of ATXN7L3B in MCF7T cells stably expressing shATXN7L3B. The mRNA levels of ATXN7L3B were

examined using quantitative real-time PCR, and protein levels of ATXN7L3B were examined using immunoblot. E. Expression levels of ER target genes, GREB1, TFF1 and IGFBP4 in AXTN7L3B depleted and control (pGIPZ) MCF7T cells. mRNA levels was examined using quantitative real-time PCR. F. ATXN7L3 reduction does not impact the colony formation of MCF7T cells. ATXN7L3B depleted or control cells were used to perform colony formation assays. After 20 days of incubation, colonies were stained with crystal violet. G. ATXN7L3B ablation impairs MCF7T cell migration. AXTN7L3B depleted or control cells were used to perform transwell migration assay. After 24 hours of incubation, cells that had migrated to the bottom of transwell chambers were stained with crystal violet and imaged. The numbers of migrated cells were counted.

5.7 Conclusions

Our data provided the first evidence that USP22 DUB activity is regulated in vivo and in vitro by ATXN7L3B through sequestration of ENY2 and destabilization of ATXN7L3. Unlike ATXN7L3, ATXN7L3B does not interact with most SAGA components, but it exhibits strong interactions with ENY2. Interactions of ENY2 and ATXN7L3 or ATXN7L3 appear to mutually exclusive. Moreover, manipulation of ATXN7L3B levels by overexpression or shRNA depletion leads to effects on H2Bub1 that are opposite to those caused by changes in ATXN7L3 levels. These effects are consistent with changes in ATXN7L3 protein levels observed upon overexpression or knockdown of ATXN7L3B. While overexpression of ATXN7L3 leads to increased DUB activity and decreased levels of H2Bub1, overexpression of ATXN7L3B leads to reduction of ATXN7L3 and increased levels of H2Bub1, consistent with a loss of DUB activity directed towards this histone modification. In contrast to the loss of ATXN7L3, overexpression of ATXN7L3B increases cytoplasmic levels of ENY2 and USP22. Similarly, depletion of ATXN7L3B leads to decreased H2Bub1 levels and increased ATXN7L3 protein levels, but decreased cytoplasmic levels of ENY2. As ENY2 is required for stabilization of ATXN7L3 within the DUB module (63), ATXN7L3B likely affects H2Bub1 levels indirectly, by changing the subcellular distribution of ENY2 and altering ENY2 availability for ATXN7L3 interaction.

Although ATXN7L3B contains the Sus1/ENY2-binding region, it lacks the ZnF-Sgf11 and SCA7 domains that reside in C-terminal region of ATXN7L3. A previous report by another group demonstrated that the ZnF-Sgf11 domain of ATXN7L3 is essential for DUB activity towards H2Bub1 in vitro (32). The ZnF-Sgf11 domain is also required for ATXN7L3 binding to nucleosomal DNA (53), and the crystal structure of the DUB module reveals an arginine cluster in the ZnF-Sgf11 domain directly interacts with ubiquitinated nucleosomes and H2A/H2B heterodimer (54). Therefore, the absence of this domain in ATXN7L3B indicates it is unlikely to interact with histones, consistent with our findings that the ATXN7L3B-DUB module cannot efficiently deubiquitinate histones in vitro and that ATXN7L3B is largely localized to the cytoplasm. Future work will determine whether the ATXN7L3B-DUB module targets non-histone, cytoplasmic proteins.

ATXN7L3B protein shows a breast cancer subtype specific expressing pattern, highly expressed in ER positive breast cancer cells. We found that depletion of ATXN7L3B inhibits migration of ER positive breast cancer cells and decrease expression of ER target genes. Considering that ATXN7L3B localizes in cytoplasm, it is likely that ATXN7L3B interferes with ER function indirectly, such as regulating the function of SAGA in the nucleus. The molecular mechanism of ATXN7L3B in ER positive breast cancer need to be further investigated.

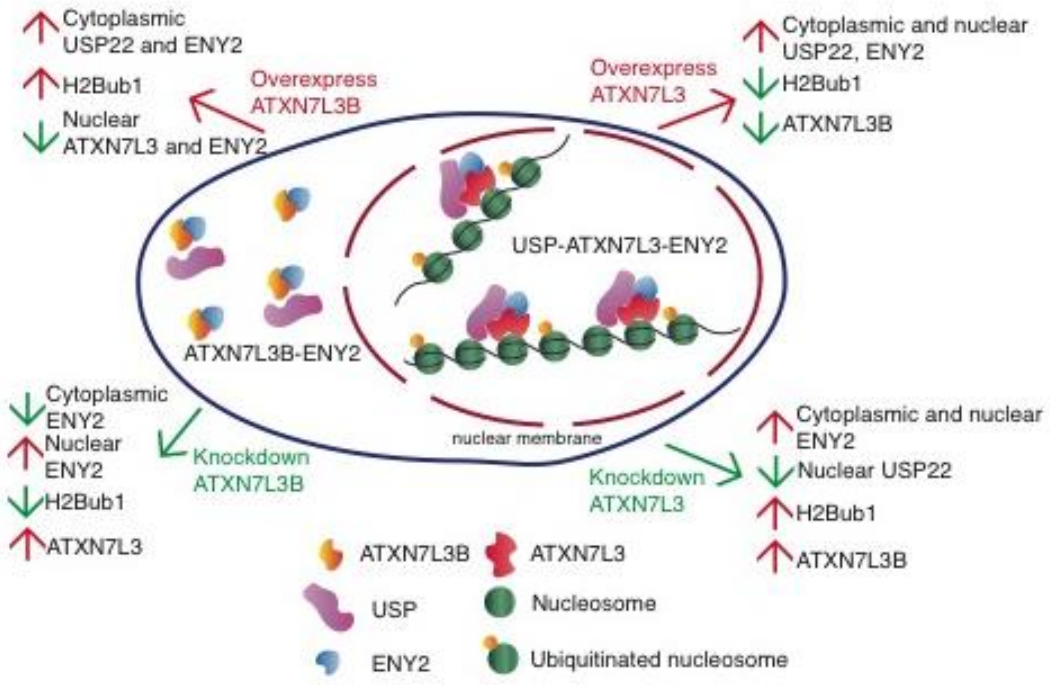


Figure 13. Regulation of protein levels of H2Bub1 and DUB components by ATXN7L3 and ATXN7L3B.

ATXN7L3 mainly resides in nucleus, while ATXN7L3B mainly localizes in cytoplasm. They both interact strongly with ENY2. Overexpression of ATXN7L3 leads to increase of cytoplasmic and nuclear USP22 and ENY2, but decrease of H2Bub1 and ATXN7L3B. Overexpression of ATXN7L3B leads to increase in H2Bub1 and cytoplasmic USP22 and ENY2, but decrease in nuclear ENY2 and ATXN7L3. Depletion of ATXN7L3 results in upregulation of ENY2 in cytoplasm and nucleus, H2Bub1, and ATXN7L3B, but decrease in nuclear USP22 level.

Depletion of ATXN7L3B decreases ENY2 in cytoplasm, H2Bub1, but increases nuclear ATXN7L3 and ENY2 protein levels.

Chapter 6: Epigenetic changes drive the molecular patterns of breast cancer subtypes.

6.1 Chromatin states define breast cancer subtypes

Using ChIP-sequencing data defining the genomic locations of histone modifications across 14 cell lines, we identified 13 different chromatin states, such as active transcriptions, heterochromatin, repressed polycomb, active enhancers, among others (Figure 14A). Using these 13 chromatin states, we are able to recapitulate the breast cancer subtypes according to their epigenetic changes. For example, the active transcription state (TxAct) genes exhibit subtype specific patterns (Figure 14B), as well as the repressed polycomb state (RepPC) (Figure 14C).

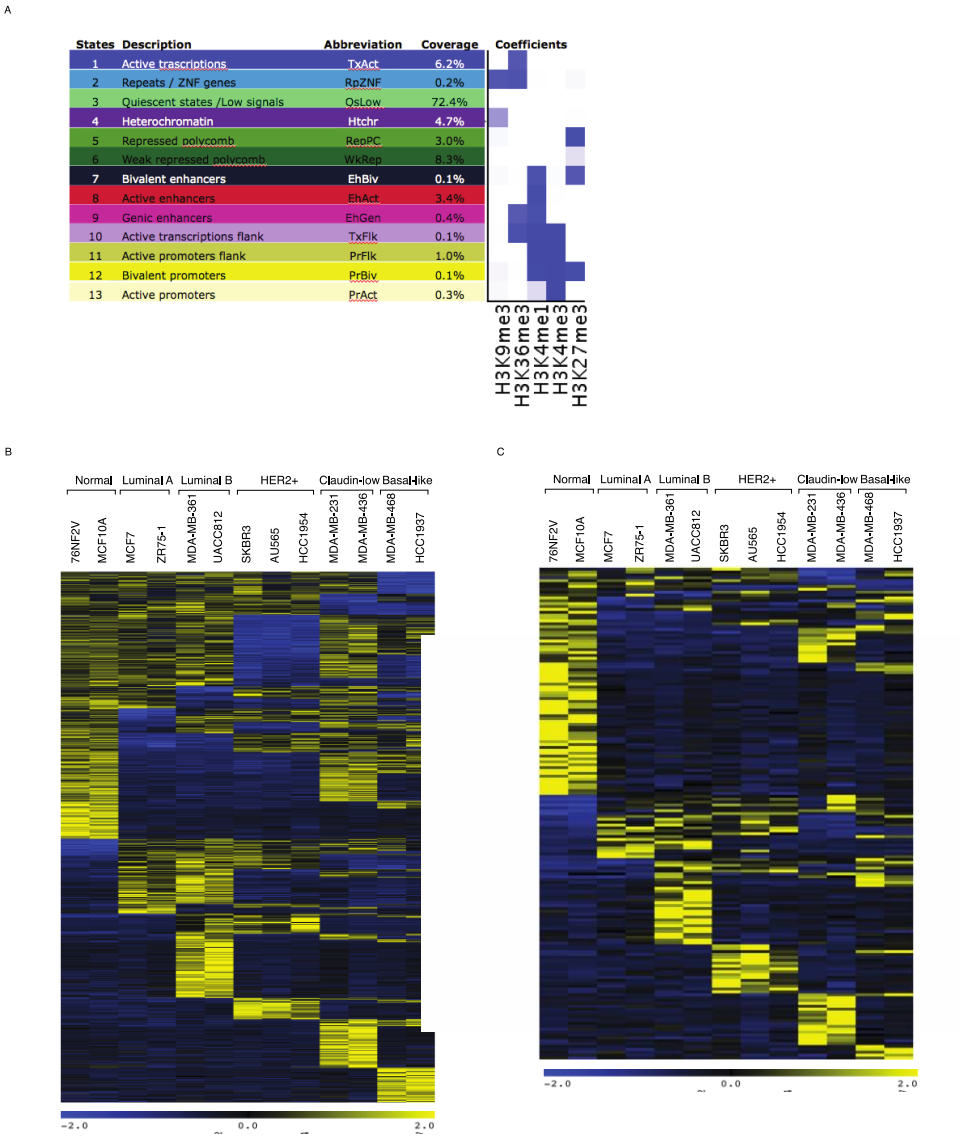


Figure 14. Chromatin states identify specific patterns of breast cancer subtypes.

A. Chromatin states identified by five histone modification marks. B.C. Subtype specific patterns identified by TxAct (B) and RepPC (C) chromatin states. Figure 14 is attributed to Dr. Yuanxin Xi.

6.2 AFAP1-AS1 is highly enriched of active transcription marks in triple-negative breast cancer cells.

After analyzing the histone modification ChIP sequencing in 14 mammary cell lines, we found that a number of genes exhibit subtype-specific enrichment of active or repressed histone modification marks. For example, NLRP gene cluster is enriched of repressed histone modification, such as H3K27me3, in breast cancer cell lines compared to two normal immortalized mammary cell lines. TSS and gene body of NAA60 and ZNF597 gene locus are enriched of active histone modifications in non-TNBC cell lines. In addition, Actin Filament Associated Protein Antisense RNA 1 (AFAP1-AS1) gene locus is highly enriched for histone modifications that indicate active transcription, such as H3K9ac and H3K79me2 (Figure 15) In TNBC cell lines.

Specifically interested in study of TNBC, we focused on exploring the function of AFAP1-AS1 in TNBC cells. The AFAP1-AS1 gene encodes a long non-coding RNA. AFAP1-AS1 has been reported to be highly expressed and predict poor prognosis in various types of cancers, including Barrett's esophagus and esophageal adenocarcinoma (67, 68), pancreatic ductal adenocarcinoma (69), lung cancer (70, 71), nasopharyngeal carcinoma (72), hepatocellular carcinoma (73, 74), and colorectal cancer (75). In Barrett's esophagus adenocarcinoma, the AFAP1-AS1 gene locus is largely hypomethylated, which results in the higher expression level compared to matched normal tissues (67). AFAP1-AS1 shares

the same gene locus with AFAP1, but AFAP1 has been reported not to be a target for AFAP1-AS1 (67). No direct targets for AFAP1-AS1 have been identified, but the epithelial marker E-cadherin expression can be repressed by AFAP1-AS1 ectopic expression (69). Meanwhile, the mesenchymal markers Vimentin, N-cadherin, Slug, Snail expression levels are elevated upon AFAP1-AS1 ectopic expression. This suggests that AFAP1-AS1 may promote proliferation, migration, or invasion of cancer cells via increasing the epithelial-mesenchymal transition (EMT) process.

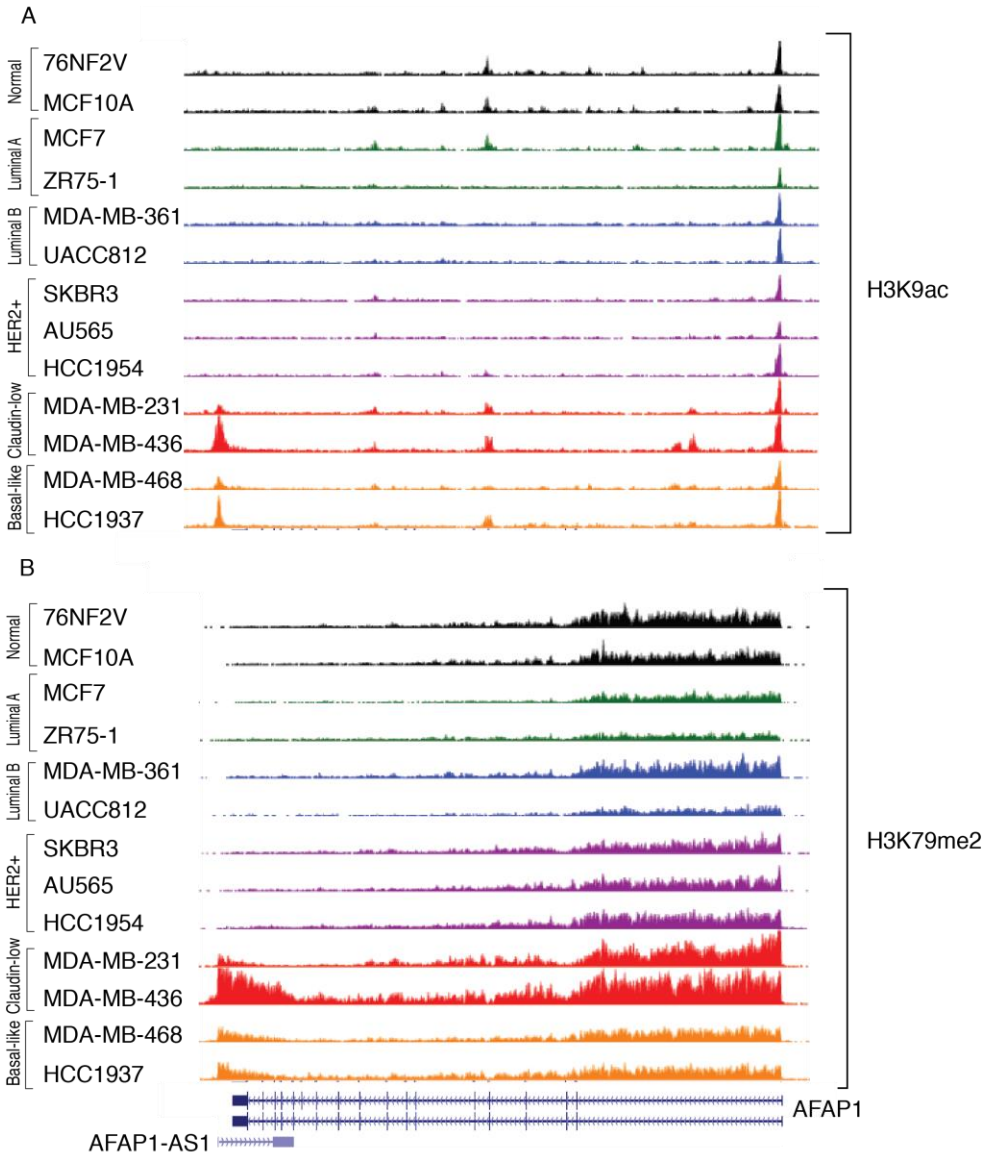


Figure 15 AFAP1-AS1 gene locus is highly enriched for active transcription marks.

A. H3K9ac enrichment at the AFAP1-AS1 and AFAP1 gene locus across cell lines. B. H3K79me2 enrichment at the AFAP1-AS1 and AFAP1 gene locus across cell lines.

6.3 AFAP1-AS1 is highly expressed in TNBC cells and tumors.

I confirmed the higher mRNA levels of AFAP1-AS1 in 4 TNBC cell lines using quantitative real-time PCR. I also confirmed that AFAP1 mRNA levels are not correlated with AFAP1-AS1 levels, indicating that AFAP1 is not regulated by AFAP1-AS1 (Figure 16A). To investigate the functions of AFAP1-AS1 in TNBC cell lines, To investigate the functions of AFAP1-AS1 in TNBC cell lines, we depleted AFAP1-AS1 expression using two siRNAs targeting AFAP1-AS1, #19 and #20, in two TNBC cell lines, MDA-MB-231 and HCC1937 (Figure 16B,E). Interestingly, knockdown of AFAP1-AS1 slows cell proliferation (Figure 16C,F) and inhibits colony formation (Figure 16D,G) of MDA-MB-231 cells and HCC1937 cells. It is yet clear the underlying mechanism of AFAP1-AS1 involved in cancer development, however, our analysis through the chromatin states signatures is the first study that demonstrates the activities of AFAP1-AS1 in breast cancers in a subtype specific manner, suggesting that AFAP1-AS1 is not involved in hormonal induced breast cancer development.

Using “human in mouse” (HIM) tumor samples, we also found that AFAP1-AS1 expression is higher in triple negative breast cancer tumors HIM2, 5, 6 compared to ER positive breast cancer tumors, HIM23 (Figure 17A, B), and well correlated with the expression of mesenchymal markers, Vimentin, Slug, and Snail, but reversely correlated with the expression of epithelial markers, E-cadherin (Figure 17C, D).

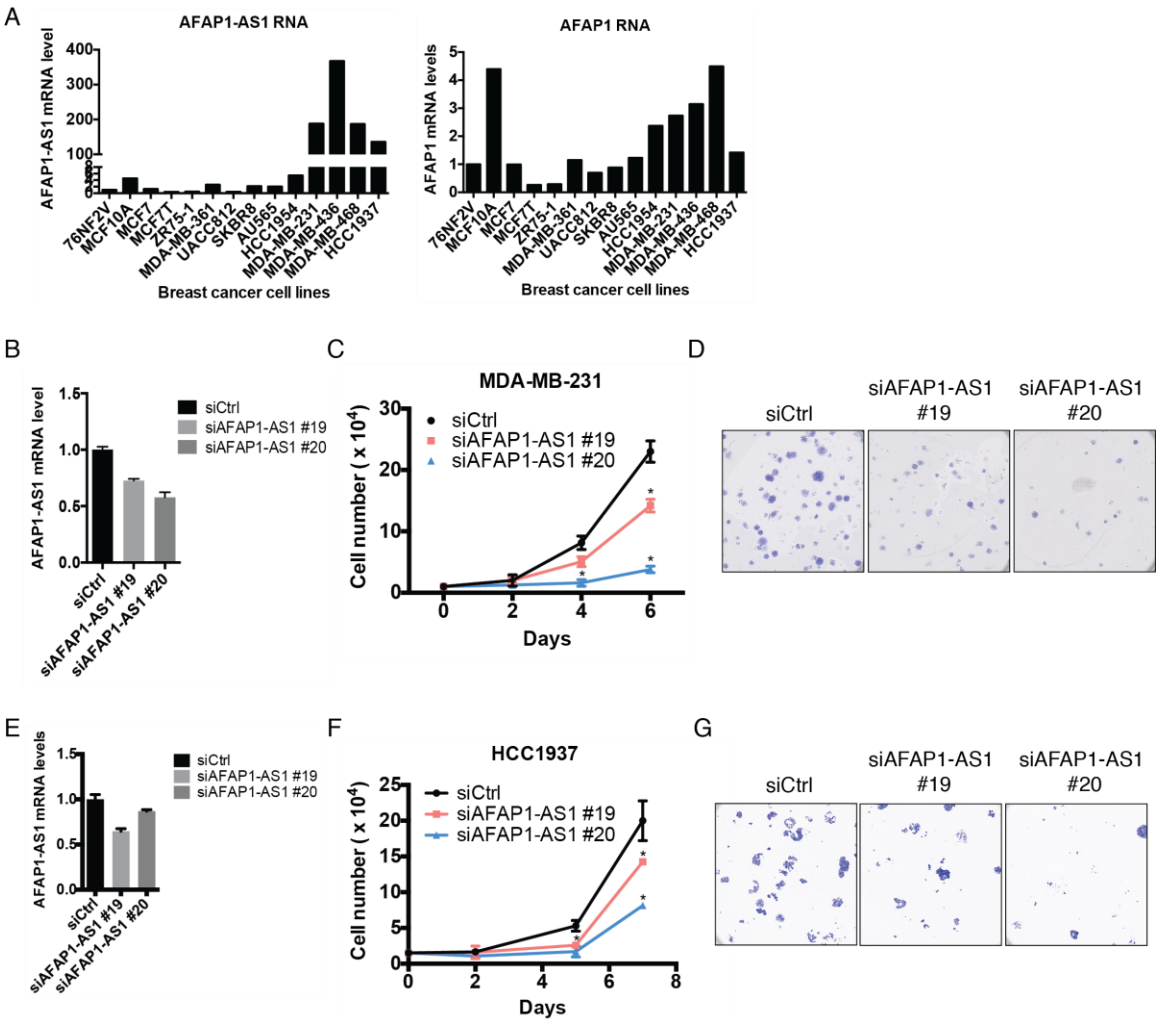


Figure 16. AFAP1-AS1 is highly expressed in TNBC cells.

A. RNAs were extracted from 14 cell lines and reverse-transcribed to cDNAs, which were used for quantitative real-time PCR to examine the mRNA levels of AFAP1-AS1 and AFAP1. B.E MDA-MB-231 and HCC1937 cells were transfected with control siRNA, or siRNA targeting AFAP1-AS1. RNAs were extracted and reverse-transcribed to cDNAs, which were used for quantitative real-time PCR to examine the mRNA levels of AFAP1-AS1. C.F. MDA-MB-231 and HCC1937 cells were transfected with control siRNA, or siRNA targeting AFAP1-AS1. 15,000 cells were plated into 12 well plates with 3 replicates for each sample. Cell

numbers were counted with hemocytometer at indicated days. D.G. MDA-MB-231 and HCC1937 cells were transfected with control siRNA, or siRNA targeting AFAP1-AS1. 300 cells were plated into 6 well plates with 2 replicates for each sample. Two weeks later, colonies were stained with crystal violet and imaged.

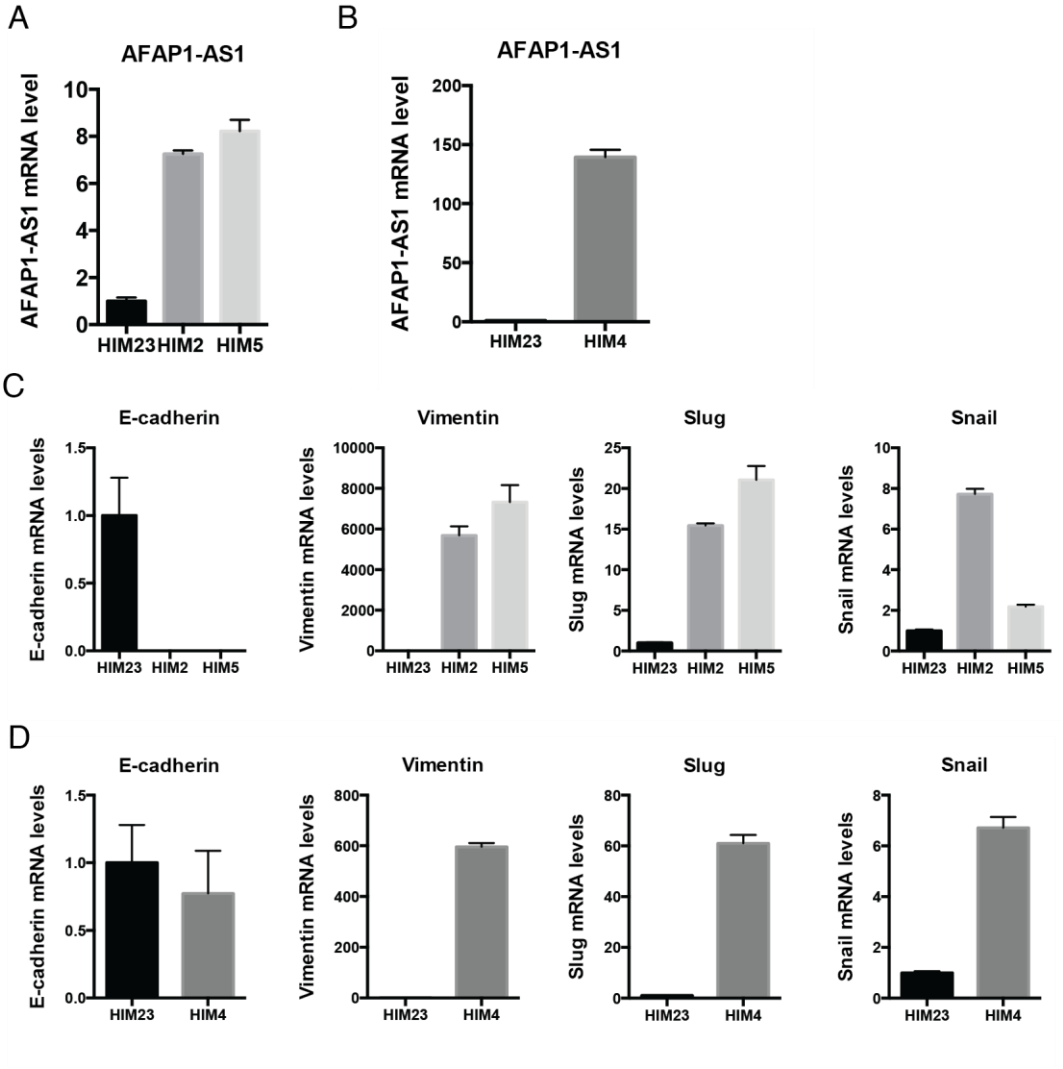


Figure 17. AFAP1-AS1 is highly expressed in TNBC tumors.

RNAs were extracted from ER positive HIM tumor (HIM23), claudin-low HIM tumors (HIM2, HIM5), and basal-like HIM tumor (HIM4). RNAs were reverse-transcribed to cDNAs, which were used for quantitative real-time PCR to examine the mRNA levels of AFAP1-AS1 (A.B), and epithelial or mesenchymal markers (C.D).

6.4 Conclusions

Our study studied the breast cancer subtypes at epigenetic levels using varied breast cancer cell lines as experimental models. Using ChIP-sequencing data defining the genomic locations of histone modifications across 14 cell lines, we are able to classify breast cancer subtypes according to the epigenetic changes.

Using ChIP-sequencing data across 14 cell lines, we identified AFAP1-AS1 as a gene that is highly expressed and enriched of active transcription markers specifically in TNBC cell lines. Depletion of AFAP1-AS1 is able to inhibit cell proliferation and colony formation of TNBC cells. In addition, the histone modification status of AFAP1-AS1 gene in TNBC cells indicates that the high expression of AFAP1-AS1 in TNBC cells is regulated by epigenetic changes.

Chapter 7: Discussion and Future Directions

Heterogeneity of breast cancer is a big obstacle for cancer therapy. About 75% of breast cancers are positive for ER and/or PR. Endocrine therapy, which blocks estrogenic signaling, is the main treatment for these patients. However, a fraction of patients fail to respond to endocrine therapy due to intrinsic or acquired resistance. As for TNBC subtype, currently, there are yet no efficient target therapies. Therefore, novel biomarkers or targets are in need to develop new therapy methods for breast cancer. The more we learned about the heterogeneity of breast cancer, the better we can provide accurate drug targets for personalized therapy.

The lower H2Bub level in claudin-low subtype is a interesting finding which suggests that H2Bub1 is a critical and consistent marker for cells with stem cell features. There are a number of DUBs and E3 ligases for H2Bub1, and more proteins that regulate the enzymes in mammalian cells. The gain or loss of function of these proteins may result from gene copy number amplification, tranlocation, mutation, or deletion. Therefore, it is a rather difficult task to identify the reasons for lower H2Bub1 level in claudin-low subtypes. At least in our study, through examining the mRNA and protein levels of USP22, USP27x, USP51, USP44, USP49, RNF20, RNF40, ATXN7L3, and ENY2, we have not found the reasons for lower H2Bub1 level in claudin-low cells. We may be able to get insight of the causes by analyzing gene expression profiles of patients from large datasets. However, it is very possible that in each cell line or tumor, there is

different reason causing lower H2Bub1 level, which makes finding the causes even more difficult.

By analyzing the differential H2Bub1 enriched genes in claudin-low versus basal-like subtypes, we identified miR200 gene locus, including miR200a, miR200b, and miR429, is significantly lower enriched of H2Bub in claudin-low cells compared to basal-like cells. Interestingly, overexpression miR200b only inhibits the cell proliferation and colony formation of claudin-low cells, indicating that miR200b can be a therapy target only for claudin-low subtype breast cancer.

microRNA, as a therapeutic entity, is in the stage of preclinical research or early clinical trials. To specifically increase miR200 levels from its upstream in claudin-low cells is another potential therapeutic method. Since miR200 gene locus is enriched of less H2Bub1 in claudin-low cells, increasing H2Bub1 level may affect the miR200 transcription status and levels in claudin-low cells. There are many ways to manipulate H2Bub1 levels in cells, such as decreasing the levels of deubiquitinases, using deubiquitinases inhbitors, or increasing the E3 ligases protein levels. This may also provide information as to figure out the reason for low global H2Bub1 level in claudin-low subtype.

Since H2Bub1 is a main substrate for DUB module in SAGA complex, we also proceed to study the DUB module-related function. We were specifically interested in a protein called ATXN7L3B, an analog of ATXN7L3. ATXN7L3B is

reported to be depleted in a family with neurodegenerative disease (76). Three members in a family, lacking the chromosome region 12q21 where ATXN7L3B lies, exhibited different degrees of motor and cognitive milestones delay, learning difficulties, and cerebellar ataxia. In another study, Tan and colleagues identified ATXN7L3B, also named lnc-SCA7, as a long non-coding RNA (77). Our work, using specific antibody targeting C-terminus of ATXN7L3B, successfully and solidly proved that ATXN7L3B gene encodes a 97 amino acids ~11kD protein. Even though the previous study failed to make this point, it showed that ATXN7L3B and ATXN7 mRNA levels are positively correlated, and both are regulated by miR-124 post-transcriptionally. As stated in the Tan's study, since miR-124 transcription is activated by SAGA complex and incorporation of polyQ-ATXN7 into SAGA reduces its activity, miR-124 expression is repressed in SCA7 disease and mutant polyQ-ATXN7 proteins accumulate and form cytotoxic nuclear inclusions. They also reported that the ATXN7L3B is highly expressed in retina, cerebellum, hippocampus, but its expression in breast tissue was not investigated.

The effects of ATXN7L3B on USP22 activity in vivo indicate that it might also impact USP22 functions in cancer. The functions of USP22 in cancer have been summarized in Chapter 1.6. Our finding that ATXN7L3B protein level is higher in ER positive breast cancer cells suggests a balance in the activity of all these proteins is likely important for normal cell function. Going forward, it will be important to determine whether overexpression of USP22, ENY2, and

ATXN7L3B as observed in cancer patients can directly promote cancer progression and metastasis. If so, these proteins may provide unique targets for small molecule inhibition as a cancer therapy.

In the future study for AFAP1-AS1, we can investigate the regulatory mechanism for high expression of AFAP1-AS1 in TNBC cells. It is possible that the mutation or high expression of certain transcription factors or co-transcription factors results in high expression of AFAP1-AS1 in TNBC cells. We can analyze the putative transcription factor-binding sites or enhancers near AFAP1-AS1 gene locus. Then ChIP assay can be adopted to verify whether the binding actually occurs. Furthermore, we can over express or knock down these identified transcription factors or related co-activators to study whether the changes in protein levels of these putative regulators can result in changes in AFAP1-AS1 levels. Meanwhile, we can also study the downstream effects of AFAP1-AS1 and the specific functions in TNBC cells. It is possible that AFAP1-AS1 specifically targets certain genes and regulates their transcription or translation. Even though AFAP1-AS1 shares the same gene locus with AFAP1, AFAP1 has been verified both in previous study and our study not to be a target for AFAP1-AS1. It would be interesting to identify the targets for AFAP1-AS1 and see whether the function of AFAP1-AS1 in TNBC is through regulating these targets. By studying the functions of AFAP1-AS1 itself, its upstream regulators, and downstream effectors, we may provide new therapy targets specifically for patients with TNBC, which currently lacks efficient target therapy.

In this study, we successfully identified breast cancer subtypes at the epigenetic levels, which indicates that epigenetic changes drive the molecular patterns of distinct subtypes. We also specifically characterized H2Bub1, miR200b, and AFAP1-AS1 in TNBC, and ATXN7L3B in ER positive breast cancer cells. All of these genes can be possible therapy targets for breast cancer patients. Besides studying the functions of ATXN7L3B in breast cancer, we also investigated its molecular roles in regulating ENY2, which exists in SAGA and other complexes, such as TREX-2. TREX-2 complex is essential in mRNA export. Silencing of ENY2 or other components results in defects in mRNA export to cytoplasm (63). In the future, we will investigate whether ATXN7L3B affects TREX-2 complex function and mRNA export.

APPENDICES

Appendix 1. List of genes highly enriched of H2Bub in basal-like cells.

Gene	TNBC-Basal HCC1937	TNBC-Basal MB468	TNBC-ClaudinLow MB231	TNBC-ClaudinLow MB436
EEFSEC	2.98	2.85	1.29	0.98
ZNF148	4.54	4.56	1.79	1.49
TMCC1	2.39	2.65	0.94	0.84
RPL32P3	3.06	3.94	1.5	1.59
ATP2C1	2.62	2.42	1	0.94
MRPL3	3.44	2.5	1.45	1.17
MRPS22	3.4	3.16	1.58	1.59
SR140	3.98	3.42	1.79	1.58
WWTR1	3.15	2.37	1.4	1
RNF13	4.04	2.32	1.18	0.75
DHX36	4.45	2.4	1.19	1.17
NMD3	4.06	2.26	1.23	1.13
EPHB3	4.44	4.1	0.91	0.47
HES1	3.94	3.28	1.21	0.58
COX17	5.19	2.39	1.13	1.67
PTPLB	3.9	2.85	1.03	1.44
UMPS	3.8	3.16	1.43	1.31
SEC61A1	3.22	3.1	1.69	1.58
RUVBL1	4.33	2.49	1.74	1.36
ACAD9	3.28	2.86	1.46	1.1
RAB43	2.78	4.32	1.68	1.16
MBD4	2.46	3.09	0.98	1.09
COPB2	3.02	2.79	1.58	1.34
ATP1B3	3.58	3.1	1.46	1.66
C3orf58	3.59	4.42	1.06	1.39
TSC22D2	6.78	3.73	1.51	1.54
EIF2A	5.09	2.66	1.23	1.08
RAP2B	4.08	5.56	1.07	1.37
SLC33A1	3.14	2.48	0.95	1.08
GMPS	6.08	3.67	1.72	1.55
TIPARP	6.5	2.98	1.68	1.7
MIR16-2	7.17	4.54	1.57	1.19
TRIM59	2.89	2.77	0.73	1.08
SMC4	4.2	2.73	1.18	1.1

SKIL	3.28	2.75	1.14	1.2
MIR1224	3.53	2.24	1.7	0.61
SNORD66	6.22	3.47	1.19	1.63
TMEM41A	5.96	2.44	1.79	1.46
SNORA4	10.2	3.19	1.01	1.53
RPL39L	2.38	2.44	1.28	1.14
BCL6	3.79	3.38	1.57	0.92
EFNB1	3.58	3.18	1.31	1.71
ZDHHC20	2.66	2.81	0.98	1.07
PSPC1	3.08	3.02	1.61	1.62
XPO4	3.56	2.49	1.75	1.48
CDK8	3.08	3.29	1.46	1.67
LNX2	2.51	2.47	1.02	1.25
USPL1	2.5	2.6	1.73	1.81
ELF1	2.85	2.45	1.41	1.43
NDFIP2	3.88	4.52	1.13	1.33
RBM26	3.59	3.08	1.3	1.48
POLR1D	2.92	2.92	1.55	1.74
POU4F1	2.4	4.56	0.47	1.01
GPR180	3.7	2.63	0.87	1.63
RAP2A	4.05	3.44	1.26	1.73
ABHD13	4.23	5.43	1.52	1.65
KDM5A	3.31	2.49	1.19	1.14
ERC1	2.86	2.24	1.14	1.72
KRT18	2.77	3.28	1.09	1.52
RAD51AP1	3.11	2.3	1.09	1.63
MIR141	3.3	2.95	0.95	1.48
KRT8	3.17	3.51	1.34	0.67
MIR196A2	2.85	3.68	0.94	0.7
MYL6	2.54	2.77	1.32	1.74
ST14	2.84	4.09	0.86	0.41
C10orf10	2.71	6.93	1.24	1.28
CYTSB	2.5	2.35	1.67	1.35
ALOX12P2	2.81	2.6	0.52	0.75
CLDN7	2.62	4.63	1.06	1.03
SNORD65	3.61	4.51	1.25	1.68
SNORD42B	2.34	2.44	1.44	1.38
SNORD42A	2.57	3.33	1.23	0.87
TOB1	3.82	4.68	1.48	1.74
SOX9	2.29	7.21	1.64	0.66
C17orf86	2.83	2.22	1.79	1.48

CBX4	2.89	2.93	1.64	1.48
C15orf61	2.56	2.25	1.42	1.28
MUDENG	2.27	2.51	1.67	1.3
NAA30	2.58	2.59	1.7	1.04
PLEKHG3	3.07	2.6	1.64	1.48
LPHN1	3.04	3.07	1.5	1.38
CRB3	2.29	2.34	1.02	0.42
WDR83	2.8	2.31	1.79	1.7
B3GNT3	2.83	3.92	0.81	0.81
SPINT2	3.26	4.38	1.23	0.38
PTPN2	2.91	2.9	1.78	1.17
CEP192	2.84	2.65	1.7	0.88
PSMG2	2.26	2.84	1.6	0.96
NDUFV2	3.93	4	1.77	1.53
CHMP1B	2.46	3.62	1.46	1.38
ZNF519	3.29	3.34	1.4	1.03
SIRPA	2.3	2.68	0.78	1.51
PCNAAS	2.77	2.92	1.33	1.04
MAFB	3.44	3.1	1.57	0.56
RIPK4	3.09	5.96	1.78	0.73
C7orf36	2.59	2.42	1.75	1.05
LRRC16A	2.45	2.79	1.14	0.69
KCNK5	3.08	3.22	1.13	1.03
LRRC1	2.37	2.37	1.76	0.78
HIST1H2AM	2.54	3.54	1.6	1.25
WWC1	2.33	3.64	1.49	1.67
C5orf38	3.41	2.6	0.84	0.85
FASTKD3	3.44	2.26	1.66	1.57
IRX2	5.53	4.88	1.2	0.73
PCDHGC5	2.45	2.21	1.62	1.75
PCDHGC4	2.83	2.56	1.52	1.65
C5orf54	2.35	2.91	1.67	1.14
PTTG1	2.7	3.41	1.75	1.59
RPL26L1	2.41	2.96	1.28	1.79
MIR575	2.77	3.49	1.36	0.79
ASXL2	2.76	3.12	1.69	1.31
EPT1	2.26	2.8	1.67	1.23
BIRC6	2.33	2.29	1.4	1.15
ADCY3	2.57	2.6	1.81	1.11
GTF3C2	2.56	3.7	1.63	1.03
TRMT61B	2.39	2.6	1.54	0.97

RTKN	3.46	3.27	1.3	1.56
SNORA41	2.55	2.87	1.56	1.66
ATP1B1	2.41	2.24	1.44	1.59
ELF3	5.11	8.01	0.98	1.46
MIR200B	5.92	18.79	0.46	0.2
MIR200A	5.17	15.05	0.72	0.12
MIR429	3.36	13.74	0.77	0.29
SFN	4.25	3.86	1.29	0.74
TACSTD2	3.92	8.09	1.07	0.75
PTGFRN	2.38	3.02	0.57	1.05
S100A14	2.31	3.39	1.08	1.04
RAB25	2.56	3.3	0.82	0.68
NCSTN	2.42	2.45	1.56	1.66
F11R	3.23	3.87	0.92	1.47
ALDH9A1	2.7	2.24	1.79	1.71
LAD1	3.36	3.13	0.81	0.93
PHLDA3	2.71	3.2	1.49	1.03
WNT9A	2.91	2.52	1.21	1.65
C1orf131	3.45	2.66	1.64	1.77
C1orf124	3.02	2.59	1.73	1.57
KCNK1	3.48	3.04	0.57	1.74
FH	2.65	2.25	1.34	1.8
OSTF1	2.64	2.25	1.31	1.35
CDKN2A	2.4	2.43	0.01	1.32
LCN2	3.91	2.21	1.05	0.94
VPS13B	3.07	2.73	1.45	0.94
STK3	4.62	2.73	1.29	0.9
SFRP1	4.01	2.61	0.49	0.48
RNF19A	3.03	4.57	1.67	1.15
GRHL2	2.43	2.36	0.53	0.37
OXR1	3.54	2.42	1.35	0.65
TRIM35	2.22	2.63	1.8	1.81
OSR2	5.64	5.16	1.14	0.83
POLR2K	5.01	4.88	1.75	1.49
ANKRD46	2.87	3.37	1.54	0.89
MAL2	5.79	4.17	0.49	1.15
GSDMC	3.08	2.23	0.61	0.95
EPPK1	4.59	5.02	1	1.8
BID	3.15	3.56	1.78	1.66
MIR659	2.73	2.34	0.05	1.63
WNT7B	2.48	6.11	0.9	0.44

PAPD5	4.59	2.26	1.46	1.28
OGFOD1	2.56	2.21	0.76	0.84
CSNK2A2	3.6	2.52	1.26	1.14
ZNRF1	3.75	3.59	1.63	1.56
AFG3L1	3.81	2.74	1.51	1.32
C16orf67	4.26	2.27	1.56	1.71
IRX3	2.39	5.88	1.27	0.53
NUDT21	5.54	3.96	1.49	1.56
AMFR	2.57	2.46	1.2	0.97
MT1X	2.44	2.84	1.02	0.98
FAM192A	4.35	2.47	1.31	1.04
CIAPIN1	3.43	2.63	1.38	1.46
POLR2C	4.03	3.39	1.76	1.64
ZNF319	2.7	3.75	0.85	1.33
SETD6	2.7	2.66	1.01	0.95
SNORA46	6.43	3.3	1.24	1.76
GOT2	4.24	2.82	1.46	1.35
AGRP	3.31	2.75	1.42	0.86
GABARAPL2	4.05	2.87	1.76	1.37
ZCCHC14	3.51	3.07	1.64	1.16
MVD	2.84	3.41	1.66	0.94
TRAPPC2L	3.78	2.62	1.57	1.48
APRT	4.11	3.24	0.99	1.47
CDT1	4.35	2.92	1.54	1.65
ZNF778	2.77	2.39	1.29	1.2
SPG7	3.38	2.36	1.7	1.45
CDK10	2.78	2.54	1.57	1.01
ZNF276	3.11	2.45	1.57	1.25
DEF8	2.43	2.37	1.67	1.65
C16orf3	3.78	2.39	1.62	1.5

Appendix 2. List of genes highly enriched of H2Bub in claudin-low cells.

Gene	TNBC-Basal HCC1937	TNBC-Basal MB468	TNBC-ClaudinLow MB231	TNBC-ClaudinLow MB436
GCN1L1	1.77	1.17	2.34	2.22
DENND5A	0.81	1.59	2.49	3.34
SNORA40	1.7	0.49	2.23	2.53
TCF7L2	1.19	1.36	2.65	2.42
CASP7	1.41	1.27	2.22	2.6
SMNDC1	1.74	1.76	2.36	3.37
PDZD8	1.72	1.22	2.39	3.02
NXN	0.49	1.36	2.44	2.47
ZBTB4	1.73	1.61	2.62	3.24
PER1	1.79	0.88	2.52	2.87
ZNF271	0.9	1.7	2.52	3.66
MIR645	1.5	0.49	3.49	2.78
IFNGR2	1.19	1.79	2.4	3.03
HIPK2	0.88	1.29	2.74	2.45
VKORC1L1	1.56	1.35	2.3	2.37
PTPN12	1.75	1.6	2.51	2.6
FLJ43663	0.89	1.68	2.32	3.33
GALNT11	1.62	1.58	2.7	2.78
ESYT2	1.69	1.07	2.31	2.89
C7orf54	1.21	1.51	2.61	2.23
AKR1B1	1.5	0.99	2.23	3.05
INSIG1	1.42	1.67	2.81	2.5
C6orf147	1.67	1.53	2.57	2.35
ADD1	0.99	1.77	2.98	2.8
STX18	1	1.72	2.22	2.41
SLAIN2	1.79	1.39	2.24	2.59
OSTC	1.6	1.26	2.22	2.61
FGFRL1	0.94	0.71	3.63	2.65
SH3BP2	0.62	1.37	2.51	2.46
MFSD10	0.97	1.77	2.4	2.9
BIN1	0.94	0.73	2.45	2.69
DUSP10	1.68	1.35	2.79	2.98
RCC1	1.74	1.64	2.44	3.19
ZNF697	0.98	1.19	3.36	5.21
ADAMTSL4	0.84	1.69	2.35	2.55
C9orf25	1.53	1.81	2.36	2.8

GPSM1	1.15	1.74	2.31	2.56
C9orf172	1.47	1.34	2.22	3.29
DPP7	1.39	1.81	2.5	2.96

Appendix 3. Distributed normalized spectral abundance factor (dNSAF) of FH-ATXN7L3, FH-ATXN7L3B, or Vector pulled down proteins.

NCBI_Gene	FH-ATXN7L3	FH-ATXN7L3B	Vector
TADA3	0.00422486	0.00002576	0
USP22	0.06286145	0.00125789	0
TAF10	0.00374286	0.0001021	0
TAF6L	0.00248553	0.00012525	0
HRNR	0.00016575	0.00001171	0
KPNA1	0.0003592	0.00004137	0
S100A9	0.00621564	0.00097623	0
SF3B3	0.00051166	0.00011888	0
SRRM2	0.0000156	0.00000404	0
LSM3	0.00084205	0.00021822	0
RPLP1	0.00075341	0.00019525	0
SNRPE	0.00046679	0.00012097	0
DUT	0.00052371	0.00013572	0
IK	0.0001542	0.00003996	0
PARK7	0.00045444	0.00011777	0
DLST	0.0000948	0.00002457	0
PSAP	0.00044819	0.00012671	0
HSPA2	0.00014345	0.00004208	0
DDX3X	0.00057201	0.0001852	0
YWHAQ	0.00154609	0.00050987	0
RPS20	0.00166334	0.00054862	0
GGCT	0.00034264	0.00011839	0
PDIA6	0.0001464	0.00005059	0
PRSS3P2	0.00069546	0.00027034	0
EIF5A	0.00058348	0.00024194	0
RPS16	0.00132363	0.00060981	0
C1QBP	0.00510155	0.00248628	0
CTTN	0.00007808	0.00004047	0
HBS1L	0.00003345	0.00001733	0
LGMN	0.00009918	0.0000514	0
SRP14	0.00047365	0.00024549	0
RPL5	0.00014459	0.00007494	0
PEPD	0.00008711	0.00004515	0
PSMA2	0.00036705	0.00019024	0
SUB1	0.00033815	0.00017526	0
CALR	0.00030895	0.00016013	0

GOT2	0.00114851	0.00059527	0
G6PD	0.00008339	0.00004322	0
RPL12	0.00052054	0.0002698	0
EEF1B2	0.00038173	0.00019785	0
SMIM4	0.00061349	0.00031797	0
PUF60	0.00015847	0.00008213	0
SERBP1	0.00015788	0.00008183	0
COPRS	0.00023339	0.00012097	0
RPS12	0.00032534	0.00016862	0
CAPZA1	0.00015016	0.00007783	0
RPS26	0.00037343	0.00019355	0
ETF1	0.00009827	0.00005093	0
CA2	0.00016517	0.00008561	0
G3BP1	0.00009216	0.00004776	0
CASP12	0.00006297	0.00003264	0
HNRNPH2	0.00054426	0.00029491	0
CIRBP	0.00099871	0.00058233	0
RPSA	0.00116459	0.00067906	0
RPS10-NUDT3	0.00051651	0.00030595	0
RPS18	0.00127138	0.00080539	0
ALYREF	0.00056934	0.0003794	0
MIF	0.00093357	0.00067742	0
ACTA2	0.0003873	0.00029274	0
UTP15	0.0000829	0.00006445	0
RPS5	0.00042102	0.00032733	0
STMN1	0.00024681	0.00019188	0
SLIRP	0.00118196	0.00091891	0
SNRPN	0.00089468	0.00069557	0
ENY2	0.36234364	0.29294913	0
PRDX6	0.00028757	0.00024842	0
LRPPRC	0.00016944	0.00015169	0
DDX39B	0.00020067	0.00018202	0
SF3B14	0.00068711	0.00062323	0
DDB1	0.00041438	0.00039049	0
HIST1H2BB	0.00102249	0.00097158	0
YWHAE	0.00129978	0.00124695	0
EIF4B	0.00003514	0.00003643	0
SFPQ	0.00015185	0.00015741	0
NHP2	0.00023858	0.00024731	0
SH3BGRL	0.00037671	0.00039049	0
SNAP29	0.00008323	0.00008627	0

ILF2	0.00016517	0.00017122	0
RPL18	0.00027009	0.00027998	0
CUTA	0.00010845	0.00011241	0
PSMB3	0.00010474	0.00010858	0
CDKN2A	0.00025715	0.00026656	0
RPL6	0.00014911	0.00015457	0
GCSH	0.00012412	0.00012866	0
PTCD3	0.00009349	0.00009691	0
KHSRP	0.0000604	0.00006261	0
TCEAL1	0.00202568	0.00223981	0
RPS9	0.00033204	0.00040156	0
FBXO38	0.00009646	0.00011999	0
HNRNPU	0.00010656	0.00013808	0
RPL31	0.0003355	0.00043473	0
GLO1	0.00023339	0.00030242	0
RPL7A	0.00016145	0.00020919	0
RPS23	0.0013514	0.00178999	0
HNRNPF	0.00069603	0.0009345	0
RBM39	0.00016907	0.00024098	0
MDH2	0.00165171	0.00250239	0
DDX21	0.00006006	0.00009339	0
CFL1	0.0002587	0.00040225	0
SRSF7	0.00019002	0.00029546	0
PRDX2	0.00021689	0.00033724	0
RPS3A	0.00008133	0.00012647	0
TIMM8A	0.00022136	0.0003442	0
MCM3	0.00015104	0.00024789	0
RPS2	0.00036642	0.00064571	0
PGAM1	0.00016907	0.00030671	0
LAMP2	0.00010449	0.00018955	0
RPS6	0.0002587	0.00049164	0
TIMM13	0.00090409	0.00175722	0
NME1-NME2	0.00020105	0.00039598	0
RPS27L	0.00025562	0.00052995	0
HBA1	0.00030243	0.00062699	0
RPS13	0.0001422	0.00029481	0
RPL24	0.00082059	0.00177214	0
MATR3	0.0001014	0.00022337	0
PRDX3	0.00018044	0.00042085	0
HSPH1	0.0001001	0.0002358	0
RPL23	0.00490794	0.01184449	0

RPS4X	0.00073479	0.00177726	0
PCBP2	0.00023726	0.00058412	0
MTHFD1	0.00004593	0.00011903	0
RPL13	0.00010176	0.00026372	0
NDUFA4	0.00053018	0.00137396	0
TCEB2	0.00036394	0.00103745	0
ATP5C1	0.00028919	0.00082437	0
CDKN2A	0.00032534	0.00092742	0
PRDX4	0.00015847	0.00045173	0
YWHAG	0.0001917	0.00055172	0
EEF1A1	0.00125487	0.00380604	0
ELAVL1	0.00006587	0.00020483	0
AASDHPPT	0.00013898	0.0004322	0
VCP	0.00007992	0.00024854	0
DDX47	0.00009438	0.00029351	0
DNAJB6	0.00052693	0.00169514	0
MCM5	0.00011701	0.00037905	0
EEF2	0.00040041	0.00129709	0
RPS14	0.0004266	0.00140034	0
EIF4A1	0.00010577	0.00035635	0
FAM207A	0.00018671	0.00062903	0
RCN2	0.00047415	0.00161494	0
USP11	0.0039912	0.01378708	0
CCT7	0.00003954	0.00014347	0
KPNB1	0.00004902	0.00017786	0
SRPX	0.00004836	0.00017546	0
NONO	0.00018235	0.00070886	0
TUBB3	0.00067426	0.00269867	0
BAG2	0.00020353	0.00084391	0
RPS27	0.00025562	0.00105991	0
USP7	0.00005845	0.00024238	0
PPP2CA	0.00013898	0.00061228	0
TCEB1	0.00038343	0.00168923	0
DNAJB1	0.00006315	0.00029459	0
PFKM	0.00010093	0.00048277	0
IPO5	0.00003852	0.00018964	0
HNRNPM	0.00005883	0.00028966	0
POLR2C	0.00015616	0.00076892	0
VIM	0.00009216	0.00045376	0
FAR1	0.00016677	0.00090761	0
ATP5B	0.00008118	0.00048387	0

AHSA1	0.00006353	0.00042804	0
RPS11	0.0001359	0.00091568	0
HADHB	0.0000906	0.00063393	0
NAP1L1	0.00010983	0.00078938	0
POLR2B	0.00005487	0.00039814	0
ATAD3A	0.00025649	0.00205591	0
SKP1	0.00026346	0.00211657	0
DEDD	0.00006752	0.00055995	0
NSUN2	0.00005867	0.00051692	0
NTPCR	0.00011301	0.0013472	0
CHP1	0.00022023	0.00279653	0
RPL38	0.00061349	0.0082673	0
CNP	0.000051	0.00071374	0
TUBA4A	0.00014379	0.00217296	0
FAM172A	0.00005162	0.00085608	0
DNAJA2	0.00020847	0.00437599	0
TUBB4A	0.00007254	0.00188942	0
HADHA	0.00002814	0.00091891	0
IMPDH2	0.00008355	0.00285478	0
ATXN7L3	0.1697357	0	0
TADA2B	0.0035787	0	0
TAF9	0.00338892	0	0
ATXN7L2	0.00324165	0	0
HIST2H2AA4	0.00264273	0	0
CCDC101	0.00263822	0	0
ATXN7	0.0025903	0	0
TAF5L	0.00226023	0	0
TADA1	0.00224337	0	0
SUPT3H	0.00223528	0	0
ATXN7L1	0.00196539	0	0
ATXN7L1	0.00176158	0	0
TRRAP	0.00154685	0	0
SF3B5	0.00124838	0	0
RPL30	0.00112029	0	0
RPS17L	0.00111337	0	0
SUPT7L	0.00103731	0	0
KAT2A	0.00091392	0	0
SUPT20H	0.00090961	0	0
MED21	0.00089468	0	0
TAF9B	0.00071289	0	0
PTBP1	0.00058561	0	0

ACAD11	0.00049551	0	0
RPS19	0.00044425	0	0
LCN2	0.00043378	0	0
SUMO1	0.00042519	0	0
RPS8	0.00041293	0	0
CTSD	0.00036482	0	0
DDT	0.00036394	0	0
CASP14	0.00035491	0	0
RRP7A	0.00030675	0	0
LOR	0.00027528	0	0
DCAF5	0.00027353	0	0
RPL29	0.00027009	0	0
TAF12	0.00026674	0	0
RRP15	0.00022843	0	0
MRPS23	0.00022602	0	0
SERPINB4	0.00022023	0	0
HNRNPC	0.00021051	0	0
KHDRBS1	0.00019388	0	0
THRAP3	0.00015739	0	0
TMEM106B	0.00015673	0	0
PPM1G	0.00015731	0	0
TPM3	0.00015068	0	0
BLMH	0.00014158	0	0
NOLC1	0.00012287	0	0
SRSF11	0.00008891	0	0
SMU1	0.00008371	0	0
ARID3B	0.00007669	0	0
INPP4B	0.00006971	0	0
DTNB	0.00006849	0	0
KAT2B	0.00006129	0	0
AP3B2	0.00003969	0	0
USP27x	0.00547978	0.00153121	0.00021091
USP51	0.00180923	0.00068216	0.00018936
HNRNPK	0.00208693	0.00060092	0.0003002
TRIM28	0.0006686	0.0006131	0.00016646
TUBB	0.00394462	0.03009542	0.00098385
HNRNPH1	0.00235126	0.00099776	0.00061911
TUBA1C	0.00219983	0.01733588	0.0007223
FLG2	0.00013471	0.00002793	0.00005813
DDX17	0.0004272	0.0000883	0.00019014
SNRPD1	0.00830018	0.00495664	0.00467196

HSPA1A	0.01488679	0.00896966	0.00923623
TUBA1A	0.00157114	0.02075566	0.00102728
S100A8	0.00369414	0.00023933	0.00298905
HSPA8	0.00907282	0.00985764	0.00752046
YWHAZ	0.00067641	0.00152961	0.00056731
RUVBL1	0.00032962	0.00207449	0.0003048
SLC25A5	0.00151314	0.00414376	0.00139924
BCLAF1	0.00017296	0	0.00015994
RPL11	0.00084441	0.00062523	0.00078085
WDR77	0.0059645	0.00250566	0.00609608
PRMT5	0.01422492	0.0103778	0.01483731
HSPA5	0.01114203	0.00194257	0.01178459
GAPDH	0.00175882	0.00113949	0.00189748
KCTD5	0.00110114	0.00057072	0.00118795
IRS4	0.00017082	0.00131034	0.00022115
HNRNPA1	0.00028861	0	0.00037363
PFN1	0.00076686	0.00015899	0.00099279
RIOK1	0.00018902	0.00005878	0.0002447
DDX5	0.00016953	0.00042051	0.00022637
NPM1	0.00204497	0.00045425	0.00283654
DNAJA1	0.00048678	0.0063074	0.00070021
CLNS1A	0.00679501	0.0014557	0.0099698
HSPA9	0.00208714	0.00075396	0.00307049
TPI1	0.00315326	0.00101173	0.00485981
DSP	0.00015706	0.0000659	0.00024206
S100A7	0.0053149	0.00077132	0.00825687
HSP90AB1	0.00011863	0.00044699	0.00019198
CCT3	0.00015759	0.00065345	0.00025503
SLC25A11	0.00027353	0.0009924	0.00044265
RPLP0	0.00054189	0.0004915	0.00087691
RPS3	0.0007069	0.0000916	0.00114396
ACTG1	0.00155745	0.00071473	0.00259449
PIP	0.00279433	0.00060981	0.00475996
C19orf10	0.00086882	0.00025732	0.00160683
PSPC1	0.00028739	0.00031919	0.00053151
DSG1	0.00020469	0.00009548	0.00039749
PCBP1	0.00036189	0.00046892	0.00078085
HSPE1	0.0035787	0.00185484	0.00817592
MYCBP	0.00229315	0.00270123	0.0053977
JUP	0.00023057	0.00005975	0.00055969
ERH	0.00165171	0.00042804	0.00400935

SERPINB3	0.00044046	0	0.00106916
SSBP1	0.00072541	0.00067677	0.00187825
AZGP1	0.00036027	0.00007469	0.00093282
RPLP2	0.00224058	0.00164516	0.00604307
TUBB2B	0.00049714	0.00154602	0.00147245
ENO1	0.00084108	0.0005385	0.00256204
MCM7	0.00005973	0.00044888	0.00019331
SERPINF1	0.00010274	0.00002662	0.00033251
HIST2H4B	0.00125081	0.00075634	0.00404827
PNKD	0.00060485	0.00094048	0.00195762
SNRPD2	0.00278344	0.00206094	0.00900866
RPS25	0.00068711	0	0.00222385
RPS7	0.00022136	0	0.00071645
KRT80	0.00015265	0.00005274	0.00065872
ARG1	0.00020005	0.00003456	0.0008633
PYCRL	0.00022523	0.00035021	0.00097196
ATP5A1	0.00017075	0.00137177	0.00082897
CNPY2	0.00047192	0.0001223	0.00229106
SET	0.00038759	0.00020089	0.00200709
TGM3	0.00015492	0.00003212	0.00080226
NACA	0.00049935	0.00082821	0.00258587
HIST1H1C	0.00191536	0.00177647	0.01044062
HIST1H1B	0.00066507	0.00039395	0.00369002
TXN	0.0017683	0.00039279	0.00981111
HSP90B1	0.00002674	0.00003355	0.00017309
PSMA7	0.00008658	0.00058338	0.00056045
RUVBL2	0.00009275	0.00175469	0.00060039
TMEM169	0.00014459	0.00014989	0.00093596
GLRX5	0.00027353	0.00021266	0.00177058
MYL12B	0.00024968	0	0.00161617
TBCA	0.00039763	0	0.0025739
SNRPD3	0.00068166	0.00008833	0.0044124
PFDN2	0.00027886	0.00144533	0.00180507
PSMD4	0.00005696	0.00008856	0.00036868
AIFM1	0.00014103	0.00089544	0.00091291
SRP9	0.00049935	0.00051763	0.00323234
CSTB	0.00043821	0	0.00283654
CCT2	0.000088	0.00034208	0.00056963
RPS28	0.00062238	0.00048387	0.00402872
ANXA6	0.000067	0.00005209	0.00043367
HNRNPD	0.00014034	0.00003637	0.00090844

TCP1	0.00015448	0.00116094	0.00099993
PARP1	0.00004235	0.00010975	0.00027414
TUBA1B	0.00014283	0	0.00102728
HSPD1	0.00018737	0.00040787	0.0014554
NCL	0.00015121	0	0.00117457
DCD	0.00448965	0.00121408	0.03537945
PFKFB3	0.00017178	0	0.00138991
MYL6	0.0004266	0.00019654	0.00368187
SOD1	0.00083658	0.00028907	0.0072203
CCT8	0.00015673	0.00022339	0.0015218
TRIM21	0.00018082	0.0003983	0.00175567
CALM3	0.00144109	0.00074692	0.01399235
NENF	0.00049935	0	0.00484851
KIF11	0.00020334	0.00031617	0.00210592
SERPINB12	0.00015905	0.00005496	0.00171593
RBBP4	0.00005064	0.00023331	0.00065562
RPL4	0.00010057	0.00075584	0.00130202
RBBP7	0.00004578	0.00026366	0.00059271
YBX1	0.00013254	0.0000687	0.00171593
HNRNPA2B1	0.00012166	0	0.00157497
RBMX	0.00005492	0	0.00071095
GAR1	0.0003958	0.00051286	0.00512408
PEBP1	0.0004593	0.00035708	0.00594613
RANBP1	0.00042731	0.00160568	0.00553197
CSE1L	0.00002211	0.00020631	0.00028628
CSTA	0.00065731	0.00022712	0.00992791
MYH10	0.0000214	0.00001109	0.00041552
NME1	0.00030328	0.00179197	0.00628207
TUBB4B	0.00007238	0.00095077	0.00160631
PPIA	0.00026027	0.00074194	0.00589657
PPIB	0.00019882	0.00010305	0.0051478
LCN1	0.000244	0	0.00631776
PKM	0.00004044	0.00020959	0.00104701
ANXA5	0.0001342	0.00027823	0.00434346
ATXN7L3B	0	0.17072197	0
PFDN5	0	0.00274613	0
MLLT11	0	0.00210215	0
HAX1	0	0.00163545	0
C12orf23	0	0.00153504	0
C19orf70	0	0.0013204	0
POLR2J3	0	0.00116129	0

NDUFB4	0	0.00111291	0
HSPB1	0	0.00108576	0
TIMM50	0	0.00107386	0
COX5A	0	0.00096452	0
SERPINH1	0	0.00090523	0
TUFM	0	0.00083162	0
CDC42EP1	0	0.00082543	0
QPCTL	0	0.00081574	0
CUL2	0	0.00080667	0
DPY30	0	0.0007869	0
FBXL12	0	0.00078518	0
RPL10	0	0.00078007	0
PSMC4	0	0.00077211	0
VBP1	0	0.00067791	0
ZNF229	0	0.00064751	0
PSMA4	0	0.0006396	0
YME1L1	0	0.00063348	0
OSGEP	0	0.0006312	0
SELRC1	0	0.00062631	0
PPP2R1A	0	0.00062353	0
KLHDC3	0	0.00061181	0
CACYBP	0	0.00060157	0
ETFA	0	0.0005878	0
TUBB6	0	0.00054874	0
PFDN1	0	0.00054733	0
UQCR10	0	0.00052995	0
MRPS17	0	0.00051365	0
PTRH2	0	0.00049739	0
EIF2S3	0	0.00049515	0
UQCRH	0	0.00048919	0
FAM96B	0	0.00047793	0
CHCHD4	0	0.00047024	0
LYPLAL1	0	0.00046958	0
HSD17B12	0	0.00046371	0
RFC4	0	0.00045988	0
TMX3	0	0.00044124	0
ATP1B3	0	0.00043878	0
PFKP	0	0.00043442	0
PRPSAP1	0	0.0004336	0
SLC25A13	0	0.00042551	0
IMPDH1	0	0.00041853	0

POLR1C	0	0.00041814	0
RPL21	0	0.00041734	0
NDUFB10	0	0.00038822	0
NDUFA13	0	0.00038643	0
PSMC5	0	0.00038376	0
PBDC1	0	0.00038211	0
NDUFS3	0	0.0003794	0
CCT6A	0	0.00037726	0
RPL27A	0	0.00037598	0
PHB2	0	0.00037221	0
POLR2E	0	0.00037097	0
CBX3	0	0.00036489	0
PFKL	0	0.00036357	0
GLUD1	0	0.000359	0
DNAJB2	0	0.00035533	0
SSR1	0	0.00035021	0
PFDN6	0	0.00034509	0
HSPA6	0	0.00034209	0
DPM1	0	0.00034243	0
RPL19	0	0.00034069	0
LPCAT4	0	0.00033982	0
NOP56	0	0.00033724	0
DRG1	0	0.00033357	0
PTPMT1	0	0.00033221	0
XRCC6	0	0.00032894	0
MGARP	0	0.0003246	0
DCAF7	0	0.00032541	0
ALDH1B1	0	0.00032289	0
RBFOX2	0	0.00032079	0
TELO2	0	0.00031911	0
FLYWCH2	0	0.00031797	0
CXorf57	0	0.00031239	0
ZNF281	0	0.00031087	0
RPL14	0	0.00031058	0
PSMC2	0	0.00030843	0
CCT4	0	0.0003061	0
PET100	0	0.00030491	0
LAS1L	0	0.00030324	0
TRUB2	0	0.0003026	0
HEATR6	0	0.00030155	0
AMBRA1	0	0.0002994	0

RSU1	0	0.0002981	0
TARDBP	0	0.0002957	0
ACTR1A	0	0.00029599	0
AURKB	0	0.00029384	0
HSBP1	0	0.00029287	0
SCO2	0	0.00029287	0
MRPL23	0	0.00029096	0
RPUSD3	0	0.00028536	0
FARSA	0	0.0002848	0
UBL4A	0	0.00028354	0
CBR4	0	0.00028175	0
SPTLC1	0	0.00028234	0
RFC3	0	0.00028135	0
HSPBP1	0	0.000279	0
MSI1	0	0.00027669	0
NIPSNAP1	0	0.00027431	0
MZT1	0	0.00027144	0
MYBBP1A	0	0.00026817	0
IRAK1	0	0.00026572	0
PSMB2	0	0.00026498	0
BBS1	0	0.00026274	0
TARS	0	0.00026168	0
MRPL28	0	0.00026084	0
TRMT2A	0	0.00025743	0
RPS24	0	0.00025682	0
PYCR1	0	0.00025584	0
BAG6	0	0.00025561	0
GRWD1	0	0.00024953	0
DIMT1	0	0.00024889	0
NUBP2	0	0.0002464	0
CYC1	0	0.0002397	0
C11orf83	0	0.00023933	0
GCN1L1	0	0.0002375	0
CDC7	0	0.00023266	0
TUBGCP3	0	0.00023313	0
BAX	0	0.00023186	0
NEDD8-MDP1	0	0.00023065	0
EEF1G	0	0.0002292	0
LDHA	0	0.00022849	0
TIMM8B	0	0.00022712	0
CTU1	0	0.00022386	0

STRAP	0	0.00022258	0
SND1	0	0.00022014	0
EMD	0	0.00021908	0
HNRNPA0	0	0.00021893	0
ATP5F1	0	0.00021736	0
SARS	0	0.00021652	0
TECR	0	0.0002168	0
SUMO3	0	0.0002161	0
CBX6	0	0.0002161	0
SDF4	0	0.0002152	0
MRPL4	0	0.00021471	0
TSEN34	0	0.0002154	0
SLC25A1	0	0.00021471	0
PIK3R2	0	0.00021402	0
FOXRED2	0	0.00021152	0
CYCS	0	0.00021198	0
MRPL37	0	0.00021048	0
LMNB1	0	0.00020891	0
ACOT8	0	0.00020932	0
TBP	0	0.00020932	0
PSMD3	0	0.00020841	0
CISD1	0	0.00020609	0
NOL9	0	0.00020609	0
SRPRB	0	0.00020533	0
PSMC3	0	0.00020281	0
PPP2R5C	0	0.00020052	0
FAM203B	0	0.00019975	0
IGBP1	0	0.00019697	0
SLC30A9	0	0.00019593	0
HOXD13	0	0.00019468	0
COX7A2	0	0.00019355	0
IKBKAP	0	0.00019217	0
NDUFA5	0	0.00019188	0
L2HGDH	0	0.00019229	0
TROVE2	0	0.00019078	0
SLC27A4	0	0.00019039	0
SLC25A6	0	0.00018835	0
XPO1	0	0.00018704	0
ATP6V1F	0	0.00018704	0
YWHAH	0	0.00018465	0
CCT5	0	0.00018514	0

MTR	0	0.00018475	0
PUSL1	0	0.00018365	0
LPCAT2	0	0.00018412	0
HLA-C	0	0.00018244	0
ATP2A2	0	0.00018157	0
LRRC59	0	0.00018125	0
HINT1	0	0.00017665	0
CAD	0	0.00017506	0
CDC42	0	0.0001748	0
RFC2	0	0.00017389	0
RFC5	0	0.00017444	0
RPP30	0	0.00017281	0
PELO	0	0.00017344	0
RARS2	0	0.00017329	0
IDH2	0	0.00017235	0
TARS2	0	0.0001705	0
RPS15A	0	0.00017122	0
POLR3D	0	0.00016777	0
SFXN4	0	0.00016512	0
RPN1	0	0.00016501	0
SLC25A40	0	0.00016463	0
	0	0.00016366	0
SEC61A1	0	0.00016366	0
PHB	0	0.00016366	0
POLD1	0	0.00016085	0
ALDH3A2	0	0.00016063	0
PDK3	0	0.0001609	0
CBWD5	0	0.00016036	0
FHL2	0	0.00015956	0
POP7	0	0.00015899	0
MAP2K7	0	0.00015937	0
HPS6	0	0.00015796	0
TOMM22	0	0.00015675	0
PAICS	0	0.00015712	0
SLC16A1	0	0.00015581	0
SLC25A10	0	0.00015511	0
YTHDF2	0	0.00015377	0
PTPLAD1	0	0.00015372	0
C2orf47	0	0.00015298	0
VPS4A	0	0.0001528	0
DNAJC7	0	0.00015245	0

DHPS	0	0.0001508	0
SHMT2	0	0.00014905	0
NUP93	0	0.00014947	0
PSMD10	0	0.00014773	0
ATP1B1	0	0.00014692	0
AAR2	0	0.00014491	0
ALG1	0	0.00014391	0
ELP3	0	0.00014242	0
MRPL50	0	0.00014087	0
LPL	0	0.00014058	0
ALDH18A1	0	0.00013999	0
UMPS	0	0.00013911	0
PSMC6	0	0.00013808	0
KRT8	0	0.00013725	0
TTC26	0	0.00013711	0
KLHDC2	0	0.00013706	0
ACY1	0	0.00013639	0
MSI2	0	0.00013572	0
GET4	0	0.00013614	0
IGF2BP3	0	0.00013455	0
TMEM33	0	0.00013517	0
GTF2I	0	0.00013382	0
TTF2	0	0.00013408	0
SPATA5L1	0	0.00013302	0
RBM14	0	0.00013308	0
PPP1CC	0	0.0001321	0
COX4I1	0	0.0001317	0
UXT	0	0.0001317	0
CDC123	0	0.00013249	0
EFTUD2	0	0.00013065	0
DHX9	0	0.00013145	0
CTU2	0	0.00012966	0
FHL3	0	0.00012941	0
CETN2	0	0.00012941	0
GTF3C5	0	0.00012866	0
NDUFA8	0	0.00012941	0
PYCR2	0	0.00012752	0
RDH13	0	0.00012841	0
FOPNL	0	0.00012792	0
NAP1L4	0	0.00012662	0
FTL	0	0.00012719	0

OXSR1	0	0.00012671	0
MCL1	0	0.00012719	0
GTF3C3	0	0.00012561	0
CDC73	0	0.00012575	0
IBA57	0	0.00012505	0
SQSTM1	0	0.00012505	0
LPCAT1	0	0.00012505	0
APRT	0	0.00012366	0
DNAJB11	0	0.00012435	0
HTRA2	0	0.00012331	0
NDUFS2	0	0.00012176	0
E2F7	0	0.00012216	0
PIK3R3	0	0.00012071	0
SSR4	0	0.00012097	0
ACLY	0	0.0001213	0
ATP6V1H	0	0.00011967	0
C12orf45	0	0.00012031	0
CDC37	0	0.00011777	0
TRABD	0	0.00011839	0
NDUFA9	0	0.00011808	0
PGAM5	0	0.00011553	0
C9orf40	0	0.00011473	0
STK38	0	0.00011488	0
SPC24	0	0.00011299	0
MAT2A	0	0.0001127	0
CANX	0	0.00011279	0
VAT1	0	0.00011327	0
CDK1	0	0.00011241	0
SAR1A	0	0.00011241	0
SRSF1	0	0.00011074	0
GDI2	0	0.00011129	0
TCERG1	0	0.00011149	0
ARCN1	0	0.00010889	0
HCFC1	0	0.00010938	0
PTOV1	0	0.00010701	0
GTPBP3	0	0.00010619	0
ACTR3C	0	0.00010599	0
WDR6	0	0.00010636	0
NUBPL	0	0.00010466	0
ATP5O	0	0.0001045	0
CDIPT	0	0.0001045	0

NDUFS7	0	0.0001045	0
DHRS7B	0	0.00010273	0
RPL10A	0	0.00010257	0
ALDOA	0	0.00010191	0
RCN1	0	0.00010087	0
SLC25A12	0	0.00010101	0
HSPA4	0	0.00010081	0
C3orf58	0	0.00010072	0
SDF2L1	0	0.00010072	0

BIBLIOGRAPHY

1. **Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D.** 2011. Global cancer statistics. CA: a cancer journal for clinicians **61:**69-90.
2. **McGuire A, Brown JA, Malone C, McLaughlin R, Kerin MJ.** 2015. Effects of age on the detection and management of breast cancer. Cancers **7:**908-929.
3. **Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D.** 2000. Molecular portraits of human breast tumours. Nature **406:**747-752.
4. **Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale AL.** 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America **98:**10869-10874.
5. **Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM.** 2007. Identification of conserved gene expression features between murine

mammary carcinoma models and human breast tumors. Genome biology **8:**R76.

6. **Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee JS, Fridlyand J, Sahin A, Agarwal R, Joy C, Liu W, Stivers D, Baggerly K, Carey M, Lluch A, Monteagudo C, He X, Weigman V, Fan C, Palazzo J, Hortobagyi GN, Nolden LK, Wang NJ, Valero V, Gray JW, Perou CM, Mills GB.** 2009. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer research **69:**4116-4124.

7. **Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM.** 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast cancer research : BCR **12:**R68.

8. **Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A, Gray JW.** 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer cell **10:**515-527.

9. **Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M, Perou CM.** 2013. Characterization of

cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. Breast cancer research and treatment **142:**237-255.

10. **Goldberg AD, Allis CD, Bernstein E.** 2007. Epigenetics: a landscape takes shape. Cell **128:**635-638.
11. **Ordway JM, Budiman MA, Korshunova Y, Maloney RK, Bedell JA, Citek RW, Bacher B, Peterson S, Rohlfing T, Hall J, Brown R, Lakey N, Doerge RW, Martienssen RA, Leon J, McPherson JD, Jeddeloh JA.** 2007. Identification of novel high-frequency DNA methylation changes in breast cancer. PloS one **2:**e1314.
12. **Strahl BD, Allis CD.** 2000. The language of covalent histone modifications. Nature **403:**41-45.
13. **Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK.** 2005. Global histone modification patterns predict risk of prostate cancer recurrence. Nature **435:**1262-1266.
14. **Seligson DB, Horvath S, McBrian MA, Mah V, Yu H, Tze S, Wang Q, Chia D, Goodglick L, Kurdistani SK.** 2009. Global levels of histone modifications predict prognosis in different cancers. The American journal of pathology **174:**1619-1628.
15. **Kim KH, Roberts CW.** 2016. Targeting EZH2 in cancer. Nature medicine **22:**128-134.
16. **Wang Y, Zhu Y, Wang Q, Hu H, Li Z, Wang D, Zhang W, Qi B, Ye J, Wu H, Jiang H, Liu L, Yang J, Cheng J.** 2016. The histone demethylase

LSD1 is a novel oncogene and therapeutic target in oral cancer. Cancer letters **374:**12-21.

17. **Wu J, Hu L, Du Y, Kong F, Pan Y.** 2015. Prognostic role of LSD1 in various cancers: evidence from a meta-analysis. OncoTargets and therapy **8:**2565-2570.
18. **Farria A, Li W, Dent SY.** 2015. KATs in cancer: functions and therapies. Oncogene **34:**4901-4913.
19. **Yoon S, Eom GH.** 2016. HDAC and HDAC Inhibitor: From Cancer to Cardiovascular Diseases. Chonnam medical journal **52:**1-11.
20. **Wang L, Dent SY.** 2014. Functions of SAGA in development and disease. Epigenomics **6:**329-339.
21. **Liu X, Tesfai J, Evrard YA, Dent SY, Martinez E.** 2003. c-Myc transformation domain recruits the human STAGA complex and requires TRRAP and GCN5 acetylase activity for transcription activation. The Journal of biological chemistry **278:**20405-20412.
22. **Zhang N, Ichikawa W, Faiola F, Lo SY, Liu X, Martinez E.** 2014. MYC interacts with the human STAGA coactivator complex via multivalent contacts with the GCN5 and TRRAP subunits. Biochimica et biophysica acta **1839:**395-405.
23. **Kuo MH, Brownell JE, Sobel RE, Ranalli TA, Cook RG, Edmondson DG, Roth SY, Allis CD.** 1996. Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. Nature **383:**269-272.

24. **Xu W, Edmondson DG, Roth SY.** 1998. Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates. Molecular and cellular biology **18:**5659-5669.

25. **Grant PA, Duggan L, Cote J, Roberts SM, Brownell JE, Candau R, Ohba R, Owen-Hughes T, Allis CD, Winston F, Berger SL, Workman JL.** 1997. Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes & development **11:**1640-1650.

26. **Xu W, Edmondson DG, Evrard YA, Wakamiya M, Behringer RR, Roth SY.** 2000. Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. Nature genetics **26:**229-232.

27. **Yamauchi T, Yamauchi J, Kuwata T, Tamura T, Yamashita T, Bae N, Westphal H, Ozato K, Nakatani Y.** 2000. Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. Proceedings of the National Academy of Sciences of the United States of America **97:**11303-11306.

28. **Bu P, Evrard YA, Lozano G, Dent SY.** 2007. Loss of Gcn5 acetyltransferase activity leads to neural tube closure defects and exencephaly in mouse embryos. Molecular and cellular biology **27:**3405-3416.

29. **Chen YC, Gatchel JR, Lewis RW, Mao CA, Grant PA, Zoghbi HY, Dent SY.** 2012. Gcn5 loss-of-function accelerates cerebellar and retinal

degeneration in a SCA7 mouse model. Human molecular genetics **21:**394-405.

30. **Zhang XY, Varthi M, Sykes SM, Phillips C, Warzecha C, Zhu W, Wyce A, Thorne AW, Berger SL, McMahon SB.** 2008. The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. Molecular cell **29:**102-111.

31. **Zhao Y, Lang G, Ito S, Bonnet J, Metzger E, Sawatsubashi S, Suzuki E, Le Guezennec X, Stunnenberg HG, Krasnov A, Georgieva SG, Schule R, Takeyama K, Kato S, Tora L, Devys D.** 2008. A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Molecular cell **29:**92-101.

32. **Lang G, Bonnet J, Umlauf D, Karmodiya K, Koffler J, Stierle M, Devys D, Tora L.** 2011. The tightly controlled deubiquitination activity of the human SAGA complex differentially modifies distinct gene regulatory elements. Molecular and cellular biology **31:**3734-3744.

33. **Kohler A, Schneider M, Cabal GG, Nehrbass U, Hurt E.** 2008. Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. Nature cell biology **10:**707-715.

34. **Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y.** 2004. Role of histone H2A ubiquitination in Polycomb silencing. Nature **431:**873-878.

35. **Kim J, Hake SB, Roeder RG.** 2005. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. Molecular cell **20:**759-770.

36. **Zhu B, Zheng Y, Pham AD, Mandal SS, Erdjument-Bromage H, Tempst P, Reinberg D.** 2005. Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Molecular cell **20:**601-611.

37. **Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M.** 2008. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nature cell biology **10:**483-488.

38. **Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, Kao CF, Pillus L, Shilatifard A, Osley MA, Berger SL.** 2003. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes & development **17:**2648-2663.

39. **Lin Z, Yang H, Kong Q, Li J, Lee SM, Gao B, Dong H, Wei J, Song J, Zhang DD, Fang D.** 2012. USP22 antagonizes p53 transcriptional activation by deubiquitinating Sirt1 to suppress cell apoptosis and is required for mouse embryonic development. Molecular cell **46:**484-494.

40. **Glinsky GV, Berezovska O, Glinskii AB.** 2005. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. The Journal of clinical investigation **115:**1503-1521.

41. **Liu YL, Yang YM, Xu H, Dong XS.** 2011. Aberrant expression of USP22 is associated with liver metastasis and poor prognosis of colorectal cancer. Journal of surgical oncology **103:**283-289.

42. **Zhang Y, Yao L, Zhang X, Ji H, Wang L, Sun S, Pang D.** 2011. Elevated expression of USP22 in correlation with poor prognosis in patients with invasive breast cancer. Journal of cancer research and clinical oncology **137:**1245-1253.

43. **Ning Z, Wang A, Liang J, Xie Y, Liu J, Feng L, Yan Q, Wang Z.** 2014. USP22 promotes the G1/S phase transition by upregulating FoxM1 expression via beta-catenin nuclear localization and is associated with poor prognosis in stage II pancreatic ductal adenocarcinoma. International journal of oncology **45:**1594-1608.

44. **Liang J, Zhang X, Xie S, Zhou X, Shi Q, Hu J, Wang W, Qi W, Yu R.** 2014. Ubiquitin-specific protease 22: a novel molecular biomarker in glioma prognosis and therapeutics. Medical oncology **31:**899.

45. **Yang M, Liu YD, Wang YY, Liu TB, Ge TT, Lou G.** 2014. Ubiquitin-specific protease 22: a novel molecular biomarker in cervical cancer prognosis and therapeutics. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine **35:**929-934.

46. **Piao S, Liu Y, Hu J, Guo F, Ma J, Sun Y, Zhang B.** 2012. USP22 is useful as a novel molecular marker for predicting disease progression and patient prognosis of oral squamous cell carcinoma. PloS one **7:**e42540.

47. **Tang B, Liang X, Tang F, Zhang J, Zeng S, Jin S, Zhou L, Kudo Y, Qi G.** 2015. Expression of USP22 and Survivin is an indicator of malignant behavior in hepatocellular carcinoma. International journal of oncology **47:**2208-2216.

48. **He Y, Jin YJ, Zhang YH, Meng HX, Zhao BS, Jiang Y, Zhu JW, Liang GY, Kong D, Jin XM.** 2015. Ubiquitin-specific peptidase 22 overexpression may promote cancer progression and poor prognosis in human gastric carcinoma. Translational research : the journal of laboratory and clinical medicine **165:**407-416.

49. **Atanassov BS, Dent SY.** 2011. USP22 regulates cell proliferation by deubiquitinating the transcriptional regulator FBP1. EMBO reports **12:**924-930.

50. **Schrecengost RS, Dean JL, Goodwin JF, Schiewer MJ, Urban MW, Stanek TJ, Sussman RT, Hicks JL, Birbe RC, Draganova-Tacheva RA, Visakorpi T, DeMarzo AM, McMahon SB, Knudsen KE.** 2014. USP22 regulates oncogenic signaling pathways to drive lethal cancer progression. Cancer research **74:**272-286.

51. **Atanassov BS, Evrard YA, Multani AS, Zhang Z, Tora L, Devys D, Chang S, Dent SY.** 2009. Gcn5 and SAGA regulate shelterin protein turnover and telomere maintenance. Molecular cell **35:**352-364.

52. **Liu YL, Jiang SX, Yang YM, Xu H, Liu JL, Wang XS.** 2012. USP22 acts as an oncogene by the activation of BMI-1-mediated INK4a/ARF pathway and Akt pathway. Cell biochemistry and biophysics **62:**229-235.

53. **Koehler C, Bonnet J, Stierle M, Romier C, Devys D, Kieffer B.** 2014. DNA binding by Sgf11 protein affects histone H2B deubiquitination by Spt-Ada-Gcn5-acetyltransferase (SAGA). The Journal of biological chemistry **289:**8989-8999.

54. **Morgan MT, Haj-Yahya M, Ringel AE, Bandi P, Brik A, Wolberger C.** 2016. Structural basis for histone H2B deubiquitination by the SAGA DUB module. Science **351:**725-728.

55. **Atanassov BS, Mohan RD, Lan X, Kuang X, Lu Y, Lin K, McIvor E, Li W, Zhang Y, Florens L, Byrum SD, Mackintosh SG, Calhoun-Davis T, Koutelou E, Wang L, Tang DG, Tackett AJ, Washburn MP, Workman JL, Dent SY.** 2016. ATXN7L3 and ENY2 Coordinate Activity of Multiple H2B Deubiquitinases Important for Cellular Proliferation and Tumor Growth. Molecular cell **62:**558-571.

56. **Lan X, Koutelou E, Schibler AC, Chen YC, Grant PA, Dent SY.** 2015. Poly(Q) Expansions in ATXN7 Affect Solubility but Not Activity of the SAGA Deubiquitinating Module. Molecular and cellular biology **35:**1777-1787.

57. **Mendez J, Stillman B.** 2000. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Molecular and cellular biology **20:**8602-8612.

58. **Fuchs G, Shema E, Vesterman R, Kotler E, Wolchinsky Z, Wilder S, Golomb L, Pribluda A, Zhang F, Haj-Yahya M, Feldmesser E, Brik A,**

Yu X, Hanna J, Aberdam D, Domany E, Oren M. 2012. RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. Molecular cell **46:**662-673.

59. **Prat A, Perou CM.** 2011. Deconstructing the molecular portraits of breast cancer. Molecular oncology **5:**5-23.

60. **Ellisdon AM, Jani D, Kohler A, Hurt E, Stewart M.** 2010. Structural basis for the interaction between yeast Spt-Ada-Gcn5 acetyltransferase (SAGA) complex components Sgf11 and Sus1. The Journal of biological chemistry **285:**3850-3856.

61. **Kohler A, Pascual-Garcia P, Llopis A, Zapater M, Posas F, Hurt E, Rodriguez-Navarro S.** 2006. The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. Molecular biology of the cell **17:**4228-4236.

62. **Atanassov BS, Mohan RD, Lan X, Kuang X, Lu Y, Lin K, Mclvor E, Li W, Zhang Y, Florens L, Byrum SD, Mackintosh SG, Davis T, Koutelou E, Wang L, Tang D, Tackett AJ, Washburn MP, Workman JL, Dent SY.** 2016. ATXN7L3 and ENY2 coordinate activity of multiple H2B deubiquitinases important for cellular proliferation and tumor growth. Molecular Cell **In press**.

63. **Umlauf D, Bonnet J, Waharte F, Fournier M, Stierle M, Fischer B, Brino L, Devys D, Tora L.** 2013. The human TREX-2 complex is stably

associated with the nuclear pore basket. Journal of cell science **126:**2656-2667.

64. **Glinsky GV.** 2006. Genomic models of metastatic cancer: functional analysis of death-from-cancer signature genes reveals aneuploid, anoikis-resistant, metastasis-enabling phenotype with altered cell cycle control and activated Polycomb Group (PcG) protein chromatin silencing pathway. Cell cycle **5:**1208-1216.

65. **Brown AM, Jeltsch JM, Roberts M, Chambon P.** 1984. Activation of pS2 gene transcription is a primary response to estrogen in the human breast cancer cell line MCF-7. Proceedings of the National Academy of Sciences of the United States of America **81:**6344-6348.

66. **Lin CY, Strom A, Vega VB, Kong SL, Yeo AL, Thomsen JS, Chan WC, Doray B, Bangarusamy DK, Ramasamy A, Vergara LA, Tang S, Chong A, Bajic VB, Miller LD, Gustafsson JA, Liu ET.** 2004. Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. Genome biology **5:**R66.

67. **Wu W, Bhagat TD, Yang X, Song JH, Cheng Y, Agarwal R, Abraham JM, Ibrahim S, Bartenstein M, Hussain Z, Suzuki M, Yu Y, Chen W, Eng C, Greally J, Verma A, Meltzer SJ.** 2013. Hypomethylation of noncoding DNA regions and overexpression of the long noncoding RNA, AFAP1-AS1, in Barrett's esophagus and esophageal adenocarcinoma. Gastroenterology **144:**956-966 e954.

68. **Zhou XL, Wang WW, Zhu WG, Yu CH, Tao GZ, Wu QQ, Song YQ, Pan P, Tong YS.** 2016. High expression of long non-coding RNA AFAP1-AS1 predicts chemoradioresistance and poor prognosis in patients with esophageal squamous cell carcinoma treated with definitive chemoradiotherapy. Molecular carcinogenesis.

69. **Ye Y, Chen J, Zhou Y, Fu Z, Zhou Q, Wang Y, Gao W, Zheng S, Zhao X, Chen T, Chen R.** 2015. High expression of AFAP1-AS1 is associated with poor survival and short-term recurrence in pancreatic ductal adenocarcinoma. Journal of translational medicine **13:**137.

70. **Zeng Z, Bo H, Gong Z, Lian Y, Li X, Zhang W, Deng H, Zhou M, Peng S, Li G, Xiong W.** 2016. AFAP1-AS1, a long noncoding RNA upregulated in lung cancer and promotes invasion and metastasis. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine **37:**729-737.

71. **Deng J, Liang Y, Liu C, He S, Wang S.** 2015. The up-regulation of long non-coding RNA AFAP1-AS1 is associated with the poor prognosis of NSCLC patients. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie **75:**8-11.

72. **Bo H, Gong Z, Zhang W, Li X, Zeng Y, Liao Q, Chen P, Shi L, Lian Y, Jing Y, Tang K, Li Z, Zhou Y, Zhou M, Xiang B, Yang J, Xiong W, Li G, Zeng Z.** 2015. Upregulated long non-coding RNA AFAP1-AS1 expression is associated with progression and poor prognosis of nasopharyngeal carcinoma. Oncotarget **6:**20404-20418.

73. **Zhang JY, Weng MZ, Song FB, Xu YG, Liu Q, Wu JY, Qin J, Jin T, Xu JM.** 2016. Long noncoding RNA AFAP1-AS1 indicates a poor prognosis of hepatocellular carcinoma and promotes cell proliferation and invasion via upregulation of the RhoA/Rac2 signaling. International journal of oncology **48:**1590-1598.

74. **Lu X, Zhou C, Li R, Liang Z, Zhai W, Zhao L, Zhang S.** 2016. Critical role for the long non-coding RNA AFAP1-AS1 in the proliferation and metastasis of hepatocellular carcinoma. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine.

75. **Wang F, Ni H, Sun F, Li M, Chen L.** 2016. Overexpression of lncRNA AFAP1-AS1 correlates with poor prognosis and promotes tumorigenesis in colorectal cancer. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie **81:**152-159.

76. **Rajakulendran S, Roberts J, Koltzenburg M, Hanna MG, Stewart H.** 2013. Deletion of chromosome 12q21 affecting KCNC2 and ATXN7L3B in a family with neurodevelopmental delay and ataxia. Journal of neurology, neurosurgery, and psychiatry **84:**1255-1257.

77. **Tan JY, Vance KW, Varela MA, Sirey T, Watson LM, Curtis HJ, Marinello M, Alves S, Steinkraus BR, Cooper S, Nesterova T, Brockdorff N, Fulga TA, Brice A, Sittler A, Oliver PL, Wood MJ, Ponting CP, Marques AC.** 2014. Cross-talking noncoding RNAs contribute to cell-specific neurodegeneration in SCA7. Nature structural & molecular biology **21:**955-961.

Vita

Wenqian Li was born in Hengshui, Hebei Province in China. She was born on 16th August 1988 to Mr. Zhigang Li and Mrs. Yuge Yin. After completing her high school at Hengshui High School, Wenqian continued her college education at Peking University Health Science Center, majored in Medical Sciences, from 2006-2011. Wenqian completed her Bachelor's dissertation project titled "investigate the effect of a synthetic agonist of Liver X Receptor on phenotypic changes of rat vascular smooth muscle cells" under the mentorship of Dr. Hong Zhang. At 2011, Wenqian joined the Ph.D. program at Graduate School of Biomedical Science in The University of Texas Health Science Center MD Anderson Cancer Center. Wenqian joined the lab of Dr. Sharon Dent at Epigenetics and Molecular Carcinogenesis program to pursue her Ph.D. dissertation project. Her project investigates the epigenetic profiles of breast cancer subtypes and studies the functions of subtype-specific expressing genes.