Long noncoding RNA AK001796 as a mediator of epithelial-to-mesenchymal transition in metastatic breast cancer

Sumayya Adib Alchalabi

University at Albany, State University of New York, sumayyaalchalabi@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Biology Commons, and the Medicine and Health Sciences Commons

Recommended Citation


https://scholarsarchive.library.albany.edu/legacy-etd/2628

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.

Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
Long Noncoding RNA AK001796 as a mediator of epithelial-to-mesenchymal transition in metastatic breast cancer

by

Sumayya A. Al-chalabi

A Dissertation

Submitted to the University at Albany, State University of New York in Partial Fulfillment of the Requirements for the Degree of Doctor of philosophy

School of Public Health
Department of Biomedical Sciences
2021
Abstract:

Long noncoding RNAs (lncRNAs) are emerging as key molecules regulating several cellular pathways and contributing to the development and progression of different diseases including cancer. The epithelial-to-mesenchymal transition (EMT), a molecular reprogramming process that plays a critical role in cancer metastasis, involves a complex interaction between protein coding and non-coding transcripts including lncRNAs. Understanding the biological significance of those lncRNAs of relatively unknown function, may help to uncover potential therapeutic targets for the aggressive subtypes of breast and other cancers.

In this study, we examined the expression profiles of more than 17,000 lncRNAs in a large set of patients’ derived breast cancers and have identified a lncRNA, AK001796, that is highly enriched in aggressive breast cancers and in several induced models of EMT. Interestingly, knockdown of this lncRNA in triple-negative breast cancer (TNBC) cell lines showed a concomitant increase in the epithelial phenotype-associated gene E-cadherin, while overexpression of AK001796 is associated with suppression of E-cadherin and upregulation of the mesenchymal markers TWIST and vimentin. Loss-of-function studies using modified antisense oligonucleotides (ASOs) suggests that this lncRNA may be a crucial player for cell viability in EMT-positive breast cancer. Further, molecular characterization of this lncRNA revealed it to be regulated by the EMT-associated transcription factor BHLHE40. To identify potential target genes for AK001796, we investigated its association with the nearby gene, BCL2L11, an apoptosis-associated protein. Interestingly, silencing of AK001796 resulted in upregulation of BCL2L11. In order to elucidate the mechanism of action of AK001796, we pulled-down an interacting protein partner, poly ADP-ribose polymerase enzyme (PARP1) and modulated its enzymatic activity. These findings suggest AK001796 is a potential therapeutic target for metastatic breast cancer.
Acknowledgements

First and foremost, praises and thanks to God, the Almighty, for his countless blessings throughout my years of study and research work.

I would like to express my sincere gratitude to my adviser, Dr. Jason Herschkowitz, for his continuous support and encouragement throughout my graduate study. His invaluable guidance and insightful feedbacks helped me to sharpen my thinking and formulate my research questions. He was supportive not only at the academic level, but also at the personal level by listening to my problems and trying to help me in every possible way.

I would also like to extend my deepest appreciation to my committee members; Dr. Douglas Conklin, Dr. JoEllen Welsh, Dr. Valerie Bolivar, and Dr. Thomas Begley for their constructive criticism and invaluable feedbacks.

Special thanks to Dr. Maneesh Kumar and Dr. Rebecca Sinnott DeVaux for their help with the experimental design and data analysis. Their extensive experience and wide knowledge helped me to shape my initial understanding of non-coding RNAs and breast cancer biology.

I also had great pleasure working with my Ph. D colleague, Ali Ropi, who always gives me a positive energy to move forward in my research and offer his help whenever I need.

I also enjoyed working with Tony Pazos, a former master’s student in Dr. Herschwitz lab, and other undergraduate and high school students who created an excellent working environment in the lab with their motivation and enthusiasm about science.

I am also grateful to Dr. Sridar Chittur and Marcy Kuentzel for their help with the sequencing experiments and their thoughtful advice. I am so thankful for all the wonderful people over the Cancer Research Center and the biomedical sciences department of SUNY Albany for their support throughout this wonderful journey.
Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. I am forever indebted to my parents for their continuous encouragement and unconditional love throughout my life. I dedicate this work to them.

I could not be able to complete this dissertation without the support of my husband, Ashraf. I greatly value his contribution to my success and appreciate his believe in me.

I owe thanks to my oldest son, Abdulrahman for helping me with taking care of his younger siblings and being the best idol for them. You really made me a proud mother.

I am also grateful to my daughter, Shahad for taking care of her little sister, Reema and being kind and patient with her brothers and classmates. You are the best big sister ever. I would also thank my son, Sameer for being the little man of the house. You always surprise me with your excellent grades and the exceptional performance at school without even asking us for help!

I cannot end without acknowledging my two babies, Reema and Adam, who were born during my Ph.D. They filled the house with joy and happiness. I feel so lucky to have such a wonderful and supportive family. This accomplishment would not have been possible without them.
# Table of Contents

Abstract ................................................................................................................................. ii

Acknowledgements ........................................................................................................... iii

Table of Contents .............................................................................................................. v

List of figures ...................................................................................................................... vii

List of tables ..................................................................................................................... viii

List of abbreviations ......................................................................................................... xi

Chapter 1: Breast Cancer, EMT and IncRNAs ................................................................. 1

1.1 Introduction ................................................................................................................... 2

1.2 Anatomy and histology of the mammary glands ....................................................... 3

1.3 Breast cancer molecular subtypes ............................................................................. 4

1.4 Epithelial to Mesenchymal Transition (EMT) ........................................................... 6

1.5 Long non-coding RNAs (LncRNAs) ........................................................................ 9

1.5.1 Transcriptional regulation by LncRNAs ................................................................. 10

1.5.2 LncRNAs regulation of splicing .......................................................................... 16

1.5.3 LncRNAs regulation of mRNAs stability ............................................................... 21

1.5.4 LncRNAs regulation of miRNAs ......................................................................... 24

1.6 LncRNA AK001796 .................................................................................................. 27

1.7 Aim and objectives .................................................................................................... 28

Chapter 2. Material and methods .................................................................................... 29

2.1 Cell lines and culture media ..................................................................................... 30

2.2 Lentiviral transfection and transduction ................................................................. 31

2.3 Gene expression analysis ......................................................................................... 32

2.4 Cellular transfection ............................................................................................... 32

2.5 RNA isolation and qRT-PCR .................................................................................. 33

2.6 Immunofluorescence staining .................................................................................. 33

2.7 Western blotting ...................................................................................................... 34

2.8 Mammosphere formation assay .............................................................................. 34

2.9 Apoptosis Assays ................................................................................................... 35

2.10 Cellular viability assays ......................................................................................... 35

2.11 Chromatin Immunoprecipitation (CHIP) .............................................................. 36
List of figures

Chapter 1

Figure 1.1 Mechanisms of regulation of transcription by lncRNAs 15
Figure 1.2 Regulation of splicing by lncRNAs 20
Figure 1.3 LncRNAs regulation of mRNAs stability 23

Chapter 3

Figure 2.1 Differential expression of lncRNAs between the claudin low and other molecular subtypes of breast cancer 43
Figure 2.2 The expression of AK001796 in patient samples relative to adjacent normal from TCGA database 44
Figure 2.3 Differential expression of lncRNAs between normal mammary epithelial cells (HMLE) cells and EMT induced models 45
Figure 2.4 Increased expression of AK001796 in other breast cancer cell lines upon induction of EMT 46
Figure 2.5 The level of expression of lncRNA AK001796 in breast cancer cell lines grouped according to their molecular subtypes 48
Figure 2.6 Gene Set Enrichment Analysis showed enrichment for the hallmark EMT with expression of AK001796 in breast cancer. 49
Figure 2.7 Gene set variation analysis (GSVA) for RNA-seq in SUM159PT cells following AK001796 depletion 51
Figure 3.1 E-cadherin expression following AK001796 silencing 53
Figure 3.2 AK001796 overexpression leads to induction of EMT 54
Figure 3.3  LncRNA AK001796 contributes in maintaining stemness in EMT high breast cancer cells  56

Figure 3.4  Induction of EMT is associated with reduced expression of BHLHE40  58

Figure 3.5  LncRNA AK001796 is negatively regulated by the transcription factor BHLHE40  59

Chapter 4

Figure 4.1  Kaplan-Mier plotter for the overall survival of breast cancer patients relative to AK001796 expression in the different subtypes of breast cancer  64

Figure 4.2  LncRNA AK001796 is essential for the survival of triple-negative breast cancer cell lines  65

Figure 4.3  AK001796 silencing is associated with increased apoptosis in EMT-high breast cancer cells.  66

Figure 4.4  Caspase 3/7 apoptotic assay in MDA-MB-231 and SUM159PT cells  68

Figure 4.5  AK001796 silencing has no significant effect on the apoptotic profile of the normal and luminal breast cancer cell lines  69

Figure 4.6  Comprehensive analysis of the expression pattern of AK001796 across all cancer types.  72

Figure 4.7  The level of expression of LncRNA AK001796 in lung cancer cell lines using BigQuery output from Cancer Cell Line Encyclopedia.  73

Figure 4.8  AK001796 silencing is associated with increase in apoptosis in lung cancer  74

Figure 4.9  AK001796 expression is correlated with apoptotic genes signatures  77
Figure 4.10 LncRNA AK001796 negatively regulates pro-apoptotic gene BCL2L11 in TNBC cells

Figure 4.11 Silencing Bim partially rescue the apoptotic profile after AK001796 depletion

Figure 4.12 Enrichment of AK001796 on the pro-apoptotic gene BCL2L11

Chapter 5

Figure 5.1 LncRNA AK001796 interaction with Poly (ADP-ribose) Polymerase 1 (PARP1)

Figure 5.2 The effect of AK001796 expression on the enzymatic activity of PARP1 (PARylation)

Figure 5.3 PARP1 play a role in regulating EMT process

Figure 5.4 Colocalization of PARP1 and AK001796

Figure 5.5 AK001796 expression decrease the sensitivity to PARPi

Chapter 6

Figure 6 Model for the mechanism of action of AK001796

Appendix

Figure S1 The nuclear localization of AK001796 and its effect on E-cadherin expression

Figure S2 Query output of cBioPortal, for the co-expression of AK001796 and A. TWIST1, B. SNAI1, C. VIM, D. CDH1.

Figure S3 Chromosome wide binding of IncRNA AK001796

Figure S4 RNA-seq panther GO analysis for the common pathways associated with AK001796 expression.
List of tables

Table 1  Breast cancer cell lines used in this study classified based on molecular subtype 31

Table 2  Antisense LNA GapmeRs oligonucleotides sequences 39

Table 3  Select Pre-designed siRNAs used in this study 39

Table 4  TaqMan primers 40

Table 5  CYBR Green qPCR primers 40

Table 6  GSEA from TCGA shows the main cancer whole marks associated with AK001796 expression 50
### List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LncRNA</td>
<td>Long non-coding RNA</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial to Mesenchymal Transition</td>
</tr>
<tr>
<td>TNBC</td>
<td>Triple negative breast Cancer</td>
</tr>
<tr>
<td>ASOs</td>
<td>Antisense oligonucleotides</td>
</tr>
<tr>
<td>BCL2L11</td>
<td>Bcl-2-like protein 11</td>
</tr>
<tr>
<td>BHLHE 40</td>
<td>Basic helix-loop-helix protein 40</td>
</tr>
<tr>
<td>PARP1</td>
<td>Poly (ADP-ribose) Polymerase 1</td>
</tr>
<tr>
<td>ER</td>
<td>Estrogen receptor</td>
</tr>
<tr>
<td>PR</td>
<td>Progesterone receptor</td>
</tr>
<tr>
<td>HER2</td>
<td>Human Epidermal growth factor Receptor 2</td>
</tr>
<tr>
<td>miRNAs</td>
<td>Micro RNAs</td>
</tr>
<tr>
<td>lincRNA</td>
<td>Long intergenic non-coding RNA</td>
</tr>
<tr>
<td>DCIS</td>
<td>Ductal carcinoma in situ</td>
</tr>
<tr>
<td>LCIS</td>
<td>Lobular carcinoma in situ</td>
</tr>
<tr>
<td>IDC</td>
<td>Invasive ductal carcinoma</td>
</tr>
<tr>
<td>ILC</td>
<td>Invasive lobular carcinoma</td>
</tr>
<tr>
<td>CL</td>
<td>Claudin low</td>
</tr>
<tr>
<td>PAM50</td>
<td>Prediction Analysis of Microarrays</td>
</tr>
<tr>
<td>HR</td>
<td>Hormonal receptor</td>
</tr>
<tr>
<td>E-cadherin</td>
<td>Epithelial-cadherin (Calcium-dependent adhesion molecule)</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Transforming Growth Factor beta</td>
</tr>
<tr>
<td>TNF-α</td>
<td>Tumor necrosis factor-α</td>
</tr>
<tr>
<td>HGF</td>
<td>Hepatocytes growth factor</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>IGF-1</td>
<td>Insulin like growth factor-1</td>
</tr>
<tr>
<td>PDGF</td>
<td>Platelet derived growth factor</td>
</tr>
<tr>
<td>ZEB1</td>
<td>Zinc finger E-box-binding homeobox 1</td>
</tr>
<tr>
<td>ZEB2</td>
<td>Zinc finger E-box-binding homeobox 2</td>
</tr>
<tr>
<td>FOXC2</td>
<td>Forkhead box protein c2</td>
</tr>
<tr>
<td>N-cadherin</td>
<td>Neural cadherin</td>
</tr>
<tr>
<td>CSCs</td>
<td>Cancer stem cells</td>
</tr>
<tr>
<td>ncRNAs</td>
<td>Non-coding RNAs</td>
</tr>
<tr>
<td>HOTAIR</td>
<td>HOX transcript antisense intergenic lncRNA</td>
</tr>
<tr>
<td>PRC1 &amp; PRC2</td>
<td>Poly comb Repressive Complex 1 &amp; 2</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of Zeste Homology 2</td>
</tr>
<tr>
<td>EED</td>
<td>Embryonic ectoderm development</td>
</tr>
<tr>
<td>SUZ12</td>
<td>Suppressor of zeste 12</td>
</tr>
<tr>
<td>RbAp46/48</td>
<td>Retinoblastoma suppressor associated protein 46/48</td>
</tr>
<tr>
<td>H3K27m3</td>
<td>Histone H3 lysine 27 trimethylation</td>
</tr>
<tr>
<td>LSD1</td>
<td>Lysine-specific histone demethylase 1A</td>
</tr>
<tr>
<td>REST</td>
<td>Restrictive element 1-silencing transcription factor</td>
</tr>
<tr>
<td>MALAT 1</td>
<td>Metastasis Associated Lung Adenocarcinoma Transcript 1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NEAT1/2</td>
<td>Nuclear-enriched abundant transcript 1/2</td>
</tr>
<tr>
<td>FOXN3</td>
<td>Forkhead box N3</td>
</tr>
<tr>
<td>MEG3</td>
<td>Maternally expressed 3</td>
</tr>
<tr>
<td>DNMTs</td>
<td>DNA methyl transferases enzymes</td>
</tr>
<tr>
<td>DBCCR1</td>
<td>Deleted in bladder cancer chromosome region 1</td>
</tr>
<tr>
<td>VIM</td>
<td>Vimentin</td>
</tr>
<tr>
<td>VIM-AS1</td>
<td>Vimentin-antisense 1</td>
</tr>
<tr>
<td>NF-Kb</td>
<td>Nuclear factor kappa light chain enhancer of activated B cells</td>
</tr>
<tr>
<td>HOXA</td>
<td>Homeobox A cluster</td>
</tr>
<tr>
<td>HIT</td>
<td>HOXA transcript activated by TGFβ</td>
</tr>
<tr>
<td>NSCLC</td>
<td>Non-small cells lung cancer</td>
</tr>
<tr>
<td>ELIT-1</td>
<td>EMT associated lncRNA induced by TGFβ-1</td>
</tr>
<tr>
<td>HIF1-α</td>
<td>Hypoxia-inducible factor 1-alpha</td>
</tr>
<tr>
<td>YB1</td>
<td>Y-box protein 1</td>
</tr>
<tr>
<td>GAS5</td>
<td>Growth arrest specific 5</td>
</tr>
<tr>
<td>GR</td>
<td>Glucocorticoid Receptor</td>
</tr>
<tr>
<td>DHFR</td>
<td>Dihydro-folate reductase</td>
</tr>
<tr>
<td>P-ETFb</td>
<td>Positive-elongation transcription elongation factor b</td>
</tr>
<tr>
<td>DDX21</td>
<td>DEAD-box RNA helicase</td>
</tr>
<tr>
<td>snRNAs</td>
<td>Small nuclear RNAs</td>
</tr>
<tr>
<td>snRNPs</td>
<td>Small nuclear ribonucleoproteins</td>
</tr>
<tr>
<td>AS</td>
<td>Alternative splicing</td>
</tr>
</tbody>
</table>
SFs  Splicing factors
SRp  Serine/ arginine-rich protein
RRM  RNA recognition motif
PPARγ Peroxisome proliferator-activated receptor gamma
SRPK1 Serine/ arginine-protein kinase 1
PP1  Protein phosphatase 1
PP2A  Protein phosphatase 2A
SFPQ  Proline- and glutamine-rich SF
PTBP2 Polypyrimidine-tract-binding protein 2
RBFOX2 RNA binding protein fox-1 homolog 2
KIF1B Kinesin family member 1B
NATs  Natural Antisense Transcripts
FASL  FAS-Ligand
ZEB2-AS1 Zing finger E-box-binding homoebox2 antisense RNA 1
IRES  Internal ribosomal entry site
asFGFR2 Antisense for fibroblast growth factor receptor 2 (FGFR2)
KDM2A Lysine-specific demethylase 2A
AREs  AU-rich elements
RBPAs RNA-binding proteins
Linc-ROR Long intergenic noncoding RNA-regulator of reprogramming
hnRNP1 Heterogeneous Nuclear Ribonucleoprotein A1
AUF1  AU-rich binding factor 1
LAST  LncRNA-assisted stabilization of transcripts
CCND1  Cyclin D1
CNBP  Cellular nucleic acid-binding protein
KLF8  Krueppel like factor 8
IGF2  Insulin-like growth factor 2
PDCD4  Programmed cell death 4
HuR  Hu-antigen R
MACC1-AS1  Metastasis associated in colon cancer protein1-antisense 1
AMPK  AMP-activated protein kinase
LncRNA OCC-1  LncRNA over expressed in colorectal cancer
β-TrCP1  β-Transducin Repeat-Containing Protein
Lnc-ASNR  Apoptosis suppressing-noncoding RNA
Bcl2  B-cell lymphoma 2
SMD  Staufen mediated decay
STAU1  Staufen
SBS  STAU1 binding site
SERPINE1  Serine proteinase inhibitor
ceRNA  Competitive endogenous RNA
HULC  Highly Upregulated in Liver Cancer
LncRNA-ATB  LncRNA- activated by TGFβ
CRC  Colorectal cancer
PNUTS  Phosphatase 1 nuclear targeting subunit
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARF6</td>
<td>ADP-ribosylation factor 6</td>
</tr>
<tr>
<td>TUG1</td>
<td>Taurine upregulated 1</td>
</tr>
<tr>
<td>CAR10</td>
<td>Chromatin-associated RNA 10</td>
</tr>
<tr>
<td>LUAD</td>
<td>Lung adenocarcinoma</td>
</tr>
<tr>
<td>TTN-AS1</td>
<td>Titin-antisense RNA1</td>
</tr>
<tr>
<td>ESCC</td>
<td>Esophageal squamous cell carcinoma</td>
</tr>
<tr>
<td>FSCN1</td>
<td>Fascin homolog1</td>
</tr>
<tr>
<td>HCP5</td>
<td>Histocompatibility leukocyte antigen complex P5</td>
</tr>
<tr>
<td>TINCR</td>
<td>Terminal differentiation-induced noncoding RNA</td>
</tr>
<tr>
<td>UCA1</td>
<td>Urothelial carcinoma-associated 1</td>
</tr>
<tr>
<td>MEG 8</td>
<td>Maternally expressed 8</td>
</tr>
<tr>
<td>JARID2</td>
<td>Jumonji, AT rich interactive domain 2</td>
</tr>
<tr>
<td>B3GALT5-AS1</td>
<td>Beta-1,3-galactosyltransferase 5-antisense1</td>
</tr>
<tr>
<td>ESCC</td>
<td>Esophageal squamous cell carcinoma</td>
</tr>
<tr>
<td>HCC</td>
<td>Hepatocellular carcinoma</td>
</tr>
<tr>
<td>STR</td>
<td>Short tandem repeat</td>
</tr>
<tr>
<td>LA</td>
<td>Luminal A</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>UNC</td>
<td>University of North Carolina</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative Real Time-Polymerase Chain Reaction</td>
</tr>
<tr>
<td>LNA</td>
<td>Locked Nucleic Acid</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>TBST</td>
<td>Tris-buffered saline Tween- 20</td>
</tr>
<tr>
<td>Gap</td>
<td>GapmeRs</td>
</tr>
<tr>
<td>CTG</td>
<td>Cell Titer-Glo assay</td>
</tr>
<tr>
<td>HMLE</td>
<td>Human Mammary epithelial cells</td>
</tr>
<tr>
<td>Caspase 3/7</td>
<td>Cysteine-dependent aspartate-directed proteases 3 &amp; 7</td>
</tr>
<tr>
<td>CHIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CHIRP</td>
<td>Chromatin Isolation by RNA Purification</td>
</tr>
<tr>
<td>RIP</td>
<td>RNA Immunoprecipitation</td>
</tr>
<tr>
<td>CDH1</td>
<td>Cadherin 1</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>ITPR1</td>
<td>Inositol 1,4,5-trisphosphate receptor type 1</td>
</tr>
<tr>
<td>NELL2</td>
<td>Neural epidermal growth factor like domain like 2</td>
</tr>
<tr>
<td>TCGA</td>
<td>The Cancer Genome Atlas</td>
</tr>
<tr>
<td>GSVA</td>
<td>Gene set variation analysis</td>
</tr>
<tr>
<td>bHLH</td>
<td>Basic Helix-Loop-Helix</td>
</tr>
<tr>
<td>STRA13</td>
<td>STimulated by Retinoic Acid 13</td>
</tr>
<tr>
<td>SHARP2</td>
<td>Split-and HAiry Related Protein 2</td>
</tr>
<tr>
<td>DEC1</td>
<td>Differentiation in Embryonic Chondrocytes 1</td>
</tr>
<tr>
<td>MOMP</td>
<td>Mitochondrial outer membrane permeabilization</td>
</tr>
<tr>
<td>APAF1</td>
<td>Apoptotic protease activating factor 1</td>
</tr>
<tr>
<td>BIM</td>
<td>Bcl2-interacting mediator of cell death</td>
</tr>
</tbody>
</table>

xvii
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BID</td>
<td>BH3-interacting domain death agonist</td>
</tr>
<tr>
<td>BAD</td>
<td>Bcl2-associated agonist of cell death</td>
</tr>
<tr>
<td>BAK</td>
<td>Bcl2 antagonist/killer</td>
</tr>
<tr>
<td>BAX</td>
<td>Bcl2 associated protein X</td>
</tr>
<tr>
<td>BCL-XL</td>
<td>B-cell lymphoma extra-large</td>
</tr>
<tr>
<td>MCL-1</td>
<td>Induced myeloid leukemia cell differentiation protein</td>
</tr>
<tr>
<td>CLOW</td>
<td>Claudin low</td>
</tr>
<tr>
<td>CTG</td>
<td>Cell Titer Glo</td>
</tr>
<tr>
<td>CCLE</td>
<td>Cancer cell lines encyclopedia</td>
</tr>
<tr>
<td>OMM</td>
<td>Outer mitochondrial membrane</td>
</tr>
<tr>
<td>RIP</td>
<td>RNA Immunoprecipitation</td>
</tr>
<tr>
<td>NAD</td>
<td>Nicotinamide-Adenine- Dinucleotide</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA binding domain</td>
</tr>
<tr>
<td>AD</td>
<td>Auto-modification domain</td>
</tr>
<tr>
<td>CAT</td>
<td>Catalytic domain</td>
</tr>
<tr>
<td>FOXD3-AS1</td>
<td>Forkhead box D3 antisense RNA 1</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor</td>
</tr>
<tr>
<td>PARylation</td>
<td>Poly ADP-ribosylation</td>
</tr>
<tr>
<td>NELL2</td>
<td>neural epidermal growth factor like repeats -like2</td>
</tr>
<tr>
<td>ITPR1</td>
<td>Inositol 1,4,5-trisphosphate receptor type 1</td>
</tr>
<tr>
<td>PARPis</td>
<td>PARP inhibitors</td>
</tr>
<tr>
<td>FDA</td>
<td>Federal Drug Administration</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Breast cancer gene 1</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous recombination</td>
</tr>
<tr>
<td>ATM</td>
<td>Ataxia telangiectasia mutated</td>
</tr>
<tr>
<td>KDM5B</td>
<td>Lysine-specific histone demethylase 5B</td>
</tr>
</tbody>
</table>
Chapter 1.

Breast Cancer, EMT and IncRNAs
1.1 Introduction:

Breast cancer is the most common cancer among women worldwide. Breast cancer is a heterogenic disease with five molecular subtypes including the luminal A, luminal B, normal like, human epidermal growth factor receptor (HER-2) enriched, basal like and the claudin low (Parker et al., 2009; Prat et al., 2010). While the luminal and HER-2 enriched subtypes can be treated with hormonal and anti-HER-2 agents, the TNBC including the basal like and the claudin low subtypes lacks a targeted therapy (Guarneri & Conte, 2009). Despite the advancement in the diagnosis and treatment of breast cancer, it represents the second leading cause of death from cancer following lung cancer in women. The main reason for death from breast cancer is metastasis (Scully, Bay, Yip, & Yu, 2012). However, the driving forces for cancer metastasis need to be better investigated in order to identify new therapeutic targets for the metastatic disease.

The EMT process plays a vital role during embryonic development and is believed to alter the epithelial characteristics of the cancerous cells into mesenchymal properties during cancer metastasis (Nieto, 2013; D. Yao, Dai, & Peng, 2011). This process is regulated by the interaction of multiple signaling pathways, transcription factors and non-coding RNAs (Taube et al., 2010). The best studied non-coding RNAs that play a critical role during EMT process are miRNAs such as members of the miR-200 family (Humphries & Yang, 2015).

Recently, long noncoding RNAs (lncRNAs) emerged as new players in cancer development and metastasis (Q. Xu et al., 2016). LncRNAs usually function through interactions with multi-protein complexes and play critical roles in regulating all levels of gene expression (Kumar, DeVaux, & Herschkowitz, 2016). They can associate with other large RNAs, microRNAs (miRNAs), DNA and proteins to impact nuclear organization, epigenetic modifications, modify mRNA stability, splicing, translation, alter protein localization and stability (Beltran et al., 2008; Mercer & Mattick, 2013). Further, lncRNAs have been found to function as oncogenes (Yuan et al., 2014), support metastasis (Taube et al., 2010) and have been directly implicated in regulating EMT in breast cancer invasion (Knudsen et al., 2012; Yuan et al., 2014).

LncRNA AK001796 (also named linc00978) is an intergenic lncRNA located on chromosome2q13. It was initially discovered in lung cancer as an oncogenic lncRNA that affect cancer cell proliferation in vitro and tumor growth in vivo (Yang et al., 2015). Later, Liu et al. found its association with cisplatin resistance in non-small cell lung cancer. Additional studies
found its association with undifferentiated and advanced staged disease in different cancers such as breast cancer, gastric cancer, esophageal cancer and hepatocellular carcinoma (X. Xu et al., 2019; H. Zhang et al., 2016; Y. Zhao et al., 2018; Zong, Shao, & An, 2019). The biological significance and the mechanisms adopted by AK001796 to induce cancer progression in breast and other cancers have not been elucidated.

1.2 Anatomy and histology of the mammary glands:

The mammary glands are modified apocrine sweat glands located on the anterior wall of the chest that are specialized for production and secretion of milk to the infant (Medina, 1996). The mammary glands are composed of 15-20 lactiferous ducts that start at the nipples and branch gradually to end up with the terminal duct lobular units (TDLU) that are also called acini that are surrounded by adipose and fibrous tissues. These ducts and lobules are lined with two layers: the inner one is simple cuboidal epithelium, and the outer layer is composed of myoepithelial cells and some bi-potent mammary stem cells (MaSCs) (Visvader, 2009). The innermost lining cells are called luminal cells that express low molecular weight keratin 7,8,18 and 19. They also express the epithelial adhesion molecule E-cadherin. The myoepithelial cells that form the outermost layer of the ducts and lobular lining characterized by having high molecular weight keratin 5/6, 14 and 17. These cells lack the hormonal receptors, and they form the origin of basal carcinoma of the breast. These epithelial cell lining forms the origin of the majority of the breast cancer (Li, Moe, & Daling, 2003). Cancer originating from the epithelial lining of the ducts and ductules is called adenocarcinoma which represents 95% of the breast cancer (Makki, 2015).

Based on the origin of the cancer cells, breast carcinomas are classified histologically into ductal or lobular carcinoma that originate from the mammary ducts or lobules. Both types could be in situ like ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS), or invasive;
invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) (Malhotra, Zhao, Band, & Band, 2010). The invasive carcinomas are further classified based on the degree of differentiation of the tissues, nuclear pleomorphism, tubular formation and mitotic index into: well differentiated (grade 1), moderately differentiated (grade 2), and poorly differentiated (grade 3) (Lester et al., 2009).

This classification was valuable for several decades; however, it depends on the histological characteristics only. The identification of the molecular subtypes gives better prognostic value and help in selecting the best targeted treatment approach.

1.3 Breast cancer molecular subtypes:

Breast cancer is a group of different diseases that have different biological features, clinical manifestations, histopathological characteristics though they affect the same anatomical organ and have a common origin (Weigelt, Geyer, & Reis-Filho, 2010). The phenotypic diversity of breast tumors is attributed to the differences in the genes’ expression between different tumors. In 2000, Perou et al. classified breast cancers into four distinct “intrinsic subtypes” according to their genes’ expression patterns (Perou et al., 2000). They put those tumors with ER+ marker and genes expression profile similar to the breast luminal epithelial cells with low molecular weight cytokeratin 8/18 under “luminal-like” category, and those tumors with genes expression profile similar to breast basal epithelial cells with cytokeratin 5/6 or 17 under “basal-like” category, and tumors which express HER2 receptor under “HER2+”, and finally those with expression profile similar to normal breast tissues with gene signature of basal cells and adipose cells under “normal-like” category (Perou et al., 2000). A year later, Sørlie et al. subclassified the luminal subtype to
luminal A, luminal B and luminal C with hierarchical clustering of gene expression pattern from 78 breast cancer samples (Sorlie et al., 2001). In 2006, Hu et al. validated the intrinsic subtypes of breast cancer classification by combining three independent microarray studies (Sørlie et al.(Sorlie et al., 2001), 2001; Van’t Veer et al(van ’t Veer et al., 2002),2002; Sotiriou et al(Sotiriou et al., 2003)) and defined five major molecular subtypes as luminal A, luminal B, HER2+/ER-, basal like and normal like (Hu et al., 2006). In 2007, a new molecular subtype has been identified by Herschkowitz et al. called the “claudin low” (CL) which is characterized by low expression of the epithelial adhesion molecules such as claudin 3, 4, 7 and E-cadherin (Herschkowitz et al., 2007). The claudin low subtype is characterized by low expression of luminal markers and high expression of mesenchymal markers (Taube et al., 2010). Furthermore, CL tumors are the most undifferentiated tumors in comparison with other molecular subtypes and enriched with stem cell characteristics (Sabatier et al., 2014). The luminal A subtype has the best prognostic value, while; the triple negative subtypes including basal and claudin low have the worst prognosis (Sorlie et al., 2001; Sorlie et al., 2003). Combining the prognostic value of the molecular subtypes using prognostic analysis of microarray (PAM50) with the histological grades give a better prediction of the outcome of the breast cancer (Coleman & Anders, 2017). The difference between the molecular subtypes affects the treatment options too. The luminal A subtypes are positive for estrogen receptors (ER) and progesterone receptors (PR) which makes it responsive to hormonal therapy. While the luminal B, that is estrogen positive and HER2 positive may respond to hormonal therapy and HER2 targeted therapy. HER2 subtype can be targeted by monoclonal antibodies as Trastuzumab or tyrosine kinase receptor inhibitor as Lapatinib. However, the triple negative breast cancer, which includes basal like and the claudin low molecular
subtypes are neither HR, nor HER2 receptors positive. Looking for a targeted therapy for these aggressive subtypes of breast cancer is of great scientific interest.

1.4 Epithelial to Mesenchymal Transition (EMT)

Most tumors arise from epithelial cells, that form the lining of the body organs and cavities, forming carcinomas. The epithelial cells characterized by the presence of specialized network of junction molecules that attach the cells to each other and to the basement membrane. Epithelial-cadherin (E-cadherin) is one of the most important junction molecules that protrudes from the surface of the epithelial cell to bind to the ectodomain of E-cadherin of the adjacent cell forming what is called adherens junction (Larsen et al., 2016). At the cytosolic side E-cadherin is bound to actin cytoskeleton through a complex of α and β-catenin giving the tensile strength to the epithelium. Loss of E-cadherin expression is considered as a marker for EMT process (Yuan et al., 2014).

EMT is defined as a molecular reprogramming process in which the epithelial cells lose their adhesion, polarity and acquire a mesenchymal characteristic (Kalluri & Neilson, 2003). During this process, the cells go through a series of changes starting from activation of some transcription factors, signaling molecules, expression of integrin and cell surface receptors, changes in cytoskeleton, resistance to apoptosis, secretion of proteases and increase in motility.

Kalluri and Weinberg described three types of EMT (Kalluri & Weinberg, 2009). The first one is associated with embryonic development during the initial implantation of the embryo in the uterus and initiation of the placenta (Vicovac & Aplin, 1996). The second type of EMT is associated with inflammatory response and fibrosis, where the epithelial cells pass through different stages of EMT undergoing what is called “partial EMT” (Kalluri & Weinberg, 2009). The third type of EMT is
the one that is associated with cancer metastasis. EMT is considered as one of the crucial mechanisms that induce malignancy of the epithelial cells (Thiery, 2002). This mechanism enables the epithelial cells to acquire migratory features, and hence leave the site of the primary tumor by invading the basement membrane and extracellular matrix (ECM) and intravasate inside the circulation, where they migrate to the site of secondary metastasis, extravasate and form a secondary tumor (Thiery & Sleeman, 2006). At the site of secondary metastasis, these migratory cells lose their mesenchymal characteristics and undergo the reverse of EMT process, which is called mesenchymal to epithelial transition (MET) (Zeisberg, Shah, & Kalluri, 2005). Besides its role in tumor metastasis, EMT process has been found to prevent senescence induced by oncogenes inducing aggressive behavior of the metastatic cells (Smit & Peeper, 2008; Weinberg, 2008).

EMT process is initiated by signaling molecules such as transforming growth factor beta (TGF-β), tumor necrosis factor-α (TNF-α), hepatocytes growth factor (HGF), epidermal growth factor (EGF), insulin like growth factor-1 (IGF-1) and platelet derived growth factor (PDGF), which activate multiple signaling pathways leading to the activation of EMT inducing transcription factors such as: TWIST1,TWIST2, Snail1, Sanil2 (SLUG), goosecoid, zing finger E-box binding homeobox 1 and 2 (ZEB1 & ZEB2), and forkhead box protein c2 (FOXC2) (Thiery, 2002), (Jechlinger, Grunert, & Beug, 2002), (Shi & Massague, 2003), (Medici, Hay, & Olsen, 2008), (Kalluri & Weinberg, 2009). Snail1 and Slug work mainly as transcription suppressors inducing EMT mainly by suppressing the epithelial marker E-cadherin and claudin, which form adherens junction and tight junction respectively (Barrallo-Gimeno & Nieto, 2005), (Scheel & Weinberg, 2012). Suppression of the epithelial markers during EMT process is accompanied by induction of mesenchymal markers such as the adhesion molecule neural cadherin (N-cadherin), vimentin
intermediate filament, and the extracellular matrix protein fibronectin (Thiery, Acloque, Huang, & Nieto, 2009).

EMT-TFs not only control the expression of adhesion molecules and other cellular machinery, but they also regulate the expression of each other’s. For example, TWIST binds directly to the promoters of SNAIL, Slug, ZEB1/2 and induces their expression. Slug activates the expression of ZEB1 and both cooperate to suppress E-cadherin (Wels, Joshi, Koefinger, Bergler, & Schaider, 2011). ZEB1 can also induce the expression of other EMT-TFs (Larsen et al., 2016).

In addition to their ability to induce metastatic characteristic of cancer cells, EMT-TFs have been found to induce tumor initiating ability of these metastatic cells “self-renewal capacity” forming what is called cancer stem cells (CSCs). This was observed when normal mammary epithelial cells were forced to express one of the EMT-TFs such as TWIST or SNAIL, the CD44 hi/CD24 low population increased (Mani et al., 2008).

Furthermore, EMT program help metastatic cancer cells to evade premature senescence and apoptosis through various mechanisms, thus increasing the aggressiveness of metastatic cancer cells. (Smit & Peeper, 2008), (Thiery et al., 2009).

EMT process is also controlled by some non-coding RNAs (ncRNAs) like miRNAs, those are short non-coding RNAs around 22 nucleotides that base pair with the 3´UTR of mRNAs targeting them for degradation or controlling their translation. Among the extensively studied miRNAs that play a crucial role in cancer initiation and metastasis is miR-200 family (miR-200a, miR-200b, miR-200c(miR-141, miR-429), which inhibit EMT by targeting ZEB1 and ZEB2 (Humphries & Yang, 2015), (Gregory et al., 2008), (Park, Gaur, Lengyel, & Peter, 2008), (Gibbons et al., 2009).

There are regulatory feedback loops between EMT transcription factors and miRNAs, where ZEB1
and SNAIL1 can bind to the promoters of miR-141 and miR-200c suppressing their expression to maintain EMT phenotype in carcinoma (Burk et al., 2008), (Thiery et al., 2009). Other ncRNAs that play a significant role in regulating the EMT process are lncRNAs.

1.5 Long non-coding RNAs (LncRNAs)

The mechanisms of genes regulations have been centered around protein coding genes, that represent only 2% of the human genome, through the central dogma (DNA is transcribed to mRNA which is translated to protein). What previously was considered as a “junk DNA” are getting more attention by scientists. It is until the last two decades, the regulatory role of ncRNAs in the developmental process of more complex organism became appreciated (Mattick, 2004). The rapid advancement of high throughput sequencing technology revealed tens of thousands of transcripts with no or little protein coding potential (Guttman et al., 2010). Non-coding transcripts involves short ncRNAs such as micro RNAs (miRNAs), siRNAs and long ncRNAs (lncRNAs).

LncRNAs are arbitrarily defined as transcripts which are 200 nucleotides or longer, with no protein coding capacity (Mercer, Dinger, & Mattick, 2009).

Based on their genomic locations, lncRNAs are classified to the following categories:

I. Long intergenic non-coding RNAs (lincRNAs): which are located between the protein coding genes.

II. Intronic lncRNAs: that are transcribed from the introns of protein coding genes.

III. Sense lncRNAs are transcribed from the sense strand of protein coding gene overlapping exons or the entire length of protein coding gene.

IV. Antisense lncRNAs are transcribed from the antisense strand for the protein coding genes and my overlap exons, introns, or the entire sequence of the protein coding genes.
V. Bidirectional lncRNAs: that are transcribed within 1kb of the promoter of protein coding gene from the opposite strand for the protein coding genes (Hermans-Beijnsberger, van Bilsen, & Schroen, 2018).

LincRNAs are more conserved than other types of lncRNAs (Guttman et al., 2010) (Guttman et al., 2009) (Khalil et al., 2009). Their transcription is also more tissue specific than protein coding genes (Derrien et al., 2012) (Cabili et al., 2011; Guttman et al., 2010). lincRNAs have been found to be epigenetically regulated in tissue specific manner; therefore, they represent excellent cellular markers (Amin et al., 2015). Furthermore, regulation of transcription of lincRNAs is similar to that of mRNAs (Guttman et al., 2010) (Guttman et al., 2009; Orom et al., 2010) (Khalil et al., 2009). The majority of lncRNAs are localized to the nucleus and chromatin (Derrien et al., 2012) and a large proportion of them regulate the expression of nearby genes (Ma, Bajic, & Zhang, 2013). They can associate with other large RNAs, DNA and proteins to impact nuclear organization, epigenetic modifications, modify mRNA stability, splicing, and translation, and alter protein localization and stability (Nguyen et al., 2020). Further, lncRNAs have been found to function as oncogenes (Yuan et al., 2014), support metastasis (Taube et al., 2010) and to participate in regulating the EMT process (Yuan et al., 2014).

The main mechanisms by which lncRNAs participate in inducing EMT process and carcinogenesis include:

1.5.1 Transcriptional regulation by lncRNAs:

LncRNAs play a crucial role in tumor progression and metastasis mainly through transcriptional regulation. LncRNAs adopt different mechanisms in regulating transcription starting from modifying chromatin architecture, binding to the DNA in a sequence specific manner, bringing
transcriptional factors and complexes to activate or suppress transcription (Derrien et al., 2012), (Kopp & Mendell, 2018).

At the epigenetic level, hundreds of lncRNAs have been found to interact directly with chromatin remodeling complexes controlling their localization and activity (Khalil et al., 2009) (Fig1. 1A). While many still need to be functionally validated, this suggests that epigenetic modification is likely a core function of lncRNAs. Among the extensively studied lncRNAs that forms a scaffold for chromatin modifying enzymes is HOTAIR (HOX transcript antisense intergenic lncRNA). HOTAIR regulates gene expression by binding to Polycomb Repressive Complex 2 (PRC2) at its 5´domain and changing its occupancy genome wide (Taube et al., 2010). PRC2 is a critical chromatin modifying complex consists of four main subunits (Enhancer of zeste homolog 2 [EZH2], embryonic ectoderm development [EED], suppressor of zeste 12 [SUZ12], and retinoblastoma suppressor associated protein 46/48 [RbAp46/48]) (Zhou et al., 2017). Many other lncRNAs bind to different components of PRC2 to induce changes in genes expression. PRC2 induces genes silencing through histone H3 lysine 27 trimethylation (H3K27m3). HOTAIR also has multiple binding sites at its 3´domain that can bind to lysine-specific histone demethylase 1A (LSD1), restrictive element 1-silencing transcription factor (REST) and REST corepressor1 (LSD1/coREST/REST) complex demethylating H3K4 (Tsai et al., 2010). HOTAIR has been found to be associated with increased risk of metastasis in breast cancer (Sorensen et al., 2013), and required for EMT and stemness of breast, colon, and cervical cancers (Padua Alves et al., 2013), (H. J. Kim et al., 2015).

H19 is another lncRNA that binds EZH2 and promotes EMT in bladder cancer through suppressing the expression of E-cadherin (A. Zhang et al., 2013). MALAT1 (Metastasis associated lung adenocarcinoma transcript 1) also known as nuclear-enriched abundant transcript 2 (NEAT2) is
another extensively studied lncRNA that interacts with chromatin remodeling complex PRC2 to regulate genes expression. MALAT1 is induced by TGFβ, promotes EMT by suppressing the expression of the epithelial marker E-cadherin through its binding to SUZ12 (Yuan et al., 2014). NEAT1 (nuclear enriched abundant transcript1), which is another lncRNA that is mainly localized to the nuclear paraspeckles, recruits the histone deacetylase SIN3A to the transcription suppressor Forkhead box N3 (FOXN3) forming a repressive complex that induce EMT and invasion in ER positive breast cancer (Zhou et al., 2017).

To explain the mechanism by which lncRNAs direct PRC2 to target genes Mondal et al mapped the binding site of MEG3 lncRNA, which also binds to EZH2. They found that MEG3 forms RNA-DNA triplexes with GA-rich regions distal to the promoters of the target genes and directs the chromatin modifiers to these genes. By this mechanism it regulates TGFβ pathway target genes (Mondal et al., 2015). This may represent a general mechanism by which lncRNAs recruit different chromatin modifying enzymes to the target genes.

The second mechanism by which lncRNAs regulate transcription at the epigenetic level is by modulating DNA methyl transferases enzymes (DNMTs) function (Fig1. 1B). LncRNA DBCCR1-003, which is derived from the locus of the tumor suppressor DBCCR1 (deleted in bladder cancer chromosome region 1), directly binds to DNMT1 and blocks its methylation of the promoter of DBCCR1 suppressing the proliferation of bladder cancer cells (H. Zhang et al., 2016).

LncRNAs can regulate transcription by additional mechanisms such as R-loop (DNA: RNA hybrid) formation, binding directly to the promotor or recruiting transcriptional activators or suppressors. The lncRNA vimentin-antisense 1 (VIM-AS1), which is transcribed head-to-head to vimentin (VIM) intermediate filament, is a good example. VIM-AS1 regulates the transcription of VIM by forming stable R-loop that reduces chromatin condensation around VIM promoter and
makes it accessible for transcription factor nuclear factor kappa light chain enhancer of activated B cells (NF-Kb) (Boque-Sastre et al., 2015) (Fig1. 1C).

LncRNAs can also recruit transcription factors to the target genes (Fig1. 1D). For example, LncRNA-HIT (HOXA transcript induced by TGFβ), which has been found to increase invasion and migration in non-small cells lung cancer (NSCLC) (H. Zhang et al., 2016), regulates transcription by binding to the transcription factor E2F1 regulating its occupancy on genes promoters (Zhou et al., 2017). ELIT-1 (EMT associated lncRNA induced by TGFβ) is a newly discovered lncRNA that also works as a scaffold by binding to SMAD3 and facilitate SMAD3/4 complex binding to the promoters of target genes promoting EMT in lung and gastric cancer (Sakai et al., 2019). LncRNA BX-111, that is induced by HIF1-α, mediates hypoxia induced EMT in pancreatic cancer by recruiting the transcription factor Y-box protein (YB1) to ZEB1 promoter (Deng et al., 2018).

Some other LncRNAs work as a decoy for transcription factors and suppress their transcriptional activity. The best studied example on that is GAS5 (growth arrest specific 5), which works as a decoy for glucocorticoid receptor (GR) (Kino, Hurt, Ichijo, Nader, & Chrousos, 2010) (Fig1. 1E). GAS5 binds to DNA binding domain of GR by mimicking the glucocorticoid receptor elements (GREs), thus competing with DNA GREs resulting in growth arrest during stress (Kino et al., 2010) GAS5 is considered tumor suppressor, and has been found to affect proliferation, apoptosis, EMT and metastasis in different types of cancer (H. Zhang et al., 2016), (Zong et al., 2019).

In addition to all these mechanisms, some LncRNAs interfere directly with the transcriptional machinery. DHFR (dihydro-folate reductase) gene transcription is regulated by DHFR lncRNA that is transcribed upstream to DHFR minor promotor. DHFR lncRNA binds to both DHFR major promotor and the major transcription factor TFIIB preventing the formation of preinitiation
complex suppressing the transcription of DHFR in quiescent cells (Martianov, Ramadass, Serra Barros, Chow, & Akoulitchev, 2007) (Nie et al., 2012) (Fig1. 1F).

The elongation of RNA polymerase II is also regulated by non-coding RNA such as 7SK snRNA that is transcribed by RNA polymerase III. 7SK suppresses the activity of positive-elongation transcription elongation factor b (P-ETFb) by forming RNA-protein complex (7SK/HEXIM/P-ETFb) that can be disassembled by the helicase activity of DDX21 (DEAD-box RNA helicase) to release the active P-ETFb (Calo et al., 2015).

These are some examples for the different mechanisms by which lncRNAs contribute to transcriptional regulation of EMT. There are many other lncRNAs that participate in regulating EMT at the transcriptional level; however, the mechanism of their action still to be elucidated.
Figure 1: Mechanisms of regulation of transcription by IncRNAs: A. Recruitment of chromatin remodeling complex PRC2 by lncRNA to induce epigenetic silencing. B. Binding to DNA methyltransferase DNMT1 and inhibiting its epigenetic modification. C. Formation of R-loop by binding of IncRNA to the DNA duplex allowing access of transcription factor (NF-kB) to access promotor of the target gene. D. Recruitment of transcription factor by IncRNA to the promotor of the target gene. E. Suppression of transcription by IncRNA GAS5, that works as a decoy GREs.
for GR. F. Binding of DHFR lncRNA to the major promoter of DHFR gene and to the TFIIB preventing the formation of preinitiation complex.

1.5.2 LncRNAs regulation of splicing:

The newly transcribed RNA molecules have coding regions (exons) and non-coding regions (introns) that are spliced out during or immediately after transcription. RNA splicing is catalyzed by spliceosomes, RNA-protein complexes composed of small nuclear RNAs (snRNAs) that assembled with small nuclear proteins to form small nuclear ribonucleoproteins (snRNPs). The splicing process can produce two or more transcripts from single RNA molecule by a process called alternative splicing (AS) increasing the proteomic complexity. This process is regulated by cis acting sequences (enhancers and silencers sites) and transacting proteins called splicing factors (SFs) (Matera & Wang, 2014). The most common splicing factors are the serine/ arginine-rich proteins (SR proteins), which have one or more RNA recognition motif (RRM) (Zong et al., 2019). Abnormalities in alternative splicing have been linked to many human diseases including cancers (Tazi, Bakkour, & Stamm, 2009).

As part of their post transcriptional genes regulation, lncRNAs are involved in the regulation of mRNA splicing mechanism. The first identified lncRNAs to be associated with RNA splicing are NEAT1 and NEAT2 (MALAT1), where they have been found to be associated with SC35 SF-containing nuclear speckles (Hutchinson et al., 2007). NEAT1 has been found to regulate AS of Peroxisome proliferator-activated receptor gamma (PPARγ) during adipogenesis through its effect on the phosphorylation of the splicing factor SRp40 (Cooper et al., 2014).

NEAT2 is mainly localized to the nuclear speckles and interacts with SR splicing factors. Tripathi et al. found that NEAT2 modulates the nuclear localization and phosphorylation status of SR
proteins (Fig1. 2A), which is required for AS regulation (Tripathi et al., 2010). MALAT1 depletion has been found to change the pattern of AS for many genes (Tripathi et al., 2010), (Bernard et al., 2010). The mechanism by which NEAT2 regulates the phosphorylation of SR proteins is expected to be through its modulation of the localization and activity of kinases like Serine/ arginine -protein kinase 1 (SRPK1), or the phosphatases; Protein Phosphatase 1 (PP1) or Protein phosphatase 2A (PP2A), that influence AS through modifying SR protein (Romero-Barrios, Legascue, Benhamed, Ariel, & Crespi, 2018).

Direct binding of NEAT2 to some splicing factor may disrupt their interaction with other proteins and hence modulate their activity. NEAT2 binding to SFPQ (proline- and glutamine-rich SF) is a good example for that. SFPQ works as a tumor suppressor by binding to the proto-oncogene PTBP2 (polypyrimidine-tract-binding protein 2), that promotes tumor growth. NEAT2 has been found to bind competitively to SFPQ releasing the proto-oncogene PTBP2 in colorectal cancer (Fig1.2B). This mechanism can explain how NEAT2 induces tumor growth and metastasis (Yuan et al., 2014).

Beside its ability to shuttle SFs between the nuclear speckles and transcription site, NEAT2 is found to increase the expression of the splicing factor SRSF1 in hepatocellular carcinoma. SRSF1 modulates the AS of a set of genes that play an important role in cancer progression and maintenance (Malakar et al., 2017). NEAT2 regulates the expression of another splicing factor called RNA binding protein fox-1 homolog 2 (RBFOX2), which modulates AS of the tumor suppressor Kinesin family member 1B (KIF1B), promoting anchorage independent survival and EMT in ovarian cancer (Gordon, Babbs, Cochrane, Bitler, & Richer, 2019).

Another lncRNA that binds to SF and affect colorectal cancer metastasis is LINC01133, which is downregulated by TGFβ. LINC01133 suppresses EMT by working as a target mimic for the
splicing factor SRSF6, which has been found to induce EMT (Kong et al., 2016). By this mechanism, LINC01133 works as a decoy for SRSF6 titrating it away from its target mRNA (Fig 1. 2C), thus suppressing EMT induction by SRSF6.

Another mechanism by which lncRNA modulate AS is mediated by a group of lncRNAs, that are transcribed in the opposite direction to some mRNA called Natural Antisense Transcripts (NATs). NAT can base pair with their sense transcripts forming RNA-RNA duplexes regulating the stability, nuclear export and splicing of mRNA (Khorkova, Myers, Hsiao, & Wahlestedt, 2014), (Romero-Barrios et al., 2018). One of the known NATs that plays important role in protecting cancer cells from programmed cell death, or apoptosis, is SAF lncRNA. SAF is transcribed antisense to FAS ligand, which initiates apoptosis by binding to FAS receptor at the cell membrane. SAF binds to FAS pre-mRNA overlapping exon6 that codes for transmembrane sequences. At the same time, it binds to splicing factor 45 (SPF45), which facilitate AS of FASL pre-mRNA leading to exon6 skipping (Fig 1. 2D). This FASΔEXO6 transcript produce soluble FAS (sFAS) that cannot induce cell death. By this mechanism SAF lncRNA protect cancer cells from apoptosis (Villamizar, Chambers, Riberdy, Persons, & Wilber, 2016).

SAF is an example of NAT that mediate exon skipping. ZEB2NAT or ZEB2-AS1 (zing finger E-box-binding homoeobox2 antisense RNA 1) is an example for a NAT that mediates intron inclusion of the transcription factor ZEB2 inducing EMT (Beltran et al., 2008). ZEB2NAT sequences is complementary to an intron on the 5´UTR of ZEB2 mRNA which has an internal ribosomal entry site (IRES), which is normally spliced out. ZEB2NAT binds to ZEB2 pre-mRNA covering a splicing donor site at that intron leading to conservation of IRES and induce ZEB2 translation (Beltran et al., 2008) (Fig 1. 2E).
Antisense lncRNAs can also regulate AS through recruitment of chromatin modifier to their sense locus. The best studied example is lncRNA asFGFR2 (antisense for fibroblast growth factor receptor 2, FGFR2), which induce cells specific AS. In epithelial cells, asFGFR2 binds to FGFR2 locus and recruits Polycomb-group proteins and H3K36 demethylase Lysine-specific demethylase 2A (KDM2a) hiding the binding site for chromatin-splicing complex. By this mechanism, asFGFR2 promotes inclusion of exon IIIb of FGFR2 in epithelial cells that is normally excluded in mesenchymal stem cells, which lack asFGFR2 transcript (Gonzalez et al., 2015) (Fig1. 2F).

From these examples, we can conclude that lncRNAs regulate alternative splicing through:

i. Regulating the transcription of some SFs
ii. Interacting with SFs modulating their nuclear localization and activity
iii. Altering the phosphorylation status of SFs
iv. Working as a decoy for splicing factors.
v. Antisense lncRNAs (NATs) form RNA-RNA duplex with sense RNA to activate or suppress splicing site (Romero-Barrios et al., 2018).
Figure 1.2: Regulation of splicing by lncRNAs: A. MALAT1 modulates localization and phosphorylation of serine arginine rich splicing factors (SR). B. MALAT1 disrupts the interaction between the splicing factor SFPQ and the proto-oncogene PTBP2 releasing PTBP2 by its interaction with SFPQ. C. LINC01133 works as a decoy for the splicing factor SRSF6 blocking its effect on target mRNAs. D. SAF NAT base pair with FAS pre-mRNA partially recruiting the splicing factor SPF45 to induce exon6 exclusion and produce soluble FAS receptor. E. ZEB2-AS1 overlap splicing donor site leading to inclusion of IRES in ZEB2 mRNA increasing its translation. F. The lncRNA asFGFR2 recruits polycomb-group proteins and KDM2a to FGFR2 pre-mRNA hiding a binding site of chromatin splicing complex, thus induce exon IIIb inclusion.
1.5.3 LncRNAs regulation of mRNAs stability:

Regulation of mRNAs stability shares a similar feature to AS regulation, being mediated by cis acting sequences like AU-rich elements (AREs) and transacting proteins such as RNA-binding proteins (RBPs) including stabilizing and destabilizing factors. LncRNAs have also been found to play a vital role in regulating mRNAs stability by different mechanisms:

The first mechanism is by recruiting different stabilizing and destabilizing RBPs to the target mRNA (Zong et al., 2019). Linc-ROR (long intergenic noncoding RNA-regulator of reprogramming) has been found to bind both stabilizing and destabilizing factors to regulate c-Myc mRNA stability. On one hand, linc-ROR interact with heterogeneous Nuclear Ribonucleoprotein 1 (hnRNP1) facilitating its binding to c-Myc mRNA and hence increases its stability. On the other hand, linc-ROR bind to the AU-rich binding factor 1 (AUF1), which is known to destabilize mRNA, and inhibits its binding to c-Myc (H. Zhang et al., 2016) (Fig 1.3A). Furthermore, Linc-ROR suppresses P53 induction by DNA damage by titrating hnRNP1 away from P53, since hnRNP1 binding to P53 increases its translation (A. Zhang et al., 2013).

Another example of a lncRNA that binds to stabilizing proteins is LAST (lncRNA-assisted stabilization of transcripts). LAST is an oncogenic lncRNA, induced by c-Myc, that promotes cell cycle progression by stabilizing CCND1 (cyclin D1) mRNA through its interaction with cellular nucleic acid-binding protein (CNBP), RNA binding protein (Zhou et al., 2017).

The second mechanism that lncRNAs adopted to regulate the stability of mRNA is by direct binding to the mRNA molecule and stabilizing it. The best example on that is LncRNA AC132217.4, that is induced by KLF8 (Krueppel like factor 8), promotes EMT in oral squamous cell carcinoma by binding to the 3´UTR of IGF2 (Insulin-like growth factor 2) mRNA increasing its stability (Zhou et al., 2017). NATs can also regulate the stability of their sense mRNAs.
PDCD4-AS1 is a lncRNA complementary to programmed cell death 4 (PDCD4) mRNA, tumor suppressor gene, that is suppressed in TNBC and metastatic diseases. PDCD4-AS1 binds to PDCD4 mRNA and inhibits binding of the decay protein HuR, thus stabilizing PDCD4 mRNA in epithelial cells (Jadaliha et al., 2018) (Fig 1.3B). Another example on NAT is MACC1-AS1 (metastasis associated in colon cancer protein1-antisense 1) lncRNA which enhances metabolic plasticity in gastric cancer by stabilizing MACC1 mRNA mediated by AMPK/lin28 pathway (Y. Zhao et al., 2018) (Fig 1.3C).

Third, some lncRNAs regulate the stability of mRNAs indirectly by modulating the stability of RBPs required for mRNAs stability. LncRNA OCC-1 (over expressed in colorectal cancer) has been found to reduce cell death in colon cancer. LncRNA OCC-1 binds directly to HuR RNP, which stabilizes thousands of mRNAs, targeting it for ubiquitination by enhancing its interaction with E3 ubiquitin ligase β-transducin repeat-containing protein 1 (β-TrCP1) (Fig 1.3D). By this mechanism lncRNA OCC-1 destabilizes mRNAs that are required for cell growth through promoting HuR degradation (Weidenfeld & Barkan, 2018).

Forth, lncRNAs can function as a decoy for RBPs that mediate mRNA degradation (R. W. Yao, Wang, & Chen, 2019). Lnc-ASNR (apoptosis suppressing-noncoding RNA) is nuclear localized lncRNA that is upregulated in cancer cells. Lnc-ASNR interacts with AUF1, which is known to induce degradation of Bcl2, apoptotic suppressor, mRNA, and reduces its cytoplasmic pool (Fig 1.3E). By this mechanism it suppresses apoptosis (Kong et al., 2016).

Fifth, lncRNAs can also target mRNA for decay like SMD (staufen mediated decay). Staufen (STAU1) is a double stranded RNA binding protein that binds to intramolecular or intermolecular dsRNAs initiating SMD. Some lncRNAs have been found to partially base pair with the 3’UTR of targets mRNAs through their Alu elements forming STAU1 binding site (SBS) and recruit
STAU1 to initiate SMD (Fig1. 3F). These lncRNAs are called 1/2sbsRNAs such as 1/2sbsRNA1 which targets serine proteinase inhibitor (SERPINE1) and FLJ21870 mRNAs for degradation (Gong & Maquat, 2011).

**Figure 1.3:** LncRNAs regulation of mRNAs stability: 

A. Linc-ROR mediates binding of hnRNP I to c-Myc mRNA and prevents AUFI from binding to c-Myc. 

B. LncRNA PDCD4-AS1 base pair partially with PDCD4 mRNA and block HuR binding site. 

C. LncRNA MACC1-AS1 binds directly to MACC1 mRNA and stabilizes it by activating AMPK inducing lin28 binding to MACC1 mRNA. 

D. LncRNA OCC-1 binds to HuR RBP and enhances its interaction with E3

---

Gong & Maquat, 2011
ubiquitin ligase β-TrCP1 targeting it to proteasomal degradation. E. LncRNA lnc-ASNR interacts with AUFL sequestering it in the nucleus and reducing its cytoplasmic pool leading to increase the stability of its target mRNA. F. \( \frac{1}{2} \) sbsRNA lncRNA base pair with Alu element at the 3´UTR of SMD target mRNAs forming STAU1 binding site (SBS). STAU1 recognizes SBS and bind to \( \frac{1}{2} \) sbsRNA/mRNA duplex initiating SMD.

1.5.4 LncRNAs regulation of miRNAs:

One of the best characterized function of lncRNAs is working as competitive endogenous RNAs (ceRNAs), which is also considered one of the mechanisms by which lncRNAs regulate mRNAs stability. These lncRNAs are also described as “molecular sponges”, where they compete with mRNA for binding miRNA (Q. Xu et al., 2016). Among the extensively studied microRNAs that play a critical role in cancer initiation and metastasis is miR-200 family (miR-200a, miR-200b, miR-200c, miR-141, miR-429), which inhibits EMT by targeting ZEB1 and ZEB2 (Humphries & Yang, 2015), (Gregory et al., 2008), (Park et al., 2008), (Gibbons et al., 2009). LncRNAs that promote EMT by working as sponges for miR-200 family include; Highly upregulated in liver cancer (HULC) (Xie, Ma, & Zhou, 2013), (Panzitt et al., 2007), which induces EMT by competing with ZEB1 for miR-200a-3p (H. Zhang et al., 2016), LncRNA-ATB (activated by TGFβ) which induces TWIST activity by targeting miR-200c (Yuan et al., 2014), (Y. Zhao et al., 2018), LncRNA-HIT, which is associated with tumor growth and metastasis in breast cancer (Richards et al., 2015). H19 binds to miR-200s and miR-138 derepressing ZEB1, ZEB2 and vimentin in lung cancer and colorectal cancer (CRC) (Zhao, Feng, Li, Ma, & Cai, 2019), (Liang et al., 2015). LncRNA-PNUTS is an example of a lncRNA that is regulated by AS and induces EMT. It is derived from PNUTS (phosphatase 1 nuclear
targeting subunit) pre-mRNA in the absence of the splicing silencer hnRNP E1, that is suppressed by TGFβ. LncRNA-PNUTS induces EMT by competitive binding to miR-205 (Grelet et al., 2017). Linc-ROR has been found to increase invasion and metastasis in TNBC by working as ceRNA for ADP-ribosylation factor 6 (ARF6), small GTPase that facilitate cellular invasion, by binding to miR-145 (Eades et al., 2015). Linc-ROR binds to other miRNAs to induce EMT like miR-205 derepressing ZEB2 (Yuan et al., 2014). TUG1 (Taurine upregulated 1) also induces EMT by working as ceRNA for different miRNAs such as miR-145, which target ZEB1 (Zhou et al., 2017). Some other lncRNAs compete with Snail1 and Snail2 for miRNAs binding like CAR10 (chromatin-associated RNA 10), which binds miR-30 and miR203 and induce metastasis of lung adenocarcinoma (LUAD) (Zong et al., 2019). LncRNA Titin-antisense RNA1 (TTN-AS1), that is associated with increased invasion and metastasis in ESCC (esophageal squamous cell carcinoma), induces EMT by competing for miR-133b binding. This stabilizes miR-133b target mRNAs such as Snail, Hur, RNA binding protein that increases mRNA stability, and FSCN1 (actin-binding protein fascine homolog1) (Y. Zhao et al., 2018). Hur stabilizes FSCN1 mRNA by binding to its 3´UTR, which further increases the expression of FSCN1 (Y. Zhao et al., 2018). FSCN1 is required for the migration and invasion of cancer cells and has been found to be associated with poor prognosis (R. W. Yao et al., 2019). Another lncRNA that competes with Snail1 and Snail2 for binding miR-203 is Histocompatibility leukocyte antigen complex P5 (HCP5) lncRNA that is induced by TGFβ, and promotes metastasis in LUAD (Zong et al., 2019). TINCR (terminal differentiation-induced noncoding RNA) induces EMT by titrating miR-125b away from Snail1 in Trastuzumab resistant breast cancer (Zong et al., 2019). Urothelial carcinoma-associated 1 (UCA1) lncRNA increases the expression of Slug in breast cancer by competitive binding of miR-1 and miR-203a (Y. Zhao et al., 2018).
In addition to regulating miRNAs function by sponging, lncRNAs can also regulate the expression of miRNAs at the epigenetic level. HOTAIR induces EMT in gastric cancer by suppressing the expression of miR34a, derepressing its target C-Met (HGF/C-Met/Snail pathway) and Snail (Yang et al., 2015). Maternally expressed 8 (MEG8) lncRNA induces EMT by recruiting EZH2 to the regulatory region of miR 34a and miR 203 increasing the expression of SaiI1 and Sail2 in lung and pancreatic cancers (Terashima, Ishimura, Wanna-Udom, & Suzuki, 2018). Maternally expressed 3 (MEG3) lncRNA recruits JARID2 (jumonji, AT rich interactive domain 2) and EZH2 to the promoters of E-cadherin and miR-200 family genes silencing them and mediates EMT induction by TGFβ in lung cancer (Terashima, Tange, Ishimura, & Suzuki, 2017). Beta-1,3-galactosyltransferase 5-AS1 (B3GALT5-AS1) is another lncRNA that induce EMT by suppressing the expression of miR-203 through direct binding to its promoter. By this mechanism B3GALT5-AS1 increase the expression of ZEB2 and Snail2 inducing EMT in colon cancer (Y. Zhao et al., 2018).

These examples show us that interplay between lncRNAs and miRNAs is pivotal for fine tuning the level of expression of EMT markers and hence regulating metastasis in different types of cancers.
1.6 LncRNA AK001796

AK001796 which is also called LINC00978 or MIR4435-2HG is a lincRNA located on chromosome 2q13 region. It was initially described in 2015 as an oncogenic lncRNA in lung cancer that is involved in resveratrol induced cellular growth inhibition of lung cancer and found to be expressed at higher level in lung cancer tissues than normal adjacent tissues in patients’ derived samples (Yang et al., 2015). Yang et al found that silencing this lincRNA is associated with decrease in cellular proliferation in vitro and diminished tumor growth in vivo (Yang et al., 2015). In another study, AK001796 found to be associated with Cisplatin resistance in NSCLC and its suppression increases the response to cisplatin treatment in lung cancer (Zhou et al., 2017).

Higher expression of AK001796 has also been found to predict poor prognosis in other cancers. In esophageal squamous cell carcinoma (ESCC), higher level of AK001796 is associated with poorly differentiated tumors, lymph node and distant metastasis (Y. Zhao et al., 2018). Furthermore, higher level of AK001796 has been found to be associated with lower overall survival rate in esophageal carcinoma (Zong et al., 2019). Other gastrointestinal (GI) tract cancers show higher level of AK001796 expression such as gastric cancer, where the level of AK001796 in the plasma of patients with gastric cancer was significantly higher than normal individuals and is found to be associated with tumor metastasis and invasion (Y. Zhao et al., 2018). Similar results were found in hepatocellular carcinoma (HCC), where AK001796 expression appeared to be correlated with tumor size, metastasis, and overall survival rate (Zhao et al., 2019).

In breast cancer, higher expression of AK001796 was described by Deng et al. (2016) to be a predictive of poor prognosis (H. Zhang et al., 2016). However, the mechanism by which this lincRNA induces tumor growth and metastasis is not well investigated.
1.7 Aim and objectives:

The association of higher AK001796 expression with advanced stages of several cancers have been proven by previous studies (Yang et al., 2015; H. Zhang et al., 2016; Zhao et al., 2019; Y. Zhao et al., 2018). However, the biological significance and the mechanism adopted by AK001796 to induce cancer progression in breast cancer has not been elucidated.

The aim of this study is to investigate the role of the lncRNA AK001796 in driving the aggressiveness of breast cancer and whether it forms a potential therapeutic target for the treatment of the highly metastatic breast cancers.

The objectives:

1. Determining the association of the lncRNA AK001796 with the EMT process and identification of the EMT markers that can be modulated by AK001796 expression.

2. Investigating the significance of AK001796 expression on the survival of metastatic breast cancer cells and its involvement in evading apoptosis.

3. Understanding the mechanism of action of the lncRNA AK001796 through localization of the genomic location of the lncRNA AK001796 and identification of the interacting protein partner/s.
Chapter 2. Material and methods
2.1 Cell lines and culture media:

Different cell lines have been used in this study. The normal human mammary epithelial cell lines: HMLE and MCF10A, TNBC cell lines: MDA-MB-231, SUM159PT, and SUM149PT, Luminal cell lines: T47D, ZR-75-1, and MCF7 (see table 1). In addition, we employed the EMT induced models, HMLE TWIST, HMLE Snail and HMLE TGFβ cells (from Mani’s lab) (Mani et al., 2008). Cell lines identities were verified by STR analysis performed at the HT Genomics Core facility, University at Albany. All these cell lines were tested and found to be free for mycoplasma using mycoplasma PCR detection kit form applied biological material (catalog number G238).

Media used are: for MDA-MB-231 and MCF7; Dulbecco’s Modified Eagle Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), 500 µg/ml of 100x Antibiotic/Antimycotic (Anti-Anti) from Gibco. For SUM159PT, and SUM149PT cell lines, Ham’s F12 medium was used supplemented with 5% (v/v) FBS, hydrocortisone 1 µg/ml, insulin 5 µg/ml, and 500 µg/ml of 100x Anti-Anti. T47D, and ZR-75-1 cells RPMI 1640 medium supplemented with 10%(v/v) FBS, and 500 µg/ml of 100x Anti-Anti.

HMLE medium is prepared from DMEM: F12 and MEGM (Mammary Epithelial Cells Growth medium) supplemented with 500 µg/ml of 100x Anti-Anti, 5 µg/ml human insulin, 10 ng/ml EGF, and 0.5 µg/ml hydrocortisone. For MCF10A, DMEM: F12 medium was used supplemented with 5% horse serum, 500 µg/ml of 100x Anti-Anti, 10 µg/ml human insulin, 0.5 µg/ml hydrocortisone, 20ng/ml EGF, and cholera toxin.

The prostate cell lines used are: DU145, that are cultured in ATCC-formulated Eagle's Minimum Essential Medium with 10% FBS, and LNCaP which grow in RPMI-1640 medium with 10% FBS.
The lung cancer cell lines used are: A549, HCC4006 and H1703. The growth media used for these cell lines are: for A549 is ATCC-formulated F-12K Medium with 10% FBS; for HCC4006 and H1703 is RPMI-1640 with 10% FBS.

Table 1. Breast cancer cell lines used in this study classified based on molecular subtype:

<table>
<thead>
<tr>
<th>Cell lines</th>
<th>ER</th>
<th>PR</th>
<th>HER2</th>
<th>Subtypes</th>
<th>Basic media</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCF7</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>LA</td>
<td>DMEM</td>
</tr>
<tr>
<td>T47D</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>LA</td>
<td>RPMI</td>
</tr>
<tr>
<td>ZR751</td>
<td>+</td>
<td>+/-</td>
<td>-</td>
<td>LA</td>
<td>RPMI</td>
</tr>
<tr>
<td>MDA-MB-231</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TNB</td>
<td>DMEM</td>
</tr>
<tr>
<td>SUM159PT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TNB</td>
<td>Ham’s F12</td>
</tr>
<tr>
<td>SUM149PT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>TNB</td>
<td>Ham’s F12</td>
</tr>
</tbody>
</table>

This table is derived from (Dai, Cheng, Bai, & Li, 2017).

2.2 Lentiviral transfection and transduction:

Lentivirus harboring ZEB1 or AK001796 are generated using FUW (Lois, Hong, Pease, Brown, & Baltimore, 2002) and pLenti puro lentiviral vectors subsequently together with packaging plasmids (2.5µg pRC-CMV-RaiI + 2.5µg HDM-Tat16 + 2.5µg HDM-HGPM2 + 2.5µg HDM-VSVG) by transient transfection of HEK293T cells. Lipofectamine 3000 (Thermofisher Scientific catalog # L3000008) is used as a transfection reagent. The supernatant containing the lentiviruses was collected in 15 ml conical tubes and filtered through 0.45 µm filter and stored at -80 °c for future transduction. For transduction, cells were incubated in 6 wells plates for 24 hours or until they become 60-70% confluent prior to transduction with the viral suspension. These cells were selected with puromycin and expanded further for future use.
2.3 Gene expression analysis

Data in Figure 2.1A, B were from UNC microarray database, 2.1A: 244Kv5 array list (236 samples were included in the analysis to compare Claudin vs others). 1B: 10 samples comparing cells underwent EMT (Snail, Twist, GSC and TGFb1) overexpressing HMLE cells) vs control. SAM analysis for both the 244Kv5 breast tumors (236 samples) and those cells underwent EMT. AK001796 are in both up list with q.value=0.

For RNA-seq analysis: Raw fastq files were aligned to the human hg38 genome using STAR v2.4.2 (Dobin et al., 2013). Transcript abundance for each sample was estimated with salmon v0.1.19 (Patro, Duggal, Love, Irizarry, & Kingsford, 2017) to quantify the transcriptome defined by gencode (v22) gene annotation. Read counts were summed over technical replicates. Gene level counts were summed across isoforms and genes with low expression (median less than 10 reads) were removed before downstream analyses. DESeq2 were used to test for differentially expressed genes between control and knockdown samples. We used GSVA method to calculate gene set enrichment scores in msigDB Hallmark gene set (v7.0), then the limma package was applied to test for differentially expressed gene sets. R ComplextHeatmap package was used to make heatmaps.

2.4 Cellular transfection:

For transfecting the cells with ASOs Gapmers or siRNA, Lipofectamine RNAiMax reagent from Invitrogen was used as the transfection reagent. The transfection mixture was prepared by mixing 50 nM Gapmers, that were diluted in 125µl of serum free media with Lipofectamine solution (7.5 µl + 125µl of Opti-MEM) per well of 6 wells plate. The transfection mixture is incubated for 20-
30 min at room temperature. 300,000 cells/well were plated in each well of 6 wells plate. The cells were incubated for 72 h (or less according to experiment conditions) at 37°C and 5% CO2.

2.5 RNA isolation and qRT-PCR:
Total RNA is isolated with E.Z.N.A Total RNA Kit 1 (OMEGA Bio-Tek) according to manufacturer’s protocol. The concentration and the quality of the RNA were measured by NanoDrop 1000 spectrophotometer. cDNA was synthesized from the extracted RNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). TaqMan Master Mix from Applied Biosystems was used to detect the expression level of IncRNA AK001796 and other genes of interest by qRT-PCR. The qRT-PCR data were analyzed with 2 -ΔΔCt method. GAPDH and 18S are used as housekeeping controls.

2.6 Immunofluorescence staining:
For immunofluorescence staining, 300,000 cells were plated on sterile coverslips in each well of six wells plate and transfected with 50nM Locked nucleic acid (LNA) Gapmers. After a 72 h incubation period, the medium was aspirated from each well and the cells were washed with 1X Phosphate buffered saline (PBS). For fixation, cells were incubated with 3.7% formaldehyde for 15 minutes, then washed three times with 1X PBS, and permeabilized by incubation with 0.5% TritonX-100 for 10 minutes. After that, the cells were washed with 1X PBS and blocked by incubation with 5% bovine serum albumin (BSA) in PBS for 10 min. The cover slips then incubated with E-cadherin antibodies, ABfinity, rabbit oligo clonal (Invitrogen catalog# 710161). The secondary antibodies that we used were goat anti-rabbit IgG Alexa Fluor 594 conjugate (Invitrogen catalog# 11037). The cells were incubated with secondary antibodies and DAPI at
1:1000 dilution in dark at 37°C for 1 h. The fluorescent signal was visualized using EVOS FL cell imaging system (Thermo Fisher Scientific).

2.7 Western blotting:
Cells’ lysates were prepared using M-per mammalian protein extraction reagent from Thermo Fisher scientific. Protein concentration was measured using BCA assay (Thermo Fisher Scientific). 40 micrograms of proteins were used from each cell lysate to be separated on 10% SDS-PAGE and transferred to PVDF membrane. The blocking buffer used is 5% BSA in tris-buffered saline Tween-20 (TBST). The primary antibodies that were used: Anti-BCL2L11 (Bim) rabbit primary antibodies from Abcam, E-Cadherin rabbit mAb (Cell signaling Technologies # 3195S), Dec1 Ab (Novus Biological # 1800), PARP-1 N-terminal antibody (Active Motif #39559) and GAPDH (Cell Signaling Technologies # 2118S) used as a loading control. Secondary Ab is Anti-rabbit, HRP- linked antibodies (Cell Signaling Technologies # 7074). The developing reagent used is SuperSignal West Dura Extended Duration Substrate from Thermo Fisher Scientific (#37071). The images obtained using iBright Imaging System (Thermo Fisher Scientific).

2.8 Mammosphere formation assay:
SUM159PT and MDA-MB-231 cells were transfected with 50nM GapmeRs and incubated for 72 h in cell culture at 37°C and 5% CO2. The dead cells were removed with the media and only the remaining viable cells were used for mammosphere formation assays. 10,000 cells were plated per well in low attachment 6 wells plate in MammoCult Human Medium from STEMCELL Technologies (catalogue # 05620) and incubated in cell culture incubator at 37°C and 5% CO2 for 7 days. After that, the mammospheres were imaged using EVOS FL cell imaging system and the
number of mammospheres were counted on IN Cell analyzer 6000. Then the primary mammospheres were collected with the media in 15 ml conical tubes and centrifuged to pellet down the mammospheres that were disrupted with 300 ul of trypsin for 2-3 minutes to get a suspension of single cells to be used for secondary mammosphere formation. 10,000 cells from this suspension were plated in Mammosphere media and incubated for another 7 days to develop secondary mammospheres.

2.9 Apoptosis Assays:

The apoptosis assays that were used in this study are: Muse Annexin V and Dead Cell Assay (Cat: MCH100105) and Muse Caspase-3/7 (Cat: MCH100108). Cells were transfected with LNA GapmeRs (Exiqon) in six wells plate and incubated for 72 h before running the apoptosis assays. A suspension of 100000 cells from each sample is mixed with the assay reagents, and the results are obtained from flowcytometric analysis on Muse Cell analyzer following manufacturer’s instructions. The remaining of the transfected cells were lysed to be used for RNA extraction and qRT-PCR to check the level of silencing the LncRNA AK001796.

2.10 Cellular viability assays:

Cellular viability was detected using Cell Titer-Glo (CTG) chemiluminescent cell viability assay (Promega). 25000 cells are plated per well in 96 wells plate and transfected with 50nM GapmerRs. Following 72 h incubation in cell culture at 37°C and 5% CO2, the plate was incubated at room temperature for 30 min. Then, the cells are mixed with 30 µl of CTG reagent according to the manufacturer’s protocol. The luminescence signal is measured using Perkin Elmer Victor 3V 1420 multilabel plate counter. Cellular viability was calculated as fold changes relative to the control.
2.11 Chromatin Immunoprecipitation (CHIP):

1. **Cross linking the cells:**

SUM159PT cells were expanded in 150mm plates (one plate per sample). When the cells reached 90% confluency, they were trypsinized and washed with 1X PBS. Then fixed with 1% formaldehyde with rotation at room temperature for 10 minutes. The cross linking was stopped by incubating with 0.125 M glycine for 5 minutes with rotation at room temperature. Followed by two washes with ice cold 1X PBS. The fixed cells were stored at -20°C.

2. **Cell lysis and chromatin shearing:**

Cells were lysed by incubating in 1 mol lysis buffer (1% SDS, 10 mM EDTA, protease inhibitors, 50 mM Tris–HCl, [pH 8]) for 10 minutes at 4 °C (in cold room) with rotation.

The chromatin was sheared with sonication at 4°C for 35 cycle (30sec on/45 sec off) to get DNA fragments between 200-500 bp. After sonication, cell debris was pelleted by centrifugation at 4°C for 10 min and the supernatant was diluted 10-fold with RIPA buffer (50 mM Tris-HCl pH8, 150 mM NaCl, 2 mM EDTA pH8, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS and protease inhibitors) and used for immunoprecipitation.

3. **Immunoprecipitation:**

The sheared diluted chromatin was incubated with 5 ug of antibodies (PARP-1 N-terminal antibody (pAb) from Active Motif or IgG form Cell Signaling #2729) over night at 4°C with rotation. The immune-complexes were collected with 50 ul of ChIP-Grade Protein G Magnetic Beads #9006 from Cell Signaling. The beads were washed tow times with CHIP lysis buffer. Then incubated with the immune complexes for 3 hours at 4°C with rotation. The samples are centrifuged at 2000g for 1 minute and the supernatant was removed. The beads were washed sequentially with 1 ml of the following buffers: low salt wash buffer (0.1% SDS, 1% Triton X-
100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris–HCl, [pH 8]), high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris–HCl, pH 8.1) and LiCl wash buffer (0.25 M LiCl, 1% Nonidet P-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris–HCl, [pH 8]). Finally, the beads were washed twice with 1 ml TE buffer (1 mM EDTA, 10 mM Tris–HCl, [pH 8]). The immunoprecipitated genomic DNA was cleared form RNA and proteins with RNase A and proteinase K respectively and extracted using PCR purification kit form Qiagen. The eluted DNA is used for library preparation for CHIP-seq or qRT-PCR to look for PARP1 target genes.

2. 12 Chromatin Isolation by RNA Purification (ChIRP):
SUM159PT cells were used in this experiment. Cells were expanded to get 60 million cells per biological replicate (20 million per sample). Cells were fixed with 1% glutaraldehyde with shaking at room temperature for 10 min and quenched with 125mM Glycine with shaking for 5 min. The fixed cells were collected with centrifugation and washed twice with ice cold PBS and stored in -20°C for future use. Then fixed cells were lysed with CHIRP lysis buffer and sonicated in 4°C water bath at setting with 20 cycles, 30 seconds ON, 45 seconds OFF for total of 100 cycles or till got fragment sizes between 200-500 on agarose gel. The sonicated cell lysates were centrifuged at 16100 g at 4°C for 10 min and the supernatants were pooled together. Then 1 ml of the supernatant was incubated with 100 pmol of biotinylated sense or antisense oligonucleotides for IncRNA AK001796. Lac Z oligos were used as a negative control. For this experiment Magna ChIRP RNA Interactome Kit (EMD Millipore #17-0494) is used following manufacturer’s protocol. The extracted RNA was subjected to next gen sequencing on an Illumina Next Seq500 platform at HT Genomics Core facility, University at Albany.
2. 13 RNA-associated protein pull-down assay:

10-20 million cells were lysed under native conditions with lysis buffer (150 mM NaCl, 10 mM Tris-HCl pH 7.4, 1 mM EDTA, 1% Triton X-100, 0.5% NP-40) supplemented with protease inhibitor cocktail and RNase inhibitor. This was followed by a brief sonication to ensure complete lysis. The cellular lysate was centrifuged briefly to remove of insoluble fractions. After preclearing with streptavidin beads the lysate was incubated with 400 pmol of each of biotinylated sense and antisense oligonucleotide corresponding to lncRNA AK001796 overnight at 4°C. Each fraction was further incubated with Dynabeads MyOne Streptavidin T1 magnetic beads for 2 h at 4°C. The tubes were placed against the magnetic stand and the supernatant was discarded. The beads bound with protein were washed 4-5 times with washing buffer (20 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1% Tween 20, 1% NP-40) and finally eluted into Laemmli buffer and run on 10% SDS-PAGE gel. The gel was stained with colloidal coomassie blue staining and after washes the unique bands were submitted for mass-spectrometry for protein identification.

2. 14 RNA Immunoprecipitation (RIP):

SUM159PT cells were fixed with 1% formaldehyde for 10 min and quenched with 125 mM Glycine. The pelleted cells were suspended with 2 ml of 1X PBS, 2 ml of nuclear isolation buffer (1.28 M sucrose, 40 mM Tris-HCl pH 7.5, 20 mM MgCl2, 4% Triton X-100) and 6 ml nuclease free H2O and incubated on ice for 20 min with intermittent mixing. The nuclei were pelleted by centrifugation at 4°C/ 1000 rpm for 15 min and resuspended in 1 ml of RIPA buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 100 U/mL RNAase inhibitor and Protease inhibitors). Chromatin was sheered by sonication at 4°C, 30 sec on/45 sec off setting for 30 cycles total. Nuclear debris were pelleted by centrifugation at 4°C/ 13000 rpm for 15 min. The
chromatin containing supernatant was incubated with PARP1 specific antibodies or IgG at 4°C with rotation overnight. The immunoprecipitated material were then incubated with protein G magnetic beads for 2 h at 4°C. The pelleted beads were washed 5X with RIP buffer followed by 1 wash with 1X PBS. After protein digestion with proteinase K, the RNA was extracted with Trizol reagent, reverse transcribed and analyzed by qRT-PCR to check the enrichment of AK001796.

**2.15 PARP Activity assay:**
SUM159PT cells were transfected with 50 nM ASOs against AK001796 and incubated in a 96 well plate for 24, 48 and 72 h. Then cells were lysed with lysis buffer and cellular lysates were plated on a histone coated 96 well plate from PARP/Apoptosis Chemiluminescent Assay (from Trevigen catalogue #4685-096-K). PARP activity was measured following the manufacturer’s protocol.

**Table 2: Antisense LNA GapmeRs:**

<table>
<thead>
<tr>
<th>GapmeR name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNCAK001796-2</td>
<td>5´- GAG TTA AGG AAA GCTT- 3´</td>
</tr>
<tr>
<td>LNCAK001796-3</td>
<td>5´- GCG AGT GAC AGA AAT - 3´</td>
</tr>
</tbody>
</table>

**Table 3: Silencer® Select Pre-designed siRNAs (Life Technologies):**

<table>
<thead>
<tr>
<th>siRNA</th>
<th>Catalogue number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative control No.1</td>
<td>4390843</td>
</tr>
<tr>
<td>BHLHE40</td>
<td>4392420, siRNA ID# s16281 and s16283.</td>
</tr>
<tr>
<td>PARP1</td>
<td>4390824</td>
</tr>
</tbody>
</table>
Table 4: TaqMan primers:

<table>
<thead>
<tr>
<th>Primer</th>
<th>Assay ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S</td>
<td>Hs99999901_s1</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Hs99999905_A1</td>
</tr>
<tr>
<td>CDH1</td>
<td>Hs01023894_m1</td>
</tr>
<tr>
<td>AK001796</td>
<td>FWD: 5´/Biosg/AGGGCAAAATCACCCCTTGT - 3´&lt;br&gt;REV: 5´/Biosq/AGAGCAAAGTCCAGGCTTG - 3´</td>
</tr>
<tr>
<td>VIM</td>
<td>Hs00958111_m1</td>
</tr>
<tr>
<td>TWIST</td>
<td>Hs01675818_s1</td>
</tr>
<tr>
<td>BCL2L11</td>
<td>Hs00708019_s1</td>
</tr>
<tr>
<td>BHLHE40</td>
<td>Hs01041212_m1</td>
</tr>
<tr>
<td>Snail</td>
<td>Hs00195591_m1</td>
</tr>
<tr>
<td>ZEB1</td>
<td>Hs00232783_m1</td>
</tr>
<tr>
<td>PARP1</td>
<td>Hs00242302_m1</td>
</tr>
</tbody>
</table>

Table 5: CYBR Green qPCR primers

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITPR1</td>
<td>FWD: 5´- TGCCTCCACAATTCTACG- 3´&lt;br&gt;REV: 5´- TGAATGTCCCACAGTTGC- 3´</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td></td>
</tr>
<tr>
<td>NELL2</td>
<td>FWD: 5´- TGAAGGGAACCACCTACC- 3´&lt;br&gt;REV: 5´- ATTTGCCATCCACATACG- 3´</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td></td>
</tr>
<tr>
<td>ITPR1</td>
<td>FWD: 5´- GAGCCCTAAGCAGCGTGTAG- 3´&lt;br&gt;REV: 5´- CTCTCCAAGACCTCCGAATG- 3´</td>
</tr>
<tr>
<td>CHIP-qPCR</td>
<td></td>
</tr>
<tr>
<td>NELL2</td>
<td>FWD: 5´- TCCCGGAGGAGGTCTCTC- 3´&lt;br&gt;REV: 5´- CGCCCGAACCTTGGTAAG- 3´</td>
</tr>
<tr>
<td>CHIP-qPCR</td>
<td></td>
</tr>
<tr>
<td>BCL2L11</td>
<td>FWD: 5´- GGGGCAAGGAGGTACTAGG- 3´&lt;br&gt;REV: 5´- TACATGACGGAGCTGGAAC- 3´</td>
</tr>
</tbody>
</table>
Chapter 3: Discovery and association of the IncRNA AK001796 with EMT
3.1 Introduction:

Targeting EMT process could be a clue for the treatment of highly metastatic breast cancers. Though extensively studied, EMT process complexity still to be investigated. Most of the research focused on the protein coding genes that regulate this process. However, non-coding RNAs including short non-coding RNAs and long non-coding RNAs emerged as master regulators of EMT at different levels. Many IncRNAs have been found to modulate EMT by different mechanisms as we mentioned in the first chapter. Many of those IncRNAs induce EMT and called “pro-EMT”, others work by suppressing EMT “anti-EMT” (Gugnoni & Ciarrocchi, 2019). Some IncRNAs are linked to specific tumors like HULC (Panzitt et al., 2007), or particular molecular mechanisms like Gas5, which works as a tumor suppressor (Zong et al., 2019).

Our goal was to identify IncRNAs specifically associated with the EMT program in breast cancer. We started with global expression analysis for IncRNAs expression in patients’ derived samples and EMT induced models. Followed by qRT-PCR validation for the differentially expressed IncRNAs to pinpoint those IncRNAs that are consistently upregulated in EMT at significant levels. LncRNAs appeared to be differentially expressed between the different molecular subtypes.

3.2 LncRNA AK001796 is upregulated under EMT in breast cancer:

Since the claudin low subtype of breast cancer is highly enriched with EMT genes signature (Taube et al., 2010), we first identified the IncRNAs abundantly expressed or depleted in this aggressive subtype of breast cancer. A large set of patients’ derived breast tumors (~200), that represent the different molecular subtypes of breast cancer, was used to examine the expression profiles of >17,000 IncRNAs using a custom designed chip from Agilent. Those IncRNAs that are expressed
in the claudin low molecular subtype versus others were clustered together as seen in the heatmap of unsupervised cluster of lncRNAs in Fig. 2.1A. AK001796 was one of those lncRNAs that were expressed at higher levels in the claudin low subtypes relative to normal and other subtypes (Fig. 2.1B).
Figure 2.1: Differential expression of lncRNAs between the claudin low and other molecular subtypes of breast cancer: A. Microarray analysis for over 200 breast tumors showing the top lncRNAs differentially expressed between the claudin low and other intrinsic subtypes of breast cancer. False discovery rate (FDR) = 0.01%. B. Boxplot comparing AK001796 expression in the claudin low and other molecular subtypes of breast cancer patients’ derived samples.

MiTranscriptome was utilized to evaluate lncRNA AK001796 expression from data published in the Cancer Genome Atlas (TCGA), the ENCODE project, and additional published sources (Iyer et al., 2015). AK001796 is significantly upregulated in breast carcinoma (12.43 FPKM) compared to adjacent normal (4.49 FPKM) (Fig. 2.2).

Figure 2.2: AK001796 expression data in adjacent normal (n=105) and primary breast cancer (n=892) from MiTranscriptome (Iyer et al., 2015) *p < 0.05, student’s t-test.
In order to pinpoint the lncRNAs specifically associated with EMT in breast cancer, we used the same lncRNA array to screen human mammary epithelial (HMLE) cells and HMLE cells that have been induced to undergo EMT through expression of EMT transcription factors, TWIST, SNAIL1, TGFβ or GSC (Mani et al., 2008) (Fig. 2.3A). We found that AK001796 is consistently upregulated under EMT conditions (Fig. 2.3B).

Figure 2.3: Differential expression of lncRNAs between HMLE and EMT induced models:
A. Heat map of 173 upregulated and 162 downregulated lncRNAs following induction of EMT by
TGFB1, Snail1, Twist, and Goosecoid (Gsc). B. qRT-PCR results for AK001796 in HMLE cells expressing different EMT inducing factors.

Furthermore, we induced EMT in three non-EMT breast cancer cell lines by overexpressing the EMT-specific transcription factor ZEB1. The level of ZEB1 overexpression was confirmed with qRT-PCR (Fig. 2.4A). We found increased expression of AK001796 corresponding to the level of ZEB1 overexpression (Fig. 2.4B). The induction of EMT by ZEB1 overexpression was confirmed by increased expression of TWIST, SNAI1 and vimentin (Fig. 2.5C &D).

Figure 2.4: Increased expression of AK001796 in other breast cancer cell lines upon induction of EMT: A. qRT-PCR for ZEB1 overexpression in different breast cancer cell lines B.
qRT-PCR results showing the level of expression of AK001796 in breast cancer cells induced to undergo EMT by ZEB1 transcription factors. C & D. qRT-PCR for the EMT markers (TWIST, SNAI1 & VIM) in ZEB1 overexpressing SUM149PT and ZR-75-1 cell lines respectively. The graphs represent average +/- SD, *p < 0.05, student’s t-test.

To compare the expression of lncRNA AK001796 between the different molecular subtypes, the levels of lncRNA AK001796 expression was examined in a large set of breast cancer cell lines using the Cancer Cell Line Encyclopedia (CCLE). We found that AK001796 is highly abundant in TNBC cell lines as compared to luminal or HER2-enriched subtypes of breast cancer (Fig. 2.5A). These results were confirmed with qRT-PCR on a panel of breast cancer cell lines grouped according to their molecular subtypes (Fig. 2.5B).
Figure 2.5: A. The level of expression of lncRNA AK001796 in breast cancer cell lines grouped according to their molecular subtypes using BigQuery output from Cancer Cell Line Encyclopedia (Tatlow & Piccolo, 2016). B. qRT-PCR for AK001796 expression in different breast cancer cell lines grouped according to their intrinsic subtype. The graph represent mean ± SD.
To answer the question whether the expression of lncRNA AK001796 is associated with EMT signature genes in breast cancer patients, the expression data from The Cancer Genome Atlas (TCGA) were used to identify genes whose expression are positively correlated with lncRNA AK001796 expression. Those genes were then used to compute overlaps with the Molecular signature data base (MsigDB) Hallmark gene sets. The most significantly enriched gene set is the EMT pathway (Fig. 2.6), in addition to other signatures for aggressive breast cancer (Table 5).

**Figure 2.6:** Gene Set Enrichment Analysis showed enrichment for the hallmark EMT (p-value = 3.34E-144) positively correlated with expression of AK001796 in breast cancer.
The previous in silico analysis was further validated with loss of function study. Where AK001796 silencing with modified antisense oligonucleotides in SUM159PT resulted in down regulation of EMT hallmark genes in addition to other hallmarks. The gene set variation analysis (GSVA) (Hanzelmann, Castelo, & Guinney, 2013) of RNA-seq data reveals that AK001796 silencing was associated with down regulation of some gene sets and signaling pathways that have significant roles in cancer metastasis such as TGFβ signaling, Notch signaling, Hedgehog signaling, angiogenesis and other pathways (Fig. 2.7B).
Figure 2.7: A. qRT-PCR for AK001796 in SUM159PT at 36 hours post transfection with GapmeRs. B. RNA-seq in SUM159PT cells following AK001796 depletion was analyzed via gene set variation analysis (GSVA) using the msigDB hallmark gene sets. Heatmap shows all hallmark gene sets with p-value less than 0.05.
3.3 Perturbation of IncRNA AK001796 leads to altered expression of EMT markers in breast cancer.

To determine if expression of AK001796 functionally supports EMT in addition to serving as a biomarker, we studied the impact of AK001796 expression on EMT gene expression. Antisense locked nucleic acid GapmeR oligonucleotides were designed to inhibit expression of AK001796 because of its nuclear localization (Fig S1A).

MDA-MB-231 and SUM159PT cells were transfected with two independent GapmeRs (Gap) targeting AK001796. Total RNA was extracted and evaluated for gene expression. Depletion of AK001796 enhanced the expression of epithelial marker gene E-cadherin in both triple-negative breast cancer cell lines (Fig. 3.1A). Using the E-cadherin-specific antibody, we observed similar results by immunofluorescence and immunoblotting (Fig. 3.1B and Fig. S1B). In a reciprocal experiment, AK001796 was overexpressed in an EMT-negative luminal cell line MCF7 (Fig 3.2A) resulting in a reduction in the expression of E-cadherin at both the mRNA and protein levels (Fig. 3.2 A and 3.2B). In addition to loss of E-cadherin, AK001796 expression increased expression of the mesenchymal markers TWIST and vimentin indicating the EMT-driving potential of IncRNA AK001796 (Fig.3.2A).

The Co-expression analysis of AK001796 (that is annotated as MIR4435-2HG) and EMT markers on cbio portal using 1108 patients’ samples from breast invasive carcinomas (TCGA, Firehose legacy) shows positive correlation with the mesenchymal markers TWIST, SNAI1, VIM and negative correlation with the epithelial marker E-cadherin coding gene, CDH1 (Fig. S2 A, B, C &D).
Figure 3.1: E-cadherin expression following AK001796 silencing: A. qRT-PCR results for the epithelial marker E-cadherin in MDA-MB-231 and SUM159PT cells after knocking down AK001796. The graph represent mean ± SD, n = 3, Student’s t test, *p < 0.05. B. Immunofluorescence staining for the epithelial markers E-cadherin in MDA-MB-231 cells after AK001796 silencing with GapmeRs 2 and 3.
Figure 3.2. AK001796 overexpression leads to induction of EMT: A. qRT-PCR showing the level of expression of AK001796, CDH1, Twist and Vimentin in MCF7 cells overexpressing AK001796. B. Immunoblotting for the epithelial marker E-cadherin in MCF7 cells overexpressing AK001796. The right panel represents quantification of three biological replicates of E-cadherin immunoblotting. Bars represent mean ± SD, n = 3, Student’s t test, *p < 0.05.
3.4 LncRNA AK001796 contributes to maintaining stemness in EMT high breast cancer cells

The EMT program has been found to induce tumor initiating ability of the metastatic cells “self-renewal capacity” forming cancer stem cells (CSCs) (Mani et al., 2008).

From the previous section, we found that the lncRNA AK001796 regulates the expression of some EMT markers. This raises the question whether AK001796 expression affects other features of the EMT process like stemness in breast cancer cells.

First, we wanted to check the growth promoting potential of lncRNA AK001796. For this purpose, colony formation assay was performed on the viable cells depleted for the lncRNA. Where SUM159PT and MDA-MB-231 cells were transfected with GapmeRs and after 72 hours the viable cells were used for this experiment. We observed a substantial reduction in the colony formation following lncRNA AK001796 depletion relative to control (Fig. 3.3A).

The mammosphere formation assay then was used to determine if AK001796 contributes to cancer cell stemness. Silencing of AK001796 resulted in a reduction in the number of primary mammospheres formed in both TNBC lines MDA-MB-231 and in SUM159PT as compared to control cells (Fig. 3.3B). Formation of secondary mammosphere, however, is a more robust feature of breast cancer stem cells. Cells from primary mammospheres were harvested and upon dissociation to single cell stage, were further seeded at limiting dilution to form secondary mammospheres. Secondary mammosphere formation is also decreased upon depletion of lncRNA AK001796 indicative of its vital role for breast cancer stemness (Fig. 3.3C and D).
A Colonies formation assay

<table>
<thead>
<tr>
<th></th>
<th>MDA-MB-231</th>
<th>SUM159PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap Ctrl</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
</tr>
<tr>
<td>Gap 2</td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
</tr>
<tr>
<td>Gap 3</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
</tbody>
</table>

B Primary Mammosphere

<table>
<thead>
<tr>
<th></th>
<th>MDA-MB-231</th>
<th>SUM159PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap Ctrl</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
</tr>
<tr>
<td>Gap 2</td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
</tr>
<tr>
<td>Gap 3</td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
</tbody>
</table>

C Secondary Mammosphere

<table>
<thead>
<tr>
<th></th>
<th>MDA-MB-231</th>
<th>SUM159PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap Ctrl</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
</tr>
<tr>
<td>Gap 2</td>
<td><img src="image15" alt="Image" /></td>
<td><img src="image16" alt="Image" /></td>
</tr>
<tr>
<td>Gap 3</td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
</tr>
</tbody>
</table>

D

<table>
<thead>
<tr>
<th></th>
<th>Gap Ctrl</th>
<th>Gap 2</th>
<th>Gap 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDA-MB-231</td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
<td><img src="image21" alt="Image" /></td>
</tr>
<tr>
<td>SUM159PT</td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
</tr>
</tbody>
</table>
Figure 3.3: LncRNA AK001796 contributes to maintaining stemness in EMT high breast cancer cells: A. Representative images of colony formation assay after silencing of AK001796 in MDA-MB-231 cells and SUM159PT. 5000 cells were plated in 6 wells plate after depletion of AK001796 using Gap 2 & 3. B. Quantification of primary mammospheres in MDA-MB-231 and SUM159PT. C. Quantification of secondary mammospheres in MDA-MB-231 and SUM159PT. The graph represent mean ± SD, n = 3, Student’s t test, *p < 0.05. D. Representative image of mammospheres in MDA-MB-231 and SUM159PT.

3.5 LncRNA AK001796 is negatively regulated by the transcription factor BHLHE40.

From the previous results we found that AK001796 is enriched under EMT conditions and affect the expression of some EMT markers. To examine the mechanism of its regulation in the context of EMT. Using the UCSC genome browser, the transcription factor BHLHE40 appears to be located on the promoter region of lncRNA AK001796 (Fig. 3.4A). BHLHE40 is a member of Basic Helix-Loop-Helix (bHLH) superfamily that possess two distinct domains; the DNA-binding domain and the helix-loop-helix domain that allow the transcription factor to form homo/heterodimeric complexes. BHLHE40 has other names like STRA13, because it is STimulated by Retinoic Acid (Boudjelal et al., 1997), SHARP2, comes from its close resemblance to Drosophila Split and HAiry Related Protein (Rossner, Dorr, Gass, Schwab, & Nave, 1997), and DEC1, which stand for Differentiation in Embryonic Chondrocytes (Shen et al., 1997). It is a transcriptional suppressor that bind directly to the E-box (5′-CACGTG-3′) and has been found to regulate different cellular processes including cellular differentiation, proliferation, apoptosis, and circadian rhythm (Ivanova, Liao, Lerman, Ivanov, & Stanbridge, 2005). We compared the expression of BHLHE40 between the normal mammary epithelial cells and the luminal breast cancer cells (MCF7 and ZR75-1) and EMT induced models and found that BHLHE40 expression is reduced on EMT induction (Fig. 3.4B & C). Accordingly, we knocked down the expression of
BHLHE40 in MCF10A (normal) and T47D (luminal) cells and found enhanced expression of AK001796 upon silencing of BHLHE40 (Fig.3.5A&B).

Figure 3.4: Induction of EMT is associated with reduced expression of BHLHE40: A. Genomic landscape of AK001796 showing the promoter binding site for BHLHE40. B. qRT-PCR
for BHLHE40 in HMLE and HMLE GFP, HMLE SNAI1, HMLE TWIST. C. qRT-PCR for BHLHE40 in MCF7 and ZR75-1 after induction of EMT with ZEB1.

Figure 3.5: LncRNA AK001796 is negatively regulated by the transcription factor BHLHE40: A. qRT-PCR analysis of AK001796, and BHLHE40 after BHLHE40 silencing with two different siRNAs in the normal breast epithelial cells MCF10A, and the luminal breast cancer cell lines T47D. B. Immunoblotting for BHLHE40 in both cell lines after knocking down BHLHE40 with siRNAs 1 and 2. GAPDH is used as internal control.
3.6 Discussion:

The higher expression of the lncRNA in the claudin low molecular subtype and the EMT induced models (Fig. 2.1B, 2.3B & 2.4A) indicates its association with the metastatic process. This is consistent with the previous studies on IncRNA AK001796 which found that its highly expressed with metastatic cancers (Zhao et al., 2019; Y. Zhao et al., 2018; Zhou et al., 2017).

Our results revealed the functional association of IncRNA AK001796 with the EMT process by its suppression of the epithelial markers E-cadherin and induction of the mesenchymal markers TWIST and vimentin (Fig 3.1 & 3.2). These findings were also observed by Fu et al. in gastric cancer (Y. Zhao et al., 2018). In addition to modulation of EMT markers, the suppression of AK001796 decreased the tumor initiating ability of EMT high breast cancer cells (Fig 3.3). From these finding, we can conclude that AK001796 may serve as a prognostic marker for more progressive disease.

The previous results also shows that the expression of AK001796 is regulated by the EMT associated transcription factor BHLHE40. The expression pattern of BHLHE40 is tumor and context specific. It has been found to be overexpressed in some tumors and downregulated in others (Ivanova et al., 2005; Zheng et al., 2009). BHLHE40 has been implicated with metastasis in several cancers. In human endometrial cancer, it was found to negatively regulate EMT program by inhibiting the expression of EMT-specific transcription factor TWIST1 (Asanoma et al., 2019; Shen et al., 1997). BHLHE40 is expressed at lower level in EMT induced models (Fig. 3.4B) and its suppression in the normal mammary cells MCF10A and the luminal breast cancer cells T47D led to increased AK001796 expression (Fig. 3.5A&B). As BHLHE40 functions as an inhibitory factor, these data suggest a mechanism by which AK001796 is induced under EMT conditions.
Chapter 4: Role of LncRNA AK001796 in evading apoptosis
4.1 Introduction:

Apoptosis is a cellular suicide or a programmed cell death that is necessary to get rid of unwanted or unhealthy cells during development or in response to stress. In normal cells, apoptosis is initiated by external or internal stimuli leading to activation of extrinsic or intrinsic pathways, respectively. Both pathways lead to activation of cysteine-dependent aspartate-directed proteases (caspases) starting with initiator caspases (caspase 2, 8, 9 and 10) which in turn activate the executioner caspases (caspase 3, 6 and 7) (Galluzzi, Lopez-Soto, Kumar, & Kroemer, 2016). Intrinsic pathway activation in response to cellular stress lead to mitochondrial outer membrane permeabilization (MOMP), that results in the release of pro-apoptotic factors such as cytochrome c from the mitochondrial intermembrane space (Fernald & Kurokawa, 2013). Then cytochrome c forms a complex or apoptosome with apoptotic protease activating factor 1 (APAF1) and the initiator caspase 9 leading to its activation that activates the executioner caspases. This process is regulated by BCL-2 family of proteins which include proapoptotic and antiapoptotic factors. Exposure to stress leads to activation of BH3 only proapoptotic proteins such as BIM (Bcl2-interacting mediator of cell death), BID (BH3-interacting domain death agonist) and BAD (Bcl2-associated agonist of cell death). Activation of these proteins cause oligomerization of multi-BH domains proteins such as BAK (Bcl2 antagonist/killer) and BAX (Bcl2 associated protein X) which lead to MOMP. The antiapoptotic members of BCL2-family of proteins such as BCL-2, BCL-XL (B-cell lymphoma extra-large), and MCL-1 (Induced myeloid leukemia cell differentiation protein) protect the cells from apoptosis by sequestering the proapoptotic proteins (Martinou & Youle, 2011; Strasser, Cory, & Adams, 2011). Although cancer cells are exposed to more cellular stresses than normal cells due to genomic instability and hypoxia, they adopt different mechanisms to resist apoptosis including suppression of BH3 only proapoptotic proteins.
and activation of antiapoptotic BCL-2 family members (Fernald & Kurokawa, 2013). In this study, we investigated the effect of AK001796 expression on the survival of cancer cells and its effect on the apoptotic response.

4.2 LncRNA AK001796 is essential for the survival of triple-negative breast cancer cell lines:

To evaluate the correlation between the expression of lncRNA AK001796 and the overall survival in all molecular subtypes of breast cancer we utilized KM-Plotter (Gyorffy et al., 2010). Partitioning patients by median expression, patients with high AK001796 expression have a significantly higher hazard ratio (2.22) and decreased overall survival in the patients exhibiting basal intrinsic subtype (p=0.023) (Fig. 4.1). To evaluate the significance of the lncRNA AK001796 for the survival of breast cancer cells, we knocked down its expression in cell lines representing all molecular subtypes of breast cancer. We transfected SUM159PT (CLOW/TNBC), MDA-MB-231 (CLOW/TNBC) SUM149PT (Basal/TNBC), ZR-75-1(luminal), T47D (luminal) (Prat et al., 2010) and MCF10A (normal) cell lines with two different GapmeRs targeting AK001796 (Gap 2 &3). Cellular viability was assessed 72 hours following GapmeRs transfection via the Cell Titer Glo (CTG) viability assay. Silencing of AK001796 significantly affected cell survival primarily in TNBC cell lines (Fig. 4.2). Interestingly, silencing of lncRNA in luminal (ZR-75-1, MCF7 & T47D) or normal cell line (MCF10A) had little effect on cellular viability, despite achieving knockdown (Fig. 4.2 &4.3A).

To investigate whether this reduction in cellular viability is due to reduced cellular proliferation or increased cell death, apoptosis was confirmed by staining with Annexin V followed by flow-cytometric measurement of cell death following AK001796 silencing. The extent of apoptosis
observed was concomitant with reduced expression level of lncRNA AK001796 in TNBC cell lines (MDA-MB-231, SUM159PT & SUM149PT) and the EMT induced model cell line (HMLE TWIST) (Fig. 4.3. A& B). Flow-cytometric analysis showed a significant amount of cell death in TNBC cell lines (Fig. 4.3.C & D) compared to luminal or normal cell lines (Fig 4.5).

**Figure 4.1**: AK001796 Kaplan-Meier plot was derived from KM-plotter (Gyorffy et al., 2010) using Affy id 232918_AT. Overall survival data is divided by median expression in the indicated subtypes of breast cancer.
LncRNA AK001796 is essential for the survival of triple-negative breast cancer cell lines

**Figure 4.2.** Relative cellular viability of the indicated breast cancer cell lines after silencing AK001796 using 50 nm GapmeRs (Gap 2 or Gap3). The upper panel represents the TNBC cell lines, while the middle and lower panel represents the normal mammary epithelial cells (MCF10A) and the luminal breast cancer cell lines (ZR-75-1, MCF7 & T47D). The graph represent mean ± SD, n = 3, Student’s t test, *p < 0.05.
AK001796 silencing is associated with increased apoptosis in EMT-high breast cancer cells
Figure 4.3. A. AK001796 expression in indicated breast cancer cells used for apoptotic assays. B. Annexin V apoptotic assay showing the apoptotic profiles in indicated breast cancer cells (MDA-MB-231 & SUM159PT) after AK001796 silencing with Gap2 and Gap 3. C & D. Graphical representation of three biological replicates of Annexin V experiments in MDA-MB-231 and SUM159PT cells lines respectively showing increase in late and total apoptosis upon AK001796 silencing.
Apoptosis was further validated by Caspase 3/7 staining followed by flowcytometry. The apoptotic profile was increased on AK001796 silencing in TNBC cell lines MDA-MB-231 and SUM159PT (Fig 4.4). Taken together these data demonstrate AK001796 expression to be critical for cell survival, specifically in TNBC.

**MDA-MB-231**

![Caspase 3/7 apoptotic assay in MDA-MB-231 cells](image)

**SUM159PT**

![Caspase 3/7 apoptotic assay in SUM159PT cells](image)

**Figure 4.** Caspase 3/7 apoptotic assay in MDA-MB-231 and SUM159PT cells that were transfected with 50nM GapmeRs C, 2, and 3 and incubated for 72 hours. These are representative images for three biological replicates.
From the expression analysis of AK001796 in different breast cancer cell lines (Fig 2.5), the normal and luminal cell lines show lower level of AK001796 expression. To investigate the effect of AK001796 silencing on cell death in these cell lines, apoptotic profiles was also analysed using annexin V staining assay following AK001796 knock down (Fig 4.5.A) in the luminal breast cancer cells T47D and ZR-75-1 and the normal mammary epithelial cells MCF10A (Fig 4.5.B). Interestingly, no significant changes in apoptotic profile were detected (Fig 4.5. C,D &E).

Figure 4. 5: AK001796 silencing has no significant effect on the apoptotic profile of the normal and luminal breast cancer cell lines: A. AK001796 expression in indicated breast cancer
cells used for apoptotic assays. B. Annexin V apoptotic assay showing the apoptotic profiles in the luminal breast cancer cells (T47D & ZR-75-1) and the normal mammary epithelial cells (MCF10A) after AK001796 silencing with Gap2 and Gap 3.
Figure 4. 5: C, D & E. Graphical representation for three biological replicates of Annexin V experiments in T47D, ZR-75-1 and MCF10A cells lines respectively showing no significant changes in apoptosis upon AK001796 silencing.

These results suggest that the level of expression of AK001796 is correlated with its importance in evading apoptosis in breast cancer cells. The higher expression of AK001796 in TNBC cells especially the EMT high cells, enhances their viability through evading apoptosis. While the low level of AK001796 expression in the normal or luminal breast cancer cells, has minimal effect on their survival.

Interestingly, a comprehensive analysis of the expression pattern of this lncRNA across all cancer types (n=6500) showed enhanced expression in tumor samples from multiple different cancers compared to their adjacent normal tissue (Fig. 4.6) (Iyer et al., 2015).
Figure 4.6 Comprehensive analysis of the expression pattern of AK001796 across all cancer types (n=6500) showed enhanced expression in tumor samples from different cancers compared to their adjacent normal tissue Mi Transcriptome (Iyer et al., 2015).

AK001796 plays a vital role not only in breast cancer, but also in other cancers including lung, gastric, esophageal, and hepatic carcinomas (X. Xu et al., 2019; Yang et al., 2015; Y. Zhao et al., 2018). The level of expression of AK001796 across the lung cancer cell lines from CCLE shows also higher level in those cells with mesenchymal like phenotype than those with epithelial phenotype (Fig 4. 7).
Figure 4.7: The level of expression of lncRNA AK001796 in lung cancer cell lines using BigQuery output from Cancer Cell Line Encyclopedia.

Mesenchymal

Epithelial
To investigate the involvement of AK001796 in evading apoptosis in lung cancer, the apoptotic profile for lung cancer cells (HCC4006, A549 and H1703) was analyzed using Annexin V staining method following AK001796 silencing (Fig 4. 7A & B). Consistent with our previous findings in breast cancer, silencing AK001796 was associated with an increase in apoptosis in lung cancer (Fig 4.8A).
Figure 4.8. A. Annexin V apoptotic assay showing the apoptotic profiles in indicated lung cancer cells (HCC4006, A459 & H1703) after AK001796 silencing with Gap2 and Gap 3. B. qRT-PCR for AK001796 expression in indicated lung cancer cells used for apoptotic assays.

4.3 LncRNA AK001796 negatively regulates pro-apoptotic BCL2L11 in TNBC cells:

Expression of AK001796 is closely tied to both EMT and aggressive breast cancer cells survival. Gene set enrichment analysis from TCGA shows the correlation between AK001796 expression and apoptosis related genes suggesting AK001796 may play a direct role in apoptotic signaling (Table 5 & Fig 4.9. A).

LncRNAs function through a wide variety of mechanisms to impact gene expression programs (Kumar et al., 2016). Nuclear lncRNAs are frequently cis acting. AK001796 is expressed from chromosome 2q13, directly downstream of BCL2L11 (Bim) (Fig 4.10. A). BCL2L11 is a pro-apoptotic BH3-only protein that facilitates Bax and Bak mediated mitochondrial release of cytochrome c and subsequent apoptosis (Gogada et al., 2013) (Willis et al., 2007). The co-expression analysis of the lincRNA AK001796 (named as MIR4435-2HG) and the BCL family of proteins using samples from breast invasive carcinoma on TCGA database (Pan cancer atlas) reveals a negative correlation with the pro-apoptotic gene BCL2L11 (Fig. 4.9. C). While the anti-apoptotic BCL family proteins such as BCL2L12 shows positive correlation (Fig 4.9.D). BCL2L12 inhibits caspase 3 and 7 and binds to P53 and prevents its association with target genes (Stehg & DePinho, 2011). Furthermore, RNA seq analysis following AK001796 silencing in SUM159PT cells shows a positive correlation with the anti-apoptotic protein BCL2 (Fig 4.9. B).
To determine if AK001796 plays a protective role for TNBC cell lines by negatively regulating the expression of BCL2L11, the level of expression of BCL2L11 was analyzed following AK001796 depletion. Indeed, silencing of AK001796 leads to induction of BCL2L11 transcription and protein accumulation in both MDA-MB-231 and SUM159PT cells (Fig 4. 10. B&C). Complementing the depletion studies, overexpression in the MCF7 cells attenuates BCL2L11 transcription (Fig 4. 12. A).

In a rescue experiment, silencing BCL2L11 after AK001796 was depleted in SUM159PT cells with Gap 3 resulted in a partial rescue of the apoptotic cell death (Fig 4.11). These results indicate that BCL2L11 is not the only apoptotic activator that mediates the anti-apoptotic effect of the lncRNA AK001796.

To determine if AK001796 may be directly regulating BCL2L11 expression we performed chromatin isolation by RNA purification (ChIRP-seq) and found AK001796 bound to the BCL2L11 gene (Fig 4.12. B). Taken together, these data suggest AK001796 plays a protective role for the TNBC cells through regulation of BCL2L11 gene expression and maybe other BCL2 family members as the CHIRP-seq results indicates.
Figure 4.9: AK001796 expression is correlated with apoptotic genes signatures: A. Gene Set Enrichment Analysis showed enrichment for the hallmark of Apoptosis positively correlated with expression of AK001796 in breast cancer. B. Relative expression of BLC2 following AK001796 silencing. C & D. Query output of cBioPortal, using Breast invasive carcinoma (TCGA Pan cancer atlas) for the co-expression of AK001796 (MIR4435-2HG) and BCL2L11 and BCL2L12.
Figure 4. 10: LncRNA AK001796 negatively regulates pro-apoptotic BCL2L11 in TNBC cells: 

A. Cartoon represents the genomic location of BCL2L11 according to LncRNA AK001796.

B. qRT-PCR results reflecting the level of expression of BCL2L11 after AK001796 depletion in MDA-MB-231 and SUM159PT cells. The graph represent mean ± SD, n = 3, Student’s t test, *p
< 0.05. C. Immunoblotting for BCL2L11 after AK001796 depletion in MDA-MB-231 and SUM159PT cells.

**Figure 4.11**: Silencing Bim partially rescue the apoptotic profile after AK001796 depletion: Caspase 3/7 apoptotic assay in SUM159PT cells that were transfected with 50nM GapmeRs C and 3 and incubated followed by siRNA for BCL2L11. These are representative images for three biological replicates.
Figure 4. 12: A. qRT-PCR for BCL2L11 in MCF7 cells overexpressing AK001796. B. ChIRP results: left panel represents the level of AK001796 pull down with indicated biotinylated antisense oligos, right panel represents the level of AK001796 enrichment at BCL2L11.
4.4 Discussion:

Detachment of the epithelial cells from the extracellular matrix (ECM) lead to activation of a form of apoptosis called anoikis (Paoli, Giannoni, & Chiarugi, 2013). The proapoptotic BH3 only members of BCL2 family of proteins such as Bid and Bim are activated upon detachment of the epithelial cell from the ECM and promote the dimerization of Bax and Bak at the outer mitochondrial membrane (OMM) (Taylor, Cullen, & Martin, 2008). Bim is the main activator of anoikis upon cellular detachment from ECM (Paoli et al., 2013).

Migratory neoplastic cells adopt different strategies to resist anoikis including constitutive activation of anti-apoptotic signals, suppression of apoptotic activator, induction of EMT (Paoli et al., 2013). The transcription factors that induce EMT have been found to activate the expression of anti-apoptotic factors such as Bcl2 which is activated by TWIST (Kwok et al., 2005). Non-coding RNAs such as miRNAs also have been found to participate in evading anoikis such as miR-221 and miR-222 (Stinson et al., 2011).

Our results shows that silencing of AK01796 leads to decrease in cellular viability and increase in apoptotic profile in breast and lung cancer cells. These results are consistent with the previous studies on lung, esophageal and hepatic carcinoma (Yang et al., 2015; Zhao et al., 2019; Y. Zhao et al., 2018). However, the association of AK001796 expression with the suppression of the apoptotic activator Bim has not been reported yet. In this study, we found that AK001796 silencing in the TNBC cells (SUM159PT & MDA-MB-231) leads to subsequent increase in Bim at mRNA and protein level and the opposite happen with overexpression. CHIRP results also revealed the enrichment of AK001796 on Bim indicating the presence of direct effect on Bim expression other than indirect effect as a consequence of activation of EMT process.
Chapter 5. Interaction of LncRNA AK001796 with Poly [ADP-ribose] polymerase 1 (PARP1)
5.1 Introduction:

The mechanisms of action of many LncRNAs were discovered by deciphering the interacting protein or proteins partners. The best example on that is HOTAIR, whose function is mediated through its interaction with PRC2 (mentioned in the first chapter) (Wu et al., 2014). Some LncRNAs once transcribed recruit proteins to nearby genes and regulate their transcription in cis (Wang et al., 2008). Others works as decoy for transcription factor blocking their interaction with target genes (Geisler & Coller, 2013).

Several methods have been developed to identify RNA-protein interactions that are either protein focused, or RNA focused depending on what molecule we want to investigate (Ferre, Colantoni, & Helmer-Citterich, 2016).

We sought to identify protein partners interacting with AK001796 to mechanistically elaborate the LncRNA function.

5.2 Identification of the interacting protein partner of AK001796:

To identify the protein/proteins complex associated with the LncRNA AK001796, biotinylated antisense probes AK001796 were used to pull down the LncRNA and any associated proteins. Pulls were performed (by Dr. Maneesh Kumar) in both MDA-MB-231 and SUM159PT cells. Associated proteins were separated by SDS-PAGE and stained with colloidal Coomassie blue. Unique bands were identified by mass-spectrometry. PARP-1 was the predominant protein found to be interacting with LncRNA AK001796. PARP1-specific antibodies are used for validation following the LncRNA isolation (Fig. 5.1.A). To complement the LncRNA pulldown, RNA Immunoprecipitation (RIP) assay was performed. PARP1 was immunoprecipitated and associated RNAs were purified and assessed by qRT-PCR. Several non-coding RNAs were also assessed as
a control for specificity. AK001796 is found specifically enriched with the PARP1 pulldown compared to IgG isotype control (Fig. 5.1.B).
Figure 5.1: LncRNA AK001796 interaction with Poly (ADP-ribose) Polymerase 1 (PARP1):

A. Immunoblotting for PARP1 in MDA-MB-231 or SUM159PT cells following AK001796 RNA pulls down with antisense oligonucleotides. B. RNA immunoprecipitation (RIP) assay using PARP1-specific antibody in SUM159PT cells. Associated RNA was evaluated by qRT-PCR for AK001796 and additional control RNAs.

5.3. AK001796 modulates enzymatic activity of PARP1:

PARP1 is a chromatin associated enzyme that catalyzes the transfer of ADP-ribose moieties from nicotinamide-adenine- dinucleotide (NAD) to variety of proteins forming linear or branched poly-ADP-ribose (PAR) chains in a process called poly ADP-ribosylation, also known as PARylation (Alemasova & Lavrik, 2019). This post translational modification alters protein-protein or protein-DNA interaction. The main target for parylation is PARP1 itself by an auto-modification reaction. PARP1 contains three structural domains: DNA binding domain (DBD), that consist of 3 zinc fingers (FI, FII and FIII), auto-modification domain (AD), that has the acceptor site for PAR moieties, and catalytic domain (CAT), responsible about PARylation (Altmeyer, Messner, Hassa, Fey, & Hottiger, 2009).

It was believed that PARP1 activity is stimulated mainly by DNA damage, where it recruits proteins involved in repair mechanisms. However, recent studies demonstrated that PARP1 activity is stimulated or inhibited by many proteins that also control the targets and the length of PAR chains (Alemasova & Lavrik, 2019). Furthermore, some lncRNAs have been found to interact with PARP1 and modulates its activity such as forkhead box D3 antisense RNA 1 (FOXD3-AS1), that suppresses neuroblastoma progression by suppressing parylation of CCCTC-
binding factor (CTCF) by PARP1 (X. Zhao et al., 2018). Understanding the regulation of PARP1 catalytic activity may help with PARP1-targeted therapy.

To determine if AK001796 functionally impacts PARP1, SUM159PT cells were depleted of the lncRNA and assessed for changes in parylation and level of PARP1. Parylation was measured 48 hrs post GapmeRs treatment as an early timepoint before depletion impacts cell viability. Depletion of AK001796 attenuates parylation as assessed by reporter assay (Fig 5.2. A) and suppressed the protein level of PARP1 (Fig 5.2.B). Furthermore, ectopic expression of AK001796 in MCF7 cells results in enhanced PARP1 expression at the protein level and activity as assessed by parylation accumulation (Fig.5.2. C). Taken together these data suggest AK001796 binds to and functionally supports PARP1 activity.
Figure 5.2: AK001796 expression affects the enzymatic activity of PARP1 (PARylation):

A. PARP1 activity assay in SUM159PT 48hrs post AK001796 silencing with Gap 2 and Gap 3. The graph represent mean ± SD, n = 3, Student’s t test, *p < 0.05. B. Immunoblotting for PARP1 in SU159PT cells after 72 hrs of AK001796 silencing with Gap2 and Gap3. C. Immunoblotting for PARP1 and PAR in MCF7 control cells and cells overexpressing AK001796.
5.4. PARP1 is involved in regulating EMT process:

The role of PARP1 in regulating EMT process is controversial. Schacke et al. support the pro-EMT effect of PARP1, where they found that induction of EMT by TGFβ increases PARylation which was opposed by PARPs inhibitors (Schacke et al., 2019). On the other hand, Pu et al. found that PARP1 has an anti-EMT effect as impairment of PARP1 activity in prostate tumorigenesis model led to induction of EMT (Pu et al., 2014). Another study by Stanisavljevic et al. found that PARP1 binds to the promoter of fibronectin in mesenchymal cells interacting with snail1 and p65 subunit of NF-KB inducing the expression of fibronectin (Stanisavljevic, Porta-de-la-Riva, Batlle, de Herreros, & Baulida, 2011). Furthermore, PARP1 has been found to regulate the transcription of Snail1 (McPhee, McDonald, Oloumi, & Dedhar, 2008) and increase its stability by post translational PARylation (Rodriguez et al., 2011).

Activation of EMT induced PARP1 expression at the mRNA level (Fig. 5.3.B) and the protein level (Fig. 5.3.A). On the other hand, silencing of PARP1 with siRNA in SUM159PT reversed the EMT phenotype mainly by suppressing the expression of Snail1 and upregulation of E-cadherin (Fig. 5.3.C).

Since AK001796 interacts with PARP1 and they both induced by EMT (Fig. 5.3.B) and regulate some EMT markers, they could work side by side to enhance survival of metastatic breast cancer cells. To answer this question, PARP1 chip in SUM159PT is used to look at common binding sites for PARP1 and AK001796. Following PARP1 chip, qRT-PCR is used to looked at some known target genes of PARP1 such as NELL2 (neural epidermal growth factor like repeats -like2) and ITPR1 (inositol 1,4,5-trisphosphate receptor type 1), which are positively regulated by PARP1 (Krishnakumar & Kraus, 2010). In addition to that, we looked at AK001796 binding site at BCL2L11 (Bim) from previous AK001796 chirp-seq. The results show more than eight folds
enrichment of PARP1 on Bim over IgG control (Fig 5.4. A). AK001796 chirp-seq also revealed that AK001796 binds to the promotor of ITPR1 and NELL2 genes. Over expression of AK001796 in MCF7 cells increased the expression of both genes (Fig. 5.4. B). These results suggest co-localization of PARP1 and AK001796 on some genomic loci to regulate genes expression.

**Figure 5.3: PARP1 play a role in regulating EMT process:** A. Immunoblotting for PARP1 and PAR in HMLE GFP and EMT induced models (TWIST, SNAI and TGFβ). B. qRT-PCR for PARP1 and AK001796 in HMLE and HMLE TWIST C. qRT-PCR for PARP1 and EMT markers
in SUM159PT 72hrs post PARP1 silencing with siRNA. The graphs represent mean ± SD, n = 3, Student’s t test, *p < 0.05.

**A** PARP1 CHIP in SUM159PT

![Graphs showing PARP1 CHIP in SUM159PT]

**Figure 5.4: Colocalization of PARP1 and AK001796:** A. qRT-PCR for PARP1 target genes (NELL2 and ITPR1) and Bim following PARP1 chip in SUM159PT cell line. The graphs represent
mean ± SD, n = 3, Student’s t test, *p < 0.05. B. qRT-PCR for NELL2 and ITPR1 in MCF7 with over expression of AK001796.

5.5 AK001796 expression contributes to resistance to PARP inhibitor Olaparib:

PARP inhibitors (PARPis) are recently approved by the Federal Drug Administration (FDA) for the treatment of BRCA1 or BRCA2 mutant ovarian cancers and metastatic breast cancers (O’Connor, 2015). The rationale for using PARP inhibitors for treatment of BRCA mutants’ cancers is the synthetic lethality that results from perturbation of PARP1 function in DNA damage repair. PARP1 get activated upon binding to single and double stranded breaks and start the assembly of multiprotein complexes that repair DNA damage (Ko & Ren, 2012). The clinical trials for PARPi revealed their efficacy not only in BRCA mutant cancers, but also in those cancers with BRCA like phenotype (BRCAness) that have abnormality in homologous recombination (HR) DNA repair mechanism such as mutation in ataxia telangiectasia mutated (ATM), that sense DNA damage and initiate HR repair (Turner, Tutt, & Ashworth, 2004; Wang & Weaver, 2011).

Most PARPi works as competitive analogues for NAD+ by binding to PARP1 catalytic subunit blocking PARP1 activity when it is bound to DNA. This will block the repair mechanism and cause subsequent cellular death.

Some patients exhibit intrinsic resistance to PARPis (Carey et al., 2018). Understanding the different mechanisms of resistance to PARPis will help to identify the conditions that restore sensitivity of cancer cells to PARPis and design better combination therapy to treat patients with PARPis more efficiently.

To investigate the effect of AK001796 expression on response to PARPis, the BRCA1 mutant TNBC cell line SUM149PT over expressing AK001796 are treated with PARP-1/2 inhibitor
Olaparib (OLA, Lynparza, AZD-2281) at different concentrations. Cellular viability is assessed using CTG assay after four days of treatment with Olaparib. The cells with AK001796 over expression show more resistance to treatment than control cells (Fig 5.5).

![Graph showing CTG assay results](image)

**Figure 5.5. AK001796 expression decrease the sensitivity to PARPi: A.** CTG assay reflecting relative viability of SUM149PT cells treated with Olaparib at 2.5, 5 and 10 uM concentrations for 4 days. **B.** Graph representation of the cellular viability of SUM149PT expressing AK001796 following treatment with Olaparib at 2.5 uM. The graphs represent mean ± SD, n = 3, Student’s t test, *p < 0.05.

**5.6 Discussion:**

PARP1 modifies a wide range of proteins including histones mainly H1 and H2B (Kraus, 2008). The highly negative poly-ADP-ribose (PAR) chains cause repulsion of histones from the DNA opening the chromatin and allow accessibility of multiple transcription factors and molecules involved in DNA repair (M. Y. Kim, Mauro, Gevry, Lis, & Kraus, 2004; Rouleau, Aubin, & Poirier, 2004). PARP1 keeps the chromatin in an open state by modifying histone demethylase lysine-specific histone demethylase 5B (KDM5B) preventing its binding to the trimethylated
histone H3K4 (Krishnakumar & Kraus, 2010). By these mechanisms PARP1 facilitate loading of RNA polymerase II on the promotors of actively transcribed genes (Krishnakumar & Kraus, 2010). In addition to that, PARylation changes the activity and localization of some transcription factors (Kanai et al., 2007) (Zerfaoui et al., 2010). PARP1 can also regulate gene expression through a non-enzymatic activity by binding directly to PARP1 specific motif on the promotors of target genes (Ko & Ren, 2012).

The level of PARP1 expression and its enzymatic activity usually increase in cancer (Shimizu et al., 2004). PARP-1 have been found to be upregulated in high grade breast cancers especially the TNBC, and to be associated with poor outcome (Rojo et al., 2012).

PARP1 regulates different cellular processes including transcription, chromatin modulation, DNA repair, alternative splicing, and cellular death by interacting and modifying many nuclear proteins. To elucidate the mechanism by which the lncRNA AK001796 mediates genes expression, we pulled down the interacting protein partner which happen to be PARP1. Induction of EMT in the normal mammary epithelial cells HMLE resulted in increase in expression of PARP1. The role of PARP1 in EMT is still controversial in different cancers. Silencing of PARP1 in SUM159PT cells led to reversal of EMT by downregulating Snail and upregulating CDH1. Our results showed that AK001796 expression is associated with increase in PARP1 activity and vice versa. To better understand the collaboration between PARP1 and AK001796, PARP1 chip is used to look at common binding genomic loci. Interestingly, the qRT-PCR analysis of PARP1 chip reveals common binding sites on the apoptotic activator Bim and some know PARP1 target genes. This suggests a possible mechanism by which AK001796 regulates genes expression by recruiting PARP1 to the target genes. Further investigations are required to explain this mechanism at the molecular level.
Chapter 6. Discussion and Conclusion
The dissemination of cancer cells from the site of origin to distal organs is a very complicated process, which makes targeting metastatic breast cancer one of the biggest challenges for clinician and scientists. As EMT process emerged as a pivotal driver for tumor dissemination, targeting molecules that induce EMT became of increasing therapeutic interests (Davis, Stewart, Thompson, & Monteith, 2014). Extensive studies have been conducted to identify transcription factors and signaling molecules that promotes EMT process. However, the role of IncRNAs in mediating EMT process is not well investigated.

In this study, we identified an EMT-associated IncRNA called AK001796, that has been found to be upregulated in poorly differentiated tumors and advanced stages diseases in lung, breast, esophageal, gastric and hepatocellular carcinomas (Yang et al., 2015; H. Zhang et al., 2016) (X. Xu et al., 2019; Y. Zhao et al., 2018; Zong et al., 2019). Our results show that the expression of AK001796 is higher in the claudin low molecular subtype of breast cancer patients’ derived samples than luminal and normal tissues. The claudin low molecular subtype is highly enriched with EMT markers, and characterized by its aggressive behavior, stemness potential and resistance to chemotherapy (Fedele, Cerchia, & Chiappetta, 2017; Prat et al., 2010). Furthermore, comparing the level of IncRNA AK001796 expression among an array of breast cancer cell lines revealed that AK001796 level is higher in the TNBC cell lines than normal and other breast cancer cell line. A finding that makes it more interesting since the TNBC including the claudin low subtype is associated with more aggressive disease and lack a targeted therapy (Aysola et al., 2013).

Induction of EMT in the mammary epithelial cells by different markers leads to upregulation of AK001796. This observation was validated in different luminal breast cancer cell lines induced to undergo EMT. Further, high expression of this IncRNA is associated with lower overall survival
rate in breast cancer patients. A result that is consistent with the previous studies in breast and esophageal carcinomas (H. Zhang et al., 2016) (Y. Zhao et al., 2018).

Our data shows that lncRNA AK001796 is not only upregulated under EMT conditions, but it also modulates the expression of some EMT markers. Silencing the lncRNA in EMT high cell lines led to an increase in the epithelial marker E-cadherin. Furthermore, overexpression of AK001796 in the EMT low luminal cell line MCF7 was associated with upregulation of the mesenchymal markers TWIST and vimentin and reduction in E-cadherin, a major whole mark for EMT. Further, we found this lncRNA itself to be negatively regulated by an EMT-associated transcription factor BHLHE40, that has been found to be down regulated during EMT in endometrial carcinoma (Asanoma et al., 2019). A loss-of-function study for BHLHE40 on several indolent breast cancer cell lines showed de-repression of the transcription of AK001796 in EMT-low cell lines.

EMT process not only increase the motility and invasiveness of metastatic cancer cells, but also help them to evade anoikis (Knudsen et al., 2012), a form of programmed cell death that is activated upon detachment from the extracellular matrix(Paoli et al., 2013). We found that lncRNA AK001796 protects metastatic breast cancer cells from anoikis by suppressing the pro-apoptotic gene BCL2L11 at transcriptional level. BCL2L11 or Bim, has been found to be suppressed in metastatic breast cancer cells (Merino et al., 2015; Woods, Yamaguchi, Lee, Bhalla, & Wang, 2007). Interestingly, lncRNA AK001796 closest neighbor is BCL2L11 located ~35 KB apart. Knocking down AK001796 led to a concomitant increase in the expression of BCL2L11 both at transcript and protein levels. We further demonstrated that after pulling down the chromatin regions associated with this lncRNA, showed its footprints on the loci for BCL2L11, indicating a direct evidence of lncRNA acting at transcriptional level.
LncRNAs regulate genes expression at the transcriptional level mainly by interacting with a protein partner. For example, lncRNA HOTAIR interact with PRC2 to regulate gene expression at epigenetic level (Taube et al., 2010). Our results show that AK001796 interacts with PARP-1 and modulates PARP-1 activity. Ectopic expression of AK001796 resulted in elevated parylation in EMT negative cells. Concomitantly, a reduction in the level of lncRNA AK001796 caused significant downregulation in PARP1 enzymatic activity underscoring the complicity of lncRNA AK001796 and PARP1 interaction. Further investigation revealed the co-localization of the lincRNA AK001796 and PARP1 on some genomic loci, which may explain the mechanism by which AK001796 regulate some genes expression to induce tumor growth and progression. PARPis were approved for the treatment of ovarian and breast cancers (O'Connor, 2015). As the case with other treatments, cancer cells acquired different mechanisms for resistance (Carey et al., 2018). The increase in the resistance to PARPi Olaparib in the TNBC cells SUM149PT overexpressing AK001796, indicates that AK001796 form a potential therapeutic target for these resistant cancers.
Figure 6: Model for the mechanism of action of the IncRNA AK001796:

AK001796 works by regulating the expression of certain genes that enhance survival and metastasis of the cancer cells probably through its interaction with PARP1. It enhances the expression of mesenchymal markers like TWIST and Vim and suppresses the expression of the epithelial marker CDH1 thus inducing EMT. It also suppresses the expression of the proapoptotic bcl2 family member Bim and induces expression of antiapoptotic protein bcl2 and by this mechanism help the metastatic cells to evade apoptosis. In addition to that it regulates PARP1 target genes ITPR1 and NELL2 which promote proliferation of the cancer cells.
Appendix:

Figure S1: The nuclear localization of AK001796 and its effect on E-cadherin expression:

A. qRT-PCR for AK001796 and two other known IncRNAs (HOTAIR & MALAT1) showing the nuclear/cytoplasmic ratios for those IncRNAs. 18s and GAPDH were used as controls.

B. Immunoblotting for E-cadherin in MDA-MB-231 cells after AK001796 silencing with Gap 2 and 3.
Figure S2: Query output of cBioPortal, using Breast invasive carcinoma (TCGA Firehose Legacy) for the co-expression of AK001796 (MIR4435-2HG) and A. TWIST1, B. SNAI1, C. VIM, D. CDH1.
Figure S3: Chromosome wide binding of IncRNA AK001796.
Figure S4: RNA-seq GO analysis for the common pathways associated with AK001796 expression.
References:


111


