CASE WESTERN RESERVE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Callie R Merry*

Candidate for the degree of Doctor of Philosophy

Committee Chair
Hung-Ying Kao

Committee Member
Eckhard Jankowsky

Committee Member
William Merrick

Committee Member
Helen Salz

Committee Member
Paul Tesar

Date of Defense
June 13, 2016

* We also certify that written approval has been obtained for any proprietary material contained therein.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>x</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xiii</td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Long non-coding RNAs</td>
<td>1</td>
</tr>
<tr>
<td><em>LncRNA categorization</em></td>
<td>2</td>
</tr>
<tr>
<td><em>LncRNA mechanisms</em></td>
<td>5</td>
</tr>
<tr>
<td><em>LncRNAs in development</em></td>
<td>8</td>
</tr>
<tr>
<td><em>LncRNAs in cancer</em></td>
<td>9</td>
</tr>
<tr>
<td><em>Summary</em></td>
<td>10</td>
</tr>
<tr>
<td>1.2 Colorectal cancer</td>
<td>11</td>
</tr>
<tr>
<td><em>Chromosome instability</em></td>
<td>13</td>
</tr>
<tr>
<td><em>DNA mismatch repair defects</em></td>
<td>14</td>
</tr>
<tr>
<td><em>CpG island methylator phenotype</em></td>
<td>14</td>
</tr>
<tr>
<td><em>DNA methylation</em></td>
<td>15</td>
</tr>
<tr>
<td><em>Summary</em></td>
<td>17</td>
</tr>
<tr>
<td>1.3 HER2-positive breast cancer</td>
<td>18</td>
</tr>
<tr>
<td>1.4 HER2-targeted therapies</td>
<td>22</td>
</tr>
<tr>
<td><em>Trastuzumab</em></td>
<td>22</td>
</tr>
<tr>
<td><em>Additional HER2-targeted therapies</em></td>
<td>25</td>
</tr>
</tbody>
</table>
1.5 Defining IncRNA expression in cancer ................................................................. 27

CHAPTER 2: DNMT1-ASSOCIATED LONG NON-CODING RNAs REGULATE
GLOBAL GENE EXPRESSION AND DNA METHYLATION IN COLON
CANCER .................................................................................................................. 28

2.1 Introduction ........................................................................................................ 31

2.2 Results ............................................................................................................... 34

Identification of DNMT1-associated IncRNAs in colon cancer cells .......... 34
DNMT1-associate IncRNA, DACOR1, is downregulated in colon cancer .... 37
DACOR1 affects DNA methylation levels at multiple sites in the human
genome .................................................................................................................. 40
DACOR1 may play a role in maintaining epithelial state of colon crypts ..... 41
DACOR1 induction reduces the clonogenic potential in colon cancer cells .. 43
DACOR1 induction affects global gene expression in colon cancer cells .... 44
DACOR1 interacts directly with chromatin at specific genomic sites ......... 48

2.3 Discussion ......................................................................................................... 50

2.4 Materials & Methods ....................................................................................... 54

CHAPTER 3: INTEGRATIVE TRANSCRIPTOME-WIDE ANALYSES REVEAL
CRITICAL HER2-REGULATED MRNAS AND LINCRNAS IN HER2+
BREAST CANCER .................................................................................................. 63

3.1 Introduction ..................................................................................................... 66

3.2 Results ............................................................................................................. 69

In vivo gene expression changes after inhibition of HER2 by trastuzumab 69
Gene expression changes in a HER2-positive breast cancer cell model after HER2 siRNA knockdown ................................................................. 73

Thousands of mRNAs and hundreds of lincRNAs are differentially expressed in HER2+ breast cancer tumors .................................................... 80

3.3 Discussion .................................................................................. 86

3.4 Materials and methods ............................................................... 90

CHAPTER 4: TRANSCRIPTOME-WIDE IDENTIFICATION OF MRNAS AND LINCARNAS ASSOCIATED WITH TRASTUZUMAB RESISTANCE IN HER2-POSITIVE BREAST CANCER .................................................................................. 96

4.1 Introduction .................................................................................. 99

4.2 Results .......................................................................................... 101

Identification of differentially expressed mRNAs and lincRNAs in trastuzumab-sensitive (TzS) vs. trastuzumab-resistant (TzR) human tumors in vivo .................................................................................................................. 101

Characterization of TzS and TzR HER2-positive cell culture model ......... 104

TzR is associated with transcriptome-wide changes in gene expression in cell culture model ............................................................................. 107

Inhibition of S100P reverses TzR in cell culture model .............................. 113

4.3 Discussion .................................................................................. 115

4.4 Materials and Methods ............................................................... 119

CHAPTER 5: DISCUSSION .................................................................. 124

5.1 Follow-up studies on DACOR1 ...................................................... 127

Mechanism and function of DACOR1 .................................................... 128
5.2 Investigation of HER2-regulated genes ................................................................. 132

Identified mRNAs ........................................................................................................ 133

Identified lincRNAs ...................................................................................................... 135

5.3 Understanding the role of S100P in trastuzumab resistance .................. 138

Functional role of S100P ............................................................................................. 139

Mechanistic role of S100P .......................................................................................... 141

Long-term studies ....................................................................................................... 144

Conclusion .................................................................................................................. 145

5.4 Conclusion .............................................................................................................. 145

APPENDIX A ................................................................................................................ 147

APPENDIX B ................................................................................................................ 160

BIBLIOGRAPHY .......................................................................................................... 162
LIST OF FIGURES

Figure 1-1. LncRNA categorization. ................................................................. 4

Figure 1-2. Chromatin remodeling mechanisms of lncRNAs. ......................... 7

Figure 1-3. Initiation of HER2 signaling. ........................................................ 21

Figure 2-1. Numerous long non-coding RNAs (lncRNAs) associate with DNMT1
in human cells. .............................................................................................. 36

Figure 2-2. The DNMT1-associated lncRNA, DACOR1, is downregulated in
colon cancer. ............................................................................................... 39

Figure 2-3. DACOR1 induction enhances DNA methylation and suppresses
growth of colon cancer cells. ........................................................................ 42

Figure 2-4. Induction of DACOR1 affects global gene expression in colon cancer
cells. ............................................................................................................ 47

Figure 2-5. DACOR1 is associated with chromatin. ........................................ 49

Figure 3-1. Inhibition of HER2 in vivo by trastuzumab results in significant
changes in the expression of mRNAs and lincRNAs in HER2-positive
breast cancer patients. .................................................................................. 72

Figure 3-2. HER2 affects the expression of mRNAs and lincRNAs in BT474 cells.
..................................................................................................................... 75

Figure 3-3. Validation of HER2-regulated lincRNAs in BT474 cells using a
second independent siRNA. ......................................................................... 77
Figure 3-4. Identification of commonly affected mRNAs and lncRNAs post-HER2 inhibition in tumors *in vivo* and post-HER2 knockdown in cell culture by siRNAs.  

Figure 3-5. Dysregulation of mRNAs and lincRNAs in HER2-positive tumors. ... 81  

Figure 3-6. Identification of 35 mRNAs that are affected in all three data sets with the expected directionality of expression. 83  

Figure 3-7. Validation of lincRNAs expression in 12 tumor vs. matched normal tissue (TCGA cohort). 85  

Figure 4-1. Identification of differentially expressed genes in trastuzumab-resistant (TzR) vs. trastuzumab-sensitive (TzS) tumors by RNA-seq. 103  

Figure 4-2. Characterization of HER2-positive trastuzumab-sensitive (TzS) and trastuzumab-resistant (TzR) BT474 breast cancer cell lines. 106  

Figure 4-3. Numerous mRNAs and lincRNAs are differentially expressed in HER2-positive TzS vs. TzR BT474 cell line. 108  

Figure 4-4. Top candidate mRNAs that were identified as associated with trastuzumab-resistant (TzR) both *in vivo* and in cell culture model. 110  

Figure 4-5. S100P is upregulated at the mRNA and protein levels in TzR cells. 112  

Figure 4-6. Knock down of S100P partially reverses TzR in cell culture. 114  

Figure 5-1. S100P alters phosphorylation of MAPK. 143
Figure A-1. Optimization of RNA co-immunoprecipitation (RIP) conditions in HCT116 cells ................................................................. 147

Figure A-2. DACOR1 (TCONS_00023265) genomic locus on human chromosome 15 ............................................................... 148

Figure A-3. SMAD3 mRNA expression in colon tumors vs matched normal tissues ........................................................................... 149

Figure A-4. Expression analysis of DACOR1 by qRT-PCR via cluster graph ... 150

Figure A-5. Expression analysis by qRT-PCR of DACOR1 induced cells by different promoters ......................................................... 151

Figure A-6. Western blot analysis of DNMT1 in transduced cell lines .......... 152

Figure A-7. E-cadherin levels after DACOR1 induction ............................. 153

Figure A-8. TJP1 expression by qRT-PCR. .............................................. 154

Figure A-9. DACOR1 induction results in decreased growth of colon cancer cells. ........................................................................ 155

Figure A-10. Induction of DACOR1 in additional lines ............................ 156

Figure A-11. DACOR1 induction with PGK promoter ................................ 157

Figure A-12. Induction of oncogenic IncRNA, TCONS_00011938 ............... 158

Figure A-13. SMAD6 and PHGDH are up-regulated in colon tumors ......... 159

Figure B-1. Cancer-associated pathways in response to HER2 knockdown in BT474 cells ................................................................. 160
Figure B-2. Identification of 35 mRNAs that are affected in all three data sets. 161
ACKNOWLEDGEMENTS

The path to obtaining my doctoral degree was possible with the help and support of many people. Firstly, I would like to thank the faculty, staff, and students of the Department of Biochemistry and Department of Genetics and Genome Sciences. I would also like to thank past and present members of my research lab including Ahmad Khalil, Victoria Moran, Sarah McMahon, Megan Forrest, and Jessica Sabers.

My committee has helped me immensely in my graduate career by providing suggestions and guidance in my research. Hung-Ying Kao, Eckhard Jankowsky, Paul Tesar, William Merrick, and Helen Salz have been imperative to my success in graduate school. Not only have they helped me to grow as a scientist and professional, but they have also provided me with enormous moral support as I negotiate graduate school.

Finally, I would like to thank my family and friends who have given me so much love and encouragement throughout my graduate career. In particular, my parents, Cheryl and Mike Woods and Steve and Vicki Merry, and my siblings have encouraged me to persevere in the achievement of my career goals while facing the challenges of graduate school. My boyfriend, Trevor Shaw, has been immensely supportive by being a constant optimist about my future. I thank you all for the love and support you have given to me.
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>ADCC</td>
<td>antibody-dependent cell cytotoxicity</td>
</tr>
<tr>
<td>CBS</td>
<td>cystathione β-synthase</td>
</tr>
<tr>
<td>CFA</td>
<td>colony formation assay</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>chromatin immunoprecipitation followed by sequencing</td>
</tr>
<tr>
<td>ChIRP-seq</td>
<td>chromatin isolation by RNA purification followed by sequencing</td>
</tr>
<tr>
<td>CIMP</td>
<td>CpG island methylator phenotype</td>
</tr>
<tr>
<td>Co-IP</td>
<td>co-immunoprecipitation</td>
</tr>
<tr>
<td>DACOR1</td>
<td>DNMT1-associated Colon Cancer Repressed IncRNA 1</td>
</tr>
<tr>
<td>DFS</td>
<td>disease-free survival</td>
</tr>
<tr>
<td>DMR</td>
<td>differentially methylated regions</td>
</tr>
<tr>
<td>DNMT1</td>
<td>DNA methyltransferase 1</td>
</tr>
<tr>
<td>EFS</td>
<td>event-free survival</td>
</tr>
<tr>
<td>ER</td>
<td>estrogen receptor</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FPKM</td>
<td>fragments per kilobase per million of mapped fragments</td>
</tr>
<tr>
<td>FST</td>
<td>follistatin</td>
</tr>
<tr>
<td>GR</td>
<td>glucocorticoid receptor</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>histone 3 lysine 27 acetylation</td>
</tr>
<tr>
<td>H3K4me1</td>
<td>histone 3 lysine 4 monomethylation</td>
</tr>
<tr>
<td>HER2+</td>
<td>HER2-positive</td>
</tr>
</tbody>
</table>
HMEC  human mammary epithelial cells
HNPPCC  Hereditary Nonpolposis Colorectal Cancer
INHBE  inhibin beta E
lincRNA  long intergenic non-coding RNA
IncRNA  long non-coding RNA
MBD  methyl binding domain
miRNA  microRNA
MMR  mismatch repair
NAT  natural antisense transcript
ncRNA  non-coding RNA
OS  overall survival
PFS  progression-free survival
PKM2  pyruvate kinase, muscle
p-MAPK  phosphorylationed MAPK
PR  progesterone receptor
qRT-PCR  quantitative real-time polymerase chain reaction
RIN  RNA integrity number
RIP-qPCR  RNA immunoprecipitation followed by qPCR
RIP-seq  RNA immunoprecipitation followed by sequencing
RNA-seq  RNA-sequencing
RNP  ribonucleoprotein
RRBS  reduced representation bisulfite sequencing
rRNA  ribosomal RNA
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>shRNA</td>
<td>short hairpin RNA</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosyl methionine</td>
</tr>
<tr>
<td>TCGA</td>
<td>The Cancer Genome Atlas</td>
</tr>
<tr>
<td>T-DM1</td>
<td>ado-trastuzumab emtansine</td>
</tr>
<tr>
<td>TJP1</td>
<td>Tight Junction Protein 1</td>
</tr>
<tr>
<td>TPP</td>
<td>time to progression</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TzR</td>
<td>trastuzumab-resistant</td>
</tr>
<tr>
<td>TzS</td>
<td>trastuzumab-sensitive</td>
</tr>
<tr>
<td>XIST</td>
<td>X inactive specific transcript</td>
</tr>
</tbody>
</table>
Long-Noncoding RNAs in Cancer

Abstract

by

CALLIE R. MERRY

The study of the genetic mutations in cancer has been a major research focus for decades. While identifying genetic mutations have allowed for advancements in the diagnosis and treatment of cancers, we can further enhance our understanding of cancer by assessing gene expression changes in the cell. In particular, the last decade has shown the emergence of the study of long non-coding (lncRNAs) in disease. While there is still much to learn about lncRNAs, many have shown functional significance through roles in regulating the epigenome. My study focuses on identifying the role of lncRNAs in colon and HER2-positive breast cancer. In my first study, I identified a lncRNA, DACOR1, that interacts with DNA methyltransferase, DNMT1, a colon cancer. The results of my study demonstrate that deregulation of DACOR1 contributes to aberrant DNA methylation and gene expression during colon tumorigenesis. In my second
study, I identified IncRNAs and mRNAs that are regulated by HER2 in HER2-positive breast cancer. These transcripts can be functional studied to further our understanding of how HER2 amplification results in tumorigenesis. Finally, I identified IncRNAs and mRNAs that become dysregulated due to the development of resistance to trastuzumab, the primary targeted therapy for HER2-positive breast cancer. Our studies suggest that the upregulation of $S100P$, an mRNA, may contribute to trastuzumab resistance in HER2-positive breast cancer. As a whole, my thesis illustrates the expression level changes of IncRNAs and mRNAs in colon and breast cancer. The focused analysis of candidate transcripts demonstrates that the changes in expression that occur can further contribute to tumorigenesis. My identification of the transcripts with altered expression in cancer have revealed new avenues to advance cancer research, as well as to advance the understanding of the role of IncRNAs in the cell.
Chapter 1: Introduction

The maintenance of proper cellular growth and function is necessary for organismal health, but genetic mutations can disrupt the ability of cells to perform essential functions for normal growth. Certain mutations allow for the development of cancer through activation of oncogenes or inhibition of tumor suppressors. These genetic changes lead to the development and advancement of cancer through the disruption of gene function and alteration of gene expression levels. In my thesis work, I am interested in defining the altered expression levels of long non-coding RNA (lncRNAs) in cancer and how these changes contribute to tumorigenesis. While lncRNAs have been identified through the use of microarray and high-throughput sequencing, researchers are still unclear on the function of these lncRNAs in cancer and the consequences of altered expression of lncRNAs on cancer tumorigenesis. My research demonstrates, in three separate studies, that lncRNAs are dysregulated in cancers and that modification of the expression level of key genes can lead to a reversal of cancer phenotypes. In the following sections, I will review what is currently known about lncRNAs and their role in cancer in order to set the stage for my studies.

1.1 Long non-coding RNA

While historically we know the importance of non-coding RNAs (ncRNAs) such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), the study of more
recently discovered classes of RNA have only recently gained momentum. One such class of ncRNAs is long non-coding RNAs (lncRNAs) that were identified on a large scale after the development of high-throughput RNA sequencing technology (Birney et al., 2007). Similarly to protein-coding genes, lncRNAs are RNA polymerase II transcripts that are 200 base pairs or greater in length and are processed through capping, splicing, and polyadenylation. However, unlike protein-coding genes, lncRNAs are poorly conserved between species (Pang et al., 2006).

**LncRNA categorization**

Long non-coding RNAs are separated into four categories based on genomic location (Figure 1-1), as not enough is known about their role in the cell to classify based on mechanism or function. Natural antisense transcripts (NATs) are transcribed on the antisense strand of a protein-coding gene and overlap at least one exon. Intronic lncRNAs are transcribed from within a single intron of a protein-coding gene. Bidirectional lncRNA transcription is initiated from the same promoter as a protein-coding gene but is transcribed from the antisense strand with no overlap in gene sequence. The final category is intervening lncRNAs (lincRNAs), which are transcribed from genomic regions inbetween protein-coding regions. LincRNAs are at least 5kb from a protein-coding gene as dictated by identification method (Guttman et al., 2009; Khalil et al., 2009; Merry et al., 2015a; Rinn and Chang, 2012). In total, GENCODE has annotated nearly 16,000
IncRNAs as of August 2015 (Derrien et al., 2012). While many IncRNAs have been identified, only a handful of them have been functionally characterized.
Figure 1-1. LncRNA categorization. Natural antisense transcripts (NATs) are transcribed on the antisense strand of a protein-coding gene. Intronic lncRNAs are transcribed from within a single intron of a protein-coding gene. Bidirectional lncRNAs are transcribed from the same promoter as a protein-coding gene but on the antisense strand. Intervening lncRNAs (lincRNAs) are transcribed from genomic regions in between protein-coding regions. Boxes indicate exons of genes. The arrow indicated transcriptional direction.
**LncRNA mechanisms**

The majority of lncRNAs are localized to the nucleus where they regulate gene expression through a variety of chromatin remodeling mechanisms such as decoy, guide, scaffold, and enhancer mechanisms (Figure 1-2) (Devaux et al., 2015; Rinn and Chang, 2012). I will briefly describe these mechanisms and then I will revisit examples of each of these mechanisms when I discuss lncRNAs that are associated with cancer. In the decoy mechanism, lncRNAs interact with transcription factors to block their binding to the DNA, thereby, inhibiting target gene expression. Guide lncRNAs recruit chromatin-modifying enzymes and localize the enzymes to the chromatin. These lncRNAs must have the capability to not only bind specialized proteins, but also interact with specific regions of DNA. The scaffolding mechanism works by bringing together proteins into a functional complex. Finally, enhancer lncRNAs work in a cis manner to initiate gene expression through chromosomal looping of the enhancer and promoter regions.

In addition to epigenetic mechanisms, lncRNAs can also act to regulate genes post-transcriptionally. About 15% of lncRNAs localize to the cytoplasm where they have been shown to regulate stability and translation of transcripts (Kapranov et al., 2007). Similar to the nuclear decoy mechanism that acts to block transcription factor from binding to the genome, lncRNAs can also act as decoys to titrate miRNAs. By functioning as a miRNA decoy, the degradation of target mRNA is prevented. Cytoplasmic lncRNAs also regulate translation through complementary pairing with a target sequence to either stabilize the
mRNA or degrade the transcript through STAU1-mediated mRNA decay (Faghihi et al., 2008; Gong and Maquat, 2011; Kretz et al., 2013; Wang et al., 2013).

While biochemical studies have already identified a variety of IncRNAs mechanisms, this does not preclude the discovery of additional mechanisms as we learn more about the role of IncRNAs.
Figure 1-2. Chromatin remodeling mechanisms of lncRNAs. Decoy lncRNAs interact with transcription factors to block binding to the DNA, inhibiting target gene expression. Guide lncRNAs recruit chromatin-modifying enzymes to localize to the chromatin. Scaffold lncRNAs work by bringing together proteins into a functional complex. Enhancer lncRNAs work in a cis manner to initiate gene expression through chromosomal looping of the enhancer and promoter regions. Blue circles indicate protein, black shape indicates lncRNA, and grey helix indicates DNA.
LncRNAs in development

Before large-scale discovery methods such as tiling arrays and RNA sequencing were used, only a few lncRNAs were identified in humans – such as XIST and H19, both of which are important for embryonic development. The identification and mechanistic understanding of these genes illustrated the importance of continued study of lncRNAs as major contributors to cellular function. XIST was identified in 1992 as the major effector of X-chromosome inactivation, which acts to silence one of the two X chromosomes in placental females for dosage compensation (Brockdorff et al., 1992; Brown et al., 1992). XIST works with other lncRNAs and protein factors to ensure proper silencing of the inactive X chromosome. The mechanism by which XIST exerts its effects on the chromosome is initiated with the expression of XIST from the inactive X chromosome. The XIST transcript then coats the inactive X chromosome and recruits transcriptionally repressive protein factors to silence gene expression (Lee, 2011). H19 is an imprinted gene that is expressed from the maternal allele and has a role in cell proliferation, although the mechanism is still not known. Increased expression of H19 has been found in many cancers including lung, breast, hepatocellular, and ovarian cancers, to name a few. (Brannan et al., 1990; Gabory et al., 2010; Ripoche et al., 1997). These functional characterizations show the importance of lncRNAs in the regulation of genomic expression and the need to further identify and characterize lncRNAs to advance the understanding of the genome.
LncRNAs in cancer

Cancer studies have identified numerous IncRNAs that are dysregulated in tumor tissue compared to normal. I have chosen to discuss 3 IncRNAs that have been shown to contribute to tumorigenesis through different nuclear mechanism of action (Figure 1-2) to illustrate the function of IncRNAs in cancer. GAS5 represents the decoy mechanism; HOTAIR works through both the guide and scaffold mechanisms; and CCAT1-L1 exerts its function through the enhancer mechanism (Gupta et al., 2010; Kino et al., 2010; Xiang et al., 2014).

GAS5 has been shown to be downregulated in breast cancer. Functional studies have indicated that normal expression of GAS5 acts to sensitize cells to apoptosis through inhibition of the glucocorticoid receptor (GR), which acts as a transcription factor (Kino et al., 2010; Mourtada-Maarabouni et al., 2009). GAS5 works as a decoy by interacting with GR through the DNA-binding domain, impeding the binding of GR to DNA (Kino et al., 2010). The decoy mechanism of GAS5 allows it to alter the expression of GR-directed gene expression.

HOTAIR is overexpressed in liver, metastatic breast, lung, and pancreatic tumors (Gupta et al., 2010; Li et al., 2013). HOTAIR is an example of both guide and scaffold IncRNA mechanisms. This IncRNA scaffolds PRC2 and LSD1 together into a complex and then localizes this complex to specific regions of the genome. This complex acts to silence the genomic region through repressive chromatin marks: PRC2 methylates H3K27 and LSD1 demethylates H3K4me2 (Tsai et al., 2010). PRC2 has been shown to associate with 20% of lincRNAs indicating that the scaffold and guide mechanisms are likely to be a common
mechanism by which IncRNAs regulate gene expression in trans (Khalil et al., 2009).

The upregulation of CCAT1-L1 leads to increased MYC expression in colorectal cancer (Xiang et al., 2014). CCAT1-L1 is transcribed specifically in colorectal cancer and acts in long-range chromatin looping between enhancers and proto-oncogene, MYC (Xiang et al., 2014). These are just a few examples of IncRNAs that are dysregulated in cancer, but they illustrate the diverse mechanisms by which they can contribute to the cancer phenotype.

Summary

Currently, we know that a large number of IncRNAs are transcribed from the genome, but only a small portion have been characterized to date. Those IncRNAs that have been identified as having a role in cancer, including HOTAIR, GAS5, and CCAT1-L1, demonstrate the importance of studying the function of IncRNAs as their altered expression levels may further contribute to the aberrant signaling in cancer cells. With the knowledge of the potential impact that IncRNAs could have in cancer, I wanted to begin my thesis work by defining the IncRNAs that may have a role in colorectal cancer. Previous studies identified that IncRNA HOTAIR interacts with PRC2 and LSD1 to enact repressive histone modifications in cancer metastasis (Tsai et al., 2010). Colorectal cancer is characterized by global hypomethylation of the DNA in combination with promoter hypermethylation. While we do not understand the mechanism by which these methylation changes occur, we do know that DNA
methyltransferase, DNMT1, is responsible for the maintenance of DNA methylation in normal cells. With the understanding that lncRNAs can act to guide enzymes to enact chromatin modifications, I wanted to identify lncRNAs that may work by a similar mechanism with DNMT1 to regulate DNA methylation in colon cancer.

1.2 Colorectal cancer

In the United States, there is an estimated 132,000 cases of colorectal cancer diagnosed each year. Nearly 50,000 deaths from colorectal cancer were estimated for 2015, making colorectal cancer the third leading cause of cancer deaths in both men and women (American Cancer Society, 2015). As with most cancers, colorectal cancer has heterogeneous pathogenesis with risk factors including heredity, lifestyle factors, and somatic mutations. Countries in Europe, North America, and Oceania have the highest incidence rate, while countries in south and central Africa and Asia have the lowest (Center et al., 2009). As evidenced by the different rates of colorectal cancer in various countries, there are contributing environmental factors, but those are not well defined. Increased age and being male are the biggest risk factors for colorectal cancer. Medically, family history, suffering from inflammatory bowel disease, and diabetes increase the risk of developing colorectal cancer. Lifestyle and diet contributions to the disease suggest that smoking, obesity, excessive alcohol consumption, and diet high in red and processed meats also increase the risk for colorectal cancer, but the data on these are not consistent (Brenner et al., 2014; Fearon, 2011).
Colorectal tumorigenesis originates from the epithelial lining of the large intestine and rectum and begins as benign polyps or adenomas, which develop over years or even decades to form tumors (Fearon and Vogelstein, 1990; Jass, 2007). This process is termed the adenoma-carcinoma sequence. Recent studies have suggested that the causative tumor cells originate from stem cells that are found at the base of colonic crypts, the finger-like invaginations of the epithelial cells (Huels and Sansom, 2015; Khalek et al., 2010). If a mutation arises in a stem cell, not only is this cell long-lived, giving the cell a chance to accumulate a ‘second hit’, it also has the ability to expand these mutations through production of differentiated cells. Mutant cells proceed to fill the whole crypt (monoclonal conversion), and then undergo fission or crypt division, which leads to the development of adenomas (Humphries and Wright, 2008).

Histopathologically, colorectal cancer is classified based on the TNM system that identifies tumor stage, lymph node involvement, and metastasis (American Joint Committee on Cancer, 2009). While this classification system is very useful, molecular pathogenesis provides a better clinical tool for prognosis and treatment (Sadanandam et al., 2013; De Sousa E Melo et al., 2013). There are three major genomic instabilities in colorectal cancer as I will discuss below: chromosome instability, DNA mismatch repair defects, and the hypermethylation of CpG islands.
Chromosome instability

Chromosome instability is found in 80-85% of sporadic colorectal cancers (Markowitz and Bertagnolli, 2009). These cancers, most commonly, develop from APC gene defects leading to the adenoma-carcinoma sequence (Kinzler and Vogelstein, 1996). Subsequent activating mutations of the KRAS oncogene and inactivating mutations of the tumor suppressor TP53 lead to tumorigenesis and is accompanied by chromosomal instability (Fearon, 2011; Lengauer et al., 1997). The hereditable syndrome, familial adenomatous polyposis coli (FAP), accounts for 1% of all colorectal cancer cases and develops as a results of a germ line mutations in the APC gene (Galiatsatos and Foulkes, 2006).

The most well studied role of APC is a tumor suppressor function that negatively regulates β-catenin in the canonical Wnt signaling pathway (Aoki and Taketo, 2007). The Wnt signaling pathway regulates cell proliferation and cell migration in the context of colon cancer. Mutant APC proteins allow for constitutively active Wnt signaling without the presence of the Wnt protein. The activation of this pathway alters signaling in the early stages of colorectal cancer by increasing cell proliferation and allowing early adenomas to develop (Fearon, 2011). As additional mutations accumulate, such as mutations that inactivate the p53 and TGF-β pathways, and mutations that activate MAPK and PI3K pathways, the disease state advances to colorectal cancer (Baker et al., 1990; Grady et al., 1999).
DNA mismatch repair defects

Hereditable colorectal cancer known as hereditary nonpolyposis colon cancer (HNPCC) or Lynch syndrome accounts for 2-4% of colorectal cancer cases (Aaltonen et al., 1998; Hampel et al., 2005; Vilar and Gruber, 2010). DNA sequencing of patients with HNPCC allowed for the identification of microsatellite changes across the genome, which is now known as high-grade microsatellite instability (Markowitz and Bertagnolli, 2009). First identified in mutant yeast strains and then validated through HNPCC patient sequencing, the underlying cause of high-grade microsatellite instability are defects in mismatch repair (MMR) genes (Fishel and Kolodner, 1995; Fishel et al., 1993). The most common mutations, which account for 70% of HNPCC patients, are in the MSH2 and MLH1 genes (Fearon, 2011; Fishel et al., 1993; Leach et al., 1993). Additional mutations have been identified in PMS1, PMS2, and GTBP/MSH6 (Fishel and Kolodner, 1995; Vilar and Gruber, 2010). The mutation of MMR genes, and subsequent loss of functional MMR proteins, allows for increased microsatellite instability as the genome cannot be repaired when polymerase slippage occurs during DNA replication (Fearon, 2011; Parsons et al., 1993).

CpG island methylator phenotype

The final molecular instability that occurs is CpG island methylator phenotype (CIMP), which occurs in 15% of sporadic colorectal cancers (Markowitz and Bertagnolli, 2009; Nosho et al., 2008). CIMP patients have the high-grade microsatellite instability phenotype, but there are no underlying
mutations to the MMR genes. Instead, patients have a loss of $MLH1$ gene expression due to hypermethylation of CpG islands in the gene promoter (Issa, 2004; Weisenberger et al., 2006). CpG islands are regions of high-density CG dinucleotides most often found in gene promoters. In normal cells, most CG dinucleotides within the genome are methylated, but CpG islands tend to be unmethylated. While in colorectal cancer there is global depletion of DNA methylation, the high level of aberrant methylation CpG islands in gene promoter leads to epigenetic silencing of gene expression (Issa, 2004).

**DNA methylation**

DNA methylation is an epigenetic factor that contributes to colorectal cancer tumorigenesis. DNA methylation is a necessary process in cells required to silence unneeded genomic regions as seen in X-chromosome inactivation, genomic imprinting, and silencing of transposable elements. The presence of methylated CpG works to repress transcription first by physically blocking the binding of transcription factors (Jones and Baylin, 2007), and secondly, by recruiting methyl-CpG-binding domain (MBD) proteins. The methylated CpGs recruit MBDs, which then act to recruit HDAC and other chromatin remodeling factors to establish heterochromatin (Zhang et al., 1999). The methylation of CpG in DNA is achieved through three known DNA methyltransferases (DNMTs): DNMT1, DNMT3A, and DNMT3B. Studies suggest that DNMT1 works to maintain methylation in adult cells by copying methylation patterns from the template strand of DNA to the newly replicated strand (Rhee et al., 2000). While
DNMT3A and DNMT3B are thought to establish \textit{de novo} methylation patterns that are maintained by DNMT1 (Okano et al., 1999).

When errors arise in DNA methylation, aberrant activation and silencing of genes can contribute to cancer pathogenesis. Alterations in DNA methylation in cancer are associated with global hypomethylation and hypermethylation of CpG islands in promoter regions (Tsai and Baylin, 2011). Global hypomethylation leads to chromosome instability, genetic mutation, and expression of oncogenic and imprinted genes (Cadieux et al., 2006; Cui et al., 2002). Hypomethylation and hypermethylation are often seen in different regions of the genome in the same tumor. Researchers have suggested that there could be an epigenetic repair mechanism that allows crosstalk between demethylation and \textit{de novo} methylation pathways. If such a mechanism exists, the hypermethylation of a gene promoter could initiate an over-compensatory hypomethylation event, or vice versa (Ehrlich, 2009).

Many cancers display global hypomethylation when compared with normal tissue including ovarian, prostrate, hepatocellular, cervical, and colon cancers (Bedford and van Helden, 1987; Ehrlich et al., 2006; Feinberg et al., 1988; Kim et al., 1994; Lin et al., 2001). While mutations in \textit{DNMT1} and \textit{DNMT3a} have been identified in cancers, the majority of hypomethylation are not associated with these mutations (Kanwal and Gupta, 2012; Ley et al., 2010; Yan et al., 2011). The main mechanism by which the genome becomes hypomethylated is still unclear. It could occur through passive or active demethylation, but most likely through a combination of both. Passive demethylation occurs when there is a
failure of DNMT1 to maintain methylation during replication. While there is no
known mechanism for active demethylation in mammals, methylation analysis in
normal and tumor tissues suggest that it does occur (Ehrlich, 2009).

Hypermethylation of a gene can contribute one hit to Knudson’s two hit
hypothesis, which states that two ‘hits’ to a gene are necessary to cause cancer
(Knudson, 1971). A defect to each allele of a gene accumulated through either
heredity or sporadically are needed for tumor development (Jones and Baylin,
2002; Shen and Laird, 2013). As mentioned previously, the hypermethylation of
the MLH1 gene promoter in colorectal cancer leads to microsatellite instability.
Other examples of hypermethylation of gene promoters that contribute to cancer
are TIMP3 (an inhibitor of Wnt signaling), RASSF1A (tumor suppressor
associated with Ras signaling), and MGMT (DNA repair) (Bachman et al., 1999;
Burbee et al., 2001; Hegi et al., 2008). These genes represent examples in which
hypermethylation can contribute one or more of the ‘hits’ leading to the
development of cancer. On the whole, the proper control of DNA methylation is
important for precise gene expression.

Summary

The goal of my first study is to identify lncRNAs that interact with DNMT1
to provide insight into the aberrant DNA methylation that is seen in colorectal
cancer, as shown in Chapter 2. The basis of this study comes from the
knowledge that colon cancer has global hypomethylation of the genome and
promoter hypermethylation that further contributes to changes in gene
expression. Additionally, IncRNAs have been shown to bind to chromatin-modifying proteins to exert changes to chromatin state. In summary, determining the interaction between IncRNAs and DNMT1 will provide the basis for continued study of IncRNAs in colon cancer.

By utilizing on the colon cancer system, I can identify IncRNAs that associate with DNMT1 to pursue an understanding of the methylation mechanism. Not all cancers have such a precise epigenetic mechanism on which to focus, but IncRNAs may still have unknown importance to tumorigenesis. For the next part of my thesis work, I wanted to identify IncRNAs that are regulated through the HER2 receptor in HER2-positive breast cancer. HER2-positive is one of the most aggressive subtypes of breast cancer. While much work has been done to identify therapies that target HER2, we do not have a full understanding of the gene expression changes that occur with overexpression of HER2. My work to determine the changes in gene expression due to HER2, in particular IncRNA gene expression changes, will allow us to more accurately study HER2-positive breast cancer in the future.

1.3 HER2-positive breast cancer

Breast cancer is a major health concern in the United States and around the world. In the US, 1 in 8 women will develop invasive breast cancer in their lifetime, with over 200,000 newly diagnosed cases each year (American Cancer Society, 2015). While the breast cancer research field has advanced treatment and detection in the last several decades, there are still 40,000 deaths due to
breast cancer each year in the US, which is the second highest cancer death rate in women after lung cancer (American Cancer Society, 2015; Siegel et al., 2012). These statistics characterize breast cancer as a single disease when, in fact, breast cancer is a heterogeneous disease that is comprised of various subtypes that are defined through the presence of receptors such as the estrogen receptor (ER), progesterone receptor (PR), and HER2. Identifying subtypes of breast cancer based through receptor status has improved the treatment of breast cancer tumors by enabling the use of receptor-targeted therapies. Due to the heterogeneous nature of breast cancer, I focused on one subtype, HER2-positive, to identify gene expression differences that may contribute to disease.

HER2-positive breast cancer accounts for 20-30% of breast cancer diagnoses and is associated with reduced time to progression (TTP) and reduced overall survival (OS) (Slamon et al., 1987, 1989). This subtype of breast cancer is characterized by the overexpression of the HER2 receptor due to an amplification event on chromosome 17 (Fukushige et al., 1986). HER2 (ERBB2 or HER2/neu) is a 185 kDa transmembrane receptor tyrosine kinase within the human epidermal growth factor family of receptors, also known as ERBB family (Stern et al., 1986). In addition to HER2, the ERBB family is made up of HER1 (ERBB1/EGFR), HER3 (ERBB3), and HER4 (ERBB4). In general, the ERBB signaling mechanism begins with a ligand binding the extracellular domain of an ERBB receptor, this induces a conformational change allowing for homo- or hetero-dimerization of the receptors. The dimerization leads to autophosphorylation of the tyrosine residue on the intracellular domain and
initiation of downstream signaling (Figure 1-3) (Yarden and Sliwkowski, 2001). Unlike the other three ERBB receptors, HER2 does not bind any known ligand but is in an open conformation to allow for dimerization (Klapper et al., 1999). HER3 is also unique in that while it binds ligands, it does not have its own kinase activity (Guy et al., 1994). The activation of HER2 initiates signaling cascades through the PI3K/AKT and MAPK pathways leading to proliferation, cell survival, and invasiveness (Arteaga and Engelman, 2014). Previous studies have identified gene expression changes through the utilization of mRNA microarrays (Le et al., 2005), but these are not exhaustive and do not focus on interrogating non-coding RNAs. My work to identify gene expression changes through the use of RNA-sequencing, will allow an unbiased identification of mRNAs and IncRNAs for further study.
Figure 1-3. Initiation of HER2 signaling. The four members of the ERBB family of receptors are shown integrated into the plasma membrane with the extracellular, transmembrane, and intracellular domains. Dimerization is shown between HER1 and HER2. HER1 is ligand-bound, inducing a conformational change to allow dimerization with constitutively active HER2. Dimerization induces autophosphorylation of the tyrosine kinase, which activates MAPK and PI3K signaling pathways to promote proliferation, cell survival, and invasiveness.
1.4 HER2-targeted therapies

As HER2-positive breast cancer is characterized by the overexpression of HER2, therapies targeting this receptor have been developed and implemented in the clinical setting. Unfortunately, the use of these therapies leads to resistance in which the targeted therapy no longer allows for clinical benefit. In my third study, my goal is to identify mRNAs and lncRNAs that are dysregulated in drug resistant samples in order to inform the field on additional treatment targets for patients displaying resistance to the standard treatments. In the following subsections, I will describe the currently available HER2-targeted therapeutics, as well as the issue of targeted therapy resistance in the treatment of HER2-positive breast cancer to illustrate the need to find alternative therapeutic targets through gene expression studies.

Trastuzumab

While chemotherapy is used as the primary treatment for HER2-positive breast cancer, there are various adjuvant treatments that were developed to target HER2 directly. In 1998, a recombinant monoclonal antibody, trastuzumab (Hercpetin®), was introduced that targets the extracellular domain of HER2 and blocks the initiation of downstream signaling (Baselga et al., 1998). Initially used for treatment of metastatic HER2-positive breast cancer, trastuzumab was approved for use as adjuvant treatment in early stage HER2-positive breast cancer in 2006 (Figueroa-Magalhães et al., 2013). A study of early-stage HER2-positive breast cancer showed that when trastuzumab was used in combination
with chemotherapy there was reduced recurrence and increased overall survival (OS) when compared to chemotherapy alone (Buzdar, 2005; Perez et al., 2011). Clinical studies have observed cardiotoxic side effects, but these can be mitigated by with the use of less cardiotoxic chemotherapy agents (Hudis, 2007; Slamon et al., 2001).

While the mechanism by which trastuzumab inhibits HER2 signaling is not fully understood, several models have been proposed. Commonly, it is suggested that trastuzumab prevents the dimerization of HER2 with other ERBB receptors, thereby, blocking downstream signaling (Juntila et al., 2009). Additionally, trastuzumab may function by preventing the shedding of the extracellular domain of HER2. When the extracellular domain is cleaved, the remaining protein (p95HER2), composed of the transmembrane domain and intracellular domain, is constitutively active (Molina et al., 2001). The third potential mechanism suggests that as an antibody, trastuzumab activates antibody dependent cell-mediated cytotoxicity (ADCC) (Gennari et al., 2004). ADCC is an action by which immune cells, most often natural killer cells, lyse antibody-flagged cells (Clynes et al., 2000). While data has been presented on each of these mechanisms it is unclear if trastuzumab works primarily through one or multiple mechanisms to inhibit HER2 signaling.

While trastuzumab has been shown to be an effective adjuvant treatment for early-stage HER2-positive breast cancer, resistance to trastuzumab has been shown to occur both through innate or acquired resistance. One study showed that 30% of early-stage and 70% of metastatic HER2-positive tumors have innate
resistance to trastuzumab (Oliveras-ferraros et al., 2010). In addition, most patients develop resistance to trastuzumab within one year of treatment (Nahta, 2012).

Just as the mechanism of action for trastuzumab inhibition of HER2 signaling is not fully understood, neither is the mechanism of trastuzumab resistance. There are four main proposed mechanisms for trastuzumab resistance in breast cancer. The first set of mechanisms focuses on the impaired binding of trastuzumab to HER2. The first example of this mechanism is HER2 epitope masking by MUC4, a membrane-associated glycoprotein, which serves to protect the epithelial cells (Nady et al., 2005). In addition, the shedding of the extracellular domain of HER2 can give rise to the constitutively active, truncated p95HER2 form of the protein (Scaltriti et al., 2007). The second mechanism suggests that changes in the downstream signaling pathway of HER2 could lead to resistance. Such examples of this mechanism include a hyperactivating mutation in \textit{PIK3CA}, or reduced expression of \textit{PTEN} (Berns et al., 2007). Additionally, increased signaling by alternative pathways, such as HER1 or ER, could counteract the inhibition of HER2 by trastuzumab (Nahta, 2012). The final proposed mechanism of resistance is the malfunction of ADCC (Clynes et al., 2000; Parihar et al., 2004; Repka et al., 2003). To combat trastuzumab resistance, other treatments targeting the HER2 receptor were developed. Three are currently approved for clinical use: lapatinib, pertuzumab, and T-DM1.
Additional HER2-targeted therapies

After the introduction of trastuzumab, additional treatments targeting the HER2 receptor began to be tested. The first was lapatinib, a tyrosine kinase inhibitor of both HER1 and HER2 (Burris, 2004). It was approved for use in HER2-positive metastatic breast cancer in 2007 (Drakaki and Hurvitz, 2015; Geyer et al., 2006). Further clinical trials have shown that in both the metastatic and early-stage setting, trastuzumab plus chemotherapy outperforms lapatinib plus chemotherapy as measured by progression-free survival (PFS) and pathologic complete response (pCR – no histological evidence of invasive cancer), respectively (Gelmon et al., 2015; Untch et al., 2012). When lapatinib was evaluated for use as a dual therapy with trastuzumab, data showed that trastuzumab plus lapatinib improved PFS and OS in metastatic breast cancer when compared to lapatinib treatment alone (Blackwell et al., 2010, 2012). Early-stage dual therapy treatment of chemotherapy in combination with trastuzumab, lapatinib, or both showed improved pCR for patients treated with trastuzumab and lapatinib compared to those treated with either separately (Guarneri et al., 2012). Although the short-term studies of trastuzumab and lapatinib dual therapy in early-stage breast cancer setting show improved pCR, the long-term benefits - such as OS, event-free survival (EFS), and disease-free survival (DFS) -- of lapatinib were not supported (de Azambuja et al., 2014; Robidoux et al., 2013).

Pertuzumab is an additionaly immunotherapy that blocks HER2 signaling by binding to the extracellular domain of HER2 to prevent dimerization with HER3 (Franklin et al., 2004). While pertuzumab and trastuzumab both bind to the
extracellular domain of HER2, they bind at different locations, thereby changing the mechanism of inhibition (Cho et al., 2003). In 2012, the FDA approved the use of pertuzumab in combination with chemotherapy and trastuzumab to treat metastatic breast cancer patients that have had no prior HER2 targeted therapies (Baselga et al., 2012; Swain et al., 2013). Additionally, pertuzumab was approved in 2013 for treatment of early-stage HER2-positive breast cancer in the adjuvant setting in combination with trastuzumab and chemotherapy (Gianni et al., 2012; Schneeweiss et al., 2013). An ongoing study (NCT01358877) is being conducted to test the effectiveness of trastuzumab and chemotherapy with and without pertuzumab (Drakaki and Hurvitz, 2015).

The final approved HER2-targeted therapy is a conjugated trastuzumab antibody, ado-trastuzumab emtansine (T-DM1). The DM1 conjugate is a microtubule inhibitor. When the antibody binds to the HER2 receptor, receptor-mediated endocytosis delivers DM1 into the cytoplasm where the inhibitory effects on microtubule assembly cause cell death (Lewis Phillips et al., 2008). T-DM1 was FDA approved in 2013 with data showing that T-DM1 had improved OS in metastatic cancers compared to the control group (Verma et al., 2012). Further studies are underway to evaluate T-DM1 in combination with other HER2-targeted therapies, as well as studies in early-stage HER2-positive breast cancer (Drakaki and Hurvitz, 2015).

The clinical studies that have been completed on these three additional HER2-targeted therapies illustrate that while these therapies are useful in combination with trastuzumab, they cannot replace trastuzumab therapy.
Additionally relevant is the fact that resistance still limits the success of these HER2-targeted therapies. While we do not definitively know the main mechanism/s by which HER2-positive breast cancer becomes resistant to trastuzumab, we can reason that if the mechanism lies outside of the HER2 receptor that additional treatments targeting that receptor will not be beneficial. As I will elaborate in Chapter 4, our data suggests that trastuzumab resistance occurs outside of the HER2 receptor. With this in mind, the most effective dual therapy treatments should target downstream effectors of the HER2 pathway. Through the identification of mRNAs and IncRNAs that are dysregulated in trastuzumab-resistant HER2-positive breast cancer, we can inform the field on possibly additional targets for therapeutic treatment.

1.5 Defining IncRNA expression in cancer

Currently, we know that there is a wide range of mechanisms by which IncRNAs exert their function in cells. In particular, researchers have identified and studied IncRNAs in cancer, and this has begun to illustrate the capability of IncRNAs in regulating gene expression in the cell. With the expanding knowledge that IncRNAs can be important molecules in regulating gene expression, it is essential to continue to identify and study IncRNAs in cancer. In the following chapters, I set out to define the altered expression of IncRNAs in colon and HER2-positive breast cancer and how these changes may contribute to tumorigenesis.
Chapter 2:

DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer

Callie R. Merry\textsuperscript{1,2}, Megan E. Forrest\textsuperscript{1}, Jessica N. Sabers\textsuperscript{1}, Lydia Beard\textsuperscript{3}, Xing-Huang Gao\textsuperscript{1}, Maria Hatzoglou\textsuperscript{1}, Mark W. Jackson\textsuperscript{3,4}, Zhenghe Wang\textsuperscript{1,3}, Sanford D. Markowitz\textsuperscript{1,3}, Ahmad M. Khalil\textsuperscript{1,2,3,*}

Affiliations: \textsuperscript{1}Department of Genetics and Genome Sciences, \textsuperscript{2}Department of Biochemistry, \textsuperscript{3}Case Comprehensive Cancer Center, \textsuperscript{4}Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

Reference:

Abstract

The cancer epigenome exhibits global loss of DNA methylation, which contributes to genomic instability and aberrant gene expression by mechanisms that are yet to be fully elucidated. We previously discovered over 3,300 long non-coding (Inc)RNAs in human cells and demonstrated that specific IncRNAs regulate gene expression via interactions with chromatin-modifying complexes. Here we tested if IncRNAs could also associate with DNA methyltransferases to regulate DNA methylation and gene expression. Using RIP-seq, we identified a subset of IncRNAs that interacts with the DNA methyltransferase DNMT1 in a colon cancer cell line, HCT116. One IncRNA, TCONS_00023265, which we named DACOR1 (DNMT1-associated Colon Cancer Repressed IncRNA 1), shows high, tissue-specific expression in the normal colon (including colon crypts) but was repressed in a panel of colon tumors and patient-derived colon cancer cell lines. We identified the genomic occupancy sites of DACOR1, which we found to significantly overlap with known differentially methylated regions (DMRs) in colon tumors. Induction of DACOR1 in colon cancer cell lines significantly reduced their ability to form colonies in vitro, suggesting a growth suppressor function. Consistent with the observed phenotype, induction of DACOR1 led to the activation of tumor-suppressor pathways and attenuation of cancer-associated metabolic pathways. Notably, DACOR1 induction resulted in down-regulation of Cystathionine β-synthase (CBS), which is known to lead to increased levels of S-adenosyl methionine (SAM) – the key methyl donor for DNA methylation. Collectively, our results demonstrate that deregulation of
DNMT1-associated IncRNAs contribute to aberrant DNA methylation and gene expression during colon tumorigenesis.
2.1 Introduction

In humans and other mammals, epigenetic modifications play critical roles in orchestrating gene expression patterns in distinct cell types throughout the life of an organism. These epigenetic modifications are regulated by the highly coordinated functions of chromatin-remodeling complexes, histone-modifying enzymes, DNA methyltransferases, and chromatin readers (Chen and Dent, 2014). However, the process by which these ubiquitous epigenetic modifiers are recruited, assembled, and stabilized at specific genomic loci in distinct cell types has yet to be fully elucidated. Increasing experimental evidence suggests potential key roles for long non-coding RNAs in recruiting and scaffolding such complexes to the genome in mammalian cells (Bertani et al., 2011; Dimitrova et al., 2014; Grote et al., 2013; Khalil et al., 2009; Klattenhoff et al., 2013; Kotake et al., 2011; Nagano et al., 2008; Pandey et al., 2008; Tsai et al., 2010; Wang et al., 2011; Yu et al., 2008).

The mammalian genome encodes thousands of long non-coding RNAs (lncRNAs) (Cabili et al., 2011; Derrien et al., 2012; Engstrom et al., 2006; Khalil et al., 2009; Okazaki et al., 2002) and studies of select lncRNAs have clearly demonstrated their critical roles in various aspects of mammalian biology (Bertani et al., 2011; Brockdorff et al., 1991; Brown et al., 1991; Carrieri et al., 2012; Ishii et al., 2006; Lanz et al., 1999; Sleutels et al., 2002; Sunwoo et al., 2009). For example, some lncRNAs are critical for embryonic development and tissue morphogenesis in vivo, as demonstrated by genetic disruptions of several lncRNAs in mice and zebrafish (Grote et al., 2013; Sauvageau et al., 2013;
Ulitsky et al., 2011). The means by which lncRNAs exert their effects have been demonstrated through several mechanisms: (i) transcription co-activators (Lanz et al., 1999); (ii) guidance of epigenetic complexes to the genome (Bertani et al., 2011; Grote et al., 2013; Klattenhoff et al., 2013; Kotake et al., 2011; Tsai et al., 2010); (iii) competing endogenous RNAs or microRNA “sponges” (Cesana et al., 2011; Johnsson et al., 2013; Poliseno et al., 2010); (iv) regulation of mRNA translation and decay (Geisler et al., 2012; Gong and Maquat, 2011; Johnsson et al., 2013); (v) decoys for transcription factors (TFs) (Kino et al., 2010); and other potential mechanisms (Morris and Mattick, 2014).

Previous studies have identified extensive genome-wide interactions between lncRNAs and several chromatin-modifying complexes, including the polycomb repressive complex 2 (PRC2) (Khalil et al., 2009; Zhao et al., 2010). These interactions are required for proper PRC2-mediated gene repression (Khalil et al., 2009; Klattenhoff et al., 2013), suggesting that deregulation of such lncRNAs could impact the ability of epigenetic-modifying complexes to regulate the epigenome and gene expression programs. Indeed, a number of studies have identified numerous lncRNAs that become highly deregulated in various human diseases, including several cancer types (Hou et al., 2014; Iyer et al., 2015; Merry et al., 2015b; Morton et al., 2014; Yildirim et al., 2013). Thus, deregulation of specific lncRNAs may result in global changes in the epigenome, including global changes in DNA methylation patterns, and deregulation of gene expression (Gu et al., 2010; Luo et al., 2014; Simmer et al., 2012; Ziller et al., 2013).
DNA methylation is an important epigenetic mark that is typically associated with repressed genes in mammalian cells (Wu and Zhang, 2014). Three distinct DNA methyltransferases (DNMT1, DNMT3a and DNMT3b) are known to regulate DNA methylation patterns in mammals. Genome-wide studies of DNA methylation in various tumors vs. matched normal tissues have demonstrated that although the promoters of some tumor suppressor genes become hypermethylated, cancer genomes are largely hypomethylated (Ehrlich, 2009; Wu and Zhang, 2014). Currently, there is a great interest in understanding how DNA methylation patterns become deregulated in human cancers, with the hope that these studies might lead to novel insights into tumorigenesis as well as the development of novel therapeutic strategies (Sharma et al., 2009). We hypothesized that a subset of lncRNAs may interact with DNA methyltransferases, and, consequently, affect their genomic occupancies and/or activities. Thus, deregulation of such lncRNAs in human cancers would result in abnormal DNA methylation patterns without any detectable mutations or changes in the expression levels of the genes that encode DNA methyltransferases.

In this manuscript, we identified specific interactions between a subset of human lncRNAs and the DNA methyltransferase DNMT1 using RNA co-immunoprecipitation followed by next-generation RNA sequencing (RIP-seq) (Figure 2-1 A). Analysis of one such lncRNA, TCONS_00023265, which we named DACOR1, revealed a critical role of this lncRNA in regulating DNA methylation and gene expression in colon cells. Furthermore, induction of DACOR1 is sufficient to suppress the growth of colon cancer cells by regulating
the expression of specific genes and pathways including cellular metabolism. Our results suggest a potential new mechanism by which the human methylome is regulated in human health and disease.

2.2 Results

*Identification of DNMT1-associated lncRNAs in colon cancer cells*

We optimized our RNA co-immunoprecipitation (RIP) protocol in the colon cancer cell line HCT116 ([Appendix A-Figure A-1](#)), and subsequently utilized it to identify potential interactions between DNMT1 and RNAs. Since there are no reliable DNMT1 antibodies that are suitable for RIP applications, we utilized a knock-in DNMT1_3X-flag HCT116 cell line to overcome this limitation (Zhang et al., 2008). First, we confirmed that DNMT1 is specifically immunoprecipitated, but not other abundant nuclear proteins such as U1-70K or histone H3 ([Figure 2-B](#)). To identify RNAs that potentially interact with DNMT1, we performed triplicate RIPs of DNMT1 using an anti-flag antibody and triplicate RIPs using an anti-IgG antibody as negative controls. We isolated co-immunoprecipitated RNAs, and quantified the small amount of DNMT1-bound RNAs. We were able to generate RNA-seq libraries from DNMT1 RIPs but not from IgG RIPs due to depletion of non-specific RNAs by several stringent washes.

Three RNA-seq libraries from three independent biological replicates of DNMT1 RIPs were sequenced and mapped to the human genome (hg19). We also sequenced total nuclear RNA (input) from HCT116 cells as a control for our RIP experiments. We generated FPKM (fragments per kilobase of exon per
million fragments mapped) values for mRNAs and IncRNAs detected in the input sample and each of the three biological replicates of DNMT1 RIP-seq. The average FPKM of each transcript in the three biological replicates of DNMT1 RIP-seq was divided by the FPKM in the input sample to generate fold changes. We identified 148 IncRNAs (14% of IncRNAs detected in the input) and 31 mRNAs (0.009% of mRNAs detected in the input) as DNMT1-associated RNAs based on a 2-fold change or higher above input (Figure 2-1 C-F). We found the highest fold change of a IncRNA associated with DNMT1 to be ~41-fold, while the highest fold change for a mRNA was only 7-fold, despite mRNAs being expressed at much higher levels than IncRNAs across all cell types (Cabili et al., 2011; Derrien et al., 2012; Iyer et al., 2015; Khalil et al., 2009). To rule out non-specific co-immunoprecipitation of highly abundant RNAs with DNMT1, we compared the expression of all DNMT1-bound vs DNMT1-unbound IncRNAs and mRNAs. We found that there was no expression bias of DNMT1-associated IncRNAs or mRNAs in comparison to unbound IncRNAs and mRNAs (Figure 2-1 G-H). Lastly, a close examination of DNMT1-associated mRNAs revealed that at least half of these mRNAs are poorly annotated transcripts with predicted open reading frames or miRNA precursors, suggesting that very few mRNAs associate with DNMT1. In summary, we have identified many IncRNAs and very small number of mRNAs that co-immunoprecipitate with DNMT1 in HCT116 cells by RIP-seq.
Figure 2-1. Numerous long non-coding RNAs (lncRNAs) associate with DNMT1 in human cells. (A) Outline of the experimental strategy utilized to identify DNMT1-associated RNAs. (B) Western blot analysis using an anti-flag-DNMT1 antibody confirms the specific immunoprecipitation (IP) of DNMT1, but not other highly abundant nuclear proteins (histone H3, U1-70K). An IP with anti-IgG antibody demonstrates that there is no detectable background. (C) Heatmap of lncRNAs in input sample vs. each of the three biological replicates of DNMT1 RIPs. We observed some variability between the three biological replicates due to stringent washes to eliminate non-specific RNAs. (D) Heatmap of mRNAs in input vs. the three biological replicates of DNMT1 RIPs. (E) Pie chart showing the number of DNMT1-associated lncRNAs vs. all lncRNAs expressed in input, approximately 14% of lncRNAs co-IP with DNMT1. (F) Pie chart showing the number of DNMT1-associated mRNAs vs. mRNAs expressed in input, approximately 0.009% of mRNAs co-IP with DNMT1. (G-H) Graphs show the expression levels of DNMT1-bound lncRNAs and mRNAs vs. non-bound lncRNAs and mRNAs in HCT116 cells. We found that DNMT1-bound lncRNAs and mRNAs show no expression bias over non-bound lncRNAs and mRNAs.
The DNMT1-associated IncRNA, DACOR1, is downregulated in colon cancer

One DNMT1-associated IncRNA, designated TCONS_00023265 (Appendix A-Fig A-2), was of interest to us due to its notable high, tissue-specific expression in normal colon tissues (Figure 2-2 A-B) and repression in colon tumors and patient-derived colon cancer cell lines (Figure 2-2 C-D). We therefore named this IncRNA DNMT1-associated Colon Cancer Repressed IncRNA 1 (DACOR1). In a panel of twelve human normal tissues, DACOR1 shows the highest expression in the colon as measured by qRT-PCR (Figure 2-2 A). We confirmed the expression of DACOR1 in the normal colon by RNA in situ hybridization, and observed DACOR1 expression in the nuclei of colon crypts, the cells from which colon cancer originates (Figure 2-2 B, large panel). We also observed that DACOR1 occupies several discrete foci in the nucleus (Figure 2-2 B, small panel). Next, we examined DACOR1 expression in a cohort of 22 colon tumors in comparison to matched normal tissue based on RNA-seq data obtained from The Cancer Genome Atlas (TCGA). This analysis revealed that DACOR1 is downregulated in colon tumors (Figure 2-2 C). We also examined the expression of the protein-coding gene SMAD3, the nearest coding gene to DACOR1, in the same TCGA cohort and found that SMAD3 shows variable expression in tumors vs. normal colon (Appendix A-Figure A-3). To further confirm that DACOR1 is downregulated in colon cancer, we examined its expression by qRT-PCR in eight normal colon samples and twenty-one patient-derived colon cancer cell lines with limited passage in culture (Figure 2-2 D and Appendix A-Figure A-4). Several of the colon cancer cell lines displayed
very low expression levels of *DACOR1* that were barely detectable by qRT-PCR, further confirming the downregulation of *DACOR1* during colon tumorigenesis (Figure 2-2 D). These intriguing observations prompted us to further investigate the potential role of *DACOR1* in colon cancer biology and its effects on DNA methylation and gene expression.
Figure 2.2. The DNMT1-associated lincRNA, DACOR1, is downregulated in colon cancer. (A) Quantitative real time PCR (qRT-PCR) of DACOR1 across a panel of human normal tissues demonstrates the high expression of DACOR1 in the colon and small intestine in comparison to other tissues examined. (B) RNA in situ hybridization confirms the expression of DACOR1 in human colon tissues and identifies the colon crypts as one of the major cell types that express it. Red arrows mark DACOR1 signal, which is shown as a chromogenic signal (brown). Close examination of colon cells (small panel) reveals that DACOR1 is retained in the nucleus and potentially interacts with chromatin. (C) Expression analysis of DACOR1 in a cohort of 22 colon cancer tumors vs 22 matched normal tissues in RNA-seq data sets obtained from TCGA demonstrates that DACOR1 is downregulated in colon tumors. (D) Examining the expression of DACOR1 by qRT-PCR in eight normal colon samples and twenty-one patient-derived colon cancer cell lines with limited passage in culture demonstrates that DACOR1 is highly repressed in in most colon cancer cells.
**DACOR1 affects DNA methylation levels at multiple sites in the human genome**

To determine the functional significance of DACOR1 association with DNMT1, we first validated the interaction of DACOR1 with DNMT1 in independent RIP experiments using RIP-qPCR (Figure 2-3 A). As a negative control, we examined the association of the highly abundant nuclear RNA U1 with DNMT1 and found no association (Figure 2-3 B). We then tested the effects of DACOR1 induction on DNA methylation in two distinct patient-derived colon cancer cell lines, V481 and V852. We transduced V481 and V852 cells with either a control or DACOR1 lentivirus, and confirmed that DACOR1 had the appropriate expression by qRT-PCR (Appendix A-Figure A-5). We isolated genomic DNA from these cell lines and equal amounts of DNA (1 μg) from each sample (n = 12) were used for DNA methylation analysis using 450K DNA methylation arrays (Illumina). These arrays cover approximately 500 thousand CpG sites out of the 28 million CpG sites in the human genome. We identified 43 and 59 specific CpG sites in V481 and V852, respectively, that become differentially methylated in response to DACOR1 expression (Figure 2-3 C). Of these sites, 42/43 (in V481) and 58/59 (in V852) displayed a gain of DNA methylation (p < 1 x 10^{-11} and 2.1 x 10^{-16}, respectively). Next, we determined if restoration of DACOR1 expression affected DNMT1 protein levels. We performed western blot analyses using a DNMT1 antibody in cells transduced with a control or DACOR1 lentivirus, and found that DNMT1 protein levels were unchanged (Appendix A-Figure A-6). In summary, DACOR1 induction appears
to enhance DNA methylation at multiple loci without affecting DNMT1 protein levels.

*DACOR1 may play a role in maintaining the epithelial state of colon crypts*

The high expression of *DACOR1* in normal colon tissues and the localization of *DACOR1* to colon crypts prompted us to examine its potential role in regulating the epithelial state of colon cells. To that end, we examined the effects of *DACOR1* induction on the levels of key epithelial markers including Tight Junction Protein 1 (TJP1) and E-cadherin in two distinct colon cancer cell lines. We found that expression of *DACOR1* led to higher levels of TJP1 protein, but not E-Cadherin ([Figure 2-3 D](#) and [Appendix A-Figure A-7](#)). To determine if the change in TJP1 is at the transcriptional or post-transcriptional level, we measured TJP1 mRNA levels by qRT-PCR in three distinct colon cancer cell lines. We found *DACOR1* expression to have no effect on TJP1 mRNA levels ([Appendix A-Figure A-8](#)) suggesting that TJP1 protein levels are regulated post-transcriptionally by *DACOR1* in colon cells. We also compared TJP1 mRNA levels in a cohort of 22 colon tumors vs 22 matched normal tissue from TCGA and found that TJP1 mRNAs levels are not significantly affected in most patients ([Appendix A-Figure A-8](#)), suggesting that TJP1 protein levels are regulated post-transcriptionally in colon tumors.
**Figure 2-3.** DACOR1 induction enhances DNA methylation and suppresses growth of colon cancer cells. (A) Validation of the interaction between DNMT1 and DACOR1 by RIP-qPCR. DACOR1 shows a 7-fold enrichment in flag-DNMT1 RIP over IgG RIP. (B) The highly abundant nuclear RNA U1 shows no enrichment in flag-DNMT1 RIP vs IgG RIP, demonstrating the specificity of our method. (C) Induction of DACOR1 expression in two distinct patient-derived colon cancer cell lines (V481 and V852) enhances DNA methylation at multiple genomic loci in trans. Blue color indicates low methylation, red color indicates high methylation. All experiments were performed in triplicates. (D) Induction of DACOR1 expression in patient-derived colon cancer cell lines (V866 and V852) results in upregulation of tight junction protein 1 (TJP1), suggesting a potential role for DACOR1 in maintaining an epithelial state of colon cells. (E) The colon cancer cell lines V481, V852, and V866 were transduced with either a control or DACOR1 lentivirus. Subsequently, colony formation assays were carried out (see methods) in each cell line. Cells with restored DACOR1 expression showed reduced colony formation, suggesting that DACOR1 is potentially a growth suppressor.
DACOR1 induction reduces the clonogenic potential of colon cancer cells

Our studies demonstrated that DACOR1 is downregulated in colon tumors and patient-derived colon cancer cell lines, but the biological significance of this repression is yet to be determined. Normal colon crypts do not propagate in tissue culture, preventing us from performing knockdown experiments of DACOR1. We therefore examined the biological effects of DACOR1 by overexpressing it in several patient-derived colon cancer cell lines. Initially, we utilized three distinct patient-derived colon cancer cell lines (V481, V852 and V866) that we transduced with either a control or a DACOR1 lentivirus. Induction of DACOR1 in these patient-derived colon cancer cell lines resulted in reduced growth of these cells (Appendix A-Figure A-9). To quantify this effect, we performed colony formation assays (CFA) using all three lines (V481, V852 and V866), and found that the induction of DACOR1 affected colony formation in V481 by ~25% (p = 0.0002), in V852 cells by ~53% (p = 0.003), and in V866 by 81% (p = 0.007) (Figure 2-3 E). The effect of DACOR1 induction, although consistent in reducing colonies, varied among the three lines as each line was derived from a distinct patient tumor, and thus have underlying genetic differences.

To test whether the effects we observed on colony formation were due to non-specific effects of overexpressing DACOR1, we performed several control experiments. First, we selected two patient-derived colon cancer cell lines, V703 and V425, that although had reduced levels of DACOR1 relative to normal colon, they still maintained some level of DACOR1 expression (Figure 2-2 D).
expression of \textit{DACOR1} in both cell lines had minor effects on colony formation of these cells, as compared to a control lentivirus (Appendix A-Figure A-10). Second, to rule out that the phenotype is due to high expression levels of \textit{DACOR1} lentivirus (CMV promoter), we cloned the full length of \textit{DACOR1} downstream of a weak PGK promoter and measured its expression levels in comparison to normal colon and control lentivirus. Using this approach, we are able to bring the overexpression level of \textit{DACOR1} closer to the expression levels observed in normal colon (Appendix A-Figure A-11). We carried out colony formation assays of control vs. \textit{DACOR1} lentivirus-transduced cells and also observed significant reduction in colony formation (Appendix A-Figure A-11). Finally, we cloned the full length of an oncogenic lncRNA, TCON_00011938, which is not associated with DNMT1, downstream of a strong CMV promoter, and found that the overexpression of this distinct lncRNA led to increased colony formation (Appendix A-Figure A-12). Collectively, these results suggest that \textit{DACOR1} induction reduces the clonogenic potential of colon cancer cells.

\textit{DACOR1} induction affects global gene expression of colon cancer cells

To gain insights into \textit{DACOR1} function, we performed RNA-seq using RNA isolated from the colon cancer cell line V852 transduced with either control or \textit{DACOR1} lentivirus and identified differentially expressed genes (Trapnell et al., 2013). We found that induction of \textit{DACOR1} affected the expression of ninety-nine genes (p < 0.05, q <0.05). Specifically, we observed that induction of \textit{DACOR1} led to the repression of several known inhibitors of TGF-\beta/BMP
signaling, including SMAD6, INHBE (inhibin beta E), and FST (follistatin), which we confirmed by qRT-PCR in two distinct colon cancer cell lines (Figure 2-4 A) (Hata et al., 1998). Previous studies have demonstrated that TGF-β/BMP signaling exerts a tumor suppressor function in the colon, and it becomes inactivated or repressed in a majority of sporadic colorectal cancers (Kodach et al., 2008). SMAD6, which is upregulated in colon tumors (Appendix A-Figure A-13) and downregulated by DACOR1, plays a major role in repressing TGF-β/BMP signaling (Hata et al., 1998).

We also found that the induction of DACOR1 led to the downregulation of several genes involved in amino acid metabolism with known roles in tumorigenesis, including PHGDH, PSAT1, CBS, and ASNS (Balasubramanian et al., 2013; Bhattacharyya et al., 2013; Possemato et al., 2011). First, we confirmed that the induction of DACOR1 leads to the repression of these genes in two distinct colon cancer cell lines, V852 and V866, by qRT-PCR (Figure 2-4 B). We subsequently confirmed the repression of PHGDH at the protein level by western blot analysis (Figure 2-4 C). PHGDH plays a key role in de novo serine biosynthesis (Amelio et al., 2014; Chaneton et al., 2012; Locasale, 2013; Possemato et al., 2011) and is highly upregulated in many colon tumors (Appendix A-Figure A-13). To determine if the repression of PHGDH by DACOR1 induction affects serine levels, we measured pyruvate kinase M2 (PKM2) activity, which is dependent on serine (Chaneton et al., 2012). Indeed, we found that DACOR1 induction leads to reduced PKM2 activity in two independent experiments (Figure 2-4 D), without affecting overall PKM2 protein
levels (Figure 2-4 C). Lastly, the repression of *Cystathionine β-synthase* (CBS) by *DACOR1* is intriguing (Figure 2-4 C), as reduced CBS levels are known to lead to increased methionine, the substrate needed to generate S-adenosyl methionine (SAM). SAM is the key methyl donor utilized by DNA methyltransferases for DNA methylation in mammalian cells. Thus, DNMT1-*DACOR1* interaction appears to indirectly regulate cellular SAM levels, and, consequently, genome-wide DNA methylation. Collectively, these findings suggest that *DACOR1* plays key roles in regulating DNA methylation, and specific tumor-suppressor and metabolic pathways in colon cells to potentially suppress colon tumorigenesis.
Figure 2-4. Induction of DACOR1 affects global gene expression in colon cancer cells. 

(A) qRT-PCR confirmations of RNA-seq data that DACOR1 represses several genes (SMAD6, FST and INHBE) involved in the repression of the TGF-beta/BMP signaling pathway. These observations suggest that induction of DACOR1 enhances the signaling of these pathways, which have tumor-suppressor effects. (B) qRT-PCR validations of RNA-seq data that DACOR1 represses the expression of key genes that are involved in amino acid biosynthesis and metabolism. (C) Western blot analyses demonstrate that DACOR1 induction leads to the repression of PHGDH and CBS but does not affect PKM2 or Actin protein levels in V852 cells. (D) Induction of DACOR1 reduces the activity of PKM2, which is known to be dependent on serine, without affecting overall PKM2 protein levels.
DACOR1 interacts directly with chromatin at specific genomic sites

To gain insights into the potential mechanism(s) by which DACOR1 could regulate gene expression and consequently cellular phenotype, we mapped the genomic occupancy of DACOR1 across the entire human genome using ChIRP-seq (Chromatin Isolation by RNA Purification followed by DNA sequencing) (Chu et al., 2011). First, we designed several biotin-modified oligonucleotides complementary to DACOR1 and confirmed that we can specifically isolate DACOR1 from cross-linked cell lysates (Figure 2-5 A). Subsequent ChIRP-seq and analysis identified 338 DACOR1 genomic occupancy sites, including 161 peaks near 150 annotated genes (multiple peaks per gene in some cases) and 177 sites in intergenic regions. As expected, we observed a peak corresponding to the genomic region of DACOR1 transcription upstream of SMAD3. We compared the genomic occupancy sites of DACOR1 near annotated genes to differentially methylated regions (DMRs) in a cohort of colon tumors vs. matched normal tissues (Simmer et al., 2012). Of the 150 annotated gene loci occupied by DACOR1, 31 sites overlap with these DMRs ($p < 3.5 \times 10^{-14}$) (Figure 2-5 B). These findings indicate that DACOR1 interacts with both DNMT1 and chromatin and, potentially, recruits and/or assembles the DNMT1 macromolecular protein complex at specific genomic sites to regulate epigenetic modifications and, consequently, the expression of specific genes and pathways (Figure 2-5 C).
Figure 2-5. **DACOR1 is associated with chromatin.** (A) Confirmation of **DACOR1** pull down from cross-linked cell lysates by specific complementary probes, in comparison to non-specific probes. (B) Intersection of **DACOR1** genome occupancy sites near annotated protein-coding genes identified in this study by ChIRP-seq and differentially methylated regions (DMRs) in colon tumors/normal colon identified by Simmer et al. (Simmer et al., 2012) reveals a significant overlap. This further supports the role of **DACOR1**, via its interaction with DNMT1, in regulating genome-wide DNA methylation. (C) A proposed model of how DNMT1-DACOR1 interactions regulate DNA methylation and gene expression. **DACOR1** interacts with specific genomic loci and potentially recruits DNMT1 to establish DNA methylation patterns and/or regulate gene expression. The DNMT1-DACOR1 axis results in modulating the expression of many genes, directly and indirectly, including those involved in amino acid metabolism. When **DACOR1** becomes repressed during colon tumorigenesis, several metabolic changes occur. For example, upregulation of PHGDH levels results in increased serine levels, and consequently increased PKM2 activity. Also, upregulation of CBS levels, which is known to cause a decrease in methionine and S-adenosyl methionine (SAM) levels, could potentially impact the levels of SAM available for DNA methylation causing a global hypomethylation.
2.3 Discussion

We previously identified specific interactions between human long non-coding RNAs (lncRNAs) and several chromatin-modifying complexes, and demonstrated that these interactions are required for regulating gene expression (Khalil et al., 2009). In this study, we identified specific interactions between the DNA methyltransferase DNMT1 and human lncRNAs, suggesting that in addition to histone modifications, DNA methylation is also indirectly regulated by lncRNAs. DNA methylation is an important epigenetic mark for the regulation of gene expression in mammalian cells from early embryonic development to fully differentiated post-mitotic cells. Our current findings that DNMT1 associates with lncRNAs suggest that these lncRNAs may influence DNMT1 genomic occupancy and/or activities, thereby indirectly regulating the methylome. Thus, deregulation of one or more of DNMT1-associated lncRNAs in human disease would lead to changes in DNA methylation patterns and potentially significant changes in gene expression without any detectable changes in DNMT1 expression levels. Indeed, we found in our studies that induction of the lncRNA DACOR1 is sufficient to change DNA methylation patterns without affecting DNMT1 protein levels in colon cancer cells. However, the mechanisms of DACOR1-mediated changes in DNA methylation patterns are not yet known. A number of potential mechanisms may be implicated, including DACOR1-mediated recruitment of DNMT1 to specific sites of the genome, similar to what has been observed of lncRNA-mediated recruitment of histone-modifying enzymes (Bertani et al., 2011; Klattenhoff et al., 2013; Tsai et al., 2010). Also, DACOR1 could affect DNMT1
activity at specific CpG sites, potentially by regulating protein components of the DNMT1 macromolecular protein complex.

We currently have several technologies to identify IncRNA-protein interactions; however, it is not currently known how IncRNAs, such as DACOR1, interact with epigenetic-modifying complexes such as histone-modifying enzymes and DNA methyltransferases. It has been proposed that secondary structure of IncRNAs plays a major role in this IncRNA-protein recognition (Khalil and Rinn, 2011; Morris and Mattick, 2014), however, the experimental evidence for this still requires much needed research. Also, it is not yet clear how epigenetic complexes bind RNAs, as they do not possess canonical RNA-binding domains similar to those found in classic RNA-binding proteins. It is possible that these proteins possess RNA-binding “domains” that are distinct from those identified in classic RNA-binding proteins, or the interaction could be indirect and mediated by RNA-binding proteins. Also, despite the recent development of technologies to map the genome-wide occupancy of nuclear IncRNAs, it is not yet known how IncRNAs can recognize specific genomic loci. Proteins that serve as intermediates between DNA and IncRNAs could also mediate this interaction.

Many IncRNAs, including DACOR1, are poorly conserved, even within mammalian species (Necsulea et al., 2014). However, several studies have now demonstrated functionality of IncRNAs despite this lack of sequence conservation (Ulitsky et al., 2011). The lack of sequence conservation for many IncRNAs makes it challenging to use model organisms to interrogate the function of IncRNAs identified in human systems and diseases. Furthermore, studies of
conserved lncRNAs in mouse models do not always recapitulate what has been observed in human cells (Dimitrova et al., 2014; Schorderet and Duboule, 2011). These observations suggest that researchers may have to utilize non-human primates or human organoids to study lncRNAs. As an example, DACOR1 would require a non-human primate such as rhesus macaque for knock out experiments to delineate its function in vivo.

Gene expression analyses of DACOR1 demonstrated that many colon tumors and colon cancer cell lines dramatically repress DACOR1 expression. Restoring DACOR1 expression into patient-derived colon cancer cells resulted in reduced growth, potentially via the modulation of several pathways. For example, DACOR1 downregulates the expression of several genes that inhibit TGF-β/BMP signaling, and thus potentially enhances TGF-β/BMP signaling, which is known to exert a tumor-suppressor activity in the colon (Hata et al., 1998; Kodach et al., 2008). DACOR1 also downregulates several genes involved in metabolism including de novo serine biosynthesis (e.g., PHGDH, PSAT1). Serine is an essential precursor for the synthesis of proteins, nucleic acid and lipids, thus, it is critical for cancer cell growth (Amelio et al., 2014; Locasale, 2013; Possemato et al., 2011). Furthermore, we found that DACOR1 induction is sufficient to attenuate pyruvate kinase M2 (PKM2) activity, which is highly dependent on serine (Chaneton et al., 2012). PKM2 has been recently implicated as a key gene in cancer metabolism (Chaneton et al., 2012; Christofk et al., 2008; Wong et al., 2013); thus, the identification of a lncRNA that attenuates its activity, although indirectly, may provide a therapeutic window in cancer biology. Lastly, DACOR1-
mediated downregulation of CBS, the deficiency of which is known to lead to increased levels of methionine and, consequently, S-adenosyl methionine (SAM), the key methyl donor utilized by DNA methyltransferases to methylate DNA, is also highly significant. These findings suggest that DNMT1, via its interaction with DACOR1, indirectly regulate the cellular levels of SAM and, subsequently, genome-wide DNA methylation.

In addition to IncRNAs, we also found a small number of mRNAs that co-immunoprecipitate with DNMT1. However, the significance of these interactions is currently unknown. We cannot rule out that some of these mRNAs have dual functions: protein-coding capacity as well as non-coding function. Previous studies have identified mRNAs that, in addition to their coding potential, function as RNA molecules to regulate a number of biological processes (Derrien et al., 2012; Lanz et al., 1999; Morris and Mattick, 2014; Poliseno et al., 2010). Also, a previous study that utilized a genome-wide computational approach to identify IncRNAs found that many “hypothetical” coding genes are indeed non-coding (Jia et al., 2010). Thus, future studies may reveal other coding transcripts that also have a non-coding function.

In summary, our current study uncovered a critical role of IncRNAs in regulating the human methylome – these findings could potentially help explain, at least in part, the genome-wide changes in DNA methylation across numerous cancer types. Furthermore, since many IncRNAs have tissue-specific expression patterns, they could serve as biomarkers of disease prognosis as well as for therapeutic strategies with potentially less side effects than coding genes.
2.4 Materials and Methods

**Sequencing data files.** All next generation RNA and DNA sequencing data files are deposited in GEO under GSE58989

**Optimization of RNA co-immunoprecipitation (RIP) in HCT116 cells:** We have previously utilized RNA co-immunoprecipitation (RIP) in human fibroblasts and HeLa cells to identify interactions between human IncRNAs and several chromatin-modifying complexes (Khalil et al., 2009; Moran et al., 2012). For this study, we optimized our RIP protocol in HCT116 cells by initially performing control experiments on a well conserved RNA-protein interaction in the spliceosome: the interaction between U1-70K protein and the small nuclear RNA U1 (Surowy et al., 1989). First, we tested an antibody against U1-70K in immunoprecipitation experiments and confirmed that this antibody specifically immunoprecipitate U1-70K protein from HCT116 cell lysate (**Appendix A-Figure A-1**). We also used an IgG antibody that should not recognize any protein as a negative control. Subsequently, we performed three independent biological replicates of U1-70K RIPs from cross-linked HCT116 cell lysate. After several stringent washes, we reversed the formaldehyde crosslinking by heat and isolated associated RNA using Trizol. Quantitative Real time PCR (qRT-PCR) analysis of U1 RNA using three distinct endogenous controls (GAPDH, 18S rRNA and CLDN3) revealed a specific interaction between U1-70K and U1 RNA
(Appendix A-Figure A-1). These results suggest that our RIP protocol is optimized in HCT116 to detect specific RNA-protein interactions.

**Immunoprecipitation (IP) of U1-70K and flag-DNMT1, and western blot analysis:** We utilized an antibody against U1-70K (Synaptic Systems, Cat # 203001) to immunoprecipitate (IP) the U1-70K protein, and anti-flag antibody to IP flag-DNMT1 from HCT116 cell lysates as follows: HCT116 cells were grown in 2 x 15 cm plates before harvesting by trypsin. An equal amount of media was added to quench the reaction and the cells were collected by centrifugation in a 15 ml conical tube at 500 g for 10 minutes. The pellets were washed twice with PBS prior to fixing in a final concentration of 0.3% formaldehyde for 15 minutes at room temperature. The reaction was quenched by adding glycine to a final concentration of 0.125 mM and incubated at room temperature for 5 minutes. The cells were pelleted by spinning at 500 g for 10 minutes and then washed twice with 1X PBS before suspending the pellets in 2.2 ml of RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl (pH 7.4), 1mM EDTA). The cells were incubated at 37°C for 30 minutes and vortexed every 5 minutes at 30-second intervals for the duration of the incubation. Samples were homogenized using a dounce homogenizer to disrupt cellular membranes. The lysate was centrifuged using a microcentrifuge at maximum speed (~13,300 RPM), and the supernatant was transferred to a new tube. 100 ul of the supernatant was taken as input, and half of remaining supernatant was incubated with an antibody against protein of interest (i.e., U1-
70K or flag-DNMT1), and the second half with an IgG antibody (negative control) overnight with rotation at 4°C. Next day, 50 ul of protein A/G magnetic beads was added to each tube and incubated for 30 minutes at room temperature with rotation. The beads, which now have the antibody and bound protein, were collected using a magnet, and washed three times with RIPA buffer and once with 1X PBS. For protein analysis by western blot, we added 100 ul of Lammeli buffer to each tube, and incubated the samples at 95°C for 5 minutes before running the samples on a denaturing SDS-PAGE gel.

RNA co-immunoprecipitation (RIP) of U1-70K and flag-DNMT1: The same protocol described above was utilized for RNA co-immunoprecipitation (RIP) of U1-70K or flag-DNMT1 from HCT116 cells. However, for the isolation of co-immunoprecipitated RNAs, we suspended the magnetic beads+antibody+protein in 100 ul of buffer C (150 mM NaCl, 50 mM Tris-HCl (pH = 7.4), 5 mM EDTA, 10 mM DTT, 1% SDS) and 10 ug of proteinase K. The samples were incubated at 42°C for 30 minutes for protein digestion, and subsequently at 65°C for 4 hours to reverse the formaldehyde crosslinking. RNA was isolated by adding 800 ul of Trizol and 200 ul of chloroform to each sample, mixed, and centrifuged at full speed for 10 minutes, and the upper clear layer (~600 ul) was transferred to a 1.5 ml tube with 600 ul of 70% ethanol. The mixture was applied to an RNeasy mini kit column (Qiagen) according to the manufacturer’s protocol. All samples were treated with DNase prior to final washes and elution with 20 ul of RNase-free water.
Analysis of RNA-seq data from RIP-seq samples: RNA sequencing libraries were made using a stranded ScriptSeq V2 (Illumina) according to manufacturer's protocol. Raw RNA-seq fastq files were aligned to UCSC human hg19 using TopHat v2.0.10. Transcript assembly was performed using Cufflinks v2.1.1. Relative transcript abundance for both mRNAs and lncRNAs was reported as fragments per kilobase of exon per million fragments mapped (fpkm). If fpkm values reported in the input sample were less than 1.0 for mRNAs and less than 0.1 for lncRNAs, the transcript was filtered as not expressed in HCT116 cells. Fold changes were then calculated as the average fpkm across RIP samples to the fpkm of the input control sample. Transcripts were identified as binding to DNMT1 if their fold change was greater than 2-fold. Heatmaps were generated using the heatmap.2 function in the gplots package (version 2.12.1) in R [R Core Development Team. (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.]

Quantitative Real time PCR (qRT-PCR): RNA was converted to cDNA using RNA to cDNA EcoDry™ Premix Random Hexamers (Clontech). TaqMan assays for GAPDH, 18s rRNA, U1, CLDN3, and DACOR1 were purchased from Life Technologies. Other primer pairs were designed using primer3 software, and most primers used were designed to span exon-exon boundaries. TaqMan Mastermix (Life Technologies) or Maxima SyBr Green/ROX qPCR Master Mix (Thermo Scientific) were used for qRT-PCR. A comparative C_{\text{T}} quantitation was
performed with a hold stage of 50°C for 2 min and 95°C for 10 min followed by 40x cycle of 95°C for 15 s and 60°C for 1 min and finally melt curve at 95°C for 15 s, 60°C for 1 min, and a ramp to 95°C at 0.3°C increments. Analysis was done using the $2^{-ΔΔC_T}$ method with GAPDH as the reference gene.

**Colony Formation Assay (CFA):** The colon cancer cell lines V481, V852, V866, V703 and V425 were transduced with either a control or DACOR1 lentivirus, and non-infected cells were eliminated by puromycin. For colony formation assays (CFA), cells were plated in either 6-well or 10cm plates in triplicates of each condition (control vs DACOR1 lentivirus). Cells were plated at 1,250, 2,500, 5,000 or 10,000 cells per well/plate, and kept under puromycin selection. Colonies were fixed with methanol/acetic acid and subsequently stained with 0.1% crystal violet solution. Plates were scanned and colonies were counted using the publically available ImageJ software (Schneider et al., 2012). Average colony counts were calculated for control and DACOR1 plates for each cell line, and a paired t-test was used to test for statistical significance.

**Illumina 450K DNA methylation arrays:** DNA was extracted from V481 and V852 cells using the DNeasy Blood and Tissue kit (Qiagen). DNA methylation profiling was performed at the Genomics Core Facility at Case Western Reserve University using the Illumina 450K HumanMethylation BeadChip (12 samples/chip). Biological triplicates from both the control and DACOR1 lentivirus transduced cells were tested in order to detect accurate methylation status. Beta
values, a ratio of the methylated/un-methylated signal, were reported and ranged from 0 (completely un-methylated) to 1 (completely methylated). In filtering probes, each cell line was analyzed separately. Reported beta values were removed if the p-value for detectable probe signal was greater than 0.05. Targets were then filtered if only a single Beta value remained in either condition. The median Beta value was calculated for control and DACOR1 samples. Targets were further filtered if the difference in the maximum Beta and minimum Beta was greater than 0.1 (10% different). Using the median Beta, sites were determined as differentially methylated if the absolute value of the delta-Beta was greater than 0.1 (>10%).

**Next generation RNA-sequencing (RNA-seq):** Six RNA samples were isolated from V852 cells transduced with either a control lentivirus (n=3) or DACOR1 lentivirus (n=3). RNAs with RNA integrity number (RIN) equal or larger than 8 (max is 10) were considered high quality and suitable for RNA-seq. Library preparation was performed using Scriptseq™ Complete Gold (Human/Mouse/Rat) (Illumina) and sequenced on an Illumina Hi-Seq2500. All six samples were run on a single flow cell, and 100 bp paired-end strand-specific sequencing reads were generated and mapped to human genome release hg19 using TopHat with 2 mismatches allowed for full-length reads. The raw reads were mapped to human genes annotated in RefSeq database using Cufflinks V2.0.2, and CuffDiff was used for identifying differentially expressed genes.
(Trapnell et al., 2013). All expression values were calculated as fragment per kilobase of exon per million of mapped fragments (FPKM).

**ChIRP-seq of DACOR1:** The ChIRP-seq protocol was carried out as previously described by Chu et al (Chu et al., 2011). Briefly, 5 x 10^8 V852 cells with DACOR1 lentivirus were first crosslinked using 1% formaldehyde for 10 minutes. The cells were spun down, suspended in Buffer A (Hepes 20mM, KCl 10mM, MgCl₂ 1.5 mM, DTT 0.5mM, 1% Empigen), and dounced before collecting the nuclei by centrifugation. The nuclei were sonicated in nuclei lysis buffer (Tris-HCl pH 7.5 20mM, EDTA 10mM, 1% SDS, 1mM DTT, protease inhibitor cocktail, RNaseOut 80 U/ml) to produce 100-500 bp DNA fragments. LiCl₂ was added at 0.5M to nuclear lysates. Equal amounts of nuclear lysates were incubated with either DACOR1 specific or non-specific DNA probes modified with a TEG linker and Biotin at their 5’ ends, and incubated for 24 hour at 37°C with rotation. Next day, RiboMinus™ streptavidin-coated magnetic beads (Life Technologies) were blocked with 800 ug/ml yeast tRNA and 800 ug/ml BSA for 1 hour at 37°C in hybridization buffer (Tris-HCl pH 7.5 5mM, EDTA 10mM, LiCl₂ 500mM) before washing and adding to nuclear lysates for 30 minutes. The beads were then washed three times with nuclear lysis buffer, wash buffer (Tris-HCl pH 7.5 5mM, EDTA 0.5mM, NaCl 1M), and PBS. The beads were suspended in 200 ul of PBS and incubated at 75°C for 5 minutes, the supernatant was collected from the beads and incubated at 65°C overnight to reverse crosslinking before extracting DNA using DNeasy Blood & Tissue Kit (Qiagen). Paired-end DNA sequencing
was performed on a HiSeq2000/2500 at Otogenetics Corporation. DNA reads were mapped against human genome (hg19) using Bowtie 2 and peak calling was performed by using MACS2. Peak annotation was completed using ChIPpeakAnno.

**PKM2 activity assay:** Cells were collected by trypsinization and pellets were washed twice by cold PBS. The pellets were then resuspended in RIPA buffer (150 mM NaCl, 1 mM EDTA, 1 mM DTT, 1% Triton X-100, 25.5 mM deoxycholic acid, 50 mM Tris-HCl, pH 7.5) and sonicated briefly at 4°C. The total extracts were subjected to PKM activity assay as described follows: reaction mixtures contain 50 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM MgCl2, 0.5 mM ADP, 0.2 mM NADH, 8 units LDH (lactate dehydrogenase from sigma) and 1mM DTT. The lysates (1-10 ug of total protein) were added to the assay mixture to reach 200 ul of the final volume in 96-well plates. The enzymatic reaction was initiated by the addition of PEP (phosphoenolpyruvic acid, 0.5 mM) as the substrate. The oxidation of NAPH was monitored at 340 nm for 3 min using a Thermo Max microplate reader (Molecular Devices). The number of units of NADH oxidation was calculated using the standard extinction coefficient of NADH (ε = 6.22 mM⁻¹ cm⁻¹). This value was then divided by the total amount of protein added in the assay giving units/mg of protein from the cell extracts. For all analyses, PKM2 activity was calculated using an amount of cell lysate where the reaction rates fell within the linear range of dependence on the concentration of lysate.
Acknowledgements. We would like to thank Ernest Chan, Ph.D. and Sarah McMahon for help with bioinformatic analyses of next generation sequencing data; Drs. Anthony Wynshaw-Boris, Thomas LaFramboise, Mark Chance, Peter Scacheri, Paul Tesar and Peter Harte for discussion of results. Next generation RNA sequencing and DNA methylation analyses were performed at CWRU Genomics Core Facility. These studies are supported by a new investigator startup funds from CWRU (AMK), and by NIH awards P50CA150964 (SM), U01 CA152756 (SM), R37-DK060596 (MH), and RO1-DK053307 (MH), and by the Clinical and Translational Science Collaborative of Cleveland, UL1TR000439 from the National Center for Advancing Translational Sciences (NCATS).

Conflict of interest. The authors declare that there are no conflicts of interest to be disclosed.
Chapter 3:

Integrative transcriptome-wide analyses reveal critical HER2-regulated mRNAs and lincRNAs in HER2+ breast cancer

Callie R. Merry\textsuperscript{1,2}, Sarah McMahon\textsuperscript{1}, Cheryl L. Thompson\textsuperscript{3,5}, Kristy L.S. Miskimen\textsuperscript{3}, Lyndsay N. Harris\textsuperscript{3,4}, and Ahmad M. Khalil\textsuperscript{1,2,3,*}

Affiliations:
Departments of \textsuperscript{1}Genetics and Genome Sciences, \textsuperscript{2}Biochemistry, \textsuperscript{3}Case Comprehensive Cancer Center, \textsuperscript{4}Medicine and \textsuperscript{5}Family Medicine and Community Health, Case Western Reserve University, Case Western Reserve University, Cleveland, OH 44106

Reference:
Abstract

Purpose: Breast cancer is a major health problem affecting millions of women worldwide. Over 200,000 new cases are diagnosed annually in the USA, with approximately 40,000 of these cases resulting in death. HER2-positive (HER2+) breast tumors, representing 20-30% of early stage breast cancer diagnoses, are characterized by the amplification of the HER2 gene. However, the critical genes and pathways that become affected by HER2 amplification in humans are yet to be specifically identified. Furthermore, it is yet to be determined if HER2 amplification also affects the expression of long intervening non-coding (linc)RNAs, which are involved in the epigenetic regulation of gene expression. Methods: We examined changes in gene expression by next generation RNA-sequencing (RNA-seq) in human tumors pre- and post- HER2 inhibition by trastuzumab in vivo, and changes in gene expression in response to HER2 knock down in cell culture models. We integrated our results with gene expression analysis of HER2+ tumors vs matched normal tissue from The Cancer Genome Atlas (TCGA). Results: The integrative analyses of these data sets led to the identification of a small set of mRNAs, and the associated biological pathways that become deregulated by HER2 amplification. Furthermore, our analyses identified three lincRNAs that become deregulated in response to HER2 amplification both in vitro and in vivo. Conclusions: Our results should provide the foundation for functional studies of these candidate mRNAs and lincRNAs to further our understanding of how HER2 amplification
results in tumorigenesis. Also, the identified lincRNAs could potentially open the door for future RNA-based biomarkers and therapeutics in HER2+ breast cancer.
3.1 Introduction

HER2-positive breast cancer is characterized by the overexpression of the HER2 receptor leading to aberrant signaling and cellular response. Cell signaling through the HER2 receptor is initiated through heterodimerization of HER2 with other ERBB family members in response to extracellular signals. Dimerization and subsequent autophosphorylation of the tyrosine kinase domain initiates a signaling cascade that ultimately leads to changes in gene expression patterns that regulate cell proliferation and growth (Mitri et al., 2012). We postulated that identifying the key genes (both coding and non-coding) that become deregulated in response to HER2 amplification could provide important insights into how HER2 amplification affects cell proliferation. Furthermore, identifying novel non-coding genes, such as lincRNAs, that become deregulated when HER2 becomes amplified may provide clues into global changes in gene expression patterns observed in HER2-positive breast cancer.

In addition to mRNAs and small non-coding RNAs, the human genome encodes over 7,600 long intervening non-coding RNAs (lincRNAs) (Cabili et al., 2011; Derrien et al., 2012; Guttman et al., 2009; Khalil et al., 2009). Gene expression studies revealed that individual lincRNAs can regulate numerous mRNA genes, mostly at genomic sites far away from the lincRNA site of transcription in a trans-regulation mechanism (Gupta et al., 2010; Huarte et al., 2010; Khalil et al., 2009; Kretz et al., 2012; Loewer et al., 2010). The impact of lincRNAs on gene expression patterns has implicated these transcripts in a wide range of biological functions including dosage compensation, genomic imprinting,
alternative splicing, nuclear organization, and regulation of mRNA translation (Carrieri et al., 2012; Clark and Mattick, 2011; Moran et al., 2012; Qureshi et al., 2010; Wang and Chang, 2011; Yildirim et al., 2013). Intriguingly, lincRNAs utilize various mechanisms to exert their effects in the cell. These mechanisms include acting as scaffolds and guides for chromatin-modifying complexes to specific regions of the genome, serving as decoys that regulate transcription factors binding to specific DNA sequences, and as microRNA “sponges” that regulate miRNA:mRNA interactions (Cesana et al., 2011; Gupta et al., 2010; Guttman et al., 2011; Huarte et al., 2010; Khalil et al., 2009; Loewer et al., 2010; Tsai et al., 2010; Ulitsky et al., 2011; Wang et al., 2011).

Recently, there has been growing interest in unraveling the roles of lincRNAs in human disease with the hope that these novel transcripts can be used as diagnostic biomarkers and/or therapeutic targets (Mattick, 2009; Niland et al., 2012; Qureshi et al., 2010; Taft et al., 2010). In addition, there is a significant interest in studying the functional roles of lincRNAs in cancer since numerous lincRNAs have been shown to be dysregulated across multiple cancer types (Fatica and Bozzoni, 2014; Iyer et al., 2015; Morris and Mattick, 2014). These studies have also shown that the expression of some lincRNAs correlates with clinical parameters, such as overall patient prognosis and metastasis (Gupta et al., 2010; Kogo et al., 2011; Niland et al., 2012; Yang et al., 2011). For example, the lincRNA \textit{HOTAIR} is highly upregulated in breast cancer, hepatocellular carcinoma, and colorectal cancers (Gupta et al., 2010; Kogo et al., 2011; Yang et al., 2011). Additionally, \textit{in vitro} and \textit{in vivo} functional studies have
implicated HOTAIR in promoting cancer metastasis (Gupta et al., 2010). The lincRNA linc-p21 is transcriptionally activated by the tumor suppressor p53 and subsequently represses a subset of p53 gene targets \textit{in trans} through binding of hnRNP-K, thereby facilitating the p53 response in cells (Huarte et al., 2010). A final example focuses on the X inactive specific transcript (XIST), which has been known for decades to be the key transcript for X chromosome inactivation (Xi) in mammalian females. Recently, it has been shown that genetic deletion of \textit{XIST} in hematopoietic stem cells results in a variety of hematological cancers and premature death in mice (Yildirim et al., 2013). Taken together, these selected examples, as well as many others, demonstrate important roles of lincRNAs in tumorigenesis and metastasis, and their potential utilization as biomarkers and/or therapeutic targets.

To further our understanding of the effects of HER2 signaling on gene expression of both mRNAs as well as novel non-coding genes such as lincRNAs, we utilized a combined \textit{in vitro} and \textit{in vivo} transcriptomic approach to pinpoint critical downstream genes. In this integrative analysis, the \textit{in vivo} data gives us the ability to identify gene expression changes that are relevant to patients. Importantly, having a view of the clinically relevant gene expression changes can provide researchers with candidate genes for therapeutic target. The \textit{in vitro} data provided through analysis of cell culture models of HER2-positive breast cancer allows us to study the genes that we identify through genetic manipulation or siRNA knockdowns. Both the \textit{in vivo} and the \textit{in vitro} aspects of this study provide necessary tools for further study and project development. The expression of the
mRNAs and lincRNAs identified through the integrative analysis were further examined in RNA-seq data sets obtained from The Cancer Genome Atlas (TCGA) project to further validate and refine our list of top candidates. Our results led to the identification of potentially key mRNAs and lincRNAs that may contribute to HER2-positive breast cancer.

3.2 Results

In vivo gene expression changes after inhibition of HER2 by trastuzumab

Previous studies have shown that HER2 inhibition in cell culture models of HER2-positive breast cancer leads to dramatic changes in the expression of mRNAs. However, it is not currently known which of these mRNAs are affected as a result of HER2 inhibition in tumors in vivo, as well as the effects of HER2 inhibition on regulatory non-coding RNAs such as lincRNAs. To identify the mRNAs and lincRNAs that become affected by HER2 inhibition in vivo, we analyzed RNA-seq data from a clinical trial that we originally designed to predict benefit from the HER2 inhibitor, trastuzumab (Herceptin®) (Clinical Trials.gov ID NCT00617942, also see methods). Trastuzumab is a monoclonal antibody drug designed to target the HER2 receptor by potentially inhibiting its heterodimerization (Junttila et al., 2009; Valabrega et al., 2007). The clinical trial accrued 80 patients of which 50 pairs of tumors could be biopsied pre-trastuzumab and post-trastuzumab (post one dose, ~10-14 days). All patients continued receiving a combination of chemotherapy and trastuzumab for 4 months, and a subset of thirteen tumor pairs representing the extremes of
response to treatment, was subjected to total RNA isolation and RNA-seq analysis. Eleven out of these thirteen patients were identified as responders to trastuzumab as measured by pathological complete response (pCR). We reasoned that since trastuzumab was an effective therapeutic drug in these eleven HER2-positive breast cancer patients, changes in gene expression pre- and post- one dose of trastuzumab would reflect mRNAs and lincRNAs regulated through HER2. Therefore, we identified differentially expressed mRNAs and lincRNAs pre- vs post- one dose of trastuzumab treatment in our RNA-seq data set of these eleven patients that were responsive to treatment as measured by pCR.

We calculated FPKM (fragments per kilobase of exon per million fragments mapped) values of each known mRNA and lincRNA in the human genome in each RNA-seq sample (22 total: 11 pre- vs 11 post- trastuzumab). We calculated an average FPKM of each mRNA and lincRNA in pre- vs post-trastuzumab, and subsequently calculated a fold change of Post/Pre for each mRNA and lincRNA. We identified 228 mRNAs and 28 lincRNAs to be differentially expressed (Figure 3-1 A-B). We also performed pathway analysis on differentially expressed mRNAs and identified five cancer-related pathways to be significantly affected (p < 0.05, p-values ranged from 0.006 – 0.00004) (Figure 3-1C). In summary, we identified a set of mRNAs and lincRNAs that become affected when HER2 signaling is inhibited in vivo. However, because of genetic and environmental differences of these 11 patients, we observed heterogeneity in the level of differential expression of both mRNAs and lincRNAs.
Thus, to further refine our list of key mRNAs and lincRNAs that are critical components of the HER2 pathway, we decided to modulate HER2 in a cell culture model.
Figure 3-1. Inhibition of HER2 in vivo by trastuzumab results in significant changes in the expression of mRNAs and lincRNAs in HER2-positive breast cancer patients. Gene expression analysis by next generation RNA sequencing (RNA-seq) was performed on RNA isolated from HER2-positive tumors pre- and post- one dose of trastuzumab. Expression of genes was calculated as FPKM values and differentially expressed genes between pre- and post-treatment conditions were identified. Differentially expressed mRNAs (A) and lincRNAs (B) in all 22 RNA-seq samples (11 pre- vs 11 post-trastuzumab) are represented by heatmaps; C. Over enrichment analysis of differentially expressed mRNAs pre- vs post-trastuzumab reveals several key pathways affected in vivo when HER2 signaling is inhibited. Number of genes and p-values are shown for each pathway.
Gene expression changes in a HER2-positive breast cancer cell model after HER2 siRNA knockdown

We began our studies in BT474 cells, which are HER2-positive breast cancer cells that have been utilized extensively as a cell culture model to study HER2-positive breast cancer. To identify mRNAs and lincRNAs that are downstream of HER2 signaling in BT474 cells, we utilized a loss-of-function approach in which we depleted HER2 using validated siRNAs and examined changes in gene expression using RNA-seq. We transfected BT474 cells with siRNAs targeting HER2, and simultaneously transfected the same number of cells with negative control siRNAs. We initially examined HER2 mRNA levels using RT-qPCR at 48 and 72 hours. At these time points, we found that the siRNAs were effective at knocking down HER2 by 80 and 87 percent, respectively (Figure 3-2 A and data not shown). Subsequently, we performed a new round of transfections at 24, 48 and 72 hours, and examined HER2 protein levels by western blot analysis. At 24 hours there was a modest reduction in HER2 protein levels, however, a significant reduction in HER2 protein levels in comparison to cells transfected with negative control siRNAs was achieved 48 hours post-transfection (data not shown and Figure 3-2B). Based on these gene expression analyses of HER2 knockdown with siRNAs, the 48 hour post-transfection time point was chosen as optimal to sufficiently deplete HER2 at both the mRNA and protein levels, while still capturing some of the early changes in mRNAs and lincRNAs expression due to loss of HER2. To that end, equal numbers of BT474 cells were transfected with either HER2 siRNAs or negative
control siRNAs in three biological replicates, and at 48 hours post-transfection total RNA was isolated, quantified, and subjected to RNA-seq (see Methods).

We found 1,015 mRNAs (303 upregulated and 712 downregulated) and 167 lincRNAs (139 upregulated and 28 downregulated) to be differentially expressed by ≥ 2-fold in the HER2-depleted BT474 cells in comparison to cells transfected with negative control siRNAs (Figure 3-2 C-D). To place our findings into context, we identified a previous study that utilized mRNA microarrays to examine the effects of HER2 inhibition by trastuzumab on mRNAs in BT474 and SKBR3 cells (Le et al., 2005). This study reported sixteen mRNA genes that are significantly downregulated in response to HER2 inhibition by trastuzumab. We found that eight of these sixteen mRNA genes were also significantly downregulated in our current study. We designed primers for those mRNA genes, and confirmed in an independent set of knockdown experiments that their expression is altered in response to HER2 depletion in BT474 breast tumor cells by RT-qPCR (Figure 3-2 E). Furthermore, pathway analysis of differentially expressed mRNAs revealed numerous affected pathways post HER2 depletion in BT474 cells (Appendix B-Figure B-1).
Figure 3-2. HER2 affects the expression of mRNAs and lincRNAs in BT474 cells. A. Real time qPCR analysis of HER2 mRNA levels in BT474 cells transfected with either control siRNAs or HER2 siRNAs (15 nM final concentration). We observed a significant knockdown of HER2 mRNA levels at 48 hours post siRNA transfections. B. Western blot analysis of HER2 protein levels in BT474 cells transfected with control siRNAs vs cells transfected with HER2 siRNAs at 48 hours post transfections. A significant reduction in HER2 protein levels is observed. Actin was used as a loading control. C-D. Gene expression analysis by RNA-seq led to the identification of differentially expressed mRNAs and lincRNAs in BT474 cells treated with either negative control (NC) or HER2 siRNAs. Heatmaps of differentially expressed mRNAs and lincRNAs between BT474 cells transfected with control siRNAs vs HER2 siRNAs are shown. E-F. Validation of RNA-seq data by RT-qPCR of eight mRNA genes and five lincRNAs in cells transfected with control siRNAs vs HER2 siRNAs. For both mRNAs and lincRNAs, we observed similar patterns in our original knockdown experiments analyzed by RNA-seq and subsequent knockdown experiments analyzed by RT-qPCR.
Since no previous studies of lincRNAs modulated in response to HER2 depletion or inhibition are available, we selected five lincRNAs from our RNA-seq data for validation by qPCR in an independent knockdown experiment in BT474 cells. We selected these lincRNAs based on fold changes and p-values and we included both up- and downregulated lincRNAs in response to HER2 depletion. Four out of the five lincRNAs showed statistically significant up- or downregulation in response to HER2 depletion by RT-qPCR similar to what we observed by RNA-seq analysis (Figure 3-2 F). To further confirm that these lincRNAs are downstream of the HER2 pathway, and that changes in their expression are not due to off-target effects of siRNAs, we utilized a second independent siRNA against HER2. First, we confirmed that this second siRNA is effective at knocking down HER2 protein levels by western blot analysis (Figure 3-3 A). Next, we performed qPCR analysis on the same 5 lincRNAs, and found that four out of the five lincRNAs show a similar response to knocking down HER2 with the second siRNA, but only three lincRNAs pass a p-value of < 0.05 (Figure 3-3 B). Also, the expression of linc-MCL1-1, which was not affected with first siRNA, was responsive to the second siRNA, similar to what we observed by RNA-seq. These experiments demonstrate that HER2 depletion in BT474 cells affect the expression of both mRNAs and lincRNAs, however, some variability is observed due to off-targets effects of siRNA-mediated depletion. Thus, to overcome these limitations we identified mRNAs and lincRNAs that are affected by both HER2 inhibition by trastuzumab in vivo and HER2 depletion in BT474 cells by siRNAs.
Figure 3-3. Validation of HER2-regulated lincRNAs in BT474 cells using a second independent siRNA. 

A. We designed and confirmed the knockdown of HER2 protein using a second independent siRNA (distinct from first siRNA used for knockdown experiments and subsequent RNA-seq analysis). Western blot analysis of HER2 protein levels in BT474 cells treated with siRNA #2 alone or in combination with first siRNA (siRNA pool) demonstrate the effectiveness of this second siRNA in knocking down HER2 protein in comparison to negative control siRNAs.

B. We examined the expression of the same 5 lincRNAs in figure 2F post-HER2 knockdown in BT474 cells with a second independent siRNA by qPCR analysis. We found that HER2 knockdown with a second independent siRNA also affects the expression of some of these lincRNAs.
To identify mRNAs and lincRNAs that are affected by both HER2 inhibition by trastuzumab in tumors and HER2 depletion in cell culture, we intersected differentially expressed mRNAs and lincRNAs that we identified post HER2 knockdown in BT474 cells with differentially expressed mRNAs and lincRNAs that we identified in response to HER2 inhibition by trastuzumab in vivo. We found 44 mRNAs and 3 lincRNAs to be common between the two data sets (Figure 3-4 A-B). Some of the 44 mRNA genes identified are key components of the PLK1 signaling pathway and E2F transcription factor network (Figure 3-4 C), which are key pathways affected in HER2-positive tumors. In summary, we have identified a small set of mRNAs and lincRNAs that are affected in response to HER2 inhibition/depletion in vivo and in cell culture suggesting an important role of not only mRNAs but also lincRNAs in HER2-positive breast cancer. To gain further insights into the potential role of these genes in HER2-positive breast cancer, we next examined their expression in a cohort of HER2-positive breast tumors and their matched normal control tissue in the Cancer Genome Atlas (TCGA) database.
Figure 3-4. Identification of commonly affected mRNAs and lincRNAs post-HER2 inhibition in tumors \textit{in vivo} and post-HER2 knockdown in cell culture by siRNAs. To identify mRNAs and lincRNAs that are likely critical targets of HER2, we intersected differentially expressed mRNAs and lincRNAs that we have identified in pre- vs post-HER2 inhibition by trastuzumab in tumors \textit{in vivo} and differentially expressed mRNAs and lincRNAs in BT474 cells treated with either control or HER2 siRNAs. We identified 44 mRNAs (A) and 3 lincRNAs (B) to overlap between the two data sets. (C) Of the 44 common mRNAs identified in our aforementioned analysis, several genes are known components of the E2F transcription factor network and PLK1 signaling pathway. These two pathways were also identified in our analysis of gene expression of HER2-positive tumors and matched normal breast tissues from TCGA, suggesting an important role for these two pathways in HER2-positive breast cancer.
Thousands of mRNAs and hundreds of lincRNAs are differentially expressed in HER2+ breast cancer tumors

By using transcriptomic analyses pre- vs. post-HER2 knockdown in cell culture and pre- vs. post-HER2 inhibition in tumors, we are able to identify a small set of lincRNAs and mRNAs that are putative members of the HER2 regulatory network. To further validate our observations, we turned to RNA-seq data from The Cancer Genome Atlas (TCGA) project to determine the expression patterns of these mRNAs and lincRNAs in HER2 tumors vs. matched normal tissue. The TCGA represents one of the most comprehensive studies of RNA expression in thousands of cancers including breast cancer [34]. We mined the TCGA RNA-seq data for HER2-positive breast tumors with matched normal tissue. In total, we identified 12 patients (24 RNA-seq samples: from 12 tumors and 12 matched normal tissue). Within this TCGA cohort, we identified 2,521 mRNAs and 283 lincRNAs to be differentially expressed between these tumors and their matched normal pairs (Figures 3-5 A-B). The pathway analysis of differentially expressed mRNAs in the TCGA data set revealed cancer-related pathways were highly enriched (Figure 3-5 C).
Figure 3-5. Dysregulation of mRNAs and lincRNAs in HER2-positive tumors. We utilized publically available RNA-seq of 12 HER2-positive tumors and 12 matched normal breast tissues to identify differentially expressed mRNAs and lincRNAs. Differentially expressed mRNAs (A) and lincRNAs (B) in HER2-positive breast cancer patient tumors vs. adjacent matched normal tissue obtained from The Cancer Genome Atlas (TCGA) are represented by heatmaps. In total, 2,521 mRNAs and 283 lincRNAs were identified as differentially expressed (see methods). C. Over enrichment analysis of differentially expressed mRNAs between tumors and matched control samples from 12 HER2-positive breast cancer patients (TCGA cohort) reveals several biological pathways, including cancer-associated pathways, that become altered due to increased HER2 expression in these tumors. Number of genes and p-values associated with each pathway are also shown.
To determine the expression patterns of the 44 mRNAs and 3 lincRNAs that are affected in response to both HER2 knockdown (in BT474 cells) and HER2 inhibition (in tumors *in vivo*), we examined their expression in the TCGA RNA-seq data. Of the 44 mRNAs identified, 35 mRNAs are also dysregulated in the TCGA cohort. We graphed the expression values of each of these 35 mRNAs in all three data sets: HER2 inhibition *in vivo* (Post-trastuzumab/Pre-trastuzumab), HER2 knockdown in BT474 cells (siHER2/siControl), and TCGA cohort (Tumor/Normal). Strikingly, each mRNA shows a similar directionality in HER2 knockdown in BT474 cells and in HER2 inhibition in tumors *in vivo*, and as expected, negatively correlated in the TCGA cohort (tumor/normal) (*Figure 3-6*). Gene names are shown in *Appendix B-Figure B-2*. 


Figure 3-6. Identification of 35 mRNAs that are affected in all three data sets with the expected directionality of expression. Of the 44 mRNAs that are affected by both HER2 inhibition in tumors by trastuzumab and HER2 knock down in BT474 cells, 35 of these mRNAs were also deregulated in HER2-positive tumors vs normal tissues (TCGA). These 35 mRNA transcripts are found to be dysregulated in all three data sets: TCGA HER2-positive (Tumor/Normal), trastuzumab clinical trial samples (Post-/Pre-trastuzumab treatment), and in BT474 cells (HER2 siRNA/control siRNA). We graphed the expression values of these 35 mRNAs in all three data sets, and strikingly each mRNA shows similar directionality in HER2 knockdown in BT474 cells and in HER2 inhibition in vivo, and as expected, negatively correlated in the TCGA cohort (tumor/normal).
We also examined the expression of the three intersected lincRNAs in our analysis of TCGA RNA-seq data. In this analysis, it was expected that a lincRNA that is upregulated in response to HER2 depletion and inhibition to be downregulated in tumor samples in comparison to matched normal tissue, and vice versa. *Linc-STARD6-2*, which we found to be downregulated when HER2 was knocked down or inhibited, is upregulated in 8/12 tumors in comparison to their matched normal tissue (*Figure 3-7 A*). *Linc-GJA1-2* and *linc-SLC39A10-10*, which become upregulated in response to HER2 knockdown or inhibition, are downregulated in 10/11 and 9/12 tumors in comparison to their matched normal tissue, respectively (*Figure 3-7 B-C*).
Figure 3-7. Validation of lincRNAs expression in 12 tumors vs. matched normal tissue (TCGA cohort). We have identified three lincRNAs that are affected by both HER2 inhibition and HER2 knockdown in tumors and BT474 cells, respectively. To determine the potential role of these lincRNAs in HER2-positive cancer, we examined their expression in HER2-positive tumors and matched normal tissues from TCGA RNA-seq data sets. The fold change of A. linc-STARD6-2, B. linc-GJA1-2, and C. linc-SLC39A10-10 was graphed for each TCGA HER2-positive patient sample (Tumor/Normal). Linc-STARD6-2 shows upregulation in 8/12 TCGA Tumor/Normal samples. linc-GJA1-2 and linc-SLC39A10-10 show downregulation in 10/11 and 9/12 tumor samples compared to matched normal controls, respectively. One patient (A1LB) did not show expression of linc-GJA1-2 in either normal or tumor tissue. Three samples, A0DZ, A1EN, A18U, have no linc-GJA1-2 expression in the tumor sample preventing calculation of the exact fold change, thus, bars were extended to the edge of the graph for these samples to indicate that there is an observable difference between tumor and normal samples.
3.3 Discussion

The amplification of the HER2 gene in 20-30% of early-stage breast cancer patients demonstrates the important role of this genetic event in breast cancer tumorigenesis. To determine the effects of HER2 amplification on gene expression, a number of previous studies have examined changes in mRNA expression in HER2-positive breast cancer cell lines in response to HER2 inhibition using mRNA microarrays. However, these microarrays are limited in both the dynamic range as well as in the number of gene targets that can be interrogated. Also, these microarrays examined a very small number of long non-coding RNAs. Furthermore, since these studies were all performed using cell lines, it was not possible to determine their relevance to what takes place in tumors in vivo. The recent advances in high-throughput sequencing technologies made it possible to examine all differentially expressed genes (coding and non-coding) in greater depth and accuracy (Wang et al., 2009). In addition to mRNAs, we can now analyze the entire transcriptome including novel classes of non-coding RNAs such as lincRNAs. Many of these non-coding RNAs are regulatory in nature, and can greatly influence mRNA gene expression patterns suggesting that their expression may also affect human health and disease (Fatica and Bozzoni, 2014; Moran et al., 2012; Niland et al., 2012; Ulitsky and Bartel, 2013). Thus, in our current study, we have utilized several RNA-seq data sets that take advantage of both cell culture models of HER2-positive breast cancer and clinically relevant tumor data in vivo to identify both lincRNAs and mRNAs that are affected in HER2-positive breast cancer. These RNA-seq data sets were
generated from RNA isolated from tumors pre- and post-inhibition of HER2 by trastuzumab \textit{in vivo}, RNA isolated from BT474 cells treated with either negative control or HER2 siRNAs and RNA isolated from HER2-positive breast tumors and their matched normal controls. By interrogating these data sets and integrating the results, we are able to identify a small set of mRNAs and lincRNAs that are affected by HER2 amplification during tumorigenesis. We believe that these mRNA and lincRNA candidates are the most promising for follow up functional studies.

In our current studies we have identified a small number of mRNAs that are clearly dysregulated in HER2-positive tumors and become significantly affected by HER2 depletion or inhibition. Importantly, several of these genes are already known to affect cell growth and proliferation and could potentially impact tumorigenesis as a result of HER2 amplification. For example, we identified the \textit{RRM2} gene, which has been previously shown to be associated with decreased survival in breast cancer (Putluri et al., 2014), to be highly upregulated in HER2-positive tumors and is significantly downregulated by both HER2 inhibition and depletion. Another key gene that we identified is \textit{TOP2A}, which is required for DNA replication as it functions as a topoisomerase. \textit{TOP2A} has been a target of several anti-cancer drugs such as etoposide and doxorubicin (Hande, 2008). A final example is \textit{PRC1}, which is highly upregulated during S and G2/M phases of mitosis is also downregulated by both HER2 knockdown and trastuzumab-mediated inhibition (Jiang et al., 1998). These few examples provide insights into how HER2 amplification leads to increased cell proliferation during tumorigenesis.
and potentially opens the door for designing combined therapies targeting both HER2 by trastuzumab and other genes by either currently approved or new drugs. In future studies, we will carry out high throughput drug screens to identify new drugs that target specific proteins identified in our current study.

Although the overexpression of HER2 in breast cancer has been known and studied for over twenty years, most studies have mainly focused on identifying proteins as therapeutic targets. However, recent advances in developing epigenetic-based therapies suggest that it is also critical to explore non-coding RNAs as drug targets. Since many lincRNAs are also known to interact with epigenetic complexes and in some cases guide them to their genomic targets (Kaneko et al., 2014; Khalil et al., 2009), this introduces the opportunity for RNA-based therapeutic approaches. There are a few advantages to targeting lincRNAs instead of epigenetic protein complexes directly. First, because many lincRNAs have tissue-specific expression, therapeutically targeting lincRNAs in various cancers may reduce negative side effects in comparison to current treatments. Additionally, since some lincRNAs direct chromatin-modifying complexes to specific genomic loci (Kaneko et al., 2014; Tsai et al., 2010), this opens up the possibility of lincRNA-mediated epigenetic reprogramming of genomic regions responsible for dysregulation in cancer (Moskalev et al., 2012). While the developments of RNA-based therapies are still in their infancy, some successes have been achieved. For example, a recent study has targeted the extra chromosome 21 in Down Syndrome for
heterochromatization and silencing by the lincRNA XIST using genome editing technologies (Jiang et al., 2013).

Therapeutic approaches that target lincRNAs could include small molecules or antagonistic oligonucleotides that would prevent lincRNA-protein interactions, since most lincRNAs function within ribonuceloprotein (RNP) complexes (Khalil and Rinn, 2011). A direct inhibition or downregulation of lincRNA expression could also be utilized, however, a major challenge in targeting lincRNAs is their low expression levels in human cells in comparison to mRNAs (Dinger et al., 2009; Fatica and Bozzoni, 2014; Geisler and Coller, 2013; Iyer et al., 2015; Mustafi et al., 2013; Wapinski and Chang, 2011). This is due to the fact that many lincRNAs are regulatory molecules, and in many cases a few copies per cell is sufficient for these lincRNAs to exert their effects (Wang et al., 2011). Thus, in our analysis of RNA-seq data sets we have set a 0.25 fpkm or higher cutoff for a lincRNA and 1 fpkm or higher for a mRNA to be called expressed to be able to capture lowly expressed lincRNAs that can still exert biological activity in cells. Our future work will focus on elucidating the functional significance of lincRNAs in breast cancer pathology.
3.4 Material and Methods

Next generation RNA-sequencing files: All raw files have been deposited in GEO: GSE60182

Clinical Trial Information: Institute: Brown University Oncology Group (Brown University, Yale University, Cedar-Sinae Center), PI: William Sikov MD; Correlative Science PI: Lyndsay Harris MD; BrUOG Study ID: BR-211B; Clinical Trials.gov ID NCT00617942.

RNA isolation and next generation RNA sequencing (RNA-seq) from tumor samples (clinical trial): Frozen biopsied cores were processed for RNA isolation using AllPrep (Qiagen), and the Ovation RNA-seq System (NuGen) was used for RNA amplification. Library preparation was performed using TruSeq v3 (Illumina) and then sequenced on an Illumina HiSeq2500. The sequenced reads were aligned to the human genome version hg19 using GSNAP (Wu and Nacu, 2010). The uniquely aligned reads were further analyzed using Cufflinks V2.0.2 and aligned to human mRNAs and lincRNA databases (Cabili et al., 2011; Guttman et al., 2009; Trapnell et al., 2012). Expression values were calculated as FPKM (fragment per kilobase of exon per million of mapped fragments) and were used to determine expression of lincRNAs and mRNAs in pre- and post-treatment samples. Transcripts were called expressed if FPKM values across either all pre-treatment samples or all post-treatment samples were ≥ 1.0 or ≥ 0.25, for mRNAs and lincRNAs, respectively. The mean expression level was
calculated and differences in expression between pre- and post-treatment samples were assessed for the 11 patients that achieved a pathological complete response (pCR) using the non-parametric Wilcoxon test for paired samples. Additionally, fold change (post/pre) were calculated to identify differentially expressed transcripts. Transcripts were deemed differentially expressed if the fold change was ≥ 2.0 or ≤ 0.5. Heatmaps were generated in R using heatmap.2, and Z-scores were scaled by row using standard Z-score calculation of log fold change.

**Cell culture of breast cancer cell lines:** Human breast cancer cell lines were grown in Hybri-Care Medium (ATCC® 46-X™) supplemented with 10% fetal bovine serum (FBS) (Bioexpress) and 100 units/ml of Penicillin and 100 μg/ml Streptomycin (Life Technologies) at 37°C with 5% CO₂.

**RNA isolation from cell lines:** RNA from BT474 cells was isolated using RNeasy® Mini Kit (Qiagen) according to the manufacturer’s protocol. An added DNase (Qiagen) treatment step was included after the first wash to remove DNA contamination of RNA preps.

**siRNA transfections:** The knockdown of HER2 was achieved through the transfection of HER2-specific siRNAs (Life Technologies, Catalogue # 4390824, s611 and s613) at a final concentration of 15 nM with Lipofectamine RNAiMax
(Life Technologies) at 7.5 μL/well of 6-well plate. Negative control siRNA #1 and #2 (Ambion Cat. # AM4611 and AM4613) were used as negative controls.

**Western blot analyses:** Protein lysates were prepared with Laemmli Sample Buffer (BioRad) and separated on a gradient 4-20% SDS-PAGE Mini-Protean® TGX™ Gels (BioRad). The gel was transferred to a nitrocellulose membrane (Thermo Scientific) and probed with primary antibodies overnight. Anti-β-actin (Ambion, AM4302, 3.1 mg/ml) was used as a loading control at a dilution of 1:1000, and anti-HER2/Erb2 (Cell Signaling, 2242S) was used to detect HER2 at a dilution of 1:500. Anti-mouse HRP (Thermo Scientific, 32230) and anti-rabbit HRP (Abcam, ab6721) were used as secondary antibodies. HRP was activated using SuperSignal® West Pico Chemiluminescent Substrate (Thermo Scientific) for autoradiography.

**Next generation RNA sequencing of BT474 cells:** After RNA was isolated from BT474 samples, we assessed the quality of RNA using BioRad Experion. RNA samples with RNA integrity number (RIN) larger than 8 (max is 10) were considered high quality and suitable for RNA-seq. Library preparation was performed using Scriptseq™ Complete Gold (Human/Mouse/Rat) (Illumina) and sequenced on Illumina Hi-Seq2500. Six samples were run on a single flow cell. We generated 100 bp paired-end strand-specific sequences, which were mapped to human genome release hg19 using TopHat with 2 mismatches allowed for full-length reads. The raw reads were mapped to human genes annotated in RefSeq
database and lincRNAs annotated in Cabili et al. (Cabili et al., 2011) using Cufflinks V2.0.2, and subsequently used for differential gene expression analysis after normalizing the values to the total mapped reads in each sample. Expression values were calculated as FPKM (fragment per kilobase of exon per million of mapped fragments) and were used to determine expression of both lincRNAs and mRNAs in HER2 knock down (KD) and negative control (NC) siRNA samples. Transcripts were considered expressed if FPKM values across either all BT474 HER2 KD samples or all NC samples were ≥ 1.0 for mRNAs and ≥ 0.25 for lincRNAs. The mean expression level was calculated, and from this statistically significant differences in expression between HER2 KD samples and NC samples was determined using a paired t-test in R. Additionally, fold changes (HER2 KD / NC) were calculated to identify differentially expressed transcripts. Transcripts were deemed differentially expressed if the fold change was ≥ 2.0 or ≤ 0.5. Heatmaps were generated in R using heatmap.2, and Z-scores were scaled by row using standard Z-score calculation of log fold change.

**Real time quantitative PCR (RT-qPCR):** Real time quantitative PCR (RT-qPCR): RNA was converted to cDNA using RNA to cDNA EcoDry™ Premix Random Hexamers (Clontech). Primers pairs were designed using primer3 software, and most primers used were designed to span exon-exon boundaries. A complete list of all primers is included. Maxima SyBr Green/ROX qPCR Master Mix (Thermo Scientific) was used for qRT-PCR. A comparative C_{T} quantitation was performed with a hold stage of 50°C for 2 min and 95°C for 10 min followed
by 40x cycle of 95°C for 15 s and 60°C for 1 min and finally melt curve at 95°C for 15 s, 60°C for 1 min, and a ramp to 95°C at 0.3°C increments. Analysis was done using the $2^{-\Delta\Delta CT}$ method with GAPDH as the reference gene [47].

**Analysis of Next generation RNA sequencing from The Cancer Genome Atlas (TCGA):** RNA sequencing (RNA-seq) fastq files from 12 tumor and 12 adjacent matched normal breast cancer pairs were obtained through The Cancer Genome Atlas (TCGA) consortium via the Cancer Genomics Hub (https://cghub.ucsc.edu) (The Cancer Genome Atlas, 2012). Sequences were aligned to UCSC hg19 with default parameters in TopHat v2.0.11 specifying an unstranded library strategy (Trapnell et al., 2012). Aligned sequences were then assembled into both mRNA and lincRNA transcripts using default parameters in Cufflinks v2.1.1 (Trapnell et al., 2012). Relative transcript abundance was reported from Cufflinks as Fragments Per Kilobase of exon per Million fragments mapped (FPKM). Transcripts were deemed as expressed if FPKM values across either all tumor samples or all matched normal samples were ≥ 1.0 for mRNAs and ≥ 0.25 for lincRNAs. The non-parametric Wilcoxon test for paired samples was used to test for statistical significance ($p < 0.05$) using the statistical software R [R Development Core Team. (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria]. Fold changes were calculated as tumor/normal, and transcripts with greater than a 2-fold change were selected as differentially expressed. Over enrichment analysis of differentially expressed mRNAs using NCBI cancer pathways was
also performed in R using fisher’s exact test. P values were corrected for multiple testing using the Bonferroni correction. Heatmaps were generated in R using heatmap.2, and Z-scores were scaled by row using standard Z-score calculation of log fold change.

**Acknowledgements:** We would like to thank Maria Sandoval, Jennifer Yori, and Ruth Keri at CWRU Pharmacology for BT474 cells and other reagents; Vinay Varadan at Case Comprehensive Cancer Center and Sheldon Bai at the RNA Center for discussion of bioinfomatic analyses; Megan E Forrest, Jessica Sabers, Thomas LaFramboise, and Anthony Wynshaw-Boris for discussion of results.

**Conflict of interest:** Callie R. Merry, Sarah McMahon, Cheryl L. Thompson, Kristy L.S. Miskimen, Lyndsay Harris, and Ahmad M Khalil declare that there is no conflict of interest to be disclosed.
Chapter 4:

Transcriptome-wide identification of mRNAs and lincRNAs associated with trastuzumab resistance in HER2-positive breast cancer

Callie R. Merry\textsuperscript{1,2}, Sarah McMahon\textsuperscript{1}, Cynthia F. Bartels\textsuperscript{1}, Alina Saiakhova\textsuperscript{1}, Peter Scacheri\textsuperscript{1,3}, Cheryl L. Thompson\textsuperscript{3,4}, Mark W Jackson\textsuperscript{3}, Lyndsay N. Harris\textsuperscript{3,4}, and Ahmad M. Khalil\textsuperscript{1,2,3}

Affiliations:
Departments of \textsuperscript{1}Genetics and Genome Sciences, \textsuperscript{2}Biochemistry, \textsuperscript{3}Case Comprehensive Cancer Center, \textsuperscript{4}Medicine and \textsuperscript{5}Nutrition, Case Western Reserve University, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106

Manuscript in preparation
Abstract

Approximately, 20-30% of early-stage breast cancer patients are HER2-positive (HER2+). These patients are characterized at the molecular level by amplification of the HER2 gene (also known as HER2/neu and ERBB2), resulting in a substantial increase in HER2 protein levels. HER2, a transmembrane tyrosine kinase receptor, is targeted therapeutically by a monoclonal antibody, trastuzumab, that binds to the extracellular domain of HER2 and inhibits its signaling activities. Although trastuzumab is effective in a substantial fraction of early-stage breast cancer patients, approximately 25% of patients have primary resistance to trastuzumab, and 70% of initial responders acquire resistance within one year. The mechanisms of trastuzumab resistance (TzR) are not currently known. To gain insights into TzR, we applied an integrative transcriptomic approach of RNA isolated from trastuzumab-sensitive and trastuzumab-resistant breast cancer tumors, and isogenic cell lines, and identified differentially expressed mRNAs and lincRNAs. This approach led to the identification of a small set of mRNAs and lincRNAs that are associated with trastuzumab resistance in cell culture and in vivo in humans. Functional analysis of a top candidate gene, S100P, demonstrated that inhibition of S100P results in reversing trastuzumab resistance. Mechanistically, S100P activation of RAGE, and downstream RAS/MEK/MAPK pathway compensates for HER2 inhibition by trastuzumab. Finally, the upregulation of S100P appears to be driven by epigenomic changes at the enhancer level. In summary, our findings demonstrate that TzR is due to activation of alternative pathways rather than
direct mutation or truncation of the HER2 receptor. Therapeutic approaches targeting S100P-RAGE axis could emerge as a powerful new therapy in treating breast cancer patients who are resistant to current therapies.
4.1 Introduction

Breast cancer is a major health problem affecting millions of women worldwide, and results in over 500,000 deaths annually. Previous studies have led to the classification of breast cancer into several subtypes, with HER2-positive (HER2+) tumors representing ~25-30% of early stage breast cancer patients’ diagnoses (Mitri et al., 2012; The Cancer Genome Atlas, 2012). HER2-positive breast cancer patients are characterized at the molecular level by an amplification of a genomic region encompassing the HER2 gene (also known as HER2/neu and ERBB2), which is a member of the ERBB family of transmembrane receptor tyrosine kinases (Mitri et al., 2012). Although no known ligands bind to the HER2 receptor itself, a number of ligands bind to other ERBB family members (e.g., HER3), and lead to the heterodimerization of these members with HER2 (Arteaga and Engelman, 2014). Heterodimerization results in autophosphorylation by the intracellular tyrosine kinase domain of HER2, and the initiation of a signaling cascade that results in the activation and repression of specific genes (Merry et al., 2015b).

A few monoclonal antibodies that bind to and inhibit HER2 are currently utilized in the clinic to treat HER2-positive breast cancer patients (Le et al., 2005; Mitri et al., 2012). Trastuzumab (Herceptin®) is a monoclonal antibody drug that binds to the extracellular domain of the HER2 receptor, and is thought to inhibit its heterodimerization, and consequently, its signaling cascade (Ali Nahit Sendur et al., 2012). Currently, trastuzumab is the standard first-line treatment for HER2-positive breast cancer patients as 70% of early stage patients respond to
a dual therapy of trastuzumab and chemotherapy. However, a significant percentage of early-stage patients (~25-30%) do not respond or benefit from trastuzumab (i.e., resistant tumors/ non-responders) by unknown mechanisms (Oliveras-ferraros et al., 2010). Furthermore, ~70% of initial responders relapse and acquire resistance to trastuzumab within one year (Ali Nahit Sendur et al., 2012; Baretta et al., 2013; Figueroa-Magalhães et al., 2013). Currently, the mechanisms of trastuzumab resistance (TzR) are not known. Although several mechanisms of TzR have been proposed, these models remain controversial (Ali Nahit Sendur et al., 2012; Baretta et al., 2013). For example, it was hypothesized that resistance occurs due to shedding of the extracellular domain of HER2, and thus, trastuzumab is no longer able to bind to HER2 (Vu and Claret, 2012). In several resistant cell lines, we have demonstrated that this is not the case – i.e. trastuzumab binds to HER2 and inhibits the autophosphorylation of HER2 intracellular domain (see results). Other models proposed that HER3 signaling could compensate for HER2 inhibition by trastuzumab. However, patients who received a combination of trastuzumab and pertuzumab, which blocks HER2-HER3 interaction, have a median progression-free survival of 5.5 months (Baselga et al., 2010).

Given the immense challenges in identifying mechanisms of TzR, we decided to take a comprehensive and unbiased approach to identify the key cellular factors that drive TzR in humans in vivo. We postulated that specific coding and non-coding genes become altered in TzR tumors that compensates for loss of HER2 signaling by trastuzumab. By applying next generation RNA-
sequencing (RNA-seq) to RNA isolated from human HER2-positive tumors and cell culture models of the disease that are either TzS or TzR, we have identified a subset of mRNAs and long intergenic non-coding RNAs that are strongly associated with TzR (Fig. 4-1 A). Functional studies of a top candidate gene, S100P, demonstrated that inhibition of S100P results in reversal of trastuzumab-resistance (TzR).

4.2 Results

Identification of differentially expressed mRNAs and lincRNAs in trastuzumab-sensitive (TzS) vs. trastuzumab-resistant (TzR) human tumors in vivo

We hypothesized that TzR tumors have a distinct gene expression profile compared to TzS tumors, and a subset of differentially expressed coding and non-coding genes contribute to the resistance phenotype. To that end, we applied RNA-seq to RNA isolated from HER2-positive tumor biopsies that were collected during a clinical trial (see methods). In this clinical trial, 50 out of 80 patient tumors were biopsied at the beginning of the trial prior to receiving trastuzumab (pre-trastuzumab). After receiving a combination of chemotherapy and trastuzumab for 4 months, a subset of thirteen tumors, representing the extremes of response to treatment, were selected for gene expression analysis by RNA-seq. Specifically, eleven of those thirteen patients were identified as responders (i.e., trastuzumab-sensitive or TzS) and two patients were identified as non-responders (i.e., trastuzumab-resistant or TzR) as measured by pathological complete response (pCR). We compared mRNA gene expression between the responders (TzS) and non-responders (TzR), and identified
differentially expressed genes between the two groups prior to receiving
treatment (pre-trastuzumab) (Fig. 4-1 A). Specifically, we identified ~1500
mRNAs (Fig. 4-1 B) and 371 lincRNAs (data not shown) that are differentially
expressed. We performed pathway analysis on differentially expressed mRNAs,
and found several pathways to be affected including those related to mammary
gland cell proliferation and development (Fig. 4-1 C). Also, RAGE receptor
binding was statistically significant and emerged as a key term (see below).
Figure 4-1. Identification of differentially expressed genes in trastuzumab-resistant (TzR) vs trastuzumab-sensitive (TzS) tumors by RNA-seq. (A) A schematic outlining overall study design. Differentially expressed genes identified in TzS vs TzR human HER2-positive tumors in vivo were intersected with differentially expressed genes identified in TzS vs. TzR BT474 cells, which led to a small list of mRNAs and lincRNAs that are associated with TzR. (B) Heatmap representation of the differentially expressed mRNAs in TzS patients (responders) vs. TzR patients (non-responders). (C) Pathway analysis of differentially expressed mRNAs in tumors in vivo. Pathway name is listed on Y-axis, percentage of genes affected in each pathway is indicated on X-axis, and p-value is indicated at the end of each bar.
Characterization of TzS and TzR HER2-positive cell culture model

To further pinpoint key mRNAs and lincRNAs that contribute to TzR, we generated TzR HER2-positive BT474 cells by chronically exposing the cell line to trastuzumab over six weeks. To confirm that we have generated BT474 cells that are resistant to trastuzumab (TzR), we first measured cell proliferation of the parental TzS and isogenic TzR cells under exposure to trastuzumab over a 96 hour time period. We found that the proliferation of parental TzS BT474 cells is significantly affected when exposed to trastuzumab within 48 hours. By contrast, the proliferation of TzR BT474 cells was not affected (Fig. 4-2 A-B).

Next, we examined the effect of trastuzumab on HER2 and the phosphorylation of HER2 (p-HER2) protein levels in TzS and TzR cells. We found TzS cells to express comparable levels of HER2 protein to TzR cells (Fig. 4-2 C). However, trastuzumab strongly inhibits p-HER2 within 24 hours of exposure (Fig. 4-2 D). Next, we examined the expression of HER2 and p-HER2 levels in both TzS and TzR cells over a 96-hour time course, and found: i. HER2 protein levels are not significantly affected in the parental TzS line by trastuzumab until the 96-hour time point (Fig. 4-2 E), ii. p-HER2 levels significantly decrease in the parental TzS line within 12 hours post-exposure to trastuzumab (Fig. 4-2 F), iii. Surprisingly, both HER2 and p-HER2 levels are significantly affected in the TzR BT474 cells at 48 hours post exposure to trastuzumab (Fig. 4-2 G-H). The decrease in HER2 and p-HER2 levels in the resistant line demonstrate that trastuzumab can still binds to and inhibit HER2 phosphorylation in TzR cells. These observations demonstrate that the
mechanism of TzR, at least in a subset of tumors, is not due to truncation or mutations of HER2 that affect trastuzumab function.
Figure 4-2. Characterization of HER2-positive trastuzumab-sensitive (TzS) and trastuzumab-resistant (TzR) BT474 breast cancer cell lines. Cell proliferation assay via cell counting with (A) one dose of 10 μg/ml trastuzumab at 0 hr and (B) 10 μg/ml trastuzumab at 0 and 48 hours show that the BT474 parental line is sensitive to trastuzumab as shown by the decrease in percentage of cells with time when treated. The trastuzumab-resistant BT474 show continued growth corresponding to the untreated parental establishing that the resistant BT474 cells do not respond to trastuzumab treatment. Western blot analysis shows that while (C) trastuzumab (10 μg/ml) does not alter the protein levels of HER2, (D) it does decrease the phosphorylation of HER2 after 48 hours, (E) HER2 protein levels stay relatively constant in parental BT474 with trastuzumab treatment over 96 hours, (F) phosphorylated HER2 decreases significantly by 24 hours with loss by 48 hours. Resistant BT474 shows a similar decrease of (G) HER2 and (H) phosphorylated HER2 at 48 hours.
TzR is associated with transcriptome-wide changes in gene expression in cell culture model

To identify genes associated with TzR in our cell culture model described above, we measured changes in gene expression between parental TzS and TzR BT474 cells by RNA-seq. We identified 233 mRNAs and 34 lincRNAs as differentially expressed (fold change ≥ 2, p ≤ 0.05) (**Figure 4-3 A-B**). To further refine our list of genes that are specifically associated with TzR, and not due to short exposure to trastuzumab, we treated the parental TzS BT474 cells with trastuzumab, and collected RNA at 48 hours post-exposure. We measured changes in gene expression by RNA-seq and identified 244 mRNAs and 27 lincRNAs that become differentially expressed as compared to mock treated cells (≥ 2 fold change and p ≤ 0.05). Subsequently, mRNAs and lincRNAs that are affected due to short exposure to trastuzumab were subtracted from mRNAs and lincRNAs that are differentially expressed in TzR vs TzS cells. This enabled us to further pinpoint mRNAs and lincRNAs that are specifically associated with long-term resistance to trastuzumab in culture. This analysis narrowed down the gene list to 201 mRNAs and 29 lincRNAs.
Numerous mRNAs and lincRNAs are differentially expressed in HER2-positive TzS vs. TzR BT474 cell line. Heatmap representation of differentially expressed (A) mRNAs and (B) lincRNAs in TzS BT474 vs TzR BT474 isogenic cell lines.

Figure 4-3.
We intersected mRNAs and lincRNAs identified in our cell culture model with differentially expressed mRNAs and lincRNAs identified in our in vivo clinical trial data set. This analysis led to the identification of 18 mRNAs and 4 lincRNAs that are associated with TzR in vivo and in cell culture (two-tailed Fischer’s exact test p<0.0001 and p = 0.0013, respectively). Expression of top mRNA genes identified in our analysis in TzS vs TzR tumors in vivo is shown using cluster graphs (Figure 4-4 A). Careful examination of these genes, their known roles in breast cancer, and our bioinformatics analysis of key pathways associated with TzR (Fig. 4-1 C), suggested a key role of S100P. S100P is part of the S100 family of proteins, which have documented roles in tumorigenesis (Bresnick et al., 2015), and is known to bind to the receptor RAGE to activate similar signaling pathways to those activated by HER2 signaling. Furthermore, analysis of S100P in Oncomine demonstrated that S100P is strongly associated with breast cancer (p = 1.8E-94, average fold change = 10.9). Furthermore, examination of RNA-seq of The Cancer Genome Atlas also demonstrated that S100P is highly upregulated in HER2-positive breast cancer (Figure 4-5 A).
Figure 4-4. Top candidate mRNAs that were identified as associated with trastuzumab-resistant (TzR) both in vivo and in cell culture model. The expression of each mRNA in each tumor sample (blue: TzS, red: TzR) is shown as FPKM value based on RNA-seq analysis. HER2 mRNA levels do not significantly change between TzS vs TzR tumors. By contrast, mRNAs identified in our analysis show striking upregulation in TzR tumors (red) vs all or the majority of TzS tumors (blue).
To further test the role of S100P, we examined S100P mRNA and protein levels in the BT474 cell line utilized for RNA-seq studies, and a second independently generated TzS vs TzR BT474 line, and SKBR3 (three isogenic TzS vs TzR cell lines total). We found S100P to be upregulated at both the mRNA (Fig. 4-5 B) and protein levels (Fig. 4-5 C). Lastly, S100P is upregulated in a cohort of breast cancer cell lines as compared to normal human mammary epithelial cells (data not shown), demonstrating that S100P is upregulated from normal to tumors, and further upregulated in resistant tumors.
Figure 4-5. S100P is upregulated at the mRNA and protein levels in TzR cells. (A) S100P is highly upregulated in HER2-positive tumors as compared to matched normal breast tissues (TCGA RNA-seq data). (B) S100P mRNA is elevated in TzR vs TzS BT474 cells. (C) S100P protein is elevated in TzR vs TzS cells as measured by western blot analysis.
To determine if inhibition of S100P would be sufficient to reverse TzR, we tested five distinct shRNAs against S100P, and identified two shRNAs that are effective at knocking down S100P (data not shown). To determine if the depletion of S100P would re-sensitize TzR cells to trastuzumab, we tested the proliferation of TzR BT474 cells that were infected with either shGFP or shS100P in mock vs trastuzumab treatments. BT474 cells infected with shGFP showed no response to trastuzumab (Fig. 4-6 A); by contrast, the proliferation of TzR BT474 cells infected with shS100P was significantly impacted when exposed to trastuzumab (Fig. 4-6 B-C). We also examined the rate of apoptosis of TzR cells in response to shRNA-mediated depletion of S100P, and found a significant increase in apoptosis in TzR cells treated with shS100P vs shGFP (data not shown).

Collectively, these data demonstrate that inhibiting S100P re-sensitizes TzR cells to trastuzumab.
Figure 4-6. Knock down of S100P partially reverses TzR in cell culture. Cell proliferation analysis of TzR cell lines using the MTS colorimetric assay demonstrated that cells treated with either mock vs trastuzumab (10 μg/ml) for 96 hours show decreased proliferation rate when S100P is knocked down with two independent shRNAs (B, C) vs a negative control shRNA targeting GFP (A).
4.3 Discussion:

Trastuzumab was originally developed and utilized as an adjuvant therapy for HER2-positive metastatic breast cancer, and currently as an adjuvant therapy in early-stage HER2-positive patients (Brufsky, 2014; Perez et al., 2011; Slamon et al., 2011, 2001). The clinical utilization of trastuzumab has revolutionized the treatment of HER2-positive breast cancer patients, however, a significant percentage of patients (~25%) do not respond to trastuzumab or develop resistance to it, by largely unknown mechanisms (Baretta et al., 2013; Vu and Claret, 2012). Because of the immense therapeutic benefits of trastuzumab to HER2-positive breast cancer patients, it is critical to identify the mechanisms of TzR. Although a number of models of TzR have been proposed, none has been experimentally demonstrated. In this manuscript, we have taken an unbiased genome-wide approach to identify genes and pathways that may drive and/or contribute to TzR. By utilizing next generation RNA-seq (RNA-seq) of tumors and cell lines that are either TzS or TzR, we identified key genes and pathways that potentially contribute to TzR. By utilizing tumor biopsies from patients that are either sensitive or resistant to trastuzumab, we were able to obtain clinically relevant data.

Many of the mRNAs that we identified as associated with TzR have been previously shown to be dysregulated in various types of cancer. MATN2 and SCGB2A2 are up-regulated in ovarian cancer (Fischer et al., 2014; Januchowski et al., 2014); S100P is up-regulated in prostrate cancer, invasive ductal carcinomas, and pancreatic cancer (Arumugam et al., 2005). LDHB is a
biomarker for triple negative breast cancer (Lehmann and Pietenpol, 2014), and TNC is associated with tamoxifen resistance in breast cancer (Helleman et al., 2008). Although the exact roles of these genes in TzR is yet to be elucidated, previous studies of these genes in cancer should provide some preliminary insights. For example, the role of S100P in activation RAGE-mediated Ras/MAPK signaling provided clues into its role in TzR (Arumugam et al., 2006; Xie et al., 2013).

S100P emerged as a top candidate for functional studies as it showed strong association with TzR both in vivo and in vitro studies. It is highly upregulated in many cancer types, and it belongs to a family of proteins, the S100 family, with well documented roles in tumorigenesis (Bresnick et al., 2015). S100 family members, including S100P, function as homodimers with a Ca^{2+}-binding EF-hand motif (Donato, 2001). Calcium binding to S100P promotes a conformational change exposing hydrophobic residues, which enables the interaction with protein binding partners (Schafer et al., 1995). S100P acts as both an intracellular and extracellular signaling molecule, and has been observed in the nucleus, cytoplasm, and extracellular matrix (Jiang et al., 2012). Previous studies of S100P in breast cancer have demonstrated that the expression of S100P correlates with breast cancer progression and decreased patient survival (Guerreiro Da Silva ID, Hu YF, Russo IH, Ao X, Salicioni AM, Yang X, 2000; Mackay et al., 2003; Schor et al., 2006; Wang et al., 2006). Also, S100P in tumorigenesis has been well documented in pancreatic cancer. These studies have demonstrated that S100P exert its oncogenic activities via interaction and
activation of the receptor RAGE (Arumugam et al., 2006; Xie et al., 2013). The extracellular ligand-binding domain of RAGE is known to bind to a number of ligands, including S100P, to initiate downstream signaling pathways that promote cell proliferation, survival, and motility (Kierdorf and Fritz, 2013; Taguchi et al., 2000). Blocking S100P interaction with RAGE is sufficient to significantly impact the growth of tumors (Arumugam et al., 2006; Taguchi et al., 2000). These findings suggest that S100P-RAGE interaction could be therapeutically blocked to reverse TzR, at least in a subset of HER2+ patients.

Our study also led to the identification of long intergenic non-coding RNAs (lincRNAs) that are dysregulated in TzR tumors and cell lines. We and others previously demonstrated critical roles of lincRNAs in regulating gene expression at the transcriptional and post-transcriptional in mammalian and other systems (Carrieri et al., 2012; Gomez et al., 2013; Gong and Maquat, 2011; Grote et al., 2013; Hou et al., 2014; Khalil et al., 2008, 2009; Merry et al., 2015c; Moran et al., 2012; Xing et al., 2014). A major mechanism utilized by lincRNAs to regulate gene expression is via the recruitment of chromatin-modifying complexes and other effector proteins to specific gene loci (Chu et al., 2011; Gomez et al., 2013; Gupta et al., 2010; Huarte et al., 2010; Khalil et al., 2009; Klattenhoff et al., 2013; Merry et al., 2015c; Schmitz et al., 2010; Tsai et al., 2010). Thus, such lincRNAs, when they become dysregulated in cancer, could have profound effect of the epigenome (Merry et al., 2015c). By targeting such lincRNAs, we could potentially modulate the epigenetic landscape of cancer cells to a normal state, at least partially (Moskalev et al., 2012). A key advantage of targeting lincRNAs...
is their tissue-specific expression, potentially reducing negative side effects. In future studies, we will assess the role of the identified top candidate lincRNAs in TzR, and their potential utility as therapeutic targets.

In summary, we have utilized both clinical HER2-positive tumor samples and cell culture models of trastuzumab-resistance to identify key coding and non-coding genes associated with acquiring resistance to trastuzumab. Our findings should provide a small list of potential therapeutic targets that can be experimentally tested. Also, our functional analysis of S100P provides novel insights into selective advantages that are acquired by resistant tumors.
4.4 Material and Methods

**Clinical Trial Information:** Institute: Brown University Oncology Group (Brown University, Yale University, Cedar-Sinai Center), PI: William Sikov MD; Correlative Science PI: Lyndsay Harris MD; BrUOG Study ID: BR-211B; Clinical Trials.gov ID NCT00617942.

**RNA isolation and next generation RNA sequencing (RNA-seq) of clinical trial tumor samples:** Patient biopsies cores were flash frozen for processing. RNA was isolated from samples using AllPrep (Qiagen) and amplified using Ovation RNA-seq System (NuGen). Library preparation of samples was completed using TruSeq v3 (Illumina) and sequenced on the Illumina HiSeq 2500 platform. Cufflinks V2.0.2 was used to annotate the aligned reads to human mRNA and lincRNA. Expression values were calculated as FPKM (fragment per kilobase of exon per million of mapped fragments). For our analysis, transcript expression was defined as if the FPKM value across either all non-responders (TzR) or all responders (TzS) samples were ≥ 1.0 for mRNAs and ≥ 0.25 for lincRNAs. The mean expression was calculated for TzR and TsS and the fold change was calculated as TzR/TzS to identify differentially expressed transcripts. Transcripts were defined as differentially expressed if the fold change was ≥ 2.0 or ≤ 0.5. Heatmaps were generated in R using heatmap.2 with Z scores scaled by row using standard Z score calculation of log fold change.
**Cell lines:** TzS cells were treated with 10μg/ml trastuzumab for at least six weeks to acquire TzR. Human breast cancer cell lines used in this study were grown in Hybri-Care Medium (ATCC_46-X™) supplemented with 10% fetal bovine serum (Bioexpress) and 100 units/ml penicillin and 100 µg/ml streptomycin (Life Technologies) at 37°C with 5% CO₂. Trastuzumab-resistant cell lines are maintained in 10 µg/ml of trastuzumab.

**Cell proliferation assay:** Cells were plated in a 96-well plate at 5000, 7500, and 10,000 cells/well with 4 replicates per density. The cells were evaluated for cell number using Operetta® High Content Imaging System (PerkinsElmer) imaging system with Harmony® High Content Imaging and Analysis Software (PerkinsElmer). Time points were taken at 0-, 24-, 48-, 72-, and 96-hour treatment time points. Percentage change normalized to the 0 time point was calculated for each well and then averaged for each sample. Cell proliferation assays were also performed using CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega) as indicated by manufacturer's instructions and read as absorbance at 490nm.

**Western blot analyses:** Protein lysates were prepared with Laemmlili Sample Buffer (BioRad) and separated on a 4-20% gradient SDS-PAGE Mini-Protean TGX™ Gels (BioRad). Gels were transferred to nitrocellulose membrane (Thermo Scientific) and probed with primary antibodies overnight at the following dilutions: anti-actin (Ambion, AM4302, 3.1 mg/ml) 1:1000, anti-HER2/Erb2 (Cell
Signaling, 2242S) 1:1000, anti-HER2-phospho (Fisher, MS-1072-P0, 200 μg/ml) 1:500, anti-p44/42 MAPK (Erk1/2) (3A7) (Cell Signaling, 9107S) 1:1000, anti-phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (Cell Signaling, 4370P) 1:500, and anti-H3 (Cell Signaling, 9715S) 1:2000. Secondary antibodies anti-mouse HRP (Thermo Scientific, 32230) and anti-rabbit HRP (Abcam, ab6721) were used at 1:10,000 dilutions. HRP was activated using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) for autoradiography.

**RNA isolation from cell lines:** RNA was isolated from cell lines using RNeasy® Mini Kit (Qiagen) according to the manufacturer’s protocol with an addition of DNase (Qiagen) treatment step after the first wash to remove DNA contamination.

**Next generation RNA-sequencing:** RNA isolated from cell lines was assessed for quality using BioRad Experion with an RNA integrity number (RIN) ≥ 8 as threshold for high quality, suitable for RNA-sequencing. Library preparation was performed using Scriptseq™ Complete Gold (Human/Mouse/Rat) (Illumina) and sequenced on Illumina HiSeq 2500 with six samples run per flow cell. The 100 bp paired-end strand-specific sequences were mapped to the human genome release hg19 using TopHat with 2 mismatches allowed for full-length reads. Raw reads were mapped to human mRNAs annotated in the RefSeq database and lincRNAs annotated in Cabili et al. (Cabili et al., 2011) using Cufflinks V2.0.2. Differential expression analysis performed after normalizing the values to the
total mapped reads in each sample. Expression values were calculated as FPKM (fragment per kilobase of exon per million of mapped fragments). Transcripts were considered expressed if FPKM values across either all TzR samples or all TzS samples were \( \geq 1.0 \) for mRNA and \( \geq 0.25 \) for lincRNAs. The mean expression values were calculated for expressed transcripts followed by fold change (TzR/TzS) with statistical significance calculated using a paired t-test. Transcripts were defined as differentially expressed if the fold change was \( \geq 2.0 \) or \( \leq 0.5 \). Heatmaps were generated in R using heatmap.2 with Z scores scaled by row using standard Z score calculation of log fold change.

**Quantitative Real-time PCR (qRT-PCR):** RNA was converted to cDNA using RNA to cDNA EcoDry™ Premix Random Hexamers (Clontech). Primer pairs were designed using primer3 software (ref: Untergrasser et al., 2007) with most spanning exon-exon boundaries. S100P forward 5’-CTCACTGAAGTCCACCTGGG-3’, S100P reverse 5’-GAAGGAGCTACCAGGCTTCC-3’, GAPDH forward 5’-GAGTCAACGGATTTGGTCGT-3’, GAPDH reverse 5’-TTGATTTTGAGGGATCTCG-3’. Maxima SyBr Green/ROX qPCR Master Mix (Thermo Scientific) was used for RT-qPCR. A comparative \( C_T \) quantitation was performed with a hold stage of 50°C for 2 min and 95°C for 10 min followed by 40x cycles of 95°C for 15 sec and 60°C for 1 min, and finally a melt curve at 95°C for 15 sec, 60°C for 1 min, and a ramp to 95°C at 0.3°C increments. Analysis was done using the \( 2^{\Delta\Delta C_T} \) method with GAPDH as the reference gene (Livak and
Apoptosis assay: Cells were plated 5000 cells/well in a 96-well plate. After 5 hours, the media was replaced and trastuzumab (10 μg/ml) was added. After 48 hours, Caspase-Glo® 3/7 Assay (Promega, G8091) reagent was added to the wells. The plate was incubated for 1 hr at room temperature and the luminescence was read by spectrophotometer. The luminescence reading was normalized to the mock treatment of each cell line to give relative apoptosis.

Acknowledgements: We would like to thank Megan Forrest, Allison Cohen, Jenny Parvani, and Kristy Miskimen.
Chapter 5: Discussion

My overall study identified IncRNAs in colon and HER2-positive breast cancer that may play a role in tumorigenesis. Each individual study of IncRNAs in colon cancer, HER2-positive breast cancer, and trastuzumab resistance in HER2-positive breast cancer has its own merit and contributions to the advancement of the IncRNA-cancer field. I will discuss the importance of each of these studies in the following subsections, but first I will examine my work as a whole.

Our lab has generated defined lists of IncRNAs in colon and HER2-positive breast cancer that others scientists can use in their own studies. In addition to the IncRNAs that we identified, I also described our integration analysis based on our HER2-positive breast cancer studies. The protocol that I developed relies on both in vitro cell culture data as well as in vivo patient data. The patient data allows us to gain a more accurate representation of the disease in patients. This is important because while these studies are in early stages, the overall goal is to identify new treatment targets for patients. With this goal in mind, it is imperative that we understand the disease that is relevant in patients and not just a cell culture model of disease. In these early stages of study, it is important to also include the cell culture model of disease because the basic functional and mechanistic studies of identified IncRNAs will be done in cell culture. The integration of both data sets allows us to identify genes that are relevant to both systems. The integrative analysis could be useful to the study of...
other diseases where there is access to patient disease samples and cell culture models of the disease. While we did not use patient RNA-sequencing data to identify IncRNAs in the colon cancer study, we did perform the follow-up studies in patient-derived cell lines, which allowed us to identify and study a patient-relevant IncRNA.

The long-term endpoint of my work is to identify new therapeutic targets to treat these cancers. The standard of treatment for dysregulated mRNAs is to target the protein product, but since IncRNAs function as RNA molecules this cannot be done. The design of IncRNA-based therapies is only in early development stages, but could be very useful for the treatment of disease. LncRNAs have been shown to be tissue-specific, therefore, targeting a lncRNA could potentially only effect the diseased tissue. In addition, since IncRNAs often regulate more than one gene, a treatment targeting a particular lncRNA could orchestrate a pervasive change in the cell. For example, if we find that decreased expression of DACOR1 leads to differential methylation and the advancement of tumorigenesis, by increasing the levels of DACOR1 we could possibly re-establish the DNA methylation patterns of a normal colon cell.

For the development of lncRNA-based therapies, we need to consider that we may need to increase or decrease the activity of the lncRNA. One therapeutic option for decreasing the activity of a lncRNA is through inhibition of binding with its protein partner through the use of antisense oligonucleotides (ASOs). The use of ASOs in cancer research is hindered by the problems of effective delivery to the tumor site, stability of the molecule, and cell toxicity, but improvements are
continuously being made to the chemistry to alleviate these complications (Esteller, 2011; Huarte, 2015). Additional agents that could be used to decrease the abundance of a lncRNA is through the use of gapmers or RNAi. Gapmers are also antisense oligonucleotides, but work by recruiting RNase H to cleave the RNA-DNA hybrid (Devaux et al., 2015). Finally, identification of a small molecule that interacts with a lncRNA to inhibit either its interaction with a binding partner or its active secondary structure could be used to decrease activity. Through the development of these technologies, we could be able to inhibit the activity of an overexpressed lncRNA in cancer.

In contrast, some lncRNAs may contribute to disease by having lower expression levels than normal. In that case, we need treatment strategies that can restore lncRNA expression. The replacement of functional cDNA copies through the use of viral vectors could be used, but there are many complications to this approach such as immune response activation, toxicity, and further mutagenesis (Wahlestedt, 2013). Another option is to use shorter versions of the lncRNA with only the function sequences intact to mimic the endogenous activity of the lncRNA (Ling et al., 2013). The shorter form of the lncRNA may be more suitable for delivery, but may not be as stable.

In addition to lncRNAs being utilized as therapeutic targets, lncRNAs could also be used as biomarkers for the identification of cancer through non-invasive methods. Ideally, lncRNA biomarkers would be detectable and stable in blood plasma. Since lncRNAs are often tissue-specific or disease-specific, they could be useful diagnostic tools. For a more detailed discussion of the potential
future study for the IncRNAs that I identified, I will look at these studies individually.

5.1 Follow-up studies on DACOR1

In my first study (Merry et al., 2015c), the goal was to identify IncRNAs that interact with DNA methyltransferase, DNMT1. I identified that IncRNA DACOR1 binds to DNMT1 and that deregulation of this IncRNA contributes to aberrant DNA methylation and gene expression in colon cancer. This is an important finding for several reasons. Firstly, I identified that IncRNAs may influence DNMT1 genomic occupancy and/or activity. When these IncRNAs are deregulated, they could further contribute to cancer tumorigenesis by altering the methylome. While the specific mechanism is not known, the overlap between DACOR1 genome occupancy and differentially methylated regions provides us with a basis to continue this study. Finally, the data showing that DACOR1-induced patient-derived colon cancer cells show a reduced tumorigenesis phenotype and enhanced tumor suppressor expression compared to the control cells lead us to explore the possibility of harnessing the DACOR1-DNMT1 interaction for therapeutic treatment.

My study has enhanced the understanding of the role that IncRNAs play in colorectal cancer and the regulation of the methylome. While I identified numerous IncRNAs that interact with DNMT1, I focused on understanding the role of DACOR1. The study has provided insight into the possible mechanism of DACOR1 in the progression of colon cancer, but in order to develop therapeutic
strategies that target DACOR1, additional research must be performed to clearly define the role that this IncRNA plays in the regulation of gene expression and DNA methylation.

Mechanism and function of DACOR1

In order to test the hypothesis that DACOR1 recruits DNMT1 to specific genomic sites to suppress colon tumorigenesis, I will first start by expanding our DNA methylation analysis. My original study used a 450K DNA methylation array to determine differentially methylated regions (DMR) and while this DNA methylation array was important for the initial inquiry, this method of methylation analysis has some limitations. These limitations can be addressed through a more comprehensive methylation analysis. As with any array, the 450K DNA methylation array has pre-designed probes that allow for only a limited number of site inquiries. In addition to the restricted number of sites that are assessed, studies have shown that there are also problems with cross-hybridization with non-target DNA (Chen et al., 2013). While whole genome bisulfite sequencing allows for the best genome coverage of the methylation sequencing options, reduced representation bisulfite sequencing (RRBS) has better coverage of the CpG islands, especially in repetitive regions (Bock et al., 2010; Harris et al., 2010). RRBS is a methylation sequencing technique that uses a restriction enzyme to target the CpG sites, regardless of methylation status. These fragments then undergo bisulfite conversion and next-generation sequencing. RRBS allows for more CpG sites to be assessed at greater depth than can be
with the methylation array while focusing on the important CpG island sequences that whole genome bisulfite sequencing can miss.

The RRBS analysis will be performed in several DACOR1- and control-induced patient-derived colon cancer cell lines to define continuity between different genetic backgrounds of patients. Ideally, I would perform the contrasting RRBS analysis in which I perform knockdowns in normal colon cells that express higher levels of DACOR1 using stably transduced control and DACOR1 shRNA lentiviruses. Unfortunately, normal colon cells do not propagate in cell culture so I cannot perform knockdown experiments on these cells. Alternatively, I could use several of the identified patient-derived colon cancer cell lines that have higher expression than other colon cancer lines, such as V867 or V868 (Fig 2-2 D) to perform the knockdown experiments.

Lastly, to further define the recruitment of DNMT1 by DACOR1, I will determine differential binding of DNMT1 through chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) (Mohammad et al., 2012). Several DACOR1-induced patient-derived colon cancer cell lines will be utilized in addition to the control cell lines to analyze the differential binding of DNMT1 with low levels of DACOR1, as seen in the control cells, and high levels of DACOR1, as seen in the induced cells. To amplify the relevance of these analyses, I will perform the ChIP-seq in populations of S-phase cells as sorted through flow cytometry, when DNMT1 expression peaks. The differential binding analysis of the occupancy of DNMT1 on the genome can be compared to the ChIRP-seq data that I previously collected on the occupancy of DACOR1 on the
genome to define the overlap between \textit{DACOR1} binding and DNMT1 binding in the presence of higher levels of \textit{DACOR1}.

Once I have completed these sequencing experiments, I will integrate our \textit{DACOR1}-induced sequencing data to identify target locations of the recruitment of DNMT1 by \textit{DACOR1} to genome. The RNA-seq data defines the target genes through differential expression analysis; RRBS will indicate promoters and gene bodies with differential methylation; and ChIP-seq and ChIRP-seq will define the occupancy of DNMT1 and \textit{DACOR1}, respectively. Together these data will allow us to define the role that \textit{DACOR1} plays in DNA methylation and subsequent gene regulation.

I have found that induction of \textit{DACOR1} reduced colony formation \textit{in vitro}, which indicates reduced tumorigenicity. To examine the tumor suppressor function of \textit{DACOR1} \textit{in vivo}, I will perform xenografts in nude mice using patient-derived colon cancer cell lines induced with a control or \textit{DACOR1} lentivirus (Xiang et al., 2014). I expect that mice injected with the \textit{DACOR1}-induced cell lines will have smaller tumors than those mice injected with the control cell line. This experiment will allow me to assess the tumor suppressor function of \textit{DACOR1} \textit{in vivo}, which is necessary for future studies that focus on identifying therapeutic agents to enhance function of \textit{DACOR1}.

\textit{Long-term studies}

The experiments that I just described are the initial experiments for the continuation of the study of \textit{DACOR1} and DNMT1 in colon cancer. These
experiments are crucial to the development of the larger project, as they will establish the functional and mechanistic knowledge of DACOR1 in colon cancer. As previously mentioned, the long-term objective of this study is to determine if we can target a IncRNA to restore proper DNA methylation in colon cancer. Through the identification and study of DACOR1, I can determine if this IncRNA would be an effective target. One major obstacle to the development of future treatment options is that DACOR1 is repressed in colon cancer, meaning to restore proper function of the molecule we need to increase the expression level. Increasing expression of a IncRNA is a more challenging task than reducing the expression, but as the field on a whole advances we will likely have better options in the future. Also, more advanced mechanistic understanding the role of DACOR1 in DNA methylation may provide us with a better alternative in which to restore DACOR1 function.

DACOR1 is just one IncRNA of 148 that we identified to interact with DNMT1. It cannot be overlooked that more than one IncRNA may be needed to effect the full restoration of the normal DNA methylation pattern. We still do not understand the sequence specificity that allows certain IncRNAs to interact with certain binding partners, but as we learn more, we could decipher this interaction. At that point, I could work to identify a group IncRNAs that have a role in DNA methylation maintenance in normal cells.

There are still many unanswered questions in understanding the interaction of IncRNAs and DNMT1 in colon cancer. As we takes steps to learn
more, this will better inform us of the ideal next steps to take as we work toward finding a treatment target for colon cancer.

5.2 Investigation of HER2-regulated genes

In my second study, I used an integrative analysis to identify genes -- mRNAs and lincRNA -- that are downstream of the HER2 receptor in breast cancer. While the study has presented exciting new potential for HER2-positive breast cancer research, there are limitations to the study. The clinical data is an important aspect of this study as it allowed me to identify patient-relevant gene expression changes. I utilized a clinical trial originally designed to identify early indicators of response to trastuzumab, and as it was intended for another study, I did not have control over the design. Of the 50 biopsied tumor pairs, only 13 were sequenced due to the high cost of RNA-sequencing at the time of the study. Ideally, for a more complete study, more tumor pairs should have been sequenced. The additionally sequenced samples would allow for further classification based on the presence or absence of the other receptors in the tumor, such as ER and PR. This would give more statistical power to the differential gene expression analysis and also allow the identification of genes that may only have altered expression depending on the presence of ER or PR.

An important step to take in the overall study of HER2-positive breast cancer is the switch from a two-dimensional (2D) culture to a three-dimensional (3D) culture. The RNA-seq analysis of the in vitro model of this study was originally performed as a two-dimensional (2D) culture. 2D cell culturing, which
allows cells to grow on a flat, plastic surface, has been pervasively used in cancer research. Advancements in cell culture methods have given rise to a three-dimensional (3D) system, which is more physiologically accurate than 2D. The 3D system allows for the formation of a microenvironment that models that of a tumor such as nutrient gradient within the tumor, intercellular cellular signaling, drug exposure, and cellular proliferation (Lovitt et al., 2014; Weigelt et al., 2014). The 3D culture environment is achieved with the use of Matrigel, which is a combination of extracellular matrix proteins that forms a gelatinous substance (Benton et al., 2009). When the cells are seeded in the Matrigel, they can form spheroid colony. Previous studies have shown that drugs perform differently between the 2D and 3D cultures. For example, HER2-positive breast cancer cell line, SKBR3, was more resistant to trastuzumab in 3D Matrigel cultures than in 2D (Weigelt et al., 2010). Due to the more accurate reflection of the tumor microenvironment, it is important that drug screening rely on the use of the 3D Matrigel culture method.

Identified mRNAs

The common treatment for HER2-positive breast cancer is the combination of chemotherapy and adjuvant trastuzumab. This has shown to be an effective combination in combating HER2-positive tumors. Unfortunately, trastuzumab resistance can impede full recovery for patients. In efforts to anticipate the need of additional treatment options, it is important that further drug targets be investigated. Through the integrative analysis, I have identified
35 mRNAs that are dysregulated after the inhibition of HER2 in HER2-positive breast cancer. These are ideal candidates for HER2-positive therapeutic targets as they are downstream of HER2 and may provide an additional option for HER2 pathway inhibition. Drugs that disrupt the function of RRM2 and TOP2A have already been approved for clinical treatment (Merry et al., 2015b). The 33 remaining protein candidates will first be tested in a restricted high-throughput drug screen to identify drugs that target specific proteins.

Our list of 33 genes that are dysregulated after inhibition of HER2 and validated as being dysregulated in HER2-positive tumors vs. normal allows me to limit the potential drugs that I screen. A brief literature search reveals that six of the genes we identified have a total of sixteen documented inhibitors. In a high throughput screen of these drugs, we will measure cell proliferation as an indicator of the drug effectiveness (Hongisto et al., 2013; Lapin et al., 2014; Liu et al., 2013; Moffat et al., 2014). I will use CellTiter 96® Aqueous One Solution Cell Proliferation Assay that measures the formation of formazan, which is produced by metabolically active cells exposed to MTS tetrazolium. The quantity of formazan is measured through absorbance at 490 nm and is directly proportional to the number of living cells (Barltrop et al., 1991).

Due to the small number of drugs that I will be screening, I will be able to employ many variables to determine the best drug across multiple settings. There are eleven commonly used HER2-positive breast cancer cell lines, of which our lab has four. By expanding our cell lines to six or more lines, I can have a range of characteristics, which vary on ER status, PR status, and tumor
source (Neve et al., 2006). This will help me identify drugs that work to inhibit growth in a wider population of tumors. I will also use normal human mammary epithelial cells (HMEC) to determine if the tested drugs affect normal cell growth.

In addition to using a variety of cell types, I will also use varying concentrations of drugs with and without the addition of trastuzumab. The addition of trastuzumab is important to assess if these drugs may potentially work better as a dual therapy with trastuzumab. Cells will be grown in the absence of drugs as a negative control to standardize the assay measurements. Finally, doxorubicin, a chemotherapeutic agent will be used as a positive control of cell growth inhibition. In the event our limited screen does not identify any drug or drug combinations that limit HER2-positive cell culture growth, I will need to expand the high-throughput screen to compounds that are not associated with identified genes through the use of the Small Molecule Drug Development Core at Case Western Reserve University. The future research targeting the mRNAs that I identified as downstream of HER2 in our study will allow us to potentially discover new and more effective treatments for early-stage HER2-positive breast cancer.

**Identified lincRNAs**

To begin our study on the function of lincRNAs in cancer, I will validate the 3 candidate lincRNAs (linc-STARD6, linc-GJA1-2, linc-SLC39A10-10) in 3D cell culture models across several HER2-positive breast cancer cell lines. Two types of validation experiments will be performed; first, I will validate the dysregulation
of the lincRNAs after HER2 siRNA knockdown using the same protocol as I used for the RNA-seq study. In addition, we will also validate changes in expression after a 48 hour exposure of trastuzumab vs. a mock exposure. Those lincRNAs that validate highly across cell lines and HER2 inhibition experiments will then be chosen for further study.

Validated lincRNA candidates will be genetically manipulated to increase or decrease expression to determine how this will affect tumorigenesis through the use of phenotypic assays. To study linc-STAR6-2, which showed upregulation in tumor samples and downregulation with HER2 inhibition, I will use knockout techniques in HER2-positive breast cancer cell lines. As siRNA knockdowns are not always efficient against lincRNAs, possibly due to their nuclear localization, we will utilize the CRISPR/Cas9 genomic editing system to delete the linc-STAR6-2 region in the genome. CRISPR/Cas9 uses RNA-guided nucleases to cleave sequence specific sites in the genome (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013; Ran et al., 2013). Linc-GJA1-2 and linc-SLC39A10-10 are downregulated in the tumor and upregulated with HER2 inhibition. The strategy for genetic manipulation is to increase expression through the creation of stable lincRNA expression cell lines achieved through lentiviral transduction of a PGK promoter vector.

Once these cell lines are established with the proper control cell lines, I can perform cancer phenotypic assays to determine if the change in expression level of these lincRNAs causes a change in the 3D model of tumorigenesis. These assays include cell proliferation, cell invasion, and apoptosis. If these
assays show a difference when \textit{linc-STAR}D6-2 is knocked out, or \textit{linc-GJA}1-2 or \textit{linc-SLC}39A10-10 are overexpressed, then I will study these lincRNAs further to identify the underlying mechanism of action. Although most lincRNAs work in the nucleus, I will verify lincRNA candidate localization through the use of RNA \textit{in situ} hybridization of primary mammary tumors or primary mammary tissue (Merry et al., 2015c). As we learn more about the potential mechanism of these lincRNAs, I can also perform RNA-seq on the genetically manipulated cell lines to uncover gene expression changes that contribute to phenotype. Additionally, I can perform RNA pull-down experiments to determine the protein or genomic binding partners of candidate lincRNAs.

As the functional roles of many lincRNAs are currently not known, my future studies will invest much time into the basic understanding of the lincRNAs. As we learn more about the role of these lincRNAs in HER2-positive breast cancer, I can better define the mechanistic experiments that will be performed. While I have restricted our initial IncRNA analysis to lincRNAs due to previous research that identified their role in disease and alterations to the epigenome (Kaneko et al., 2014; Khalil et al., 2009; Mattick, 2009; Niland et al., 2012; Tsai et al., 2010), I can expand our analysis to all categories of IncRNAs. As our understanding of lincRNA in HER2-positive breast cancer increases and I identify their mechanistic contributions to disease, I will work to develop lincRNA-targeted therapies to treat HER2-positive breast cancer.
5.3 Understanding the role of S100P in trastuzumab resistance

In my final study, I examined gene expression changes that occur with trastuzumab resistance in HER2-positive breast cancer through the use of an integrative analysis of in vivo and in vitro samples. I identified mRNAs and lincRNAs that are differentially expressed in trastuzumab-resistant (TzR) HER2-positive breast cancer and narrowed our focus to one gene, $S100P$. While $S100P$ is an mRNA, I chose to continue my research on this gene because of its previous association with other cancers and the promising outlook of it as a therapeutic target in cancer.

My research begins to explore the mechanistic role of S100P as compensatory signaling of the HER2 pathway. In pancreatic cancer studies, S100P has been shown to interact with RAGE to initiate RAS/MEK/MAPK signaling, and preliminary studies suggest that this same mechanism is utilized by tumors to bypass trastuzumab treatment of HER2-positive breast cancer. I also show in initial studies that the knockdown of S100P may help to reverse trastuzumab resistance in HER2-positive breast cancer. While more work needs to be done to determine the exact mechanism of S100P and its contributions to trastuzumab resistance in HER2-positive breast cancer, I have made progress in identifying new therapeutic targets for patients with TzR HER2-positive breast cancer.

While my approach allowed us to identify S100P and other candidate mRNAs and lincRNAs that may be associated with trastuzumab resistance in HER2-positive breast cancer, my results are limited by the fact that we only had
two TzR patient samples. Due to this, we were not able to calculate p-values for the clinical data to narrow our list of candidates. With a larger number of both TzS (trastuzumab-sensitive) and TzR clinical samples available, I could increase the power of the study to provide even more accurate results. Despite these limitations, I was able to identify both coding and noncoding genes that are associated with trastuzumab resistance in HER2-positive breast cancer.

Cromolyn has been identified to bind S100P to inhibit its interaction with RAGE and subsequent downstream signaling. The traditional use of cromolyn is for the prophylactic treatment of asthma. The use of cromolyn in S100P expressing pancreatic cell lines has been shown to significantly inhibit cell proliferation, invasion, and NFκB activity. It also was shown to inhibit tumor growth in a mouse model (Arumugam et al., 2006; Taguchi et al., 2000). While the identification of inhibitory effects of cromolyn on pancreatic cancer hints at a possible treatment for S100P overexpressing cancers, I still need to determine if S100P acts through RAGE in HER2-positive breast cancer. My findings show that knocking down S100P can reverse trastuzumab resistance in cell models, but more research needs to be done to determine if cromolyn or S100P antibody treatment would be sufficient to inhibit the function of S100P in HER2-positive breast cancer.

Functional role of S100P

The long-term goal of my study is to identify a therapeutic target to reverse trastuzumab resistance in HER2-positive breast cancer. Since the study is in the
initial phases of target identification, I first need to validate our phenotypic results in an additional cell line to verify that the results I see are consistent across multiple isogenic TzS and TzR cell lines. In the BT474 TzS and TzR cell lines, I saw reduced cell proliferation and increased apoptosis in cells that expressed shS100P and were treated with trastuzumab when compared to mock treatment. These phenotypic assays will also be performed in a second pair of isogenic TzS and TzR cell line, SKBR3, with lentiviral shS100P and shGFP control, with and without trastuzumab. If I see the same results in two TzR HER2-positive breast cancer cell lines, this gives me the confidence to move onto mouse studies. In addition to the cell proliferation and apoptosis assays, I can also perform invasion assays to determine if the reduction of S100P in combination with trastuzumab is able to inhibit cell migration and metastasis. The use of these phenotypic assays in multiple cell lines will provide the data to determine if the inhibition of S100P is able to re-sensitize TzR HER2-positive breast cancer cells to trastuzumab.

Once we establish the phenotypic consequences of S100P inhibition in vitro, we can look at the tumor growth in vivo. To study S100P inhibition by shRNA in vivo, I will perform xenografts in nude mice with an implanted estrogen-release pellet. We will be testing shS100P vs. shGFP control expressing BT474 TzR cells with and without treatment with trastuzumab. All cells will stably express luciferase, which will be used to measure tumor growth through luminescence imaging (Mezzanotte et al., 2010). Cells will be counted and 5 x 10^6 cells will be mixed with Matrigel before they are orthotopically injected into the mammary fat pads. Trastuzumab will be administered intraperitoneally twice
a week at the concentration of 10 mg/kg; PBS will be used as the control vehicle (Arpino et al., 2007; Fridman et al., 2012; Zhang et al., 2011). The mice will be imaged once a week for tumor growth and after 6-8 weeks the mice will be sacrificed and the tumors excised for measurement. This data will allow us to learn if the inhibition of S100P is sufficient to re-sensitize cells to trastuzumab in mice.

**Mechanistic role of S100P**

While we are establishing the ability of S100P inhibition to alter the cancer phenotype in vitro and in vivo, I can work to determine the pathway by which S100P signaling occurs. Previous research in pancreatic cancer has established that S100P binds to the receptor, RAGE, leading to activation of the Ras/MEK/MAPK and other signaling pathways (Arumugam et al., 2004). The Ras/MEK/MAPK pathway is a key pathway that can also be activated by HER2 signaling. To test the hypothesis that the S100P-RAGE interaction leads to activation of Ras/MEK/MAPK, and thus compensates for HER2 inhibition by trastuzumab, I will perform western blot analyses on key proteins in this pathway.

My preliminary work has examined the effect of recombinant S100P on MAPK phosphorylation (p-MAPK) in TzS cells as follows: BT474 cells (TzS) were treated with trastuzumab alone or trastuzumab + recombinant S100P protein, and the levels of pMAPK, MAPK and histone H3 were measured by western blot. I found that trastuzumab treatment results in reducing p-MAPK levels, but the addition of S100P protein is sufficient to restore it (Figure 5-1 A). To further
expand on these preliminary results, I will determine specific details of treatment. All experiments will be done with serum starvation for 2 hours followed by treatment in low serum (1%). I will test S100P recombinant protein through a time trial to determine if and when the increase in p-MAPK signal is seen after S100P is administered without trastuzumab. Once I have established the timing of the pathway stimulation, I can rerun the BT474 TzS western blot (Figure 5-1 A) with mock, S100P, trastuzumab, and trastuzumab + S100P lanes to strengthen the results. These experiments can also be done in the SKBR3 TzS cell lines as well to establish the pattern across cell lines.

Additionally, I have also examined the effect of knocking down S100P on p-MAPK in TzR cells. I found that TzR cells harboring shGFP had minor reduction of p-MAPK when treated with trastuzumab, however, TzR cells harboring shS100P showed significant reduction of p-MAPK when treated with trastuzumab (Figure 5-1 B). I can also perform these experiments in the SKBR3 TzR cell lines.
Figure 5-1. S100P alters phosphorylation of MAPK. (A) In BT474 TzS cells, trastuzumab decreases p-MAPK and the addition of S100P restores some signal. (B) The addition of trastuzumab to BT474 TzR cell lines expressing shS100P causes a significant reduction in p-MAPK signal. H3 and HER2 were used as loading controls.
After I establish if S100P does alter the MAPK signaling pathway through assessment of the p-MAPK signal via western blots, I need to determine if this pathway is activated through the interaction of S100P with RAGE. The first step is to identify if S100P is directly binding to RAGE in HER2-positive breast cancer. I will use a series of co-immunoprecipitation experiments to determine if S100P and RAGE are interacting (Arumugam et al., 2004). Next, I need to try to block the interaction of S100P and RAGE. As discussed previously, cromolyn has been shown to inhibit the interaction of S100P and RAGE, unfortunately, high concentrations are needed to see an effect (Arumugam et al., 2013). Cromolyn is also able to bind to other proteins in the S100 family (Okada et al., 2002).

Long-term studies

As the use of immunotherapy in cancer treatments is continuing to advance, I would like to develop an antibody that targets S100P to inhibit its interaction with RAGE. Immunotherapy would be very useful in this case as S100P is part of a large family of proteins. By developing a neutralizing antibody that only targets S100P at its point of interaction with RAGE, I could inhibit its activation of the RAGE signaling pathway without blocking other functions or other S100 family proteins. After the development and testing of specificity of this antibody, I could continue our studies using the antibody with the hopes that it could one day be used clinical treatment of TzR HER2-positive breast cancer.

The above-mentioned studies all focus on the downstream aspects of S100P. It is also important to determine the causative reason for the increase of
S100P expression in TzE HER2-positive breast cancer. To identify the potential mechanism(s) that drive increased S100P expression in TzR tumors and cell lines, I will examine global changes in the epigenome. Specifically, I will analyze changes in histone H3 lysine 4 monomethylation (H3K4me1) and H3 lysine 27 acetylation (H3K27ac), two key histone modifications associated with active enhancers (Akhtar-Zaidi et al., 2012). Once I identify genome-wide changes in H3K4me1 and H3K27ac between TzS and TzR cells, I will specifically examine enhancers surrounding top candidate genes identified in our RNA-seq analysis, including S100P. These findings will indicate if trastuzumab resistance is wired at the epigenetic level.

Conclusion

My research identifying S100P and the study of its potential causative relationship with trastuzumab resistance in HER2-positive breast cancer is a promising path for the treatment of this cancer. My suggested future studies will concentrated on expanding our understanding of S100P in trastuzumab resistance in HER2-positive breast cancer with the intent of moving onto drug development to re-sensitizing HER2-positive breast cancer to trastuzumab.

5.4 Conclusion

The hallmark of cancer is abnormal cells growth, which can be reached through the dysfunction of many molecular mechanisms. This creates a great deal of complexity in defining what treatments will work to effectively treat the
tumor. For this reason, the field of cancer research has to be multifaceted in an attempt to combat this disease at different stages and through different mechanisms. My research contributes to the knowledge of colon and HER2-positive breast cancer transcriptomes through identification of lncRNA and mRNAs. The identification and initial functional studies are steps to understand the mechanism of action and the contribution to health and disease. While much work needs to be done to further define treatment targets, my work provides the necessary basis for the development of better and more accurate cancer treatment options.
Figure A-1. Optimization of RNA co-immunoprecipitation (RIP) conditions in HCT116 cells. (A) We carried out immunoprecipitation (IP) experiments using an antibody against U1-70K and confirmed that we are specifically IP U1-70K by western blot analysis. An antibody against IgG was used as a negative control. (B) We examined the RNA levels of U1 RNA in three independent biological replicates of U1-70K RIPs and found specific and strong enrichment in comparison to RIPs with a non-specific antibody (IgG) using three distinct endogenous controls. These results demonstrate that our RIP experimental conditions are optimized to detect specific protein-RNAs interactions in HCT116 cells.
Figure A-2. DACOR1 (TCONS_00023265) genomic locus on human chromosome 15. Snapshot of the UCSC genome browser showing the genomic region that encodes DACOR1 (Red arrow) on chromosome 15. DACOR1 has three exons. The nearest annotated protein-coding gene to DACOR1 is SMAD3.
Figure A-3. SMAD3 mRNA expression in colon tumors vs matched normal tissues. SMAD3 mRNA shows variable expression between colon tumors and matched normal tissues with no clear trend. On average, SMAD3 is slightly higher in normal tissues.
Figure A-4. Expression analysis of DACOR1 by qRT-PCR via cluster graph. Expression analysis of DACOR1 by qRT-PCR in normal colon vs. patient-derived colon cancer cell lines represented as a cluster graph.
Figure A-5. Expression analysis by qRT-PCR of DACOR1 induced cells by different promoters. Expression analysis by qRT-PCR of DACOR1 in normal colon, two colon cancer cell lines transduced with a control lentivirus, and same two cell lines transduced with a DACOR1 lentivirus.
Figure A-6. Western blot analysis of DNMT1 in transduced cell lines. V852 and V866 cells, which were transduced with either a control or DACOR1 lentivirus. No differences in DNMT1 protein levels are observed, suggesting that DACOR1 affects DNMT1-mediated DNA methylation by other mechanism(s).
Figure A-7. E-cadherin levels after DACOR1 induction. DACOR1 expression has no effect on E-Cadherin protein levels in colon cancer cells.
**Figure A-8. TJP1 expression by qRT-PCR.** (A) TJP1 mRNA levels do not change in response to DACOR1 induction in colon cancer cell lines, despite an observed increase in TJP1 protein levels suggesting a post-transcriptional regulation of TJP1 by DACOR1. (B) TJP1 mRNA levels are not significantly affected in majority of colon tumors vs. matched normal tissue (unchanged in 17/22 samples) supporting our observations in cell culture.
Figure A-9. DACOR1 induction results in decreased growth of colon cancer cells. A field view of colon cancer cells that were transduced with either a control or DACOR1 lentivirus. We quantified the effect of DACOR1 on the growth of colon cancer cells using colony formation assays.
Figure A-10. Induction of DACOR1 in additional lines. A-B. qRT-PCR expression analysis of DACOR1 in the colon cancer cell lines V703 and V425 post transduction with either a control or DACOR1 lentivirus (CMV promoter); C-D. DACOR1 has minor effects on colony formation in V703 and V425 cells. These are colon cancer cell lines that maintain some endogenous levels of DACOR1 expression (see figure 2D).
Figure A-11. DACOR1 induction with PGK promoter. A. qRT-PCR expression analysis of DACOR1 in normal colon, the colon cancer cell lines V866 with either a control or DACOR1 lentivirus (Pgk promoter); B. DACOR1 induction using a weak Pgk promoter is sufficient to reduce colony formation in the colon cell line V866.
Figure A-12. Induction of an oncogenic lncRNA, TCONS_00011938 (A) qRT-PCR expression analysis of the lncRNA TCONS_00011938 in the colon cancer cell line V481 with either a control or TCONS_00011938 lentivirus (CMV promoter). (B) TCONS_00011938 induction is sufficient to enhance colony formation in the colon cell line V481. This finding further supports the conclusion that the induction of the lncRNA DACOR1 effect on suppressing colon cancer growth is not simply due to a RNA toxicity effect.
Figure A-13. SMAD6 and PHGDH are up-regulated in colon tumors. The mRNA levels of SMAD6 and PHGDH are highly up-regulated in a cohort of 22 colon tumors in comparison to matched normal tissues (TCGA RNA-seq).
APPENDIX B

Figure B-1. Cancer-associated pathways are affected in response to HER2 knockdown in BT474 cells.
Figure B-2. Identification of 35 mRNAs that are affected in all three data sets
Bibliography


American Joint Committee on Cancer (2009). Colon and Rectum Cancer Staging.


wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. Cancer Res. 66, 8469–8476.


Jass, J.R. (2007). Classification of colorectal cancer based on correlation of
clinical, morphological and molecular features. Histopathology 50, 113–130.


HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. Cancer Res. 71, 6320–6326.


Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases
Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257.


Protoc. 7, 562–578.


