Post-transcriptional Regulation of Cancer Traits and Gene Expression in a Genetically Defined, Primary Cell-derived Model of Breast Tumorigenesis

by

Laura Simone Bisogno

University Program in Genetics and Genomics
Duke University

Date:_______________

Approved:

___________________________

Jack Keene, Advisor

___________________________

Jen Tsan Chi, Chair

___________________________

Christopher Nicchitta

___________________________

Debra Silver

___________________________

Kris Wood

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University Program in Genetics and Genomics in the Graduate School of Duke University

2017
ABSTRACT

Post-transcriptional Regulation of Cancer Traits and Gene Expression in a Genetically Defined, Primary Cell-derived Model of Breast Tumorigenesis

by

Laura Simone Bisogno

University Program in Genetics and Genomics
Duke University

Date: ______________________

Approved:

__________________________________________________________________________
Jack Keene, Advisor

__________________________________________________________________________
Jen Tsan Chi, Chair

__________________________________________________________________________
Christopher Nicchitta

__________________________________________________________________________
Debra Silver

__________________________________________________________________________
Kris Wood

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University Program in Genetics and Genomics in the Graduate School of Duke University

2017
Abstract

Post-transcriptional events are crucial determinants of gene expression, and aberrant expression patterns resulting from misregulation are evident in many pathological states. Cancer has traditionally been viewed as being driven by aberrant transcriptional regulation and signaling events, though, over the past several years, many RNA binding proteins and non-coding RNAs have emerged as critical players in tumor development. It is now recognized that regulation of post-transcriptional processes, such as mRNA stability and translation, robustly influence cancer-related gene expression patterns of proto-oncogenes, growth factors, cytokines, and cell cycle regulators. Despite its recognized importance, mechanisms of post-transcriptional regulation that influence molecular pathways at the mRNA level are understudied in the context of tumorigenesis. Additionally, cancer cells are derived from normal cells that often evolve step-wise and progressively to a neoplastic state, and the involvement of post-transcriptional regulation has not been looked at in the context of tumor initiation and step-wise progression. Thus, more studies are needed in order to fully understand the post-transcriptional mechanisms activated by cancer driver mutations that coordinate tumor initiation and progression.

In this dissertation, we aimed to elucidate mechanisms of post-transcriptional regulation coordinating tumorigenesis. We first established a genetically defined,
primary cell-derived model of breast cancer initiation and progression. In this model, normal human mammary epithelial cells were immortalized through the expression of hTERT, p53<sup>DD</sup>, cyclin D1, CDK4<sup>R24C</sup> and c-MYC<sup>T58A</sup>, and subsequently converted to a tumorigenic state through expression of oncogenic H-RAS<sup>G12V</sup>. Using RNA-sequencing and real-time PCR arrays, we comprehensively quantified changes in mRNA abundance, miRNA expression and alternative splicing in this system, and revealed thousands of changes during immortalization and relatively few changes during RAS transformation. Moreover, pre-malignant, immortalized cells had expression signatures consistent with an epithelial-to-mesenchymal transition (EMT), but they expressed low levels of mesenchymal protein markers and were non-invasive. Activation of RAS in these pre-malignant cells induced an invasive phenotype without major changes in global mRNA expression. Consistent with post-transcriptional mechanisms, RAS increased protein levels of Vimentin and N-cadherin without changing mRNA levels.

We then sought to investigate a mechanism of this RAS-induced post-transcriptional regulation. We used a method developed in our lab called Digestion-Optimized Ribonucleoprotein Immunoprecipitation coupled with RNA-sequencing (DO-RIP-seq) to identify and quantify transcriptome-wide binding sites for the RNA binding protein HuR. Our study is the first to identify and quantify transcriptome wide binding sites for any RBP during tumorigenesis, and we report that HuR quantitatively,
but not qualitatively, changed association at individual mRNA binding sites during RAS transformation. We identified a GU-rich secondary motif associated with a decrease in HuR binding during transformation. Furthermore, our data suggest that HuR may cooperate with the CELF1 protein to positively regulate the translation of a subset of mRNAs and promote the EMT phenotype. We generated HuR CRISPR knockout cell lines and demonstrated that HuR expression was necessary for the maintenance of cancer traits, including proliferation, anchorage independent growth, migration and invasion, but it does not regulate mRNA stability in this context. Lastly, we identified a binding site position dependent mechanism by which HuR regulates alternative polyadenylation of mRNAs encoding proteins involved in cancer-related processes.

In conclusion, our findings indicate that EMT-associated invasion can be initiated through two sequential stages: transcriptional priming followed by oncogenic RAS-triggered post-transcriptional regulation. The HuR RNA binding protein is important for maintaining the cancer phenotypes induced by oncogenic RAS, and regulation by HuR may be, at least in part, determined by a GU-rich secondary motif as well as cooperation with the CELF1 RNA binding protein.
# Contents

Abstract ................................................................................................................................. iv

List of Tables .......................................................................................................................... xiii

List of Figures ........................................................................................................................ xiv

List of Abbreviations .......................................................................................................... xvii

Acknowledgements ............................................................................................................. xix

1. Introduction ...................................................................................................................... 1

   1.1 Post-transcriptional Regulation of Gene Expression ............................................. 1

   1.2 Analysis of ribonucleoprotein complexes ............................................................. 3

   1.3 Post-transcriptional regulation of gene expression in cancer ......................... 6

       1.3.1 Hu Family of RBPs/ HuR ........................................................................... 8

          1.3.1.1 HuR expression and localization in cancers ......................................... 9

          1.3.1.2 HuR regulation of cancer-related mRNA targets ............................. 10

       1.3.2 The CELF family of RBPs ......................................................................... 13

       1.3.3 microRNAs ................................................................................................. 16

   1.4 mRNA processing events: setting the stage for post-transcriptional regulation .. 19

       1.4.1 Alternative Polyadenylation ..................................................................... 20

       1.4.2 Alternative Splicing .................................................................................. 23

   1.5 The Hallmarks of Cancer: stepwise progression and evolution ....................... 25

   1.6 The Epithelial-to-Mesenchymal Transition ......................................................... 27

   1.7 Cell culture systems in which to study progressive tumorigenesis ................. 28
1.8 Overview of Contents..................................................................................................................30

2. Materials and Methods...................................................................................................................32

2.1 Cell Culture .................................................................................................................................32

2.2 Establishing a primary cell-derived system of tumorigenesis .............................................32

2.2.1 Generation of amphotropic retroviruses.............................................................................32

2.2.2 Retroviral transductions .......................................................................................................33

2.2.3 Generation of stable cell lines ..............................................................................................34

2.2.4 Horizontal Spread Assay ....................................................................................................34

2.3 Real-time PCR ............................................................................................................................34

2.4 Soft Agar Assay ..........................................................................................................................37

2.5 Flow Cytometry ..........................................................................................................................37

2.5.1 Cell Cycle Analysis ...............................................................................................................37

2.5.2 Live Cell Staining and Sorting .............................................................................................38

2.6 Wound Healing Scratch Assay ................................................................................................38

2.7 Matrigel Assay .............................................................................................................................39

2.8 Western Blots ...............................................................................................................................39

2.9 RNA sequencing ..........................................................................................................................40

2.10 Alternative splicing ....................................................................................................................41

2.11 microRNA profiling ....................................................................................................................41

2.12 Digestion Optimized Ribonucleoprotein Immunoprecipitation and RNA-sequencing (DO-RIP-seq) .........................................................................................................................42

2.12.1 Preparation of cell lysates ..................................................................................................42
4.3.1 Large changes in gene expression occur during immortalization, while relatively few changes occur during RAS transformation ................................................................. 68

4.3.2 Immortalized and transformed cells have a mesenchymal gene expression signature .................................................................................................................................. 74

4.4 Alternative Splicing Analysis .......................................................................................................................................................................................... 76

4.4.1 Large changes in alternative splicing occur during immortalization, but not during oncogenic RAS transformation .................................................................................................................. 76

4.4.2 Immortalized cells have acquired an EMT splicing signature that is maintained during RAS transformation .................................................................................................................................. 78

4.5 microRNA expression .................................................................................................................................................................................................. 83

4.6 Discussion .......................................................................................................................................................................................................................... 88

4.6.1 Prognostic implications and drug targets ......................................................................................................................................................... 89

4.6.2 Post-transcriptional regulation downstream of oncogenic RAS ......................................................................................................................... 91

4.6.3 EMT transitions ................................................................................................................................................................................................ 94

4.6.4 Changes in miRNAs during cancer progression ........................................................................................................................................... 95

5. Binding site identification and quantification provides mechanistic insight into HuR regulation during tumorigenesis ................................................................. 96

5.1 Background and Significance ......................................................................................................................................................................................... 96

5.2 DO-RIP-seq to identify and quantify RBP binding sites transcriptome-wide .............................................................................................................. 97

5.2.1 Identification and quantification of HuR binding sites ........................................................................................................................................... 97

5.2.2 Identification and quantification of CELF1 binding sites ............................................................................................................................... 102

5.2.3 Comparison of HuR and CELF1 binding sites .............................................................................................................................................. 103

5.2.4 Identification and quantification of HuR and CELF1 mRNA targets at the whole transcript level .......................................................................................................................... 106

x
5.3 Analysis of HuR Knockout TFO cells
5.3.1 Generation of HuR Knockout transformed HMECs using the CRISPR-Cas9 genome editing system
5.3.2 The effect of HuR knockout on transformed HMEC cell growth
5.3.3 The effect of HuR knockout on transformed HMEC migration and invasion
5.3.4 RNA sequencing analysis of HuR CRISPR KO cells
5.3.5 The effect of HuR knockout on miRNA-mediated suppression and miRNA expression
5.4 HuR binding site position-dependent regulation of alternative polyadenylation
5.4.1 Analysis of HuR binding position relative to regions important for APA regulation
5.4.2 Analysis of HuR binding position on 3'UTR isoform abundance
5.4.3 Analysis of HuR binding position on protein expression
5.4.4 HuR binding position and association with core APA machinery
5.4.5 Gene ontology analysis of mRNAs with HuR binding sites upstream or downstream of the poly(A) signal sequence
5.5 Discussion
5.5.1 The presence of a GU secondary motif and cooperative regulation with CELF1 may be two separate mechanisms that determine HuR RNPs during oncogenic RAS transformation
5.5.2 HuR maintains cancer phenotypes through mechanisms other than mRNA stability
5.5.3 Therapeutic Implications
6. Conclusions and Future Directions
6.1 Elucidating genetic alterations that lead to EMT-related transcriptional priming ................................................................. 159

6.2 Determining post-transcriptional mechanisms induced by oncogenic RAS transformation .......................................................... 161

6.3 Understanding mechanisms of regulation by HuR in RAS-transformed HMECs ........................................................................ 162

6.4 HuR as a therapeutic target ...................................................................................................................................................... 167

6.5 Determining post-transcriptional regulation coordinating the immortalization of primary cells .................................................................. 169

References ........................................................................................................................................................................... 172

Biography ........................................................................................................................................................................ 202
List of Tables

Table 1: pBabe plasmids..................................................................................................................33

Table 2: Real-time PCR primer sequences used to check transgene expression..................35

Table 3: Primer sequences of APA distal and universal primers for real time PCR ........36

Table 4: All miRNAs that significantly changed expression during the primary to immortalized transition.................................................................................................................................87

Table 5: All miRNAs that significantly changed expression during the immortalized to transformed transition..................................................................................................................................................88

Table 6: List of validated HuR Knockout clones and controls used in experiments........118

Table 7: All genes that changed significantly (q<0.05) after HuR KO ........................................130
List of Figures

Figure 1: RNA binding proteins regulate each hallmark of cancer ........................................8

Figure 2: Individually validated mRNA targets of HuR form cancer regulons ..................13

Figure 3: Intracellular pathways, as determined by the Counter laboratory, altered to induce immortalization and oncogenic transformation in normal human cells ..........30

Figure 4: Workflow to establish and validate a primary cell-derived system of tumorigenesis. ...............................................................................................................................51

Figure 5: Confirmation of transgene expression. ........................................................................54

Figure 6: Anchorage independent growth ..................................................................................55

Figure 7: Cell cycle analysis .......................................................................................................56

Figure 8: Cell morphology ........................................................................................................56

Figure 9: HMEC immortalization increases RNA expression of RBPs. .................................58

Figure 10: Cancer-related RBPs are regulated during immortalization .................................59

Figure 11: Migration and Invasion ............................................................................................67

Figure 12: Western blots of canonical epithelial and mesenchymal markers .......................68

Figure 13: RNA-seq analysis of transition states. .................................................................70

Figure 14: Gene Ontology Analysis .........................................................................................73

Figure 15: Epithelial and mesenchymal gene expression signatures. .................................75

Figure 16: Expression levels of mRNA encoding N-cadherin and Vimentin .......................76

Figure 17: Alternative splicing analysis ....................................................................................77

Figure 18: Breakdown of the types of reproducible alternative splicing changes ...............78

Figure 19: EMT alternative splicing signatures .......................................................................80
Figure 20: Well characterized EMT alternative splicing events are observed during immortalization..........................................................82

Figure 21: Novel EMT alternative splicing events ..........................................................83

Figure 22: microRNA dynamics in cancer transition states.................................86

Figure 23: Analysis of HuR binding sites .................................................................99

Figure 24: IGV visualizations of examples of sites that change quantitatively with HuR during RAS transformation .................................................................100

Figure 25: Change in HuR binding during RAS transformation is associated with a GU-rich secondary motif .................................................................101

Figure 26: Analysis of CELF1 binding sites..............................................................103

Figure 27: Overlapping HuR and CELF1 binding sites ........................................104

Figure 28: CELF1 overlapping binding sites and changes in HuR association........105

Figure 29: Correlation Matrix of R values on Log (RPKM), Spearman ..................107

Figure 30: Scatter plots of HuR and CELF1 RSLs ...................................................107

Figure 31: Histograms of p-values ........................................................................108

Figure 32: Analysis of HuR and CELF1 overlapping mRNA targets .......................109

Figure 33: CDF plot of mRNA length versus targeting .........................................110

Figure 34: Gene ontology analysis for all HuR mRNA targets, all CELF1 mRNA targets, mRNA targets of both HuR and CELF1 (overlapping), and non-targets (neither) ..........112

Figure 35: Gene Set Enrichment Analysis for mRNA targets of A) HuR and B) CELF1 113

Figure 36: Generation of HuR KO cell lines ...........................................................115

Figure 37: CRISPR/Cas9 system abrogates HuR protein expression ......................116

Figure 38: Representative visualizations of indels with the TIDE webtool ...............117
Figure 39: Doubling time of HuR KO cells

Figure 40: Cell Cycle Analysis of TFO, HuR WT controls, and HuR KO clones

Figure 41: FITC-annexin V and DAPI staining to assess live, apoptotic and dead cell populations

Figure 42: Soft agar assay in HuR WT and KO clonal cell lines

Figure 43: Quantification of migratory competence after HuR depletion

Figure 44: Quantification of invasive capacity after HuR depletion

Figure 45: Immunoblot of N-cadherin and Vimentin in HuR WT and KO clonal cell lines

Figure 46: RNA-sequencing analysis in HuR WT and HuR KO TFO HMECs

Figure 47: Cumulative distribution function (CDF) plots showing changes in expression levels after HuR KO according to HuR LOD scores or number of binding sites

Figure 48: CDF plots showing the response of HuR and CELF1 targets to HuR depletion

Figure 49: Gene Ontology analysis of all mRNAs that changed expression in response to HuR KO

Figure 50: Analysis of miRNA regulation in HuR KO cells

Figure 51: Histogram of HuR binding sites around the poly(A) signal sequence

Figure 52: Global evidence for HuR binding site position dependent APA

Figure 53: Change in 3'UTR isoform abundance after HuR depletion

Figure 54: Luciferase Constructs

Figure 55: Luciferase Reporter Assay

Figure 56: CSTF64 RIP-PCR
Figure 57: Gene ontology analysis for mRNAs with HuR binding sites upstream or downstream of the poly(A) signal sequence.

List of Abbreviations

APA- alternative polyadenylation

ARE- AU-rich element

CDF- cumulative distribution function

CELF- CUG-binding protein ELAV-Like Family

CLIP- crosslinking immunoprecipitation

CPSF- cleavage and polyadenylation specificity factor

CRISPR- clustered regularly interspaced short palindromic repeats

CstF- cleavage stimulation factor

DO-RIP-seq- digestion optimized ribonucleoprotein immunoprecipitation coupled with deep sequencing

DSE- downstream element

ELAV/Hu- embryonic lethal abnormal vision/ Hu proteins

EMT- epithelial-to-mesenchymal transitin

ESRP- epithelial splicing regulatory protein

FACS- fluorescence activated cell sorting

GO- gene ontology

GRE- GU-rich element
GSEA- gene set enrichment analysis
HMEC- human mammary epithelial cell
IMO- immortalized human mammary epithelial cell
IP- immunoprecipitation
KO- knockout
LOD- logarithm of odds score
MiRNA- microRNA
PRIM- primary human mammary epithelial cell
PTR- post-transcriptional regulation
RBP- RNA binding protein
RIP-seq- ribonucleoprotein immunoprecipitation coupled with deep sequencing
RISC- RNA induced silencing complex
RNP- ribonucleoprotein complex
RPM- reads per million mapped reads
RRM- RNA recognition motif
RSL- RIP-seq-like score
TFO- transformed human mammary epithelial cell
USE- upstream element
UTR- untranslated region
Acknowledgements

First and foremost I would like to thank my advisor, Jack Keene, for continuously supporting my ideas and my research, and for always being enthusiastic about my work, even when the results were negative. I would like to especially acknowledge Matt Friedersdorf for assisting with the analyses in this dissertation. I would also like to thank previous members of the Keene lab- Jeff Blackinton, Cindo Nicholson, Marshall Thompson and Kyle Mansfield- for providing mentorship, advice and creating a collegial work environment. Also thank you to my committee for providing valuable advice and encouragement.

Thanks to DiSean Kendall for materials and advice on setting up my cell culture model, Eric Lykken and Qi-Jing Li for help with the microRNA arrays, the Ali-Osman lab for assistance with the xCELLigence RTCA system, the Caron lab for assistance with the Clariostar luminometer, the Silver and Pickup labs for use of their microscopes, and the Nicchitta lab for help with translation methods.

Thanks to my family, especially my parents, for supporting my academic endeavors since the very beginning. Special thanks to Joe for your unconditional love, support, and countless sacrifices that you made for the sake of my education and career; and to Joelle for helping me learn to prioritize and teaching me what work/life balance really means. I could not have done this without either of you.
1. Introduction

1.1 Post-transcriptional Regulation of Gene Expression

Precise control of gene expression is achieved through regulation at both the transcriptional and post-transcriptional levels (Keene, 2001, 2007). While many studies focus on steady state mRNA levels to measure gene expression, it has become increasingly clear that this does not provide an accurate picture of the complex regulatory features of gene expression occurring within a cell. Post-transcriptional regulation (PTR) of mRNA expression is controlled and coordinated by trans-acting RNA binding proteins (RBPs), and also non-coding RNAs, which bind to cis regulatory elements contained within nascent transcripts (Keene, 2007). Tightly regulated PTR events include splicing, nuclear export, localization, mRNA stability, and ultimately translation (Keene, 2007; Maniatis and Reed, 2002; Morris et al., 2010; Orphanides and Reinberg, 2002). Large numbers of mRNAs undergo PTR, as evidenced by the fact that steady state protein levels do not always correlate with steady state mRNA levels (Ideker et al., 2001; Mansfield and Keene, 2009; Vogel and Marcotte, 2012). RBPs in eukaryotes have been estimated to outnumber DNA binding proteins (Gerber et al., 2004; Keene, 2001), are highly expressed when compared to any other class of genes (Kechavarzi and Janga, 2014), and are, as a class, more conserved than transcription factors in metazoans (Alie et al., 2015), highlighting the importance of PTR.
There have been thousands of RBPs identified in vertebrates to date, each of which acts on various levels of mRNA metabolism (Glisovic et al., 2008). RBPs recognize their mRNA targets through sequence- or structure-specific RNA binding domains. These include the well-characterized RNA Recognition Motif (RRM), the K-homology domain, zinc finger domain, the Piwi/Argonaute/Zwille domain, the double stranded RNA binding domain, and the Arginine rich domain (Glisovic et al., 2008). These domains recognize specific cis elements typically found in untranslated regions (UTRs), but also found in open reading frames. Each protein can bind many mRNAs, and each mRNA can bind multiple proteins, allowing for the formation of complex RNA-protein networks (Keene, 2007).

Many studies in several organisms, including yeast, fruit flies, trypanosomes and mammals, have suggested that functionally related mRNAs are coordinately regulated by specific RBPs into ribonucleoprotein (RNP) complexes to form ‘posttranscriptional RNA regulons,’ a concept first proposed and demonstrated by the Keene lab (Keene, 2007, 2010; Morris et al., 2010; Tenenbaum et al., 2002). This concept can be equated to bacterial DNA operons, where genes that function in similar pathways are physically linked in order to allow for their rapid, coordinated expression (Jacob et al., 1960). However, in eukaryotic cells, DNA operons are rare, and it has been traditionally thought that gene expression is primarily regulated at the level of transcription through the use of common DNA promotor elements. It is now widely recognized that
coordinated regulation of gene expression also occurs post-transcriptionally (Keene, 2007). RNA regulons allow for the combinatorial regulation of functionally related mRNAs through post-transcriptional events, leading to appropriate spatiotemporal expression and rapid responses to environmental signals by governing the co-expression of mRNAs with common functions (Keene, 2007, 2010; Mukherjee et al., 2011). Thus, RBPs play crucial, evolutionarily conserved, roles in efficiently coordinating proteomic responses leading to cell survival and adaptation in response to stressful conditions, such as those found in the tumor microenvironment during cancer progression. Rapid protein expression changes can be coordinated from mRNAs that are already transcribed, allowing for an adaptive gene expression program while transcription is halted or unchanged.

1.2 Analysis of ribonucleoprotein complexes

Many methods to determine RNA-protein interactions have been described, each with their own distinct benefits. The first global, in vivo method is the commonly used Ribonucleoprotein Immunoprecipitation coupled with microarray or deep sequencing (RIP-chip/seq) (Keene et al., 2006). RNP complexes are immunoprecipitated by incubating cell lysates with beads coated with antibody specific to the RBP of interest, and mRNAs bound to an RBP are determined using a microarray or, more commonly in recent years, deep sequencing. Importantly, a negative immunoprecipitation (IP), often an IgG control, is used to normalize the RIP data, allowing for the quantification of
binding. By normalizing for mRNA abundance and background binding, condition
specific probabilities of each RBP-mRNA interaction are generated (Morris et al., 2010;
Mukherjee et al., 2009). Under the original conditions described by our lab, no RNA re-
association has been observed (Mukherjee et al., 2011; Mukherjee et al., 2009;
Tenenbaum et al., 2000; Tenenbaum et al., 2002). While this method has been
successfully used by many labs and in many organisms to identify both mRNA targets
of RBPs as well as dynamic remodeling of RNPs (Calaluce et al., 2010; Mazan-Mamczaras
et al., 2008a; Mazan-Mamczarz et al., 2008b; Mazan-Mamczarz et al., 2011; Mukherjee et
al., 2009), one drawback is that this method does not allow for the identification of
specific binding site locations.

In order to identify binding sites, our lab developed methods that involve
crosslinking (Keene 2003 U.S. patent application 6635422). Many labs have since applied
adaptations, and crosslinking immunoprecipitation (CLIP) methods are now commonly
used to identify transcriptome wide binding sites (Hafner et al., 2010; Konig et al., 2010;
Licatalosi et al., 2008; Ule et al., 2003; Zhang and Darnell, 2011). However, our lab has
recently demonstrated that CLIP procedures contain background binding, which is not
typically accounted for in analyses, as it is assumed that stringent washing procedures
eliminate all background (Friedersdorf and Keene, 2014). This background can be used
to improve binding site calling, enabling for improvements on identifying bona fide
binding sites from background signal, however, even with this improvement, CLIP methods do not actually quantify binding sites (Friedersdorf and Keene, 2014).

Quantification of binding sites is important in understanding PTR. It has been demonstrated that many RBPs “sample” mRNAs without ever having a regulatory effect (Lapointe et al., 2015), and thus CLIP methods may “lock” a protein in to a binding site through crosslinking. Moreover, individual RBPs can have multiple binding sites within a given mRNA (Mukherjee et al., 2011), and both identification and quantification of these sites are essential to differentiate between regulatory and non-regulatory sites. Quantification is necessary if one is to study mRNP remodeling under dynamic biological conditions (Iruarrizaga-Lejarreta et al., 2012; Mazan-Mamczarz et al., 2008a; Mazan-Mamczarz et al., 2008b; Mukherjee et al., 2009; Tenenbaum et al., 2000).

In order to solve this problem of binding site level quantification, our lab recently developed a method called Digestion-Optimized (DO) RIP-seq (Nicholson et al., 2017; Nicholson et al., 2016). Prior to the immunoprecipitation step, cell lysates are digested with a nuclease under tightly controlled, optimized conditions to yield RNA fragments that were protected by RBPs. Fragments that are bound to an RBP of interest are recovered from immunoprecipitations and deep-sequenced. Analysis of the fragments is similar to RIP-procedures in that it involves normalization of the fragments to a negative immunoprecipitation control or to input mRNA samples to quantitatively determine binding site enrichment. With this method, one can identify global binding
sites, distinguish weak from strong binding sites, and quantify changes in site usage under different conditions. Additionally, by looking at enriched regulatory motifs within binding regions, one can start to distinguish combinatorial interactions between multiple RBPs. Thus, DO-RIP-seq is a powerful approach that can reveal changes at the binding site level and relate it to remodeling of RNA regulons, and potentially functionally outcome, on a global scale across dynamic biological states.

**1.3 Post-transcriptional regulation of gene expression in cancer**

Given the importance of PTR in regulating gene expression, it is not surprising that it is an important mechanism regulating tumor progression. Cancers progress through the accumulation of genetic alterations that ultimately drive malignancy. Although each case can result in different phenotypes and clinical outcomes, there are a few major hallmarks that the majority of normal human cells must acquire in order to progress to malignancy (Hanahan and Weinberg, 2000, 2011). Cancer cells evolve progressively toward a neoplastic state by acquiring cell autonomous growth signals, insensitivity to growth inhibitory signals, evasion of apoptosis, a limitless replicative potential, induction of angiogenesis, and activation of tissue invasion and metastatic dissemination, which is what ultimately kills most patients.

Many studies have shown that PTR is an important determinant of gene expression in both normal cellular processes and pathological states. It is now widely recognized that RBPs robustly influence cancer-related gene expression patterns, as they
regulate many mRNAs encoding proto-oncogenes, growth factors, cytokines, and cell cycle regulators (Abdelmohsen and Gorospe, 2010; Lukong et al., 2008). The majority of oncogenes and genes involved in proliferation have particularly large UTRs (Lopez de Silanes et al., 2007), providing further evidence that they are likely to be tightly regulated by RBPs and non-coding RNAs. Cancer has traditionally been viewed as being driven by aberrant transcriptional regulation and signaling events, though, over the past several years, many RBPs and non-coding RNAs have emerged as critical players in tumor development (Figure 1) (Abdelmohsen and Gorospe, 2010; Lukong et al., 2008; Wurth, 2012; Wurth and Gebauer, 2015). Mutations in RBPs or alterations in their expression levels can have a large impact on gene expression programs, and expression levels of certain subsets of RBPs have been shown to be different in cancer versus normal tissues and cells (Bisogno and Keene, 2017; Galante et al., 2009; Kechavarzi and Janga, 2014). Additionally, some RBPs, although not necessarily directly targeting proto-oncogenes and growth factors, can have profound effects on tumorigenesis through altering RNA processing, such as alternative splicing or alternative polyadenylation (Mayr and Bartel, 2009; Oltean and Bates, 2014). A few RBPs that are especially relevant to cancer, and to this dissertation, are discussed below.
RNA-binding proteins regulate cancer hallmarks

**Figure 1: RNA binding proteins regulate each hallmark of cancer**

Examples of well-studied RBPs known to directly regulate major hallmarks of cancer. Reviewed in Abdelmohsen and Gorospe 2010; Oltean and Bates 2014; Wurth 2012; Wurth and Gebauer 2015.

### 1.3.1 Hu Family of RBPs/ HuR

One of the most extensively studied RBP families is the highly conserved embryonic lethal abnormal vision (ELAV)/Hu protein family. Three Hu family proteins, HuB, HuC and HuD, are predominantly cytoplasmic and neuron specific, and were originally identified as tumor antigens in individuals with lung carcinomas that had developed paraneoplastic neurological disorders, suggesting a role for the Hu proteins in the regulation of growth regulatory genes (Keene, 1999). The fourth Hu protein, HuR (also called HuA), is ubiquitous and predominantly nuclear, although nucleo-
cytoplasmic shuttling is strongly linked to its role in regulating mRNA stabilization and translation (Fan and Steitz, 1998a, b; Ma et al., 1996).

Each Hu protein contains three RRM s through which they bind to specific mRNAs, as well as a variable hinge region between RRM s 2 and 3 which is important for protein localization and shuttling (Burd and Dreyfuss, 1994; Levine et al., 1993; Ma et al., 1996; Nabors et al., 2001; Query et al., 1989). Hu proteins preferentially bind to AU-rich elements (AREs), highly conserved repetitive sequences typically found in 3'UTRs of normally labile mRNAs (e.g. cytokines, proto-oncogenes) (Caput et al., 1986; Gao et al., 1994; Shaw and Kamen, 1986). While AREs are generally destabilizing elements (Barreau et al., 2005; Shaw and Kamen, 1986), early studies of HuB revealed that it stabilizes messages through binding the ARE (Gao and Keene, 1996; Jain et al., 1995; Levine et al., 1993). Since then, many studies have corroborated these findings and demonstrated that Hu proteins typically act to increase mRNA stability and/or promote translation of mRNA targets. The association between the ubiquitously expressed HuR protein and cancer has been studied extensively (Abdelmohsen and Gorospe, 2010).

1.3.1.1 HuR expression and localization in cancers

HuR is primarily nuclear, but its stabilizing effects are closely linked to its translocation to the cytoplasm (Fan and Steitz, 1998b). Increased cytoplasmic accumulation of HuR has been linked to advanced tumor grade and poor clinical outcomes in several types of cancers, including pancreatic, colon, ovarian and breast
(Abdelmohsen and Gorospe, 2010; Lopez de Silanes et al., 2003; Raspaglio et al., 2010; Yi et al., 2009). In invasive ductal breast carcinoma and hereditary breast cancer, cytoplasmic HuR accumulation is an independent marker of reduced patient survival and is associated with high tumor grade (Heinonen et al., 2005; Heinonen et al., 2007; Wang et al., 2013). HuR is also overexpressed in pre-malignant lesions and immortalized cell lines, and inhibition of HuR resulted in reduced tumorigenic properties in vitro (Heinonen et al., 2011). In breast cancer cell lines, it has been reported that increases in HuR abundance as well as increases in cytoplasmic localization, either with or without corresponding increases in total HuR levels, are associated with cancer (Al-Ahmadi et al., 2013; Calaluce et al., 2010; Guo and Hartley, 2006). HuR mRNA levels and stability were demonstrated to be higher in the MCF7 epithelial adenocarcinoma cell line when compared to the immortalized MCF10A cell line, and even higher in the highly tumorigenic MDA-MB-231 breast cancer cell line (Al-Ahmadi et al., 2013; Guo and Hartley, 2006). However, contradictory studies reported higher HuR protein levels in MCF7 compared to MDA-MB-231 cells (Hostetter et al., 2008), while others reported no difference in expression levels (Calaluce et al., 2010).

1.3.1.2 HuR regulation of cancer-related mRNA targets

Increased levels of cytoplasmic HuR have been linked to the increased stability of transcripts encoding cancer-related proteins (Abdelmohsen and Gorospe, 2010; Denkert et al., 2004; Fan and Steitz, 1998b; Liang et al., 2012). Of the hallmarks of cancer
described by Hanahan and Weinberg, HuR has been demonstrated to target mRNAs encoding proteins involved in each step of malignant transformation (Abdelmohsen and Gorospe, 2010; Kotta-Loizou et al., 2016; Lukong et al., 2008; Simone and Keene, 2013) (Figure 2). Thus, HuR may contribute to tumorigenesis by aberrantly regulating cancer-related targets.

Several studies have looked at global HuR mRNA binding targets. Array analysis of colon cancer cells either over- or under-expressing HuR demonstrated that HuR differentially regulates genes encoding proteins involved in proliferation, angiogenesis and tissue invasion, suggesting that it plays an important role in cancer by altering the stability/translation of mRNAs encoding proteins involved in malignant progression (Lopez de Silanes et al., 2004). Ribonomic analyses comparing low and high tumorigenic MCF7 cells (with and without overexpression of the MCT-1 oncogene) demonstrated that HuR differentially associates with 1676 mRNAs in these two cell lines, losing association with mRNAs encoding proteins involved in preventing tumorigenesis and gaining association with mRNAs encoding cancer related mRNAs with increasing malignancy (Mazan-Mamczarz et al., 2008a). Similar findings were obtained through the comparison of immortalized MCF10A cells and cells transformed with MCT-1 oncogene, suggesting that HuR acts as a downstream effector of oncogenes (Mazan-Mamczarz et al., 2008b). Another study used RIP-chip to compare HuR targets in MCF7 and MDA-MB-231 cell lines and found that subsets of mRNAs involved in
cancer-related pathways, including epithelial cell differentiation, vasculature development and signal transduction, were differentially regulated (Calaluze et al., 2010). A study from our lab revealed that treatment of leukemia-derived T-cells with phorbol esters and mitogens resulted in coordinate changes in RNP associated mRNAs (Mukherjee et al., 2009). HuR RNPs are also remodeled in response to DNA damage (Mazan-Mamczarz et al., 2011).

While it is clear that HuR functions as a modulator of the proliferative gene expression program through regulating the stability and translation of mRNAs encoding growth-related proteins, the molecular mechanisms underlying these functional effects remain unclear. Mechanistic studies aimed at determining how HuR stabilizes mRNA targets have been scarce. Transcripts can associate with many trans regulators with often opposing activities (Lal et al., 2004). Investigating mechanisms of competition and cooperation among these factors is important in fully understanding complex post-transcriptional networks in disease states.
Cancer Hallmarks

<table>
<thead>
<tr>
<th>Cell Proliferation</th>
<th>Cell Proliferation (continued)</th>
<th>Resisting Cell Death</th>
<th>Angiogenesis</th>
<th>Invasion and Metastasis</th>
<th>Inflammation and Immune Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCNA1</td>
<td>CDK7</td>
<td>HIF1α</td>
<td>CD9</td>
<td>TGFβ</td>
<td></td>
</tr>
<tr>
<td>CCNB1</td>
<td>EGF</td>
<td>VEGF</td>
<td>GATA3</td>
<td>TNFα</td>
<td></td>
</tr>
<tr>
<td>CCND1</td>
<td>YES1</td>
<td>COX2</td>
<td>MMP9</td>
<td>CCL8</td>
<td></td>
</tr>
<tr>
<td>CCNE1</td>
<td>IGF1R</td>
<td>THBS1</td>
<td>MTA1</td>
<td>CSF1R</td>
<td></td>
</tr>
<tr>
<td>FOS</td>
<td>EBB2</td>
<td>IL8</td>
<td>SNAI1</td>
<td>COX2</td>
<td></td>
</tr>
<tr>
<td>HMG2A</td>
<td>CALM2</td>
<td>RHOB</td>
<td>PDGFC</td>
<td>IL2</td>
<td></td>
</tr>
<tr>
<td>eIF4E</td>
<td>SOCS3</td>
<td>RUNX1</td>
<td>uPA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eIF4EBP2</td>
<td>CXCR4</td>
<td></td>
<td>uPAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTNNB1</td>
<td>ESR1</td>
<td></td>
<td>WNT5A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN1B1/p27</td>
<td>O-MYC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOXA5</td>
<td>RAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>STAT3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AP1</td>
<td>GM-CSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individually validated mRNA targets of HuR

Figure 2: Individually validated mRNA targets of HuR form cancer regulons

List includes both positive and negative regulators of cancer. Some of these mRNAs (e.g. c-Myc) can be included in more than one hallmark, but are only listed once for the sake of simplicity. Included are the original hallmarks as described by Hanahan and Weinberg in 2000, as well as the two “emerging hallmarks” described by them in 2011: Evading Immune destruction and reprogramming of energy metabolism. Many more mRNAs that encode functions within these acquired hallmarks have been identified using ribonomics approaches, as mentioned in the text.

1.3.2 The CELF family of RBPs

Similar to AREs, the GU-rich element (GRE) is a highly evolutionarily conserved motif that plays important roles in post-transcriptional regulation of mRNAs, including those involved in proliferation, apoptosis and cell motility (Lee et al., 2010). This GU-rich motif is recognized by several RBPs, including the six members of the CUG binding protein (CUGBP) ELAV-Like Family (CELF) (Dasgupta and Ladd, 2012). CELF proteins bind to GREs and typically repress mRNA translation, but also have known roles in
alternative splicing, deadenylation and mRNA degradation (Dasgupta and Ladd, 2012).

The GRE consensus sequence, UGUUUGUUUGU, was originally identified as being a decay element in the 3’UTRs of many labile mRNAs in primary human T-cells (Vlasova et al., 2008). In addition to the GRE consensus sequence, CELF1 has been demonstrated to preferentially bind to a GU-repeating sequence (UGUGUGUGUGU) enriched in introns and 3’UTRs, as identified by RIP-chip studies (Rattenbacher et al., 2010) and SELEX-based methods (Lambert et al., 2014; Marquis et al., 2006; Ray et al., 2013).

Binding sites in introns generally regulate splicing (Daughters et al., 2009), while binding sites in 3’UTRs typically result in mRNA decay (Masuda et al., 2012). Disruption of regulation by CELF proteins may play a role in many developmental diseases and cancers. For example, CELF6 was identified in a screen for proteins that could rescue KRAS suppression in oncogene addicted colon cancer cell lines (Shao et al., 2014). A transposon based genetic screen in mice colorectal cancer showed that CELF1 was a candidate to drive tumorigenesis when mutated (Starr et al., 2009). RIP-chip studies in HeLa cells showed that mRNAs encoding proteins involved in proliferation, apoptosis and migration were enriched in CELF1 RNP’s (Rattenbacher et al., 2010). CELF1 is associated with poor survival in non-small cell lung cancer (Wu et al., 2013), and depletion of CELF1 in non-small cell lung cancer cell lines increased protein levels of pro-apoptotic factors as well as E-cadherin (Gao et al., 2015). CELF1 has also been shown to negatively regulate messages encoding cyclin dependent kinase 4,
tumor necrosis factor, TNF-receptor 1b, Occludin, Myc, and transcription factors c-Jun and Junb (Liu et al., 2015; Rattenbacher et al., 2010; Vlasova et al., 2008; Yu et al., 2013b; Zhang et al., 2008). A more recent study demonstrated that CELF1 regulates an RNA regulon containing the mRNA encoding the signal recognition particles (SRP), and CELF1 knockdown resulted in higher SRP expression and reduced migration in vitro (Russo et al., 2017).

Although CELF1 normally acts as a negative regulator of expression, potentially through the direct recruitment of PARN deadenylase (Moraes et al., 2006) or localization of mRNAs to processing bodies (Yu et al., 2016; Yu et al., 2013b), it has been shown to positively regulate expression levels in certain contexts. CELF1 positively regulates Survivin, a member of the Inhibitor of Apoptosis Protein family, expression in esophageal cancer cells (Chang et al., 2012). It has also been demonstrated to promote translation of p21 in human fibroblasts and HeLa cells treated with the chemotherapeutic bortezomib (Gareau et al., 2011; Iakova et al., 2004). A more recent study uncovered a role for CELF1 in positively regulating the translation of several mRNAs involved in promoting metastasis (Chaudhury et al., 2016). Understanding the contexts and mechanisms by which CELF1 either represses or promotes translation is of interest.

Many mRNAs identified as being enriched in CELF1 RIP-chips (Lee et al., 2010) were also identified as enriched in HuR IPs (Mukherjee et al., 2009). Since HuR is
typically an mRNA stabilizer, and CELF1 is a destabilizer, this suggests a mechanism of global competition between these two RBPs. In intestinal epithelium, HuR and CELF1 compete for binding to the same 3'UTR elements of Occludin (Yu et al., 2013b), Myc (Liu et al., 2015), and E-cadherin (Yu et al., 2016), and regulate their translation in opposite directions. Thus, CELF1 is part of a complex PTR regulatory network that is important in disease progression, and understanding mechanisms by which CELF1 regulates protein expression in combination with other RBPs could provide new therapeutic avenues.

1.3.3 microRNAs

MicroRNAs (miRNA) are small non-coding RNAs of about 19-25nts in length that modulate many important biological processes through binding to complementary sequences typically in the 3'UTR of mRNA targets, which results in translation inhibition and/or mRNA degradation (Aravin and Tuschl, 2005; Lewis et al., 2005; Pillai et al., 2007). Like RBPs, miRNAs can regulate mRNA expression and protein function via coordinated and combinatorial mechanisms: each miRNA can target multiple, often functionally related, mRNAs, and each mRNA can be regulated by multiple miRNAs (Agami, 2010).

The regulatory effects of miRNAs are carried out through the RNA-induced silencing complex (RISC), also referred to as the microRNP, a complex of miRs and RBPs, including the catalytic Argonaute proteins (Ago) (Bartel, 2004). There is evidence
to suggest functional interplay between the RISC/miRNA complexes and other RBPs (Simone and Keene, 2013). For example, the cationic amino acid transporter 1 (CAT1) protein is repressed by miR-122, but this repression can be reversed in a HuR-dependent manner during induced stress through a mechanism involving recruitment of CAT1 mRNA from processing bodies to actively translating polysomes (Bhattacharyya et al., 2006; Filipowicz et al., 2008). A similar mechanism was reported by the Gorospe lab demonstrating that HuR attenuates the miR-548c-3p-mediated repression of TOP2A (Srikantan et al., 2011). Alterations in TOP2A levels have been observed as a mechanism of TOP2A-targeted drug resistance (Burgess et al., 2008). The Gorospe lab also showed that HuR can outcompete and displace miR-494 from nucleolin mRNA, which encodes an RBP that positively regulates mRNAs encoding anti-apoptotic and proliferation factors (Tominaga et al., 2011). In line with this, global studies from our lab have indicated that HuR mRNA targets are among the most concentrated mRNA targets of miRNAs, and models support a mechanism of direct competition when both HuR and miRNA binding sites are proximal (Mukherjee et al., 2011). Thus, miRNAs form complex post-transcriptional regulatory networks that can influence the expression of cancer-related mRNAs.

Global changes in miRNA expression levels are associated with cancer progression. In general, miRNAs are suppressed during tumorigenesis, but there are a few miRNAs, termed oncomiRs, that are positively associated with disease (Dedes et al.,
2011; Faggad et al., 2010; Thomson et al., 2006). One example is the miR-17-92 cluster, which contains six miRNAs that are transcribed from a locus frequently amplified in cancers (Diosdado et al., 2009; Hayashita et al., 2005; Ota et al., 2004; Rinaldi et al., 2007). The most oncogenic miRNA in this cluster is miR-19, which inhibits the expression of the PTEN tumor suppressor (Olive et al., 2009). Another example is miR-155, which is upregulated in cancers, including lymphomas and breast cancer, and correlates with Myc overexpression and overall poor survival (Eis et al., 2005; Kong et al., 2014; Metzler et al., 2004; Tam et al., 2002). MiR-221 and miR-222 target the p27 tumor suppressor (le Sage et al., 2007). MiR-10b is overexpressed in metastatic cells and breast cancer tissue, and it has been demonstrated to be an enhancer of invasion and metastasis, in part through targeting HOX10D (Ma et al., 2010; Ma et al., 2007; Ma and Weinberg, 2008).

Many miRNAs are considered to be tumor suppressive, and mRNAs functioning in the acquisition of each major hallmark of cancer as described by Hanahan and Weinberg are regulated by multiple miRNAs (Grammatikakis et al., 2013). For example, miR-15a and miR-16 can function as tumor suppressors by targeting the anti-apoptotic gene BCL-2 (Cimmino et al., 2005). Both of these microRNAs are frequently deleted or downregulated in many cancers, including B-cell chronic lymphocytic leukemia and pituitary adenomas (Bottoni et al., 2005; Calin et al., 2002). Interestingly, cytoplasmic HuR protein binds directly to miR-16 to prevent degradation of Cox-2 mRNA, a prostaglandin synthase commonly elevated in several types of cancer (Young et al.,
miR-331-3p suppresses the mRNA encoding Her2/neu oncogene, and HuR can rescue this suppression by hindering mRNA association with the microRNP complex, but not the miRNA (Epis et al., 2011). It has also been suggested that HuR itself may be a direct functional target of tumor suppressive miRNAs. For example, miR-125a and miR-519 negatively regulate HuR, are inversely correlated with HuR levels, and enforced expression of these miRNAs decreases both HuR protein levels and tumorigenicity in athymic mice (Abdelmohsen et al., 2010; Abdelmohsen et al., 2008; Guo et al., 2009). Therefore, not only does HuR alter miRNA-mediated suppression, but miRNAs also can mediate HuR-regulated gene expression, and thus, cell proliferation and potentially tumorigenesis. Additionally, a recent study mapped the RBP-pre-miRNA interactome and found that ~180 RBPs interacted with pre-miRNAs (Treiber et al., 2017). Therefore, RBPs can also regulate miRNA biogenesis. In summary, RBPs and miRNAs form complex, dynamic post-transcriptional regulatory networks that may directly regulate cancer traits.

1.4 mRNA processing events: setting the stage for post-transcriptional regulation

In eukaryotes, mature mRNAs are produced from precursor mRNAs through co-transcriptional and post-transcriptional processing mechanisms. These processes include the addition of a 7-methyl guanosine cap to the 5’end, removal of introns through splicing, and endonucleolytic cleavage followed by addition of a 3’ poly(A) tail (polyadenylation) (Gruber et al., 2014). A given transcript may be alternatively spliced
or alternatively polyadenylated to produce multiple mRNA isoforms. Since many RBP and miRNA recognition elements are contained within 3’UTRs and introns, these RNA processing events set the stage for post-transcriptional regulation.

1.4.1 Alternative Polyadenylation

It has recently been discovered that over 50% of human genes can undergo alternative 3’UTR cleavage and polyadenylation (APA), which results in mRNA isoforms that differ in 3’UTR length (Lianoglou et al., 2013). Shorter 3’UTR isoforms often lack cis regulatory elements, and thus they may escape regulation by RBPs and miRNAs, often resulting in greater protein production (Akman et al., 2012; Bava et al., 2013; Mayr and Bartel, 2009; Sandberg et al., 2008; Singh et al., 2009; Wiestner et al., 2007). While APA is generally thought to influence mRNA translation, not protein function, a recent paper has demonstrated that APA can also determine protein function without changing the amino acid sequence (Berkovits and Mayr, 2015). Christine Mayr’s group demonstrated that the long 3’UTR isoform of CD47 acts as a scaffold to recruit the RBP HuR, which acts as an adaptor to recruit the SET protein to the site of translation in order to facilitate association of SET with newly translated CD47 cytoplasmic domains, which is important for CD47 translocation to the plasma membrane (Berkovits and Mayr, 2015). Therefore, in addition to regulating protein abundance, APA may also influence protein localization and function.
Deregulation of APA has been reported in the context of cancer. For example, Mayr and Bartel reported that the short 3’UTR isoform of the proto-oncogene IGF2BP1 is more stable than the long isoform, and that the short isoform results in greater oncogenic transformation (Mayr and Bartel, 2009). Similarly, the cell cycle regulators CCND1 and CDC6 have been shown to have shortened 3’UTRs in lymphoma and breast cancer, respectively (Akman et al., 2012; Rosenwald et al., 2003). Additional studies have demonstrated that global 3’UTR shortening is associated with proliferative states (Elkon et al., 2012; Sandberg et al., 2008), although other studies found global changes in both directions (Lianoglou et al., 2013; Singh et al., 2009), suggesting some gene and cell-type specificity.

Given that APA has important functional consequences, it is not surprising that it is a highly regulated process requiring multi-protein complexes. Cleavage and polyadenylation is carried out by core 3’-end processing machinery. The main components of this complex are the cleavage and polyadenylation specificity factor (CPSF), FIP1 and WDR33, which recognize a conserved AAUAAA poly(A) signal sequence that peaks about 21 nucleotides upstream of the 3’cleavage site and binds to a region upstream of this site, and the cleavage stimulation factor (CstF) and cleavage factors I and II, which binds to GU- or U-rich elements 10-30 nucleotides downstream of the 3’cleavage site (Gruber et al., 2014; Martin et al., 2012; Yao et al., 2012). Once cleaved,
the 3’ end is polyadenylated by the nuclear polyadenylate polymerase (PAP) (Wahle, 1991).

Many RBPs have been implicated in the decision to use one poly(A) site over another, either through recruiting or blocking core APA machinery proteins. For example, hNRNP H and CPEB1 have been shown to bind upstream of poly(A) sites and recruit polyadenylation core proteins (Bava et al., 2013; Katz et al., 2010). In contrast, HuR has been shown to bind downstream of poly(A) sites and promote proximal poly(A) read-through, possibly through blocking CstF complex binding, resulting in longer 3’UTRs (Dai et al., 2012; Hilgers et al., 2012; Slevin et al., 2007; Soller and White, 2003; Zhu et al., 2007). HuR has also been shown to auto regulate its own APA (Dai et al., 2012; Mansfield and Keene, 2012). FUS and MBLN proteins also have described roles in APA (Batra et al., 2014; Masuda et al., 2015). Therefore, not only does APA determine regulation by RBPs and miRNAs through generating 3’UTRs that either possess or lack cis binding elements, but it is also tightly regulated by many RBPs themselves. Recent 3’end sequencing experiments under different dynamic conditions have revealed there are still undiscovered mechanisms regulating APA (Gruber et al., 2014). Uncovering the complexities of these mechanisms will be important to understanding post-transcriptional regulation in disease states.
1.4.2 Alternative Splicing

During pre-mRNA splicing, which is carried out by a core RNP complex termed the ‘spliceosome,’ introns are removed and exons are joined together to form mature mRNA (Oltean and Bates, 2014). Alternative splicing, a process by which mRNAs can be differentially spliced, significantly broadens the potential number of gene products that can be produced through generating distinct mRNA isoforms that vary in both coding and non-coding regions. It has been estimated that about 95% of multi-exon human genes are alternatively spliced (Pan et al., 2008). Not only can alternative splicing result in differences in mRNA stability and translation, but also localization, protein-protein interactions, post-translational modifications and ultimately function of the final protein product (Oltean and Bates, 2014).

Regulation of alternative splicing requires the recruitment of RBPs and formation of RBP-RNA complexes, the dysregulation of which is associated with many disease states, including cancer (Oltean and Bates, 2014). RBPs regulate alternative splicing events, many of which directly impact processes involved in tumorigenesis. For example, RBM5 has been shown to regulate alternative splicing of the Fas receptor through promoting expression of a membrane-bound, pro-apoptotic form of the protein (Bonnal et al., 2008). HuR has also been reported to be directly involved in Fas exon 6 skipping, which promotes the expression of a soluble, anti-apoptotic form (Izquierdo, 2008).
More recent studies have revealed that RBPs can affect alternative splicing of networks of functionally related mRNAs. For example, epithelial splicing regulatory proteins 1 and 2 (ESRP1 and 2) regulate a broad splicing switch that, along with transcriptional and epigenetic regulation, mediates the transition between an epithelial and a mesenchymal cell, a process closely linked to metastatic dissemination (Shapiro et al., 2011; Yang et al., 2016). ESRP proteins were discovered in a screen to identify regulators of alternative splicing of epithelial and mesenchymal specific isoforms of the fibroblast growth factor receptor 2 (FGFR2) (Warzecha et al., 2009), and since their discovery it’s been revealed that they regulate over 1,000 epithelial related alternative splicing events in combination with other RBPs including RBM47, and Quaking (Shapiro et al., 2011; Yang et al., 2016). In contrast, the RBP RBFOX2 promotes a more mesenchymal alternative splicing signature (Shapiro et al., 2011), although in some cases it can cooperate with ESRPs (Dittmar et al., 2012). Additionally, global studies done by our lab revealed that HuR binding sites are enriched in introns within the first 50 nucleotides of 3’ splice sites, which was associated with both exon-inclusion and exon-exclusion events (Mukherjee et al., 2011).

The mechanisms that determine alternative splicing events associated with cancer progression are only just beginning to be elucidated. Understanding how alternative splicing is regulated and coordinated during cancer progression as well as
the contributions of specific alternatively spliced isoforms to cancer progression could lead to new therapeutics and prognostics based on alternative splicing.

1.5 The Hallmarks of Cancer: stepwise progression and evolution

The progression of tumors through the acquisition of hallmarks described by Hanahan and Weinberg can, in part, be explained as a model of Darwinian evolution, in which novel genetic mutations confer a growth or survival advantage (Gerlinger et al., 2012; Gupta et al., 2005). This paradigm implies that tumors become progressively more malignant over time through the selection and expansion of cells that contain advantageous mutations. Indeed, many of the most common alterations observed in cancers confer selective advantages. For example, loss of p53 function results in apoptotic resistance and progression through the cell cycle (Willis et al., 2004), pRb loss results in insensitivity to growth inhibitory signals (Hanahan and Weinberg, 2000), and gain of function of oncogenes, such as c-Myc and Ras, enhances proliferation (Hanahan and Weinberg, 2000, 2011). However, metastatic traits do not provide a clear survival advantage within the primary tumor. Therefore, if metastasis were selected for by the Darwinian evolution model, it would be expected to be a very rare event. However, this is not the case. Accordingly, it has been hypothesized that metastatic traits are not necessarily late stage events in tumor progression, but rather can be determined early in the process, either within the cells of origin or within the primary tumor (Gupta et al., 2005).
In support of this hypothesis, genomic studies have shown that metastases from a patient, while often genetically similar to their primary tumor of origin, sometimes contain even fewer genetic abnormalities than the primary tumor, consistent with having been disseminated at an earlier stage of progression (Naxerova and Jain, 2015). In addition, metastases have been detected in patients before the primary tumor is identified, or even in the absence of a detectable tumor of origin (Friberg and Nystrom, 2015; Varadhachary, 2007). Mouse models have supported clinical observations of early metastatic spread. For example, mammary cancer cells were detected in both lung tissue and bone marrow in mice with pre-malignant atypical ductal hyperplasia (Husemann et al., 2008). Cells in these pre-malignant lesions showed increased expression levels of cathepsins, matrix metalloproteases, and Twist mRNA, suggesting that the acquisition of a gene expression program necessary for invasion is an early event. Likewise, using a transgenic mouse model of pancreatic ductal adenocarcinoma, invasive cells with Zeb1 expression were observed at the pre-malignant pancreatic intraepithelial neoplasia stage (Rhim et al., 2012). While these studies make it clear that metastatic programs can be acquired early on in multistep tumorigenesis, the role of global gene expression and the precise genetic alterations that induce pre-malignant cells to activate such programs are poorly understood.
1.6 The Epithelial-to-Mesenchymal Transition

The epithelial-to-mesenchymal transition (EMT), the process by which epithelial cells lose many of their epithelial properties and gain mesenchymal markers and characteristics, is centrally involved in metastatic dissemination (Lambert et al., 2017; Nieto et al., 2016; Shibue and Weinberg, 2017). This program involves a reorganization of the cytoskeleton and a loss of junctions and apical-basal polarity. It also results in the activation of a signaling mechanism that promotes invasion and motility as well as cooperation between the tumor cell and the surrounding microenvironment, all of which are necessary for metastatic dissemination. This process is reversible, and a mesenchymal-to-epithelial transition (MET) is involved in the seeding of new secondary tumors at distant sites.

While transcriptional and alternative splicing profiles associated with EMT are well documented (Lamouille et al., 2014; Shapiro et al., 2011; Warzecha et al., 2010; Yang et al., 2016), the genetic alterations and regulatory mechanisms that induce premalignant cells to activate EMT programs are less understood. There is much evidence that EMT is not a binary process, but is rather a dynamic program with multiple intermediary stages of partial, or incomplete, EMT (Nieto et al., 2016). In fact, the EMT program is usually only activated partially in human carcinomas, and partial EMT cells are observed in circulating tumor cells from breast cancer patients with each of the three major histological subtypes (Yu et al., 2013a). Thus, it is becoming increasingly clear that
EMT and gain of metastatic potential is a much more complex process than initially appreciated. EMT contributes to chemotherapy and immunotherapy resistance (Shibue and Weinberg, 2017), and therefore specifically targeting EMT is of clinical interest. However, EMT is a dynamic and reversible process, and the MET is known to be involved in metastatic colonization. Therefore, a full reversal back to an epithelial cell may not be ideal. Understanding how cells regulate the early stages of EMT and transitions to different EMT states is critical.

1.7 Cell culture systems in which to study progressive tumorigenesis

Cancers are derived from normal cells that evolve step-wise and progressively to a neoplastic state, and the involvement of PTR in this progression has not been looked at in the context of tumor initiation and step-wise progression. The earliest stages of oncogenesis, the processes by which a normal cell becomes progressively tumorigenic, are the most difficult to study. Comparisons of patient tumor samples to normal matched tissues do not give information about the early stages of transformation due to a lack of isogenic intermediate stages. While mouse models can help to illuminate genetic alterations that lead to cancer, many alterations characteristic of human cancers do not yield the same cancers in mice (Hooper, 1998; O'Hayer and Counter, 2006; Rangarajan et al., 2004). Cell culture models typically involve cell lines that are already immortalized, and thus abnormal. In order to study PTR during tumor initiation and step-wise cancer progression, we need a model system that recapitulates the
immortalization and subsequent transformation of a normal human cell. Work in the Weinberg laboratory demonstrated that normal human epithelial cells can be converted to a tumorigenic state through the expression of the viral proteins SV40 T-Ag and t-Ag and mammalian hTERT and Ras$^{\text{G12V}}$ (Counter et al., 1998; Hahn et al., 1999). Subsequent work in the Counter laboratory expanded on these findings, and identified a core set of proteins that together drive the process of tumorigenesis (Kendall et al., 2006; Kendall et al., 2005). The introduction of hTERT, p53$^{\text{DD}}$, cyclin D1, CDK4$^{\text{R24C}}$ and c-MYC$^{\text{T58A}}$ to primary epithelial cells immortalizes the cell, while subsequent expression of Ras$^{\text{G12V}}$ converts the cell to a fully tumorigenic state (Figure 3). Using this system, well-defined genetic models of tumorigenesis can be established for most epithelial cells, providing a cell culture model for the investigation of the earliest stages of tumorigenesis.
Figure 3: Intracellular pathways, as determined by the Counter laboratory, altered to induce immortalization and oncogenic transformation in normal human cells

1.8 Overview of Contents

In this dissertation, I set out to understand mechanisms of post-transcriptional regulation coordinating the early stages of tumor initiation and progression. Chapter 2 provides detailed explanations of all Materials and Methods used in this work. In chapter 3, I discuss the generation, optimization, validation and characterization of a primary cell-derived, genetically defined system of primary human mammary epithelial cell immortalization and oncogenic RAS transformation. I then show that this system can be used to uncover mechanisms of post-transcriptional regulation during
tumorigenesis that may be missed by other model systems. In **chapter 4**, I describe the comprehensive quantification of changes in mRNA abundance, alternative splicing and miRNA expression during each transition stage in this isogenic system, and I demonstrate that changes in global gene expression that are typically associated with malignancy occur prior to oncogenic transformation. This is the first comprehensive mRNA expression study in this cell culture model. Furthermore, I demonstrate that EMT-associated invasion can be initiated through two sequential stages: transcriptional priming followed by oncogenic RAS-triggered post-transcriptional regulation. In **chapter 5**, I explore a potential mechanism of RAS-triggered post-transcriptional regulation by RBPs. Using DO-RIP-seq, I identified and quantified HuR and CELF1 binding sites transcriptome-wide. This is the first study to quantify global RBP binding sites in dynamic biological conditions. I show that RNPs are not drastically remodeled during RAS transformation, but rather binding to a subset of binding sites changes quantitatively. I identify a GU-rich secondary motif that determines a decrease in HuR binding during transformation, and I suggest that HuR and CELF1 synergistically regulate a subset of mRNAs to promote the EMT phenotype. Additionally, I show that HuR is necessary for the maintenance of the RAS-induced cancer traits, but that it does not regulate mRNA stability in this context. I then identify a role for HuR in regulating APA via a binding site position dependent mechanism. Finally, in **Chapter 6** I discuss the overall conclusions reached in this dissertation as well as future directions.
2. Materials and Methods

2.1 Cell Culture

Human Mammary Epithelial Cell (HMECs) were obtained from Lonza at p9 and maintained in Mammary Epithelial Growth Medium (MEGM; Lonza, Walkersville, MD, USA). Immortalized (IMO) and Transformed (TFO) HMECs were maintained in MEGM+10% fetal bovine serum (FBS). Cells were passaged at 50-70% confluency using a Reagent Pack (Lonza) according to the manufacturer’s subculturing protocol. All HMECs were authenticated by analysis at the Duke University DNA Analysis Facility. 293T/17 packaging cells were obtained from the Duke University Cell Culture Facility and maintained in Dulbecco’s Modified Eagle’s Medium (ThermoFisher Scientific, Grand Island, NY, USA)+10% FBS. All cells tested mycoplasma negative.

2.2 Establishing a primary cell-derived system of tumorigenesis

2.2.1 Generation of amphotropic retroviruses

293T/17 cells were co-transfected with p0467/pcl-10A (gift of Dr. Chris Counter) and individual pBabe plasmids containing transgenes of interest (Table 1) using FuGENE-6 (Roche, Basel, Switzerland). 24 hours later, the transfection was repeated. 8 hours after the second transfection, the media was replaced with fresh MEGM, and cells were incubated in this new media for 48 hours. Amphotropic retrovirus-containing media was harvested from the cells and filtered through a sterile 0.45μm Acrodisc with an HT tuffryn membrane (VWR, Radnor, PA). All media was snap-frozen in liquid
nitrogen and stored at -80 C. Fresh MEGM was added to the 293T/17 cells, and 12 hours later media was harvested, filtered and snap frozen as done previously.

### Table 1: pBabe plasmids

<table>
<thead>
<tr>
<th>Transgene</th>
<th>Selection Marker</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>hTERT</td>
<td>Hygromycin</td>
<td>addgene plasmid #1773, gift of Dr. Bob Weinberg (Counter et al., 1998)</td>
</tr>
<tr>
<td>p53DD</td>
<td>G418</td>
<td>gift of Dr. Chris Counter (Hahn et al., 2002; Kendall et al., 2005)</td>
</tr>
<tr>
<td>Cyclin D1</td>
<td>Puromycin</td>
<td>addgene plasmid #9050, gift of Dr. William Hahn</td>
</tr>
<tr>
<td>CDK4R24C</td>
<td>Zeocin</td>
<td>gift of Dr. Chris Counter (Hahn et al., 2002; Kendall et al., 2005)</td>
</tr>
<tr>
<td>C-mycTS8A</td>
<td>Blasticidin</td>
<td>gift of Dr. Chris Counter (Kendall et al., 2005; Yeh et al., 2004)</td>
</tr>
<tr>
<td>H-RASG12V</td>
<td>YFP</td>
<td>gift of Dr. Chris Counter (Hahn et al., 1999)</td>
</tr>
<tr>
<td>Control</td>
<td>GFP</td>
<td>addgene plasmid #10668, gift of Dr. William Hahn</td>
</tr>
</tbody>
</table>

### 2.2.2 Retroviral transductions

Media containing the pBabe-hTERT amphotropic retroviruses was thawed at 37C. Polybrene (hexadimethrine bromide, Sigma-Aldrich, St. Louis, MO) was added to a concentration of 4µg/ml, and the media was added to primary HMECs. 12 hours later, the procedure was repeated with the second pBabe-hTERT amphotropic retrovirus containing media. This procedure was repeated every 12 hours until the HMECs were transduced with each plasmid listed in Table 1. Cells without H-RAS were saved as the
immortalized (IMO) cell line. Cells that were RAS-transformed are referred to as the transformed (TFO) cell line.

2.2.3 Generation of stable cell lines

Cells were selected sequentially with the following antibiotics, with a 5 day recovery between drug selections: Hygromycin: 80µg/ml for 7 days; G418: 250µg/ml for 10 days; Puromycin: 0.5µg/ml for 5 days; Zeocin: 800µg/ml for 8 days; Blasticidin: 4.5µg/ml for 7 days. RAS-transformed cells were sorted by flow cytometry for YFP expression.

2.2.4 Horizontal Spread Assay

Horizontal spread assays were performed as described in Bisogno and Keene, 2017. Briefly, conditioned media was collected from both IMO and TFO cells, filtered through a 0.45µm acrodisc filter and polybrene was added to a final concentration of 4µg/ml. Conditioned media was added to 293 cells. Protocol was repeated 24 hours later. 293 cells were then placed under antibiotic selection as detailed above, and media was replaced every 3 days until all 293 cells died.

2.3 Real-time PCR

Total RNA was isolated from cells using TriSure (Bioline, Luckenwalde Germany) and reverse transcribed using the iScript cDNA synthesis kit (Bio-Rad, Hercules, CA), both according to the manufacturer’s recommendations. Quantitative real-time PCR was performed using the Roche Lightcycler with Sybr green detection
(Invitrogen, Carlsbad, CA). Primers (Listed in Tables 2 and 3) were either designed to span introns or RNA was treated with DNaseI (New England Biolabs, Ipswich, MA) before reverse transcription, and a no RT control was used. Amplification of a single product was confirmed by melting curve analysis, and the ΔΔCt analysis method was used. Data are reported as either an agarose gel of the PCR end product or as mean and standard deviation with p-values calculated using a standard t-test.

**Table 2: Real-time PCR primer sequences used to check transgene expression.**

<table>
<thead>
<tr>
<th>Transgene</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>GapDH</td>
<td>F 5'-CAT GTT CGT CAT GGG TGT GAA CCA-3'</td>
<td>5'-AGT GAT GGC ATG GAC TGT GGT CAT-3'</td>
<td>Bisogno and Keene, 2017</td>
</tr>
<tr>
<td>hTERT</td>
<td>5'-GAG GTG CAG AGC GAC TAC-3'</td>
<td>5'- TCC ACA CCC TAA CTG ACA CAC A-3'</td>
<td>Kendall et al. 2005</td>
</tr>
<tr>
<td>p53&lt;sup&gt;DD&lt;/sup&gt;</td>
<td>5'-GCT CAC TCC AGC TAC CTG AAG A-3'</td>
<td>5'- TCC ACA CCC TAA CTG ACA CAC A-3'</td>
<td>Bisogno and Keene, 2017</td>
</tr>
<tr>
<td>Cyclin D1</td>
<td>5'-CCC AGC AGA ACA TGG ACC C-3'</td>
<td>5'-TTC TGC CTC GGG AG-3'</td>
<td>Bisogno and Keene 2017</td>
</tr>
<tr>
<td>CDK4&lt;sup&gt;R24C&lt;/sup&gt;</td>
<td>5' GAC TGG CCT CGA GAT GT A-3'</td>
<td>5'-TAC TTC TGC CTG CTG GGG-3'</td>
<td>Kendall et al. 2005</td>
</tr>
<tr>
<td>C-Myc&lt;sup&gt;T58A&lt;/sup&gt;</td>
<td>5'-ACT CGC TGC TGT CCT CCG A-3'</td>
<td>5'-GAG TGA GAC GTG GCA CCT CTG A-3'</td>
<td>Bisogno and Keene, 2017</td>
</tr>
<tr>
<td>HRAS&lt;sup&gt;G12V&lt;/sup&gt;</td>
<td>5'- GCA TCC CCT ACA TCG AGA-3'</td>
<td>5'-TAC TTC TGC CTG CTG GG-3'</td>
<td>Bisogno and Keene, 2017</td>
</tr>
</tbody>
</table>
### Table 3: Primer sequences of APA distal and universal primers for real time PCR

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Universal Forward</th>
<th>Universal Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>5’TCAGGGCAGGTATCCACCTA 3’</td>
<td>5’CGGGATTGGCTAAACTCCCA 3’</td>
</tr>
<tr>
<td>TMEM55a</td>
<td>5’AATGCAGGTGAGAGTGTCTAGC3’</td>
<td>5’GACTCCCAAAGTCTTGAAACGAT3’</td>
</tr>
<tr>
<td>RPN1</td>
<td>5’GGCAGATTGGGTGAGTAGTGG 3’</td>
<td>5’GGACAAACGGCAAACTCACA</td>
</tr>
<tr>
<td>CCNF</td>
<td>5’GTCCCTGCACACTGCGAG 3’</td>
<td>5’AGGTGTTCACATTGACGCCATC 3’</td>
</tr>
<tr>
<td>KIF11</td>
<td>5’GGCATTAACACACTGGAGAGT 3’</td>
<td>5’ATTCGCAACCCCCCAAATGAA 3’</td>
</tr>
<tr>
<td>LAMC2</td>
<td>5’AGTGTGACTGGTTGCACTC 3’</td>
<td>5’TTTGACAAGTGCTGGTG 3’</td>
</tr>
</tbody>
</table>

### Distal Forward | Distal Reverse

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Distal Forward</th>
<th>Distal Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>5’AGCTTATCTCGTGTCAATTGTC3’</td>
<td>5’AGCTAAGCCCCACCCITGATAC 3’</td>
</tr>
<tr>
<td>TMEM55a</td>
<td>5’TGGAATCAGCTCTCGCAGCT3’</td>
<td>5’AGTCCTATCTCGAAACTCACAAT3</td>
</tr>
<tr>
<td>RPN1</td>
<td>5’TTCCATTTGTCACACACGCT 3’</td>
<td>5’GCCCTGGGTGTATATCGCCTT 3’</td>
</tr>
<tr>
<td>CCNF</td>
<td>5’CCTCAATGCCACCTGCACACACA 3’</td>
<td>5’CCGACTGCTGTCTCTTC 3’</td>
</tr>
<tr>
<td>KIF11</td>
<td>5’TGGTGCACAAATGTAAAGG3’</td>
<td>5’GGGCACAGGCTTAGTC 3’</td>
</tr>
<tr>
<td>LAMC2</td>
<td>5’TGCAATTCCAGCTGACTCTG3’</td>
<td>5’CCATAGTCTTCCGAGCCAGG 3’</td>
</tr>
</tbody>
</table>
2.4 Soft Agar Assay

For each cell line to be tested, 0.75mL of 2x MEM (Thermo Fisher Scientific) was combined with 0.25mL sterile water and incubated at 37C for 15-20 minutes. 1.8% agarose was added to 0.6% and mixture was plated in one well of a 6-well plate. Agarose was allowed to solidify at 37C for 30-60 minutes. Cell lines were resuspended at 25,000-50,000 cells in 0.5ml sterile water, and combined with 0.75mL of 2x MEM and 0.25mL 1.8% agarose. Mixture was plated on top of the bottom agarose layer and incubated at room temperature for 30 minutes before moving to a 37C incubator. 1x media was added as needed to prevent drying, wells were imaged after 3 weeks. Four distinct fields were photographed and colonies were counted. Data is reported as the mean and standard deviation of at least three biological replicates, and p-values were calculated using a standard t-test.

2.5 Flow Cytometry

2.5.1 Cell Cycle Analysis

Cells were harvested with trypsin/EDTA at 50% confluent, and 100,000 cells were resuspended in 300µl PBS. Cells were then fixed in 700µl of 100% ethanol at -20C. Ethanol solution was removed, and fixed cells were resuspended in 500µl PBS. 1µg/ml Propidium Iodide (Thermo Fisher Scientific) and 200µg/ml RNAseA were added, and cells were incubated for 15 minutes at room temperature before being stored at 4C. All samples were analyzed within five days by the Duke Flow Cytometry Core Facility.
Results are reported as the mean and standard deviation of cell lines tested. P-values were calculated using a t-test.

### 2.5.2 Live Cell Staining and Sorting

Cells were harvested with trypsin/EDTA at 75-90% confluent and washed in cold PBS. 500,000 cells were resuspended in 100µl 1x annexin binding buffer (Thermo Fisher Scientific). Cells were stained with 5µl FITC-annexin V (Thermo Fisher Scientific) and 0.1µg/ml DAPI (Thermo Fisher Scientific). Cells were incubated for 15 minutes in the dark. 400µl of annexin binding buffer (Thermo Fisher Scientific) was added and stained cells were analyzed immediately by the Duke Flow Cytometry Core Facility. Results are reported as the mean and standard deviation of the percent of apoptotic (FITC positive), live (FITC and DAPI negative) and dead (DAPI positive) cells in each cell line tested.

### 2.6 Wound Healing Scratch Assay

Cells were grown in a single monolayer in 6-well plates. Cells were serum starved overnight and during experiment to control for proliferation. A single scratch was made with a p200 pipette tip and the scratch was photographed in 3 distinct places. After 24 hours the scratch was imaged in the same 3 locations, and percent wound closure was quantified using ImageJ’s MRI wound healing tool. Data is reported as the mean and standard deviation of 4 replicates. P-values were calculated using a t-test.
2.7 *Matrigel Assay*

Matrigel membrane matrix (Corning, Corning, NY, USA), was thawed on ice overnight and diluted to 300µg/ml in cold coating buffer (0.01M Tris pH 8, 0.7% NaCl, filtered). 100µl of Matrigel was added to 6.5mm Transwell with 8.0µm Polycarbonate membrane inserts (Corning), and allowed to solidify at 37°C for about 2 hours. After gelling, inserts were washed gently with warm serum free media. 600µl of media with serum was added to the lower chamber. Cells were harvested with Trypsin/EDTA, washed 3 times with serum free media, and resuspended in serum free media at ~50,000 cells/ml. 200µl of cell suspension was added to the upper chamber. No Matrigel control inserts were used in every experiment. After a 12 hour incubation, cells that did not invade were scraped off the top chamber with a cotton swab. Cells were fixed for 2 minutes in methanol, and stained with 1% Toluidine Blue in 1% Borax for 2 minutes. Cells were washed with distilled water twice for 2 minutes and allowed to air dry. 4-6 distinct fields were photographed and cells were counted. Data is reported as the mean and standard deviation of 3 replicates, quantified as percent that migrated through Matrigel vs. through the membrane without Matrigel. P-values were calculated using a t-test.

2.8 *Western Blots*

Whole cell lysates were prepared by scraping into cold 0.1% SDS lysis buffer. Protein was cleared by centrifugation, quantified with a Bradford assay, and boiled in
Laemmli loading buffer. Protein was then size separated on a criterion SDS PAGE gel (Biorad, Hercules, CA, USA). Protein was transferred onto nitrocellulose membrane, and transfer was verified using Ponceau S staining. Membranes were blocked with 5% Non-fat milk in TBS+0.1% Tween20 (TBST) for one hour at room temperature, incubated with primary antibody at 4C overnight, washed thoroughly with TBST, incubated with secondary antibody for one hour at room temperature, washed extensively with TBST, and then developed using SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher Scientific), and exposed to film.

The following primary antibodies were obtained from Santa Cruz (Dallas, TX, USA): N-cadherin, sc-271386, 1:250; Vimentin, sc-6260, 1:250; E-cadherin, sc-21791, 1:500; GapDH, sc-47724, 1:1000, B-tubulin, sc-5274, 1:1000. CSTF64 antibody was from Bethyl laboratories and used at a 1:1000 dilution. HuR was detected using 3A2 serum 1:500. The anti-mouse and anti-rabbit HRP conjugated secondary antibodies were purchased from Jackson ImmunoResearch (West Grove, PA, USA) and used at a 1:20000 dilution.

2.9 RNA sequencing

Total RNA was obtained from cells using the TRIsure reagent (Bioline) according to the manufacturer’s protocol. RNA quantity was analyzed using a NanoDrop spectrophotometer, and ribosomal RNA was depleted using the Ribozero Gold kit (Epicentre, Madison, WI, USA). 50ng of rRNA depleted RNA was used as input for sequencing libraries, which were made using the ScriptSeq v2 RNA-seq library.
preparation kit (Epicentre) according to the manufacturer’s instructions. 16 amplification cycles were used in the final PCR amplification step. Libraries were checked with a BioAnalyzer, and sequenced on the Illumina Hi-Seq 2000/2500 platform (125bp PE) or the Illumina Hi-Seq 4000 platform (150bp PE) at Duke University’s Center for Genomic and Computational Biology. We obtained approximately 50-80 million reads per library. Reads were mapped to the human genome (hg19) using TopHat2 (Kim et al., 2013), and significant changes in gene expression patterns were determined using Cufflinks/Cuffdiff (Trapnell et al., 2010). Gene ontology was analyzed using GOrilla (Eden et al., 2007; Eden et al., 2009) and semantic clustering of GO terms was performed with REVIGO (Supek et al., 2011).

2.10 Alternative splicing

Alternative splicing was analyzed using paired-end RNA-sequencing data and the VAST-tools program as previously described (Irimia et al., 2014).

2.11 microRNA profiling

Total RNA was obtained from cells using the TRIsure reagent (Bioline) according to the manufacturer’s protocol. RNA quantity was analyzed using a NanoDrop spectrophotometer. A poly(A) tail was added to 1µg of total RNA using Epicentre’s Poly(A) Polymerase Tailing Kit. PolyA tailed RNA was reverse transcribed using qScript reverse transcriptase (Quanta Biosciences, Gaithersburg, MD, USA) and a custom universal tag coupled to oligo(dt) (Lykken and Li, 2016). cDNA was diluted in Tricine-
EDTA. miRNA quantity in each sample was assessed using qPCR arrays using PerfeCTa SYBR Green SuperMix (Quanta Biosciences) and miRNA specific primers (Lykken and Li, 2016). Arrays were performed in duplicate, and data is reported as delta ct values normalized using global mean normalization.

2.12 Digestion Optimized Ribonucleoprotein Immunoprecipitation and RNA-sequencing (DO-RIP-seq)

DO-RIP-seq experiments were performed as described in detail by our lab (Nicholson et al., 2016). A brief description of the methods with details specific to DO-RIP in TFO and IMO HMECs is described below.

2.12.1 Preparation of cell lysates

5-6 15cm plates (~80% confluent) of IMO and TFO cells were used as DO-RIP inputs. Cell lysates were harvested in polysome lysis buffer without MgCl₂ (PLB: 100mM KCl, 10mM HEPES (pH 7.0), 0.5% NP40, 1mM DTT, 100 units/mL RNase Out (Thermo Fisher Scientific), 1x Protease Inhibitor Cocktail (Roche)) and frozen at -80°C. Lysates were thawed on ice and precleared at 15,000 xg for 15 minutes prior to use.

2.12.2 Digestion of cell lysates

CaCl₂ was added to a 5mM final concentration. Lysates were digested with 2µl of micrococcal nuclease (New England Biolabs) per 15cm plate of cells for 5 minutes at 30°C, as determined through optimization procedures described previously (Nicholson et al.,...
Reactions were stopped after 5 minutes through the addition of EGTA to 10mM final concentration. ~5% of digested RNA was saved as “input” normalizing controls.

2.12.3 Preparation of magnetic beads

60µl Protein G Dynabeads (Invitrogen) were pre-coated with 350µl 3A2 HuR serum or 10µl CELF1 (CUGBP1, Mouse Monoclonal, MBL) overnight at 4C. Beads were washed 4x with NT2 buffer (50mM Tris-HCl (pH 7.4), 150mM NaCl, 1mM MgCl2, 0.05% NP40) and resuspended in 750µl NT2.

2.12.4 Immunoprecipitation

150µl of digested lysate was added to the washed beads along with 1mM DTT, 20mM EDTA and 100 units of RNase Out. IP reaction was rotated for 4 hours at 4C. Beads were then washed 4x with cold NT2, and RNA was isolated from the beads using TriSure reagent (Bioline) according to the manufacturer’s protocol.

2.12.5 Preparation, sequencing and analysis

RNA was dephosphorylated with antarctic phosphatase (New England Biolabs), and radiolabeled with [γ-P32] using a T4 Polynucleotide kinase reaction (New England Biolabs). RNA was separated on a 15% polyacrylamide TBE-urea gel at room temperature, and RNA between 25-75 nucleotides was extracted and eluted from the gel as described previously (Nicholson et al., 2016). Sequencing libraries were generated and validated using the NEBNext small RNA library prep protocol (New England Biolabs) as described (Nicholson et al., 2016). “Input” normalizing controls were first depleted of
ribosomal RNA using the RiboZero Gold Kit (Illumina). Libraries were sequenced on the Illumina Hi-Seq 2000/2500 platform (100bp, single end). DO-RIP-seq libraries were analyzed as described previously (Nicholson et al., 2017; Nicholson et al., 2016).

2.13 CRISPR Knockout of the HuR RBP

Custom plasmids containing guide RNAs targeting HuR fused with Cas9 and GFP protein were ordered from Horizon Discovery. Both guide RNAs targeted exon 2. gRNA E1: TAATGGTTATGAAGACCACA (sense) gRNA E2: CCTGGGTCATGTTCTGAGGG (antisense). IMO and TFO HMECs were transfected with each guide RNA using Lipofectamine 2000 (Thermo Fisher Scientific). 24 hours after transfection, cells were sorted with flow cytometry into 96-well plates. All surviving colonies were screened for HuR protein expression with an immunoblot (HuR 3A2 serum 1:500) and mutations were confirmed with Sanger sequencing of HuR exon 2 from genomic DNA of the clones. Primer sequences: Forward: 5’ACAAGAGCAATGGACAGAGTT 3’ Reverse: 5’TGCAGTTACTAGTTTTGCCTCA 3’. Mutations were analyzed by aligning sequences to the WT exon 2 sequence using NCBI BLAST align tool. Editing efficiency in each clone was assessed and indels were identified using the TIDE webtool (Brinkman et al., 2014).

2.14 Measuring proliferation rates in real time

Proliferation rates were measured using an xCELLigence Real Time Cell Analyzer (RTCA) system (ACEA Biosciences, San Diego, CA), a computer controlled
real-time automated electro-sensing platform. 500 or 1000 TFO cells were plated in 96-well plates with gold microelectrode coated wells (E-plate 96; ACEA Biosciences, Inc). After plating, cells were allowed to settle for 30 minutes at 37°C in a 5% humidified CO₂ chamber. 30 minutes after plating, proliferation rates were monitored in real time by the RTCA system for 5 days, with impedance measurements taken every 15 minutes. The impedance is proportional to the number of cells on the surface of the well. Impedance readings were transposed by the RTCA software into cell number and plotted against time to generate growth curves. Doubling time was calculated by dividing the slope of the exponential growth phase by the constant k=0.693 (Ali-Osman, 1996; Lidsky et al., 2014).

2.15 Luciferase Plasmids

TFO HMEC genomic DNA was isolated using the QiaAmp DNA mini kit (Qiagen, Hilden, Germany), and full length 3’UTRs of RPN1 and TMEM55a were isolated from gDNA using primers that incorporated a 5’ BamH1 and 3’ EcoRV. PCR products were TA cloned into pGEMT (Promega, Madison, WI) before being subcloned into the MCS of PCDNA3 mammalian expression vectors containing firefly luciferase. Specific HuR binding sites of interest were mutated using New England Biolab’s Q5 site directed mutagenesis kit according to the manufacturer’s protocol. All constructs were confirmed with Sanger sequencing.
3’UTR cloning primers:

RPN1 Forward: 5’ TAATTGGATCCCCCTGCCCGCATCCT 3’ (BAMH1); Reverse: 5’
GCGGCGATATCAATGTTCACAGTCCCTGCTTT 3’ (EcoRV)

TMEM55a Forward: 5’ TCCTTGGATCCGGTGTATGGATTCACTGA 3’ (BAMH1);
Reverse: 5’ CGGGGGATATCTGGAAATAATCTTTATTAG 3’ (EcoRV)

Mutagenesis Primers:

RPN1 Forward: 5’TCCTTAAGAAAAACTTTTTTTTTTTTTTTCC 3’; Reverse: 5’
CTTTAAGGACAAACGGCAAC 3’- This primer set resulted in a deletion of a 26nt T-rich region identified as being a strong HuR binding site.

TMEM55a Forward: 5’ AGTTATTTCATTACCTGCTGATA 3’; Reverse: 5’
CAAATCCAAATCATCATCTATAAGAGATG 3’- This primer set resulted in a deletion of a 46nt T-rich stretch identified as being a strong HuR binding region. It also resulted in a single nucleotide T to G conversion (underlined) in the middle of a T-rich stretch.

2.16 Luciferase Assays

Equimolar quantities of luciferase reporter vectors were co-transfected with a Renilla luciferase expression vector (pRL) into HuR WT and HuR KO TFO cells using Lipofectamine2000 reagent (Thermo Fisher Scientific). PCDNA3.1 luciferase empty vector (EV) +/- pRL were used as controls. 24 hours after transfection, cells were harvested using passive lysis buffer (Promega). Dual luciferase assays were performed on a CLARIOstar luminometer (BMG Labtech, Ortenberg, Germany) using the Dual-
Luciferase Reporter 1000 Assay System (Promega). The ratios of Firefly to Renilla signal were normalized to PCDNA3.1 EV, and data is reported as log 2 fold change of normalized signal of the mutant constructs compared to the wild type constructs. Data are reported as mean and standard deviation, and p-values were calculated using a standard t-test.

2.17 Ribonucleoprotein Immunoprecipitation (RIP)

Cell lysates used for RIPs were harvested in polysome lysis buffer (PLB: 100mM KCl, 5mM MgCl₂, 10mM HEPES (pH 7.0), 0.5% NP40, 1mM DTT, 100 units/mL RNase Out (Thermo Fisher Scientific), 1x Protease Inhibitor Cocktail (Roche)) and frozen at -80°C. Protein G Dynabeads (Invitrogen) were pre-coated with 1µg CSTF64 antibody (Rabbit Polyclonal, Bethyl Laboratories, Montgomery, TX) or Normal Mouse Serum (NMS) negative control overnight at 4°C. Lysates were thawed on ice and precleared at 15,000 xg for 15 minutes. Beads were washed 4x with NT2 buffer (50mM Tris-HCl (pH 7.4), 150mM NaCl, 1mM MgCl₂, 0.05% NP40), resuspended in 750µl NT2, and tumbled with 150µl cleared lysate, 1mM DTT, 20mM EDTA and 100 units of RNase Out for 4 hours at 4°C. Beads were then washed 4x with cold NT2, and RNA was isolated from the beads using TriSure reagent (Bioline) according to the manufacturer’s protocol.
3. Generation and validation of a primary cell-derived isogenic model of tumorigenesis and demonstration of its utility in studying post-transcriptional regulation

3.1 Background and Significance

Global studies have identified many RBPs that are significantly misexpressed in tumors compared to normal tissues (Galante et al., 2009; Kechavarzi and Janga, 2014), and several studies have suggested that RBPs dynamically and differentially regulate target mRNAs in different states and contexts (Mazan-Mamczarz et al., 2008a; Mazan-Mamczarz et al., 2008b; Mukherjee et al., 2011; Mukherjee et al., 2009; Papadaki et al., 2009). However, cancer cells are derived from normal cells that often evolve step-wise and progressively to a neoplastic state, and the involvement of PTR in this progression has not been looked at in the context of tumor initiation and step-wise progression. Thus, more studies are needed in order to fully understand the PTR regulators and downstream genetic programs activated by cancer driver mutations that coordinate tumor origins, evolution and progression.

Studies of tumorigenesis typically involve the use of cancer cell lines. While cell line models have been informative and have demonstrated the importance of transcriptional and post-transcriptional control of cancer (Chaudhury et al., 2016; Mazan-Mamczarz et al., 2008a; Mazan-Mamczarz et al., 2008b), they typically involve cell lines that are already immortalized, and thus they may already be well along the path to becoming a cancer cell. For example, it is common to set up an isogenic system of
late stage tumor formation by adding an oncogene, such as H-RAS\textsuperscript{G12V}, to an already immortalized cell line, such as the MCF10A human mammary epithelial cell line. However, one major drawback of such an isogenic system is that it cannot be used to study the earliest stages of oncogenesis, the process by which a normal cell becomes progressively tumorigenic, since there are no normal isogenic primary cells to compare. Likewise, comparisons between patient tumor tissues and normal matched tissues do not provide information about the intermediate stages of transformation. In addition, certain cancer cell lines, such as the ever-popular MCF7 breast cancer cell line, have been demonstrated to evolve and adapt differently over time (Lee et al., 2015). Thus, findings from one laboratory may not be replicable in the same cell line in another laboratory due to accumulated mutations in these lines. While mouse models can help to illuminate genetic alterations that lead to cancer, many alterations characteristic of human cancers do not yield the same cancers in mice (Hooper, 1998; O'Hayer and Counter, 2006; Rangarajan et al., 2004). The primary cell-derived model system of step-wise immortalization and tumorigenesis depicted in Figure 3 offers a system to study early stages of tumorigenesis without confounding factors due to underlying cell line differences. Additionally, rather than using standard cancer cell lines as pre-clinical models, establishing multiple donor primary cell-derived isogenic lines in which to conduct initial studies could be highly beneficial in studying individuals with different genetic backgrounds. Despite this benefit, relatively few laboratories have taken
advantage of the cell culture models developed by the Weinberg and Counter laboratories. We therefore decided to use this system to study PTR during tumor initiation and step-wise progression. Unfortunately, when we went to set up this system, these cell lines were not available to be shared between laboratories due in part to the fact that primary cells have extremely limited passage numbers and are in short supply. While the immortalized and transformed lines could be shared if available, the main advantage of this system, the comparison to primary cells, would be lost. In addition, if one is to make comparisons between different individuals of various genetic backgrounds, multiple primary cells from several donors must be obtained and step-wise transformed. Therefore, the cell lines must be engineered by the laboratory wishing to use them. However, the process to make these cell lines is time consuming and laborious. We therefore decided to streamline and simplify the protocol and validation steps.

The process of engineering a stable isogenic system in the absence of viral genes has been described (Kendall et al., 2005; O'Hayer and Counter, 2006). We have simplified this method with a few useful modifications to streamline the numerous steps of transfection and infection involved in establishing this system. Details of the modified protocol can be found in Chapter 2 of this dissertation and in a recent publication (Bisogno and Keene, 2017). In this chapter, we briefly describe our modifications to the method and optimizations to the cell growth conditions. A workflow of the procedure

50
for establishing the isogenic model is outlined in Figure 4. Additionally, we characterize several oncogenic properties of this cell system. We then use RNA-sequencing to reveal large changes in RBPs during the early transition to immortalization, demonstrating the utility of this system for capturing early PTR events that may be missed by other model systems.

Figure 4: Workflow to establish and validate a primary cell-derived system of tumorigenesis.

hTERT, p53DD, cyclin D1, CDK4R24C and c-MycT58A are ectopically expressed in normal human primary cells to generate immortalized, intermediary cells. Immortalized cells are subsequently transformed with HRASG12V. All transfections are done through generating amphotropic retroviruses in a 293T packaging cell line and infecting the cells of interest. Stable cells are selected through either antibiotic selections or fluorescent sorting. Cells are then checked with a horizontal spread assay to ensure they do not shed recombinant retrovirus. Transgene expression and cell phenotype is then validated with qRT-PCR and soft agar assays. Once verified, many global gene expression studies, drug tests, and post-transcriptional assays can be carried out in this system.
3.2 A streamlined method for establishing a primary cell-derived cell culture model of tumorigenesis

We reconstructed the genetically defined cell culture model of stepwise neoplastic transformation originally developed by the Weinberg lab and modified by the Counter lab (Bisogno and Keene, 2017; Hahn et al., 1999; Kendall et al., 2005). Primary HMECs (PRIM) were immortalized through the sequential, stable expression of hTERT, p53DD, cyclinD1, CDK4R24C, and C-MYC\textsuperscript{T58A}. This pre-malignant, immortalized cell line (IMO) was used to create a third, transformed cell line (TFO) through the expression of H-RasG12V.

We first tried to sequentially express and select for each transgene individually as outlined in a methods paper by O’Hayer and Counter (O’Hayer and Counter, 2006). However, this method was specific to the SV40 system of immortalization and subsequent transformation. When used to express multiple mammalian transgenes, we found that PRIM cells senesced before becoming immortalized. We therefore sought to establish a more rapid method for immortalizing the cells. Our modified method is described in detail in the Materials and Methods section and in a recent publication (Bisogno and Keene, 2017). In short, all amphotropic retroviruses were made at one time and frozen in liquid nitrogen. Then, PRIM cells were sequentially infected with retroviruses containing each transgene without selection and without splitting cells in between each infection. With this method, PRIM cells were rapidly immortalized in just 9 days. After cells were immortalized, we made stable cell lines with antibiotic selection.
We then RAS-transformed the IMO cells with YFP-tagged H-RAS^{G12V}, and sorted with flow cytometry to enrich for RAS-positive cells. All cell lines were checked with a horizontal spread assay to make sure that they did not shed recombinant retroviruses. Once the cells were immortalized, we added 10% FBS to the HMEC MEGM media; we found that IMO and TFO cells grew in spheroids when cultured in serum free media.

### 3.3 Validation of transgene expression and tumorigenic potential

After generating this system from normal HMECs, we confirmed expression of all transgenes with quantitative real-time PCR. Cells were retested for transgene expression after being maintained in selection-free media for 20-25 additional passages post-selection, and all transgenes were still expressed (Figure 5A). As an additional verification, transformed cells were over 80% fluorescent at p46 (Figure 5B). This confirmed that, once stable cells were made, cells continued to express the transgenes of interest when selective media is removed. Therefore, one does not need to maintain these cells in selective antibiotics once proper selection is carried out. This is likely due to the fact that all transgenes confer a growth and survival advantage to cells.

We used soft agar assays, an *in vitro* proxy for tumor forming potential, to confirm that only transformed cells exhibit anchorage independent growth (Figure 6). Likewise, this assay was performed initially and then repeated at 20-25 additional passages. Results were replicated, indicating that the phenotype was stable over time. Additionally, we observed that IMO and TFO cells had increased proliferation.
compared to PRIM cells, and we demonstrated that IMO and TFO cells have a greater proportion in S phase and fewer in G2/M phase (Figure 7). Interestingly, PRIM cells grew in distinct clusters, whereas IMO and TFO cells had a scattered pattern (Figure 8), suggesting altered cell adhesion during immortalization.

Figure 5: Confirmation of transgene expression.

A) All transfections were validated by transgene confirmation using quantitative RT-PCR for the PRIM, IMO, and TFO cell lines, and a “no template” control (ctrl). The imo and tfo cells were maintained in antibiotic-free media for 20-25 passages before reconfirming transgene expression. The final PCR products are shown. B) Transformed cells were sorted using FACS at p46. As seen in the table, the selected cells maintained over 80% fluorescence.
Figure 6: Anchorage independent growth

Soft agar assays were used to test for anchorage independent growth. Representative images of cell fields are shown to scale for each of the three cell lines. The data were quantified by photographing 4 fields of 4 independent experiments. The mean and standard deviations of the 4 fields were calculated as number of colonies per field as shown in the graph. These images confirm that the RAS-transformed cells form large spheroidal foci, the expected phenotypic property.
Cell Cycle Analysis of Primary (PRIM), Immortalized (IMO), and Transformed (TFO) HMECs. Data is reported as percent of cells in each phase, * p<0.05.

Phase contrast micrographs of Primary (PRIM), Immortalized (IMO), and Transformed (TFO) HMECs at low confluency (top) and high confluency (bottom).
3.4 RBP{s} associated with malignancy are upregulated during immortalization

In order to determine if this system could be useful as a model for studying PTR during the early stages of tumorigenesis, we assessed the expression dynamics of RBP{s} in this system using RNA-sequencing. Total RNA was extracted from each cell type, and sequencing libraries were made using Epicentre’s Scriptseq v2 kit with 16 amplification cycles. Samples were sequenced on an Illumina Hi-Seq 2000 (one replicate; 125bp PE) and an Illumina Hi-Seq 4000 (two replicates; 150bp PE). There were approximately 60-80 million reads per library. Reads were mapped to the human genome (hg19) using TopHat2 (Kim et al. 2013), and significant changes in gene expression patterns were determined using Cufflinks/Cuffdiff (Trapnell et al. 2010).

GO category analysis revealed significant enrichment in upregulated genes for post-transcriptional processes, including nuclear part, polyA binding, RNA processing, RNA splicing and Ribonucleoprotein complex during immortalization (Figure 9). These changes were significant during immortalization but not during transformation.

To see if our findings could have potential clinical significance, we looked for expression levels of RBP{s} identified by two recent in silico studies as over-expressed in a variety of patient tumors compared to normal tissues (Galante et al., 2009; Kechavarzi and Janga, 2014). We combined published lists of RBP{s} that were strongly upregulated across 9 different cancers profiled in the TCGA database compared to normal tissues sequenced as part of the Human BodyMap 2.0 project (Kechavarzi and Janga, 2014), or that were
significantly upregulated specifically in breast tumors compared to normal breast cells as determined using SAGE data and ONCOMINE microarray datasets (Galante et al., 2009). The sum of these two data sets was 58 significantly upregulated RBPs. We then filtered this dataset for those RBPs that had a mean FPKM value greater than 1 for all three conditions across all of our RNA-seq replicates. This list consisted of 46 expressed RBPs, 26 of which had significant changes in mRNA expression during immortalization and none of which changed significantly during transformation (Figure 10).

Figure 9: HMEC immortalization increases RNA expression of RBPs.

GO category analysis of RNA-seq data shows significant changes in RNA regulatory processes during immortalization. Horizontal bars indicate statistically significant increases during the transition from primary HMECs to immortalized HMECs. These changes in expression were not observed during RAS transformation.
Figure 10: Cancer-related RBPs are regulated during immortalization.

RBPs previously demonstrated to be upregulated in patient tumors show statistically significant expression changes during immortalization, but not transformation. Log 2 fold change mRNA levels during immortalization (Prim to Imo) and transformation (Imo to Tfo) for a subset of RBPs found in previous studies to have significant RNA expression changes in tumor samples compared to normal tissue.

Many of the RBPs identified as being alternatively expressed in tumors versus normal tissues have previously demonstrated roles in cancer. For example, TIA1 promotes stress granule assembly and has several putative cancer-relevant targets (Anderson and Kedersha, 2002; Wigington et al., 2015). CSTF2, an alternative
polyadenylation factor, is upregulated, and it has been suggested that cancer cells exhibit high incidences of alternative polyadenylation (Mayr and Bartel, 2009).

Interestingly, mRNA encoding the ADAR RNA editing enzyme is upregulated as well. A-to-I editing is highly prevalent in tumor cells compared to normal cells, and ADAR silencing reduces proliferation and increases apoptosis (Fumagalli et al., 2015).

In addition, we identified many other RBPs that have altered RNA expression in immortalized cells compared to primary cells. Notably, we found ESRP1, an epithelial splicing factor known to be suppressed in metastatic cells (Warzecha et al., 2010), to be significantly downregulated. In contrast, Sam68/KHDRBS1, an RBP that regulates a cancer-promoting alternative splicing program (Wurth, 2012), is significantly upregulated. Other examples of RBPs previously demonstrated to be cancer-promoting that are upregulated in immortalized cells include ELAVL1 (Abdelmohsen and Gorospe, 2010), LARP1 (Hopkins et al., 2016), and hnRNPM (Xu et al., 2014).

These data suggest that RBP expression differences that are important in human tumors can occur early in the process, before transformation. These RBPs are therefore high priority candidates for future studies of tumor initiation. It is likely that previous studies that used model systems that begin with immortalized cell lines rather than primary cells for comparison could fail to identify important changes in RBP dynamics that regulate tumor initiation and progression.
3.5 Discussion

The earliest stages of oncogenesis are the most difficult to study. We established and optimized a streamlined protocol for generating and validating a primary cell-derived isogenic model of the early stages of tumor initiation using HMECs. This model should be adaptable to most types of human cells, allowing for the study of the early stages of many different tumor types for which normal primary cells are available. When selecting primary cells of origin, the amount of material that will be needed in downstream applications should be taken into consideration. Primary cells must be able to divide several times in cell culture if one is to produce enough material for downstream methods such as RNA-seq and RIP/CLIP-seq methods. We used commercially available HMECs that had undergone post-stasis selection and had spontaneously lost expression of p16\(^{INK4}\) (Huschtscha et al., 1998), allowing them to undergo several additional population doublings.

We intentionally did not use any viral genes to immortalize the cells, as was originally done by the Weinberg laboratory (Hahn et al., 1999). SV40 is not believed to be causative in the formation of human tumors (Gazdar et al., 2002; Shah, 2007). Additionally, expression of SV40 has been demonstrated to induce DNA damage response pathways (Boichuk et al., 2010; Hein et al., 2009) and to target several host pathways in addition to the minimal perturbations needed for immortalization (Ali and DeCaprio, 2001; Skoczylas et al., 2004). Therefore, while SV40-containing isogenic
systems, such as the common HMLE lines, may be attractive to researchers due to the fact that they are readily available and relatively easy to culture, they may not accurately recapitulate in vivo changes during the onset of human tumorigenesis.

Several published studies have also utilized isogenic models without introducing viral oncoproteins to study tumorigenesis. The immortalized MCF10A cell line has been used in many labs to study changes in PTR after perturbations such as TGFβ induced epithelial to mesenchymal transition (EMT) or oncogenic transformation. These studies have been useful for revealing important mechanisms of PTR. However, these studies are limited to understanding PTR during later stages of tumor progression, as these cell line models do not allow direct comparison with the normal primary source. Our finding that the most significant changes in genes encoding RBPs are observed during the transition step from normal to immortalized, not during the step to transformation, suggests that PTR plays an important role in the very early stages of tumor formation. These changes could be missed in many other standard cell line models.

Numerous diseases, including cancers, have been associated with altered expression and functions of RBPs (Castello et al., 2012; Lukong et al., 2008; Wurth, 2012). We previously estimated there are approximately 1500 RBPs in mammalian cells (Keene, 2001). Recently, there has been significant progress made towards mass discovery of RBPs, and over 800 have been identified to date (Castello et al., 2012). Additionally, efforts have been made to understand and quantify dynamic changes in RBP expression,
targeting and regulation in dynamic settings, including disease states, stress conditions and drug responses (Castello et al., 2012; Castello et al., 2016; He et al., 2016; Mukherjee et al., 2011; Mukherjee et al., 2009; Munschauer et al., 2014; Nicholson et al., 2017; Strein et al., 2014). This primary cell-derived isogenic system, when combined with next-generation sequencing, is a valuable approach for identifying pertinent RBP expression changes occurring during biologically dynamic states that are otherwise difficult to identify. Many RBPs experimentally determined to be involved in tumorigenesis were initially identified due to aberrant expression levels in cancers (Galante et al., 2009). Therefore, the RBPs identified as being dynamic in this system (Figure 10) open the door for future studies of PTR in tumor initiation and progression. In addition, parallel comparisons of RBP dynamics in a cohort of primary cell-derived isogenic cancer systems could provide insights into how patients with diverse genetic backgrounds respond differently to drugs targeting PTR.
4. Gene expression dynamics during tumorigenesis revealed that RNA signatures and phenotypes occur during distinct transition states

4.1 Background and Significance

It is well accepted that changes in gene expression are abundant during cancer progression, and changes in the cancer transcriptome have been extensively profiled (Wesolowski and Ramaswamy, 2011; Wouters et al., 2017). However, these populations of cells are generally heterogeneous with overlapping states of progression towards malignancy, and this can lead to imprecise assessments of which genes are truly active or inactive at any given time during the progressive evolution of a tumor. Furthermore, within these heterogeneous populations are many mutations that are not necessarily driving tumorigenesis, thus understanding how contributions of different mutations result in a given cancer phenotype and genotype has been difficult. One approach to this dilemma has been to use isogenic cell culture models in which specific genetic mutations are introduced into non-malignant cell lines in order to drive cancer progression. However, these studies generally start with cell lines that are immortalized, and thus already on the pathway to becoming malignant. An ideal cell culture model is one that captures multiple transitions of a normal epithelial cell as it progressively evolves towards a malignant state. The Counter lab previously demonstrated that the RAS transformed HMECs, but not the pre-malignant line, form tumors in immunocompromised mice, and, although these tumors did not appear to metastasize,
the tumors were surrounded by areas with local tissue invasion (Kendall et al., 2005). Invasion is associated with a characteristic EMT, whereby cells lose epithelial properties and gain mesenchymal properties. EMT activates a signaling mechanism that leads to invasion and motility and promotes cooperation between the tumor cell and the surrounding microenvironment (Nieto et al., 2016). Additionally, our observation that IMO and TFO cells grew in a scattered pattern, in contrast to PRIM cells which grew in clusters, suggested altered cell adhesion, a property of EMT. We therefore hypothesized that this genetically defined cell culture system of stepwise tumorigenesis may provide a model with which to study discrete transition states during both tumor evolution and EMT. Therefore, in this chapter we extensively characterized and profiled the primary-derived isogenic cell system described in Chapter 3. This is the first comprehensive gene expression analysis in a primary cell-derived model of stepwise breast tumorigenesis.

4.2 Analysis of Migration and Invasion

We first measured migratory properties of IMO and TFO cells using a wound-healing scratch assay. Cells were grown to confluency and a single scratch in the cell monolayer was quantified by degree of closure after 24 hours. We found that both IMO and TFO cells were migratory, but had no significant differences in migration rate (Figure 11a). We also quantified invasion through a Matrigel membrane, and, in contrast to migration, RAS-transformation increased invasiveness by ~3-fold compared to IMO
(Figure 11b). As expected, PRIM cells were unable to migrate through a membrane with or without Matrigel, indicating that they were neither migratory nor invasive.

We then tested for protein expression of canonical epithelial marker, E-cadherin, and mesenchymal markers, Vimentin and N-cadherin (Figure 12). As expected, PRIM cells robustly expressed E-cadherin and lacked N-cadherin expression, but PRIM cells did express Vimentin protein. In contrast, the migratory IMO cells had lost expression of E-cadherin, but had low expression of mesenchymal markers; very low levels of N-cadherin protein were detected in IMO cells, and Vimentin levels were lower than in the PRIM cells. TFO cells lacked E-cadherin expression and had high expression levels of Vimentin and N-cadherin when compared to PRIM and IMO cells. These results demonstrated that the IMO cell line exhibited an intermediate EMT phenotype, while oncogenic RAS-transformation induced a strong mesenchymal phenotype. Thus, this primary cell-derived model of tumorigenesis is a model of progressive EMT.
Figure 11: Migration and Invasion

Immortalized cells exhibit an intermediate EMT phenotype, while RAS transformed cells have undergone a fully invasive EMT. A) Wound-healing scratch assay. Representative photos and quantification of all biological replicates (n=4). B) Transwell assay with and without Matrigel. Shown are representative images and the mean and standard deviation of 3 biological replicates, quantified as percentage that migrated through Matrigel vs. through the membrane without Matrigel.
4.3 mRNA expression analysis

4.3.1 Large changes in gene expression occur during immortalization, while relatively few changes occur during RAS transformation

The reprogramming of gene expression is known to be a key step in the transdifferentiation of epithelial cells to mesenchymal cells during EMT (Lamouille et al., 2014). In order to compare the gene expression changes between a normal epithelial cell and a migratory cell and an invasive cell, we extracted total RNA from PRIM, IMO and TFO cells for deep sequencing. We performed paired-end RNA-seq for each sample.
in biological triplicates, as described in Chapters 2 and 3. Surprisingly, the mRNA expression changes comparing the PRIM to IMO transition with the PRIM to TFO transition were very similar (Figures 13a and 13b). The largest number of significant gene expression changes was observed during the early IMO transition (3809 genes changed, FDR q-val<0.05), with relatively few significant changes occurring in the TFO cells (188 genes changed, FDR q-val<0.05) (Figures 13a and 13b). Of these few mRNAs that changed expression during RAS-transformation, an interesting pattern emerged when compared to the changes that occurred in the preceding transition. The majority of RAS-transformation regulated mRNAs were also expressed during immortalization, and within this subset of mRNAs that changed during both transitions, most responded in opposing directions (Figures 13c and 13d). For example, approximately half of all transcripts that increased expression during transformation also showed a decrease at the preceding step of immortalization. Interestingly, this suggests that oncogenic RAS transformation induces unexpected changes in mRNA abundance that cause a partial reversion back to the mRNA expression levels observed in the primary cells.
Figure 13: RNA-seq analysis of transition states.
Most mRNA abundance changes in this model occur early, during immortalization and are maintained throughout transformation. A) Heatmap depicts changes between primary cells and immortalized cells (PRIM to IMO) and between primary to transformed cells (PRIM to TFO). The heatmap was generated by k-means clustering of log2 fold changes (relative to PRIM) for all genes with FPKM >1 in at least one condition. Significant changes (q<0.05) are indicated as black marks in the columns next to the heatmap, with the IMO changes referring to early changes (from PRIM to IMO) while TFO changes refer to late changes (from IMO to TFO).

B) Number of genes significantly increased (red) or decreased (blue) during immortalization and transformation. C) mRNAs regulated by RAS (TFO) are preceded by opposing changes in IMO. mRNAs significantly (q<0.05) increased (All TFO Upregulated) or decreased (All TFO Downregulated) from IMO to TFO were classified based on the preceding change that occurred during immortalization. Red indicates mRNAs that significantly increase from PRIM to IMO, blue indicates mRNAs that significantly decrease from PRIM to IMO and gray do not change significantly from PRIM to IMO. Expected changes assume independence and are based on the proportion of changes of all mRNAs from IMO to TFO. D) Average log2 fold-change in relative abundance for categories of mRNAs regulated during TFO. Colors are the same as in C).

Thickness of lines is relative to the proportion of all changes.

To see how these changes in mRNA abundance relate to changes in cellular function we performed Gene Ontology analysis and found enrichment for many GO categories, including those expected to change given the transgenes overexpressed in this system (Figure 14a). Specifically, among mRNAs induced during immortalization we saw enrichment for ‘telomere maintenance’, ‘cell proliferation’, ‘mitotic cell cycle process’, ‘chromosome organization’, and ‘metabolic process’. These processes are known to be controlled by transgenes introduced into the IMO cells (Figure 3).

Furthermore, among mRNAs that decreased during the early IMO transition, we saw enrichment for ‘programmed cell death’ and ‘negative regulation of cell proliferation,’ which are characteristically shut down by dominant negative p53 (Willis et al., 2004).
Therefore, many of the mRNAs that were regulated up or down during immortalization were expected based on the known pathways that were perturbed in these cells.

Importantly, we observed enrichment of upregulated mRNAs involved in ‘epithelial cell migration’ and enrichment of downregulated mRNAs involved in ‘cell adhesion,’ in line with our observation that IMO cells are migratory.

For mRNAs induced during the transition to transformation, we found enrichment for categories including ‘regulation of cell death’ and ‘signal transduction by p53 class mediator’ (Figure 14a). Among mRNAs that decreased during transformation we observed enrichment for ‘extracellular matrix organization,’ ‘cell adhesion,’ and ‘regulation of epithelial to mesenchymal transition’ (Figure 14a). We unexpectedly found that these enriched categories were primarily driven by the subset of transformation regulated mRNAs that had changed in the opposite direction during immortalization (Figure 14b). Interestingly, all downregulated mRNAs that were enriched for EMT-related GO categories in the transformed cells were upregulated during the preceding step (PRIM to IMO). Moreover, many of the genes upregulated during immortalization and downregulated during transformation would be expected to positively correlate with invasion. For example, TWIST1, TGFβ1 and MMP7 mRNAs all have demonstrated roles in invasion (Nieto et al., 2016; Shao et al., 2017) and are included in this subset.
Figure 14: Gene Ontology Analysis

A) Enriched GO categories for genes increasing (top row) or decreasing (bottom row) during immortalization (left panels) and transformation (right panels). B) Enriched GO categories for RAS regulated mRNAs (shown in right panels of A), further categorized by preceding changes (PRIM to IMO).
4.3.2 Immortalized and transformed cells have a mesenchymal gene expression signature

Given that mRNAs encoding proteins involved in the regulation of EMT were upregulated during the early transition and slightly downregulated during the later transition, we decided to classify PRIM, IMO and TFO cells based on epithelial and mesenchymal gene expression signatures. There is considerable evidence indicating that EMT is not binary, but instead functions as a gradient with multiple intermediary stages of partial, or incomplete, EMT (Nieto et al., 2016). To quantify the interplay between EMT progression and cancer progression, Thiery and colleagues developed a method for EMT scoring based on transcriptomics (Tan et al., 2014). EMT signature scores are calculated using a two-sample Kolmogorov-Smirnov test that is based on multiple gene expression signatures (Tan et al., 2014). The EMT signature score is reported as a value between -1 and +1, with more positive scores indicating a more mesenchymal state. We used this method to score PRIM, IMO and TFO HMECs. PRIM HMECs were assigned a score of 0.22905 (p=0.05737), IMO cells were scored as 0.71391 (p=2.109e-15), and TFO cells were given a score of 0.53635 (p=8.02e-08) (Figure 15). Thus, immortalization results in a highly mesenchymal gene expression signature that is not further increased, but rather slightly decreased, during RAS transformation.

It is important to note that expression of transcripts encoding mesenchymal markers N-cadherin and vimentin are unchanged during RAS transformation (Figure 16), even though protein levels are significantly increased. Our data therefore suggest a
model whereby gene expression changes are acquired early in the tumorigenesis process, before the full EMT phenotype is evident. TFO cells have relatively few significant gene expression changes compared to IMO cells. However, the RAS-induced mRNA changes are slightly less epithelial than the immortalized cells, despite the significant increase in malignancy, invasion, and expression of EMT markers.

![Graph showing epithelial and mesenchymal gene expression signatures](image)

**Figure 15: Epithelial and mesenchymal gene expression signatures.**

Immortalization results in a decrease in an epithelial and increase in a mesenchymal gene expression signatures. EMT signature scores for primary, immortalized and transformed HMECs, calculated using a two-sample Kolmogorov-Smirnov test that is based on previously reported gene expression signatures (Tan et al. 2014). The EMT signature score is reported as a value between -1 and +1, with more positive scores indicating a more mesenchymal state. The X-axis shows ranked normalized expression, with “1” being the most abundant and “0” being the least abundant.
4.4 Alternative Splicing Analysis

4.4.1 Large changes in alternative splicing occur during immortalization, but not during oncogenic RAS transformation

Transcripts encoding proteins involved in RNA processing were upregulated during the transition from PRIM to IMO (Figure 14a, top left). Given that alternative splicing is an important RNA processing event, we measured changes in splicing genes. As a group, core splicing machinery components (U1Sm, U1, U2, tri-snRNP, and EJC complexes) were significantly upregulated during immortalization, while no further expression changes were observed for this group during RAS transformation (Figure 17a). This is in agreement with a recent report that splicing factor genes are generally overexpressed in breast cancer cells compared to normal tissues (Sveen et al., 2016). However, our findings suggest that this increase occurs during immortalization rather than transformation. We next analyzed global changes in alternative splicing during immortalization and transformation using VAST-TOOLS (Irimia et al., 2014). Indeed,
immortalization induced a number of significant alternative splicing changes, while additional splicing differences during RAS transformation were undetectable (Figure 17b). We conclude that early increases in expression of splicing factors leads to a drastic change in the alternative splicing program during immortalization and this alternative splicing program is maintained during the transition to transformation. This suggests that alternative splicing associated with cancer progression may be an early event that precedes malignancy.

Figure 17: Alternative splicing analysis

Large changes in alternative splicing occur during immortalization, but not during transformation. A) Expression of mRNA encoding core splicing components increase significantly during immortalization. Heatmap depicts changes from PRIM to IMO and from PRIM to TFO. B) Significant changes in alternative splicing are observed during immortalization. dPSI= change in percent spliced in. Blue is significant reproducible alternative splicing events from PRIM to IMO and red is significant reproducible alternative splicing events from IMO to TFO.
4.4.2 Immortalized cells have acquired an EMT splicing signature that is maintained during RAS transformation

The majority of changes in alternative splicing were spliced exons (Figure 18), which is similar to what others have observed during EMT transitions (Yang et al., 2016). As described in Chapter 3, we also observed decreases in the epithelial splicing RBPs, ESRP1 and ESRP2, further suggesting that an EMT-related alternative splicing program may be regulated during the IMO transition from PRIM cells. It was previously reported that EMT drives an alternative splicing program associated with multiple tumor types (Shapiro et al., 2011; Yang et al., 2016). We assessed the extent of EMT-specific alternative splicing during immortalization and transformation by comparing our datasets to previously reported EMT splicing signatures (Figures 19a and 19b). We observed close agreement between these datasets, and many well-described
mesenchymal isoforms, including those for CD44, ENAH (also known as MENA) and p120-catenin, were identified in our system (Figure 20). Thus, these findings indicate that the IMO-induced alternative splicing program is, at least in part, inducing an EMT alternative splicing program.

ESRP1 and ESRP2 are key RBP regulators of alternative splicing during EMT (Warzecha et al., 2010; Yang et al., 2016), and we observed large decreases in mRNA encoding these proteins during immortalization. ESRP1 decreased nearly 400-fold and ESRP2 decreased ~2.7 fold during immortalization. To determine the extent to which the ESRPs are regulating alternative splicing during immortalization, we compared our splicing datasets with a previously reported splicing signature induced by ESRP knockdown (Yang et al., 2016). Indeed, we observed many of the same alternative splicing events during immortalization (Figure 19c). As a control, we compared our datasets to a signature induced by RBM47 knockdown (Yang et al., 2016). RBM47 is another major regulator of EMT alternative splicing; however, mRNA encoding this protein was not detected in this system. We observed poor correlation between our datasets and the RBM47 dataset (Figure 19c).
Figure 19: EMT alternative splicing signatures

Immortalization induces an EMT-related alternative splicing signature that is maintained during transformation. A) Comparison of alternative splicing changes induced during immortalization to previously reported EMT-signatures (Shapiro et al. 2011; Yang et al. 2016). B) Comparison of alternative splicing events reported by Shapiro et al. to Immortalization (top) and transformation (bottom). C) Comparison to alternative splicing events reported by Yang et al. to an ESRP KD signature (top) and an RBM47 KD signature (bottom). dPSI= change in percent spliced in.

Our data clearly indicate vast changes in alternative splicing of EMT-related mRNAs during immortalization. In addition to these isoforms that are known to be regulated during EMT, we identified a number of other splicing events that may be important in cell migration. However, predicting how alternative splicing impacts the function of a protein is notoriously difficult. To address this, we searched the literature to see if the identified splicing events have been previously described to impact protein
function. We found that many of these events identified during immortalization would be expected to significantly alter properties characteristic of EMT (Figure 21). For example, the alternatively spliced EDA isoform of fibronectin (FN1) is prominent in contexts of cell migration, and the EDA isoform has demonstrated roles in cell adhesion (Manabe et al., 1997). MYO1B, which encodes unconventional myosin, is involved in regulating cell motility. Alternative splicing of this gene yields proteins with lever arms of different lengths that regulate myosin step size and, thus, are likely to increase motility rates (Laakso et al., 2010). Furthermore, we observed alternative splicing of TLN1, which connects integrins to the actin cytoskeleton at focal adhesions, in a region known to regulate its recruitment to the membrane (Goult et al., 2013). TLN1 has a well-described role in regulating focal adhesion dynamics, cell migration and invasion; alternative splicing of this gene has not been explored as a mechanism of its regulation in this context. We also observed alternative splicing of DLG1 in a region that allows interactions with kinases, including Crk, which regulates cell adhesion, spreading, and migration (McLaughlin et al., 2002). Overall, our data suggest that these alternative splicing events may significantly contribute to a cell’s migratory phenotype, and thus, warrant future investigation.
Figure 20: Well characterized EMT alternative splicing events are observed during immortalization.
Figure 21: Novel EMT alternative splicing events

The plots in the left panels show the two joint posterior distributions over psi. Each replicate is plotted as dots below the histogram. The right panels show the probability (y-axis) of delta psi being greater than a given value (x-axis). The red line is the 95% confidence interval.

4.5 microRNA expression

MiRNAs are important post-transcriptional regulators that modulate biological processes, including cancer formation and metastasis, through either mRNA degradation or translational repression (Di Leva and Croce, 2013). To determine the extent that miRNAs could be modulating the gene expression program during HMEC immortalization and transformation, we used a qRT-PCR array to quantify 340 of the most common and conserved miRNAs. Similar to our results described above for
mRNA expression dynamics, we observed the highest number of miRNA expression changes during immortalization, with 20 miRNAs significantly increasing or decreasing expression during immortalization and only 6 changing during transformation (Figure 22a and Tables 4 and 5).

Additionally, the miRNAs that changed expression during immortalization changed to a much larger degree than those that changed during transformation. The median log 2 fold changes for miRNAs that significantly increased or decreased were 2.3 and -3.5, respectively, during immortalization, and 1.3 and -1.1 during transformation (Figure 22a and Tables 4 and 5).

During both immortalization and transformation, we observed changes in miRNAs with previously described roles in EMT. For example, both miR-200a and miR-10b play well-documented roles in migration and invasion (Bouyssou et al., 2014; Di Leva and Croce, 2013). MiR-200c is also a well-documented suppressor of metastasis in many cancer types, and it directly targets Zeb1 and Zeb2, transcriptional repressors of E-cadherin (Shibue and Weinberg, 2017). Moreover, miR-205 has been shown to suppress metastasis and EMT in prostate cancer (Chao et al., 2014; Tucci et al., 2012), breast cancer (Chao et al., 2014), and gastric cancer (Xu et al., 2016). MiR-205 displayed the largest changes out of all miRNAs, with a log 2 fold change of -14.85. Importantly, we observed that mRNA targets of miR-205 increased in expression during immortalization, indicating that this miRNA acts to destabilize mRNA targets in the primary cells (Figure
Therefore, our data suggest that miR-205 down regulation is an important regulatory mechanism governing the transition to a mesenchymal gene expression signature and migratory phenotype during immortalization.
Figure 22: microRNA dynamics in cancer transition states.

A) Expression of canonical EMT miRNAs is altered during immortalization. Dashed lines represent a 2-fold change. Red dots mean significantly increased. Green dots are significantly increased during immortalization (left panel) or transformation (right panel). Data is reported as Ct values normalized to controls. B) mRNA targets of miR-205 increase in abundance during immortalization. CDF plots comparing change in mRNA abundance for miR-205 targets and non-targets. TargetScan 7.1 conserved sites were used to identify mRNA targets.
Table 4: All miRNAs that significantly changed expression during the primary to immortalized transition

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Log 2 Fold Change PRIM-IMO</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-155</td>
<td>4.7752</td>
<td>0.0209205</td>
</tr>
<tr>
<td>miR-338-3p</td>
<td>3.2063</td>
<td>0.0258609</td>
</tr>
<tr>
<td>miR-523</td>
<td>2.6137</td>
<td>0.0182128</td>
</tr>
<tr>
<td>miR-451</td>
<td>2.48888</td>
<td>0.0221886</td>
</tr>
<tr>
<td>miR-524-5p</td>
<td>2.3147</td>
<td>0.0193478</td>
</tr>
<tr>
<td>miR-198</td>
<td>2.3027</td>
<td>0.0276387</td>
</tr>
<tr>
<td>miR-339-3p</td>
<td>2.0427</td>
<td>0.0304607</td>
</tr>
<tr>
<td>miR-508-3p</td>
<td>1.2517</td>
<td>0.0234453</td>
</tr>
<tr>
<td>miR-524-3p</td>
<td>1.2492</td>
<td>0.0397081</td>
</tr>
<tr>
<td>miR-380-5p</td>
<td>1.23938</td>
<td>0.0066989</td>
</tr>
<tr>
<td>miR-26a-1*</td>
<td>1.10587</td>
<td>0.0428535</td>
</tr>
<tr>
<td>miR-205</td>
<td>-14.8531</td>
<td>0.0043466</td>
</tr>
<tr>
<td>miR-200c</td>
<td>-8.17162</td>
<td>0.0442563</td>
</tr>
<tr>
<td>hsa-miR-718</td>
<td>-4.47912</td>
<td>0.0220238</td>
</tr>
<tr>
<td>miR-375</td>
<td>-4.0617</td>
<td>0.0135942</td>
</tr>
<tr>
<td>miR-183</td>
<td>-3.53463</td>
<td>0.0343314</td>
</tr>
<tr>
<td>miR-187</td>
<td>-3.30912</td>
<td>0.0450006</td>
</tr>
<tr>
<td>hsa-miR-298</td>
<td>-2.08512</td>
<td>0.032853</td>
</tr>
<tr>
<td>miR-429</td>
<td>-1.9303</td>
<td>0.0347827</td>
</tr>
<tr>
<td>miR-584</td>
<td>-1.1428</td>
<td>0.0128254</td>
</tr>
</tbody>
</table>
Table 5: All miRNAs that significantly changed expression during the immortalized to transformed transition

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Log 2 Fold Change IMO-TFO</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-380-3p</td>
<td>1.0378</td>
<td>0.03810413</td>
</tr>
<tr>
<td>miR-146a</td>
<td>1.29137</td>
<td>0.01173079</td>
</tr>
<tr>
<td>miR-200a</td>
<td>1.3703</td>
<td>0.01841264</td>
</tr>
<tr>
<td>miR-10b</td>
<td>1.73888</td>
<td>0.00550824</td>
</tr>
<tr>
<td>miR-30c</td>
<td>-1.13212</td>
<td>0.02468893</td>
</tr>
<tr>
<td>miR-425*</td>
<td>-1.06562</td>
<td>0.0338408</td>
</tr>
</tbody>
</table>

4.6 Discussion

We utilized an isogenic, primary cell-derived model of breast cancer progression to better understand the genetic alterations and regulatory mechanisms that promote tumorigenesis and EMT. Using a genetically defined model system enabled us to narrow down the pathways in which perturbation may activate an EMT expression program. By activating telomere maintenance, mutating the p53 and pRb tumor suppressor pathways, and expressing mutated c-Myc in IMO cells, we have found that an EMT transcriptional program is activated. However, pre-malignant cells were non-invasive and did not have robust protein expression of mesenchymal markers, suggesting that they represent a partial EMT state.

Oncogenic RAS transformation was found to increase both invasion and protein expression of N-cadherin and Vimentin protein without causing large changes in mRNA and miRNA expression levels or alternative splicing events. Our work suggests a
mechanism of progressive EMT whereby pre-malignant cells are primed for a full EMT at the mRNA level before they become malignant. Interestingly, when H-RAS\textsuperscript{G12V} was expressed in primary HMECs expressing only hTERT, the cells were not transformed (Elenbaas et al., 2001). It is possible that hTERT in the absence of the other four transgenes does not sufficiently prime pre-malignant cells at the transcriptional level. The sum of this data indicates that transcriptionally primed pre-malignant cells can be induced by oncogenic RAS transformation to become both malignant and invasive simultaneously through post-transcriptional mechanisms.

4.6.1 Prognostic implications and drug targets

Given that only four defined pathways are altered during immortalization in this system, IMO cells can serve as a pivotal landmark for future studies to determine which pathways, or combination of pathways, results in activation of the EMT program. Understanding which specific genetic alterations contribute to early activation of the EMT program may reveal therapeutic targets that prevent the activation of metastatic drivers in pre-malignant lesions.

Additionally, by analyzing alternative splicing during primary cell immortalization, we identified changes in alternative splicing that may be necessary for mRNA priming prior to the transition to malignant transformation. Indeed, we identified both previously described and novel EMT-related alternatively spliced isoforms (Figures 20 and 21). For example, expression of Mammalian ENAbled
(MENA/ENAH) isoforms plays a crucial role in regulating cell migration and invasion (Oudin et al., 2016; Roussos et al., 2011a; Roussos et al., 2011b). MENA has two distinct alternative splicing events, an invasion suppressive cassette exon, Mena11a, and a separate pro-invasive cassette exon, MenaINV. The ratio of the expression levels of the MenaINV and Mena11a isoforms has been successfully used to calculate metastatic risk in breast cancer patients (Rohan et al., 2014). In our HMEC system we observed a nearly complete loss of the invasion suppressing Mena11a in IMO and TFO cells, while MenaINV was present but not changing in all cell lines thus indicating the ratio MenaINV to Mena11a increases dramatically at the IMO stage. Our data indicate that specific EMT-associated alternatively spliced isoforms are acquired by pre-malignant cells. Importantly, we identified alternative splicing events that have not previously been described in the context of EMT, many of which are likely to influence EMT-related phenotypes, including MYO1B, TLN1, DLG1 and FN1. Given that transcripts encoding splicing factors are widely upregulated across many cancer types (Sveen et al., 2016), and epithelial splicing networks have been reported in both breast cancer and non-small cell lung cancer cells lines (Shapiro et al., 2011; Yang et al., 2016), it is likely that this epithelial splicing regulatory network is widespread across many tumor types. Therefore, these novel EMT-associated isoforms may also have prognostic value in calculating metastatic potential in both pre-malignant and malignant cells.
4.6.2 Post-transcriptional regulation downstream of oncogenic RAS

Precise control of gene expression is known to involve regulation at the transcriptional, post-transcriptional and post-translational levels (Keene, 2001, 2007; Schwanhausser et al., 2011). While many studies have relied on steady state mRNA levels as a proxy for gene expression, it has become increasingly clear that RNA levels alone may not provide a complete picture of the complex regulatory processes occurring during cancer cell progression. Indeed, large numbers of transcripts are regulated at the post-transcriptional and post-translational levels, as evidenced by the fact that steady state protein levels often do not correlate with steady state mRNA levels (Griffin et al., 2002; Ideker et al., 2001; Keene and Tenenbaum, 2002; Mansfield and Keene, 2009). In this study, we used RNA-seq in an isogenic system to quantify changes in the transcriptome during tumorigenesis. It is important to note that RNA-seq measures a combination of both transcription and mRNA decay, both of which are highly dynamic, and does not allow one to distinguish their relative contributions without additional assays. However, any occurrence that influences mRNA levels, including epigenetic modifications and changes in mRNA stability, would be reflected in the transcriptomic data. In our data, gene expression levels in the IMO and TFO cells were not drastically different (Figure 13), yet their invasive phenotypes are critically different (Figure 11). Therefore, it is apparent that RNA expression quantified by RNA-seq does not directly indicate a functional outcome. Our study suggests that RAS transformation likely
controls invasion through post-transcriptional mechanisms. Understanding the downstream mechanisms by which RAS post-transcriptionally regulates gene expression and functional outcome could provide therapeutic approaches to targeting metastasis. Our work has eliminated alternative splicing, mRNA stability and regulation by miRNAs as potential post-transcriptional mechanisms downstream of oncogenic RAS signaling, as we did not observe significant changes in splicing, mRNA abundance or miRNA expression levels during transformation.

Several studies have demonstrated a central role of translational regulation as an effector of oncogenic signaling. In glioblastoma, RAS and Akt signaling have minimal effects on the transcriptome, while significantly altering mRNA translation (Rajasekhar et al., 2003). Furthermore, TGFβ-induced EMT induces phosphorylation of eIF4E, which in turn regulates translation of Snail and Mmp-3 (Robichaud et al., 2015). In fact, several other studies have demonstrated roles for translational RNA regulons in the acquisition of cancer-related phenotypes downstream of major oncogenic signaling pathways (Jung et al., 2014; Romeo et al., 2013; Shahbazian et al., 2010; Truitt and Ruggero, 2016; Tsukumo et al., 2016; Wurth et al., 2016). Thus, it is possible that the effects of oncogenic RAS signaling observed during HMEC transformation induces acquisition of both anchorage independent growth and an invasive phenotype at the level of translational regulation. For example, a recent study established a role for the CELF1 RNA-binding protein in regulating EMT-associated translational regulation (Chaudhury et al., 2016).
Another CELF family member, CELF6, has been demonstrated to rescue KRAS suppression in oncogene addicted cells (Shao et al., 2014). In light of these studies, our findings suggest the intriguing possibility that RAS signaling may regulate translation through altering CELF family mRNA targets. Therefore, it will be important to determine post-transcriptional response and remodeling of translational RNA regulons induced by RAS in this system.

Many studies have linked RAS signaling pathway to metastasis. Recent reports using mouse models of mammary cancer indicate that a subpopulation of Her2+ early lesion cells activate a migratory phenotype prior to proliferation and tumor growth (Harper et al., 2016; Hosseini et al., 2016). Also, MCF10a immortalized mammary cells overexpressing Her2 showed altered cellular architecture and invasive phenotype (Harper et al., 2016). Her2/neu is an upstream activator of RAS (Yoh et al., 2016), suggesting a potential mechanistic link. Additionally, expression of the RAS oncogene in MCF10a cells increased invasion through a mechanism involving a decrease in expression of ΔNP63 at the transcriptional level (Yoh et al., 2016). Interestingly, ΔNP63 is decreased during the PRIM to IMO transition in our HMEC system, suggesting that it may be regulated transcriptionally as part of a broader mechanism of EMT transcriptional remodeling in breast cancer, independent of oncogenic RAS. In our system, RAS induces invasion via a transcription-independent mechanism, indicating
complex, diverse regulatory mechanisms by which cells can achieve the same functional outcome.

### 4.6.3 EMT transitions

The EMT program is usually only partially activated in human carcinomas, and studies conducted in both in cell culture and patient samples have used expression of epithelial/mesenchymal protein markers or RNA abundance to evaluate degree of EMT (Nieto et al., 2016; Tan et al., 2014). In our cell culture model, immortalized cells lost expression of E-cadherin, yet they did not robustly express N-cadherin or Vimentin (Figure 12). However, if we were to just measure RNA abundance they would be indistinguishable from our invasive TFO cells, which robustly expressed N-cadherin and Vimentin. While transcriptomic profiling of clinical samples may be useful in determining patient prognosis, our data suggest that this may not always be an accurate proxy for invasive potential. We show that post-transcriptional mechanisms may control transitions between different degrees of EMT partiality.

We observed a slight decrease in the mesenchymal signature score in the transition of IMO to TFO (Figure 15) despite seeing a massive increase in protein expression of mesenchymal markers (Figure 12). This result has intriguing parallels to another slightly incongruous result seen by Thiery and colleagues in breast cancer: a higher mesenchymal score correlates with better patient outcome (Tan et al., 2014). This
may suggest cells with high mesenchymal mRNA expression patterns have not undergone full EMT including the post-transcriptional mechanisms we report here.

### 4.6.4 Changes in miRNAs during cancer progression

We also observed large changes in miRNA expression during immortalization, and only modest expression changes in a few miRNAs during RAS transformation.

**Figure 22 and Tables 4 and 5.** There is much abundant recent evidence to suggest a crucial role for miRNAs in regulating EMT, and many miRNAs (miR-200a, miR-200c, miR-429, miR-205, miR-10b, and others) known to be involved in EMT were identified in our system as being significantly altered (Bouyssou et al., 2014; Di Leva and Croce, 2013). Despite the reported roles for these miRNAs in EMT, we only observed corresponding expression changes for mRNA targets of miR-205 during immortalization, and we did not observe any significant changes in the predicted targets for the few miRNAs regulated during RAS transformation. It is possible that other differentially expressed miRNAs only regulate only a few mRNAs in this system, or that they regulate mRNA targets at the level of translation.
5. Binding site identification and quantification provides mechanistic insight into HuR regulation during tumorigenesis

5.1 Background and Significance

Regulation by the HuR RBP is known to be altered in many cancers (Abdelmohsen and Gorospe, 2010; Yuan et al., 2010). Global changes in HuR bound mRNAs have been demonstrated during the transformation of immortalized epithelial cells and in the progression to an advanced malignant state, implicating functional remodeling of RNPs in both of these processes (Calaluce et al., 2010; Mazan-Mamczarz et al., 2008a; Mazan-Mamczarz et al., 2008b; Mazan-Mamczarz et al., 2011). While it is clear that HuR functions as a modulator of the proliferative gene expression program through regulating the stability and translation of mRNAs encoding growth-related proteins, the molecular mechanisms underlying these functional effects remain unclear.

While RIP procedures identify and quantify mRNA targets transcriptome wide, they do not distinguish the contributions of individual binding sites. Indeed, CLIP methods have demonstrated that many mRNA targets have multiple HuR binding sites per message (Mukherjee et al., 2011). The significance of the contribution of each individual binding site to functional outcome is unknown. The main drawback to CLIP methods is that they do not consider RNA abundance and, thus, are not quantitative. The ability to quantify binding is essential for comparing dynamic biological conditions, such as tumorigenesis. Our lab has recently developed a method called DO-RIP-seq,
which is capable of both identifying and quantifying binding sites on a global scale (Nicholson et al., 2017; Nicholson et al., 2016). This method combines aspects of both RIP and CLIP techniques to yield a technique capable of quantifying RBP-RNA interactions globally at both the whole transcript and individual binding site levels. The quantification accurately reflects the relative strength of each binding site.

The data presented in Chapter 4 suggested a mechanism of oncogenic RAS-triggered post-transcriptional regulation. Given HuR’s important role in tumorigenesis, in this chapter we investigate a role for HuR in tumorigenesis in the HMEC system. We applied DO-RIP-seq to IMO and TFO HMECs to identify and quantify binding sites. We then use these binding sites to gain mechanistic insights into HuR’s regulation in RAS-transformed cells.

5.2 **DO-RIP-seq to identify and quantify RBP binding sites transcriptome-wide**

5.2.1 Identification and quantification of HuR binding sites

We performed DO-RIP-seq in IMO and TFO HMECs as described in the Materials and Methods section and in a recent publication from our lab (Nicholson et al., 2016). Lysates were digested with micrococcal nuclease under optimized conditions before performing HuR RNP IP and sequencing 20-70 nucleotide RBP protected mRNA fragments as described. Sequencing reads were normalized to total IP “input” RNA libraries to account for RNA abundance and allow for quantification. HuR binding sites were identified by quantifying enrichment values of HuR-bound reads over read values
from the input libraries at 5 nucleotide intervals across the transcriptome. Log of odds scores (LOD) were generated using Gaussian mixture modeling (GMM) to calculate the probability of association for each 5 nucleotide interval. Three biological replicates of each condition were used to calculate p-values (single tailed paired t-test), and 5 nucleotide intervals with LOD >0 and p<0.05 were identified as binding sites.

Our goal was to use DO-RIP-seq to understand HuR dynamics during the transition from normal cell to a fully transformed cell. Since the method requires a large amount of input material, we were unable to obtain satisfactory DO-RIP data from PRIM HMECs. We therefore compared dynamics during only the transition from premalignant (IMO) to RAS-transformed (TFO) HMECs. We identified 91,579 HuR binding sites in both IMO and TFO cells. We first characterized the binding site features identified by DO-RIP-seq. In both IMO and TFO cells, HuR preferentially bound to U-rich sequences in 3’UTRs and introns, consistent with what is known about HuR binding (Figure 23a and 23b). The HuR consensus motif and genomic locations of the binding sites did not change during oncogenic RAS transformation. Surprisingly, 99.8% of our identified HuR binding sites were found in both IMO and TFO cells, indicating that binding sites generally did not completely gain or lose HuR association during oncogenic RAS transformation (Figure 23c). While 15 binding sites were considered unique in IMO and 159 binding sites were considered unique in TFO, these binding sites were generally ones that had LOD scores just below 0 in one condition and just above 0
in another condition, suggesting that they may be a result of our stringent LOD >0 cut off and may not actually be significantly changed.

Although we did not observe qualitative changes in HuR binding, LOD scores in IMO and TFO cells were not tightly correlated (R=0.7314, Figure 23d), suggesting that binding sites change quantitatively during transformation. Indeed, many examples of these quantitative changes could be visualized for many individual messages (Figure 24). Our lab has previously shown that this quantification reflects the relative strengths of binding (Nicholson et al., 2017). Therefore, our data suggest that HuR changes its relative binding strengths at specific binding sites during transformation.

Figure 23: Analysis of HuR binding sites

A) Bitmap of the most frequently observed 8-mer in HuR binding sites. B) Breakdown of the location of identified HuR binding sites. C) Number of binding sites with LOD>0 in each condition (blue circle=IMO; red circle=TFO) D) Scatterplot of Median LOD IMO and Median LOD TFO, R=0.7314
We then sought to use these quantitative changes to try to understand the mechanism by which HuR determines the strength of the binding site. We searched within binding site regions for enriched secondary motifs, and we identified a GU-rich motif (Figure 25a). Sites that contained this motif were more likely to decrease association with HuR during RAS transformation, whereas sites that had U-rich secondary motifs were more likely to gain association with HuR (Figure 25a). In fact, ~80% of all sites that increased association with HuR during transformation were U-rich.

On the other hand, over half of all messages containing a GU-rich secondary motif decreased association with HuR during RAS transformation. The presence of this GU secondary motif suggested a mechanism whereby another GU sequence binding RBP
may be acting in competition with HuR. By this model, RAS transformation would result in association with a GU-motif binding RBP with a message, which would compete off HuR (Figure 25b).

Figure 25: Change in HuR binding during RAS transformation is associated with a GU-rich secondary motif

A) Bitmap of the most enriched secondary motif contained within HuR binding sites. HuR binding sites containing enriched secondary motifs were compared for change in binding strength from HuR DO-RIP IMO vs TFO. B) RBP competition model suggested by the data shown in A.
5.2.2 Identification and quantification of CELF1 binding sites

Several RBPs bind to GU-motifs. However, after searching the literature, only one family of RNA binding proteins, the CELF family, was both expressed in our HMEC system and known to bind to this motif. CELF1 has demonstrated functional links to cancer (Rattenbacher et al., 2010; Starr et al., 2009; Wu et al., 2013), and it has been shown to compete with HuR for binding to E-cadherin, Occludin and Myc (Liu et al., 2015; Yu et al., 2016; Yu et al., 2013b). Therefore we decided to investigate the hypothesis that CELF1 could be competing HuR off of other messages on a global scale in response to RAS transformation.

We performed DO-RIP-seq to identify and quantify global binding sites for the CELF1 protein in IMO and TFO cells. In both IMO and TFO cells, CELF1 preferentially bound to GU-rich sequences in introns and 3’UTRs, consistent with what is known about CELF1 binding (Figures 26a and 26b). The CELF1 consensus motif and genomic locations of the binding sites did not change during oncogenic RAS transformation. Similar to what we observed with HuR binding, the majority of CELF1 binding sites did not qualitatively change during RAS transformation (Figure 26c), and we found evidence that many CELF1 binding sites change quantitatively (Figure 26d).
5.2.3 Comparison of HuR and CELF1 binding sites

To investigate our hypothesis that HuR and CELF1 regulate mRNA targets in a combinatorial manner, we first looked to see if binding sites overlapped. In both IMO and TFO cells, ~55% of HuR binding sites overlapped with CELF1 binding sites, while ~31% of CELF1 sites overlapped HuR sites in both conditions (Figures 27a and 27b).
Based on our model (Figure 25b), we hypothesized that overlapping sites would decrease association with HuR and gain association with CELF1 during RAS transformation. We therefore asked how overlapping binding sites changed in association with either HuR or CELF1 during transformation. Surprisingly, sites that had increased CELF1 LOD scores during transformation had slight increases in HuR LOD scores during transformation, although they did not drastically change when
compared to all overlapping sites (Figure 28a). In contrast, binding sites that had decreased CELF1 LOD scores during transformation also had decreased HuR LOD scores, indicating that overlapping binding sites that lose association with CELF1 are also more likely to lose association with HuR (Figure 28a). We then limited our analysis to the 3’UTR. CELF1 sites that overlapped HuR in the 3’UTR did not change in HuR LOD during transformation (Figure 28b). This suggests that our hypothesized model where CELF1 binding competes HuR off of binding sites in the 3’UTR is incorrect. In fact, HuR and CELF1 may actually regulate binding sites in a synergistic manner.

Figure 28: CELF1 overlapping binding sites and changes in HuR association

A) CDF plot showing the change of LOD in HuR during RAS transformation for all CELF1 overlapping binding sites (black), all overlapping sites that decrease in CELF1 association during transformation (red) and all overlapping sites that increase in CELF1 association during transformation (green). B) CDF plot showing the change of LOD in HuR during RAS transformation for all CELF1/HuR overlapping binding sites in the 3’UTR (yellow) and all CELF1 3’UTR sites that do not overlap HuR binding sites (black).
5.2.4 Identification and quantification of HuR and CELF1 mRNA targets at the whole transcript level

With DO-RIP-seq data, not only can we calculate relative strengths of RBP binding at individual binding sites, but we can also calculate probability of binding to entire transcripts. Since many RBP targets have multiple binding sites with different LOD scores at each binding site, quantifying at the whole transcript level allows us to more accurately look at the functional enrichments of messages that increase or decrease association during transformation. To quantify binding to total transcripts, reads per million (RPM) were calculated for the exonic reads of each gene. RIP-seq-like scores (RSLs) were calculated as the logarithmic difference between each gene’s RPM in the HuR and input libraries (Nicholson et al., 2016).

In agreement with our binding site-level analysis, HuR DO-RIPs in IMO and TFO cells correlated strongly with each other, indicating that whole messages did not change association with HuR during RAS transformation (Figures 29 and 30). We observed similar results for CELF1, suggesting that CELF1 does not significantly change association with entire transcripts during transformation (Figures 29 and 30). HuR mRNA targets and CELF1 targets correlate better with each other than they do with input samples, indicating some overlap in associated mRNAs (Figure 29). A global test of p-values further demonstrated that there are no significant differences between whole mRNA targets of HuR during RAS transformation, but that there are differences between HuR and CELF1 targets in IMO cells (Figure 31).
There is strong correlation between HuR and CELF1 targets during RAS transformation (top). Targets of HuR and CELF1 in IMO or TFO cells are similar, but less tightly correlated (bottom).
Comparison of HuR IMO to HuR TFO: p-values are randomly and uniformly distributed, suggesting that the null hypothesis is true (mRNA targets in IMO and TFO cells are not significantly different). B) Comparison of HuR IMO to CELF IMO: p-values are moderately anti-conservative, suggesting that there are significant similarities between HuR and CELF1 mRNA targets, but there are also some differences, and C) Comparison of HuR IMO and TFO to HuR inputs: this bimodal pattern indicates that DO-RIP samples separate out targets and non-targets.

1586 transcripts had HuR LOD scores >0 in IMO HMECs and 1536 transcripts had LOD scores >0 in TFO HMECs. 1427 transcripts had LOD >0 in both IMO and TFO (Figure 32a). Although 109 mRNAs had LOD >0 in only TFO and 159 mRNAs had LOD>0 in only IMO, when we looked at the histograms of the LOD scores, we saw that the two populations of targets and non-targets separate out, but that the left distribution of the target peak runs below 0 (Figure 32b). Therefore, the cut-off of LOD>0 is probably very conservative. The mRNAs that were considered different between IMO and TFO had LOD scores just below 0 in one condition and just above 0 in another condition, so it is likely that these differences are not significant, and that, although they are considered
different between the conditions based on our harsh cut-off, they are actually targeted by HuR in both IMO and TFO HMECs.

Figure 32: Analysis of HuR and CELF1 overlapping mRNA targets

A) Venn diagram showing number and overlap of mRNA targets of HuR and CELF1 in IMO cells and HuR targets in TFO cells. B) Histogram of all LOD scores for HuR. Red is the distribution of CELF1 mRNA targets C) Histogram of all LOD scores for CELF1. Red is the distribution of CELF1 mRNA targets.

1753 transcripts had CELF1 LOD scores $>$0 in IMO HMECs (Figure 32a). We did not see large changes in mRNA targets in TFO cells, but given that we do not have
replicates for CELF1 TFO DO-RIP-seq, we do not consider the TFO targets further. Of
the CELF1-associated mRNAs, 62% were also HuR targets (Figure 32a). Accordingly,
when we looked where targets of either RBP fall in a histogram of RSLs of the other
RBP, we saw significant enrichment of RBP targets within the higher LOD scores,
further supporting that mRNA targets of one RBP are highly likely to be targets of both
RBPs (Figure 32b). We hypothesized that this observation could be driven by long
3'UTRs. Indeed, mRNA length was correlated with targeting by both HuR and CELF1,
suggesting that longer mRNAs are more likely to contain regulatory elements for
multiple RBPs (Figure 33).

Figure 33: CDF plot of mRNA length versus targeting

We used RSLs to determine if HuR or CELF1 bind to mRNAs that function on
specific pathways or processes. Among both HuR and CELF1 mRNA targets, we
observed significant enrichment for DNA transcription and RNA metabolism related
pathways (Figure 34). GO analysis with overlapping targets yielded similar enrichments (Figure 34). Non-targets were enriched for mRNAs involved in processes such as ribosomal RNA processing, structural constituent of the ribosome, translation initiation, non-coding RNA metabolism, and protein targeting to the ER, indicating that HuR and CELF1 likely do not regulate these processes (Figure 34).

Finally, we then ranked all expressed genes according to either HuR or CELF1 LOD scores, and we performed Gene Set Enrichment Analysis (GSEA) on these ranked lists (Mootha et al., 2003; Subramanian et al., 2005). For the HuR list, GSEA analysis revealed positive enrichment scores for similar GO categories observed for the entire unranked list of HuR targets, specifically DNA and Nucleic Acid Binding, but also for categories not enriched for the entire set, specifically cell adhesion-related categories (Figure 35a). We also saw a positive correlation for gene sets related to cancer, including sets of genes that were upregulated in response to VEGFA, B-catenin, TGFB, and MEK signaling. This data suggests that HuR targets are enriched for cancer-relevant mRNAs when compared to non-targets. For the CELF1 list, we observed positive correlation with cell-adhesion-related gene sets, as well as KRAS and TGFB signaling genes (Figure 35b). We also observed a strong positive enrichment for several gene sets related to transmembrane transporter activity, suggesting that CELF1 may be involved in this process. For both RBPs, negatively enriched gene sets were largely related to ribosomes.
and non-coding RNAs, further validating that these are largely not regulated by these RBP's.

Figure 34: Gene ontology analysis for all HuR mRNA targets, all CELF1 mRNA targets, mRNA targets of both HuR and CELF1 (overlapping), and non-targets (neither).
Figure 35: Gene Set Enrichment Analysis for mRNA targets of A) HuR and B) CELF1
5.3 Analysis of HuR Knockout TFO cells

In order to determine if HuR binding sites identified by DO-RIP-seq are functional, we depleted HuR from TFO HMECs. Here we describe the generation of HuR CRISPR Knockout (KO) cell lines. We assessed the contribution of HuR to the cancer phenotype and sequenced cells to determine the effect of HuR depletion on mRNA expression levels.

5.3.1 Generation of HuR Knockout transformed HMECs using the CRISPR-Cas9 genome editing system

We transfected TFO HMECs with plasmids containing Cas9 fused to GFP and one of two different guide RNAs designed to target exon 2 of HuR (named E1 and E2) (Figure 36a). Twenty-four hours after transfection, cells were single cell sorted into 96-well plates by fluorescence activated cell sorting (FACs) (Figure 36b). For the E1 gRNA, 13/96 IMO clones survived clonal expansion, while 22/96 TFO clones survived expansion. For the E2 gRNA, no IMO cells survived expansion and 23/96 TFO clones survived clonal expansion. Expanded single cell colonies were screened for inactivation of HuR protein with an immunoblot (Figure 37), and mutations were confirmed with Sanger sequencing (Figure 38). Cells that had been identified as being GFP+ with FACs, but were confirmed to be wild type for HuR, were used as controls in subsequent experiments. All clones used in subsequent experiments are listed in Table 6. We obtained 5 HuR KO clones with the E1 gRNA and 10 HuR KO clones with the E2 gRNA. Of note, we attempted to knockout HuR in IMO HMECs using this same procedure, yet
we were unable to obtain viable HuR knockout lines. Therefore, the rest of the described experiments focus on the TFO cells.

**Figure 36: Generation of HuR KO cell lines**

A) Schematic of CRISPR/Cas9 single plasmid. The U6 promoter drives the expression of the guide RNA and the guide RNA scaffold, which includes a 42nt Cas9 binding RNA structure and a 40nt transcription terminator. The CMV promoter drives the expression of a Cas9 that is fused to GFP via a 2A peptide, which mediates cotranslational cleavage. B) YFP/GFP double positive single cells were sorted with FACs. Over 80% of cells were YFP+, indicating expression of the oncogenic RAS transgene, while less than 5% of cells were GFP+, indicating a transfection efficiency of 4.5-5%. ~4% of cells were both YFP+ and GFP+, and these cells were single cell sorted, and colonies were expanded for further analysis.
Figure 37: CRISPR/Cas9 system abrogates HuR protein expression.

Western blot for HuR in TFO clonal cell lines transfected with CRISPR/Cas9 vectors containing one of 2 guide RNAs designed to exon 2 of HuR (E1 or E2 gRNA). HuR positive clones were used as controls.
Figure 38: Representative visualizations of indels with the TIDE webtool.

Tracking of indels by decomposition (TIDE) (Brinkman et al., 2014) to visualize Sanger sequences. The top graphs depict the spectrum of indels and their frequencies: the x-axis shows deletions and insertions and the y-axis shows the percentage of sequences with a given indel. The bottom graphs show the visualization of aberrant signal (y-axis) in control (WT) sequence (black) and test (KO) sequence (green). The vertical blue line indicates where the CRISPR mutation is expected to occur based on the gRNA sequence. A) Wild-type clone, B) E1-10 KO clone has a homozygous +1 insertion, C) E2-20 KO clone has a homozygous +1 insertion, D) E2-17 clone has a heterozygous -2 deletion and a heterozygous +1 insertion.
Table 6: List of validated HuR Knockout clones and controls used in experiments.

<table>
<thead>
<tr>
<th>Description</th>
<th>Name</th>
<th>HuR Mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type controls</td>
<td>E1-1</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>E1-2</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>E1-17</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>E2-6</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>E2-9</td>
<td>None</td>
</tr>
<tr>
<td>E1 gRNA Knockouts</td>
<td>E1-4</td>
<td>Homozygous 60nt deletion starting at nt 16 of E1 gRNA site</td>
</tr>
<tr>
<td></td>
<td>E1-10</td>
<td>Homozygous deletion of nt 17 of E1 gRNA site</td>
</tr>
<tr>
<td></td>
<td>E1-12</td>
<td>Homozygous deletion of nt 17 of E1 gRNA site</td>
</tr>
<tr>
<td></td>
<td>E1-14</td>
<td>Homozygous deletion of nt 17 of E1 gRNA site</td>
</tr>
<tr>
<td>E2 gRNA Knockouts</td>
<td>E2-10</td>
<td>Homozygous insertion between nts 17 and 18 of E2 gRNA site</td>
</tr>
<tr>
<td></td>
<td>E2-11</td>
<td>Homozygous insertion between nts 17 and 18 of E2 gRNA site</td>
</tr>
<tr>
<td></td>
<td>E2-14</td>
<td>Homozygous deletion of nt 17 of E2 gRNA site</td>
</tr>
<tr>
<td></td>
<td>E2-17</td>
<td>Heterozygous -2 deletion; heterozygous +1 insertion in E2 gRNA site</td>
</tr>
<tr>
<td></td>
<td>E2-20</td>
<td>Homozygous insertion between nts 16 and 17 of E2 gRNA site</td>
</tr>
</tbody>
</table>
5.3.2 The effect of HuR knockout on transformed HMEC cell growth

To determine the effect of HuR on cancer phenotypes, we first analyzed proliferation rates of HuR WT and KO cell lines. We measured proliferation rates of TFO, TFO WT clones, TFO E1 KO clones, and TFO E2 KO clones using the xCELLigence RTCA system to measure cell impedance, which is directly proportional to cell number, in real time over the course of 5 days. We then plotted cell number against time, and used the slope of the exponential growth phase to calculate doubling time for each cell line (Ali-Osman, 1996; Lidsky et al., 2014). As expected, the doubling time for TFO cells (~33.3 hours) and TFO WT clones (~35.1 hours) did not significantly differ, indicating that proliferation rates of each single cell within a heterogeneous WT HuR population are not significantly different. Contrastingly, HuR deficient clones exhibited a 2-3 fold increase in doubling times, indicating a reduced proliferation rate (Figure 39).
Figure 39: Doubling time of HuR KO cells.

Proliferation rates of TFO, HuR WT clones, and E1/E2 HuR KO clones were measured using an xCELLigence Real Time Cell Analyzer System. Cell lines tested: TFO totals in triplicate; WT: E1-1, E2-6; E1 KO: E1-4, E1-10, E1-14; E2 KO: E2-10, E2-11, E2-17, E2-20. Shown are the average of 5 technical replicates of each cell line, reported as the mean and standard deviation of biological replicates. P-values were calculated using a t test.

To see if changes in the cell cycle could explain the reduced proliferation rate, we fixed cells in ethanol, stained with DAPI, and analyzed cell cycle with FACs analysis. HuR KO cells showed a small, but significant, decrease in number of cells in the G2-M phase, and even smaller, insignificant, increases in the number of cells in the G1 and S phases (Figure 40). This data suggests that HuR depletion may have a slight impact on the cell cycle, specifically on mitosis. However, the reduced proliferation rates observed in the HuR KO cells is not explained by a decrease or block in synthesis.
Figure 40: Cell Cycle Analysis of TFO, HuR WT controls, and HuR KO clones.

FACs analysis of cell cycle. Data is reported as mean and standard deviation of percentage of cells in each G0/G1, S or G2/M. P-values were calculated using a t test. Cell lines tested: TFO total, E1-1 WT, E2-9 WT, E1-10 KO, E2-20 KO. N=2.

To see if increased cell death was contributing to the reduced proliferation rate, we assessed the percent of live, dead and apoptotic cells. Cells were stained with FITC-annexin V and DAPI, and analyzed with FACs analysis. Compared to HuR WT cells, HuR KO cells had significantly less live cells and more dead cells, although no significant change in the apoptotic cell population was observed (Figure 41). It is important to point out that a large percentage of cells from both the WT and KO cell lines were determined to be apoptotic. However, this is likely due to some spectral overlap between FITC and YFP. Since the TFO cells contain a YFP-tagged oncogenic RAS, the percentage of apoptotic cells is likely overestimated using this assay. In
conclusion, the reduced proliferation observed in the HuR KO cells may be due to increased cell death, and it remains to be determined if this increase in cell death is due to apoptosis and/or necrosis.

Finally, we measured anchorage independent growth with a soft agar assay. HuR KO with both gRNAs significantly reduced colony formation compared to both TFO totals and WT clonal controls (Figure 42). This suggests that HuR is necessary for anchorage independent growth, and may be necessary for tumor formation.

![Figure 41: FITC-annexin V and DAPI staining to assess live, apoptotic and dead cell populations.](image)

FACs analysis. Data is reported as mean and standard deviation of percentage of live, apoptotic and dead cells. P-values were calculated using a t test. Cell lines tested: TFO total, E1-1 WT, E2-9 WT, E1-10 KO, E2-20 KO. N=2.
Figure 42: Soft agar assay in HuR WT and KO clonal cell lines.

Soft agar assays were used to test for anchorage independent growth. Representative images of cell fields are shown to scale for WT and KO cell lines. The data were quantified by photographing 4 fields in each well. N=8. Data is presented as the number of colonies counted per well. *p<0.05. Cell lines tested: TFO total, E1-1 WT, E1-2 WT, E1-10 KO, E1-14 KO, E2-11 KO, E2-17 KO.

5.2.3 The effect of HuR knockout on transformed HMEC migration and invasion

Previous studies have demonstrated that HuR knockdown results in reduced migratory and/or invasive properties in HeLa cells, 184B5Me cells (in vitro ductal in situ carcinoma model), and wounded intestinal epithelium (Dormoy-Raclet et al., 2007; Heinonen et al., 2011; Zhuang et al., 2013). Given that our TFO cell line is both migratory and invasive, we decided to see if complete HuR knockout affects these properties in TFO HMECs.

We quantified migration using a wound healing scratch assay. HuR WT and KO cells were grown to confluency, a single scratch was made in the cell monolayer, and
scratch closure was quantified after 24 hours. Wound closure ranged from ~65-80% for both TFO total and TFO E1 and E2 WT clonal cell lines (Figure 43). In contrast, HuR depleted cells showed a significant reduction in wound closure. The E1 and E2 KO cell lines exhibited a wound closure of ~25% and ~45%, respectively (Figure 43). We then quantified invasion through a Transwell coated with a Matrigel membrane, normalized to migration without Matrigel. HuR depletion significantly decreased invasion of TFO cells, with the KO cells exhibiting a 26-37% reduction in invasive capacity compared to the TFO WT cells (Figure 44). Despite these large changes in EMT-associated phenotypes, expression of N-cadherin and Vimentin mesenchymal markers did not change in HuR KO cells compared to HuR WT cells (Figure 45). This suggests that HuR KO cells are not reverting back to a completely epithelial-like state, but that they are in an intermediate EMT state.
Figure 43: Quantification of migratory competence after HuR depletion

Wound-healing scratch assay. Representative photos and quantification of all biological replicates. Cell lines tested: TFO totals (n=4), E1-1 WT (n=4), E1-10 KO (n=3), E2-9 WT (n=4), E2-20 KO (n=3).
Figure 44: Quantification of invasive capacity after HuR depletion. Transwell assay with and without Matrigel. Shown are representative images and the mean and standard deviation of 3 biological replicates, quantified as percentage that migrated through Matrigel vs. through the membrane without Matrigel. Cell lines tested: TFO total, E1-10 KO, E2-20 KO.
5.3.4 RNA sequencing analysis of HuR CRISPR KO cells

Given the drastic phenotype of the HuR KO cells, we analyzed the transcriptomes of both WT and KO clonal population with RNA sequencing. We sequenced 4 WT cell lines and 7 KO cell lines on the Illumina Hi-Seq 2000/2500 platform (125bp PE; E1-1WT, E2-6WT, E1-4KO, E2-10KO, E2-20KO) or the Illumina Hi-Seq 4000 platform (150bp PE; E1-17WT, E2-9WT, E1-10KO, E1-14KO, E1-12KO, E1-14KO) at Duke University’s Center for Genomic and Computational Biology. We obtained approximately 50-75 million reads per library. Surprisingly, HuR ablation did not significantly alter the transcriptome, and, in fact, HuR WT and HuR KO cell lines had a very tight correlation (R=0.985) (Figure 46). In total 23 transcripts out of ~10,000 expressed genes changed significantly with a q-value <0.05 and log 2 fold change of at least 1 (Table 7). Out of these 23, 9 decreased expression and 14 increased expression in
response to HuR KO. 6/9 of the mRNAs that decreased expression and 6/14 of the mRNAs that decreased expression had a HuR LOD score >0. A few of these genes have known roles in the observed KO phenotypes. For example, MARCKS is a filamentous actin crosslinking protein that plays a role in cell shape and motility (Hartwig et al., 1992), and it is positively associated with metastasis in both lung and colon cancer (Chen et al., 2014; Rombouts et al., 2013). TGFB protein family members can have anti-proliferative functions through upregulating p21 and p15, although they can also be pro-invasive depending on the context (Leivonen and Kahari, 2007). The cell cycle regulator CCND2 is upregulated in response to HuR KO, and this gene is often lost in human breast carcinomas (Evron et al., 2001). A recent study found elevated ENPP1 expression in metastatic breast cancer cell lines compared to non-metastatic cell lines as well as metastatic breast tumors compared to normal epithelium (Lau et al., 2013). However, given the drastic phenotypes observed, we would expect more drastic changes to the gene expression program in response to HuR KO. It has been previously demonstrated that individual cells within a cancer cell population exhibit high transcriptome variability (Nguyen et al., 2016). It is possible that the clonal nature of these WT and KO cell lines results in variability that is greater than the response to HuR. Therefore, we adjusted significance values for the remaining analyses.
Figure 46: RNA-sequencing analysis in HuR WT and HuR KO TFO HMECs

Comparison of RNA-sequencing data in HuR WT TFO HMEC clones (x-axis; n=4) and HuR KO TFO HMEC clones (y-axis; n=7). Blue denotes an increase of at least 2-fold and red denotes a decrease of at least 2-fold in HuR KO.
Table 7: All genes that changed significantly (q<0.05) after HuR KO

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Log 2 Fold Change</th>
<th>HuR target</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAM46C</td>
<td>5.71356</td>
<td>No</td>
</tr>
<tr>
<td>SAA1</td>
<td>5.04869</td>
<td>No</td>
</tr>
<tr>
<td>CES1</td>
<td>4.86389</td>
<td>No</td>
</tr>
<tr>
<td>KRT5</td>
<td>4.51134</td>
<td>No</td>
</tr>
<tr>
<td>IL8</td>
<td>3.72139</td>
<td>Yes</td>
</tr>
<tr>
<td>ID4</td>
<td>3.57572</td>
<td>No</td>
</tr>
<tr>
<td>RTN4RL1</td>
<td>3.50345</td>
<td>No</td>
</tr>
<tr>
<td>TGFB3</td>
<td>2.99568</td>
<td>Yes</td>
</tr>
<tr>
<td>C3</td>
<td>2.74736</td>
<td>No</td>
</tr>
<tr>
<td>COL6A1</td>
<td>2.55727</td>
<td>No</td>
</tr>
<tr>
<td>TGFA</td>
<td>2.40479</td>
<td>No</td>
</tr>
<tr>
<td>TNS4</td>
<td>2.29604</td>
<td>No</td>
</tr>
<tr>
<td>BIRC3</td>
<td>1.54459</td>
<td>Yes</td>
</tr>
<tr>
<td>CCND2</td>
<td>1.54459</td>
<td>Yes</td>
</tr>
<tr>
<td>MARCKS</td>
<td>-1.7085</td>
<td>Yes</td>
</tr>
<tr>
<td>MLLT3</td>
<td>-1.85854</td>
<td>Yes</td>
</tr>
<tr>
<td>FBN2</td>
<td>-2.3049</td>
<td>Yes</td>
</tr>
<tr>
<td>ENPP1</td>
<td>-2.40621</td>
<td>Yes</td>
</tr>
<tr>
<td>MXRA8</td>
<td>-2.55284</td>
<td>No</td>
</tr>
<tr>
<td>FAM171B</td>
<td>-2.59999</td>
<td>Yes</td>
</tr>
<tr>
<td>MAN1A1</td>
<td>-2.75986</td>
<td>Yes</td>
</tr>
<tr>
<td>RNF128</td>
<td>-3.47617</td>
<td>No</td>
</tr>
<tr>
<td>MEF2C</td>
<td>-3.59256</td>
<td>No</td>
</tr>
</tbody>
</table>

To see if HuR targets respond to HuR KO more than non-targets, we plotted the log 2 fold change of mRNAs according to either their LOD scores or number of binding sites (Figure 47). Not taking into consideration significance, mRNAs with higher LOD scores or more binding sites per mRNA showed a stronger response to HuR depletion.
Both higher LOD scores and higher binding sites correlated with a decrease in mRNA levels, which is the expected response based on HuR’s canonical role as an mRNA stabilizer. Interestingly, mRNA targets of CELF1 decreased expression in response to HuR KO, despite not having HuR binding sites (Figure 48). Moreover, targets of both CELF1 and HuR decreased more than HuR targets alone (Figure 48), further suggesting combinatorial interactions between HuR and CELF1.

![Cumulative distribution function (CDF) plots showing changes in expression levels after HuR KO according to HuR LOD scores or number of binding sites.](image)

**Figure 47:** Cumulative distribution function (CDF) plots showing changes in expression levels after HuR KO according to HuR LOD scores or number of binding sites.
Figure 48: CDF plots showing the response of HuR and CELF1 targets to HuR depletion

When changing the significance value from \( q<0.05 \) to \( p<0.05 \), 349 mRNAs changed in response to HuR KO, with 225 increasing and 124 decreasing expression. Of these mRNAs that changed, 99 mRNAs that increased expression were HuR targets and 82 that decreased expression were HuR targets. For mRNAs that changed expression, we observed GO category enrichment for categories related to the observed phenotypes, including regulation of growth, biological adhesion and extracellular matrix organization. GO category analysis on HuR targets that changed expression in the HuR KOs showed similar enrichments.
Figure 49: Gene Ontology analysis of all mRNAs that changed expression in response to HuR KO

A) mRNAs that increased expression in HuR KOs p<0.05; B) mRNAs that decreased expression in HuR KOs p<0.05 C) HuR targets that changed expression after HuR KO p<0.05

5.3.5 The effect of HuR knockout on miRNA-mediated suppression and miRNA expression

Previous studies have suggested that HuR can act both synergistically and antagonistically with miRNAs (Simone and Keene, 2013). Therefore, we analyzed the response of HuR targets with miRNA sites within 10nt of identified binding sites versus mRNA targets with no nearby miRNA sites. HuR stabilized mRNAs just as well, if not better, when binding near miRNA sites (Figure 50a), suggesting that HuR binding may antagonize miRNA-mediated suppression of RNA abundance. HuR’s stabilizing effect for sites near miRNA seed matches may be a result of directly blocking miRNA/RISC binding and/or competition for functional outcome.

We then tested HuR’s effect on miRNA abundance using a qRT-PCR miRNA array profiling system. HuR KO resulted in a significant decrease in miR-10b abundance (Figure 50b). MiR-10b has well documented roles in metastasis (Ma et al., 2010; Ma et al., 2007; Ma and Weinberg, 2008), and therefore HuR protein may regulate migration and invasion, at least in part, by regulating miR-10b expression. In addition, HuR and miR-10b targets significantly overlap, similar to what is typically seen for HuR and many other miRNAs (Figure 50c). We obtained a list of predicted miR-10b targets from
TargetScan (Agarwal et al., 2015), and compared this list to mRNAs that responded to HuR KO. No predicted mRNA targets of miR-10b significantly changed expression after HuR KO (p<0.05 and L2FC 2), suggesting that, if miR-10b plays a role in migration or invasion in this system, it is functioning at the level of translation and not mRNA decay.

Figure 50: Analysis of miRNA regulation in HuR KO cells

A) CDF plot showing changes in expression levels after HuR KO for mRNAs with and without miRNA sites <=10nt from the HuR site. B) miRNA array expression analysis. Dashed lines represent a 2-fold change. Data is reported as Ct values normalized to controls. C) Percent of microRNA targets that are also HuR targets. miR-10b targets are all predicted targets in TargetScan, HuR targets are targets identified by DO-RIP-seq. Bar in middle is the average overlap with predicted targets of miR-205, miR-122 and miR-34a. HuR-random overlap is the average overlap with 3 random gene sets the same size as the ones used for the miR-overlap analysis.
5.4 HuR binding site position-dependent regulation of alternative polyadenylation

Our DO-RIP data clearly demonstrate that HuR typically binds to multiple sites within a target message. It is likely that only some of these sites are regulatory. In addition, HuR has many known roles in post-transcriptional regulation, and it is likely that specific binding sites are important for different aspects of regulation. Moreover, our RNA-seq data of HuR KO cells indicates that, while HuR targets do respond more to HuR KO than non-targets, and the size of this response correlates with number and strength of binding sites, the effect of HuR KO on target mRNA abundance is small. HuR KO results in drastic phenotypic changes. Therefore, we hypothesized that HuR’s main effect in TFO HMECs is not on mRNA stability, but on something that would be more subtly reflected in the total mRNA levels. Transcripts that are regulated at the level of alternative polyadenylation (APA) are not generally regulated at the level of mRNA abundance. Therefore, we explored a potential role for HuR in regulating APA in TFO HMECs. In this section, we identify a role for HuR alternative polyadenylation that is determined by HuR’s binding position relative to the polyadenylation (poly(A)) signal sequence.

5.4.1 Analysis of HuR binding position relative to regions important for APA regulation

Based on binding sites identified in TFO HMECs by DO-RIP-seq, we found HuR binding to be enriched in 3’UTRs 50 nucleotides upstream and 25 nucleotides
downstream from the proximal 3’ end termini (**Figure 51**). Specifically, 1483 mRNAs had HuR binding sites 25 nucleotides downstream (DSE) and 2135 mRNAs had HuR binding sites 50 nucleotides upstream (USE). The human cleavage and polyadenylation core machinery complexes are known to bind to both of these regions, and thus they are critical for regulating polyadenylation (Gruber et al., 2014; Martin et al., 2012; Yao et al., 2012).

**Figure 51: Histogram of HuR binding sites around the poly(A) signal sequence**

Red line depicts the number of HuR binding sites compared to a shuffled binding sites control (black line) beginning at the position relative to the most proximal 3’ terminus (dashed vertical line), in nucleotides, as shown on the X-axis. The region highlighted in green is where the CPSF complex binds and the region highlighted in purple is where the CstF complex binds. Both of these regions were determined by previously published CLIP experiments (Martin et al. 2012; Yao et al. 2012). The region highlighted in orange is the location of the proximal poly(A) signal, which is most commonly AATAAA and generally begins -21nts from the 3’ terminus (Gruber et al. 2014).
We next investigated whether or not HuR binding site position around the poly(A) site correlated with poly(A) site usage. To determine poly(A) site usage we used a previously published 3’-end RNA sequencing dataset obtained from RAS transformed MCF10a (immortalized breast) cells (Lianoglou et al., 2013). We reasoned that, since this is a similar system of isogenic transformation, the relative 3’UTR isoform ratios may be similar to those expressed in TFO HMECs. We found that transcripts with high proximal poly(A) site usage more frequently had HuR binding sites upstream of the poly(A) site when compared to transcripts with downstream HuR binding sites, which were more likely to skip the proximal poly(A) site (Figure 52a). Furthermore, for messages with a strong LOD at the USE, proximal poly(A) site usage was higher, whereas for messages with a strong LOD at the DSE, proximal poly(A) site usage was lower (Figure 52b). HuR has been demonstrated to promote long 3’UTR isoform usage in certain mRNAs through binding downstream of poly(A) cleavage sites (Hilgers et al., 2012; Soller and White, 2003; Zhu et al., 2007). However, our global data suggest a model whereby HuR can also promote short isoform usage when bound upstream of the poly(A) site (Figure 52c).
Figure 52: Global evidence for HuR binding site position dependent APA

A) Histogram showing percentage of proximal poly(A) site usage on the X-axis, as determined by 3′-end sequencing in RAS transformed MCF10a cells (Lianoglou et al. 2013), and frequency of HuR binding sites either upstream (orange) or downstream (blue) of the proximal poly(A) site on the Y-axis, as determined by HuR DO-RIP-seq in TFO HMECs. B) CDF plots of proximal poly(A) site usage and HuR LOD scores for messages where HuR binding sites were identified at the USE (left) or DSE (right). Black lines are messages where the proximal poly(A) site was used >=20% of the time, and red lines are messages where the proximal poly(A) site was used <20% of the time. C) Model of HuR binding site position dependent APA derived from global evidence.
5.4.2 Analysis of HuR binding position on 3’UTR isoform abundance

To test the model of HuR position-dependent polyA site choice, we selected candidate mRNAs from our DO-RIP seq data: two mRNAs with HuR binding sites at the DSE (AES and TMEM55a), two mRNAs with HuR binding sites at the USE (RPN1 and CCNF), as well as two negative controls- one HuR target with no USE or DSE binding sites (KIF11) and one mRNA that is not a HuR target (LAMC). Each of the mRNAs selected as candidates had only two expressed 3’UTR isoforms (referred to as long isoform and short isoform). We designed two sets of RT-PCR primers to the 3’UTRs of these genes: one set specific to the region upstream of the proximal poly(A) site (universal primers) and one set specific to the region downstream of the proximal poly(A) site (distal primers) (Figure 53). We performed RT-PCR and compared expression of long isoforms after HuR KO, normalized to the universal primers in order to control for mRNA abundance. For messages with HuR binding sites at the DSE (AES and TMEM55a), the abundance of the long isoform decreased after HuR knockout (Figure 53). Conversely, the abundance of the long isoform increased upon HuR knockout when HuR bound to the USE (RPN1 and CCNF) (Figure 53). The sum of this data provides evidence that HuR expression is important for APA of a subset of mRNA targets, and that HuR binding sites positioned near the proximal poly(A) signal sequence may be important in regulating 3’UTR isoform abundance.
Universal and distal primers were designed upstream and downstream, respectively, of the proximal 3' terminus of candidate mRNAs with downstream HuR binding sites (DSE) or upstream HuR binding sites (USE). Two mRNA controls, one HuR target with no USE or DSE binding sites (KIF11) and one mRNA that is not a HuR target (LAMC) were also tested. Shown is the log2 fold change of the long isoform in HuR KO TFO cells compared to HuR WT TFO HMECs, normalized to universal primers. N=3.

5.4.3 Analysis of HuR binding position on protein expression

It has been shown for specific genes that shorter 3'UTRs can produce more protein than longer 3'UTR isoforms (Mayr and Bartel, 2009), most likely due to loss of
negative regulatory elements contained within longer 3’UTRs. To test the contribution of HuR binding site position around the proximal poly(A) site to protein expression levels, we cloned the full length 3’UTRs of TMEM55a (HuR binds DSE) and RPN1 (HuR binds USE) into firefly luciferase reporter vectors. We then used site directed mutagenesis to mutate the HuR binding sites in the DSE and USE, respectively (Figure 54). Constructs were co-transfected with a Renilla luciferase expression vector into both HuR WT and HuR KO TFO HMECs and assayed using a dual luciferase reporter assay.

Mutating the DSE HuR binding site of TMEM55a resulted in a significant increase in luciferase signal compared to that when a wild type TMEM55a 3’UTR was expressed in HuR competent cells (Figure 55). In contrast, the differences in luciferase signal when either the mutated or wild type TMEM55a 3’UTR constructs were expressed in HuR KO cells was non-significant. Mutating the USE HuR binding site of RPN1 resulted in a decrease in luciferase signal compared to the wild type 3’UTR construct, although this decrease did not reach significance (Figure 55). In conclusion, mutating HuR binding sites around the poly(A) site effects target protein expression, but the effect is small.
HuR WT or KO TFO HMECs were co-transfected with Firefly luciferase reporter constructs depicted in Figure XX and Renilla luciferase expression vectors. Luciferase assays were normalized to cells expressing an empty expression vector. Shown is the log2 fold change of normalized signal of the mutant constructs compared to the wild type constructs. Data are reported as mean and standard deviation, and p-values were calculated using a standard t-test.
5.4.4 HuR binding position and association with core APA machinery

The recruitment mechanisms for the core APA machinery complexes are poorly understood. To determine whether or not HuR binding directly influences recruitment of APA complexes, we immunoprecipitated the APA factor CSTF64 in both HuR WT and HuR KO TFO cells, isolated protein-bound RNA and performed RT-PCR with distal and universal primers for TMEM55a, RPN1 and controls (Figure 56). The long isoform of TMEM55a was significantly decreased in the CSTF64-bound mRNA pool after HuR KO (Figure 56b). Contrastingly, the long isoform of RPN1 was significantly increased in the CSTF64-bound mRNA pool after HuR KO (Figure 56b). This data suggests that HuR may play a role in the recruitment of the CSTF complex.
A) CSTF64 western blot for RIP inputs, depleted supernatants, and immunoprecipitated protein for CSTF64 RIP and Normal Mouse Serum negative RIPs

B) Quantitative real-time PCR analysis of CSTF64 associated mRNAs. Shown is the mean and standard deviation of the log2 fold change of the enrichment of the long isoform in the CSTF64 IP after HuR KO, normalized to negative IP; p<0.05.

5.4.5 Gene ontology analysis of mRNAs with HuR binding sites upstream or downstream of the poly(A) signal sequence

Finally, we performed GO analysis to determine if mRNAs with HuR binding sites at USE or DSE are enriched in specific functional categories. We found that mRNAs
with USE or DSE HuR binding sites are enriched for categories related to mRNA processing and metabolism, suggesting that HuR may regulate 3’UTR length, both lengthening and shortening, for mRNAs involved in these processes (Figure 57).

Interestingly, mRNAs with USE HuR binding sites are enriched for categories related to cell cycle and cell polarity (Figure 57, Left panel), suggesting that HuR may promote the shortening of 3’UTRs, potentially increasing the protein expression, of genes involved in proliferation and motility, both of which are related to the observed HuR KO phenotypes. With the exceptions of CCND2 and TGFA, none of the genes in these categories changed their mRNA abundance after HuR KO, suggesting that alternative APA may be an independent mechanism by which HuR can regulate gene expression.

Figure 57: Gene ontology analysis for mRNAs with HuR binding sites upstream or downstream of the poly(A) signal sequence
5.5 Discussion

5.5.1 The presence of a GU secondary motif and cooperative regulation with CELF1 may be two separate mechanisms that determine HuR RNPs during oncogenic RAS transformation

In this chapter, we report the first global quantification of HuR binding sites in a dynamic system. In both IMO and TFO cells, HuR bound to the same U-rich consensus sequences primarily in 3’UTRs, comparable to other studies of HuR binding sites in other cell lines (Figure 23) (Lebedeva et al., 2011; Mukherjee et al., 2011; Nicholson et al., 2017). Individual binding sites did not gain or lose association with HuR during oncogenic RAS transformation, but rather sites quantitatively changed in the strength of HuR association (Figure 23). Mechanistic studies aimed at determining how HuR stabilizes mRNA targets have been scarce. Therefore, we sought to use the quantitative differences in binding sites to identify a mechanism of HuR regulation. We searched for secondary motif enrichment within HuR binding sites, and identified a GU-rich motif significantly enriched within binding site regions (Figure 25). This was the most significantly enriched secondary motif, and it was contained within 1% of HuR binding sites. Importantly, almost 60% of binding sites containing this secondary motif lost association with HuR during RAS transformation.

Transcripts can associate with many trans regulators with often opposing activities (Lal et al., 2004), and investigating mechanisms of competition and cooperation among these factors is important in fully understanding complex post-transcriptional
networks in disease states. We hypothesized that our observation could be explained by a mechanism in which another RBP binds to the GU-rich motif and competes with HuR. Indeed, this mechanism of competition has been previously described to occur between HuR and CELF1, which binds to GU-rich sequences for binding to mRNA encoding Occludin, c-Myc, and E-cadherin (Liu et al., 2015; Yu et al., 2016; Yu et al., 2013b). Therefore, we used DO-RIP-seq to see if competition between HuR and CELF1 is a widespread occurrence. We identified and quantified CELF1 binding sites transcriptome wide and compared to HuR binding sites. Indeed, CELF1 preferentially bound to the GU-rich motif and ~31% of its binding sites overlapped with HuR binding sites (Figures 26 and 27). However, our results suggested that CELF1 is not competing with HuR for binding to the GU-rich motif, but rather may be acting synergistically with HuR (Figure 28).

While many studies have reported CELF1 to be an mRNA destabilizer (Dasgupta and Ladd, 2012), a recent study from Joel Neilson’s lab reported a role for CELF1 in positively regulating the translation of ten mRNAs encoding proteins necessary and sufficient for EMT (Chaudhury et al., 2016). Interestingly, six out of the ten mRNAs identified in this study (EGR3, JUNB, PPARC1A, SEMA6D, SSBP2 and TICAM2) are also HuR targets in our study. The other four mRNAs (DUSP2, FOSB, PAD12 and SNAI1) have little to no expression in our system. Moreover, we demonstrated that HuR depletion in transformed cells reduced migration and invasion (Figures 43 and 44), and
mRNAs that were regulated by both HuR and CELF1 decreased in expression more than targets of HuR alone (Figure 48). This data, in combination with the study from Joel Neilson’s group, suggests the fascinating possibility that HuR and CELF1 cooperate to regulate the EMT phenotype by positively influencing the translation of a subset of functionally related mRNAs.

5.5.2 HuR maintains cancer phenotypes through mechanisms other than mRNA stability

Despite the dramatic HuR knockout phenotype, only a few mRNAs significantly changed expression in response to HuR depletion (Figure 46 and Table 7). Other global studies of mRNA expression levels in HuR depleted cell lines showed similar findings. For example, while HuR knockdown in HeLa cells reduced proliferation, adhesion, migration, and invasion, only 18 mRNAs were significantly downregulated upon HuR depletion (Dormoy-Raclet et al., 2007). This study found that HuR directly bound to B-actin encoding mRNA, which decreased expression 3-fold when HuR was depleted. They suggest that HuR’s effect on migration and invasion was due to B-actin stabilization and translation, which would be expected to affect cytoskeletal functions. Interestingly, in TFO HMECs, HuR binds very strongly to B-actin encoding mRNA. However, neither mRNA levels nor protein levels of B-actin are altered in HuR KO TFO HMECs, which suggests HuR affects migration and invasion via a different mechanism in our cell system. Indeed, although we also only identified very few mRNAs that were
significantly up or down regulated after HuR depletion, none of the mRNAs identified in our study were identified as being altered in HeLa cells.

A more recent study demonstrated that HuR KD reduced invasion and migration in activated microglia (Matsye et al., 2017). Only 172 mRNAs changed expression levels in response to HuR KD, and these mRNAs were enriched for proliferation, migration and inflammatory response, although these changes in expression levels were demonstrated to be through the regulation of promoter activity, not mRNA stability. MMP12 was important for HuR’s role in migration and invasion. We did not observe changes in MMP12 expression levels in HuR KO TFO HMECs.

The Brody lab has extensively studied the effect of HuR depletion in pancreatic ductal adenocarcinoma (PDA). HuR CRISPR KO in a PDA cell line reduced proliferation, increased cell death, altered the cell cycle, and reduced anchorage independent growth (Lal et al., 2017), similar to what we observed in our HMEC model. Furthermore, HuR KO also resulted in higher sensitivity to chemotherapy and glucose deprivation and reduced tumor formation in mice (Lal et al., 2017). HuR shRNA knockdown in PDA cells resulted in expression changes for 87 mRNAs (Jimbo et al., 2015). The most significant of these was TGFB1, which increased 2.46-fold. The rest of the significantly changing mRNAs exhibited very small fold changes, with the average significant change being around 25% up or down. Therefore, HuR KD in a cell culture
model of PDA had small effects on the transcriptome while having large phenotypic effects, similar to what we observed in our breast cancer cell culture model.

Interestingly, although HuR KD/KO resulted in similar phenotype changes in HeLa cells (Dormoy-Raclet et al., 2007), activated microglia (Matsye et al., 2017), PDA cells (Jimbo et al., 2015; Lal et al., 2017), and RAS transformed HMECs (this chapter), the mRNAs identified as being significantly up or downregulated in each experiment did not overlap (with the one exception of CDK2, which was identified as being downregulated after KD in both HeLa and PDA cells). Given that HuR has been shown to regulate groups of functionally related mRNAs, this is surprising. The sum of these studies suggests that HuR is necessary for the maintenance of cancer-related traits, but not at the level of mRNA stability.

Despite lack of significance, RNA-sequencing of multiple replicates of HuR KO and wild-type clonally expanded cell lines demonstrated that HuR targets trended towards decreasing abundance after HuR depletion, and the size of this effect was directly related to both the LOD score and the number of binding sites (Figure 47). This suggests that HuR is regulating mRNA targets in RAS transformed HMECs, but that the primary mechanism of this regulation is at a level that would be expected to be only subtly reflected in the transcriptome. One possibility is that HuR is regulating translation. In fact, there is evidence to support a role for HuR in translation without corresponding changes to mRNA levels (Leandersson et al., 2006; Meng et al., 2008;
Saunus et al., 2008; Yan et al., 2012). Another possibility is alternative polyadenylation. Changes in APA regulation may influence the ratios of long to short isoform usage, which would not be strongly reflected in total RNA-sequencing data due to the method not being quantitative with respect to 3’UTR isoform abundance.

Our DO-RIP-seq data revealed a striking binding pattern for HuR around proximal poly(A) sites. Binding sites were significantly enriched 50 nucleotides upstream and 25 nucleotides downstream from the proximal 3’ end termini (Figure 51). These positions correspond to the precise positions of CPSF and CstF complex binding, as previously determined by CLIP experiments (Martin et al., 2012; Yao et al., 2012). Globally, mRNAs where HuR bound upstream of the proximal poly(A) site were more likely to use that site, whereas the proximal poly(A) site was more likely to be skipped when HuR bound downstream (Figure 52). This suggested a mechanism by which HuR may directly interact with the core APA machinery and function in either recruiting or blocking cleavage and polyadenylation complexes. We further explored this mechanism for candidate mRNA with upstream or downstream binding sites. When HuR was depleted from cells, mRNAs with upstream binding sites showed an increase in long 3’UTR isoform usage, whereas mRNAs with downstream binding sites showed a decrease in long 3’UTR isoform usage (Figure 53), validating a dual-role for HuR in APA site decisions that is dependent upon binding site position around the 3’ termini.
Position-dependent mechanisms of poly(A) site choice have been identified for a few other RBPs. In a murine model of myotonic dystrophy, when bound to a region upstream of the poly(A) site, the muscleblind-like (MBLN) RBP enhanced use of the proximal site. However, when the MBLN binding site overlapped the poly(A) site, it repressed use of that site through blocking the CPSF complex (Batra et al., 2014). In neuronal cells, FUS binding downstream of poly(A) sites recruited CPSF160 and promoted polyadenylation. In contrast, when it bound upstream of the poly(A) site it stalled RNA polymerase II and reduced gene expression (Masuda et al., 2015). HnRNP1 has been demonstrated to suppress APA when bound to a downstream element, and promote APA when bound to an upstream element (Hall-Pogar et al., 2007). While HuR was previously known to bind to regions downstream of the poly(A) site and promote 3’UTR lengthening (Dai et al., 2012; Hilgers et al., 2012; Slevin et al., 2007; Soller and White, 2003; Zhu et al., 2007), our data indicate that, similar to MBLN and hnRNP1, it can also bind to regions upstream of proximal poly(A) sites, which results in 3’UTR shortening. We suggest that this could be a widespread mechanism by which HuR regulates cancer traits. Indeed, mRNAs with HuR binding sites around USE and DSE include are enriched for those involved in cell cycle regulation, cell division, and mRNA processing (Figure 57).

Mutating HuR binding sites upstream and downstream of the poly(A) site had an effect on protein expression (Figure 55). Mutating downstream HuR binding sites
resulted in significantly higher luciferase reporter gene expression and mutating upstream HuR binding sites resulted in lower expression. However, although these changes in protein expression were significant, they were very small, with the largest change being a log 2 fold change of 0.4. An and colleagues reported that long and short 3’UTR isoforms produce comparable amounts of protein that differ in their function due to differences in localization (An et al., 2008). It remains to be determined if these small effects on protein expression are significant.

We suggest that HuR regulates APA through a mechanism involving recruitment of CSTF64 when bound downstream of the poly(A) site and blocking CSTF64 recruitment when bound upstream of the poly(A) site. When HuR was depleted from cells, the long isoform of messages with downstream binding sites showed reduced enrichment in CSTF64 IPs, and messages with upstream binding sites showed increased enrichment of the long isoform in CSTF64 IPs (Figure 56). One potential caveat is that, although the IP-associated isoforms are normalized for total mRNA abundance, we cannot completely distinguish whether or not changes in long isoform association with CSTF64 is simply due to changes in long isoform abundance. However, another study has also suggested that HuR prevents CSTF64 binding (Zhu et al., 2007). Furthermore, the fraction of long 3’UTR isoforms may be underestimated in our studies since we cannot completely distinguish whether lower expression of long isoforms is due to decreased usage of the proximal poly(A) site or higher mRNA instability in
comparison to the short 3’UTR isoforms. However, the sum of our data provide clues that differential poly(A) site usage is an important HuR-dependent mechanisms of regulation that will be the subject of future investigations.

5.5.3 Therapeutic Implications

Although we did not observe remodeling of HuR RNPs during RAS transformation (Figures 29 and 30), HuR was necessary for the maintenance of tumorigenic properties, including anchorage independent growth, proliferation, migration and invasion (Figures 39-44). HuR CRISPR KO cells showed a significant reduction in migration and invasion, yet still robustly expressed N-cadherin and Vimentin proteins (Figure 45), suggesting that they have not reverted back to an epithelial state, but rather are intermediate. Activation of an EMT program has been linked to chemotherapeutic resistance, and therefore new therapeutic approaches to targeting this program are of interest (Shibue and Weinberg, 2017). Our data suggest that targeting HuR may result in the simultaneous loss of tumor initiating capacity and metastatic traits, while maintaining cells in an intermediate EMT state. While inducing the reverse of EMT, the so-called mesenchymal-to-epithelial transition, is a potentially useful therapeutic approach that has been demonstrated to sensitize cells to chemotherapy (Pattabiraman et al., 2016), the precise timing of this reversal will be important. Reversion back to an epithelial state is associated with the last stage of metastasis- secondary tumor formation. Therefore, targeting HuR and reverting cells
back to an intermediate, non-metastatic state may provide a way to avoid positively impacting colony formation.

Another interesting finding was that IMO HMECs were not viable following HuR depletion. HuR has been shown to play an essential role in mouse embryogenesis and postnatal development. Elavl1/ HuR null mice exhibit embryonic lethality due to defects in placental, spleen and bone development (Katsanou et al., 2009). The Hla lab demonstrated an essential role for HuR in progenitor cell survival in mice (Ghosh et al., 2009). Postnatal deletion of HuR resulted in hematopoietic organ atrophy, loss of intestinal villi, obstructive enterocolitis and lethality within 10 days (Ghosh et al., 2009). Progenitor cells underwent apoptosis, whereas quiescent stem cells and differentiated cells were unaffected (Ghosh et al., 2009). While many studies, including our own, have demonstrated that cancer cells are viable when HuR is depleted, the intermediate state of the IMO cells may be less differentiated (i.e. they are neither fully epithelial nor mesenchymal), and therefore HuR depletion in these intermediate cells may be lethal. More studies are needed to determine why HuR KO is lethal in immortalized HMECs, but HuR may provide a potential therapeutic target for targeting cells that are in an intermediate, or primed, EMT state.

Finally, HuR depletion resulted in a significant reduction of miR-10b expression levels (Figure 50). MiR-10b is well-studied in the context of EMT (Ma et al., 2010; Ma et al., 2007; Ma and Weinberg, 2008), and our data suggest a novel role for HuR in
maintaining miR-10b expression levels in metastatic cells. MiR-10b is transcriptionally upregulated by Twist (Yang et al., 2004), which is targeted by HuR in our cell system, thus suggesting a potential indirect mechanism by which HuR regulates miR-10b expression. Importantly, therapeutic knockdown of miR-10b has been demonstrated to inhibit lung metastasis in a mouse mammary tumor model (Ma et al., 2010), and therefore HuR’s role in maintaining migration and invasion in TFO HMECs may be, at least in part, due to regulation of this miRNA. While antagomirs targeting miR-10b have been effective in mouse models (Ma et al., 2010), targeting HuR may provide an alternative way to target miR-10b and the metastatic phenotype.
6. Conclusions and Future Directions

Tumors are highly heterogeneous, and a global understanding of the complex molecular mechanisms regulating breast cancer origin and progression are still lacking. In this dissertation, we aimed to gain a better understanding of these mechanisms. We developed a streamlined protocol for generating a cell culture model of stepwise tumorigenesis, beginning with normal human mammary epithelial cells and progressing through immortalization and RAS transformation. We quantified mRNA expression, alternative splicing, and miRNA expression changes, and explored mechanisms of post-transcriptional coordination by the RBP HuR. One of the major findings of this work is that EMT can be initiated through two sequential stages: transcriptional priming followed by oncogenic RAS-triggered post-transcriptional regulation. Additionally, we demonstrated that HuR is necessary for maintenance of cancer traits. Our study is the first to identify and quantify transcriptome wide binding sites for any RBP during tumorigenesis, and we report that HuR quantitatively, but not qualitatively, changed association at individual mRNA binding sites during RAS transformation. We explored mechanisms that may determine changes in quantitative binding, and revealed that the presence of a GU-secondary motif and cooperative regulation with CELF1 may be two separate mechanisms that control HuR RNPs during oncogenic RAS transformation. Additionally, HuR maintained the RAS-induced cancer phenotypes through mechanisms other than mRNA stability. One potential mechanism is alternative
polyadenylation, and we demonstrated that HuR can define 3'UTR length dependent upon its binding site position relative to the 3'terminus. While this work significantly adds to our knowledge of post-transcriptional regulation during tumorigenesis, there are many more questions that need to be answered. In the rest of chapter, I discuss future directions of interest as they relate to the major findings presented in this work.

6.1 Elucidating genetic alterations that lead to EMT-related transcriptional priming

Activation of EMT is known to occur early during tumorigenesis, but the genetic alterations that induce pre-malignant cells to activate an EMT program are poorly understood. While the acquisition of most cancer hallmarks described by Hanahan and Weinberg can be explained as a model of Darwinian evolution, in which genetic mutations confer a selective advantage, activation of migration and invasion does not provide a survival advantage to the primary tumor (Gerlinger et al., 2012; Gupta et al., 2005; Hanahan and Weinberg, 2011). Therefore, it has been postulated that mutations that contribute to metastatic competence may be harbored within genetic lesions for another trait that is selected for within the primary tumor (Gupta et al., 2005). Using a genetically defined model system enabled us to narrow down the pathways in which perturbation or mutation may activate an EMT expression program. By activating telomere maintenance, mutating the p53 and pRb tumor suppressor pathways, and expressing mutated c-Myc, we have found that an EMT transcriptional program is activated. These immortalized HMECs can now serve as a foundation for future studies
to determine which specific pathway, or combination of pathways, results in the activation of the EMT transcriptional program. It is likely that hTERT is not sufficient for our observed transcriptional priming, as HMECs only expressing hTERT could not be RAS-transformed (Elenbaas et al., 2001). Therefore, future studies should prioritize the investigation of the p53, pRb and c-Myc pathways as potential activators of the EMT transcriptional program.

Each of these pathways, or any combination of these, could reasonably induce EMT-related transcriptional priming. P53 directly activates both miR-205 and miR-200c expression, two of the most significantly downregulated miRNAs during immortalization (Chang et al., 2011; Piovan et al., 2012). MiR-205 has been shown to suppress metastasis and EMT in prostate cancer (Tucci et al., 2012), breast cancer (Chao et al., 2014), and gastric cancer (Xu et al., 2016). MiR-200c is also a well-documented suppressor of metastasis in many cancer types, and it directly targets Zeb1 and Zeb2, transcriptional repressors of E-cadherin (Shibue and Weinberg, 2017). Moreover, wild-type p53 expression has been associated with EMT suppression in a number of cancers, and p53 transcriptionally activates a number of genes involved in negatively regulating metastatic traits (Powell et al., 2014). Thus, the p53 pathway is a candidate for regulating our observed EMT transcriptional priming. Likewise, c-Myc has also been associated with EMT. MCF10A cells overexpressing c-Myc had reduced E-cadherin and increased N-cadherin expression (Cho et al., 2010a). A recent study using a mouse model of breast
cancer metastasis found that the majority of early disseminated circulating tumor cells were pRb negative (Harper et al., 2016), suggesting that suppressing pRb may be important in EMT. Understanding if and how perturbation of these pathways function together to induce EMT transcriptional priming is of future interest. Knowing precisely which pathways are responsible for EMT-priming could lead to new ways to target EMT in pre-malignant lesions, prior to the gain of tumor forming or metastatic potential.

6.2 Determining post-transcriptional mechanisms induced by oncogenic RAS transformation

Our data indicate that transcriptionally primed pre-malignant cells can be induced by oncogenic RAS to become both malignant and invasive simultaneously through post-transcriptional mechanisms. Understanding the downstream mechanisms by which RAS post-transcriptionally regulates gene expression and functional outcome could provide therapeutic approaches to targeting metastasis. Our work has eliminated alternative splicing, mRNA stability and regulation by miRNAs as potential post-transcriptional mechanisms downstream of oncogenic RAS signaling, as we did not observe significant changes in splicing, mRNA abundance or miRNA expression levels during transformation. We therefore propose that the post-transcriptional regulation triggered by RAS is at the level of translational or post-translational control. We have demonstrated that oncogenic RAS expression increases protein expression of N-cadherin and Vimentin without altering their mRNA levels. We also explored the possibility that these proteins are regulated by the RBP HuR, which targets both of these messages.
However, HuR KO cells still robustly express both of these proteins. Understanding the mechanism by which the expression of these two proteins, as well as others regulated by RAS in the context of EMT and tumorigenesis, will be of interest. Changes in translation during RAS transformation could be determined through polysome gradients coupled with RT-PCR for candidate genes or deep-sequencing, or through ribosome profiling methods to determine global changes in translation during RAS transformation (Reid et al., 2015). However, our preliminary data with sucrose cushions suggest that N-cadherin and Vimentin translation is not different between immortalized and transformed cells (data not shown). In agreement with our preliminary data, polysome gradient data suggests that translation of N-cadherin and Vimentin does not change when MCF10A cells are treated with TGFB to induce an EMT (Chaudhury et al., 2016). Therefore, these two proteins may be regulated post-translationally. Given the large phenotypic differences between IMO and TFO HMECs, it is likely that both translational and post-translational mechanisms are cooperating in this system to induce the full EMT phenotype. Elucidating these precise mechanisms could provide therapeutic avenues for targeting the transition to an invasive state.

6.3 Understanding mechanisms of regulation by HuR in RAS-transformed HMECs

We explored the possibility that RAS-transformation could be altering regulation by the RBP HuR. Using the DO-RIP-seq method, we identified and quantified HuR binding sites transcriptome wide in both IMO and TFO HMECs. During RAS-
transformation, we did not observe significant remodeling of HuR RNPs. However, we did observe quantitative changes at individual binding sites. The mechanism by which HuR stabilizes an mRNA is not well-understood. It could be that HuR directly inhibits decay enzymes, or it may compete with instability factors such as miRNAs and other RBPs (Srikantan and Gorospe, 2012). Precise mechanisms are likely contextual, but there is much evidence to support the latter competition model (Lal et al., 2004; Simone and Keene, 2013). We identified a GU-rich secondary motif enriched within binding site regions that decreased association with HuR during transformation, and we hypothesized that another GRE binding RBP could be competing HuR off of overlapping mRNA targets during transformation. Based on the literature, we thought that CELF1 would be a good candidate RBP to investigate. However, the integration of CELF1 and HuR DO-RIP-seq data suggested a mechanism of synergism between HuR and CELF1. Therefore, CELF1 is not explaining the decrease in association with HuR and GU-containing binding sites. Future studies will focus on identifying the GU-binding RBP.

A recent study used RNAcompete, an in vitro method to determine sequence preferences, to identify sequence preferences for 207 RBPs (Ray et al., 2013). Data from this study can be used to identify candidate RBPs that may be competing with HuR for GU-rich binding sites during RAS transformation. Indeed, CELF family RBPs were found to have a strong preference for GU-rich sequences, and it is possible that another
CELF protein could be competing with HuR. CELF6, for example, has been shown to be able to rescue K-RAS-induced senescence (Shao et al., 2014); however, CELF6 mRNA is expressed at very low levels in our HMEC system. Other possible candidates identified through RNAcompete include RBM38 and RBM24. RBM24 shows very low mRNA expression in our system, but RBM38 is expressed. Interestingly, RBM38 has been recently shown to prevent TGFB-induced EMT in breast cancer cells (Wu et al., 2017), and it was previously suggested in the literature that HuR and RBM38 cooperate to regulate CDKN1A, which encodes p21 (Cho et al., 2010b). It is possible that HuR and RBM38 could act competitively as well, and thus, RBM38 may be a good candidate to look at for competition with HuR in the context of EMT.

HuR and CELF1 have been shown to compete for binding to mRNA targets (Liu et al., 2015; Yu et al., 2016; Yu et al., 2013b), but our data suggest that these two proteins bind to mRNA targets in a synergistic manner. Given that CELF1 has recently been shown to be necessary for EMT through the positive translational regulation of pro-EMT factors (Chaudhury et al., 2016) that are also targeted by HuR in our dataset, and our data indicate that HuR is necessary for maintaining an EMT phenotype, it is logical to suggest that these two proteins cooperate to regulate translation necessary for EMT. Therefore, future studies will focus on better understanding this mechanism. One approach would be to look at protein expression of the CELF1 regulated mRNAs identified by Joel Neilson’s group in our HuR CRISPR KO TFO HMECs. DO-RIP-seq can
also be used to globally study combinatorial regulation, and DO-RIP-seq experiments for each of these RBP s in the absence of the other would tell us if binding to a subset of co-regulated messages is dependent on the presence of both proteins.

Another novel discovery from the work presented in this dissertation is that HuR can determine both lengthening and shortening of 3’UTRs depending on binding site position. We have demonstrated this for a few candidate mRNAs, and presented global evidence that suggests this may be a widespread mechanism during transformation. This global conclusion is based on our DO-RIP-seq data combined with 3’end-sequencing data obtained by another laboratory in RAS-transformed MCF10A cells (Lianoglou et al., 2013). Interestingly, the Leslie and Mayr labs found that, during RAS-transformation of MCF10A cells, genes that changed their 3’UTR length were enriched for genes associated with cellular migration (Lianoglou et al., 2013). In line with this, we showed that mRNAs with HuR binding sites upstream of the proximal poly(A) signal sequence were enriched for categories related to cell polarity. It is plausible that APA is a widespread regulatory mechanism during RAS-transformation that is, at least in part, regulated by HuR. The integration of 3’-end sequencing experiments in IMO, TFO, and HuR CRISPR KO cell lines would determine if HuR’s position-dependent regulation of APA is a global mechanism of regulation that coordinates either the acquisition or maintenance of cancer traits in this cell system.
HuR does not globally regulate mRNA stability in RAS transformed HMECs. Although HuR has been reported in the literature to stabilize mRNA, many of these studies were conducted under stress conditions such as drug treatment or hypoxia (Chand et al., 2017; Hostetter et al., 2008; Levy, 1998; Li et al., 2013; Sheflin et al., 2004; Yang et al., 2013), and these studies looked at very specific transcripts rather than global stability (Srikantan and Gorospe, 2012). In addition, many studies have identified a role for HuR in protein expression without corresponding changes in mRNA levels (Durie et al., 2011; Lal et al., 2005; Mazan-Mamczarz et al., 2011; Yan et al., 2012) The possibility remains that HuR-dependent regulation of mRNA stability could occur primarily during stress conditions. Indeed, studies have shown that under conditions commonly encountered in the tumor microenvironment, such as hypoxia and oxidative stress, the nuclear RBP HuR is shuttled to the cytoplasm, where it stabilizes mRNAs encoding proteins involved in the stress response through the formation of RNPs (Gorospe et al., 2011; Masuda et al., 2009). Future studies could investigate the involvement of HuR in regulating a response to stresses commonly found in the tumor microenvironment, such as hypoxia.

In the context of tumorigenesis in our HMEC system, HuR’s main effect is likely on regulating translation. Future studies of how HuR affects the global translatome are of interest. Integrating global translation datasets from HuR KO cells, such as ribosome profiling or polysome gradients coupled with RNA-sequencing, with translation data
from IMO and TFO cells will determine subsets of messages that are regulated by HuR at the level of translation during RAS transformation. Moreover, integrating this data with DO-RIP-seq and APA data will create a comprehensive picture of regulation of subsets of mRNAs by HuR important in RAS transformation.

Another interesting finding from our work was that HuR binding may antagonize miRNA-mediated suppression of mRNA abundance in TFO HMECs. HuR stabilized mRNAs just as well when binding near miRNA sites, suggesting that HuR’s stabilizing effect, when binding near miRNA seed matches, may be a result of blocking miRNA/RISC binding or function. Several studies reporting HuR’s antagonistic and synergistic effects on miRNA function for single mRNAs have been reported (Bhattacharyya et al., 2006; Filipowicz et al., 2008; Meisner and Filipowicz, 2011; Simone and Keene, 2013; Srikantan et al., 2011; Tominaga et al., 2011). Global evidence suggests that HuR can antagonize the destabilizing effect of miRNAs when bound at overlapping binding sites (Lu et al., 2014; Mukherjee et al., 2011; Nicholson et al., 2017). Our data corroborate these other global studies, and understanding mechanisms of competition with HuR and the RISC in the context of tumorigenesis are of future interest.

6.4 HuR as a therapeutic target

Activation of an EMT program is thought to directly contribute to chemoresistance through inducing changes associated with ‘stemness’ (i.e. cancer stem cells), and clinical studies have found strong correlations between EMT-associated gene
expression programs and chemoresistance (Fischer et al., 2015; Shibue and Weinberg, 2017). HuR has also been implicated in drug resistance to a number of chemotherapeutics. For example, two recent studies in pancreatic cancer demonstrated that HuR plays a role in cancer cell survival and gemcitabine resistance (Jakstaite et al., 2015; Zarei et al., 2017). HuR has also been implicated in resistance to doxycycline (Filippova et al., 2011), paclitaxel (Janakiraman et al., 2017), tamoxifen (Hostetter et al., 2008), and doxorubicin (Latorre et al., 2014; Latorre et al., 2012). Our finding that HuR is necessary for maintenance of the full EMT program induced by oncogenic RAS brings up the intriguing possibility that HuR could be involved in EMT-dependent drug resistance mechanisms.

In conventional cancer therapies, only one gene or pathway is targeted at a time. Targeting HuR, or other post-transcriptional regulators, would target multiple signaling pathways at one time, reducing the potential for cancer cells to develop resistance through exploiting redundancies in cellular signaling. Our data suggest that targeting HuR may result in the simultaneous loss of tumor initiating capacity and metastatic traits, while reducing potential therapy resistance. Future studies to identify drugs that target HuR are of interest.

High-throughput screens have identified agents against HuR, and a number of chemicals, including MS-444, dehydromutactin, okicenone and 15,16-dihydotanshinone, have been shown to disrupt HuR-mRNA interactions (D'Agostino et
Moreover, 15,16-dihydrotanshinone inhibited migration in MDA-MB-231 cells, which was rescued by HuR overexpression (D’Agostino et al., 2015). Therefore, these would be good candidates to test in RAS TFO HMECs.

Phosphorylation of HuR on serines 221 and 318 by protein kinase C delta was shown to be necessary for doxorubicin sensitivity (Latorre et al., 2014). HuR is phosphorylated by Chk2 (Abdelmohsen et al., 2007; Yu et al., 2011), Mapk14 (Lafarga et al., 2009), protein kinase C alpha (Doller et al., 2008; Doller et al., 2007; Doller et al., 2010; Doller et al., 2011), Tyrosine-protein kinase JAK3 (Yoon et al., 2014), and cyclin-dependent kinase 1 (Kim et al., 2008), and these post-translational modifications affect HuR localization as well as regulation of mRNA targets. For example, it was found that rottlerin, a protein kinase inhibitor, reduced HuR phosphorylation by protein kinase C alpha at serine 318, which resulted in reduced expression of HuR mRNA targets and impaired migration and invasion in colon cancer cells (Doller et al., 2011). Therefore, identifying post-translational modifications on HuR in the context of RAS-transformation will be important in order to determine other ways to target HuR.

6.5 Determining post-transcriptional regulation coordinating the immortalization of primary cells

One of the original aims of this work was to determine post-transcriptional regulatory events important for tumor initiation. As shown in chapter 3, gene expression changes for many RBPs, including those known to be involved in cancer, occurred
during the transition from a normal cell to an immortalized cell, with no significant changes in overall gene expression during the final stage of transformation. This suggests that post-transcriptional mechanisms, in cooperation with transcriptional reprogramming, may be responsible for the EMT-related transcriptional priming. Therefore, future studies will focus on understanding post-transcriptional mechanisms regulating this early transition. Doing so will first entail optimizing DO-RIP-seq to require less input material, since primary cells have proliferation limits. We have done some of this optimization and obtained one replicate of HuR and CELF1 DO-RIP-seqs in primary cells. This preliminary DO-RIP-seq data for both HuR and CELF1 suggest that both HuR and CELF1 mRNPs in normal cells are drastically different from those in immortalized and primary cells. However, these differences could be driven by expression differences between the cell lines. Therefore, primary DO-RIP-seq replicates will be normalized to total mRNA abundance and integrated with other data sets, such as 4SU-seq to measure global mRNA synthesis and decay rates (Dolken et al., 2008) and ribosome profiling data to measure global translation (Reid et al., 2015). Integrating quantitative binding site data with these functional data sets will distinguish between regulated and non-regulated binding sites. It is also possible that the expression changes in RBPss during immortalization are part of the transcriptome priming, and that post-transcriptional regulation is not dynamic during immortalization, but rather the cells are primed for post-transcriptional mechanisms.
Taken together, our work has begun to elucidate the complexities of post-transcriptional regulation during cancer progression and EMT transitions. Our work suggests many more experiments to further distinguish precise mechanisms of regulation of gene expression during tumorigenesis, and could lead to the discovery of novel therapeutics to target malignancy and EMT.
References


174
mRNA by HuR facilitates DNA repair and resistance to PARP inhibitors. Cancer research.


protein HuR impairs malignant characteristics of pancreatic ductal adenocarcinoma cells. Oncotarget 6, 27312-27331.


elements mediate mRNA decay by binding to CUG-binding protein 1. Mol Cell 29, 263-270.


Biography

Laura Simone Bisogno was born October 14, 1987 in Red Bank, NJ. Laura graduated from Montclair State University summa cum laude in May 2010 with a Bachelor of Science degree in Molecular Biology.

Publications:


Bisogno, L.S., Friedersdorf, M.B., Keene, J.D. (2017) Oncogenic Ras transformation drives malignancy and invasion post-transcriptionally. Manuscript Submitted for Publication

Awards:

Marcy Speer Memorial Fellowship, 2011-2012

Ruth L. Kirschstein National Research Service Award (F31), National Institutes of Health, National Cancer Institute, 2014-2017