Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

2016

Investigating the Role of Muscleblind-Like 1 in the Suppression of Breast Cancer Progression

Lisa Fish

INVESTIGATING THE ROLE OF MUSCLEBLIND-LIKE 1 IN THE
SUPPRESSION OF BREAST CANCER PROGRESSION

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Lisa Fish
June 2016

INVESTIGATING THE ROLE OF MUSCLEBLIND-LIKE 1 IN THE SUPPRESSION OF BREAST CANCER PROGRESSION

Lisa Fish, Ph.D.
The Rockefeller University 2016

Breast cancer is a prevalent disease. Metastatic disease accounts for the majority of deaths from breast cancer, as patients with distant metastatic disease have a much worse prognosis than those with localized disease. In order to better understand why some breast cancers metastasize and others do not, it is critical to identify and elucidate the determinants of breast cancer progression. Post-transcriptional control of gene expression plays a central role in modulating transcriptional output. The interactions between messenger RNA cis-regulatory elements and trans-factors control coordinated gene expression states. Post-transcriptional regulatory programs that enhance metastatic capacity are selected for during cancer progression. In this study, the RNA binding protein Muscleblind-like 1 (MBNL1) is identified as a novel suppressor of breast cancer metastasis. MBNL1 loss-of-function contributes to the pathogenesis of myotonic dystrophy, a human genetic disease, but has no reported role in tumorigenesis or cancer progression. In this study, MBNL1 is identified as a suppressor of breast cancer metastasis. MBNL1 was found to suppress metastasis of human breast cancer cells in a xenograft mouse model. Additionally, MBNL1 transcript levels are significantly correlated with metastasis-free survival of breast cancer patients. MBNL1 depletion was also found to enhance the invasion and trans-endothelial migration capacity of breast cancer cells. Identification of endogenous MBNL1 protein-RNA interactions in breast cancer cells was carried out using HITS-CLIP. Transcriptome-wide analysis of MBNL1-dependent transcript stability and

MBNL1 HITS-CLIP data revealed that globally, transcripts directly bound by MBNL1 are stabilized by MBNL1. Two transcripts, DBNL and TACC1, were identified as transcripts that were bound by MBNL1 and also destabilized upon MBNL1 depletion. Both DBNL and TACC1, when overexpressed in breast cancer cells depleted of MBNL1, were found to reverse the pro-invasive and metastatic colonization phenotypes observed upon MBNL1 depletion. Therefore, DBNL and TACC1 were identified as modulators of the metastasis suppressive effect of MBNL1 in breast cancer.

ACKNOWLEDGEMENTS

I also would like to thank my family and friends, who have always been a constant source of encouragement and support during grad school.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1. Breast cancer incidence, classification and clinical outcomes

Breast cancer is a common disease. Worldwide, there were an estimated 1,671,000 new cases diagnosed and 522,000 deaths from breast cancer in 2012 (Ferlay et al. 2012). In the United States, breast cancer is the most commonly diagnosed cancer. There will be an estimated 231,840 new breast cancer cases diagnosed and an estimated 40,290 deaths from breast cancer in 2015 (Table 1.1). The lifetime risk of being diagnosed with breast cancer for a woman in the US is 12.3% (Howlader et al. 2015).

Table 1.1 Incidence of common cancers in the United States for 2015

Cancer Type	New Cases	Deaths
Breast	231,840	40,290
Lung	221,200	158,040
Prostate	220,800	27,540
Colorectal	132,700	49,700
Bladder	74,000	16,000

Clinically, breast cancer can be classified by stage, grade, hormone receptor/HER2 status, and gene expression signature. Stage is defined by the size of the primary tumor and the degree to which it has spread to other sites in the body. The most commonly used staging system is called TNM (Tumor, regional lymph Node, Metastasis). This system assesses primary tumor size, degree of regional lymph node involvement and presence of cancer metastasis to distant organs in order to determine the cancer stage. Stages range from I to IV. Stage I breast cancers are small, have not detectably spread to regional lymph

nodes and have not metastasized, and stage IV cancers are those that have metastasized to distant organs. Stages II and III have intermediate levels of cancer size and spread (Sobin et al. 2010). Grade refers to the appearance of the primary tumor tissue and assesses how similar to normal breast tissue the tumor tissue appears. In the commonly used Nottingham grading system, each of the following factors contributes to the final assigned tumor grade: 1) The percent of the tumor tissue retaining normal milk duct structures. 2) The size and shape of the nucleus in the cancer cells. 3) The number of actively dividing cells. Hormone receptor/HER2 status refers to the expression of the estrogen receptor (ER), progesterone receptor (PR) and of the erb-b2 receptor tyrosine kinase 2 (ERBB2, also known as HER2/neu) by breast cancer cells. The expression of ER, PR and HER2 is typically assessed by immunohistochemistry, and the cancer is then designated as being positive or negative for expression of each of these factors. The expression of these factors are important for deciding the treatment of a breast cancer, as the effectiveness of drugs commonly used to treat breast cancer is largely determined by cancer cell ER, PR and HER2 expression. Breast cancers can also be categorized by their gene expression signature, also known as their intrinsic molecular subtype. Transcriptomic profiling of large sets of breast cancers has shown that breast cancers can also be divided into groups based on gene expression signatures (Bertucci et al. 2000; Perou et al. 2000; Sorlie et al. 2001; Sorlie et al. 2003). The molecular subtypes of breast cancer based on gene expression profiling have been

identified as luminal A, luminal B, HER2-enriched, basal-like, normal-like and claudin-low, the last two of which are rare. The luminal and basal subgroups are so named because their patterns of gene expression are similar to that of normal breast luminal and basal tissues. Classification of breast cancers by gene expression subtype, in addition to classification by stage, grade and receptor/HER2 status, can add information that is useful for prognosis and treatment (Kittaneh et al. 2013). Each of these criteria for classifying breast cancer is correlated with survival outcomes, but the stage and receptor status are currently the primary determinants of which therapies are used for treatment. However, gene expression signature information is increasingly being used to develop and test new targeted therapies (Eroles et al. 2012).

Depending on how it is classified using the systems described above, a breast cancer may be treated with surgery, radiation and non-specific cytotoxic chemotherapies or anti-angiogenic therapies. However, a primary consideration for breast cancer treatment is the receptor status of the cancer, as cancers that are ER/PR positive, and/or HER2 positive are treated differently from triple negative/basal-like cancers. This is because specific targeted therapies exist for hormone receptor positive and HER2 positive breast cancers, and these treatments confer a survival advantage over treatment with non-specific therapies. There is a lack of targeted therapies for triple negative/basal-like breast cancers, as these cancers do not express the factors that are targeted by these specific drugs. The established treatment for hormone receptor positive

breast cancers is administration of drugs aimed at reducing estrogen signaling, thereby reducing tumor growth. These drugs include tamoxifen, a small molecule that binds to ER and inhibits estrogen signaling in breast tissue. In post-menopausal women, aromatase inhibitors, including the small molecules anastrozole, exemestane and letrozole, are used to inhibit the actions of aromatase. The enzymatic activity of aromatase converts androgens into estrogen, and anti-aromatase therapy reduces systemic estrogen levels (Miller et al. 2014). Additionally, in pre-menopausal women, drugs can be used to suppress ovarian estrogen production in order to reduce systemic estrogen levels, although the effectiveness of this therapy in addition to tamoxifen therapy is not dramatic (Francis et al. 2015). The established therapy for HER2 positive breast cancers is trastuzaumab (Herceptin), an antibody directed against the extracellular domain of HER2. Trastuzaumab is effective in improving the survival of patients with HER2 expressing breast cancers (Piccart-Gebhart et al. 2005; Smith et al. 2007). However, breast cancers that have metastasized to distant organs remain difficult to treat, and women with breast cancer that has not spread from the primary site, or has spread regionally have a much higher 5-year survival rate compared to those who have cancer that has spread to distant organs (Figure 1.1A). Therefore, the majority of deaths from breast cancer are due to metastatic disease.
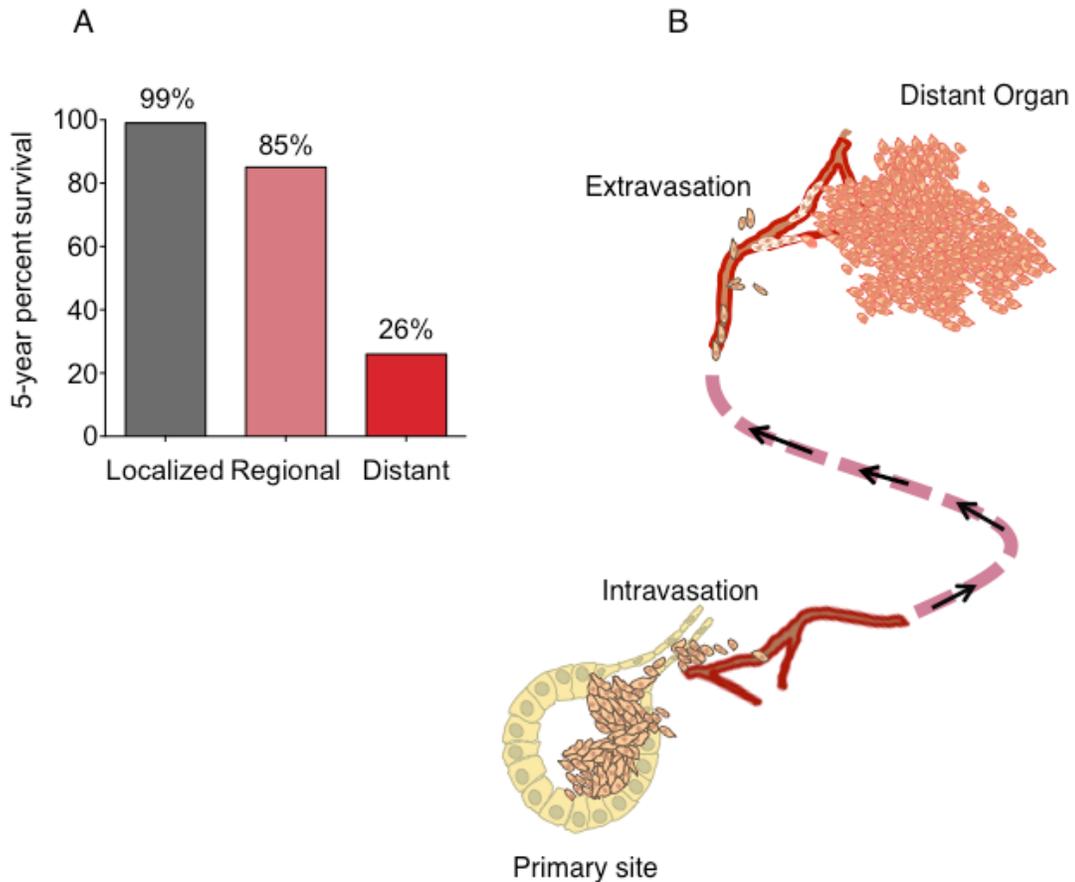
Figure 1.1 Breast cancer survival and metastatic progression
(A) The 5-year survival rate of breast cancer patients in the United States,
divided by level of spread of the cancer. (B) Illustration depicting the metastatic
cascade.

2. Cancer metastasis

From the above data, it is clear that the survival rate of breast cancer
patients is dependent on the extent of cancer progression (Figure 1.1A). Because
of this, it is important to understand the mechanisms by which cancer cells can
successfully spread through the body and colonize distant organs, a process
called metastasis. In this process, a cancer cell from the primary tumor moves

away from the primary tumor and into the surrounding tissue, intravasates into blood vessels, survives in circulation, extravasates out of blood vessels, and successfully survives and proliferates in a secondary organ (Chiang and Massague 2008) (Figure 1.1B). It is unlikely that a cancer cell from the primary tumor has the capacity to carry out all of these steps. Indeed, very few cancer cells from a primary tumor make it through all of the steps of the metastatic process. In animal models of metastasis, less than 0.01% of cancer cells in circulation result in macroscopic metastases (Chambers et al. 2002; Luzzi et al. 1998). In addition to the qualities intrinsic to the cancer cell, both the microenvironment of the primary tumor and the microenvironment of the distal organ to which the cancer has metastasized play roles in modulating the likelihood of metastasis (Joyce and Pollard 2009). Therefore, because of the complex nature and clinical importance of metastasis, it is of great interest to identify molecules that contribute to this process.

The first step of metastasis is cancer cell movement into blood vessels, a process called intravasation. Cancer cells may intravasate into vessels that are in direct contact with the tumor, or the cancer cell may first invade into the surrounding tissue and then intravasate into blood vessels. Cancer cell invasion can be promoted by local environmental stresses such as hypoxia, pH level and nutrient deprivation as well as the presence of various stromal cell types and the composition of the extracellular matrix (Chang and Erler 2014; Quail and Joyce 2013). Cancer cells invade by moving through the extracellular matrix and normal

6

tissue that surrounds the primary tumor. These cancer cells can then intravasate into systemic circulation by moving through the layers of endothelial cells that comprise the blood vessel walls. Normally, endothelial cells act as a barrier to the movement of cells in to or out of the bloodstream, and therefore cancer cell movement across the endothelial cell layer is an aberrant occurrence. Cancer cells may intravasate through active and passive mechanisms. Tumor vasculature is abnormal, with leaky endothelial cell junctions and abnormal coverage by pericytes, features that allow for easier cancer cell intravasation (Dudley 2012; Xian et al. 2006). Cancer cells may also actively intravasate into vessels through either paracellular (movement through the junctions between endothelial cells), or transcellular (movement directly through the body of an endothelial cell) mechanisms. Many molecules that enhance invasion generally also enhance intravasation. Molecules expressed by breast cancer cells that are important for paracellular intravasation include N-WASP, which reorganizes the actin cytoskeleton by promoting invadopodia formation, and also enhances invasion and intravasation (Gligorijevic et al. 2012). Increased cancer cell expression of the matrix metalloproteinase MT4-MMP disrupts blood vessel integrity in the tumor environment, which allows for increased cancer cell intravasation (Chabottaux et al. 2009). Another metalloproteinase, ADAM12, is expressed in breast cancer tumor vasculature but not in the vasculature of normal breast tissue. The ectodomain shedding of endothelial cell vascular endothelial cadherin and of the angiopoietin 1 receptor TIE2 is mediated by

ADAM12, and could play a role in promoting cancer cell intravasation by disrupting endothelial cell junctions (Frohlich et al. 2013). Microenvirnomental factors can also affect intravasation: macrophages can enhance breast cancer cell intravasation by inducing cancer cell expression of Mena (an epidermal growth factor-responsive cell migration protein) through colony-stimulating factor-1 receptor signaling (Keirsse et al. 2014). Transcellular intravasation is another method breast cancer cells can use to move into vessels, and a role for endothelial cell myosin light chain kinase has been demonstrated in this process (Khuon et al. 2010). Once successfully intravasated, cancer cells in circulation encounter and must survive multiple stresses, including the mechanical shear stress of the bloodstream, detachment from the ECM, and destruction by immune cells.

Cancer cells then move out of circulation in a process called extravasation. During extravasation, cancer cell must again traverse the endothelial layer of blood vessels. The first step of extravasation is the adhesion of cancer cells to the endothelial cells of a blood vessel. This adhesion can occur after cancer cell arrest in capillaries due to size constraints, but can also involve cancer cell rolling on the vessel wall before firm adhesion (Geng et al. 2012; Kienast et al. 2010; Stoletov et al. 2010). Adhesion involves a large number of factors on both the cancer and endothelial cells. Initial attachment can occur through the interactions of cancer cell expressed N-cadherin, Sialyl Lewis A and/or CD44 with endothelial cell expressed N-cadherin and/or E-selectin.

Stabilization of cancer cell-endothelial cell interaction further depends on integrins, CD44 and/or Mucin 1 (Reymond et al. 2013). The cancer cell then migrates through the endothelial cell layer and into the surrounding tissue. Once the cancer cell has extravasated and is located in a distant organ, the environment of that organ and the intrinsic programing of the cancer cell both contribute to the suppression or promotion of cancer cell survival, proliferation, and ultimately formation of metastatic nodules in that distal organ.

3. Post-transcriptional regulation of gene expression

For metastasis to occur, cancer cells must acquire cellular phenotypes that enable their migration into distant organs, where their subsequent survival and proliferation lead to end-organ failure. The enhanced invasiveness and migratory capacity necessary for metastasis can be generated through the coordinated expression of specific gene sets. Post-transcriptional regulation of gene expression is a method of doing this, and refers to processes that affect the expression of a messenger RNA (mRNA) after transcription. The post-transcriptional control of mRNA expression is determined by interactions between regulatory cis-elements and trans-factors. Therefore, regulated modifications to cis-elements and interacting trans-factors determine the identity, localization and expression of a transcript. These processes include alternative splicing, alternative polyadenylation site selection, RNA editing, post-transcriptional nucleotide modification, and non-templated polyadenylation of the 3' end of a mRNA (Li and Mason 2014; Liu et al. 2014; Norbury 2013). Alterations to

regulatory trans-factors can also affect post-transcriptional gene expression. The major types of trans-factors are RNA binding proteins and microRNAs, both of which carry out their regulatory roles through direct interactions with transcripts. Importantly, a single trans-factor can interact with and regulate the expression of multiple transcripts, thereby coordinating the expression of gene sets.

Pre-mRNA alternative splicing and alternative polyadenylation site selection

Pre-mRNA alternative splicing is a major method of transcriptome diversification. Although there are approximately 20,000 genes in the human genome, analysis of high-throughput RNA sequencing data has revealed that from 95-100% of human genes express alternatively spliced transcript variants (Pan et al. 2008; Wang et al. 2008). Interestingly, the amount of alternatively spliced transcripts increases with increasing organism complexity, and therefore the prevalence of alternative splicing in humans is thought to be necessary for diversifying the human proteome (Nilsen and Graveley 2010). Pre-mRNA alternative splicing is the process by which specific exons and introns are actively included or excluded from a transcript. Four major types of alternative splicing exist: exon inclusion, exon exclusion, mutual exon exclusion, and intron retention. Alternative splicing adds a regulatory layer to the basal splicing of all intron-containing transcripts. The basic mechanism of splicing, where introns are removed and exons are joined together, is controlled by cis-elements in the pre-mRNA. The conserved elements that are recognized by the spliceosome are

exon-intron junctions at the 5' and 3' ends of introns, the branch point sequence, located upstream of the 3' splice site, and the polypyrimidine tract, which is located between the branch point sequence and the 3' splice site. The 5' splice site has a conserved GU dinucleotide and the 3' splice site has a conserved AG dinucleotide. These elements are then bound in a defined order by the small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4/U6, and U5, which are central components of the spliceosome. The spliceosome is a large RNA-protein complex that catalyzes the removal of introns and joining of exons (Lee and Rio 2015). In eukaryotes, exon definition is the major mechanism used by the basal splicing machinery.

Alternative splicing is directed by cis-regulatory sequence elements and trans-factors. These regulatory elements are divided into four types: exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs). The best-studied classes of proteins that bind to these regulatory elements and influence exon choice are the ubiquitously expressed members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family (Huelga et al. 2012), and proteins in the serine-arginine rich (SR) protein family (Anko 2014). Another class of proteins that play roles in regulating alternative splicing are those that exhibit tissue-specific expression patterns and are involved in tissue differentiation. These include proteins from the NOVA (neuro-oncological ventral antigen)(Ule et al 2003; Zhang et al. 2010), RBFOX (Fox RNA-binding)(Weyn-Vanhentenryck et al. 2014;

Yeo et al. 2009; Zhang et al. 2008), CELF (CUG binding, Elav-like), and MBNL (Muscleblind-like) (Kalsotra et al. 2008; Ladd et al. 2001; Wang et al. 2012, Wang et al. 2015) families. These proteins control alternative splicing by binding in the upstream intron, downstream intron or in the regulated exon itself to control exon inclusion or skipping. Some of these factors act to recruit spliceosome machinery, while others act to block the actions of the spliceosome. These regulatory factors have also been shown to antagonize each other's functions (Zhu et al. 2001). It is thought that the developmental stage and tissue-specific expression patterns of these factors dictate their activity.

Although deadenylation and decapping accounts for bulk mRNA turnover, cis-elements and trans-factors allow for the modulation of transcript stability. Cis-elements that regulate mRNA decay are commonly found in 3'UTRs. Many such elements have been identified, and are usually defined by sequence. However, cis-elements defined by structure rather than sequence have been identified as modulators of mRNA decay (Goodarzi et al. 2012). Both sequence-specific and structural elements interact with trans-factors, which then can affect the stability of the transcript through recruitment or inhibition of mRNA decay machinery.

In addition to alternative splicing, pre-mRNAs are subject to alternative cleavage and polyadenylation (APA). APA produces transcripts with different 3' ends, which can affect post-transcriptional gene regulation. APA is a prevalent mode of regulation, as approximately 70% of human transcripts have alternative poly(A) containing products as assessed by RNA sequencing methods (Derti et

al. 2012). The location of many regulatory elements, 3'UTRs are enriched in sites recognized and bound by trans-factors. These interactions can control the expression of that transcript. Therefore, shortening of a 3'UTR through APA can allow escape from regulation by specific trans-factors. This type of dysregulation has been observed in cancers (Mayr and Bartel 2009). In a subset of breast cancers, among transcripts with significantly different 3'UTR lengths, shorter 3'UTRs were associated with shorter overall patient survival (Lembo et al. 2012).

The control of the length of the poly(A) tail is another important method the cell uses to regulate stability and maintain steady-state levels of transcripts. Normally, poly(A) tails of 100-250 nucleotides are added to the 3' end of each mRNA by the actions of poly(A) polymerase (PARP) and nuclear poly(A) binding protein (PABPN1). The presence of a poly(A) tail allows a transcript to be exported to the cytoplasm and engage with the translation machinery. Once in the cytoplasm, the length of a transcript's poly(A) tail is then controlled by the opposing actions of deadenylases and adenylases. Generally, the PAN2/3 complex (poly(A) nuclease 2/3) is the initial factor that trims the poly(A) tail of an mRNA, and its activity is stimulated by PABP (poly(A) binding protein). Further deadenylation is carried out by the CCR4-NOT complex, which can be inhibited or stimulated by the binding of transcript-specific RBPs. Once a transcript has been deadenylated to a certain point, RNA degradation occurs (Norbury 2013).

Molecular mechanisms of mRNA degradation

In eukaryotic cells, a combination of factors dictate the stability of messenger RNAs (mRNAs). Ribonucleases control mRNA degradation through defined pathways and trans-acting factors and cis-acting elements control the rate of mRNA decay. Two critical cis-acting determinants of mRNA stability are incorporated co-transcriptionally: the 7-methyl-guanosine cap (m^7G) which is added to the 5' end of the transcript, and the poly(A) tail which is added to the 3' end of the transcript. These two additions protect the transcript from immediate destruction by exonucleases, and their removal is crucial to normal turnover of mRNAs (Yamashita et al. 2005). Transcripts are constantly subjected to the actions of ribonucleases so as to maintain steady-state levels of mRNA in the cell. Transcripts can be further subject to regulated decay, both by cis-elements in their sequence and in response to various stimuli that regulate the level of trans-acting factors such as RBPs and miRNAs.

Messenger RNAs are degraded by the enzymatic actions of ribonucleases, and these RNA degradation enzymes carry out their actions in a specific order. Here, a general overview of the major factors in these pathways in eukaryotic cells is given. The primary pathway of mRNA decay is the deadenylation-dependent pathway. In this pathway, the first step of mRNA degradation is removal of the poly(A) tail. This requires the action of deadenylases, which are 3'-5' exoribonucleases with a high preference for a poly(A) substrate. Deadenylases and poly(A) polymerases dynamically control poly(A) tail length in the cytoplasm, and poly(A) tail length is a major determinant

14

of mRNA stability. Generally, the poly(A) tail of a transcript is bound by PABP

(poly(A) binding protein) in the cytoplasm, and PABP interacts with eIF4G, a

component of the eIF4F complex which binds the m^{7}G cap at the 5' end of the

transcript. This interaction promotes the recruitment of the translation initiation

complex as well as protects the transcript from degradation. Once deadenylases

have shortened the poly(A) tail below a threshold length, PABP no longer

interacts with the tail and the transcript is left unprotected from further actions of

ribonucleases (Kim et al. 2006; Tharun and Parker 2001; Wilusz et al. 2001). The

major deadenylase containing complexes that are active in metazoans are

CCR4-CAF1-NOT, PAN2-PAN3 and PARN (Yan 2014). After removal of the

poly(A) tail by the actions of a deadenylase complex, the remaining RNA can be

degraded in either a 3'-5' or 5'-3' direction. 3'-5' degradation occurs through the

action of the exosome, a large multi-subunit complex with exonuclease activity

(Houseley and Tollervey 2009). The m^{7}G cap is then removed by DcpS

(scavenger decapping enzyme). Messenger RNA can also be degraded in the 5'-

3' direction after deadenylation. The first step in this process is removal of the

m^{7}G cap by the decapping complex, which requires the enzymatic action of

DCP2 (van Dijk et al. 2002; Wang et al. 2002). The decapping complex also

contains proteins that enhance decapping. In metazoans, these cofactors are

DCP1, EDC3, EDC4, PAT1B, DDX6, and the LSM1-7 complex. These factors are

not all required for decapping of all mRNAs, but can act in a transcript specific

manner (Arribas-Layton et al. 2013; Fenger-Gron et al. 2005, Ling et al. 2011). After decapping, the remaining RNA is degraded 5'-3' by the exonuclease XRN1.

Messenger RNA may also be degraded in a deadenylation-independent manner. However, this pathway is thought to play a very minor role in bulk mRNA decay in eukaryotic cells. The existence of this pathway in yeast has been demonstrated by studying the regulated decay of the RPS28B transcript. Initiation of this pathway requires the recruitment of the decapping complex component EDC3 to the transcript, which in turn recruits the decapping enzyme DCP2. After removal of the m^7G cap, the RNA is degraded 5'-3' by XRN1 (Badis et al. 2004).

An additional type of mRNA decay utilizes endoribonucleases. This method of mRNA degradation differs from the pathways described above because removal of the poly(A) tail or the 5' cap is not necessary for initiation of decay. Instead, endonucleases act on a capped, polyadenylated mRNA and cleave it. The resulting RNA pieces are then destroyed by the actions of the exoribonucleases XRN1 and the exosome. Instances of endoribonuclease-mediated mRNA decay in metazoans have been reported, but have not been heavily studied. Generally, these endonucleases have highly specific transcript targets, their action is activated by external stimuli, their actions can be localized, and they cleave at specific sequences. This is in contrast to the deadenylation and decapping that facilitates constitutive bulk mRNA turnover. The best studied of these endonucleases is PMR1, which was discovered in *Xenopus* extracts as

the endonuclease responsible for deadenylation independent decay of the albumin transcript in response to estrogen (Yang and Schoenberg 2004). In human cells, the PMR1 endonuclease is associated with polysomes, and localization of PMR1 to the leading edge of MCF7 breast cancer cells enhances cell motility in a scratch assay (Gu et al. 2012). Other proteins that have demonstrated endonuclease activity on specific mRNA substrates in metazoans are IRE1/ERN1 (endoplasmic reticulum to nucleus signaling 1), ZC3H12A (zinc finger CCCH-type containing 12A), SMG6 (SMG6 nonsense mediated mRNA decay factor), and APE1 (apurinic/apyrimidinic DNA endonuclease 1)(Schoenberg 2011). SMG6 acts in the nonsense mediated decay pathway (Eberle et al. 2009), and therefore typically acts as a quality control factor rather than a regulator of the decay of specific mRNAs in response to stimulus.

mRNA stability regulatory pathways

Transcripts that do not contain open reading frames (ORFs) are subject to active degradation by the cell to ensure the production of functional proteins. The recognition of transcripts without ORFs occurs through multiple mechanisms, depending on the problem with the transcript. Together, these are referred to as surveillance pathways. These are 1) nonsense mediated decay (NMD), which occurs when an mRNA has a premature termination codon (Kervestin and Jacobson 2012). 2) nonstop mediated decay (NSD), which happens when an mRNA does not contain an in-frame termination codon (Frischmeyer et al. 2002), and 3) No-go decay (NGD), which takes place when ribosomes stall at the 5' end

of an mRNA (Doma and Parker 2006). These pathways of decay are triggered by problems in the processing or sequence of the transcript. Interestingly, NMD can also participate in the regulated decay of transcripts in a process called alternative splicing coupled NMD (AS-NMD). AS-NMD acts through selective degradation of mis-spliced mRNAs. Types of alternative splicing that trigger NMD are retention of a premature termination codon-containing exon or intron, splicing that leads to an elongated 3'UTR, and splicing that results in the presence of an upstream open-reading frame (Sibley et al. 2014). Alternative splicing can also lead to intron retention. In addition to triggering NMD in the cytoplasm, this can lead to a block on the export of a transcript to the cytoplasm, and thereby induce transcript degradation by the nuclear exosome (Ge and Porse 2013).

Cytoplasmic RNA granules and mRNA stability

The subcellular localization of a transcript also contributes to its stability. Processing bodies (P-bodies) and stress granules are two major types of cytoplasmic RNA granules that are sites of regulated mRNA decay and storage. Stress granules form in response to various stresses. These stress stimuli include oxidative stress, hypoxia, amino acid starvation, endoplasmic reticulum stress, heat shock, UV irradiation and viral infection. These different stimuli each induce the activity of specific kinases that then phosphorylate eIF2-alpha (eukaryotic initiation factor 2) (Kedersha et al. 1999). This phosphorylation event halts translation initiation, but does not affect actively elongating ribosomes. Stress granules contain translation initiation factors, 40S ribosomal subunits and

18

mRNAs bound to the translation pre-initiation machinery. A large number of other RBPs have also been found in stress granules, and the specific composition of a given stress granule is dependent on the identity of the cell and the type of stress. Another characteristic of mRNAs in stress granules is that they retain their poly(A) tails and remain in a closed loop formation with the cap complex (Anderson and Kedersha 2008). Components of stress granules disperse once the stress is removed, and translation of the mRNAs can resume. Stress granule formation may play a role in cancer progression. In a mouse xenograft model of osteosarcoma metastasis to the lung, depletion of the stress granule nucleating factor G3BP1 reduced both number of stress granules per cancer cell and reduced the number of metastatic nodules in the lung (Somasekharan et al. 2015). Although this is a correlative finding, it is consistent with stress granule formation promoting metastasis.

A second major type of cytoplasmic RNA granule is processing bodies, which have distinct components from stress granules. However, P-bodies and stress granules also have many overlapping components and functions, and these protein components, as well as RNA transcripts, can move dynamically between the two granule types (Kedersha et al. 2005). P-body components include mRNA degradation, mRNA surveillance and miRNA-mediated gene silencing factors. The decapping enzyme DCP2 is a characteristic component of these foci. Unlike stress granules, P-bodies are constitutively present. Although they contain mRNA decay machinery, and are thought to be the sites of mRNA

decay, the presence of P-bodies is not required for global or specialized RNA

decay to occur (Eulalio et al. 2007; Stoecklin et al. 2006).

4. RNA binding proteins, tumorigenesis and cancer progression

RNA interacting proteins are abundant, with ~550 mRNA-interacting

proteins common to both HeLa and HEK293 cells (Baltz et al. 2012; Castello et

al. 2012). Through their interactions with RNA, these proteins mediate a variety

of regulatory processes that have wide-ranging downstream effects in

development, homeostasis and pathological processes. The majority of RNA

binding proteins bind multiple RNA transcripts and/or other RNA molecules and

in doing so can impact the regulation of those RNAs. Given their ability to

coordinately regulate sets of transcripts, proteins belonging to this group have the

potential to confer attributes that affect the metastatic potential of cancer cells.

Indeed, recent studies have noted such effects on breast cancer progression for

a variety of RNA-binding proteins. The metastasis modulating effects of these

RNA-binding proteins have been shown to be generated through different post-

transcriptional regulatory steps, including direct modulation of transcript stability

by TARBP2 (Goodarzi et al. 2014) and RBM47 (Vanharanta et al. 2014), as well

as regulation of alternative splicing patterns by HNRNPM (Xu et al. 2014). The

role of RNA binding proteins in breast cancer tumorigenesis, in contrast to cancer

progression, has been more heavily studied. Among the best-studied RBPs in

breast tumorigenesis are ELAVL1 (embryonic lethal abnormal vision-like 1, also

known as HuR) and ZFP36 (zinc finger protein 36, also known as TTP). Both are

AU rich element (ARE) -binding proteins, and ELAVL1 is commonly upregulated in breast cancer, while ZFP36 is commonly downregulated (Sanduja et al. 2010; Abdelmohsen and Gorospe 2010).

5. Muscleblind-like family: Identification and structure

One RNA binding protein that has been implicated in human genetic disease and in the regulation of alternative splicing is Muscleblind-like 1 (MBNL1). The Muscleblind-like family of proteins is present in organisms ranging from worms to humans. Muscleblind-like proteins do not exist in bacteria, fungi or plants, and only occur in metazoans. Invertebrates have one muscleblind-like gene, while the majority of vertebrates have three muscleblind-like genes (Pascual et al. 2006). Members of this family contain highly conserved cysteine-cysteine-cysteine-histidine (CCCH) zinc finger domains with which they bind RNA. These CCCH zinc finger domains are common, and are found in over 1000 proteins. CCCH zinc finger domains bind specific sequence motifs in RNA. Consistent with this observation, SELEX studies have determined that MBNL1 binds the RNA motif YGCY (Goers et al. 2010), while CLIP-seq studies have identified an enrichment of UGC and GCU containing 4mers in MBNL1-bound sites in mouse myoblasts (Wang et al. 2012). Crystal structures of the MBNL1 CCCH zinc finger domains, both while bound to RNA with the sequence CGCUGU, and while not bound to RNA, have been solved (Teplova and Patel 2008). These structures show that the two pairs of zinc fingers have a symmetric fold where the two zinc fingers in each pair are connected by a linker that orients

their RNA binding surfaces away from each other, making the RNA fold into an antiparallel form. This structure as well as other *in vitro* biochemical binding data (Goers et al. 2010) suggests a model where the zinc fingers of MBNL1 interact with the 5'-GC steps in RNA that folds into hairpin structures with bulged and unpaired uridines.

MBNL proteins in development

Work on a *Drosophila melanogaster* mutant led to the first characterization of a muscleblind family member, with the observation that in *Drosophila*, muscleblind is required for normal development of photoreceptor cells as well as for normal muscle differentiation and attachment (Begemann et al. 1997; Artero et al. 1998). In mammals, the three members of the Muscleblind-like family are MBNL1, MBNL2 and MBNL3, and all contain four tandem CCCH zinc finger domains. These highly similar proteins exhibit tissue-specific expression patterns. MBNL1 is the most highly expressed member of the family, and is detected in adult skeletal muscle, heart, brain, intestine, kidney, liver, lung and placenta. Mice with genetic knockout of MBNL1 develop muscle weakness, cataracts and cardiac pathology (Dixon et al. 2015; Kanadia et al. 2003). Loss of MBNL1 function results in a shift to fetal and embryonic stem cell-like transcript splicing patterns, and also impairs erythroid terminal differentiation (Cheng et al. 2014; Han et al. 2013; Lin et al. 2006). MBNL2, like MBNL1, is detectable in most adult tissues, with the highest expression in brain, and has decreased expression in heart and skeletal muscle compared to MBNL1. Mice null for MBNL2 do not

display abnormal muscle development, but they exhibit sleep disturbances, impaired learning and are prone to seizures (Charizanis et al. 2012). Of the three MBNL members, MBNL3 has the lowest expression levels. MBNL3 is primarily expressed during embryonic development, but is also detectable in adult lung, spleen and testis. Interestingly MBNL3 expression is decreased in myoblasts when they are induced to differentiate, and overexpression of MBNL3 suppresses myoblast differentiation (Kanadia et al. 2003; Lee et al. 2007; Squillace et al. 2002). These expression patterns and developmental phenotypes are consistent with a model where MBNL1 and MBNL2 promote normal tissue differentiation while MBNL3 acts as a suppressor of differentiation. Together, these MBNL1-dependent developmental phenotypes and the lack of MBNL orthologs in single cell organisms suggest that MBNL family proteins function primarily to ensure correct differentiation of tissues.

MBNL molecular mechanisms

MBNL1 has a well-characterized role in the regulation of pre-mRNA splicing. The first pre-mRNAs found to have MBNL1-dependent alternative splicing were identified because these transcripts had been found to have altered splice forms in myotonic dystrophy, a disease in which MBNL1 has a pathogenic role, as discussed in the next section. These first identified alternatively spliced genes were TNNT2 (cardiac troponin T type 2) and INSR (insulin receptor) (Ho et al. 2004). In a study examining the mechanism of MBNL1-regulated alternative splicing of the TNNT2 pre-mRNA, MBNL1 was found to bind to a stem-loop

23

structure in the polypyrimidine tract of the intron upstream of the alternative exon

in TNNT2 pre-mRNA. MBNL1 binding was found to inhibit the formation of the

spliceosome complex by modulating the local RNA structure, which then

interfered with U2AF2 binding and led to exon skipping (Warf et al. 2009). More

recently, genome-wide sequencing and computational analyses have found

many more MBNL family protein dependent splicing events (Charizanis et al.

2012; Du et al. 2010; Wang et al. 2012). Mechanistically, MBNL protein splicing

function is dependent on where the protein binds relative to the alternative exon.

MBNL1 and MBNL2 binding on the alternative exon or within the upstream intron

usually leads to skipping of the alternative exon (Charizanis et al. 2012; Du et al.

2010; Wang et al. 2012). MBNL1 binding to the downstream intron promotes

retention of the alternative exon. In addition to the cis-acting sequences in a

transcript, trans-acting factors can also influence the binding of MBNL1 to RNA.

A study has shown that the DEAD-box RNA helicase DDX5 promotes MBNL1

binding to the stem-loop regulatory element in the TNNT2 pre-mRNA (Laurent et

al. 2012). Interestingly, another DEAD-box helicase, DDX6, has been found to

interact with and unwind CUG repeat expansion RNA hairpins, leading to a

decrease of MBNL1 binding to these RNA structures (Pettersson et al. 2014).

MBNL1 co-immunoprecipitates with several proteins in human myoblasts,

including hnRNP H,H2, H3, F, A2/B1, K, L, DDX5, DDX17 and DHX9, and these

may act in concert with MBNL1 to affect RNA splicing (Paul et al. 2011). More

definitively, CLIP-seq analysis of MBNL and CELF1 determined that these

proteins have opposing effects on the splicing of a large set of alternative exons (Wang et al. 2015).

MBNL1 has also been found to developmentally regulate alternative poly(A) site selection in mouse embryonic fibroblasts through directly binding to 3'UTRs (Batra et al. 2014). CLIP-seq analysis of MBNL1 and MBNL2 in mouse myoblasts in combination with cellular fractionation and transcriptomic profiling led to a proposed model wherein MBNL1 binding to 3'UTRs promotes localization and cytoskeletal trafficking of transcripts, which then promotes secretion of their protein products (Wang et al. 2012). Another study analyzed CLIP-seq of MBNL1 in mouse myoblasts and proposed that MBNL1 binding destabilizes transcripts (Masuda et al. 2012). Clearly, MBNL1 has multiple mechanisms through which it post-transcriptionally regulates gene expression.

6. MBNL and human genetic diseases

Myotonic dystrophy

To date, most studies on MBNL family proteins have investigated the mechanism through which they contribute to the pathology of myotonic dystrophy, a human genetic disease. Myotonic dystrophy is the most common muscular dystrophy that affects adults, with an estimated global incidence of 1 case per 8,000 people. People affected with this disease exhibit muscle weakness and atrophy, cardiac conduction defects, cataracts, diabetes, male frontal balding, various behavioral changes and hypersomnia (Udd and Krahe 2012). Two forms of the disease exist, myotonic dystrophy type 1 (MD1) and

myotonic dystrophy type 2 (MD2), which are classified according to the genic location of the causative mutation. In both types, the causative mutation is a nucleotide repeat expansion in a non-coding region of a gene. In type 1, this is a $(CTG)_n$ repeat expansion in the 3'UTR of the *DMPK* (Dystrophia Myotonica Protein Kinase) gene (Aslanidis et al. 1992; Brook et al. 1992; Harley et al. 1992; Mahadevan et al. 1992). In type 2 disease this is a $(CCTG)_n$ repeat expansion in the first intron of the *CNBP* (CCHC-type zinc finger, Nucleic acid Binding Protein) gene (Liquori et al. 2001).

The mechanism through which these repeat expansions lead to the pathology associated with the disease has been intensely studied. The prevailing model for how these repeat expansions cause the pathologic features of myotonic dystrophy is an RNA gain-of-function mechanism (Figure 1.2A). In this model, transcripts containing these repeat expansions accumulate in the nucleus, where they bind proteins, effectively sequestering these proteins and preventing their normal actions in the cell. Several lines of evidence point to this RNA gain-of-function model being a major contributor to myotonic dystrophy pathogenesis. Mice null for *DMPK* exhibit late onset myopathy, but do not display myotonia and cataracts, two phenotypes characteristic of myotonic dystrophy (Reddy et al. 1996). Also, although *SIX5 (*Sine oculis-related homeo box 5), the gene located immediately downstream of *DMPK,* has reduced expression levels in MD1 due to
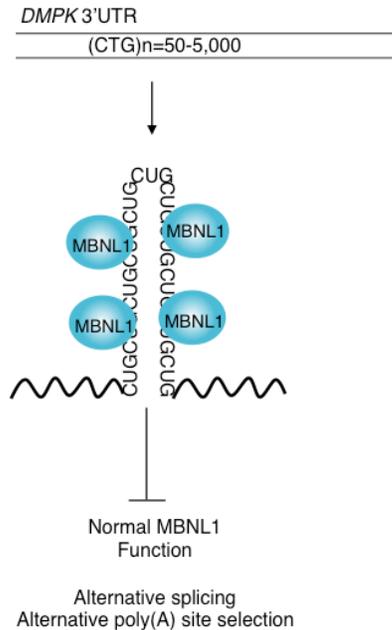
Figure 1.2. RNA gain-of-function model for myotonic dystrophy 1 pathogenesis
Illustration depicting the toxic gain-of function RNA model for the pathogenesis of
myotonic dystrophy 1. An expansion of the number of CTG trinucleotide repeats
in the 3'UTR of the *DMPK* gene, which normally contains 5-37 CTG repeats,
leads to the accumulation of DMPK transcripts with CUG trinucleotide repeat
expansions. This RNA then has a gain-of-function by binding and sequestering
MBNL1, preventing the normal functions of MBNL1 in the cell.

the expanded $(CTG)_n$ repeat in the 3'UTR of *DMPK* (Klesert et al. 1997), mice

null or heterozygous for *SIX5* develop cataracts but no abnormal muscle

phenotype (Klesert et al. 2000; Sarkar et al. 2000), again, suggesting that SIX5

levels are not sufficient to phenocopy the symptoms of myotonic dystrophy. The

evidence in favor of the RNA gain-of-function model includes a characteristic

accumulation of nuclear foci in muscle samples from individuals affected with

myotonic dystrophy, and that these nuclear foci contain the expanded repeat

transcripts (Liquori et al. 2001; Mankodi et al. 2001; Taneja et al. 1995). The

identification of MBNL proteins as $(CUG)_n$ repeat binding proteins (Miller et al.

2000) and the discovery that MBNL proteins colocalize with the $(CUG)_n$

expanded repeat-containing RNA in nuclear foci further supports this model

(Fardaei et al. 2001; Mankodi et al. 2001). Another critical finding in support of

this model was that *MBNL1* null mice recapitulate many of the phenotypes

characteristic of myotonic dystrophy, including myotonia and cataract

development (Kanadia et al. 2003).

It is thought that the pathogenesis of myotonic dystrophy is caused by the

dysregulation of the alternative splicing of a set of transcripts normally regulated

by MBNL proteins. In favor of this spliceopathy model, a study compared global

splice forms in *MBNL1* null mice with those in a mouse model of MD1, where

animals were engineered to express a transgene consisting of 250 CUG repeats

in the human skeletal muscle actin gene 3'UTR (Mankodi et al. 2000). This

showed a ~80% overlap in splice forms in the two mice (Du et al. 2010). Taken

together, the evidence supports a model for myotonic dystrophy pathogenesis

where expanded repeats in non-coding genic regions result in the accumulation

of "toxic" repeat-containing RNAs that bind MBNL proteins and prevent them

from carrying out their usual functions.

Spinocerebellar ataxia 8

Although their role in myotonic dystrophy is the most heavily studied, there

is evidence that MBNL1 functions in the pathogenesis of another human genetic

disease, spinocerebellar ataxia 8 (SCA8). SCA8 is a slowly progressing

neurodegenerative disease that affects the cerebellum. The causative SCA8

mutation is a CAG CTG repeat expansion in the overlapping genes *ATXN8OS*

and *ATXN8,* which are located on opposite DNA strands (Koob et al. 1999). This

expansion results in an expanded run of glutamines in the protein product of

ATXN8, and a CUG trinucleotide repeat expansion in the non-coding ATXN8OS

transcript (Moseley et al. 2006). In tissue samples from both individuals affected

with SCA8 and from mice engineered to express the ATXN8OS transcript

containing expanded CUG repeats, ATXN8OS transcripts with expanded CUG

repeats were observed in ribonuclear inclusions. Interestingly, MBNL1 was also

localized in these inclusions in specifically the molecular layer interneurons and

deep cerebellar nuclei (Daughters et al. 2009).

Fragile X-associated tremor/ataxia syndrome

MBNL1 has been found in nuclear inclusions specific to Fragile X-

associated tremor/ataxia syndrome (FXTAS), which is another late onset

neurodegenerative disease. FXTAS affects the cerebellum and white matter of

the brain. In males, there is a 1 in 3,000 lifetime risk of developing FXTAS,

although the incidence of the causative mutation is higher (Jacquemont et al.

2004). The causative mutation is an expanded CGG repeat in the 5'UTR of the

FMR1 gene. When this trinucleotide is repeated more than 200 times, it leads to

loss of *FMR1* expression and to fragile X syndrome, a more severe condition

than FXTAS. However, when the CGG trinucleotide is repeated 55 to 200 times

(normal repeat number is 5 to 40), it can lead to FXTAS. This intermediate

number of CGG repeats leads to an increase in the level of FMR1 mRNA,

although there are normal to reduced levels of the FMR1 protein (Hagerman et

al. 2001; Hagerman and Hagerman 2015). A characteristic of FXTAS is the

presence of ubiquitin positive nuclear inclusions in neurons and astrocytes. In

samples from FXTAS patients, these inclusions were isolated and mass

spectrometry analysis found, among other proteins, the presence of MBNL1 and

the RNA binding protein HNRNPA2/B1 (Iwahashi et al. 2006). In FXTAS patient

brain samples these repeat expansion containing FMR1 transcripts co-localized

with the RNA binding proteins SAM68, HNRNPG and MBNL1 in nuclear foci.

However, in this study MBNL1 retained some ability to regulate the splicing of

some of its targets (Sellier et al. 2010). Therefore, although the role of MBNL1 in

myotonic dystrophy has been the most extensively studied, MBNL1 may also

play a role in other human genetic diseases.

CHAPTER 2. MBNL1 and Breast Cancer Progression

1. Identification of MBNL1 as a suppressor of breast cancer progression

The effect of MBNL1 on breast cancer metastasis was originally examined

because MBNL1 was a candidate miR-335 target transcript. MiR-335 has been

identified as a suppressor of breast cancer metastasis (Tavazoie et al. 2008).

The MBNL1 transcript has a miR-335 complementary site in its 3'UTR, however,

overexpression of miR-335 in breast cancer cells did not modulate the level of

MBNL1 protein (Figure 2.1A). Therefore, the seed sequence complementarity of

miR-335 to the MBNL1 transcript is not sufficient to repress MBNL1 expression in these cells. Although MBNL1 levels were unaffected by miR-335 overexpression in breast cancer cells, the effect of MBNL1 on breast cancer metastasis was tested. To assay the effect of MBNL1, a xenograft mouse model was used. In this assay, RNAi-mediated depletion of MBNL1 using two independent short hairpin RNAs (shRNAs) robustly increased the metastatic colonization capacity of MDA-231 human breast cancer cells in tail-vein metastatic colonization assays (Figure 2.1B). Moreover, depletion of MBNL1 also significantly enhanced (~15-fold) metastatic colonization by highly metastatic LM2 cells, which is an *in vivo* selected sub-population of the MDA-231 cell line (Minn et al. 2005) (Figure 2.1C). To test if the effect of MBNL1 on metastasis was cell line specific, the effect of MBNL1 knockdown on the metastatic capacity of CN34 cells was tested. The CN34 cell line is an independent and minimally passaged human breast cancer cell line (Tavazoie et al. 2008). Again, MBNL1 knockdown in this independent cell line significantly increased (~100-fold) lung colonization by the CN34 breast cancer cells (Figure 2.2A). These findings reveal that endogenous MBNL1 can act as a suppressor of breast cancer metastatic colonization.
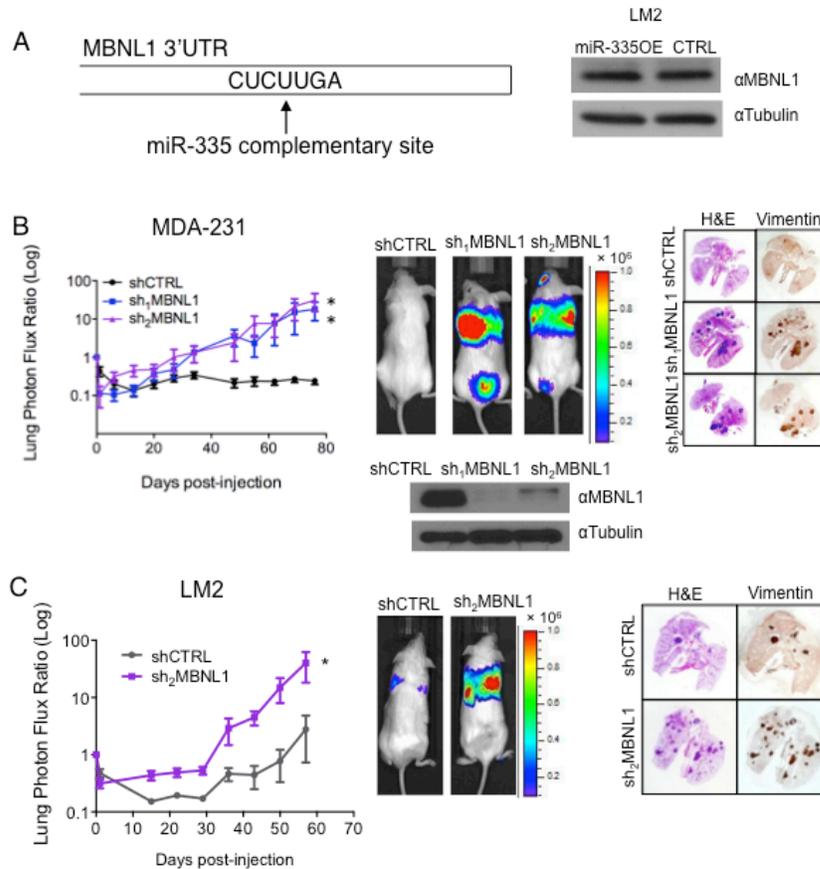
Figure 2.1 MBNL1 identification and effect on metastatic lung colonization
(A) Schematic depicting the miR-355 complementary site in the 3'UTR of the
MBNL1 transcript. Western blot for MBNL1 in whole cell lysate from LM2 breast
cancer cells overexpressing miR-335. Tubulin was used as a loading control. (B)
1×10^5 MDA-231 cells expressing either of two independent shRNAs targeting
MBNL1 or a control shRNA were intravenously injected into NOD-Scid mice.
Lung colonization was assessed by bioluminescence imaging and histology at
indicated timepoints. Representative H&E-stained lungs shown correspond to
day 83 post-injection. N=10-11. (C) 2×10^4 LM2 cells expressing an shRNA
targeting MBNL1 or a control shRNA were intravenously injected into NOD-Scid
mice and lung colonization was monitored by bioluminescence imaging.
Representative H&E and vimentin stained lungs correspond to day 64 post-
injection. N=5.

As injection of cancer cells into venous circulation primarily tests their capacity to colonize the lung, the impact of MBNL1 depletion on systemic metastasis to multiple organs was tested next. This was accomplished through intra-cardiac injection of cancer cells into the arterial systemic circulation of immunodeficient mice, which allows the cancer cells to pass through many organs before going through the lungs. This assay revealed a significant increase in the colonization of multiple distal organs, including brain, lung, and bone, by MBNL1 depleted cells relative to control cells (Figure 2.2C). This was interesting as these organs, as well as the liver, are the preferred sites of metastatic colonization of human breast cancers (Lee 1983). The effect of MBNL1 overexpression on lung metastatic colonization was also assayed. Overexpression of MBNL1 in highly metastatic LM2 cells significantly decreased their lung colonization capacity (Figure 2.2B). Together, these experiments show that MBNL1 is a robust suppressor of breast cancer metastasis to multiple distal organs.

Figure 2.2 MBNL1 suppresses breast cancer metastatic lung colonization

(A) 7.5×10^4 CN34 cells expressing either an shRNA targeting MBNL1 or a control shRNA were intravenously injected into NOD-Scid mice and lung colonization was monitored by bioluminescence imaging. Representative vimentin stained lungs shown correspond to day 78 post-injection. N=6. (B)1 x 10^5 LM2 cells stably expressing exogenous MBNL1 or a control vector were injected intravenously into NOD-Scid mice, lung colonization was monitored by bioluminescence imaging. Representative H&E and vimentin stained lungs shown correspond to day 37 post-injection N=10. (C) After intracardiac injection into athymic nude mice of 5×10^4 MDA-231 cells expressing a control hairpin or an shRNA targeting MBNL1, systemic metastasis was monitored over time by bioluminescence imaging. N=10. Bioluminescence signal quantification and gross bone histology corresponding to bone metastasis by 5×10^5 MDA control cells or MBNL1 knockdown cells 35 days after intracardiac injection of the cells. N=10.
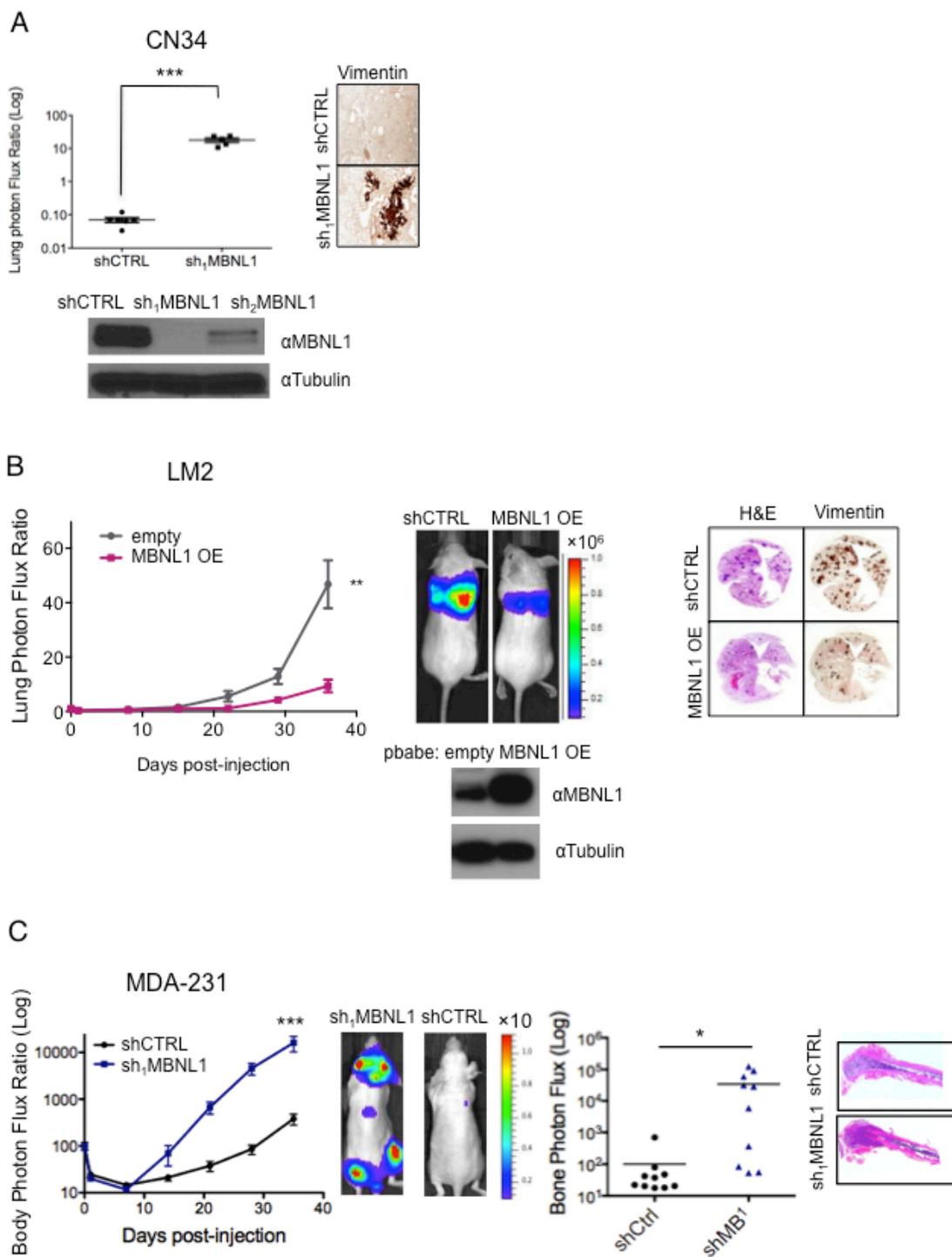
2. Expression and clinical correlation of MBNL in breast cancer

To test if an association exists between MBNL transcript levels and clinical outcome in breast cancer patients, clinical data and microarray analysis of primary tumors from breast cancer patients were analyzed. This analysis revealed that there was a significant correlation between reduced MBNL1 transcript levels in tumors and elevated metastatic relapse rates (Figure 2.3B). Furthermore, breast cancer metastases expressed significantly lower levels of MBNL1 transcript relative to primary tumors (Figure 2.3A). However, there was no significant correlation between MBNL2 or MBNL3 transcript levels and distant metastasis free survival (Figure 2.3 C,D) (Gyorffy et al. 2010). Together, these data reveal a clinical association between MBNL1 expression and metastasis-free survival in breast cancer patients. The two other human MBNL proteins, MBNL2 and MBNL3 do have high sequence similarity to MBNL1. However, MBNL1 is the most abundant MBNL transcript in breast cancer tumors, as assessed by RNA sequencing data from The Cancer Genome Atlas (TCGA) collection (Figure 2.3E). Therefore, it is reasonable that MBNL1 might play the dominant role in mediating MBNL family protein effects on breast cancer progression. Furthermore, in the cell lines used in this study, MBNL1 depletion by RNAi did not downregulate MBNL2 transcript levels (Figure 2.3F), therefore MBNL1 depletion alone is sufficient for the effects observed on metastasis suppression.

Figure 2.3 Expression and clinical correlation of MBNL in breast cancer

(A) Dot-plot representation of MBNL1 expression in a set of 117 primary breast cancers and 36 distal metastases from previously published datasets. Cancer samples were transcriptomically profiled using a common platform, and normalized to allow for intra and inter-cohort comparisons. N=153. (B-D) Kaplan-Meier survival curves depicting probability of distant metastasis free survival (DMFS) for breast cancer patients with tumors expressing high (red) or low (black) levels of each indicated MBNL family member transcript. High and low expression levels of each transcript were determined by dividing at the median. N=1609. P-value is based on the Mantel-Cox log-rank test. (E) MBNL family RNA levels (RPKM) in a set of breast tumors from RNA sequencing data (TCGA). N=1098. (F) Relative MBNL1 and MBNL2 transcript levels in MDA-231 shCTRL and shMBNL1 cells as assessed by qRT-PCR. HPRT1 was used as an endogenous control. P-values based on a two-way Student's t-test. Data are shown as mean ±SEM .

A

P < 0.0001

MBNL1 Expression

Primaries   Metastases

B

MBNL1

DMFS

high expression
low expression

N = 1609
P = 0.024

time (months)

C

MBNL2

DMFS

high expression
low expression

N = 1609
P = 0.24

time (months)

D

MBNL3

DMFS

high expression
low expression

N = 1609
P = 0.20

time (months)

E

relative RNA expression

MBNL1   MBNL2   MBNL3

N=1098

F

relative RNA (% shCTRL)

MBNL2   MBNL1

shCTRL
sh₁MBNL1
sh₂MBNL1
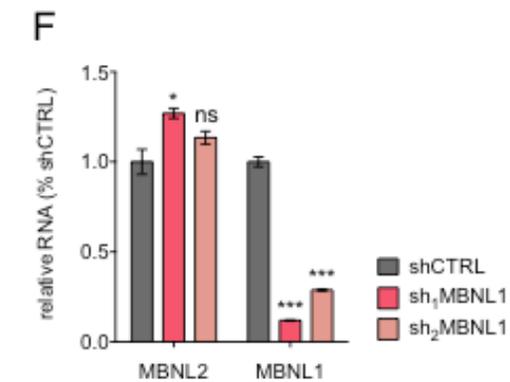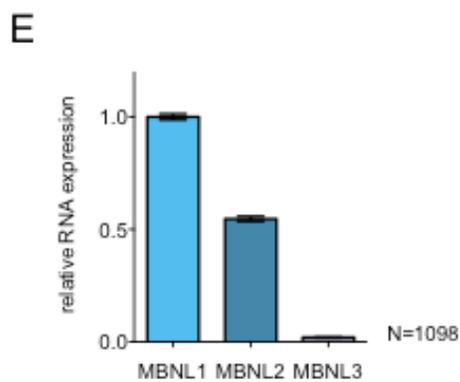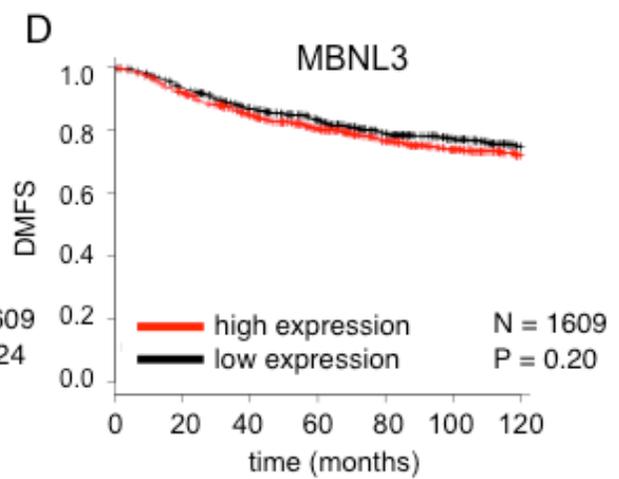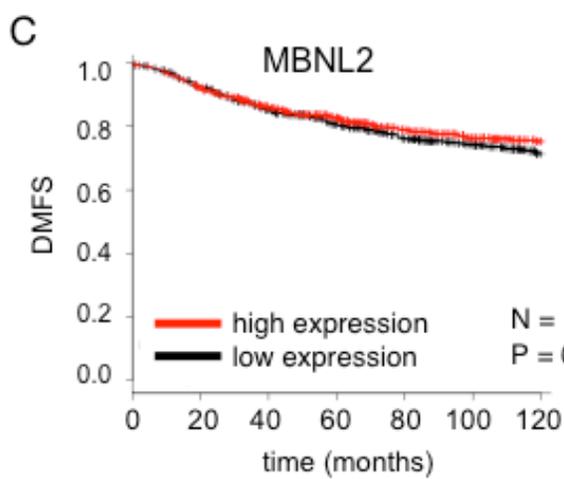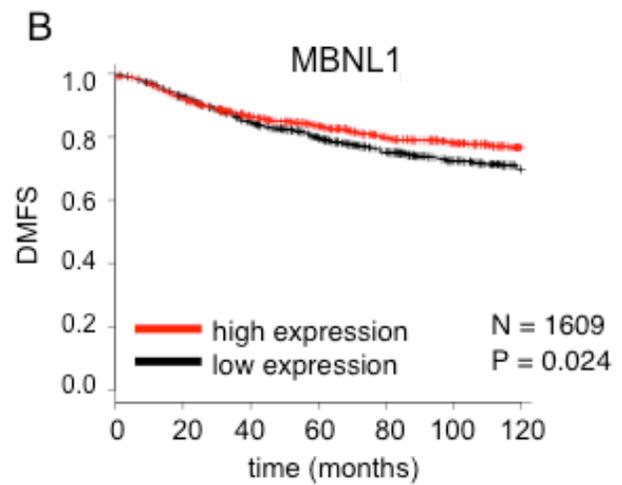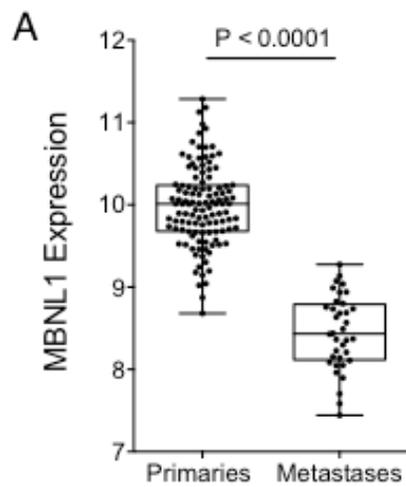
3. Effect of MBNL1 on metastasis-associated cellular phenotypes

Next, to define the cellular phenotypes altered in cells depleted of MBNL1, assays for a variety of metastasis-associated phenotypes were carried out. To test if cell proliferation was a major contributor to the enhanced metastatic colonization observed upon MBNL1 depletion, cell count assays were carried out. The viable number of cells was counted at five days post plating. ShRNA-mediated knockdown of MBNL1 did not enhance *in vitro* cell proliferation. Instead, MBNL1 depletion reduced the number of cells *in vitro* (Figure 2.4A). Furthermore, MDA-231 cells depleted of MBNL1 did not exhibit enhanced primary tumor growth rates *in vivo* (Figure 2.4B). Therefore, the enhanced *in vivo* metastasis phenotype exhibited by MBNL1 depleted cells is unlikely to be caused by increased proliferation or growth rates.

To identify potential phenotypes exhibited by MBNL1 depleted cells that could enhance metastatic activity, the ability of breast cancer cells depleted of MBNL1 to invade through matrigel was analyzed. This transwell matrigel invasion assay is a commonly used model of cancer cell invasion through the basement membrane. The major components of matrigel are laminin, entactin and collagen IV, all major structural proteins of the basement membrane (Hughes et al. 2010). Invasion through matrigel was significantly enhanced in cells depleted of MBNL1 relative to control cells in both MDA-231 and CN34 breast cancer cell populations (Figure 2.4E,G). Furthermore, this increase in invasiveness was abrogated upon stable overexpression of MBNL1 in cells depleted of MBNL1 (Figure 2.4F).

Another phenotype required for efficient metastasis to organs such as lung or brain, which contain endothelial barriers, is enhanced trans-endothelial migration capacity (Reymond et al. 2013). To assay this, a monolayer of human umbilical vein endothelial cells (HUVECs) was formed on a transwell insert, and cancer cells were seeded on top of the HUVEC monolayer. The number of cancer cells migrating through the HUVEC monolayer was then quantified. This assay found that cancer cells depleted of MBNL1 displayed enhanced trans-endothelial migration capacity relative to control cells (Figure 2.4D). Importantly, MBNL1 depletion did not enhance general migratory capacity of cancer cells in the absence of matrigel or endothelial cells as a barrier (Figure 2.4C). These results reveal that MBNL1 impedes both the invasion and trans-endothelial migration capacity of MDA-231 breast cancer cells, and suggests that these processes, fundamental to metastatic progression, contribute to the *in vivo* metastasis suppressive role of MBNL1.

Figure 2.4 Effect of MBNL1 on cell proliferation, migration, transwell invasion and trans-endothelial migration

(A) Cell proliferation of MDA-231 cells expressing shRNAs targeting MBNL1 or a control shRNA was assessed by seeding 2.5×10^4 cells and quantifying the viable number of cells after five days. N=3. (B) 5×10^5 MDA-231 cells expressing an shRNA targeting MBNL1 or a control shRNA were inject into bilateral mammary fat pads of NOD-Scid mice. Once palpable, tumor size was assessed by caliper measurements. (C) Cell migration capacity of MDA-231 MBNL1-knockdown cells compared to control cells was assessed by seeding 5×10^4 cells in a boyden chamber with 0.8um pores. After 12 hours, the number of cells migrated to the basal side of each insert was quantified. N=6-7. (D) Trans-endothelial migration capacity of MDA-231 cell with MBNL1 knockdown compared to control cells was assessed by seeding 5×10^4 cells on a trans-well insert with 0.3um pores coated with a monolayer of HUVECs. After 12 hours, the number of cells migrating through the HUVEC monolayer was quantified by counting the number of cells on the bottom side of each insert. N=6. (E) MDA-231 cells with MBNL1 knockdown or control cells were seeded at 5×10^4 cells per matrigel-coated insert with 0.8m pores. After 20 hours, the number of cells invaded onto the basal side of each insert was quantified. N=12-13. (F) MDA-231 cells with MBNL1 knockdown and stable overexpression of the MBNL1 open reading frame were subjected to transwell invasion assays. N=5. (G) Transwell invasion assays were performed using CN34 cells with MBNL1 knockdown or control cells. N=5-6. For all, p-values based on a two-way Student's t-test. Data are shown as mean ±SEM .
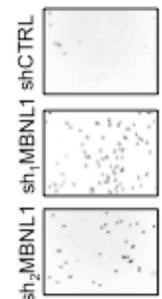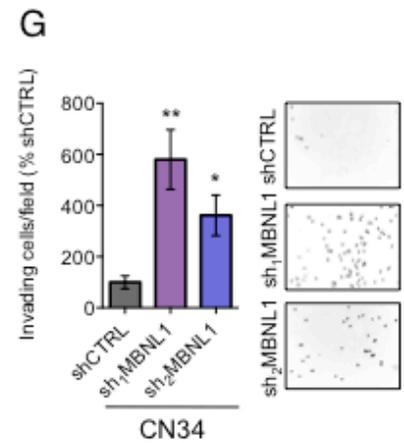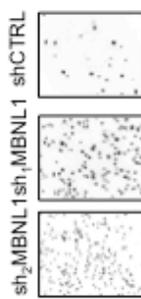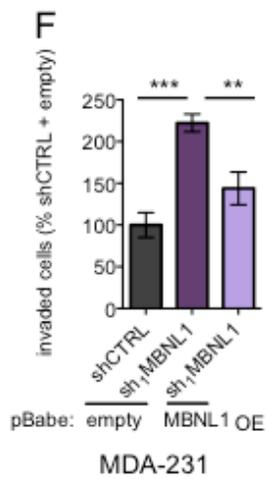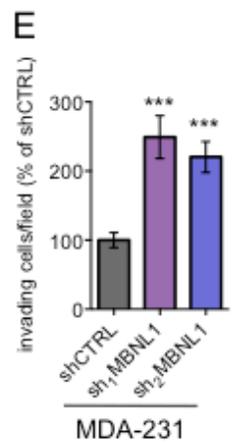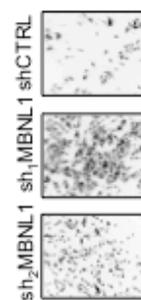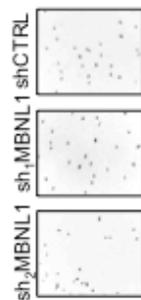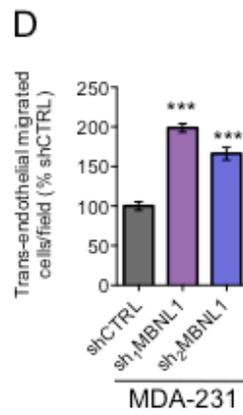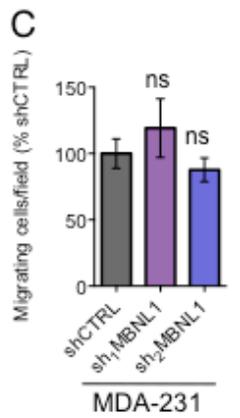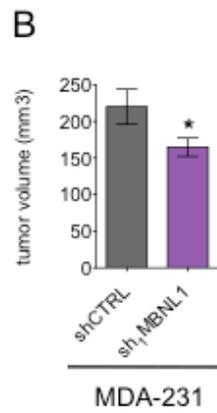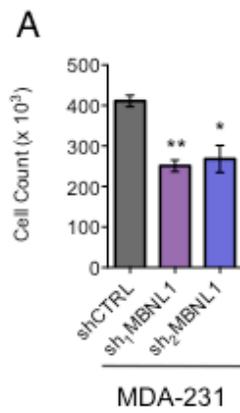
CHAPTER 3: IDENTIFICATION AND ANALYSIS OF DIRECT MBNL1-RNA INTERACTIONS

1. Using HITS-CLIP to identify MBNL1-bound RNAs in breast cancer cells

Given that MBNL1 is an RNA binding protein, and to further investigate the molecular mechanism by which MBNL1 mediates metastasis suppression, high throughput sequencing-crosslinking immunoprecipitation (HITS-CLIP) was performed to identify RNAs that directly interact with MBNL1 in breast cancer cells (Figure 3.1A). These MBNL1-interacting RNAs were potential mediators of the metastasis suppressive effects of MBNL1, as RNAs bound by MBNL1 may also be directly regulated by MBNL1. HITS-CLIP is used to identify endogenous RNA targets of RNA binding proteins and uses ultra-violet crosslinking on whole cells or tissues to capture *in vivo* RNA binding protein-RNA interactions (Licatalosi et al. 2008). Analysis of the MBNL1 HITS-CLIP sequencing library was carried out using the CIMS algorithm, which incorporates information provided by the crosslink-induced nucleotide deletions present at the site of RNA-protein interaction, a mutation type that is a consequence of the UV radiation used to crosslink RNA and proteins in the HITS-CLIP procedure (Zhang et al. 2011). Using this program, statistically significant MBNL1-bound regions were identified, and FIRE (Elemento et al. 2007), a computational method based on mutual information, was used to find any enriched nucleotide motifs in those clusters. This analysis found that MBNL1-bound CLIP tag clusters were significantly enriched for the GCUU motif (Figure 3.1B), which is in agreement with the YGCY

motif previously determined as the optimal MBNL1 binding motif by SELEX

(Goers et al. 2010). This finding is also in agreement with data from HITS-CLIP

for MBNL1 in mouse myoblasts, where MBNL1 binding sites were found to be

enriched for UGC and GCU containing 4mers (Wang et al. 2012). Analysis of the

relative genic distribution of MBNL1 binding sites revealed that, as expected from

its known role as a regulator of alternative splicing, the majority of MBNL1

binding sites occurred in introns. However, MBNL1 interacted with many sites in

coding and untranslated regions (Figure 3.1C), consistent with recent findings

from HITS-CLIP studies of MBNL1 in mouse myoblasts (Masuda et al. 2012;

Wang et al. 2012).
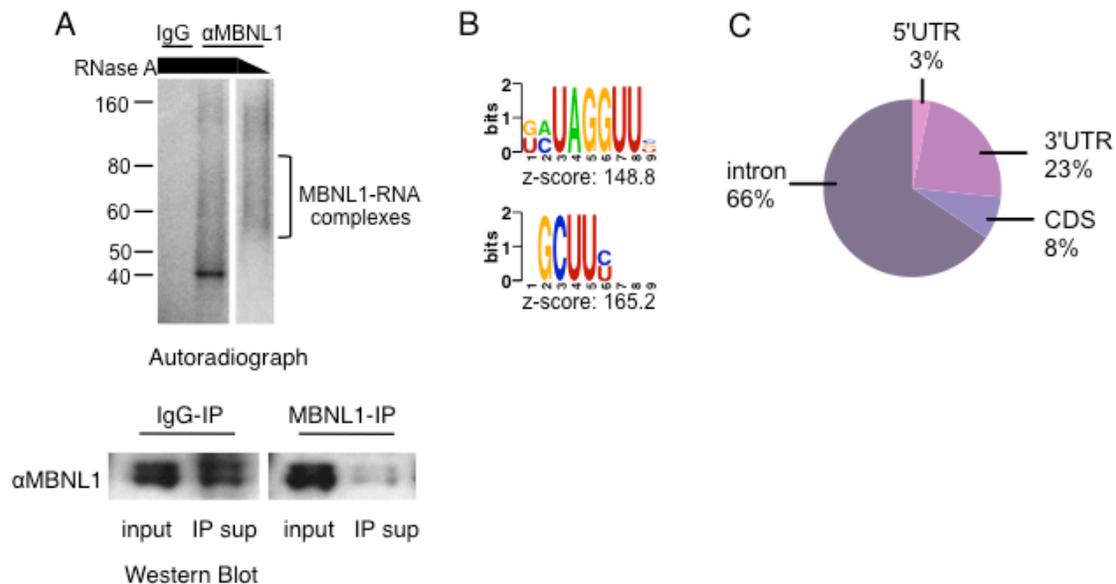
Figure 3.1 HITS-CLIP of MBNL1 in breast cancer cells

(A) ^{32}P-labeled RNA-MBNL1 complexes were visualized by autoradiography after treatment with low or high concentrations of RNase, and immunoprecipitation with anti-MBNL1 or IgG and separation by SDS-PAGE. Western blot for MBNL1 in equal volumes of lysate used for MBNL1 HITS-CLIP before (input) and after (IP sup) immunoprecipitation with MBNL1 antibody or IgG conjugated beads. (B) Nucleotide motifs significantly enriched in the MBNL1 bound HITS-CLIP CIMS clusters. (C) Pie chart depicting percent of MBNL1 HITS-CLIP clusters mapping to the indicated genic regions.

2. Analysis of MBNL1-regulated alternative splicing events in breast cancer cells

MBNL1 has been well characterized as a regulator of alternative splicing (Pascual et al. 2006; Wang et al. 2012). Therefore, the effect of MBNL1 depletion on alternative splicing in breast cancer cells was examined. To do this, high-throughput transcriptome sequencing was carried out in MDA-231 cells stably expressing a control shRNA or expressing two independent MBNL1-targeting shRNAs. The computational tool MISO (Katz et al. 2010), a program that infers relative levels of transcript isoforms using RNA sequencing data, was then used to identify MBNL1-dependent changes in transcript isoforms in breast cancer cells. Alternative exons that had MBNL1-dependent changes in abundance were identified by MISO analysis of RNA sequencing data. These were then compared with the MBNL1 HITS-CLIP data set to obtain a set of transcripts that had MBNL1-dependent differences in exon inclusion and were also bound by MBNL1 in the HITS-CLIP data, and therefore likely to be directly regulated by MBNL1 binding. The criteria used to arrive at this set of MBNL1-dependent alternative exons was that the event had a bayes factor of ≥10 and a change in the percent spliced in exon ($\Delta\Psi$) of ≥|0.2| in each of the two independent MBNL1-targeting shRNA expressing cell lines compared to the control cell line, as assessed by MISO (Table 3.1).

Table 3.1 MISO predicted MBNL1-dependent alternative exons

refseq ID	symbol	alternative exon	BF sh$_1$MBNL1	BF sh$_2$MBNL1	ΔΨ MISO sh$_1$MBNL1	ΔΨ MISO sh$_2$MBNL1	ΔΨ RT-PCR sh$_1$MBNL1	ΔΨ RT-PCR sh$_2$MBNL1
NM_015850	FGFR1	chr8:38287200:38287466	1.0E+12	1.0E+12	0.57	0.36	0.39	0.18
NM_001077207	SEC31A	chr4:83763338:83763634	1.0E+12	1.0E+12	0.28	0.21	0.26	0.20
NM_002840	PTPRF	chr1:44063419:44063724	3.2E+02	3.3E+08	0.37	0.41	0.09	0.10
NM_017802	HEATR2	chr7:819590:819781	2.9E+01	9.9E+01	0.33	0.33	0.06	0.06
NM_032434	ZNF512	chr2:27806293:27806583	6.8E+01	1.5E+01	0.31	0.22	0.06	0.06
NM_015442	CNOT10	chr3:32778902:32778982	5.3E+01	3.3E+01	0.44	0.37	0.03	0.04
NM_014371	AKAP8L	chr19:15514286:15514366	1.5E+01	2.2E+05	0.43	0.49	0.00	0.00
NM_004082	DCTN1	chr2:74593586:74593687	5.4E+06	1.3E+03	0.30	0.24	0.00	0.00
NM_005336	HDLBP	chr2:242194789:242194926	1.0E+12	1.0E+12	0.24	0.21	0.00	0.00
NM_002116	HLA-A	chr6:29912277:29912393	1.0E+12	1.0E+12	-0.27	-0.24	0.00	0.00
NM_002116	HLA-A	chr6:29912277:29912411	1.0E+12	1.0E+12	-0.29	-0.25	0.00	0.00
NM_001040694	INCENP	chr11:61908972:61908983	1.0E+12	2.3E+08	-0.31	-0.27	0.00	0.00
NM_001040446	MTMR12	chr5:32235068:32235235	2.5E+02	1.2E+03	0.31	0.33	0.00	0.00
NM_001243984	RELA	chr11:65429158:65429306	8.1E+02	4.2E+02	0.33	0.36	0.00	0.00
NM_003200	TCF3	chr19:1619315:1619473	2.8E+02	1.0E+12	0.27	0.37	0.00	0.00
NM_003200	TCF3	chr19:1619315:1619470	1.4E+01	6.0E+02	0.24	0.34	0.00	0.00
NM_001122956	DBNL	chr7:44092466:44092550	4.7E+01	2.9E+04	0.37	0.37	0.00	0.00
NR_027850	MTX2	chr2:177134944:177135147	1.3E+07	1.8E+09	-0.21	-0.20	-0.01	-0.03
NM_130847	AMOTL1	chr11:94528177:94528326	5.7E+02	6.0E+01	0.32	0.29	-0.02	-0.02
NM_021019	MYL6	chr12:56554410:56554454	1.0E+12	1.0E+12	-0.34	-0.22	-0.30	-0.23
NM_001008660	PICALM	chr11:85689113:85689136	1.0E+12	1.0E+12	-0.26	-0.25	-0.38	-0.29

This analysis revealed 21 MBNL1-dependent alternative exons that were in transcripts also bound by MBNL1, with 71% of these exons exhibiting inclusion and the remainder exhibiting exclusion upon MBNL1 knockdown. Transcripts that had MISO predicted MBNL1-dependent changes in exon inclusion and had MBNL1 HITS-CLIP binding sites were validated by RT-PCR. Although many of the MISO-predicted transcript variants were validated by RT-PCR, several did not validate. In these cases, only one of the predicted transcript variants was detected by RT-PCR. However, of the genes that had detectable levels of both MISO-predicted transcript variants by RT-PCR analysis, there was good correlation between the MISO calculated $\Delta\Psi$ and the $\Delta\Psi$ assessed by RT-PCR (Figure 3.2B).

Figure 3.2 MBNL1-dependent skipped exons

(A) MBNL1 interacts with sites near MBNL1-skipped exons. For the indicated transcripts, the skipped exon and its upstream and downstream intron and exon are shown, with mapped reads from experimental replicates of MBNL1 HITS-CLIP. (B) Correlation plots of ΔΨ values calculated by MISO compared to ΔΨ values calculated by semi-quantitative RT-PCR for MBNL1-dependent alternative exons for genes that had both transcript forms expressed as assessed by RT-PCR. Shown are separate plots for each independent MBNL1-targeting shRNA compared to shCTRL. Pearson correlation coefficient is shown for each plot.

A

FGFR1

SEC31A

MYL6

PICALM

B
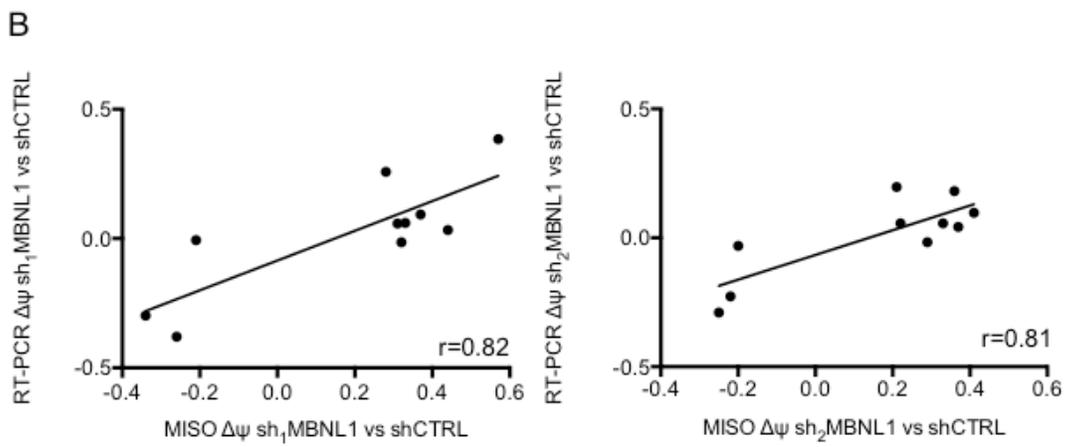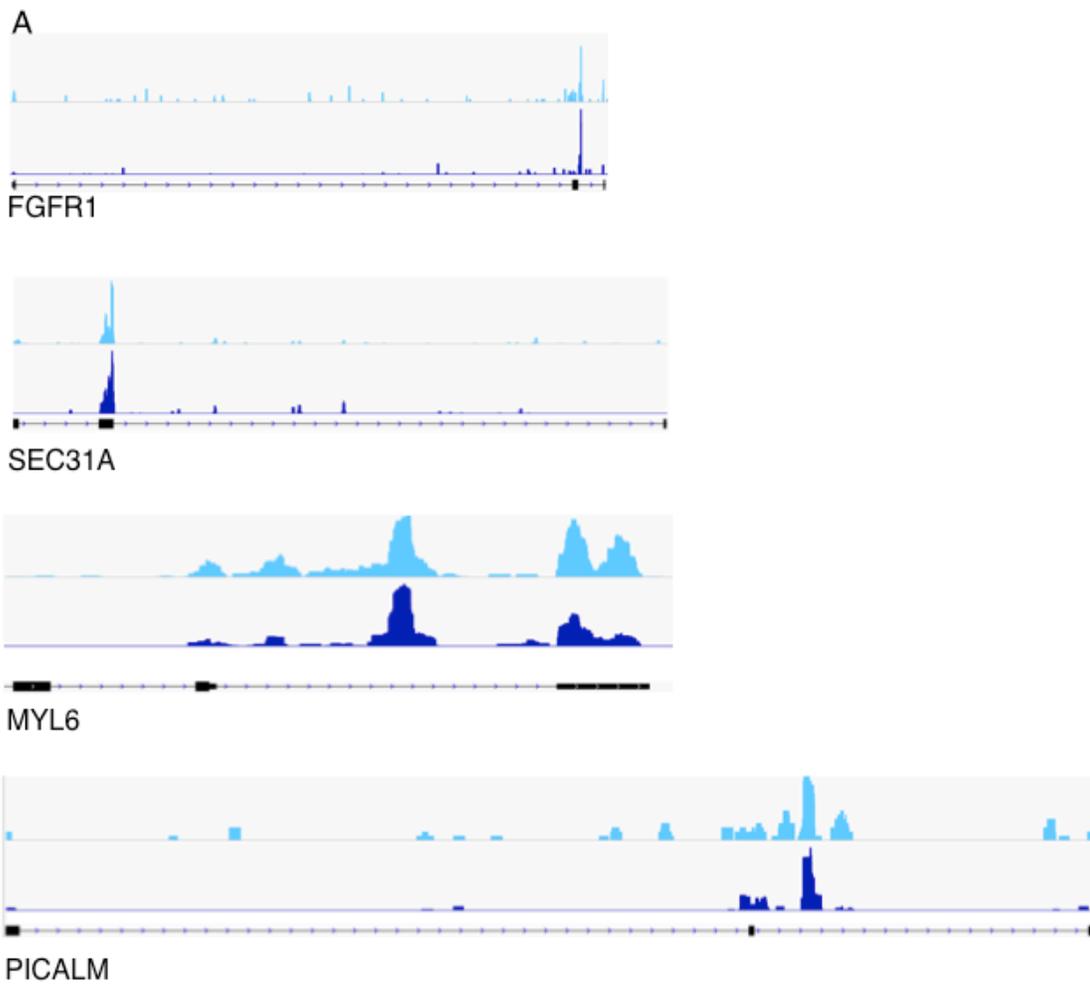
This analysis found four genes, FGFR1 (fibroblast growth factor receptor 1), SEC31A (SEC31 homolog A), MYL6 (myosin light chain 6), and PICALM (phosphatidylinositol binding clathrin assembly protein) that had large (> 30%) changes in exon exclusion in MBNL1-knockdown cells compared to control cells by both MISO and by RT-PCR analysis, and were bound by MBNL1 (Table 3.1, Figure 3.2A, Figure 3.3A). Two of these transcripts, FGFR1 and SEC31A, exhibited increased inclusion of a specific exon when MBNL1 was depleted. Two others, MYL6 and PICALM, displayed increased exclusion of an exon upon MBNL1 depletion. Furthermore, in MDA-231 cells expressing a MBNL1-targeting shRNA, overexpression of MBNL1 resulted in the reversion of these splice forms (Figure 3.3 A). MBNL1-dependent changes in relative transcript isoform abundance were then tested to see if they could affect the transwell invasion phenotype observed in MBNL1 depleted cells. In this set of experiments, the transcript variant that had an increase in exon inclusion was specifically depleted using siRNA. For the transcripts that had an increase in exon inclusion upon MBNL1 knockdown (FGFR1, SEC31A), these longer isoforms were experimentally depleted using siRNA in MBNL1 knockdown cells. Conversely, MYL6 and PICALM had decreased exon inclusion in MBNL1 knockdown cells, and therefore their longer isoforms were depleted using siRNA in shCTRL cells. The modulation of these transcript isoforms in these cells lines did not have significant effect on the transwell invasion capacity of the cells (Figure 3.3B).

Figure 3.3 Semi-quantitative RT-PCR analysis of MISO-predicted MBNL1-

dependent alternative splicing events and effect on *in vitro* invasion

(A) Semi-quantitative RT-PCR analysis of MBNL1-dependent transcript isoforms

predicted by MISO in RNA-seq data. Shown are RT-PCR products visualized by

agarose gel analysis. Percent exon inclusion relative to shCTRL or to MBNL1

overexpressing cells was calculated using Image J. Cartoons represent the exon

included (contains the red exon) and the skipped exon forms of each indicated

transcript, with the DNA length in basepairs indicated for each transcript variant.

(B) MDA-231 cells expressing shCTRL or sh$_1$MBNL1 were transiently transfected

with siRNAs targeting the MBNL1-dependent exon of the indicated transcript.

Transwell invasion assays were then carried out. N=6. Data are shown as mean

±S.E.M.

A

MYL6

shCTRL sh₁MBNL1 sh₂MBNL1

324
279

inclusion (%):   1   0.40   0.54

pBabe: empty MBNL1 OE

0.94   1

PICALM

199
175

inclusion (%):   1   0.45   0.58

0.78   1

FGFR1

535
268

inclusion (%):   1   3.04   1.96

1.36   1

SEC31A

488
146

inclusion (%):   1   1.41   1.31

1.19   1

B
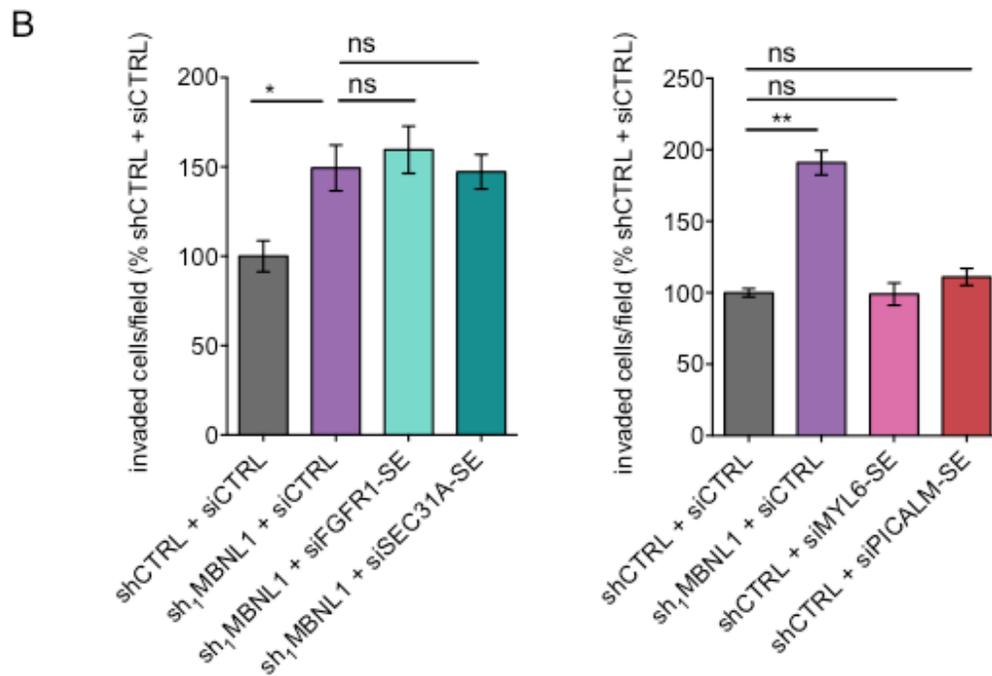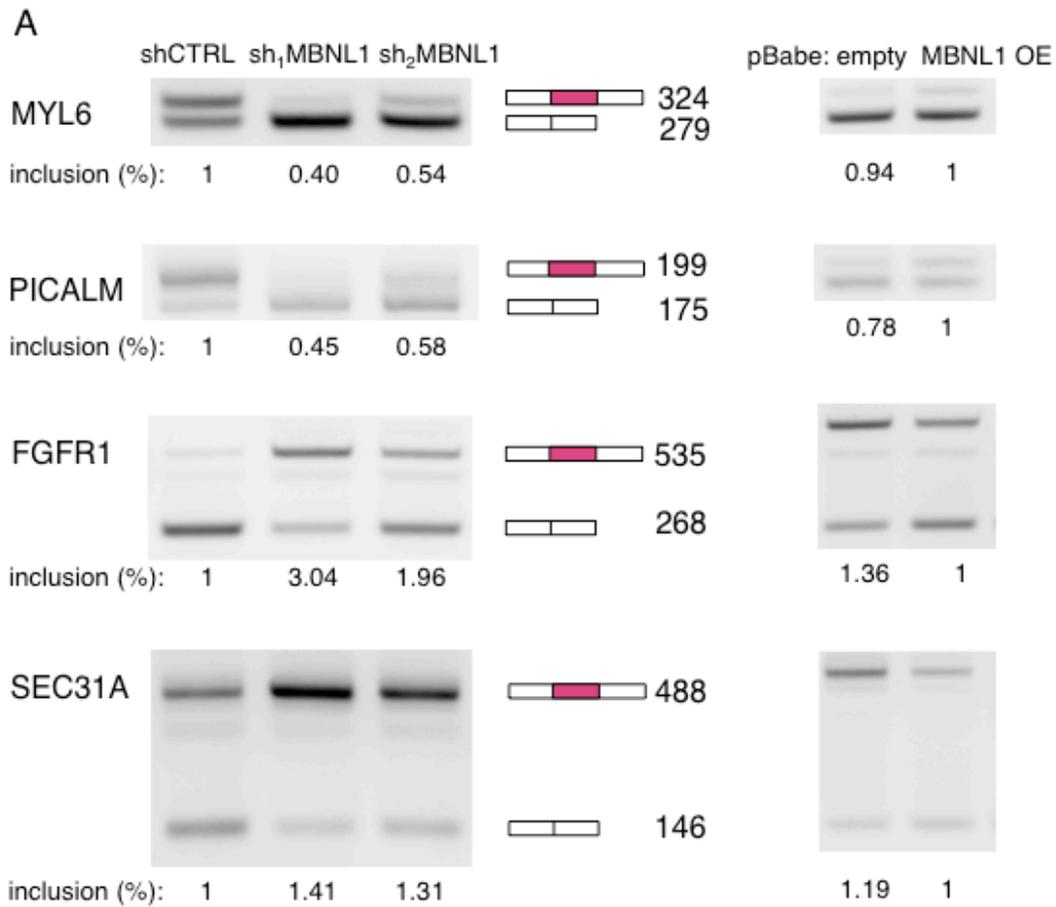
This result does not exclude the possibility that regulation of alternative splicing by MBNL1 contributes to invasion and metastasis. However, in the analysis of the group of transcripts identified here, of those that had both direct MBNL1-dependent differences in transcript isoform expression and the greatest changes in isoform expression, none individually affected the invasion capacity of MDA-231 breast cancer cells.

CHAPTER 4. MBNL1 REGULATION OF TRANSCRIPT STABILITY AND IDENTIFICATION OF MBNL1-STABILIZED METASTASIS SUPPRESSOR TRANSCRIPTS

1. Analysis of MBNL1 effect on global transcript stability

A major type of regulatory activity for RNA binding proteins is control of transcript stability. There is some evidence that MBNL1 levels can modulate the stability of transcripts in mouse myoblasts (Masuda et al. 2012), but the effect of MBNL1 on transcript stability in breast cancer cells has not been reported. Therefore, to determine if MBNL1 depletion alters the stability of transcripts to which it binds, a transcriptome-wide analysis of RNA stability was conducted. In this experiment, breast cancer cells with MBNL1 depleted or control cells, were treated with alpha-amanitin to inhibit polymerase II-dependent transcription, and RNA was isolated. Relative transcript levels were determined by transcriptomic profiling. Relative distribution of MBNL1-bound transcripts in the stability dataset was assessed using TEISER (Goodarzi et al. 2014). This analysis showed that MBNL1-bound transcripts were enriched among transcripts that had decreased

stability upon MBNL1 knockdown (Figure 4.1A). This is consistent with a model

where MBNL1-bound transcripts are stabilized by MBNL1 binding.
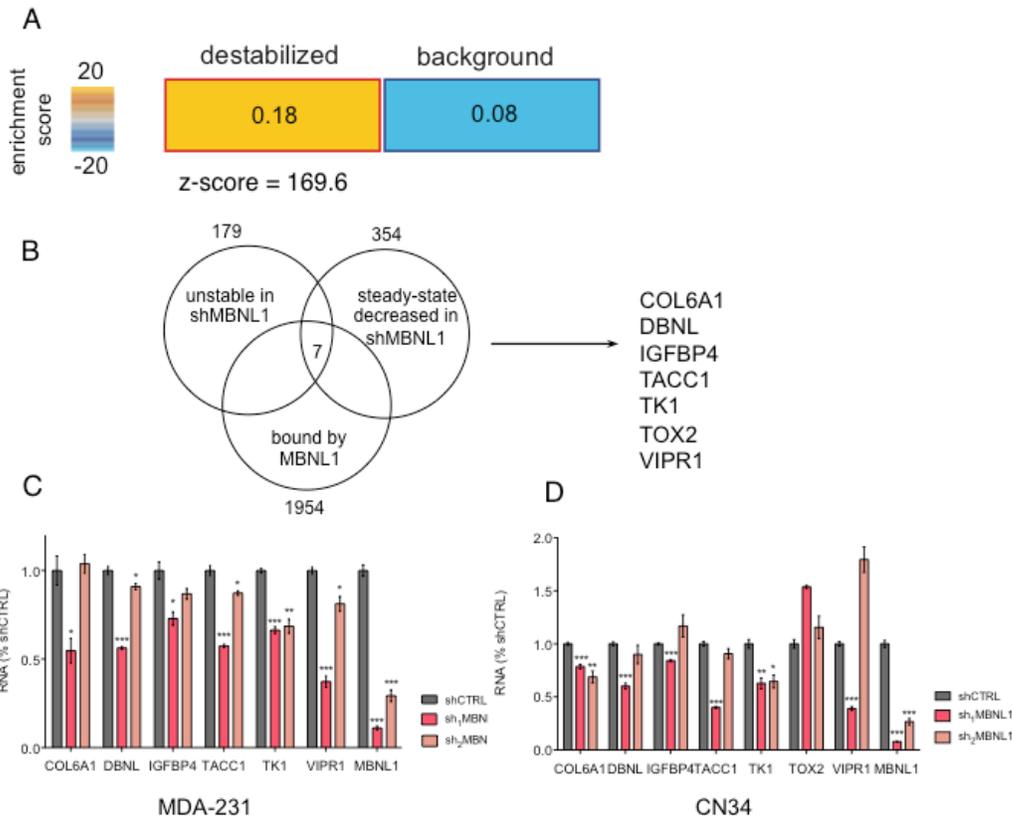
Figure 4.1 Identification of MBNL1-stabilized transcripts

(A) Differential stability of transcripts in MDA-231 MBNL1-depleted compared to control cells as assessed by alpha-amanitin treatment and microarray profiling. Transcripts were ordered by their log2 fold-change in MBNL1-knockdown cells compared to control cells at nine hours after alpha-amanitin treatment normalized to pre-alpha-amanitin treated transcript levels. MBNL1-bound transcripts, as assessed by HITS-CLIP, were significantly enriched among transcripts destabilized by MBNL1-knockdown. (B) Venn diagram of transcripts stabilized and bound by MBNL1. (C) qRT-PCR of candidate MBNL1 target transcripts in MDA-231 shMBNL1 and shCTRL cells at steady-state. HPRT1 was used as an endogenous control. N=3. (D) qRT-PCR of putative MBNL1 target transcripts in CN34 shMBNL1 and shCTRL cells at steady-state. HPRT1 was used as an endogenous control.

Next, the identification of genes directly bound and regulated by MBNL1 that also mediate MBNL1's effects on metastasis was carried out. To do this, transcripts that were directly bound by MBNL1 in the HITS-CLIP dataset, were stabilized by MBNL1 based on the alpha-amanitin analysis, and had reduced steady-state abundance upon MBLN1 knockdown, were identified. This analysis identified a group of seven transcripts as potential regulators of breast cancer metastasis downstream of MBNL1 (Figure 4.1B). The steady-state levels of these transcripts were assessed by qRT-PCR in MBNL1-knockdown compared to control cells in the MDA-231 and CN34 cancer cell lines. Three of these transcripts exhibited reduced steady-state levels upon MBNL1 knockdown in both cell lines and using two independent MBNL1-targeting shRNAs: DBNL, TACC1 and TK1 (Figure 4.1C,D). The reduced stability of these transcripts after alpha-amanitin treatment and upon MBNL1-depletion in MDA-231 and CN34 cells was verified by qRT-PCR (Figure 4.2A,B). The stabilization of these transcripts by MBNL1 was also tested using an independent method of transcription inhibition. In this assay, dichlorobenzimidazole 1-β-D-ribofuranoside (DRB), an inhibitor of CDK9, was used to inhibit transcription. Breast cancer cells were treated with DRB and RNA was isolated at various time-points for qRT-PCR analysis. DBNL, TACC1, and TK1 transcripts exhibited shorter half-lives in MBNL1-depleted cells relative to control cells, while COL6A1, IGFBP4 and VIPR1 transcripts did not exhibit significant changes in decay rates in MBNL1-depleted cells relative to control cells in this assay (Figure 4.2C). These findings demonstrate that, in

breast cancer cells, MBNL1 depletion reduces the stability of a subset of

transcripts to which it binds.
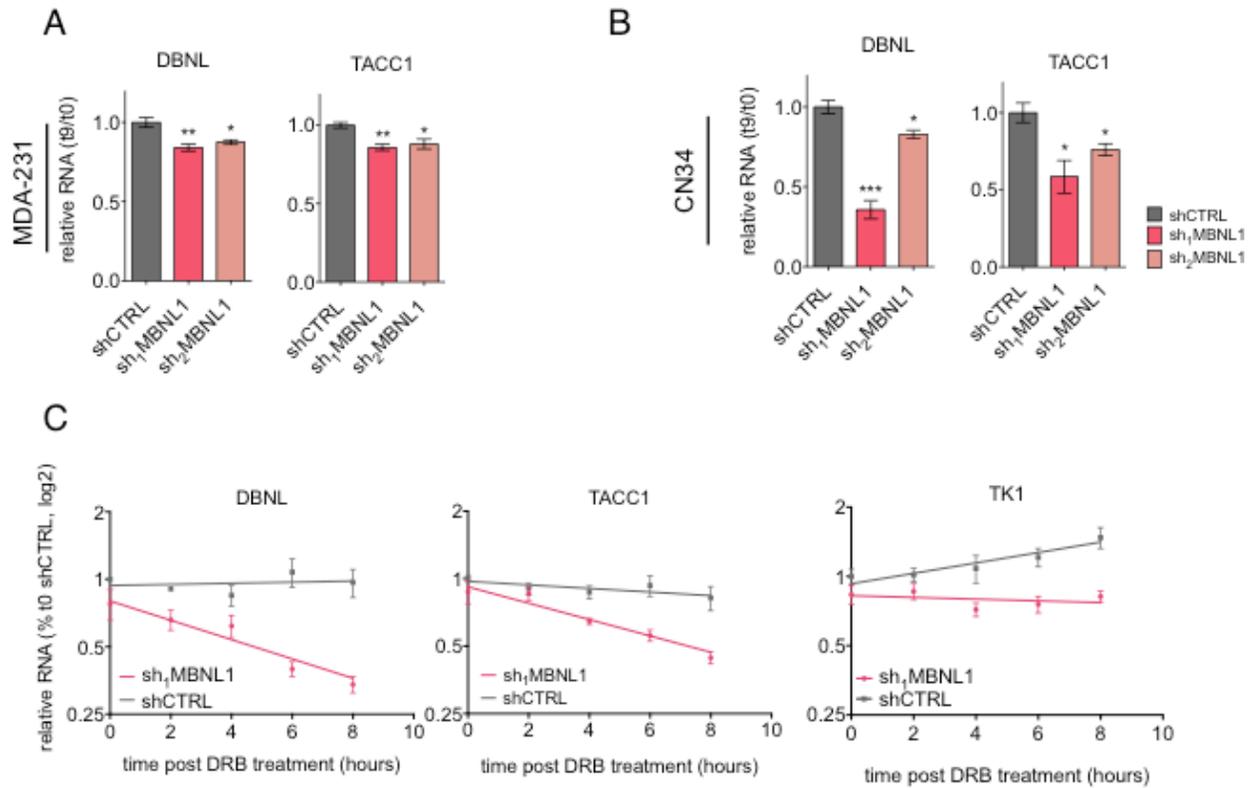
Figure 4.2 Validation of MBNL1-dependent transcript stability

qRT-PCR of putative MBNL1 target transcripts in (A) MDA-231 and (B) CN34
shMBNL1 and shCTRL cells after nine hours of alpha-amanitin treatment. 18S
was used as an endogenous control. N=3. (C) qRT-PCR of putative MBNL1
target transcripts in MDA-231 shMBNL1 and shCTRL cells at times indicated
after treatment with DRB. N=3. For all, P-values are from a one-way Student's t-
test. Data are shown as mean ±S.E.M.

2. Effects of DBNL and TACC1 on metastatic invasion in breast cancer cells

MBNL1 depletion leads to the reduced stability and subsequently the reduced abundance of DBNL, TACC1 and TK1 transcripts. To test if rescuing the expression of these genes in cells depleted of MBNL1 could reverse the enhanced metastatic phenotype observed, each of these genes was stably overexpressed in MBNL1-depleted cells, using overexpression of mCherry as a control gene. These cells were then subjected to *in vitro* transwell invasion assays, and a significant decrease was observed in the invasion capacity of the cell lines overexpressing DBNL and TACC1 relative to cells expressing mCherry control (Figure 4.3A). However, TK1 over-expression in this context did not significantly reduce the invasiveness of MBNL1-depleted cells. As DBNL and TACC1 overexpression could both compensate for MBNL1-knockdown in transwell invasion assays, elucidating the roles of DBNL and TACC1 in repressing invasion and metastasis downstream of MBNL1 was the next step. Also, and consistent with these findings, overexpressing DBNL and TACC1 in the CN34-Lm1a line, an independent highly metastatic breast cancer cell population, also suppressed cell invasiveness (Figure 4.3B).
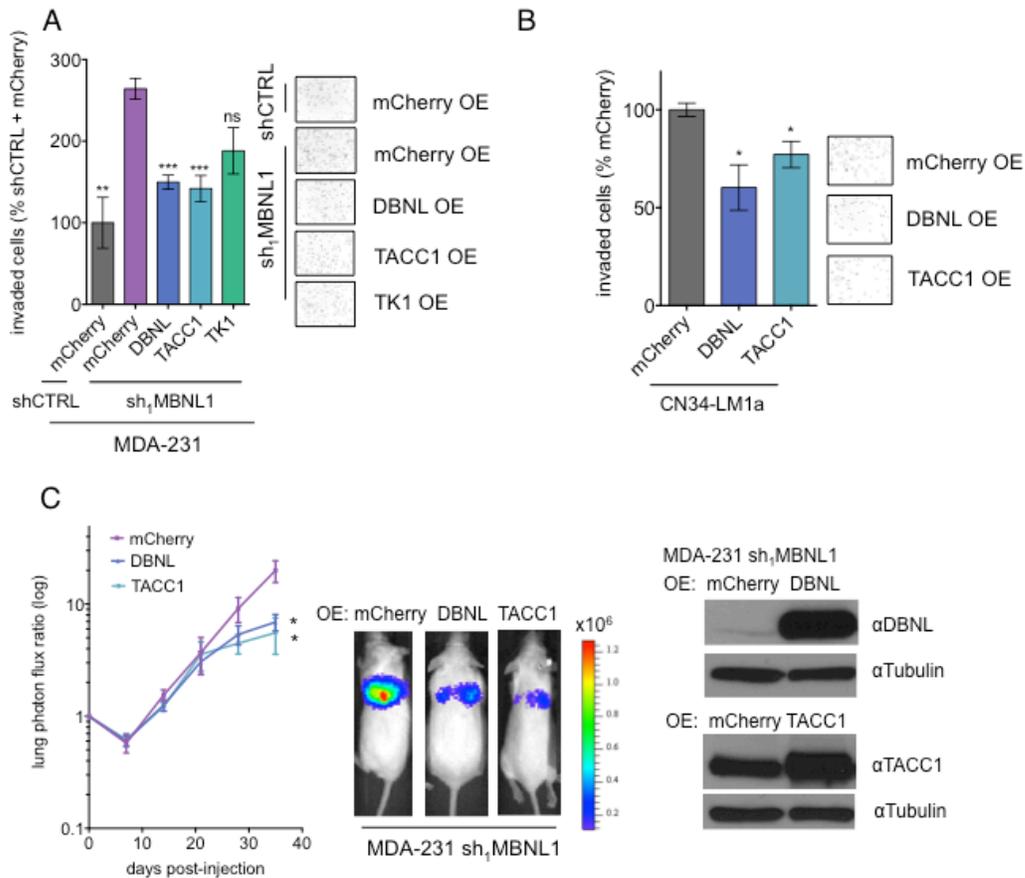
Figure 4.3 Effect of DBNL and TACC1 on invasion and metastatic lung colonization

(A) Candidate MBNL1-regulated genes were stably overexpressed in shMBNL1 MDA-231 cells. mCherry was overexpressed as a control. 5×10^4 cells were seeded in transwell matrigel invasion chambers and the number of cells invading to the basal side of the insert was quantified after 20 hours. N=6. (B) CN34-LM1a cells stably overexpressing DBNL, TACC1 or mCherry were assessed for transwell invasion capacity. N=6. (C) 2×10^4 MDA-231 cells stably overexpressing DBNL, TACC1 or mCherry were injected into the venous circulation of NSG mice. Lung colonization was monitored by bioluminescence imaging. N=6. Western blot showing levels of DBNL and TACC1 in whole cell lysate from MDA-231 sh_1MBNL1 cells stably overexpressing the indicated factors. Data are shown as mean ±SEM.

To determine if these genes impacted metastatic lung colonization *in vivo*, breast cancer cell with shRNA-mediated stable MBNL1 knockdown were engineered to stably overexpress DBNL, TACC1 or mCherry. These cells were injected into the venous circulation of immunodeficient mice. Overexpressing DBNL and TACC1 in MBNL1-depleted cells significantly abrogated the enhanced metastatic phenotype conferred by MBNL1 depletion (Figure 4.3C). These findings demonstrate that overexpression of either DBNL or TACC1 is sufficient to repress the metastatic lung colonization of MBNL1-depleted breast cancer cells *in vivo,* as well as to suppress transwell invasion *in vitro*.

One commonly used mechanism of regulating mRNA stability is alternative poly(A) site choice (APA), which leads to 3'UTR length alterations. Both DBNL and TACC1 have extensive binding of MBNL1 in their 3'UTRs, a location which could control poly(A) site selection (Figure 4.4A). The length of the 3'UTRs of DBNL and TACC1 were assessed by qRT-PCR, using primers specific to different part of these 3'UTRs. In this analysis, both DBNL and TACC1 3'UTR lengths were unchanged upon MBNL1 depletion (Figure 4.4B,C).
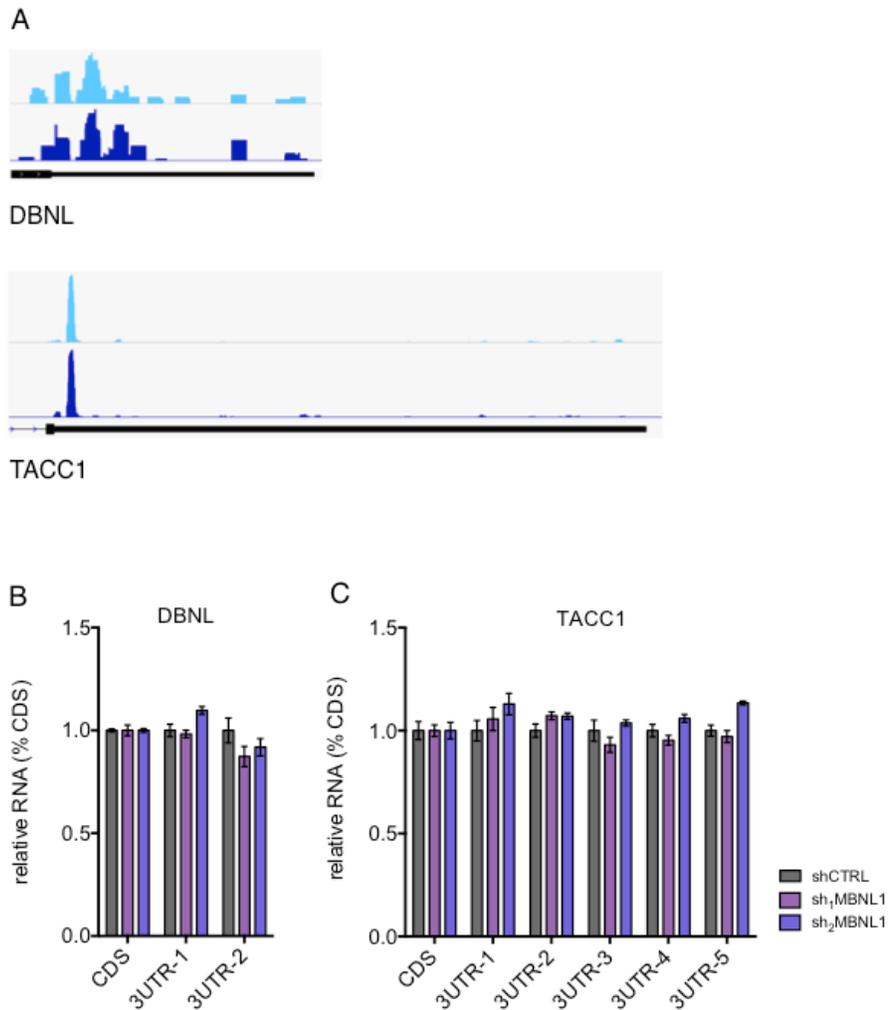
Figure 4.4 MBNL1 binding on 3'UTRs of DBNL and TACC1 and 3'UTR length (A) MBNL1 interacts with the 3'UTRs DBNL and TACC1. The last exon of the indicated transcripts are shown with mapped reads from experimental replicates of MBNL1 HITS-CLIP. (B) qRT-PCR analysis of DBNL and (C) TACC1 3'UTR length in MDA-231 cells with stable knockdown of MBNL1 or cells expressing a control shRNA. 3'UTR primers are numbered as increasingly distal from the stop codon. N=3. Data are shown as mean ±S.E.M.

As poly(A) tail length is a major determinant of mRNA stability, the relative poly(A) tail lengths of DBNL and TACC1 were assessed. To do this, a method employing 3' tagging of adenylated transcripts and subsequent PCR amplification of the poly(A) tail using transcript specific primers, was carried out (Janicke et al. 2012). This assay revealed no difference in the poly(A) tail length of DBNL and TACC1 transcripts upon MBNL1 depletion (Figure 4.5A-D). This result suggested that MBNL1 does not regulate the stability of these transcripts through controlling their adenylation or deadenylation. These data are consistent with MBNL1 binding to the 3'UTRs of DBNL and TACC1 and regulating their stability in a poly(A) tail length- and APA- independent manner.
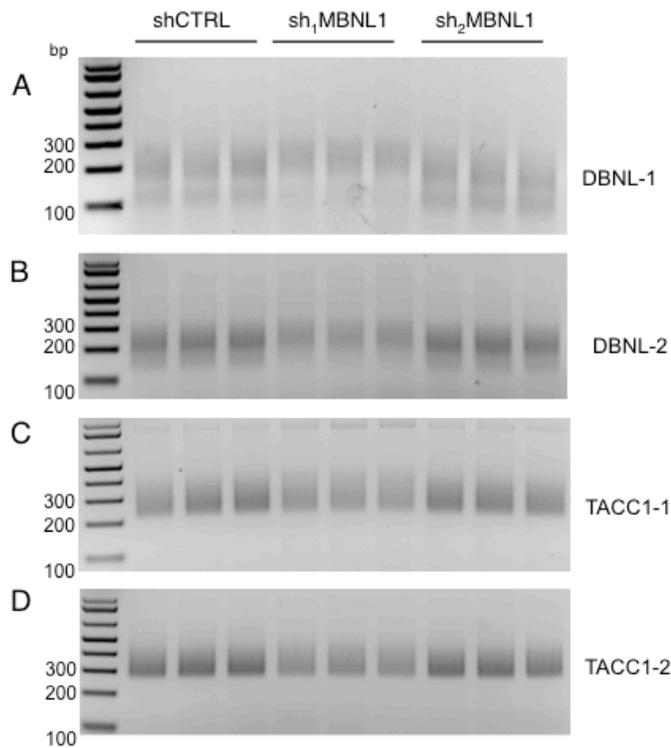
Figure 4.5 Poly(A) tail length of DBNL and TACC1 transcripts

The poly(A) tail length of DBNL and TACC1 transcripts was assessed in MDA-231 shCTRL and MBNL1 knockdown cells as described in methods. The PCR products were run on a 2% agarose-TAE gel and stained with ethidium bromide. (A,B) DBNL poly(A) tail length using two independent gene-specific primers (DBNL-1, DBNL-2). (C,D) TACC1 poly(A) tail length using two independent gene-specific primers (TACC1-1, TACC1-2).

3. Clinical correlation of DBNL and TACC1 in breast cancer outcome

TACC1 expression can be inferred from affymetrix microarray datasets as this transcript is represented by an informative probe. Analysis of a large dataset comprising 3455 tumor samples representing multiple cohorts revealed a significant correlation between TACC1 expression and relapse-free survival in breast cancer patients (Figure 4.6A), a finding consistent with TACC1 acting as a suppressor of breast cancer metastasis. However, DBNL is not represented by a probe on the microarray platform used to analyze these large breast cancer datasets. Therefore, to determine whether there exists an association between DBNL and breast cancer progression, qRT-PCR was performed on a panel of publicly available cDNAs (Origene) generated from breast tumor-derived RNA. Consistent with it having a metastasis suppressor role, metastatic stage IV tumors expressed significantly lower levels of DBNL relative to stage I and stage II tumors (Figure 4.6B).
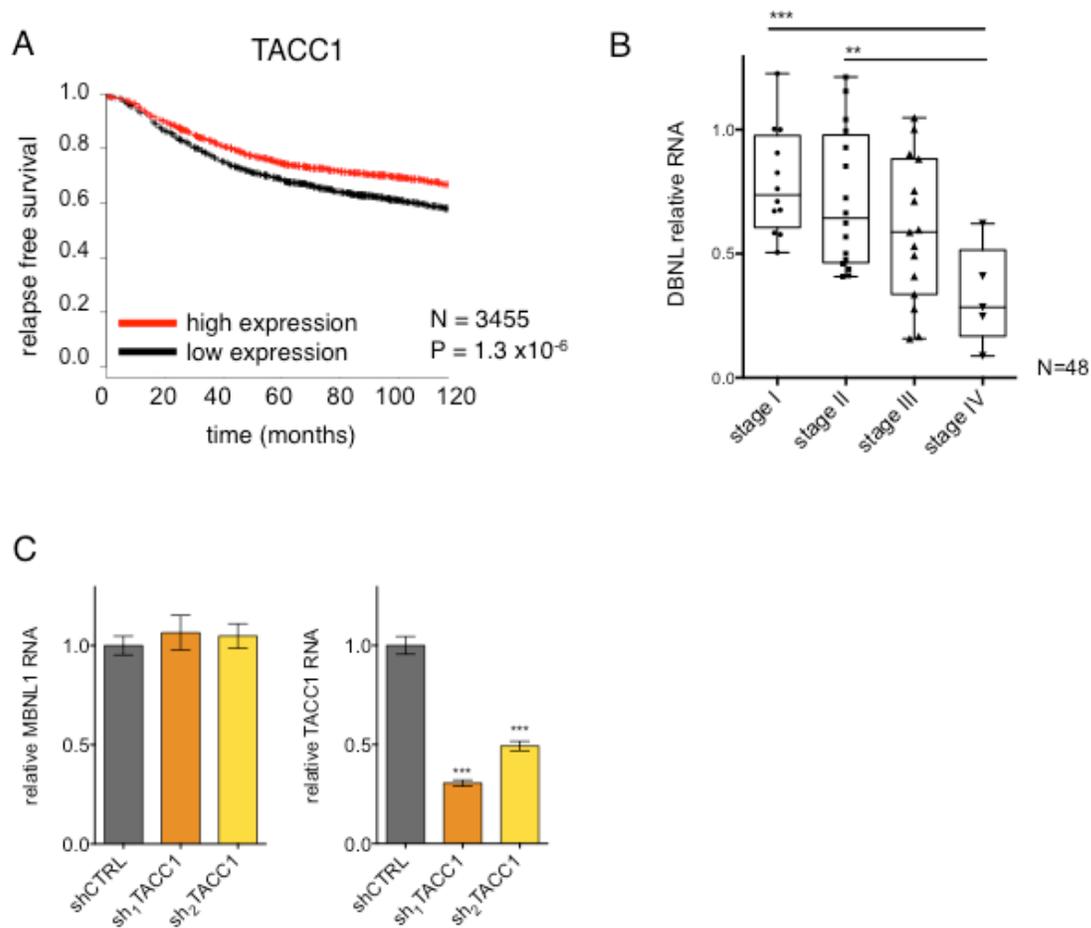
Figure 4.6 Clinical correlation of DBNL and TACC1 expression in breast cancer
(A) Kaplan-Meier curve showing relapse free survival of breast cancer patients
with tumors expressing high (red) or low (black) levels of the TACC1 transcript.
N=3455. (B) Levels of DBNL were assessed by qRT-PCR from RNA obtained
from a panel of staged breast cancer tumors. Beta-actin was used as an
endogenous control. N=48. (C) qRT-PCR for MBNL1 and TACC1 in MDA-231
cells stably expressing two independent TACC1-targeting shRNAs, HPRT1 was
used as an endogenous control. P-values are based on Student's t-test unless
otherwise noted. Data are shown as mean ±SEM.

Interestingly, data suggests that MBNL1 is upstream of TACC1 in this regulatory relationship, as shRNA-mediated depletion of TACC1 did not affect MBNL1 transcript levels in MDA-231 cells (Figure 4.6C). However, MBNL1 could conceivably be regulated at the transcriptional level by TACC1, as TACC1 has been shown to have effects on the transcriptional activity of the retinoic acid receptor alpha and thyroid receptor, and also interacts with YEATS4/GAS41, which is part of a complex that acetylates histones (Guyot et al. 2010; Lauffart et al. 2002). Together, this data is consistent with MBNL1 interacting with TACC1 and DBNL transcripts to stabilize them, thereby suppressing invasion and metastasis of MDA-231 breast cancer cells. DBNL expression is reduced in stage IV breast cancer compared to stage I and stage II tumors. TACC1 transcript levels are positively correlated with significantly longer relapse-free survival of breast cancer patients, suggesting that TACC1 acts as a breast cancer metastasis suppressor in human breast cancers.

4. Effect of secreted proteins on MBNL1-mediated phenotypes

Cancer cells secrete factors that are known to regulate their spread from a primary tumor and to promote colonization of distant organs. As MBNL1 depletion enhanced *in vitro* invasion and trans-endothelial migration, and both of these processes can be modulated by cancer cell-mediated secretion of molecules into the extracellular environment, it seemed an intriguing possibility that MBNL1 might regulate the expression levels of secreted pro-metastatic molecules. A study in mouse myoblasts has also found a role for MBNL protein

localization of transcripts that contain a signal peptide sequence (Wang et al. 2012). To test the possibility that MBNL1 depletion impedes the invasion capacity of breast cancer cells through regulation of secreted factors, the invasion capacity of cells treated either with media that had been conditioned by MBNL1-depleted cells or with media that had been conditioned by cells expressing a control shRNA was assessed. In this assay, breast cancer cells seeded in conditioned media from MBNL1-depleted cells displayed significantly greater invasion capacity relative to cells seeded in conditioned media from control cells (Figure 4.7A). This finding suggested that factors specifically secreted by MBNL1-depleted cells could contribute to the increased cell invasiveness.

To identify any molecule(s) regulated by MBNL1 that also mediated this pro-invasive effect, stable isotope labeling by amino acids in cell culture (SILAC) and mass spectrometry (MS) was conducted to identify proteins with different levels in media conditioned by control or MBNL1-depleted cells. This experiment revealed five proteins with higher abundance in MBNL1-depleted cell conditioned media relative to control cell conditioned media. These proteins are CTGF, LOXL2, QSOX1, APLP2 and NPC2 (Table 4.1). Three of these factors, CTGF (connective tissue growth factor), LOXL2 (lysyl oxidase-like 2), and QSOX1 (quiescin Q6 sulfhydryl oxidase 1), have been previously implicated in cell migration and invasion. LOXL2 and QSOX1 have also been associated with breast cancer progression (Ahn et al. 2013; Barker et al. 2011; Chen et al. 2007; Katchman et al. 2013).

Figure 4.7 Identification of secreted proteins upregulated by MBNL1 depletion

(A) Conditioned media from MDA-231 cell with MBNL1 knockdown or shCTRL

cells was used to treat MDA-231 shCTRL cells that were subjected to transwell

invasion assays. N=5-6. (B) Western blot of conditioned media from MDA-231

and CN34 cells with MBNL1 knockdown or control cells, all in biological triplicate.

Bar graphs are quantitation of LOXL2 and QSOX1 signal from the Western blots.

(C) ELISA for CTGF in conditioned media from MDA-231 cells with MBNL1

knockdown or control cells. N=3. Data are shown as mean ±S.E.M.

A

invaded cells/field (% shCTRL)

conditioned media: shCTRL / sh₁MBNL1

MDAshCTRL cells
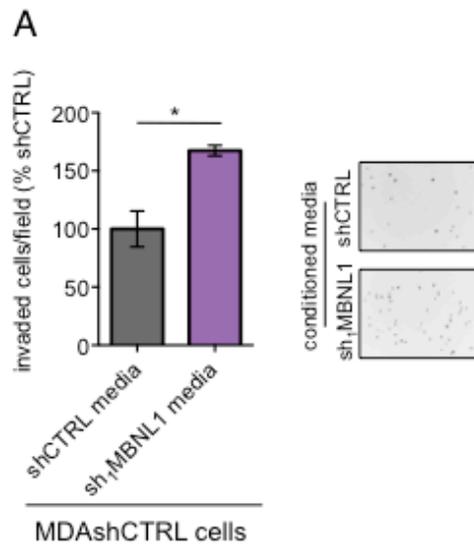
Table 4.1 Proteins identified from SILAC analysis of conditioned media from MDA-231 cells ($sh_1MBNL/shCTRL$ fold change)

Protein	Fold change
NPC2	2.05
QSOX1	1.97
CTGF	1.77
LOXL2	1.62
APLP2	1.29

B

LOXL2 — MDA-231, CN34
QSOX1 — MDA-231, CN34

shCTRL, sh₁MBNL1, sh₂MBNL1

LOXL2 (relative protein) — MDA-231 / CN34: shCTRL, sh₁MBNL1, sh₂MBNL2

QSOX1 (relative protein) — MDA-231 / CN34: shCTRL, sh₁MBNL1, sh₂MBNL2

C

CTGF — relative CTGF protein concentration: shCTRL, sh₁MBNL1, sh₂MBNL1
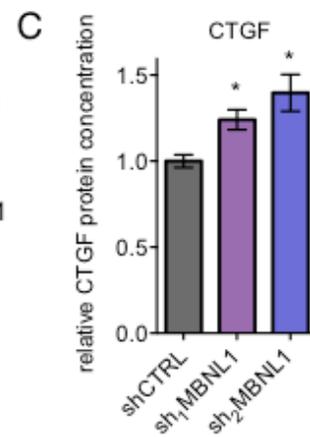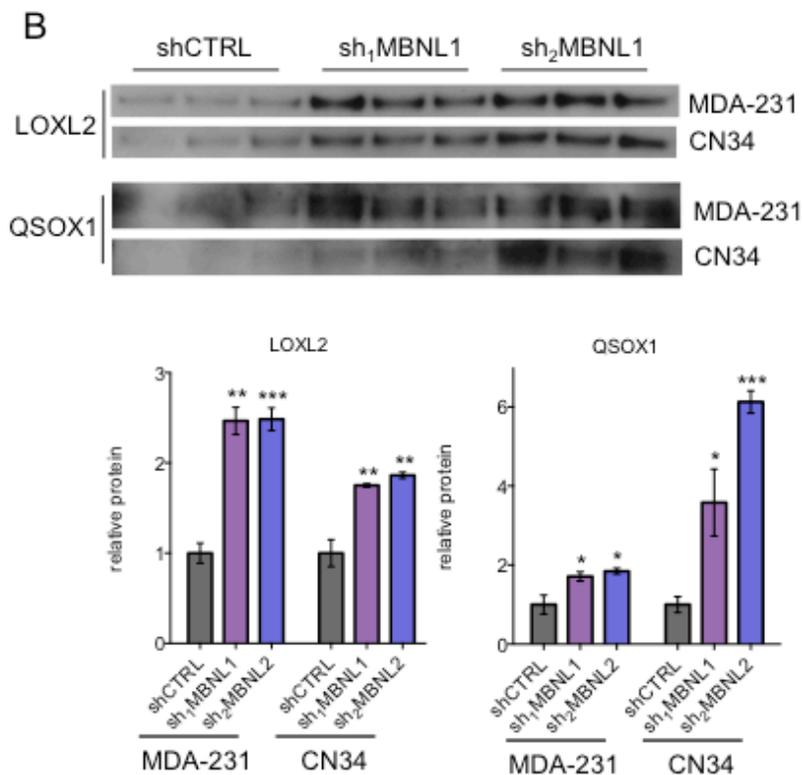
To validate the up-regulation of these proteins in conditioned media from MBNL1-knockdown cells compared to control cell conditioned media, Western blotting for LOXL2 and QSOX1 was carried out, while an ELISA was used to assay CTGF levels (Figure 4.7B,C).

To assess if these factors could affect the invasion capacity of cells, siRNA-mediated knockdown was used to deplete CTGF, LOXL2, and QSOX1 alone, as well as each of these transcripts in combination with the others. Depletion of any one of these factors alone did not affect the invasive capacity of breast cancer cells (Figure 4.8A). However, depletion of LOXL2 or QSOX1 in combination with depletion of CTGF led to a decrease in the invasion capacity of MDA-231 cells (Figure 4.8B). This data is consistent with a role for the secreted factors LOXL2, QSOX1 and CTGF in mediating the effect of MBNL1 on breast cancer cell invasion.
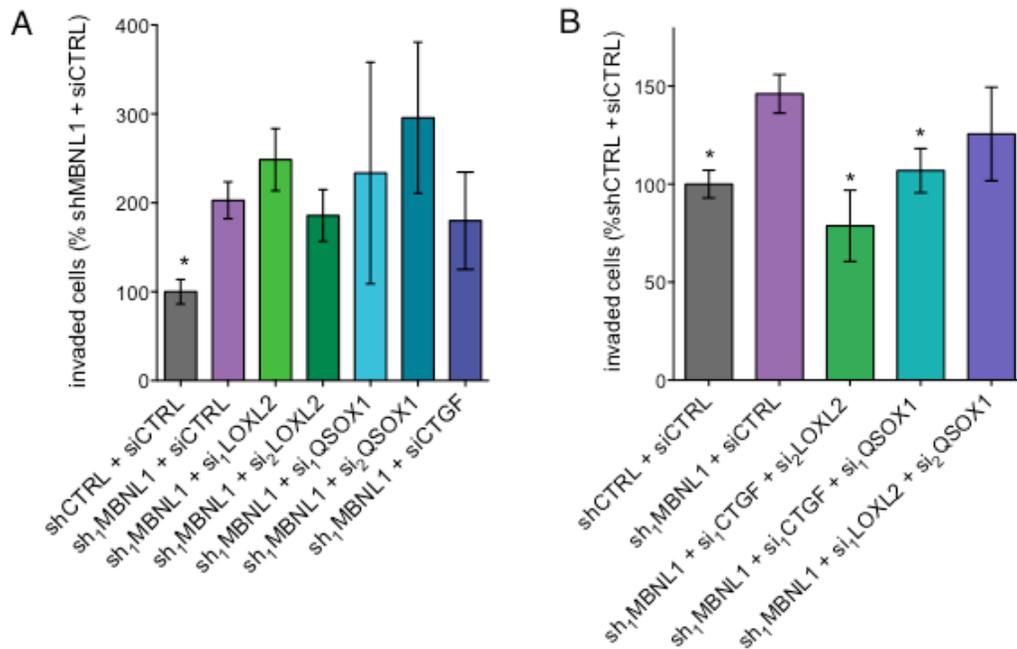
Figure 4.8 Effect of secreted proteins on *in vitro* invasion

(A) MDA-231 cells depleted of MBNL1 in combination with CTGF, LOXL2 or
QSOX1 were subjected to transwell invasion assays. N=5-6. (B) MDA-231 cells
depleted of MBNL1 in combination with the indicated factors were subjected to
transwell invasion assays. N=6. Data are shown as mean ±S.E.M.

CHAPTER 5. DISCUSSION

1. MBNL1 regulation of messenger RNA expression

This study identified MBNL1 as a suppressor of breast cancer progression, and discovered that part of this effect is mediated by MBNL1 regulation of DBNL and TACC1 transcript stability. Interestingly, factors secreted by MBNL1-depleted cells were found to moderately enhance cancer cell invasion. The mechanism through which these secreted factors are upregulated in MBNL1 depleted cells is presently unclear. MBNL1 was also found to directly bind and regulate a set of alternatively spliced transcripts in breast cancer cells. However, RNAi-mediated modulation of the levels of these transcript variants had no significant effect on the *in vitro* invasion capacity of breast cancer cells. It is possible that some of these MBNL1-dependant alternative transcript variants contribute to the regulation of metastasis, but may work in concert with each other so that modulation of a single splicing event does not have an effect on invasion. Therefore, the primary mediator of MBNL1-mediated breast cancer metastasis suppression identified in this study is the stabilization of a subset of transcripts by MBNL1 binding (Figure 5.1)
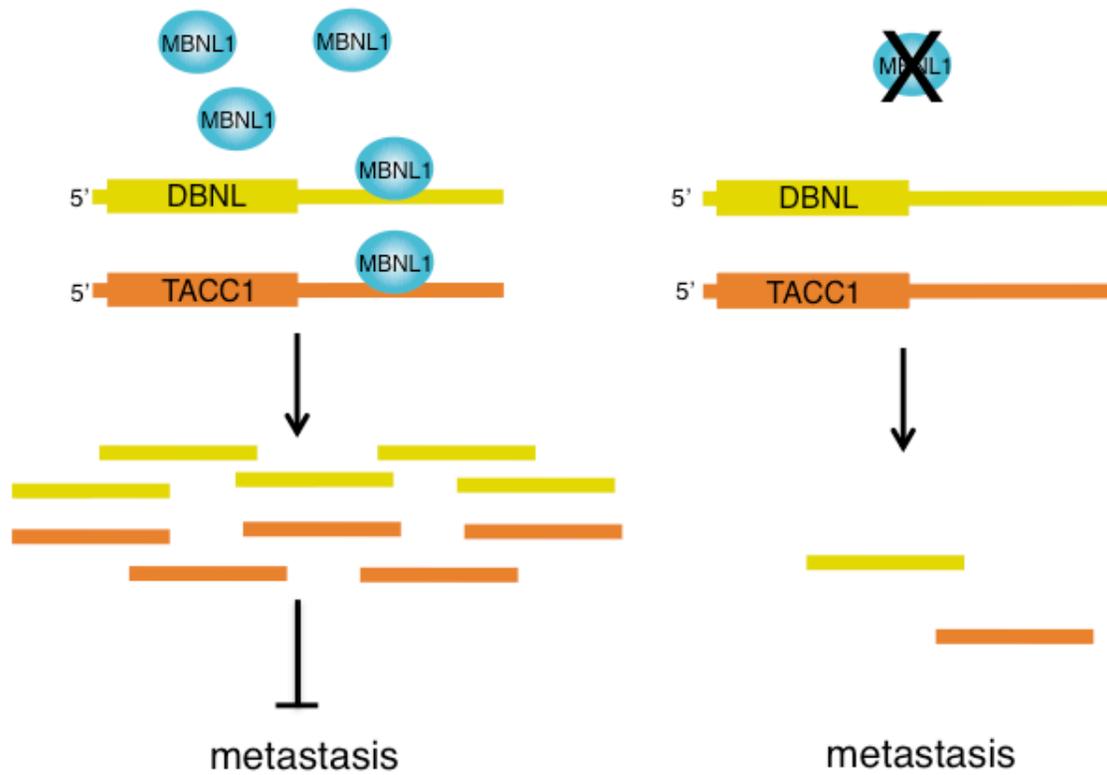
Figure 5.1 Model of MBNL1 regulation of transcript stability that contributes to breast cancer metastasis

Depiction of MBNL1 binding to the 3'UTRs of DBNL and TACC1 to increase their stability and thereby contribute the suppression of breast cancer cell invasion and metastasis.

An open question is the mechanism by which MBNL1 affects the stability of mRNAs in breast cancer cells. MBNL1 is known to affect pre-mRNA processing by contributing to both alternative exon and alternative poly(A) site selection. A recent study reported a regulatory role for MBNL proteins in polyadenylation site choice on 3'UTRs in mouse embryonic fibroblasts, thereby controlling 3'UTR length (Batra et al. 2014). However, in the current study, there

was no observed difference in the 3'UTR length of DBNL and TACC1 transcripts upon MBNL1 depletion, as assessed by qRT-PCR. There was also no observed change in the length of the poly(A) tails of DBNL and TACC1 transcripts when MBNL1 was depleted. These results suggest that APA and poly(A) tail lengthening are not mechanisms through which MBNL1 regulates the stability of these transcripts, although these regulatory mechanisms may be relevant to the control of the stability of other MBNL1-stabilized transcripts.

One possibility for the mode of MBNL1 regulation of mRNA stability that is consistent with this data is MBNL1-dependent localization of mRNAs to stress granules and/or P-bodies. MBNL1 has been observed to localize to arsenite-induced stress granules in HeLa and COS-7 cells (Onishi et al. 2008). Cancer cells undergoing the steps necessary for metastatic progression encounter stresses that are known to induce stress granule formation, including hypoxia and nutrient deprivation. It is possible that a function of MBNL1 in breast cancer cells is to localize transcripts to stress granules, where they would be protected from decay and then could rapidly resume translation upon removal of the stress. Therefore, when MBNL1 is depleted these transcripts would be subjected to increased degradation. Similarly, MBNL1 could bind transcripts in P-bodies to protect them from decay. This would be a mechanism similar to one observed to regulate the poly(A) tail length-independent stability of the HIF1A transcript via binding by PAN2 (Bett et al. 2013). In this model for MBNL1 stabilization of transcripts, MBNL1 would stabilize a subset of its mRNA targets by localizing

them to cytoplasmic RNA granules where they would be protected from decay factors.

2. Regulation of MBNL1 expression and activity

Little is known about the regulation of MBNL1 expression. The most direct data addressing this question comes from studies in *Drosophila*, where it was shown that MEF2 (myocyte enhancer factor 2) drives transcription of muscleblind, and tissue-specific enhancer elements of muscleblind expression were also identified, although these sequences are not conserved in humans (Artero et al. 1998; Bargiela et al. 2014). Regulation of MBNL1 levels and activity in cancer is a clinically relevant question because of MBNL1's role in suppressing breast cancer metastasis identified in this study. In breast cancer, there are many possibilities for regulation of MBNL1 transcript levels. These include decreased transcription of the gene through copy number loss at the *MBNL1* genetic locus, mutation or epigenetic silencing of promoter elements required for transcription of *MBNL1*, or through the loss-of-function of transcription factors that transcribe the *MBNL1* gene. MBNL1 could also be down-regulated at the mRNA level through the alteration of an as yet to be defined cis-regulatory element that controls MBNL1 transcript stability, or through levels of a trans-acting factor that controls MBNL1 transcript stability.

The clinical data available for examination in this study only contained data for MBNL1 transcript levels, and levels of MBNL1 mRNA do correlate with metastasis-free survival of breast cancer patients (Figure 2.3D). However,

MBNL1 protein loss-of-function is another formal possibility for down-regulation of MBNL1 activity. One mechanism for this would be the mutation of amino acids in the MBNL1 protein that affected either its RNA binding ability or its interactions with other proteins. Another interesting possibility is regulation of MBNL1 protein activity by a toxic RNA gain-of-function mechanism. In this model, cancer cells would have unchanged MBNL1 transcript and protein levels, but inhibition of MBNL1 activity would occur through sequestration of the MBNL1 protein by an RNA containing expanded CUG repeats, in a mechanism similar to that observed in myotonic dystrophy. It is plausible that a subset of cancer cells could acquire just such a trinucleotide repeat expansion, which would then lead to the accumulation of a transcript with a large number of CUG repeats, which could then bind MBNL1. Acquisition of expanded CAG trinucleotide repeats at multiple loci in breast cancer has been observed (Pizzi et al. 2007). Therefore, it is conceivable that such a mechanism could also generate CUG repeat expansions in breast cancers. This could lead to the suppression of MBNL1 activity in cancer cells and induce acquisition of enhanced metastatic capacity. Interestingly, a higher incidence of primary tumor development in individuals affected with myotonic dystrophy has been observed, and this increase seems to be specific to the disease itself, as unaffected family members do not have an increased cancer incidence (Gadalla et al. 2011; Lund et al. 2014; Win et al. 2012). There are no reports on the relative incidence of metastatic disease in individuals with myotonic dystrophy, but this would be an interesting area for future study. There

have also been multiple case reports that have examined samples from individuals with myotonic dystrophy who also had cancer of various types, and found an expansion of the CUG repeat in the *DMPK* gene in tumor cells compared to normal tissue (Akiyama et al. 2008; Banuls et al. 2004; Jinnai et al. 1999; Kinoshita et al. 1997; Kinoshita et al. 2002; Ogata et al. 1998; Osanai et al. 2000). This illustrates the fact that cancers that develop from a wide variety of tissues have the ability to generate additional expansions in CUG repeats, which could contribute to the down-regulation of MBNL1 and possibly increase the metastatic capacity of the cancer cells.

3. Roles of DBNL and TACC1 in cancer progression

In this study, expression of DBNL and TACC1 counteracted the MBNL1-depedent effects on cancer cell invasion and metastatic lung colonization. The cell biological processes affected by DBNL and TACC1 could conceivably mediate these phenotypes. A role for DBNL in suppressing breast cancer metastasis has not been previously reported. DBNL (drebrin-like protein) is an F-actin binding protein and has roles in mammalian development, endocytosis and the immune response (Connert et al. 2006; Haag et al. 2012; Hepper et al. 2012; Kessels et al. 2000; Larbolette et al. 1999; Schymeinsky et al. 2009). DBNL has an N-terminal F-actin binding domain and C-terminal proline rich and SH3 domains. The DBNL SH3 domain has been reported to interact with factors involved in endocytosis, cell motility, and neuronal morphology development (Fenster et al. 2003; Hou et al. 2003; Kessels et al. 2001; Pinyol et al. 2007). In a

mouse genetic knockout of DBNL, animals have adult onset organ abnormalities, including enlarged spleens, dilation and fibrosis of heart chambers, and lung emphysema. In adulthood, these animals also develop impaired motor skills, which may be due to the observation that they have impaired synaptic vesicle recycling of hippocampal boutons (Connert et al. 2006).

Although there is no reported role for DBNL in cancer progression, a recent study showed that DBNL depletion in Src-transformed fibroblasts regulates podosome rosette formation and, interestingly, increases cell invasiveness (Boateng et al. 2012). Although this effect was observed in transformed fibroblasts, this finding is consistent with decreased DBNL levels in breast cancer cells enhancing their invasive capacity. Another process relevant to metastasis, adhesion to blood vessel walls, has been shown to be affected by DBNL in neutrophils. Intriguingly, DBNL has been shown to promote firm neutrophil adhesion *in vitro* under physiologic shear stress conditions and to also promote adhesion and extravasation *in vivo*. Depletion of DBNL resulted in a reduced number of integrin-beta-2 clusters in high-affinity conformation. Furthermore, DBNL was necessary for neutrophil crawling under flow conditions (Hepper et al. 2012; Schymeinsky et al. 2009). Cancer cells go through a process similar to this during extravasation. Reduction in the level of DBNL in breast cancer cells, if consistent with the function of this protein in neutrophils, would decrease adhesion and intraluminal crawling of a cancer cell. This could allow a cancer cell to travel farther in the bloodstream and then only adhere to

endothelial cells when physically forced to, such as arrest in small capillaries. This decrease in adhesion might allow for a cancer cell to spread farther in the body.

TACC1 (transforming acidic coiled-coil containing protein 1) is a member of the TACC family, which also contains TACC2 and TACC3. All three of these proteins are characterized by the presence of a highly conserved, highly acidic 200 amino acid long coiled-coil domain at their C-termini (Still et al. 2004). Although the down-regulation of TACC1 has been previously reported in breast cancer (Conte et al. 2003), the role of this molecule in mediating this process remains undefined. TACC1 has been reported to associate with proteins involved in diverse cellular processes, including those involved in cell division, specifically mitotic spindle dynamics (chTOG/CKAP5, Aurora A, B and C), transcription (GAS41/YEATS4, thyroid hormone receptor, retinoid acid receptor alpha, FHL3), and RNA processing (LSM7, SNRPG, TDRD7) (Conte et al. 2003; Gabillard et al. 2011; Guyot et al. 2010; Lauffart et al. 2002; Lauffart et al. 2007). In adult humans, transcript levels of TACC1 are highest in the brain, placenta, skeletal muscle, spleen, testes, and uterus, and TACC1 levels are higher in the developing mouse embryo than in adult mouse tissue (Lauffart et al. 2006). No mouse genetic knockout model of TACC1 has been reported. Experiments to assess the localization of TACC1 in HeLa cells found that during mitosis, TACC1 is present at the spindle midzone during anaphase, is localized at the midbody

80

during cytokinesis, while during interphase TACC1 relocalizes to the nucleolus (Delaval et al. 2004).

TACC1 was originally identified as a molecule involved in breast cancer because the *TACC1* gene is located on 8p11, a chromosomal region commonly amplified in breast cancers. Furthermore, it was found that overexpression of TACC1 in mouse fibroblasts resulted in cellular transformation and anchorage independent growth, suggesting that TACC1 has a role in oncogenesis (Still et al. 1999). Additionally, ectopic overexpression of TACC1 in a mouse pten heterozygous background increased the number of mammary tumors formed (Cully et al. 2005). However, in contrast to this, TACC1 mRNA has been shown to be down-regulated in breast tumors compared to normal breast tissue, which supports a suppressive role for this molecule in breast cancer progression (Conte et al. 2003). In agreement with this, analysis of patient data in this current study revealed that TACC1 transcript levels are significantly correlated with relapse-free survival in breast cancer patients (Figure 4.2A). These data can be explained if TACC1 acts as both a promoter of tumorigenesis but a suppressor of cancer progression. Recently, rare but recurrent fusions of FGFR and TACC genes have been reported in glioblastoma multiforme. Fusions of FGFR1-TACC1 and FGFR3-TACC3 were observed in a small subset of glioblastomas, with the tyrosine kinase domain of the FGFR gene fused to the acidic coiled-coil domains of the TACC gene. The resulting fusion proteins were found to promote tumor growth (Singh et al. 2012).

Taken together, the literature on TACC1 is not in agreement on a single function for this protein. TACC1 may be involved in MAPK signaling pathways by regulating the subcellular localization of phospho-ERK. Through its interactions with YEATS4/GAS41, a member of the NuA4 acetyltransferase complex, TACC1 may also modulate histone acetylation and transcription. It has been implicated as a regulator of mitotic exit through its interactions with Aurora B at the midbody during cytokinesis, and as a regulator of RNA processing through its interactions with LSM7 and SNRPG. The interaction of TACC1 with these RNA processing factors is especially intriguing given the role of MBNL1 in RNA processing. LSM7 is a component of both the LSM1-7 complex and the LSM2-8 complex, which mediate cytoplasmic mRNA decapping, and nuclear pre-mRNA splicing and decay, respectively (Tharun 2009). SNRPG (small nuclear ribonucleoprotein G) is a protein component of the snRNPs that make up the core spliceosome machinery. It is therefore possible that downregulation of TACC1 by MBNL1 results in the dysregulation of splicing and stability of transcripts not directly bound by MBNL1. It would also be interesting to test the effect of DBNL and TACC1 on modulating the levels of the secreted factors that are upregulated in MBNL1-knockdown cells. This might reveal a mechanistic link between MBNL1 and these downstream secreted factors.

4. MBNL1-dependent phenotypes and cancer progression

In this study, MBNL1 was found to suppress the invasion and trans-endothelial migration capacity of breast cancer cells. Both of these processes

occur during metastatic progression, and programmed cell invasion is also critical during normal development. In fact, MBNL1 has been reported to negatively regulate endocardial cell invasion during chick heart development (Vajda et al. 2009). This illustrates that MBNL1 suppresses cell invasion during specific normal developmental processes, and suggests breast cancer cells may be commandeering this developmental function of MBNL1 to enhance their invasive capacity. Another intriguing finding from the current study is that MBNL1 depletion results in a significant reduction in cell growth, both in culture and in primary tumor growth assays (Figure 2.4A,B). This is consistent with MBNL1 acting as a promoter of primary tumor growth but a suppressor of metastasis. MBNL1 levels are significantly correlated with distant metastasis-free survival in breast cancer. Therefore, even if MBNL1 acts to promote primary tumor growth, it could be associated with better clinical outcomes in breast cancer because it suppresses cancer metastasis to distant organs. This finding is consistent with a study that analyzed the outcomes of breast cancer patients that had node positive disease and had very small or large primary breast tumors. Patients with very small primary tumors had significantly increased breast cancer specific mortality compared to those with large tumors (Wo et al. 2011). Therefore, MBNL1 may act to suppress cancer cell invasion while simultaneously acting to increase growth of the primary tumor. Given this, it is possible that down-regulation of MBNL1 in small tumors might be an indicator of cancer aggressiveness. It is also interesting to note that generally, MBNL1 acts as a

promoter of tissue differentiation. Depletion of MBNL1 and MBNL2 leads to a

reversion to embryonic cell-like splicing patterns, and MBNL1 depletion leads to

an increased efficiency of induced pluripotent stem cell reprograming (Han et al.

2013; Lin et al. 2006). MBNL1 is also required for terminal differentiation of red

blood cells (Cheng et al. 2014). Together, these data are consistent with MBNL1

acting as a promoter of tissue differentiation. Therefore, it is interesting that in

breast cancer cells, depletion of MBNL1 leads to an increase in metastatic

capacity, as plasticity of gene expression can be advantageous during metastatic

progression.

CHAPTER 6. METHODS AND MATERIALS

Animal Experiments

All mouse studies were conducted according to a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the Rockefeller University. For tail-vein colonization assays, 6-8 week old age matched female NOD-scid mice were used. Cancer cells were engineered to stably overexpress luciferase (Ponomarev et al. 2004), and these cells were injected into the tail vein of each mouse. For systemic metastasis assays, 8 week old female athymic mice were used, and cancer cells were injected into arterial circulation. The location and number of cancer cells was monitored by bioluminescence in living animals over times. Statistical significance was based on a one-tailed Mann-Whitney test. *In vivo* tumor growth assays were performed by injecting 5×10^5 cells mixed with 100ul of matrigel into bilateral mammary fat pats of 7 week old female NOD-Scid mice. Tumor volume was assessed by caliper measurements.

Cell Culture

All cell lines were maintained at 37°C, 5% CO_2. The 293LTV and MDA-MB-231 and derivative cell lines were cultured in Dulbecco's modified eagle medium supplemented with 10% fetal bovine serum, 2mM L-glutamine, 1mM sodium pyruvate, 100 units/mL penicillin, 100 ug/mL streptomycin, and amphotericin B (base media and supplements all Life Technologies). HUVECs were obtained from ATCC and maintained in EGM-2 media (CC-3162 Lonza) supplemented with 2% fetal bovine serum.

RNAi-Mediated Gene Knockdown

Lentivirus was produced by transfecting 293LTV cells that had been seeded in 6-well plates and allowed to become 60% confluent with 2ug pLKO.1 shRNA containing vector, 2 ug vector psPAX2 and 1 ug vector pMD2.G. TransIT-293T (Mirus Bio) was used for transfection according to the manufacturer's protocol. Virus was harvested at 48 hours post-transfection and passed through a 0.45um syringe filter to remove 293T cells.

Cancer cells were transduced by incubation with the lentivirus for 6 hours in the presence of 8ug/mL polybrene. After transduction, media was changed to normal growth media. Selection was started 48 hours after transduction by adding puromycin to a final concentration of 1.5ug/ml. Cell were kept under selection for 48 hours.

For siRNA mediated gene knockdown, cells were seeded at 1.5×10^5 per well in 6-well plates. 20 hours after seeding, cells were transfected with 90 pmol of siRNA using lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions.

Sequences targeted by shRNAs:

sh_1MBNL1: 5'-GCCAACCAGAUACCCAUAAUA

sh_2MBNL1: 5'-GCCUGCUUUGAUUCAUUGAAA

sh_1TACC1: 5'-CCACGUCAUGUGGUCAGAAAU

sh_2TACC1: 5'-GAAGGCAAAGUCGCGUUUAAU

shCTRL: 5'-CAACAAGAUGAAGAGCACCAA

Sequences targeted by siRNAs:

siCTGF: 5'-UACCAGCAGAAAGGUUAGUAUCAUCAG

siFGFR1: 5'-ACACACCACCUACUUCUCCGUCAAUGU

si$_1$LOXL2: 5'-CUGGAGCAGCACCAAGAGCCAGUCUUG

si$_2$LOXL2: 5'-UUGGAGGACACAGAAUGUGAAGGAGAC

siMYL6-SE: 5'-AAGCGUUUGUGAGGCAUAUCCUGUCGG

siPICALM-SE: 5'-AACAAUGAAUGGCAUGCAUUUUCCACA

si$_1$QSOX1: 5'-AAGCAACAUCAUCCUGGACUUCCCUGC

si$_2$QSOX1: 5'-CUGGAGGAGAUUGAUGGAUUCUUUGCG

siSEC31A-SE: 5'-CAGCCGUAUCCCUUCGGAACAGGGG

siCTRL: 5'-CGUUAAUCGCGUAUAAUACGCGUAU

Retroviral and Lentiviral Mediated Protein Overexpression

Retrovirus was produced by transfecting 293T cells seeded in 6-well plates with 2 ug of pBabe containing the ORF of interest, 2 ug Gag/Pol and 1 ug VSV-G. Transfection was carried out using transIT-293T (Mirus Bio) according to the manufacturer's protocol. Virus was harvested 48 hours post-transfection and passed through a 0.45um syringe filter to remove cells. For carboxy terminal flag-tagged MBNL1 overexpression, target cells were transduced by incubation with the retrovirus for 6 hours in the presence of 8 ug/mL polybrene. At this time, media was changed to normal growth media. Selection was started 48 hours after transduction by adding puromycin to a final concentration of 1.5 ug/mL or hygromycin to a final concentration of 350 ug/mL. Puromycin selection was

maintained for 48 hours, while hygromycin selection was carried out for 11-14 days.

Lentivirus was produced by transfecting 293LTV cells seeded in 6-well plates with 2 ug pLX304 containing the ORF of interest with a carboxy-terminal V5 tag, 2 ug vector psPAX2 and 1 ug vector pMD2.G. TransIT-293T (Mirus Bio) was used to transfect the DNA per the manufacturer's protocol. Virus was harvested at 48 hours post-transfection and passed through a 0.45um syringe filter to remove 293T cells. For c-terminal V5-tagged DBNL and TACC1 overexpression, target cancer cells were transduced by incubation with the lentivirus for 6 hours in the presence of 8 ug/mL polybrene. After transduction, media was changed to normal growth media. Selection was started 48 hours after transduction by adding puromycin to a final concentration of 1.5 ug/ml. Cells were typically kept under selection for 48 hours, or until cells that had not been transduced as a control were all dead.

Transcriptomic Sequencing and Alternative Splicing Analysis

Whole transcriptome sequencing libraries were constructed using the ScriptSeq v2 kit (Epicentre) per the manufacturer's instructions. The input RNA was isolated with a spin column based kit, including an on-column DNase treatment per the manufacturer's instructions (Norgen), and then depleted of rRNA using the Ribo-Zero kit (Epicentre). The libraries were sequenced at the Rockefeller University Genomics Resource center on a HiSeq2000 (Illumina). Reads were first trimmed to remove linker sequences and low-quality bases, using cutadapt (v1.2.1).

Tophat2 (v2.0.8) was then used to map the reads to the human transcriptome (RefSeq transcriptome index, hg19). Cufflinks (v2.0.2) was then used to estimate RPKM values and compare shControl and shMBNL1 samples.

For alternative splicing analysis, mapped reads from each of the two MBNL1-targeting shRNA expressing MDA-231 cell lines were compared to the shCTRL expressing cell line using MISO (v0.4.9) (Katz et al. 2010), in conjunction with the provided annotations for skipped exons, to quantitate modulations in alternative splicing events.

RT-PCR analysis of MISO-predicted transcript variants

Semi-quantitative real-time PCR was carried out by first isolating RNA and synthesizing cDNA. This cDNA was used as the template for PCR to assess the relative abundance of transcript isoforms. The PCR products were resolved on 1.5% or 2% agarose-TAE gels, as determined by the product sizes to be separated. The PCR products were then visualized by ethidium bromide staining. Quantitation of the bands was done using Image J software (NIH).

Table 6.1 Primers used for RT-PCR analysis of MISO-predicted skipped exons

AKAP8L_SE_F	TGG AAC TTG GAA CTC TGG GA
AKAP8L_SE_R	ACT CGG GGA TGA TGT TCT GG
AMOTL1_SE_F	GTG AAT GGG GTT GAT TGT CCG
AMOTL1_SE_R	AGG AAG TTT GGG GAG TGG AA
CNOT10_SE_F	GCG AAA GCA GTG AAA CTT GC
CNOT10_SE_R	GTC TGA TCC TTG GTC CTG CT
DBNL_SE_F	AGT GAA GGA CCC CAA CTC TG
DBNL_SE_R	ACG GCA TTG GTC TTC TGG TA
DCTN1_SE_F	CTG CAA GAA GAT CCG AAG GC
DCTN1_SE_R	CTT CGA GCT TCA AAC CCA GG
FGFR1_SE_F	AAG GAG GAT CGA GCT CAC TG
FGFR1_SE_R	CCA ATA TGG AGC TAC GGG GT
HDLBP_SE_F	GCT GTG GAG AGG CTA GAA GT
HDLBP_SE_R	TCT GCC CTT TCT TGC CAA TG
HEATR2_SE_F	AGG ATT CGA AGA TGA CGC GA
HEATR2_SE_R	CAC AAA CTC TGG CTC AGG TG
INCENP_SE_F	CAG TGC AGA GGA ACC AGA TG
INCENP_SE_R	GAT TCT CCA GGC GCT GC
MTMR12_SE_F	AAA AGG AGT GGG TCA TGG GT
MTMR12_SE_R	TCG TTG ATG TTT GAA GCG CA

MTX2_SE_F	CTC GTT AAC TGC CGA GAG C
MTX2_SE_R	GCA TTT TCA GGC CAA GGT TC
MYL6_SE_F	AGA CAG TGG CCA AGA ACA AG
MYL6_SE_R	ATT CAC ACA GGG AAA GGC AC
PICALM_SE_F	ACC AAC AAC CGC TTG GAA TG
PICALM_SE_R	AAA GGG GTT TGG AGG TCT CA
PTPRF_SE_F	CGA AGA CCA ACA GCA CAA GG
PTPRF_SE_R	GCA CGT AGG TGA CCT GGT AG
RELA_SE_F	AGC ACA GAT ACC ACC AAG AC
RELA_SE_R	CTT GGA AGG GGT TGT TGT TG
SEC31A_SE_F	TCC GGG TTT CAT AAT GCA TGG
SEC31A_SE_R	TGT TCA AAG CTG GAG GGT CA
TCF3_SE_F	GAC GGG GGT CTC CAC G
TCF3_SE_R	CAC TGT AGG AGT CGG GAG G
ZNF512_SE_F	GAT GTC TTC CAG ACT CGG TG
ZNF512_SE_R	TCG ACA TGA GAA GTA GCA GC

Quantitative Real-Time PCR

To carry out quantitative real time PCR (qRT-PCR), RNA was isolated using a total RNA isolation kit, including an on-column DNase treatment (Norgen). Synthesis of complementary DNA (cDNA) was carried out using the superscript III reverse transcriptase kit (Life Technologies) using an oligodT for priming unless otherwise specified. For each cDNA synthesis reaction, an equal mass of total RNA reverse transcribed for all samples to be compared. Typically 1-2ug per 20ul reaction was used. Synthesis was carried out per the manufactuer's instructions, including an RNase H digestion step. qRT-PCR was carried out in technical quadruplicates, using fast syber green master mix (Applied Biosystems) and fluorescence was monitored using a 7900HT Fast Real Time instrument (Applied Biosystems). Data were analyzed using the delta delta Ct method. Endogenous control transcript used for normalization are indicated in each figure.

Primers used for qRT-PCR

MBNL1: 5'-CTGCCCAATACCAGGTCAAC/5'-GGGGAAGTACAGCTTGAGGA

MBNL2: 5'-TGCCCAGCAGATGCAATTTA/5'-GGACCTACAGGGAAAGTGGG

DBNL: 5'-AAGGCTTCAGGTGCCAACTA/5'-GACACGGCATTGGTCTTCTG

TACC1: 5'-GCCTCAGCGAATCAGACAAG/5'-TGCCGGGTCTCTTCGTATTT

mCherry: 5'-CCTGTCCCCTCAGTTCATGT/5'-GTCCTCGAAGTTCATCACGC

TACC1-3UTR-1: 5'-GCCCTGGAAGAAACCCTAGA/

5'-AACCCAAACTCAGCAGCCTA

TACC1-3UTR-2: 5'-GGCCATTAACCCCAACATGG/

5'-TGCATGGATTTGGGTTTGCC

TACC1-3UTR-3: 5'-CCCTTAAGAACCTGACCCCA/

5'-TCCTCATAACGGTCATGGCT

TACC1-3UTR-4: 5'-TCCAGCCAGTTACCCTTTCA/

5'-GGGCAGTTTACACTCCCTGT

TACC1-3UTR-5: 5'-GAAGACCCATCCCCTAGTGC/

5'-GCATGCTAAGAGGCACAGAA

DBNL-3UTR-1: 5'-CCCCTCTCAGACATGGCTTC/

5'-CACTGGGGGTCCTATTCCTG

DBNL-3UTR-2: 5'-TTGGCAGCAGGGAATTTGTC/

5'-TGCATCCCCACTTCCCATAG

HPRT1: 5'-ATGACCAGTCAACAGGGGAC/5'-CTGCATTGTTTTGCCAGTGTC

18S: 5'-AGCGAAAGCATTTGCCAAGA/5'-TATGGTCGGAACTACGACGG

Matrigel Invasion Assay

First, to induce invasiveness, cancer cells were serum-starved in media

containing 0.2% FBS for 20 hours. Cancer cells with siRNA-induced gene

knockdown were starved starting at 48 hours post siRNA transfection. The

starved cancer cells were seeded at 5 x 10^4 cells per well in matrigel coated

invasion chambers with 0.8um pore size (BD Biosciences) that had been pre-

equilibrated in 0.2% FBS containing media. The cells were incubated for 20

hours at 37°C. After the cells had been allowed to invade for 20 hours, the

matrigel-coated inserts were washed with PBS. Using a q-tip the cells on the top

side of each insert were scraped off. The inserts were fixed in 4%

paraformaldehyde for 15 minutes at 37°C, then cut out and mounted onto slides

using vectashield with DAPI (Vector Laboratories). The number of cells invaded

through the matrigel was quantified by imaging each insert using an inverted

fluorescence microscope (Zeiss Axiovert 40 CFL). Five images were taken per

insert at 10x magnification. ImageJ (NIH) was used to quantify the number of

invaded cells.

Trans-Endothelial Migration Assay

First, monolayers of HUVECs were prepared by seeding $5x10^4$ HUVECs in EGM-

2 media on collagen coated HTS Fluoroblok with 0.3um pores transwell inserts

(Corning). After monolayer formation, cancer cells that had been serum-starved

in media containing 0.2% FBS for 20 hours and pulse labeled with cell tracker

green CMFDA dye (Life Technologies) were seeded on top of the HUVEC

monolayers at 5×10^4 cancer cells per well, in EGM2 media containing 0.2%

FBS. After 12 hours, the number of cancer cells migrating through the HUVEC

monolayer was quantified by imaging each insert using an inverted fluorescence

microscope (Zeiss Axiovert 40 CFL). Four images were taken per insert and

ImageJ (NIH) was used to quantify the number of cell migrated through the

HUVEC monolayer.

Cell Proliferation Assay

To assay the effect of MBNL1 knockdown on cell proliferation, 2.5×10^4 cells were seeded in 6-well plates in triplicate. After 5 days, the number of viable cells was assessed using trypan blue, which is excluded from living cells.

Western Blotting

Whole cell lysate was prepared by washing trypsinized cells 2x with cold PBS, resuspending the cell pellet in ice-cold RIPA buffer (25mM Tris-HCl pH 7.6, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) with 1x complete protease inhibitor cocktail (Roche) and incubating on ice for 20 minutes. The lysate was sonicated to shear DNA and then cleared by spinning at 4°C in a microcentrifuge at maximum speed for 15 minutes. The total protein concentration of the resulting cleared lysate was determined using the BCA assay method (Thermo Scientific). Equal amounts of total protein were run on 4-12% bis-tris NuPAGE gels in MOPS buffer (Life Technologies) under reducing conditions. Proteins were transferred to 0.2um PVDF (Millipore) and blocked with 5% nonfat milk in PBS. Antibodies used in this study include custom made polyclonal rabbit anti-MBNL1 (Yenzyme, raised to aa 363-378 of NP_066368), anti-flag (Sigma F1804), anti-DBNL (Abcam ab86708) anti-TACC1 (Abcam ab187358), anti-QSOX1 (Sigma SAB2700031), anti-LOXL2 (Abcam ab113201) . The blots were incubated with horseradish peroxidase conjugated secondary antibodies and developed using the ECL method (Thermo Scientific).

High-Throughput Sequencing and Crosslinking Immunoprecipitation

(HITS-CLIP)

HITS-CLIP for endogenous MBNL1 was carried out as previously described

(Jensen and Darnell, 2008), with some modifications, as described in detail

below.

Crosslinking of cells

For each immunoprecipitation, ~$5x10^6$ cells were used. MDA-231 cells were

grown to sub-confluence before crosslinking and harvesting. Cells were rinsed 1x

with 5 mL cold PBS, the wash aspirated, then an additional 2 mL cold PBS was

added to just cover the cells. A Biorad genelinker was used to irradiate each

plate with 400mJ/cm^2 UV (254nm) to crosslink RNA to proteins. Plates were

placed on ice and collected into a falcon tube by scraping. Cells were pelleted at

2000 rpm for 5 minutes at 4°C, and the PBS aspirated. Cells were resuspended

in 1mL cold PBS and transferred to a microcentrifuge tube. Cells were pelleted at

14,000 x g for 30 seconds at 4°C, and the PBS aspirated. The washed,

crosslinked cell pellets were immediately placed at -80°C.

Immunoprecipitation

For each immunoprecipitation,100ul of beads were used. Antibody-conjugated

beads were prepared by washing protein A dynabeads (Life Technologies) 1x

with PBS plus 0.02% tween-20, then resuspending beads in PBS plus 0.02%

tween-20 to 2x the original bead volume. To each 100ul of washed beads, either

15ug of rabbit IgG (Cell Signaling Technology) or 25ul of anti-MBNL1 antibody

(0.6mg/mL)(Yenzyme) was added. The beads and antibodies were rotated at room temperature for 30 minutes, then washed 3x with PBS plus 0.02% tween.

The crosslinked and frozen cell pellets were resuspend in 800uL cold low salt wash buffer (1x phosphate buffered saline (PBS), 0.1% Sodium Dodecyl Sulfate (SDS), 0.5% sodium deoxycholate, 0.5% IGEPAL CA-630) + 15ul RNasin RNase inhibitor (40 units/ul, Promega) + 1x protease inhibitor cocktail (Roche EDTA-free protease inhibitor tablets). To digest DNA and reduce viscosity of the lysate, 30ul DNaseI RQ1(1 unit/ul, Promega) was added to the lysate and incubated at 37°C for 5 minutes at 1000 rpm in a thermomixer. Lysate was treated with high and low concentrations of RNase A. For the low RNase treatment, RNase A (USB) was added to a final concentration of 30 pg/ul, while for the high RNase treatment, RNAse A was added to a final concentration of 1500 pg/ul. After RNase addition, tubes were incubated at 37°C for 5 minutes, then immediately placed on ice. Cell debris was then cleared from the lysate by spinning at 16,000xg at 4°C for 20 minutes. The cleared supernatant was then pipetted into clean tubes. The antibody-conjugated dynabeads were then added to the cleared lysate, and the mixture rotated end-over-end at 4°C for 2 hours. After the immunoprecipitation, the beads were collected on a magnet and the supernatant removed. The beads were then washed in the following order (all washes were ice cold): 2x low salt wash buffer, 2x high salt wash buffer (5X PBS, 0.1% SDS, 0.5% sodium deoxycholate, 0.5% IGEPAL CA-630), and finally 2x PNK buffer (50mM Tris-HCl pH 7.4, 10mM $MgCl_2$, 0.5% IGEPAL CA-630).

Labeling and gel resolution of MBNL1-RNA complexes

In order to visualize the MBNL1 protein-RNA complexes, a radioactively labeled RNA oligo was ligated to the immunoprecipitated protein-RNA complexes. To prepare the labeled oligo the 3' RNA linker oligo 5'-GUG UCA GUC ACU UCC AGC GG-puromycin was 5' ^{32}P labeled using T4 PNK (NEB) and **γ-**^{32}P ATP (Perkin Elmer). The 3' puromycin is added to the oligo to block unwanted ligation products at later steps in the protocol. After end labeling, excess nucleotides were removed from the reaction by running it through a G-25 sephadex spin column (GE life sciences). To prepare the RNA in the immunoprecipitated protein-RNA complexes for ligation, the immunoprecipitation was treated with calf alkaline intestinal phosphatase (CIP) (Roche) to dephosphorylate the RNA ends. After CIP treatment, the immunoprecipitations were washed in the following order, (all washes were ice cold): 1x with PNK buffer, 1x with PNK + EGTA buffer (50mM Tris-HCl pH 7.4, 20mM EGTA, 0.5% IGEPAL CA-630), 2x with PNK buffer. The ^{32}P-labled RNA oligo was then ligated to the protein complexes by adding the following to the beads after removing all of the last wash from the previous step: 5ul 10x T4 RNA ligase buffer, 5ul BSA (1mg/mL), 5ul 10mM ATP, 1.25ul T4 RNA ligase (5 units/ul) (Fermentas), 1.25ul RNasin, 6.25ul ^{32}P-labeled 3' linker oligo (7.5 pmol total), H$_2$O to 50ul. Reactions were incubated at 16°C for 1 hour in a thermomixer with shaking at 1000 rpm for 30 seconds every 15 minutes. After 1 hour, 48 pmol of the 3' linker oligo (5' phosphorylated with cold

ATP) was added to the ligation mixture and the mixture was incubated at 16°C for 16 hours.

After ligation of the 3' linker oligo, the immunoprecipitations were washed in the following order (all washes were ice cold): 1x with low salt wash buffer, 1x with high salt wash buffer, 3x with PNK buffer. The 5' ends of the protein-RNA complexes were then phosphorylated using T4 PNK, and then washed 3X with ice cold PNK buffer. These labeled protein-RNA complexes were then eluted from the beads by resuspending the beads in 40ul of 1x NuPAGE loading buffer (Life Technologies) and heating at 70 deg C in a thermomixer at 1000 rpm for 10 minutes. The beads were immediately collected on a magnet, and the eluted complexes were run on a 4-12% bis-tris NuPAGE gel (Life Technologies) in 1x MOPS running buffer (Life Technologies) at 150 volts in the cold room. The complexes were then transferred onto BA-85 nitrocellulose (Whatman) in cold 1x NuPAGE transfer buffer with 10% methanol at 100 volts in the cold room. To visualize the RNA-protein complexes, the membrane was then exposed to film and the film was developed.

In order to size select the RNA to make the HITS-CLIP library, the bands migrating at ~55-70 kDa in the low RNase treated lanes were cut from the nitrocellulose. This is because MBNL1 alone migrates at ~40 kDa, and the average molecular weight of 50 nucleotide long RNA, the minimum size for informative sequencing analysis, is ~15 kDa. The nitrocellulose was cut into smaller pieces and treated with proteinase K and then phenol chloroform

extracted to digest the protein and isolate the RNA. The aqueous layer was precipitated by adding ethanol:isopropanol (1:1), glycoblue, and sodium acetate, then incubating overnight at -20°C. The RNA pellet was washed 2x with 75% ethanol then air dried, then resuspended in 5.9ul H_2O.

<u>5' linker ligation, cDNA synthesis and PCR amplification MBNL1 bound RNA</u>

A 5' linker RNA oligo was then ligated to the isolated RNA by adding the following components to the RNA: 1ul 10x T4 RNA ligase buffer, 1ul BSA (0.2mg/mL), 1ul 10mM ATP, 0.1ul T4 RNA ligase, 1ul of the 5' linker RNA oligo at 20 pmol/ul (5'-AGGGAGGACGAUGCGG). The ligation mixture was incubated at 16°C overnight in a PCR tube in a thermocycler. The ligation reaction was then treated with DNase to remove any residual DNA. This mixture was then phenol chloroform extracted and the RNA recovered from the aqueous layer by ethanol:isopropanol precipitation as above. The RNA pellet was washed 2x with 75% ethanol and then air dried. The RNA pellet was resuspended in 10ul H_2O.

From this RNA, cDNA was synthesized by adding the following to 8ul of the resuspended RNA: 2ul DP3 primer at 5pmol/ul (5'-CCGCTGGAAGTGACTGACAC), 3ul 3mM dNTPs. As a control for DNA contamination, a parallel cDNA synthesis without the reverse transcriptase was done with the 2ul remaining of the resuspended RNA. The mixtures were incubated at 65°C for 5 minutes, then at 4°C for 1 minute. To this, the following was added: 1ul 0.1 M DTT, 4ul 5x superscript reverse transcriptase buffer, 1ul RNasin, 1ul superscript III (Life Technologies) or 1ul H_2O (for the no reverse

transcriptase control reactions). The reactions were then incubated at 50°C for 45 minutes, then 55°C for 15 minutes, then 90°C for 5 minutes, then placed on ice.

This cDNA was then amplified by PCR using accuprime Pfx supermix (Life Technologies and the primer pair: DP5: 5'-AGGGAGGACGATGCGG/ DP3: 5'-CCGCTGGAAGTGACTGACAC. The optimal number of cycles was determined by running 22-30 cycles of the follow program: 1. 95°C 2 min 2. 95°C 20 sec 3. 58°C 30 sec 4. 68°C 20 sec. repeat 2-4 for 22-20 cycles 5. 68°C 5 min. Formamide loading dye (2x solution: 95% deionized formamide, 5% 100mM EDTA pH 7.5, bromophenol blue for tracking) was added to the resulting PCR products were run on a 10% TBE-polyacrylamide urea gel and visualized by post-staining with SYBR gold (Life Technologies). Amplisize molecular ruler (Biorad) was run alongside the PCR products as a size standard. PCR products running at ~90-140 basepairs were cut from the gel and the DNA isolated from the gel slice using the Qiagen gel extraction kit per instructions for extracting DNA fragments from polyacrylamide gels. DNA was eluted from the spin column with 30ul EB.

Next, this eluted DNA was further PCR amplified and Illumina sequencing compatible sequences added. To each PCR tube, the following was added: 27ul Accuprime Pfx supermix (Life Technologies), 0.5ul 20uM DSFP5 (5'-AATGATACGGCGACCACCGACTATGGATACTTAGTCAGGGAGGACGATGCGG), 0.5ul 20uM DSFP3(5'-CAAGCAGAAGACGGCATACGACCGCTGGAAGTGACTGACAC), 3ul of DNA

isolated in previous step. The reactions were then cycled according to the following program: 1. 95°C 2 min 2. 95°C 20 sec 3. 58°C 30 sec 4. 68°C 40 sec. repeat 2-4 for 6-10 cycles 5. 68°C 5 min. The products were run on a 2% metaphor agarose-TBE gel and visualized by ethidium bromide staining. PCR products running between 150-170 basepairs were gel extracted using the Qiagen gel extraction kit. DNA was eluted from the spin column with 30ul EB.

<u>Library sequencing and computational analysis</u>

The biological duplicate MBNL1 HITS-CLIP libraries were sequenced at the Rockefeller University Genomics Resource Center on an Illumina HiSeq 2000 instrument with 50 bp single-end run length. The sequencing reads were first trimmed to remove linker sequences and low-quality bases, using cutadapt (v1.2.1) with parameters -q 15 and -m 25. The Resulting reads were then aligned to the human genome (build hg19) using bowtie2 (v2.1.0). The analytical package CIMS (Zhang et al. 2011) was then used to identify CLIP peaks in each of the biological replicates (FDR<10%). The peaks from the two replicates were then overlapped (intersectBed) to generate a high-confidence list of MBNL1 binding sites. Sequences were extracted and analyzed for motif discovery using ChIPseeqer (Giannopoulou and Elemento 2011). Briefly, a randomized background set was generated using a 1st-order Markov Model to control for length and dinucleotide frequency in the real binding sites. The resulting 'real' and 'scrambled' sequences were then analyzed using FIRE (Elemento et al. 2007) to identify the best representation of the MBNL1 binding site.

Transcript Stability Assays

For the alpha-amanitin microarray data, MDA-231 MBNL1 knockdown or control cells were treated with 10ug/mL alpha-amanitin (Sigma). Nine hours after alpha-amanitin addition, RNA was isolated from the cells using a total RNA isolation kit with on-column DNase treatment (Norgen). RNA was labeled using the TargetAmp nano labeling kit according to the manufactuer's protocol (Epicentre), and hybridized to Illumina beadchip arrays. Lumi package (R) was used to normalize and compare Illumina HT12 Beadchip raw signals for shControl and sh_1MBNL1 samples at 0 and 9 hours post alpha-amanitin treatment. The difference between the sh_1MBNBL1/shControl logFCs were used as a measure of stability. To discretize the input values, transcripts with logFC difference of less than -0.1 were labeled as "destabilized". The computational tool TEISER, in non-discovery mode, was then used to assess the stability of MBNL1-bound transcripts in this dataset (as described in Goodarzi et al. 2014).

Cells were seeded at $2x10^5$ per well in 6-well plates. 18 hours after seeding, 5,6- Dichlorobenzimidazole 1-beta-D-ribofuranoside (DRB)(Sigma) was added to the cells to a final concentration of 100μM. RNA was isolated at 0, 2, 4, 6 and 8 hours after DRB addition using a total RNA isolation kit with on-column DNase treatment (Norgen). Relative levels of the transcripts of interest were assessed by qRT-PCR, using 18S as the endogenous control. Half-life calculations were done using the formula $t_{1/2}=ln2/k_{decay}$, where the decay constant

was determined by plotting the data on a semilog scale and using non-linear regression to find the best fit line (Graphpad Prism).

Poly(A) Tail Length Assay

Length of the poly(A) tails was measured as described in (Janicke et al. 2012). Total RNA was isolated using a spin-column based total RNA extraction kit, including an on-column DNase-treatment step (Norgen). An anchor primer (5'-GCGAGCTCCGCGGCCGCGTTTTTTTTTTTT) was annealed to poly(A) sequences by mixing the following in a PCR tube: 1ul 100uM ePAT anchor primer, 1ug total RNA, H_2O to a final volume of 8ul. This was incubated at 80°C for 5min, cooled to 25°C, and then Klenow polymerase was used to fill in the overhang by adding the following to the mix: 1ul Klenow polyermerase (-exo) (5U/ul, NEB), 4ul 5X superscript III buffer, 1ul 0.1M DTT, 1ul 10mM dNTPs, 4ul H_2O, 1ul RNase out. The mixture was then incubated at 25°C for 1 hour, then the Klenow inactivated by incubating at 80°C for 10 min, then cooled to 55°C. While holding at 55°C, 1ul superscript III reverse transcriptase (200U/ul, Life Technologies) was added and then the mixture incubated at 55°C for another hour and the enzyme heat inactivated by incubating at 80°C for 10 min.

The poly(A) tail length of each transcript of interest was then assessed by using this 3' tagged cDNA as the template from which to PCR amplify the region between the anchor primer specific sequence and a sequence 100-300 nucleotides upstream from the 3' end of, and specific to, the transcript of interest. Oligonucleotides used for poly(A) tail length assay:

Universal reverse primer: 5'-GCGCCGGCGCCTCGA

DBNL-1 forward: 5'-TGTGTGCCTCAACTGATTCTGACTTCAGG

DBNL-2 forward: 5'-CGGCTCACTCGATGCCTGCAGG

TACC1-1 forward: 5'-GCGTTGACTACTGCTATCAGGATTGTGTTGTG

TACC1-2 forward: 5'-GGGTGCCATGAAGTGTGTGAGGAGC

PCR was carried out by mixing the following: 2.5ul 10x reaction buffer, 0.5ul

10mM dNTPs, 1ul MgSO4, 1ul 5uM gene specific forward primer (see above),

1ul 5uM universal reverse primer, 1ul 3' tagged cDNA (as synthesized in the first

step), 0.1ul hi fidelity platinum taq (Life Technologies), 17.9ul H_2O. The reactions

were cycled using the following protocol: 1. 94°C 2 min 2. 94°C 15 sec 3. 58°C

30 sec 4. 68°C 35 sec. repeat 2-4 x 30. 5. 68°C 5 min. The resulting PCR

products were resolved by running on a 2% agarose-TAE gel and visualized by

ethidium bromide staining.

MBNL1, DBNL and TACC1 Expression Correlation with Clinical Outcome

To test the correlation of MBNL1, MBNL2, MBNL3 and TACC1 expression levels

with clinical outcome in breast cancer, the kmplot meta data set was used

(Gyorffy et al. 2010). Expression levels of the genes were divided into high and

low based on the median expression level, and a censoring threshold of 10 years

was set. P-values were based on the Mantel-Cox log-rank test. DBNL expression

correlation with clinical outcome in breast cancer was assessed by qRT-PCR in a

set of 48 staged breast tumor samples, using beta-actin as an endogenous

control (Origene, breast cancer panel 3).

Conditioned Media Preparation

Conditioned media for assessment by Western blotting was prepared by seeding 5×10^4 cancer cells per well in 24 well plates. After 18 hours, the cells were washed twice with PBS, and 0.2mL of serum free DMEM complete media was added (DMEM media supplemented with 2mM L-glutamine, 1mM sodium pyruvate, 100 units/mL penicillin, 100 ug/mL streptomycin, and amphotericin B). After 24 hours, the media was harvested and spun at 5×10^3 rpm for 10 minutes to pellet any cells. The supernatant was removed and equal volumes were assayed by Western blotting or ELISA. Extracellular CTGF levels in conditioned media from MDA-231 cells with MBNL1 knockdown or control cells were quantified using a CTGF ELISA kit (PeproTech). Conditioned media for treatment of cancer cells in transwell invasion assays was prepared in the same manner.

SILAC-Mass Spectrometry

Proteins from MDA-231 MBNL1 knockdown cells were isotopically labeled by growing in media supplemented with isotopically labeled $[^{13}C_6]$-L-lysine and $[^{13}C_6,^{15}N_4]$-L-Arginine by passaging cells for ten days in DMEM-flex media supplemented with the labeled amino acids along with 10% dialyzed FBS, L-glutamine, penicillin, streptomycin and amphotericin (Life Technologies). In parallel, MDA-231 cells expressing a control shRNA were passaged identically, but in media containing normal lysine and arginine. To harvest secreted proteins, plates were washed twice in PBS, and media containing all of the above components except FBS was added. After 20 hours, the media from the plates

was harvested and equal volumes were mixed. The proteins were separated by

SDS-PAGE, and bands were excised and subjected to mass spectrometry.

REFERENCES

Abdelmohsen K, Gorospe M. 2010. Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdisciplinary Reviews: RNA* 1:214–29.

Ahn SG, Dong SM, Oshima A, Kim WH, Lee HM, Lee SA, Kwon SH, Lee JH, Lee JM, Jeong J. 2013. LOXL2 expression is associated with invasiveness and negatively influences survival in breast cancer patients. *Breast Cancer Res Treat* 141: 89-99.

Akiyama M, Yuza Y, Yokokawa Y, Yokoi K, Ariga M, Eto Y. 2008. Differences in CTG triplet repeat expansion in leukemic cells and normal lymphocytes from a 14-year-old female with congenital myotonic dystrophy. *Pediatr Blood Cancer* 51: 563-565.

Anderson P, Kedersha N. 2008. Stress granules: the Tao of RNA triage. *Trends Biochem Sci* 33: 141-150.

Anko ML. 2014. Regulation of gene expression programmes by serine-arginine rich splicing factors. *Semin Cell Dev Biol* 32: 11-21.

Arribas-Layton M, Wu D, Lykke-Andersen J, Song H. 2013. Structural and functional control of the eukaryotic mRNA decapping machinery. *Biochim Biophys Acta* 1829: 580-589.

Artero R, Prokop A, Paricio N, Begemann G, Pueyo I, Mlodzik M, Perez-Alonso M, Baylies MK. 1998. The muscleblind gene participates in the organization of Z-bands and epidermal attachments of Drosophila muscles and is regulated by Dmef2. *Dev Biol* 195: 131-143.

Aslanidis C, Jansen G, Amemiya C, Shutler G, Mahadevan M, Tsilfidis C, Chen C, Alleman J, Wormskamp NG, Vooijs M, et al. 1992. Cloning of the essential myotonic dystrophy region and mapping of the putative defect. Nature *355*; 548-551.

Badis G, Saveanu C, Fromont-Racine M, Jacquier A. 2004. Targeted mRNA degradation by deadenylation-independent decapping. *Mol Cell* 15: 5-15.

Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46: 674-690.

Banuls J, Botella R, Palau F, Ramon R, Diaz C, Paya A, Carnero L, Vergara G. 2004. Tissue and tumor mosaicism of the myotonin protein kinase gene

trinucleotide repeat in a patient with multiple basal cell carcinomas associated with myotonic dystrophy. *J Am Acad Dermatol* 50: S1-3.

Bargiela A, Llamusi B, Cerro-Herreros E, Artero R. 2014. Two enhancers control transcription of Drosophila muscleblind in the embryonic somatic musculature and in the central nervous system. *PloS ONE* 9: e93125.

Barker HE, Chang J, Cox TR, Lang G, Bird D, Nicolau M, Evans HR, Gartland A, Erler JT. 2011. LOXL2-mediated matrix remodeling in metastasis and mammary gland involution. *Cancer Res* 71: 1561-1572.

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, et al. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483: 603-607.

Batra R, Charizanis K, Manchanda M, Mohan A, Li M, Finn DJ, Goodwin M, Zhang C, Sobczak K, Thornton CA, et al. 2014. Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol Cell* 56: 311-322.

Bertucci F, Houlgatte R, Benziane A, Granjeaud S, Adelaide J, Tagett R, Loriod B, Jacquemier J, Viens P, Jordan B, et al. 2000. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet* 9: 2981-2991.

Bett JS, Ibrahim AF, Garg AK, Kelly V, Pedrioli P, Rocha S, Hay RT. 2013. The P-body component USP52/PAN2 is a novel regulator of HIF1A mRNA stability. *Biochem J* 451: 185-194.

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, et al. 1992. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68: 799-808.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149: 1393-1406.

Chabottaux V, Ricaud S, Host L, Blacher S, Paye A, Thiry M, Garofalakis A, Pestourie C, Gombert K, Bruyere F, et al. 2009. Membrane-type 4 matrix metalloproteinase (MT4-MMP) induces lung metastasis by alteration of primary breast tumour vascular architecture. *J Cell Mol Med* 13: 4002-4013.

Chambers AF, Groom AC, MacDonald IC. 2002. Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* 2: 563-572.

Chang J, Erler J. 2014. Hypoxia-mediated metastasis. *Adv Exp Med Biol* 772: 55-81.

Charizanis K, Lee K-Y, Batra R, Goodwin M, Zhang C, Yuan Y, Shiue L, Cline M, Scotti M, Xia G, et al. 2012. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* 75: 437-450.

Chen PS, Wang MY, Wu SN, Su JL, Hong CC, Chuang SE, Chen MW, Hua KT, Wu YL, Cha ST, et al. 2007. CTGF enhances the motility of breast cancer cells via an integrin-alphavbeta3-ERK1/2-dependent S100A4-upregulated pathway. *J Cell Sci* 120: 2053-2065.

Cheng AW, Shi J, Wong P, Luo KL, Trepman P, Wang ET, Choi H, Burge CB, Lodish HF. 2014. Muscleblind-like 1 (Mbnl1) regulates pre-mRNA alternative splicing during terminal erythropoiesis. *Blood* 124: 598-610.

Chiang AC, Massague J. 2008. Molecular basis of metastasis. *N Engl J Med* 359: 2814-2823.

Connert S, Wienand S, Thiel C, Krikunova M, Glyvuk N, Tsytsyura Y, Hilfiker-Kleiner D, Bartsch JW, Klingauf J, Wienands J. 2006. SH3P7/mAbp1 deficiency leads to tissue and behavioral abnormalities and impaired vesicle transport. *EMBO J* 25: 1611-1622.

Conte N, Delaval B, Ginestier C, Ferrand A, Isnardon D, Larroque C, Prigent C, Seraphin B, Jacquemier J, Birnbaum D. 2003. TACC1-chTOG-Aurora A protein complex in breast cancer. *Oncogene* 22: 8102-8116.

Cully M, Shiu J, Piekorz RP, Muller WJ, Done SJ, Mak TK. 2005. Transforming acidic coiled coil 1 promotes transformation and mammary tumorigenesis. *Cancer Res* 65: 10363-10370.

Daughters RS, Tuttle DL, Gao W, Ikeda Y, Moseley ML, Ebner TJ, Swanson MS, Ranum LP. 2009. RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet* 5: e1000600.

Delaval B, Ferrand A, Conte N, Larroque C, Hernandez-Verdun D, Prigent C, Birnbaum D. 2004. Aurora B-TACC1 protein complex in cytokinesis. *Oncogene* 23: 4516-4522.

Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Res* 22: 1173-1183.

Dixon DM, Choi J, El-Ghazali A, Park SY, Roos KP, Jordan MC, Fishbein MC, Comai L, Reddy S. 2015. Loss of muscleblind-like 1 results in cardiac pathology and persistence of embryonic splice isoforms. *Sci Rep* 5: 9042.

Doma MK, Parker R. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* 440: 561-564.

Du H, Cline M, Osborne R, Tuttle D, Clark T, Donohue J, Hall M, Shiue L, Swanson M, Thornton C, et al. 2010. Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat Struct Mol Biol* 17: 187-193.

Dudley AC. 2012. Tumor endothelial cells. Cold Spring Harb Perspect Med 2: a006536.

Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH. 2009. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* 16: 49-55.

Elemento O, Slonim N, Tavazoie S. 2007. A universal framework for regulatory element discovery across all genomes and data types. *Mol Cell* 28: 337-350.

Eroles P, Bosch A, Perez-Fidalgo JA, Lluch A. 2012. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev* 38: 698-707.

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E. 2007. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol* 27: 3970-3981.

Fardaei M, Larkin K, Brook JD, Hamshere MG. 2001. In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res* 29: 2766-2771.

Fenger-Gron M, Fillman C, Norrild B, Lykke-Andersen J. 2005. Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol Cell* 20: 905-915.

Fenster SD, Kessels MM, Qualmann B, Chung WJ, Nash J, Gundelfinger ED, Garner CC. 2003. Interactions between Piccolo and the actin/dynamin-binding protein Abp1 link vesicle endocytosis to presynaptic active zones. *J Biol Chem* 278: 20268–20277.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013. http://globocan.iarc.fr, accessed on 10 May 2015.

Francis PA, Regan MM, Fleming GF, Lang I, Ciruelos E, Bellet M, Bonnefoi HR, Climent MA, Da Prada GA, Burstein HJ, et al. 2015. Adjuvant ovarian suppression in premenopausal breast cancer. *N Engl J Med* 372: 436-446.

Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. 2002. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* 295: 2258-2261.

Frohlich C, Klitgaard M, Noer JB, Kotzsch A, Nehammer C, Kronqvist P, Berthelsen J, Blobel C, Kveiborg M, Albrechtsen R, et al. 2013. ADAM12 is expressed in the tumour vasculature and mediates ectodomain shedding of several membrane-anchored endothelial proteins. *Biochem J* 452: 97-109.

Gadalla SM, Lund M, Pfeiffer RM, Gortz S, Mueller CM, Moxley RT, Kristinsson SY, Bjorkhom M, Shebl FM, Hilbert JE, et al. 2011. Cancer risk among patients with myotonic muscular dystrophy. *JAMA* 306: 2480-2486.

Ge Y, Porse BT. 2013. The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression. *Bioessays* 36: 236-243.

Geng Y, Marshall JR, King MR. 2012. Glycomechanics of the metastatic cascade: tumor cell-endothelial cell interactions in the circulation. *Ann Biomed Eng* 40: 790-805.

Giannopoulou EG, Elemento O. 2011. An integrated ChIP-seq analysis platform with customizable workflows. *BMC Bioinformatics* 12:277.

Gligorijevic B, Wyckoff J, Yamaguchi H, Want Y, Roussos ET, Condeelis J. 2012. N-WASP-mediated invadopodium formation is involved in intravasation and lug metastasis of mammary tumors. *J Cell Sci* 125: 724-734.

Goers ES, Purcell J, Voelker RB, Gates DP, Berglund JA. 2010. MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res* 38: 2467-2484.

Goodarzi H, Najafabadi HS, Oikonomou P, Greco TM, Fish L, Salavati R, Cristea IM, Tavazoie S. 2012. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* 485: 264-268.

Goodarzi H, Zhang S, Buss CG, Fish L, Tavazoie S, Tavazoie SF. 2014. Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. *Nature* 513: 256-260.

Gu SQ, Bakthavachalu B, Han J, Patil DP, Otsuka Y, Guda C, Schoenberg DR. 2012. Identification of the human PMR1 mRNA endonuclease as an alternatively processed product of the gene for peroxidasin-like protein. *RNA* 18: 1186-1196.

Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. 2010. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients. *Breast Cancer Res Treatment* 123:725-31.

Haag N, Schwintzer L, Ahuja R, Koch N, Grimm J, Heuer H, Qualmann B, Kessels MM. 2012. The actin nucleator Cobl is crucial for Purkinje cell development and works in close conjunction with the F-actin binding protein Abp1. *J Neurosci* 32: 17842-17856.

Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, Grigsby J, Gage B, Hagerman PJ. 2001. 2001. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* 57: 127-130.

Hagerman PJ, Hagerman RJ. 2015. Fragile X-associated tremor/ataxia syndrome. *Ann NY Acad Sci* 1338: 58-70.

Han H, Irimia M, Ross PJ, Sung HK, Alipanahi B, David L, Golipour A, Gabut M, Michael IP, Nachman EN, et al. 2013. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* 498: 241-245.

Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, Harper PS, Housman DE, Shaw DJ. 1992. Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 355: 545-546.

Hepper I, Schymeinsky J, Weckbach LT, Jakob SM, Frommhold D, Sixt M, Laschinger M, Sperandio M, Walzog B. 2012. The mammalian actin-binding

protein 1 is critical for spreading and intraluminal crawling of neutrophils under flow conditions. *J Immunol* 188: 4590-4601.

Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS, Cooper, TA. 2004. Muscleblind proteins regulate alternative splicing. *EMBO J* 23: 3103-3112.

Hou P, Estrada L, Kinley AW, Parsons JT, Vojtek A, Gorski JL. 2003. Fgd1, the Cdc42 GEF responsible for Faciogenital Dysplasia, directly interacts with cortactin and mAbp1 to modulate cell shape. *Human Mol Gen* 12: 1981–1993.

Houseley J, Tollervey D. 2009. The many pathways of RNA degradation. *Cell* 136: 763-776.

Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, et al. 1975-2012. SEER cancer statistic review. National Cancer Institute. http://seer.cancer.gov/csr/1975_2012/, based on November 2014 SEER data submission, posted to the SEER web site April 2015, accessed 29 April 2015.

Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S, et al. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* 1: 167-178.

Hughes CS, Postovit LM, Lajoie GA. 2010. Matrigel: a complex protein mixture required for optimal growth of cell culture. *Proteomics* 10: 1886-1890.

Iwahashi CK, Yasui DH, An HJ, Greco CM, Tassone F, Nannen K, Babineau B, Lebrilla CB, Hagerman RJ, Hagerman PJ. 2006. Protein composition of the intranuclear inclusions of FXTAS. *Brain* 129: 256-271.

Jacquemont S, Farzin F, Hall D, Leehey M, Tassone F, Gane L, Zhang L, Grigsby J, Jardini T, Lewin F, et al. 2004. Aging in individuals with the FMR1 mutation. *Am J Ment Retard* 109: 154-164.

Janicke A, Vancuylenberg J, Boag PR, Traven A, Beilharz TH. 2012. ePAT: a simple method to tag adenylated RNA to measure poly(A)-tail length and other 3'RACE applications. *RNA* 18: 1289-1295.

Jensen KB, Darnell RB. 2008. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol Biol* 488, 85-98.

Jinnai K, Sugio T, Mitani M, Hashimoto K, Takahashi K. 1999. Elongation of (CTG)n repeats in myotonic dystrophy protein kinase gene in tumors associated with myotonic dystrophy patients. *Muscle Nerve* 22: 1271-1274.

Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, Cooper TA. 2008. A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc Natl Acad Sci USA* 105: 20333-20338.

Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, Timmers AM, Hauswirth WW, Swanson MS. 2003. A muscleblind knockout model for myotonic dystrophy. *Science* 302, 1978-1980.

Katchman BA, Ocal IT, Cunliffe HE, Chang YH, Hostetter G, Watanabe A, LoBello J, Lake DF. 2013. Expression of quiescin sulfhydryl oxidase 1 is associated with a highly invasive phenotype and correlates with a poor prognosis in Luminal B breast cancer. *Breast Cancer Res* 15: R28.

Katz Y, Wang ET, Airoldi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* 7: 1009-1015.

Kedersha NL, Gupta M, Li W, Miller I, Anderson P. 1999. RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. *J Cell Biol* 147: 1431-1442.

Kedersha N, Stoecklin G, Ayodele M, Yacono P, Lykke-Andersen J, Fritzler MJ, Scheuner D, Kaufman RJ, Golan DE, Anderson P. 2005. Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J Cell Biol* 169: 871-884.

Kessels MM, Engqvist-Goldstein AE, Drubin DG. 2000. Association of mouse actin-binding protein 1 (mAbp1/SH3P7), an Src kinase target, with dynamic regions of the cortical actin cytoskeleton in response to Rac1 activation. *Mol Biol Cell* 11: 393-412.

Kessels MM, Engqvist-Goldstein AE, Drubin DG, Qualmann B. 2001. Mammalian Abp1, a signal-responsive F-actin-binding protein, links the actin cytoskeleton to endocytosis via the GTPase dynamin. *J Cell Biol* 153: 351–366.

Khuon S, Liang L, Dettman RW, Sporn PH, Wysolmerski RB, Chew TL. 2010. Myosin light chain kinase mediates transcellular intravasation of breast cancer cells through the underlying endothelial cells: a three-dimensional FRET study. *J Cell Sci* 123: 431-440.

Kienast Y, von Baumgarten L, Fuhrmann M, Klinkert WE, Goldbrunner R, Herms J, Winkler F. 2010. Real-time imaging reveals the single steps of brain metastasis formation. *Nat Med* 16: 116-122.

Kinoshita M, Igarashi A, Komori T, Tamura H, Hayashi M, Kinoshita K, Deguchi T, Hirose K. 1997. Differences in CTG triplet repeat expansions in an ovarian cancer and cyst from a patient with myotonic dystrophy. *Muscle Nerve* 20: 622-624.

Kinoshita M, Osanai R, Kikkawa M, Adachi A, Ohtake T, Komori T, Hashimoto K, Itoyama S, Mitarai T, Hirose K. 2002. A patient with myotonic dystrophy type 1 (DM1) accompanied by laryngeal and renal cell carcinomas had a small CTG triplet repeat expansion but no somatic instability in normal tissues. *Intern Med* 41: 312-318.

Kittaneh M, Montero AJ, Gluck S. 2013. Molecular profiling for breast cancer: a comprehensive review. *Biomark Cancer* 5: 61-70.

Klesert TR, Otten AD, Bird TD, Tapscott SJ. 1997. Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat Genet* 16, 402-406.

Klesert TR, Cho DH, Clark JI, Maylie J, Adelman J, Snider L, Yuen EC, Soriano P, Tapscott SJ. 2000. Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nat Genet* 25, 105-109.

Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LP. 1999. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet* 21: 379-384.

Ladd AN, Charlet N, Cooper TA. 2001. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell Biol* 21: 1285-1296.

Larbolette O, Wollscheid B, Schweikert J, Nielsen PJ, Wienands J. 1999. SH3P7 is a cytoskeleton adapter protein and is coupled to signal transduction from lymphocyte antigen receptors. *Mol Cell Biol* 18: 1539-1546.

Lauffart B, Dimatteo A, Vaughan MM, Cincotta MA, Black JD, Still IH. 2006. Temporal and spatial expression of TACC1 in the mouse and human. *Dev Dyn* 235: 1638-1647.

Lauffart B, Howell SJ, Tasch JE, Cowell JK, Still IH. 2002. Interaction of the transforming acidic coiled-coil 1 (TACC1) protein with ch-TOG and GAS41/NuBI1

suggests multiple TACC1-containing protein complexes in human cells. *Biochem J* 363: 195-200.

Lauffart B, Sondarva GV, Gangisetty O, Cincotta M, Still IH. 2007. Interaction of TACC proteins with the FHL family: implications for ERK signaling. *J Cell Commun Signal* 1: 5-15.

Laurent FX, Sureau A, Klein AF, Trouslard F, Gasnier E, Furling D, Marie J. 2012. New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. *Nucleic Acids Res* 40: 3159-3171.

Lee KS, Squillace RM, Wang EH. 2007. Expression pattern of muscleblind-like proteins differs in differentiating myoblasts. *Biochem Biophys Res Commun* 361: 151-155.

Lee Y, Rio DC. 2015. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu Rev Biochem* 84: 31.1-31.33.

Lee YT. 1983. Breast carcinoma: pattern of metastasis at autopsy. *J Surg Oncol* 23: 175-180.

Lembo A, Di Cunto F, Provero P. 2012. Shortening of 3'UTRs correlates with poor prognosis in breast and lung cancer. *PLoS One* 7: e31129.

Li S, Mason CE. 2014. The pivotal regulatory landscape of RNA modifications. *Annu Rev Genomics Hum Genet* 15: 127-150.

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, et al. 2008. HITS-CLIP yield genome-wide insights into brain alternative RNA processing. *Nature* 456: 464-469.

Lin X, Miller JW, Mankodi A, Kanadia RN, Yuan Y, Moxley RT, Swanson MS, Thornton CA. 2006. Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum Mol Genet* 15: 2087-2097.

Ling SH, Qamra R, Song H. 2011. Structural and functional insights into eukaryotic mRNA decapping. *Wiley Interdiscip Rev RNA* 2: 193-208.

Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP. 2001. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 293: 865-867.

Liu H, Ma CP, Chen YT, Schuyler SC, Schang KP, Tan BC. 2014. Functional impact of RNA editing and ADARs on regulation of gene expression: perspectives from deep sequencing studies. *Cell Biosci* 4: 44.

Lund M, Diaz LJ, Gortz S, Feenstra B, Duno M, Juncker I, Eiberg H, Vissing J, Wohlfahrt J, Melbye M. 2014. Risk of cancer in relatives of patients with myotonic dystrophy: a population-based cohort study. *Eur J Neurol* 21: 1192-1197.

Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, Groom AC. 1998. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* 153: 865-873.

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, et al. 1992. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 255: 1253-1255.

Mankodi A, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS, Thornton CA. 2001. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum Mol Genet* 10: 2165-2170.

Mankodi A, Logigian E, Callahan L, McClain C, White R, Henderson D, Krym M, Thornton C. 2000. Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* 289: 1769-1773.

Masuda A, Andersen HS, Doktor TK, Okamoto T, Ito M, Andresen BS, Ohno K. 2012. CUGBP1 and MBNL1 preferentially bind to 3'UTRs and facilitate mRNA decay. *Sci Rep* 2:209.

Mayr C, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138: 673-684.

Miller E, Lee HJ, Lulla A, Hernandez L, Gokare P, Lim B. 2014. Current treatment of early breast cancer: adjuvant and neoadjuvant therapy. *F1000Res* 3: 198.

Miller JW, Urbinati CR, Teng-Umnuay P, Stenberg MG, Byrne BJ, Thornton CA, Swanson MS. 2000. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J* 19: 4439-4448.

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB., Gerald WL, Massague J. 2005. Genes that mediate breast cancer metastasis to lung. *Nature* 436: 518-524.

Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS, Chen G, Weatherspoon MR, Clark HB, Ebner TJ, et al. 2006. Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat Genet* 38: 758-769.

Nilsen TW Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463: 457-463.

Norbury CJ. 2013. Cytoplasmic RNA: a case of the tail wagging the dog. *Nat Rev Mol Cell Biol* 14: 643-653.

Ogata K, Takahashi A, Oguchi N, Ishitoya J, Fuse S, Shimpo T. Somatic mosaicism of p(CTG)n expansion in a case of myotonic dystrophy with parotid tumor. 1998. *Rinsho Shinkeigaku* 38: 736-738.

Onishi H, Kino Y, Morita T, Futai E, Sasagawa N, Ishiura S. 2008. MBNL1 associates with YB-1 in cytoplasmic stress granules. *J Neurosci Res* 86: 1994-2002.

Osanai R, Kinoshita M, Hirose K, Homma T, Kawabata I. 2000. CTG triplet repeat expansion in a laryngeal carcinoma from a patient with myotonic dystrophy. 2000. *Muscle Nerve* 23: 804-806.

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40: 1413-1415.

Paul S, Dansithong W, Jog SP, Holt I, Mittal S, Brook JD, Morris GE, Comai L, Reddy S. 2011. Expanded CUG repeats dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex. *J Biol Chem* 286: 38427-38438.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. 2000. Molecular portraits of human breast tumours. *Nature* 406: 747-752.

Pettersson OJ, Aagaard L, Andrejeva D, Thomsen R, Jensen TG, Damgaard CK. 2014. DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophia myotonica type 1. *Nucleic Acids Res* 42: 7186-7200.

Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, Gianni L, Baselga J, Bell R, Jackisch C, et al. 2005. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 353: 1659-1672.

Pignatelli J, Goswami S, Jones JG, Rohan TE, Pieri E, Chen X, Adler E, Cox D, Maleki S, Bresnick A, et al. 2014. Invasive breast carcinoma cells from patients exhibit MenaINV- and macrophage-dependent transendothelial migration. Sci Signal *7*: ra112.

Pinyol R, Haeckel A, Ritter A, Qualmann B, Kessels MM. 2007. Regulation of N-WASP and the Arp2/3 complex by Abp1 controls neuronal morphology. *PLoS ONE* 2: 400.

Pizzi C, Di Maio M, Daniele S, Mastranzo P, Spagnoletti I, Limite G, Pettinato G, Monticelli A, Cocozza S, Contegiacomo A. 2007. Triplet repeat instability correlates with dinucleotide instability in primary breast cancer. *Oncol Rep* 17: 193-199.

Ponomarev V, Doubrovin M, Serganova I, Vider J, Shavrin A, Beresten T, Ivanova A, Ageyeva L, Tourkova V, Balatoni J, et al. 2004. A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur J Nucl Med Mol Imaging* 31: 740-751.

Quail DF, Joyce JA. 2013. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* 19: 1423-1437.

Reddy S, Smith DB, Rich MM, Leferovich JM, Reilly P, Davis BM, Tran K, Rayburn H, Bronson R, Cros D, et al. 1996. Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat Genet* 13: 325-335.

Reymond N, d'Agua BB, Ridley AJ. 2013. Crossing the endothelial barrier during metastasis. *Nat Rev Cancer* 13: 858-870.

Sanduja S, Blanco FF, Dixon DA. 2010. The roles of TTP and BRF proteins in regulated mRNA decay. *Wiley Interdiscip Rev RNA* 2: 42–57.

Sarkar PS, Appukuttan B, Han J, Ito Y, Ai C, Tsai W, Chai Y, Stout JT, Reddy S. 2000. Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. *Nat Genet* 25: 110-114.

Schoenberg DR. 2011. Mechanisms of endonuclease-mediated mRNA decay. *Wiley Interdiscip Rev RNA* 2: 582-600.

Schymeinsky J, Gerstl R, Mannigel I, Niedung K, Frommhold D, Panthel K, Heesemann J, Sixt M, Quast T, Kolanus W, et al. 2009. A fundamental role of mAbp1 in neutrophils: impact on beta(2) integrin-mediated phagocytosis and adhesion in vivo. *Blood* 114: 4209-4220.

Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, et al. 2010. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J* 29: 1248-1261.

Sibley CR. 2014. Regulation of gene expression through production of unstable mRNA isoforms. *Biochem Soc Trans* 42: 1196-1205.

Singh D, Chan JM, Zoppoli P, Niola F, Sullivan R, Castano A, Liu EM, Reichel J, Porrati P, Pellegatta S, et al. 2012. Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* 337: 1231-1235.

Smith I, Procter M, Gelber RD, Guillaume S, Feyereislova A, Dowsett M, Goldhirsch A, Untch M, Mariani G, Baselga J, et al. 2007. 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. *Lancet* 369: 29-36.

Sobin LH, Gospodarowicz MK, Wittekind C. 2010. Breast Tumours. TNM Online. 181-193.

Somasekharan SP, El-Naggar A, Leprivier G, Cheng H, Hajee S, Grunewald TG, Zhang F, Ng T, Delattre O, Evdokimova V, et al. 2015. YB-1 regulates stress granule formation and tumor progression by translationally activating G3BP1. *J Cell Biol* 208: 913-929.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci* 98: 10869-10874.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S. 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci* 100: 8418-8423.

Squillace RM, Chenault DM, Wang EH. 2002. Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev Biol* 250: 218-230.

Still IH, Hamilton M, Vince P, Wolfman A, Cowell JK. 1999. Cloning of TACC1, an embryonically expressed, potentially transforming coiled coil containing gene, from the 8p11 breast cancer amplicon. *Oncogene* 18: 4032-4038.

Still IH, Vettaikkorumakankauv AK, DiMatteo A, Liang P. 2004. Structure-function evolution of the transforming acidic coiled coil genes revealed by analysis of phylogenetically diverse organisms. *BMC Evol Biol* 4:16.

Stoecklin G, Mayo T, Anderson P. 2006. ARE-mRNA degradation requires the 5'-3' decay pathway. *EMBO Rep* 7: 72-77.

Stoletov K, Kato H, Zardouzian E, Kelber J, Yang J, Shattil S, Klemke R. 2010. Visualizing extravasation dynamics of metastatic tumor cells. *J Cell Sci* 123: 2332-2341.

Taneja KL, McCurrach M, Schalling M, Housman D, Singer RH. 1995. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol* 128: 995-1002.

Tavazoie SF, Alarcon C, Oskarsson T, Padua D, Wang Q, Bos PD, Gerald WL, Massague J. 2008. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* 451: 147-152.

Teplova M, Patel DJ. 2008. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat Struct Mol Biol* 15: 1343-1351.

Tharun S. 2009. Roles of eukaryotic Lsm proteins in the regulation of mRNA function. *Int Rev Cell Mol Biol* 272: 149-189.

Tharun S, Parker R. 2001. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol Cell* 8: 1075-1083.

Tomecki R, Dziembowski A. 2010. Novel endoribonucleases as central players in various pathways of eukaryotic RNA metabolism. *RNA* 16: 1692-1724.

Udd B, Krahe R. 2012. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. *Lancet Neurol* 10: 891-905.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302: 1212-1215.

Vajda NA, Brimacombe KR, LeMasters KE, Ladd AN. 2009. Muscleblind-like 1 is a negative regulator of TGF-beta-dependent epithelial-mesenchymal transition of atrioventricular canal endocardial cells. *Dev Dyn* 238: 3266-3272.

van Dijk E, Cougot N, Meyer S, Babajko S, Wahle E, Seraphin B. 2002. Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J* 21: 6915-6924.

Vanharanta S, Marney CB, Shu W, Valiente M, Zou Y, Mele A, Darnell RB, Massague J. 2014. Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer. *Elife* 3: e02734-e02734.

Wang ET, Cody NAL, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, et al. 2012. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* 150: 710-724.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470-476.

Wang ET, Ward AJ, Cherone JM, Giudice J, Wang TT, Treacy DJ, Lambert NJ, Freese P, Saxena T, Cooper TA, et al. 2015. Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res* 25: 1-14.

Wang Z, Jiao X, Carr-Schmid A, Kiledjian M. 2002. The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proc Natl Acad Sci USA* 99: 12663-12668.

Warf MB, Diegel JV, von Hippel PH, Berglund JA. 2009. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc Natl Acad Sci USA* 106: 9203-9208.

Weyn-Vanhentenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ, et al. 2014. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* 6: 1139-1152.

Wilusz CJ, Gao M, Jones CL, Wilusz J, Peltz SW. 2001. Poly(A)-binding proteins regulated both mRNA deadenylation and decapping in yeast cytoplasmic extracts. *RNA* 7: 1416-1424.

Win AK, Perattur PG, Pulido JS, Pulido CM, Lindor NM. 2012. Increased cancer risks in myotonic dystrophy. *Mayo Clin Proc* 87: 130-135.

Wo JY, Chen K, Neville BA, Lin NU, Punglia RS. 2011. Effect of very small tumor size on cancer-specific mortality in node-positive breast cancer. *J Clin Oncol* 29: 2619-2627.

Xian X, Hakansson J, Stahlberg A, Lindblom P, Betsholtz C, Gerhardt H, Semb H. 2006. Pericytes limit tumor cell metastasis. *J Clin Invest* 116: 642-651.

Xu Y, Gao XD, Lee JH, Huang H, Tan H, Ahn J, Reinke LM, Peter ME, Feng Y, Gius D, et al. 2014. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes Dev* 28: 1191-1203.

Yamashita A, Chang TC, Yamashita Y, Zhu W, Zhong Z, Chen CY, Shyu AB. 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat Struct Mol Biol* 12: 1054-1063.

Yan YB. 2014. Deadenylation: enzymes, regulation, and functional implications. *Wiley Interdiscip Rev RNA* 5: 421-443.

Yang F, Schoenberg DR. 2004. Endonuclease-mediated mRNA decay involves the selective targeting of PMR1 to polyribosome-bound substrate mRNA. *Mol Cell* 14: 435-445.

Yeo GW, Coufal NG, Liang TY, Peng GE, Fu GE, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* 16: 130-137.

Zhang C, Darnell RB. 2011. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 29: 607-614.

Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB. 2010. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* 329: 439-443.

Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ. 2008. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev* 22: 2550-2563.

Zhu J, Mayeda A, Krainer AR. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8: 1351-1361.