Investigating the molecular mechanisms and functions of the Musashi-2 RNA-binding protein

By

Christopher Bennett

B.S., Texas A&M University 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Molecular, Cellular, and Developmental Biology

2016

This thesis entitled:
Investigating the molecular mechanisms and functions of the Musashi-2 RNA-binding protein
Written by Christopher Bennett.
Has been approved by the
Department of Molecular, Cellular and Developmental Biology

_____
Robin Dowell, Ph.D. (Chair)

_____
Rui Yi, Ph.D. (Advisor)

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Abstract

Bennett, Christopher (Ph.D., Molecular, Cellular and Developmental Biology)

Investigating the molecular mechanisms and functions of the Musashi-2 RNA-binding protein

Thesis directed by Associate Professor Rui Yi

The Musashi (Msi) family of RNA-binding proteins is post-transcriptional regulators of gene expression. They were discovered in 1994 as being required for Drosophila sensory organ development. Since then, Msi proteins have been found to enhance cell proliferation and maintain stem cell identities in a multitude of mammalian tissues. In addition, overexpression of Msi proteins is often observed in many types of human cancers, most prominently the widely expressed Msi family member, Musashi-2 (*Msi2*). *Msi2* plays oncogenic roles in hematopoietic, neural, and gastrointestinal tissues. However, *Msi2* has received little attention in other tissues in which it is expressed, such as in stratified epithelium. Additionally, the fundamental questions concerning *Msi2* such as target recognition, molecular mechanisms of *Msi2*-mediated gene regulation and cell type-specific functions are poorly understood.

To understand the functions of *Msi2* in mouse skin, I identified *Msi2* targets using High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) in keratinocytes, which are basal progenitor cells that express high levels of *Msi2*. Analysis of these data revealed a strong preference of *Msi2* for 3'UTR binding at regions enriched for a trinucleotide UAG core motif. These data were used in conjunction with RNA-seq and ribosome profiling data in *Msi2*-depleted keratinocytes

and *Msi2*-normal keratinocytes to reveal a possible mechanism of target regulation. I found that loss of *Msi2* increased RNA stability without altering translational efficiency, consistent with *Msi2*'s role as an inhibitor of gene expression. From these datasets, I generated a list of high confidence, *Msi2*-regulated targets in keratinocytes. Among these *Msi2* targets, I found that genes involved in the regulation of proliferation, cell survival and migration were enriched. Indeed, the *Msi2*-depleted keratinocytes showed reduced levels of cell proliferation and survival, but increased migration when compared to the keratinocytes with normal levels of *Msi2*. These results are consistent with observations in wounded skin, where migrating keratinocytes proximal to the wound showed reduced levels of *Msi2* in contrast to the normal levels observed in normal keratinocytes distal to the wounded skin. Collectively, these data provide new insights into the targets, molecular mechanisms, and functions of *Msi2*, and identify an unexpected function of *Msi2* in restricting epithelial cell migration in keratinocytes.

Acknowledgements

I would like to thank my advisor, Dr. Rui Yi, for his support and patience during my training as a scientist, and for allowing me to work on this project. His openness to new ideas and techniques allowed me to develop skills in experimental and computational biology. He stuck with me through all the troubles I had and allowed me to develop my own approach to science.

I would also like to thank my committee members: Dr. Robin Dowell, Dr. Tom Blumenthal, Dr. Joaquin Espinosa, and Dr. Norman Pace. They provided great advice and support during my tenure as a graduate student, and pushed me to better myself, my ideas, and my project.

The entire Yi Lab has provided great advice, support, and enjoyment during my journey through graduate school. I'd especially like to acknowledge Dr. Kent Riemondy and Dr. Jerome Lee for their contributions to my project and development as a scientist. This project was done in collaboration with Dr. Kent Riemondy who helped complete some experiments (marked in figure legends) and refine the bioinformatics to publishable quality.

Last but not the least, I would like to thank all of my friends and family members who have supported and encouraged me during my trials and tribulations in grad school. Without them it would have been a longer and less enjoyable process.

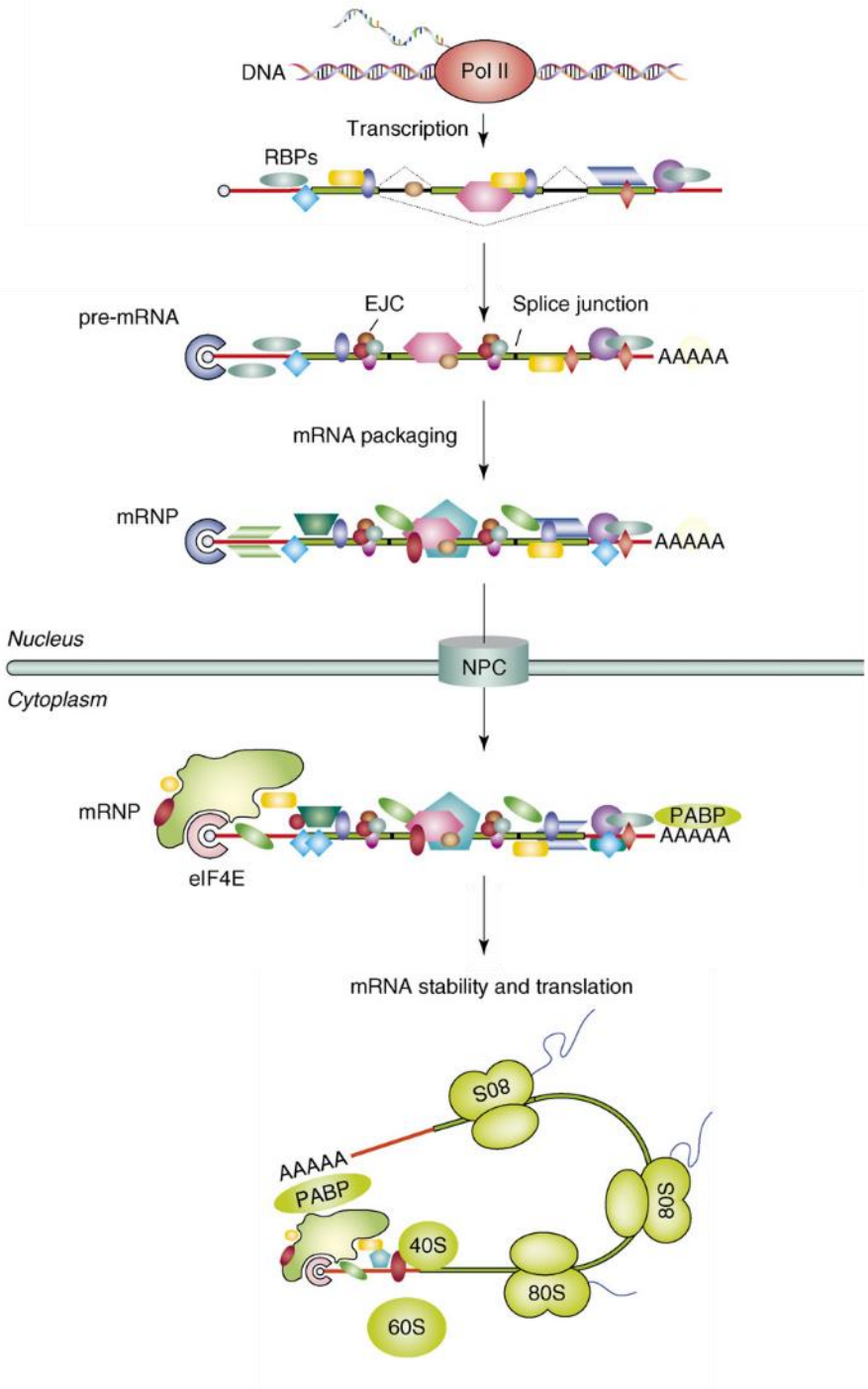
TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

RNA-binding proteins play crucial roles in post-transcriptional gene regulation. Due to the role of RNA-binding proteins in regulating gene expression, perturbations in their expression and function can be found in many disease states such as cancers or developmental defects (Lukong et al., 2008). Recent work has begun to illuminate the diverse regulatory mechanisms performed by various RNA-binding proteins and the processes they modulate, but much remains unknown (Gerstberger et al., 2014; Lunde et al., 2007; Nawy, 2014). Mechanisms and functions of many RNA-binding proteins have yet to be elucidated in mammalian cells. The RNA-binding protein, Musashi-2 (*Msi2*), has recently gained attention for its oncogenic potential (Kharas et al., 2010; Kuang et al., 2013; Moore, 2010). While research has begun to explore *Msi2*'s molecular mechanisms and binding interactions, there are still many questions that have not been sufficiently addressed (Li et al., 2015; Park et al., 2015; Sutherland et al., 2015a; Taggart et al., 2016). Additionally, *Msi2* is expressed in a wide range of tissues but many current studies have focused on *Msi2*-mediated regulation in hematopoietic and gastrointestinal systems (Li et al., 2015; Park et al., 2015; Sugiyama-Nakagiri et al., 2006). To better understand *Msi2*'s functions, I examined the binding, regulation, and modulated processes of this protein in the epithelial progenitor cells, keratinocytes.

Figure 1: Functions of RNA-binding proteins in Eukaryotes



Schematic of RNA-binding proteins that associate with mRNA throughout the steps from transcription to translation of the message. Red strands on the mRNA indicate 5' and 3' UTR while green indicates coding region. Adapted with permission from Elsevier (Lukong et al., 2008)

1.1 Overview of RNA-binding proteins and modes of regulation

RNAs perform many critical functions in biological systems (Lukong et al., 2008). One of the most well-known and characterized functions is to transmit information coded in the genome to the ribosome for protein synthesis in the form of messenger RNAs (mRNAs). Control of these messages is critical to ensure faithful timing and level of gene expression, thus, many regulatory mechanisms have evolved (Lukong et al., 2008). One such mechanism is RNA-binding protein mediated post-transcriptional control of gene expression. RNA-binding proteins can associate with a diverse array of mRNA targets in order to exert control over gene expression via a number of mechanisms, including modulating stability, localization, and translation, along with many others. (Gerstberger et al., 2014). Additionally, many different RNA-binding proteins can bind a single mRNA, each performing a specific function and influencing the output of the message (Figure 1) (Lukong et al., 2008). Individual RNA-binding proteins can have a distinct combination of binding characteristics, molecular mechanisms, and interactions (Lunde et al., 2007).

1.1.1 RNA-binding Proteins

RNA-binding proteins are a diverse class of regulators that function by binding RNA at various points throughout the life of the RNA molecule (Lukong et al., 2008). From the time an RNA is transcribed to its ultimate degradation it is bound by RNA-binding proteins (Figure 1) (Lukong et al., 2008). These proteins can perform various functions, including regulation of mRNAs through modulating translation, stability, localization, and other mechanisms (Gerstberger et al., 2014; Lukong et al., 2008).

RNA-binding proteins have evolved many different mechanisms to perform these molecular functions (Lukong et al., 2008). For example, translation can be regulated by inhibiting initiation or elongation depending on the factors that the RNA-binding protein interacts with and the domains present (Gebauer and Hentze, 2004). Furthermore, many RNA-binding proteins perform multiple functions in the same or different cell types (Guo et al., 2010; Iatsenko et al., 2013). A good example is the Argonaute RNA-binding proteins, involved in the microRNA (miRNA) pathway (Kawasaki and Taira, 2004; Volpe et al., 2002). Argonautes can function by destabilizing transcripts, re-localizing transcripts, or inhibiting translation (Guo et al., 2010; Kim et al., 2006; Pare et al., 2009; Younger and Corey, 2011).

RNA-binding proteins interact with RNA via an RNA-binding domain. Large families of RNA-binding domains have been identified, characterized, and predicted (Lunde et al., 2007). The mechanisms by which each family recognizes RNA can vary depending on structures of the RNA recognizing domains. Some domains, like that the PAZ domain in the Argonaute family, interact with the RNA phosphodiester backbone (Yan et al., 2003). These domains are generally nonspecific binders, allowing the protein to interact with many RNAs. Other domains, like the RNA-recognition motif (RRM) in the Musashi family and hnRNP A1, have evolved a protein structure that allows for precise recognition of a specific RNA sequence through base stacking, hydrogen bonds, or hydrophobic interactions (Ding et al., 1999; Lunde et al., 2007; Maris et al., 2005; Ohyama et al., 2011). Still other domains, such as the double-stranded RNA-binding motif of the RNA-binding protein, DGCR8, recognize a specific RNA secondary or tertiary structure (Lambert et al., 2014; Quick-Cleveland et al., 2014).

Many post-transcriptional regulating RNA-binding proteins interact with the 3' untranslated region (3'UTR) of mRNAs (Jia et al., 2013; Vindry et al., 2014). One likely reason for this is that it allows the protein to stay associated with its targets longer without being stripped off by a translating or scanning ribosome, affording a longer residence time for the protein to perform its function.

1.1.2 Regulation of Stability

Mature mRNAs can be stabilized by RNA-binding proteins that protect the transcript from endo- and exonucleases (Ye and Blelloch, 2014). Poly-A binding proteins bind to the poly-A tail and directly control stability of the transcript by protecting the 3' end from degradation (Eliseeva et al., 2013; Glisovic et al., 2008). Cap-binding proteins protect the 5' cap of the transcript from decapping enzymes and prevent exonuclease degradation (Topisirovic et al., 2011). Different RNA-binding proteins can enhance or destabilize these proteins, thus further modulating the stability of the target transcript. For example, Argonaute can remove poly-A binding proteins from the tail, accelerating the shortening of the Poly-A tail (Huntzinger et al., 2012). The resulting transcript would be unprotected from endogenous RNases and rapidly degraded.

RNA stability is typically measured by using an inducible gene-expression system or treating with transcription inhibitors such as Actinomycin D and assaying RNA abundance over time to generate a decay curve from which RNA half-life can be calculated (Chen et al., 2008). However, this approach is generally limited to a handful of targets. To measure RNA stability in a high throughput manner, it usually requires

techniques such as RNA sequencing or microarray. RNA sequencing, however, only measures RNA steady state, which includes the rate of transcription as a variable.

1.1.3 RNA Localization

Another method of RNA regulation by RNA-binding proteins is controlling localization of a transcript. RNA-binding proteins can bring mRNA transcripts to or from sites of translation inhibition, such as RNA granules (Glisovic et al., 2008; Liu et al., 2005). Once at these sites the RNA can be degraded within the granule or it can be stored until repression is relieved (Pare et al., 2009; Zhang and Zhang, 2013).

Thus, this means of regulation can cause a change in the steady state levels of RNA. Such effects, when a system is perturbed, can be readily measured by RNA-seq. The ribosome occupancy of a transcript can also change in these loci. When the translation of a transcript is inhibited, it can be reflected by the reduced number of ribosomes occupied on the transcript. A technique called ribosome profiling can readily detect the depletion of ribosome protected fragments in a genome-wide scale. Isolation of RNA-granules accompanied by RNA-seq, immunofluorescence, *in situ*, or other equivalent detection methods can inform on whether or not the targets are being sequestered.

1.1.4 mRNA Translation

Many RNA-binding proteins have evolved to control translation as. Translation comprises three phases: initiation, elongation, and termination. Translation initiation occurs when initiation factors (eIF) recognize the cap of an mRNA and recruit factors and the small ribosomal subunit and build a scanning ribosome complex (Gebauer and

Hentze, 2004). This complex scans through the 5' untranslated region (5'UTR) until it recognizes an AUG codon in the correct context (Figure 1). In eukaryotes, an 80S ribosome is assembled, initiation factors dissociate, and elongation factors proceed with translation elongation (Clancy and Brown, 2008). This continues until a stop codon is reached and recognized by termination factors. The newly translated protein is then released and the ribosomal subunits dissociate to be reused in another round of translation. In highly expressed transcripts, the 5' and 3' end of a transcript are physically associated with one another to circularize the transcript, thus promoting efficient translation initiation and ribosome recycling (Figure 1) (Gebauer and Hentze, 2004). Translation can be controlled or modulated at any of these steps through a variety of mechanisms including control via RNA-binding proteins (Kong and Lasko, 2012). For example, translation initiation can be inhibited by blocking circularization of the 5' and 3' ends of the transcript (Kong and Lasko, 2012).

Polyribosome fractionation is a method by which translation of a targeted RNA can be assessed. However, new sequencing technologies, like ribosome profiling, make it possible to isolate ribosome protected fragments and assess translation rates for every transcript (Ingolia, 2014). When combined with RNA-seq, translation efficiency, ribosomes per transcript, can be calculated to assess whether observed changes are a result of RNA abundance changes or translation changes.

1.1.5 RNA-binding Protein Target Recognition

RNA-binding proteins use a variety of methods to target their transcripts. Most use a variation of known RNA-binding domains that have evolved to interact with a

unique portion of their RNA target (Lunde et al., 2007). Some domains have evolved to recognize a specific sequence of bases forming a recognition motif (Lunde et al., 2007; Maris et al., 2005). Depending on the motif degeneracy allowed and its commonality, this can provide specific or promiscuous target binding. Other domains, however, interact with the phosphodiester backbone of RNA rather than a sequence of bases. This method of interaction can lead to promiscuous binding to any RNA or, like in the case of Argonautes miRNA interactions, facilitates RNA-RNA recognition to provide target specificity and modularity to the protein function (Song et al., 2003; Wang et al., 2008; Yan et al., 2003). In some cases RNA-binding proteins require a specific structural context to recognize their targets (Quick-Cleveland et al., 2014; Wang et al., 2011). This can be in the form of a double-stranded RNA, a hairpin, etc. These diverse ways by which RNA-binding proteins can interact with their targets, and the subtle protein domain differences that mediate them, allow for dynamic affinities to RNA elements. This diversity highlights a fundamental problem: how do we determine which RNAs a particular RNA-binding protein targets?

Many different methods have been developed over the years to address this problem. A few examples include: yeast two-hybrid screens which can detect direct RNA-protein interactions, RNA-immunoprecipitation (RIP) where the RNA-binding protein is immunoprecipitated with its associated RNA target, and more recently, bioinformatic approaches and/or Systematic Evolution of Ligands by Exponential Enrichment SELEX. These methods have led to important advances in our understanding of RNA-binding proteins; however, they suffer from various limitations including false positives/negatives and/or ability to find exact binding sites, etc. To

overcome these limitations, cross-linked immunoprecipitation (CLIP) methods have been developed (Darnell, 2012; Wang et al., 2009a; Zhang and Darnell, 2011). CLIP methods are used to identify exact RNA-binding protein targets by directly crosslinking the RNA and protein together. While CLIP has its own limitations, it has greatly improved the identification of true RNA-binding protein targets and requirements/elements for binding. The power of this method comes from its ability to find any RNA-binding protein of interest and its mRNA and non-coding RNA (ncRNA) targets. The most common interactions of interest are those involving messenger RNAs (mRNA), though this may be changing as ncRNA function and regulation research becomes more widespread.

1.1.6 Remarks

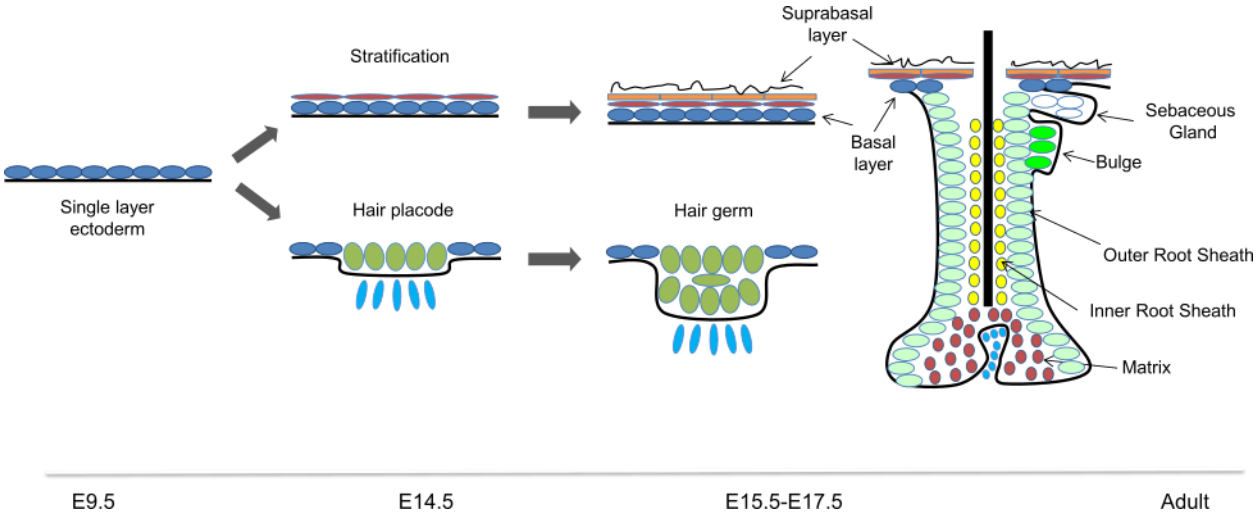
Identification of RNA targets and mechanisms of RNA-binding proteins have been areas of increasing focus. However, many important RNA-binding proteins remain inadequately characterized. With recent developments in sequencing and profiling technologies, RNA-binding protein targets and target regulation can be assessed transcriptome-wide. Thus, it has become easier than ever to study new and poorly characterized RNA-binding proteins.

1.2 Murine skin as a model for studying RNA-binding proteins

Mammalian skin is a complex organ composed of stratified epithelium in the epidermis connected to hair follicles and sweat glands that extend into the fatty dermis which provides insulation and structure (Figure 2) (Blanpain and Fuchs, 2006). Skin

performs the critical function of serving as a barricade between the organism and the environment (Blanpain and Fuchs, 2006). It forms a waterproof barrier around the organism, prevents desiccation, and protects the internal tissues from damage, environmental changes, and pathogens. In order to ensure normal organ function, the skin maintains a large reservoir of stem cells to regenerate and maintain the organ.

Figure 2: Model of Skin and Hair Follicle development



Schematic of Skin development starting at embryonic day 9 following the development of the epidermis (top) and hair follicle (bottom). Proliferative cell populations include: single layer ectoderm, basal progenitor cells (basal layer), hair placode, hair germ, and matrix. Stem cell populations include: single layer ectoderm, basal progenitor cells (basal layer), hair placode, hair germ, and bulge

One of the stem cell populations is the keratinocytes, localized to the basal layer of the epidermis. These basal progenitor and stem cells are responsible for normal maintenance and replenishment of the. They are also, in part, responsible for tissue repair when damage occurs (Arwert et al., 2012). Wound repair involves a complex set of regulatory factors and fine-tuned control of proliferation and migration. Upon wounding, keratinocytes start to proliferate and migrate to the site of the wound. At this

point they interact and coordinate with fibroblasts to initiate the process of re-epithelization and restore barrier function to the skin (Barrientos et al., 2008). The two cell types coordinate and communicate with each other by releasing and responding to different growth factors and secreted molecules.

1.2.1 Skin Development and Structure

Skin development in mice begins around embryonic day 9 (E9), when epidermal progenitor cells are generated and begin to express Keratin 5 and Keratin 14 (Figure 2) (Byrne et al., 1994). At embryonic day 14 the epidermal progenitor cells begin to undergo asymmetric division, with the proliferative stem cell population maintaining contact with the basal membrane. The population of cells that detach from the basal lamina begins to differentiate and express Keratin 1 and Keratin 10 in place of Keratin 5 and Keratin 14 (Lechler and Fuchs, 2005). This process, termed stratification, is one of the first steps in generating a mature skin structure. At around embryonic day 18 the waterproof barrier is formed when the stratum corneum is constructed from suprabasal layers in the epidermis. The cells in this region become cornified, a process whereby the cellular nucleus is lost and the cells consist of crosslinked protein and lipid membranes to perform the protective functions associated with the skin (Koster and Roop, 2007). This layer is constantly shed and replenished by the basal progenitor cells throughout the lifespan of the organism in order to maintain barrier function. The basal progenitor cells are maintained and constantly give rise to differentiating cells to replenish the skin.

1.2.2 Hair Follicle Morphogenesis and Structure

Around the time stratification is occurring in the epidermis, hair follicle morphogenesis begins through activation of the β-catenin/Wnt signaling pathway and crosstalk between the epithelial progenitor cells and dermal cell populations (Figure 2) (Huelsken et al., 2001; Sennett and Rendl, 2012). These new hair follicle progenitor cells begin to migrate down into the dermis and differentiate into various populations of cells with defined localizations, including hair follicle stem cells, located in the bulge region (Cotsarelis et al., 1990). Throughout the lifetime of the adult animal the hair follicles cycle through different stages of growth and destruction, with different cell populations requiring different rates of proliferation in the hair follicle (Schneider et al., 2009).

1.2.3 Skin as a Model

Murine skin is an established model for studying cell dynamics, stress responses, and development (Blanpain and Fuchs, 2009). Complex regulatory networks coordinate these processes and function to maintain a steady state functionality of the skin. Additionally, the skin provides readily accessible cell populations that can be isolated, manipulated and studied. The diversity of the cell populations and processes performed in the skin provide an excellent platform for studying gene functions, and for studying the roles of RNA-binding proteins involved in post-transcriptional regulation. With this system, the role of RNA-binding proteins in stem cell dynamics, cellular proliferation, differentiation, stress responses and a whole host of other processes can be assessed.

1.2.4 Proliferation, Differentiation, and Migration in the Epidermis

Proliferation and differentiation are tightly controlled processes in the basal progenitor cells of the epidermis. Keratinocytes in this region express a number of factors, such as c-Myc and ΔNp63, that promote or maintain proliferation and inhibit differentiation (Alani et al., 1999; Bull et al., 2001; Fan and Khavari, 1999; Liefer et al., 2000). Additionally, it had been proposed that contact with the basal lamina helps maintain proliferation and progenitor identity (Vaezi et al., 2002). Integrins and laminins are critical components that help maintain contact to this region, and thus promote the progenitor identity and inhibit differentiation (Frye et al., 2003). Inhibition of Notch signaling is another way keratinocytes maintain proliferation since activated Notch signaling is critical to promote epidermal differentiation (Blanpain and Fuchs, 2006; Rangarajan et al., 2001).

Another tightly controlled process in keratinocytes is migration. Keratinocyte migration is inhibited by tight interactions with the basal lamina and to each other in normal tissue (Heng, 2011; Santoro et al., 2003). Migration is stimulated upon wounding when keratinocytes receive a signal such as *Hbegf* and loosen their connection to the basal lamina and switch integrin expression from α6β4 to α3β1 integrin (Barrientos et al., 2008; Santoro et al., 2003). Control of each of these processes is critical to maintaining normal epithelial homeostasis, with many check points in place to ensure faithful function.

1.3 Musashi functions in neural, gastrointestinal and hematopoietic systems

The Musashi (Msi) protein family is a set of conserved RNA-binding proteins with important developmental functions in the neural, hematopoietic, and gastrointestinal systems in many organisms including worm, fly, mouse and human (Ito et al., 2010; Kharas et al., 2010; Nakamura et al., 1994; Park et al., 2014; Sakakibara et al., 2002; Wang et al., 2015). Musashi comprises two known family members, *Msi1* and *Msi2*. *Msi1* was first characterized as being critical for sensory hair formation in Drosophila, with subsequent studies identifying and describing its roles in murine neural development and gastrointestinal maintenance (Nakamura et al., 1994; Sakakibara et al., 1996). *Msi2* was later discovered in mouse neural tissue in a screen of *Msi1* homologues (Sakakibara et al., 2001). Similar to *Msi1*, *Msi2* is involved in maintenance and development in neural and gastrointestinal systems in addition to being involved in hematopoietic function (Katz et al., 2014; Kharas et al., 2010; Sakakibara et al., 2001). Previous work on Msi proteins primarily focused on *Msi1*, with limited studies on *Msi2*. However, interest in *Msi2* started peaking when a correlation was noted between *Msi2* protein levels and malignancy of various tumors, discussed below (Mu et al., 2013). This family of proteins has increasingly received attention in literature due to their developmental roles and their oncogenic potential.

1.3.1 Msi1 in the Neuronal Tissues

Msi1, the first Musashi member discovered, was characterized as being involved in sensory hair development in Drosophila (Nakamura et al., 1994). Specifically, loss of *Msi1* led to loss of neural identity in the Drosophila hair. The result was a twin hair

phenotype, thus the name Musashi, after the legendary Japanese samurai Miyamoto Musashi, who purportedly wielded two swords in battle. Later, *Msi1* was found to be required for neural progenitor cell maintenance in mouse neurons (Dobson et al., 2008; Imai et al., 2001; Sakakibara et al., 1996). Loss of *Msi1* in mouse brains leads to mental defect through attenuated brain development via loss of asymmetric cell divisions and loss of the stem cell populations. Additionally, *Msi1* controls neural midline crossing in the brain (Kuwako et al., 2010). This brain development defect was observed for the *C. elegans* Msi homolog as well. In *C. elegans* Msi plays roles in memory and male mating behavior by regulating actin dynamics and neural maintenance, respectively (Hadziselimovic et al., 2014; Yoda et al., 2000).

Msi2 was discovered seven years after *Msi1*, in a screen for *Msi1* homologues (Sakakibara et al., 2001). Like *Msi1*, *Msi2* was found to be important in neural development. Studies on both proteins in the central nervous tissue revealed that *Msi2* maintains a similar expression pattern and presents a similar regulatory role as *Msi1* (Sakakibara et al., 2002). However, *Msi2* was found to have a broader expression pattern than that of *Msi1*, where it maintains a different set of functions. Thus, in neurons, Msi proteins maintain similar functions in regulating development and maintaining stem cell or proliferative dynamics. In support of the critical aforementioned roles, over-abundance of Msi proteins have been found in a variety of neural tumors (Cox et al., 2013; Nikpour et al., 2011; Uren et al., 2015; Vo et al., 2012). Interestingly, examination of *Msi1* in glioblastoma demonstrated that *Msi1* increased migration as its loss caused a decrease in average velocity of cells (Uren et al., 2015).

1.3.2 Msi2 in Hematopoietic System

Not long after its discovery, *Msi2* was linked to myeloid leukemia (Barbouti et al., 2003). Analysis of different leukemia patients found that those with higher levels of *Msi2* had a poorer prognosis for recovery and a higher chance of relapse (Han et al., 2015; Kaeda et al., 2014; Mu et al., 2013). This spurred investigation into the role *Msi2* plays in hematopoietic tissue. Studies using marrow transplants in mice found that loss of *Msi2* in the hematopoietic lineage results in stem cell depletion and a breakdown of normal tissue functions (Andrés-Aguayo et al., 2011; Kharas et al., 2010; Park et al., 2014). In studies in which mice whose hematopoietic system was destroyed through irradiation and grafted with donor marrow lacking or overexpressing *Msi2*, it was found that marrow lacking *Msi2* could not efficiently graft, and caused the mice to become anemic and unable to replenish their blood due to loss of the stem cell population. Conversely, overexpression of *Msi2* drove pathologic cellular proliferation. Mice that had been grafted with *Msi2*-overexpressing marrow developed blood problems associated with poorly differentiated blood cells. While the animals did not lack a stem cell population, they were unable to generate properly differentiated cell populations. Indeed, overexpression of *Msi2* could drive a more malignant leukemia (Ito et al., 2010; Kharas et al., 2010). Thus, as in neurons, *Msi2* controls proper stem cell maintenance and proliferation in the hematopoietic system, and can contribute to oncogenesis. This observation led to the study of *Msi2* in other systems where it is highly expressed.

1.3.3 Msi in the Gastrointestinal System

Msi2 was found to be highly expressed in crypt cells in the intestine (Katz et al., 2014). Perturbations in its expression in this environment showed similar effects to that of the hematopoietic system (Katz et al., 2014; Li et al., 2015; Wang et al., 2015a). Loss of *Msi2* led to gastrointestinal crypt deterioration and an inability to maintain the high rate of proliferation required for crypt maintenance. Conversely, overexpression of *Msi2* led to polyp formation from over proliferation. Additionally, *Msi1* has been linked to more malignant gastrointestinal cancers (Ji Eun Choi, Jun Sang Bae, Ju Hyung Lee, Kyu Yun Jang, Myoung Ja Chung, 2014; Ravindran and Devaraj, 2014). A recent study demonstrated that *Msi1* and *Msi2* are functionally redundant in intestine (Li et al., 2015).

1.3.4 Msi in Pluripotent Embryonic Cells and Spermatogenesis

The function of Msi proteins in embryonic tissues was observed in Xenopus oocytes as being involved in temporal regulation of target mRNAs (Charlesworth et al., 2006). Subsequent studies in mammalian embryonic stem cells demonstrated that *Msi2* plays a critical role in maintaining an embryonic stem cell state (Wuebben et al., 2012). Without *Msi2*, embryonic stem cells differentiated, lost the ability to efficiently proliferate, and thus were lost throughout subsequent passages in culture, highlighting the importance of *Msi2* in maintaining a proliferative stem cell state.

Even prior to fertilization, Msi proteins play critical roles in proper spermatogenesis (ErLin et al., 2015; Fraser et al., 2008; Sutherland et al., 2000, 2014). Loss of Msi proteins leads to fertility defects in mice and *C. elegans* characterized by loss of gonad structure and function. Taken as a whole, these data show that Msi

proteins play critical roles in cellular proliferation and demonstrate their oncogenic potential. Additionally, Msi proteins maintain stem cell identity and prevent differentiation (Siddall et al., 2006). Alterations in Msi expression can be detrimental to normal tissue maintenance and can lead to abnormal cell growth or a loss of stem cell populations. Additionally, data have indicated that Msi proteins can serve other functions outside of their well characterized roles in proliferation and stem cell maintenance (Uren et al., 2015). Thus, it is possible that Msi proteins may maintain different functions in tissues where they have not been adequately characterized.

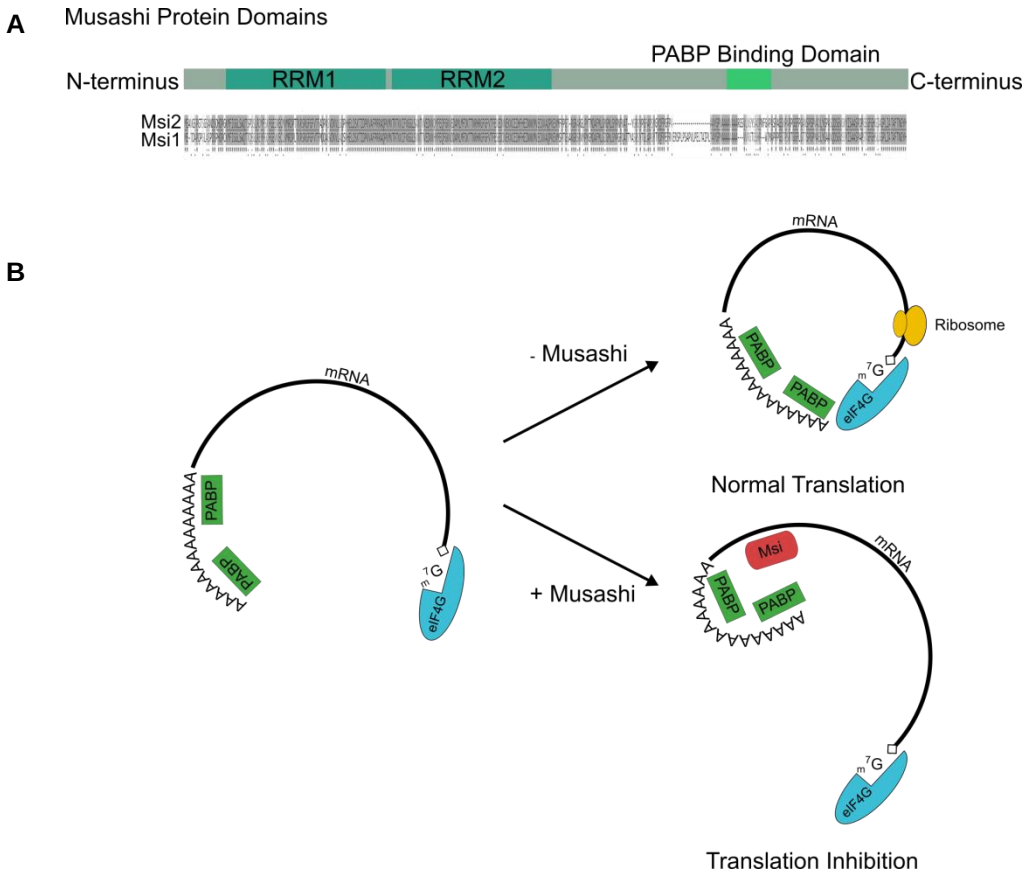
1.4 Mechanism of Musashi function

Phylogenetic analysis indicates that the Msi family is most closely related to the Dazap family of RNA-binding proteins involved in developmental processes, and can be found in most bilaterian animals (Kerner et al., 2011). The Msi family is characterized by containing a conserved N-terminus harboring the RNA-binding domain with two RNA-recognition motifs (RRM), and a less conserved C-terminus where regulatory activity is presumed to reside (Figure 3A) (Moore, 2010; Sutherland et al., 2013). *Msi1* has one known isoform whereas *Msi2* has at least two (Wuebben et al., 2012). Neither isoform appears to be dominant in the tissue types studied. One isoform is a splice variant, in which a region of the protein is omitted from the middle of the C-terminal domain. The two isoforms have overlapping functions; however, the splice variants may still contain independent functions that have yet to be characterized (Wuebben et al., 2012). Additionally, each isoform is potentially phosphorylated, as demonstrated by

incubating cellular lysates with phosphatases and observing a reduction in *Msi2*'s molecular weight (Sakakibara et al., 2001).

Recent work has begun to examine the roles, binding, and mechanisms of *Msi1* and *Msi2* in different tissue types (Li et al., 2015; Uren et al., 2015; Wang et al., 2015a). Additionally, a study investigated potential Msi inhibitors in an attempt to minimize the oncogenic potential of Msi proteins (Minuesa et al., 2014). Most work to date has focused on *Msi1*, based on the assumption that what is found for *Msi1* will hold true for *Msi2*. However, an increasing body of literature demonstrates that *Msi2* can function differently from *Msi1* (Katz et al., 2014; Wang et al., 2015a).

Figure 3: Mechanism of Musashi-mediated translational repression



(A) Diagram of Musashi protein domains with Msi2 versus Msi1 protein alignment and model of Musashi mechanism of action. Musashi proteins contain two RNA recognition motifs (RRM) in their conserved N-terminus and a Poly-A binding protein (PABP) binding domain in their divergent C-terminus. (B) Model of Msi mechanism of regulation. With Musashi binding, transcript poly-A tail is circularized to Musashi protein and away from the cap, preventing efficient ribosome initiation. Without Musashi binding transcript is circularized and efficient ribosome initiation can occur.

1.4.1 Msi RRMs and Binding

A number of studies have begun identifying the targets of Msi proteins to determine how they regulate their cellular functions (Park et al., 2014, 2015; Wang et al., 2015a). As mentioned previously, Msi proteins use two RNA-Recognition Motifs (RRMs), RRM1 and RRM2, to bind to and recognize their targets (Figure 3A) (Nakamura et al., 1994). The two RRMs can bind RNA independently of each other and have little protein structural modelling between them for binding of one to orient the binding of the other (Miyanoiri et al., 2003; Nagata et al., 1999; Ohyama et al., 2011). Due to this lack of structure, it is speculated that both RRMs can independently interact with motifs at length to facilitate more efficient binding, thus, adding a complication to motif prediction. RRM1 is the dominant binding module and predominantly recognizes a GUAG motif through base stacking interactions and hydrogen bonding (Nagata et al., 1999). RRM2 adds stability to target binding by recognizing a UAG with lower affinity using similar interactions to those of RRM1.

Msi1's preferred binding motif was identified, through SELEX, to be G/AU$_{(n)}$AGU (Imai et al., 2001). Of particular interest is the variable length of the uridine stretch in the first part of the motif. Some reports have indicated that the uridine stretch can be 1-3 residues long but others indicate it could be much longer (Ohyama et al., 2011).

Further biochemical studies demonstrated that both Msi proteins require a core UAG to be able to bind (Zearfoss et al., 2014). Mutations to any of these residues were shown to abolish binding *in vitro*, with smaller effects to binding constants by mutations outside of the core motif. The similar binding requirement between *Msi1* and *Msi2* is not surprising due to the high degree of similarity between their RRM domains and the critical residues in the domains. This is supported by the observation that ablation of both *Msi1* and *Msi2* in neural organoids synergistically leads to deterioration of the stem cell state (Sakakibara et al., 2002). Interestingly, the UAG core motif is not detected from the published Msi-CLIP studies (Wang et al., 2015). This implies that: the UAG is not involved with binding *in vivo*, there is a secondary structural element required, or that there is too much noise in the CLIP data to detect a *bona fide* recognition motif. A few studies support a secondary structural requirement for Msi binding by finding the motif or Msi binding site residing at the terminal loop of a stem loop (Charlesworth et al., 2006; Imai et al., 2001). This is supported by the observation that *Msi2* co-regulates a miRNA with HuR by binding the terminal loop of the pre-miRNA stem loop (Choudhury et al., 2013).

1.4.2 Msi Targets

Many of the Msi identified targets are involved in proliferation (Battelli et al., 2006; Gao et al.; Imai et al., 2001; Liu et al., 2014). By repressing negative proliferation regulators Msi proteins can promote cell growth and survival, consistent with Msi's oncogenic activities. One of the first studies demonstrated that NOTCH-signaling inhibitor numb was a target of *Msi1* (Imai et al., 2001; Nishimoto and Okano, 2010).

Additionally, cell cycle inhibitor p21 was also determined to be an Msi target (Battelli et al., 2006). Recent studies have begun to identify Msi targets using CLIP methods (Park et al., 2014, 2015; Vo et al., 2012; Wang et al., 2015a). These data further support the role of Msi proteins in targeting genes associated with proliferation. Additionally, important signaling regulators have been identified, such as Jag1 and PTEN (Chen et al., 2014; Wang et al., 2015a). Much of this data indicates that Msi proteins, like other post-transcriptional regulators, bind to elements in the 3'UTR of transcripts (Park et al., 2014; Uren et al., 2015; Vo et al., 2012). This is common for many RNA-binding proteins (Vindry et al., 2014). This data provided insight into how Msi proteins regulated stem cell dynamics and why they demonstrate oncogenic activity.

1.4.3 Mechanism of Msi Regulation

Some of the first studies on *Msi1* determined that Msi regulates its targets by inhibiting translation initiation without significantly altering RNA abundance (Battelli et al., 2006; Imai et al., 2001; Kawahara et al., 2008). *Msi1* does so by interacting with poly-A binding protein (PABP) and titrating PABP from the 5' cap to inhibit mRNA circularization, thus, inhibiting efficient translation initiation (Figure 3B) (Kawahara et al., 2008). *Msi1* contains a PABP-binding domain, which allows it to preferentially recruit the poly-A tail to itself rather than the cap. This interaction is more stable than that of PABP:cap interaction, causing more favorable interaction to occur with itself. Reporter assays on *Msi2* targets suggest a similar mechanism of action (Sakakibara et al., 2001). Reporter protein levels were decreased but mRNA levels were not drastically altered. Indeed, *Msi2* contains a predicted PABP-binding domain in its C terminus, suggesting it

may interact with PABPs. However, the majority of sequence differences between *Msi1* and *Msi2* are located in their C termini, leaving the possibility that *Msi2* employs alternative modes of regulation (Figure 3A). In support of this finding, one group noticed that the mRNA levels changed slightly in response to loss of *Msi2* (Katz et al., 2014). However, this discrepancy with the presumed function of *Msi2* was dismissed as a secondary effect from sequencing or on loss of ribosome occupancy.

1.4.4 Regulation of Msi Proteins

With what is known about *Msi1* biology and oncogenic potential, a set of studies sought to explore the upstream processes regulating *Msi1* as opposed to how *Msi1* regulates downstream targets (Arumugam et al., 2012a, 2012b; Clingman et al., 2014; Kawase et al., 2014; Sutherland et al., 2015b; Vo et al., 2012). One study found that, in neural progenitor cells, *Msi1* binds to retinoic acid upon neural differentiation through a lipid-binding pocket near its RRMs (Clingman et al., 2014). This binding efficiently blocked *Msi1* target recognition and regulation, providing a way to rapidly turn off *Msi1* protein function and promote subsequent cellular differentiation. While this report demonstrated a way for cells to inhibit protein function, another group demonstrated that Rfx neural transcription factors may be involved in transcriptional regulation of *Msi1* (Kawase et al., 2014). Yet another study found that *Msi1* can inhibit its own mRNA and prevent inappropriate protein accumulation (Arumugam et al., 2012a). Our understanding of how Msi protein levels and functions are regulated is limited in contrast to what we know about Msi's regulation of downstream targets.

1.5 Unknowns in the Musashi 2 field

The data presented in the previous section paint a model of Msi function and mechanism of action. Msi proteins regulate stem cell states and proliferation by inhibiting the translation of negative regulators of cellular proliferation (Battelli et al., 2006; Gao et al.; Imai et al., 2001; Liu et al., 2014). They do this by binding to the 3' UTR of transcripts with a UAG-containing motif, in a possible secondary structure, and titrating PABP from the 5' cap of the transcript (Moore, 2010; Sutherland et al., 2013). However, these data are most supportive for *Msi1*. There is less data indicating *Msi2* follows this model. First, no one has formally demonstrated that *Msi2* circularizes the poly-A tail to itself or inhibits translation initiation as in the case of *Msi1*. In fact, there is some indication that *Msi2* may regulate RNA levels rather than translation, but this is not yet confirmed (Katz et al., 2014).

Data from *in vitro* derived *Msi2* motif suggest that *Msi2* regulates a similar set of targets as *Msi1* (Zearfoss et al., 2014). However, while there does appear to be some level of redundancy between *Msi1* and *Msi2*, an argument can be made based on embryonic stem cell knockout data that *Msi2* can regulate a distinct set of targets from that of *Msi1* (Sakakibara et al., 2002; Wuebben et al., 2012). Additionally, *Msi1* and *Msi2* display a unique expression profile in various tissues and cell types (Sugiyama-Nakagiri et al., 2006). While there is a common theme to the processes regulated by Msi proteins, there is a possibility they can regulate novel pathways and targets in cell types where they have not been characterized. This view is supported by *Msi2* regulation of migration associated gene Jag1 in the gastrointestinal system (Katz et al.,

2014). *Msi1* is not known to regulate Jag1, leading to the possibility that Jag1 is unique to *Msi2* regulation.

Additionally, it can has been demonstrated that HuR and *Msi2* cooperate to regulate miR-7 biogenesis in neural cells (Choudhury et al., 2013). The implication is that *Msi2* may be able to perform other functions or mediate novel protein-protein interactions. To date, there is no information on which proteins interact with *Msi2* or whether *Msi2* interacts with anything else at all. Nor is there any data on cellular mechanisms regulating *Msi2* expression or activity. Taken as a whole, the data characterizing *Msi2* are lacking. While recent studies are beginning to look into *Msi2*, questions still remain and more studies are required.

1.6 Summary and Aims for Thesis

Despite the critical functions of *Msi2* in tumorigenesis and animal development, many important questions involving *Msi2* remain unanswered. While many studies have focused on the role of *Msi2* in neural, hematopoietic, and gastrointestinal systems, few researchers have looked in other tissues where *Msi2* is expressed, like skin epithelium (Ito et al., 2010; Kharas et al., 2010; Nakamura et al., 1994; Park et al., 2014; Sakakibara et al., 2002; Wang et al., 2015). Interestingly, previous work in the Yi lab identified *Msi2* as a target of miR203 and an important regulator of keratinocyte, epithelial progenitor cell, colony forming ability (Jackson et al., 2013a). It is possible that *Msi2* performs distinct functions or regulates novel pathways in the skin. Additionally, it is not fully understood how *Msi2* regulates its targets, what processes it can regulate, or what targets it binds and what elements characterize that binding.

There are many ways to approach these problems. Here I will describe a transcriptome-wide sequencing approach using HITS-CLIP, ribosome profiling and RNA sequencing to begin to answer these questions. This is the first study to use all three of these sequencing technologies to characterize an RNA-binding protein in mouse skin.

In Chapter 2, I describe the expression of *Msi2* in keratinocytes and skin, and describe the binding interactions mediated by *Msi2* as determined through HITS-CLIP. These data provide insight into the rules and mechanisms that govern *Msi2* binding and target recognition. In Chapter 3, I present RNA-seq and Ribo-seq analysis coupled with target information to determine *Msi2*'s mode of target regulation. Analysis of these data reveals whether *Msi2* regulates its targets through RNA stability or, as is the case for *Msi1*, translational repression. In the same line of analysis, I identify the pathways that *Msi2* regulates in epidermal cells. In Chapter 4, I explore the phenotypic consequences of *Msi2* loss. My results reveal how loss of *Msi2* affects the targeted processes in keratinocytes. Finally, in Chapter 5, I discuss my results in light of current knowledge of *Msi2*, what questions have been addressed, and what has been learned. I conclude by suggesting further lines of experimentation to extend upon the framework established within this study and the growing field of Msi biology.

CHAPTER 2

TRANSCRIPTOME-WIDE IDENTIFICATION OF *MSI2* TARGETS

2.1 Introduction

Identification of RNA-binding protein targets has been an evolving process over the decades (König et al., 2012). In the early days, target identification largely revolved around intuitive guess work, gel shifts, and other involved processes. These techniques were effective at identifying a handful of targets but lacked breadth in transcriptome-wide determination. RNA-immunoprecipitation (RIP) coupled with micro arrays (RIP-chip), qPCR (RIP-qPCR) or RNA sequencing (RIP-seq) allowed for broader detection of targets that could be identified and characterized. However, these techniques and others like it suffer from potentially high background and false positives and false negatives due to the stochastic nature of binding between RNA-binding proteins and their targets. Recently, new techniques, called Cross-Linked Immunoprecipitation (CLIP), have been developed to preserve native RNA-protein interactions (Kishore et al., 2011). A variety of CLIP methods that have now been developed, each with varying degrees of successes and applications (König et al., 2012).

2.1.1 HITS-CLIP

One of the more well-known CLIP techniques, called HITS-CLIP, was developed in 2009 (Wang et al., 2009a). It was originally designed to detect Argonaute:miRNA target transcripts through direct crosslinking of the protein with its multiple mRNA targets. Many groups have adapted HITS-CLIP to their RNA-binding protein of interest (Bracken et al., 2014; Chi et al., 2009; Macias et al., 2012; Weyn-Vanhentenryck et al.,

2014). HITS-CLIP involves using ultraviolet light to directly crosslink RNA and proteins preserving RNA:protein interactions (Figure 4).

Figure 4: Diagram of steps in HITS-CLIP protocol

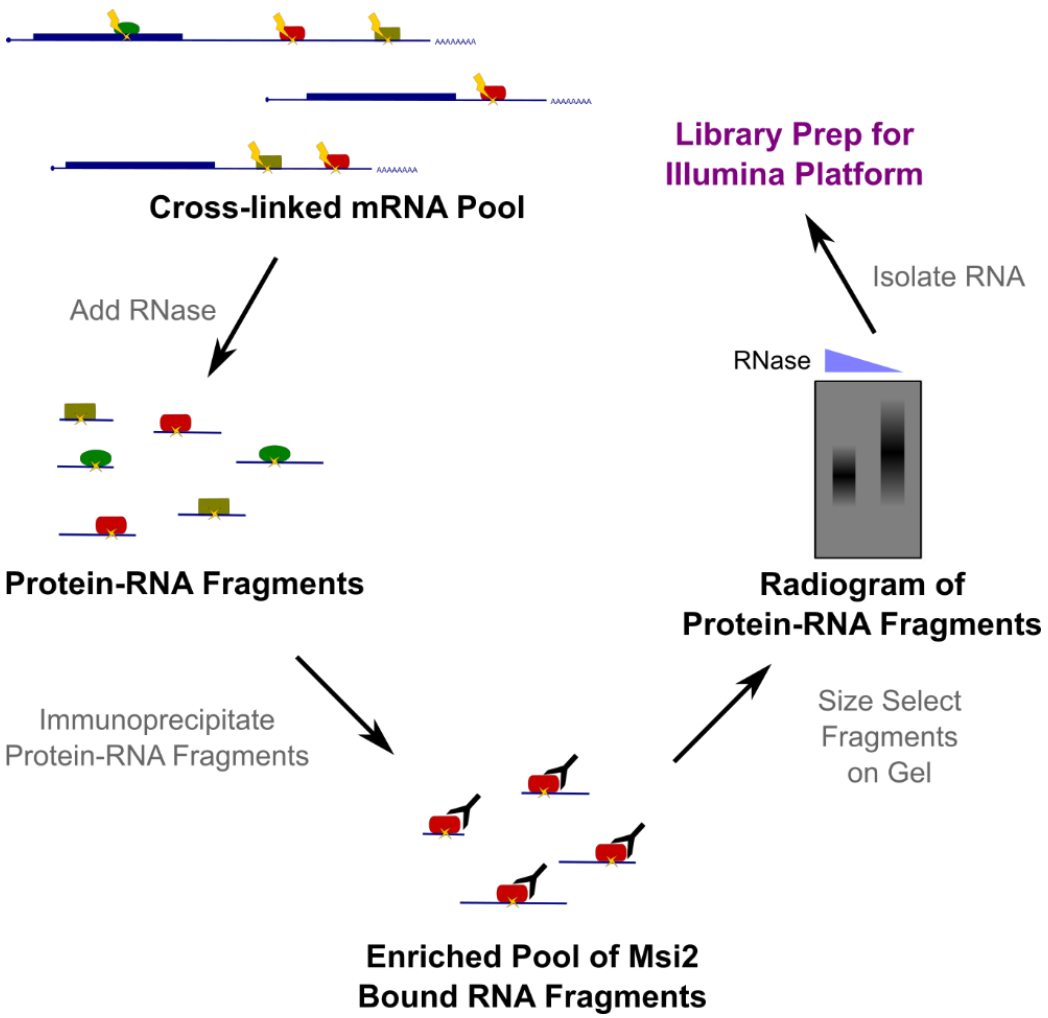


Diagram of steps in HITS-CLIP protocol. RNA binding proteins and mRNA are cross-linked using UVC. The RNA is degraded using RNase leaving behind RNA:Protein hybrids. The protein of interest is immunoprecipitated and protein:RNA fragments are size selected. The RNA is isolated and turned into libraries for sequencing on an Illumina Platform.

The crosslinking between the RNA-binding protein and the RNA is such that *in vivo* interactions are maintained and can be identified. The RNA-binding protein of interest is then immunoprecipitated, excess RNA is degraded using RNases, and the protein is

removed from the crosslinked RNA using protease. The resulting RNA fragments are then sequenced to identify the binding sites for the RNA-binding protein. The localization of the interactions can be assessed from mapping the fragments to the genome, e.g., 3'UTR, coding sequence, etc. Additionally, if the RNA is sufficiently cleaved down to ~20-50 nucleotide sizes, HITS-CLIP can determine precise binding sites. This provides information into the regions the RNA-binding protein primarily prefers to bind, and can be extended to show motifs enriched for certain RNA-binding proteins. HITS-CLIP can be performed on endogenous RNA-binding proteins and can be used to assay the transcriptome-wide array of binding partners. This technique can still suffer from high background and libraries with low fragment complexity, but it provides a better view than all other techniques of the direct RNA interactions mediated by an RNA-binding protein.

2.1.2 Previous Msi2 CLIP

To date, HITS-CLIP has been used on many different RNA-binding proteins, including Argonautes, Rbfox, and DGCR8 (Chi et al., 2009; Macias et al., 2012; Weyn-Vanhentenryck et al., 2014). Recent studies have been published using *Msi2*-HITS-CLIP (Park et al., 2014, 2015). However, while their data are valuable and provide some insight into the interactions mediated by *Msi2*, these studies suffer high background signal in their data and thus have not been able to identify a coherent set of motifs. Additionally, these studies performed CLIP in hematopoietic systems, leaving the opportunity to identify new *Msi2* targets within a different transcriptional landscape.

In this chapter I characterize the expression of *Msi2* in skin and identify *Msi2* targets using HITS-CLIP in keratinocytes.

2.1.3 Chapter Summary

The full breadth of *Msi2* RNA binding is unknown. Additionally, it is unknown what transcript regions and context *Msi2* prefers to bind. The first step I took was to identify and isolate a population of *Msi2*-expressing cells from mouse skin for further study. Immunostaining was used to assay *Msi2* expression in different skin cell populations at embryonic day 14 (E14), 18 (E18), postnatal day 6 (P6), and in adult. The basal keratinocyte population was identified as expressing high levels of *Msi2* throughout development. These cells were isolated and cultured as described in the methods. *Msi2*-HITS-CLIP was then performed on the cultured keratinocytes to identify *Msi2* targets. A low background autoradiogram signal confirmed that *Msi2* can be CLIPed in keratinocytes. Libraries were generated and the data were filtered for biological complexity and read numbers at various cut-offs to determine the optimum filters to use. From this filtered dataset *Msi2* binding characteristics were generated. I found that *Msi2* predominantly interacts with 3' UTRs of potential mRNA targets. De novo motif searching in these regions identified an enrichment of multiple UAG-containing motifs.

2.2 Materials and Methods

2.2.1 Animals

Mice were bred and housed according to the guidelines of the IACUC at a pathogen-free facility at the University of Colorado (Boulder, CO, USA). Tissues were

harvested for immunofluorescence as previously described (Yi et al., 2006). All procedures involving mice were approved under IACUC protocol 1408.01.

2.2.2 Keratinocyte Cell Culture, Viral Infection, and Growth Assays

Keratinocytes were isolated from neonatal epidermis as previously described (Riemondy et al., 2015). Keratinocytes were initially plated on a NIH-3T3 subclone J2 feeder layer for three passages, after which continuously proliferating keratinocytes were propagated in the absence of feeder layer support. Keratinocytes were grown in the presence of E-Low Media with 0.05 mM Ca^{++} at 37°C in 5% CO_2. Lentiviral shRNA knockdowns were performed using PLKO shRNA constructs (Appendix 11). Lentivirus was generated by transiently transfecting HEK-293FT cells with pVSVG, psPAX.2, and relevant PLKO constructs at a 1 : 2.5 : 4 mass ratio using MiRus Bio LT1 (Mirus Bio LLC). Retroviral *Msi2* overexpression was performed using MIGR-Vector and MIGR-*Msi2* constructs (Jackson et al., 2013b). Retrovirus was generated by transiently transfecting HEK-293FT cells with pCL-Eco, pAdvantage, and relevant MIGR constructs at a 1:1.7:8.7 mass ratio using MiRus Bio LT1. Viral supernatant was harvested 24-72 hours post transfection and filtered with a 0.45 μM filter. Puromycin was added (2 μg/ml) to select for shRNA producing keratinocytes. Cellular growth curves were calculated by trypsinizing cultures and counting the number of cells post-plating on a hemocytometer. Colony formation assays were performed by plating 1,000 cells per well of a 6 well plate in triplicate and culturing for 5-10 days. Colony density was calculated using imageJ (images set to 8-bit to convert to grayscale, threshold set and applied across all colony images to convert to black (colony) and white (background), wells were selected with

the circle tool and particles were analyzed to generate area covered by colonies. Colony area was normalized to PLKO-shScr/MIGR-Vec control).

2.2.3 Immunofluorescence and Immunocytochemistry

OCT skin tissue sections (10 μM) were fixed for 10 minutes with 4% paraformaldehyde at room temperature, washed three times for 5 minutes with PBS, and blocked for 10 minutes using Gelatin Block (0.1% Triton X-100, 2% gelatin, 2.5% normal goat serum, 2.5% normal donkey serum, and 1% BSA in PBS). Sections were incubated with the following primary antibodies in blocking buffer overnight at 4°C: Keratin 5 (Krt5), E-cadherin (Cdh1), and Musashi-2 (*Msi2*) (See Table S2 for antibody information). Following three 10-minute washes in PBS, sections were incubated with appropriate Alexa Fluor secondary antibodies (1:2000) for 1 hour and washed two times for 5 minutes with PBS. Sections were incubated with Hoechst Dye for 10 minutes and mounted in VectaShield anti-fade solution.

For focal adhesion staining, keratinocytes were sparsely plated on 1 μg/ml fibronectin-coated #1.5 thickness coverslips. Cells were fixed in 4% paraformaldehyde with 0.5% Triton-X-100 in PBS at 37˚C for 10 minutes, followed by three PBS washes and storage at 4˚C for at least overnight. Cells were next blocked for 10 minutes in the Gelatin Blocking buffer at room temperature, then incubated with mouse anti-vinculin antibody in Gelatin Blocking buffer for 2 hours at room temperature (See Table S2 for antibody information). Coverslips were next washed 3 times with PBS for 10 minutes, followed by co-incubation with Alexa Fluor 488 anti-mouse IgG1a secondary antibody at 1:2000 and phalloidin-Alexa Fluor 647 at 1:50 in Gelatin Blocking buffer for 1 hour.

Coverslips were washed twice with PBS for 10 minutes followed by incubation with Hoechst Dye for 10 minutes and mounted with VectaShield anti-fade.

2.2.4 Msi2 HITS-CLIP

Msi2 HITS-CLIP was performed as previously described for Ago2-HITS-CLIP with minor modifications (Riemondy et al., 2015). Briefly, 15 cm dishes of mouse keratinocytes were irradiated with 200 mJ/cm using 254 nm UVC light, harvested by scraping, and stored at -80°C. After lysis, the lysates were then treated with 10 μL per mL lysate Turbo DNase (Thermo Fisher), 5 μL per mL lysate RNase OUT (Thermo Fisher), and partially digested with 10 μL of either a 1:1, 1:20, 1:50 or a 1:75 dilution of an RnaseA/T1 mix (Sigma/Ambion 1x mix = 3.33μl RNase-A (2 μg/μl) with 6.66 μl Rnase-T1 (1 U/μl). Crosslinked *Msi2* was immunoprecipitated for 2 hours at 4°C using 5 μg of an anti-musashi 2 antibody (Appendix 10) complexed with Protein-G Dynabeads (Thermo Fisher). After end-labeling, 5'adaptor ligation, and phosphatase treatment, *Msi2*-RNA complexes were resolved on a 10% Novex Bis-Tris gel (Thermo Fisher) and transferred to a nitrocellulose membrane (GE Healthcare Life Sciences). The nitrocellulose was subsequently exposed to X-ray film and a phosphor screen overnight. Protein-RNA complexes migrating between 40-60 kDa for the 1:1 and 1:20 RNase dilutions or 70-160+ kDa for the 1:20, 1:50, and 1:75 RNase dilutions were isolated from the nitrocellulose. RNA was extracted by Proteinase K treatment followed by acidic phenol-chloroform extraction and ethanol precipitation. After ligation and reverse transcription, the cDNA was subjected to 20 cycles of PCR, purified on a 10% native PAGE gel, and subjected to 1x100 sequencing on an Illumina HiSeq 2000.

2.2.5 Msi2-HITS-CLIP peak identification

Raw reads were first filtered to keep only high quality sequences (CIMS package, fastq_filter.pl, settings = -f mean:0-24:20), then trimmed to remove 5' and 3' adapter sequences (Cutadapt, default settings) (Appendix 9 for oligo sequences). Sequence duplicates were next discarded to remove reads arising from PCR duplicates, followed by removal of the 2 NN's introduced by the adaptors from the 5' and 3' ends. Reads were next aligned to the mouse genome (mm10) using Novoalign (autothreshold alignment with -s 1 setting), requiring a minimum 20-nucleotide alignment length. Uniquely mapped alignments were next filtered to remove duplicate alignments with identical start and stop coordinates, which likely represented PCR duplicates with sequencing errors or PCR introduced mutations. Continuous alignments were next clustered to identify *Msi2* binding sites (Bedtools merge, setting = -s). Binding sites were then split to resolve multiple summits within long peaks using PeakSplitter (-v 0.6). Peaks were filtered to keep only peaks with at least 10 total reads, with at least one read present in at least 4 of 5 libraries (Biological Complexity ≥4). Alignment statistics for each library are shown in (Appendix 1). Peaks were annotated to the genome using combined Gencode (VM4) and UCSC (mm10) annotations downloaded from the UCSC table browser, with regions selected based on at least a single nucleotide overlap with the first 3'UTRs, followed by CDS, 5'UTRs, ncRNAs, introns from protein coding genes, and lastly intergenic regions (bedtools intersect, setting = –s –u). ncRNA regions were defined by selecting Gencode transcripts with a non-coding transcript class ID, and by selecting UCSC knownGene transcripts with identical cdsStart and cdsEnd coordinates.

3'UTR regions were extended 5 kbp or up to the next downstream genomic feature if a feature occurred closer than 5 kbp downstream. Peak summits were identified by calculating reads coverage across the peak, then by identifying the region with the maximum reads coverage. If the read coverage was equivalent across a region larger than a single nucleotide, then the middle of the region was selected as the summit (Bedtools coverage, setting = -s –d). For all analysis steps that do not reference a particular software package, in-house R, Python, or Bash scripts were used to perform the analysis.

2.2.6 RIP-qPCR

RNA Immunoprecipitation followed by qPCR was performed as follows. One 15 cm plate of cultured keratinocytes was harvested via scraping and stored at -80˚C until lysis with non-denaturing lysis buffer (2mM EDTA, 20mM Tris-HCl, 137mM NaCl, 1% NP-40, 1x Pierce Protease Inhibitor Cocktail, 40 U/µL RNaseOUT). The lysate was split and added to 25µL Protein G Dynabeads (Thermo Fisher) pre-bound with 5 µg anti-*Msi2* antibody or Rabbit IgG (Appendix 10). The antibody-bead, lysate mix was allowed to incubate for 2 hours at 4˚C. The supernatant was removed and the beads were washed three times with NT2 buffer (50mM Tris-HCl, 150mM NaCl, 1mM $MgCl_2$, 0.05% NP-40) before the RNA was extracted using TRizol (Thermo Fisher). An aliquot of total lysate and supernatant was taken for RNA extraction via TRizol and western blotting for determining the efficiency of the IP. RNA samples were subsequently used for qPCR (Appendix 9). Relative enrichment was calculated as enrichment over IgG control using *Gapdh* and *Hprt* as non-target controls.

2.2.7 Motif Searching

Msi2 HITS-CLIP 3'UTR binding sites were used to identify de novo motifs. A region extending 20 nucleotides 5' and 3' of the peak summit was selected for motif searching. De novo motif searching for 3-9mers was performed using HOMER with sequence auto-normalization methods disabled with a supplied background 3'UTR database (findMotifsGenome.pl–size given –norevopp –rna –noweight –nlen 0 –bg –chopify). De novo motif searching within 5'UTR and CDS regions was performed in the same fashion. For intronic regions the background database was subset to included only 20,000 randomly selected regions to reduce computational burden for the motif searching. High resolution motif logos were generated by converting the HOMER motif output with Weblogo. The top motifs generated for each length (3-9mer) were defined as *Msi2* binding motifs (UAG, UAGU, UAGUA, UUAGUA, UAGUAGU, UAGUAGUA, GUAGUAGUA). The consensus sequence for each motif was then used to plot the positional distribution of motifs surrounding each peak summit, and to classify binding sites as motif containing or not motif containing. Positional motif densities were computed by enumerating consensus motif match start positions across an interval 110 nucleotides 5' and 3' of the peak summit and plotted with a bin size of 5 nucleotides. Randomized 3'UTR backgrounds were constructed by shuffling the genomic intervals throughout 3'UTR coordinates (bedtools shuffle, run individually for each strand). For assessing the multiplicity of UAG motifs within *Msi2* binding sites, the number of UAGs was enumerated in 50-nucleotide windows +/- 225 nucleotides from the peak summit position. 225 nucleotides of sequence 5' and 3' of the 450-nucleotide window were

analyzed as flanking regions as a background control in addition to randomly selected 3'UTR regions, which were defined as described above.

2.2.8 Metagene Profiles

Msi2 binding profiles along a scaled average mRNA were computed using a Python script. Briefly, 5'UTR, CDS, and 3'UTR regions were extracted from protein-coding Gencode transcripts that contained 5'UTR, CDS, and 3'UTR attributes and were expressed in keratinocytes (BaseMean >= 10 in shScr libraries). The longest transcript for each region was taken as representative for each gene. Reads coverage was then calculated across each exonic region excluding intronic sequences (pyBedtools bed12tobed6 and coverage, setting = -s –d). Reads coverage vectors for each region (5'UTR, CDS, and 3'UTR) per transcript were then constructed and normalized by library size to reads per million mapped, averaged across 100 bins (Scipy binned_statistic) and averaged over all transcripts. Peak coverage vectors were computed analogously and averaged over all transcripts without normalization for library size.

2.2.9 Analysis of publically available datasets:

Msi2-HITS-CLIP analyzed datasets were downloaded from GSE64388 and GSE62115 GEO datasets. For GSE64388 datasets, all 3'UTR HITS-CLIP peaks identified in all wild-type and *Msi2* transgenic HITS-CLIP datasets were used for the analysis. For GSE62115 datasets, human gene symbols were converted to mouse homologs using a database downloaded from Ensemble BioMart, and only genes with

3’UTR localized *Msi2* binding sites were used for the analysis. GO analysis was performed using DAVID with Gene symbols as input.

2.2.10 Statistical Analyses

All statistics were computed using either R or Excel. Statistical tests performed for each experiment are listed in the figure legends. For two sample, student t-tests, normality was assumed and one-way or two-way equal variance tests were performed. For qPCR experiments, the mean of all technical replicates was computed and used to represent one independent value for statistical testing. For analyses of empirical cumulative distributions, two-way Kolmogorov-Smirnov tests were used unless indicated in the figure legend. Sample sizes were chosen based on prior knowledge of variability in each assay. Neither I nor collaborators were blinded to sample identity during sample collection, processing, and analysis.

Data Access:

Raw sequencing and analyzed data are available as a GEO super series with accession number GSE71333.

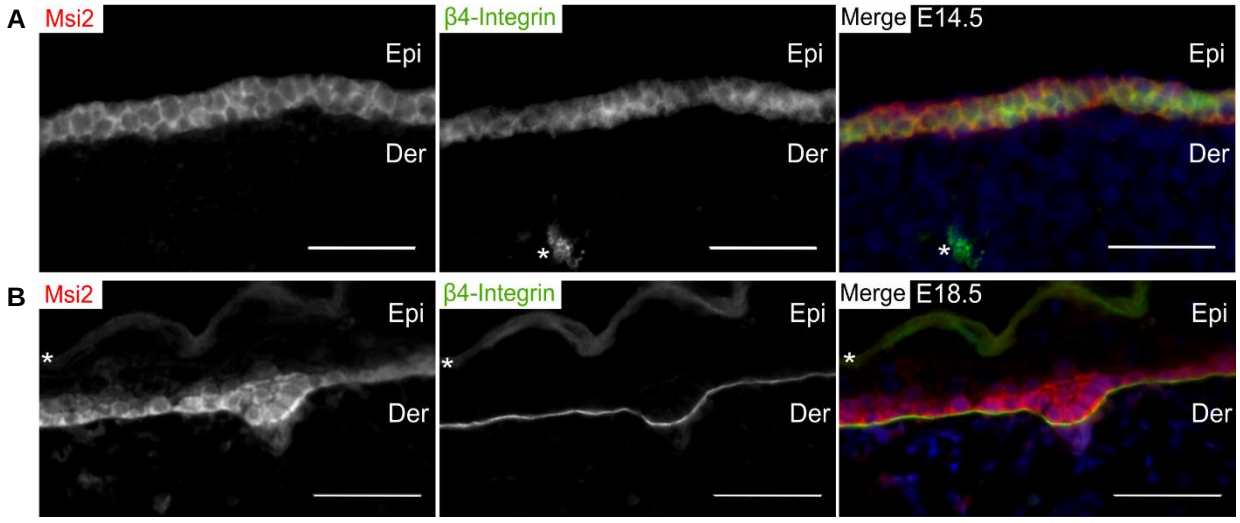
2.3 Results

2.3.1 Msi2 Localization in Skin

Previous work in the Yi lab identified *Msi2* as a target of miR-203 and an important regulator of keratinocyte, epithelial progenitor cells, colony forming ability

(Jackson et al., 2013a). However, the full expression pattern of *Msi2* in developing or adult skin has not been assessed.

Figure 5: Msi2 is expressed in proliferative cell populations in developing skin



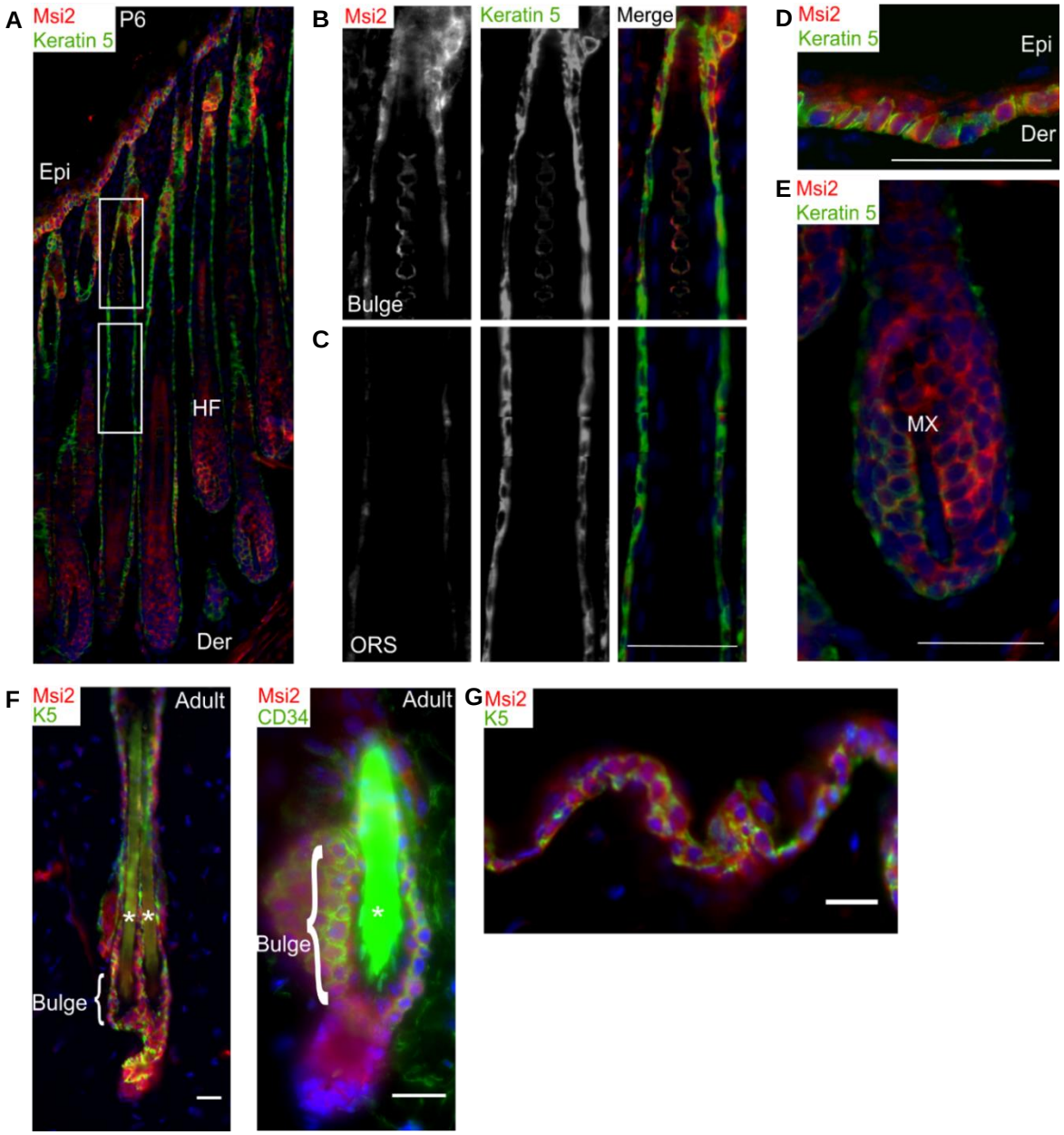
(A) Immunostaining of *Msi2* (red) and *Beta-4-integrin* (green), marking basal lamina, in sections of embryonic day 14.5 (E14.5) developing skin. (B) Immunostaining as in (A) in embryonic day 18.5 (E18.5) post stratification. Asterisks (*) mark nonspecific staining. Epi = Epidermis. Der = Dermis. Scale bar = 50 µm. Blue = dapi staining of nuclei

I hypothesized that *Msi2* is expressed throughout skin development in proliferative progenitor cells. In order to identify important *Msi2*-containing cell populations and elucidate a role for *Msi2* in skin development (Sutherland et al., 2013), I performed immunostaining throughout different developmental stages beginning at embryonic day 14 (E14). At this time point, a single layer of epidermal progenitor cells expresses *Msi2* (Figure 5A). After stratification at E18.5 *Msi2* expression is restricted to the basal layer of cells and excluded from the differentiated layers (Figure 5B). Additionally, *Msi2* is highly expressed in the early hair bud that will develop into a hair

follicle. These findings support the hypothesis that *Msi2* is required in progenitor cell populations during skin development.

These data, while informative, do not indicate whether *Msi2* is expressed in progenitor cell populations in prenatal or adult animals, when keratinocytes can be isolated. Staining at P6 and in adult animals indicates that *Msi2* is highly expressed in skin progenitor and stem cells *in vivo* in post-natal and adult animals (Figure 6). While *Msi2* is again expressed in the epidermal progenitor populations in P6 and adult tissues (Figure 6D, G), it also appears to be expressed in hair follicle stem cells (Figure 6B, F). Interestingly, *Msi2* expression is lost in the differentiated outer root sheath of the hair follicle but is regained in the highly proliferative but transiently amplifying matrix cells (Figure 6C, E). Previous studies demonstrate that *Msi2* is important for maintaining stem cell identity (Moore, 2010; Sutherland et al., 2013). From the expression pattern of *Msi2* this may be true for skin stem cells as well, thus it is of interest to study the function and mechanism of *Msi2* regulation in keratinocytes. Fortunately, keratinocytes can be isolated and cultured *ex vivo* and provide a strong system for studying the molecular mechanisms of a protein of interest, in this case *Msi2*. Since *Msi2* expression is restricted to keratinocytes in the epidermis and there is precedence in the Yi lab for its importance, keratinocytes were isolated and cultured for subsequent study.

Figure 6: Msi2 is expressed in proliferative skin cell population and skin stem cells



(A-E) *Keratin 5* (green), a marker of basal progenitor cells, and *Msi2* (red) immunostaining of postnatal day 6 (P6) skin of the (B) bulge stem cell compartment (C) outer root sheath progenitors (D) interfollicular epidermis (E) transit-amplifying cells in the Matrix. (F) Immunostaining of *Msi2* (red) and *Keratin 5* (K5, green) or *CD34* (green), a bulge marker, of the bulge stem cell compartment. (G) Immunostaining of *Msi2* (red) and *Keratin 5* (K5, green) in adult epidermis (Done by Kent Riemondy). Asterisks (*) mark nonspecific staining. Epi = Epidermis. Der = Dermis. Mx = Matrix. HF = Hair follicle. ORS = outer root sheath. Scale bar = 50 µm. Blue = dapi staining of nuclei.

2.3.2 Msi1 Expression in Keratinocytes

Since some evidence suggests that *Msi1* can compensate for *Msi2*, I wanted to know if *Msi1* is expressed in keratinocytes (Li et al., 2015). I used previously published RNA-seq data from total epidermis (Beronja et al., 2013) to find the relative expression of *Msi1* versus *Msi2* (Table 1). I found that *Msi2* is more than 9-fold more abundant than *Msi1*, suggesting that *Msi2* is the dominant Msi protein in the epidermis and should receive little compensation from *Msi1*. Furthermore, similar results were observed in RNA-seq and ribosome profiling (Ribo-seq) data from cultured keratinocytes used in this study (see Chapter 3), in which *Msi2* is ~12-fold more abundant than *Msi1* (Table 1). Lastly, a previous study in the Yi lab generated genome-wide proteomic data from neonatal murine epidermis (Wang et al., 2012). These mass spectrometry data demonstrated that *Msi2* protein is detectable, whereas *Msi1* is at undetectable levels (Table 1). These data support an important role for *Msi2* in skin epidermal progenitor cells and indicates that compensatory effects from *Msi1* will not confound any results. Thus, I did not further pursue a role for *Msi1*.

Table 1: Msi2 is the dominant Msi in keratinocytes

	RNA-seq Total Epidermis FPKM	RNA-seq Keratinocyte FPKM	Ribo-seq Keratinocyte RPKM	Mass Spec Total Epidermis Spectral Counts
Msi1	1.65	0.35	0.96	N/D
Msi2	15.08	4.53	21.38	4
$^{Msi1}/_{Msi2}$ *Ratio*	9.1	12.92	22.2	N/A

RNA-seq, Ribo-seq and Mass Spectrometry measurements of *Msi1* and *Msi2* in keratinocytes and total epidermal samples

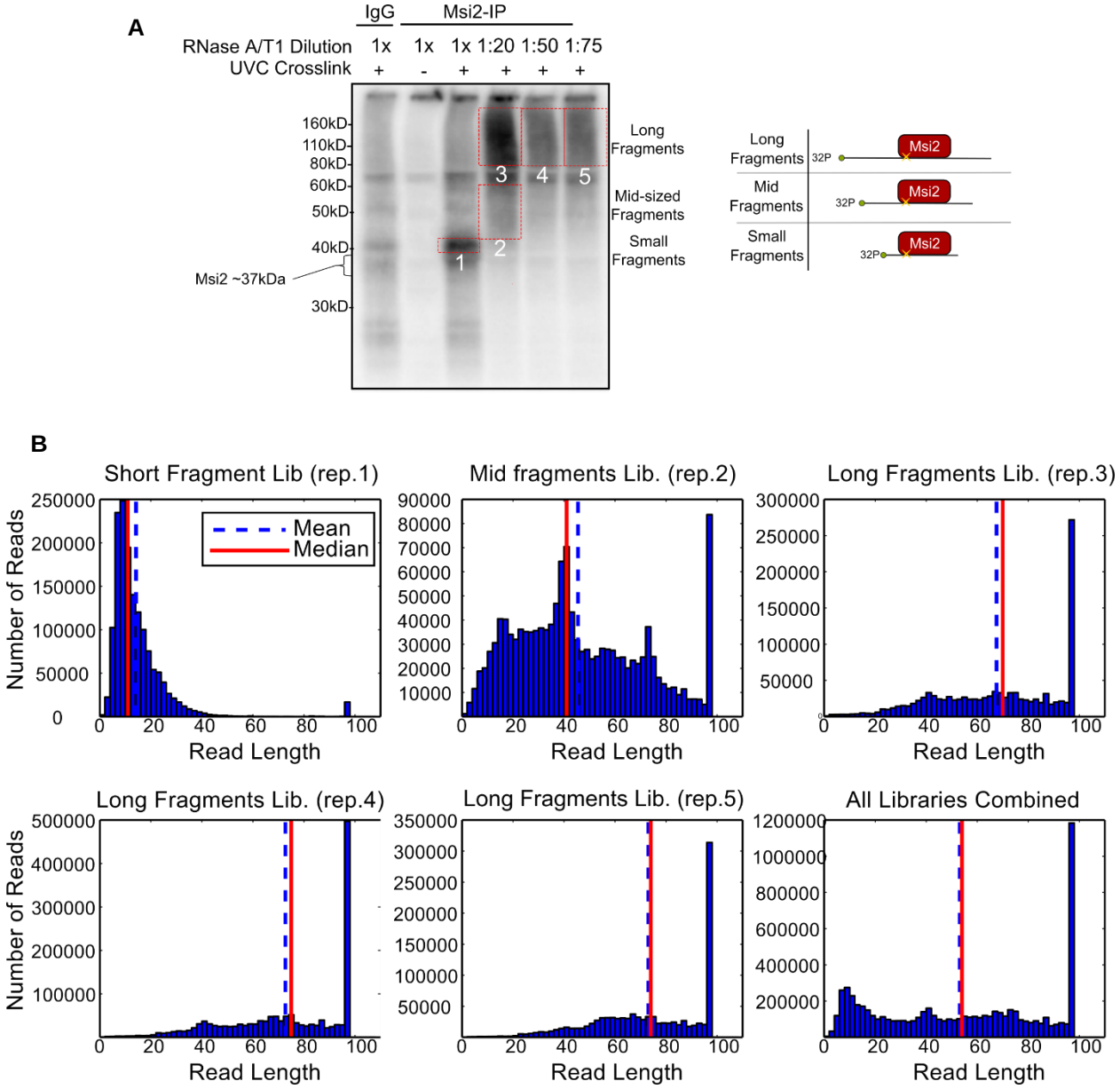
2.3.3 Msi2-HITS-CLIP Library

The first step I took to understand the function of *Msi2* in mouse keratinocytes was to identify its target mRNA. To this end, I used HITS-CLIP to identify all *Msi2* binding targets transcriptome-wide. UVC-linked RNA fragments crosslinked to *Msi2* were isolated from keratinocyte lysates treated with different concentrations of RNase A/T1 and radiolabeled for monitoring throughout the protocol. The isolated fragments were then visualized on nitrocellulose membranes, used to capture RNA-protein hybrids (Figure 7A). One expects to see shifts in radiolabeled RNA-protein hybrids correlating with RNase concentrations, with high RNase-treated samples showing smaller fragments compared to lower RNase concentrations. This result would give me confidence that what is being observed is RNA crosslinked to *Msi2*.

High RNase treatment resulted in a relatively uniform band with an apparent molecular weight of ~40kD, slightly larger than the anticipated 37 kD of *Msi2*, and decreasing RNase treatments produced broader smears between ~60-110 kD, confirming capture of an RNase-sensitive *Msi2*-RNA complex (Figure 7A). Five libraries with varying sized *Msi2*-associated RNA fragments were sequenced in order to obtain more replicates for target identification (Figure 7A). Additionally, this allowed for libraries of varying sized RNA fragments: short, medium, and long. The long fragments should contain more uniquely assigned reads and can be useful for target identification, while the short and mid fragments can be more useful in determining the exact *Msi2* binding site. Indeed, analysis of the read lengths received from sequencing showed that the small RNA fragment library had an average length of less than 20 nucleotides,

the mid-length library centered around a mean length of ~50 nucleotides, and the long libraries were >100 nucleotides (Figure 7B).

Figure 7: HITS-CLIP libraries show varying read lengths



(A) Autoradiogram of the ^{32}P labeled *Msi2*-RNA complexes isolated and used for library preparation. The red boxes indicate regions of the nitrocellulose membrane excised for library preparation. The white numbers indicate the replicate number of the library. Diagram of Small, Mid, and Long fragments denote length of RNA bound to *Msi2*. (B) Reads distribution for each replicate or

all the libraries in aggregate. The reads displayed were first processed prior to plotting to remove low quality sequence, trimmed to remove 5' and 3' adaptors, collapsed to remove PCR duplicates, and further trimmed to remove 5' and 3' NN randomized dinucleotides introduced by the adaptors. Red bars indicate median read lengths. Blue bars indicate mean read lengths.

Each library yielded between 4-8 million reads for a total of 34 million reads (Appendix 1). The small RNA fragment library had only ~60,000 unique read alignments, most being smaller than the required 20 nucleotides, and could not be aligned. The mid-length and long libraries each produced on average ~800,000 unique read alignments. Many of the reads in each library were removed during positional and sequence collapsing, indicating that the libraries were low complexity. Nevertheless, ~350,000 unique cDNA reads, unambiguously mapped to the mouse genome, were recovered for peak generation.

2.3.4 Msi2 Transcriptome Binding

The largest portion of the mapped reads aligned to 3'UTRs (36.47%), followed by intronic, noncoding RNAs, intergenic regions, and coding regions (CDS) (Figure 8A). This indicated that most of the binding may be occurring in mRNA 3'UTRs and provided confidence that true *Msi2* binding events were captured. The data from all five libraries were merged, and peaks were generated from overlapping reads. ~25,000 peaks were identified, many containing only two reads. The peak alignment distribution did not significantly change from the distribution observed from read alignment alone (Appendix 2).

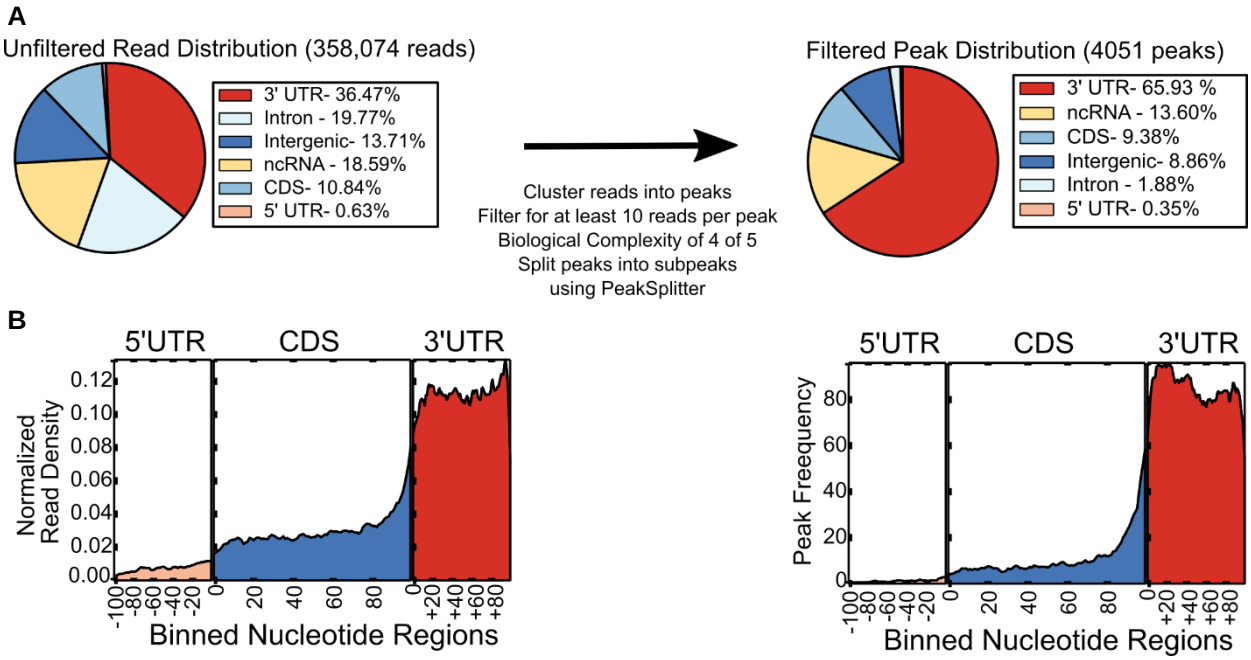
In order to confidently identify *Msi2* binding sites and targets, and remove background noise, the data were refined and filtered. To define an impartial filter for the

data, multiple total read counts per peak (10, 20, 30, 40, and 50 reads in each peak) and biological complexities (BC) (minimum number of libraries that must contain at least one read; BC 0, BC 3, BC 4, and BC 5) were tested simultaneously. Additionally, a side effect of the longer reads in some of the libraries led to what appeared to be overlapping peaks (Appendix 3), thus various PeakSplitter parameters (used to separate the peaks based on the percent decrease in the coverage or trough between peaks) were tested as well. As expected, the peak numbers dropped dramatically by increasing stringency in filtering (Appendix 2). Additionally, the percentages in peak alignments shifted more in favor of 3'UTRs after requiring a minimum of 10 reads per peak. Peak splitting increased the number of peaks in each alignment locus but did not alter the alignment percentages (data not shown). Thus, peak splitting was set at default for PeakSplitter (see methods). Additionally, none of the filters significantly increased signal from the RNA-seq and Ribo-seq (Appendix 4, Chapter 3), thus, the RNA-seq and Ribo-seq were excluded as a criterion for filtering. Subsequently, filters were set to be the lowest stringency where the percent of 3'UTR mapped peaks was stable and did not dramatically change upon increased stringency. Thus, a set of confident peaks was identified by requiring 4 out of 5 libraries to have at least one read per peak and the peak to contain a minimum of 10 total reads.

Peaks located in 3'UTRs stood out as the dominant region before and after peak filtering (65.93%) whereas peaks covering intronic and intergenic regions were reduced to 8.86% and 1.88% respectively (Figure 8A), suggesting that intronic and intergenic binding events may be the result of spurious binding or alignment artifacts. Overall, 75.66% of *Msi2*-HITS-CLIP peaks were located in mRNAs, with 87.14% of those peaks

located in the 3'UTR. To visualize this observation, metagene plots demonstrated that the read and peak distribution across mRNA transcripts were strongly enriched in the 3'UTR region (Figure 8B). Within the 3'UTR, neither the reads distribution nor peak distribution showed a strong preference toward either the stop codon or the 3' terminus of the transcript (Figure 8B). These data indicate that *Msi2* primarily binds the 3' UTR of target mRNAs, consistent with what is expected of a post-translational RNA regulator. Thus, *Msi2* targets were defined by having a 3'UTR HITS-CLIP peak leading to 1199 identified targets.

Figure 8: *Msi2* predominantly binds 3'UTRs



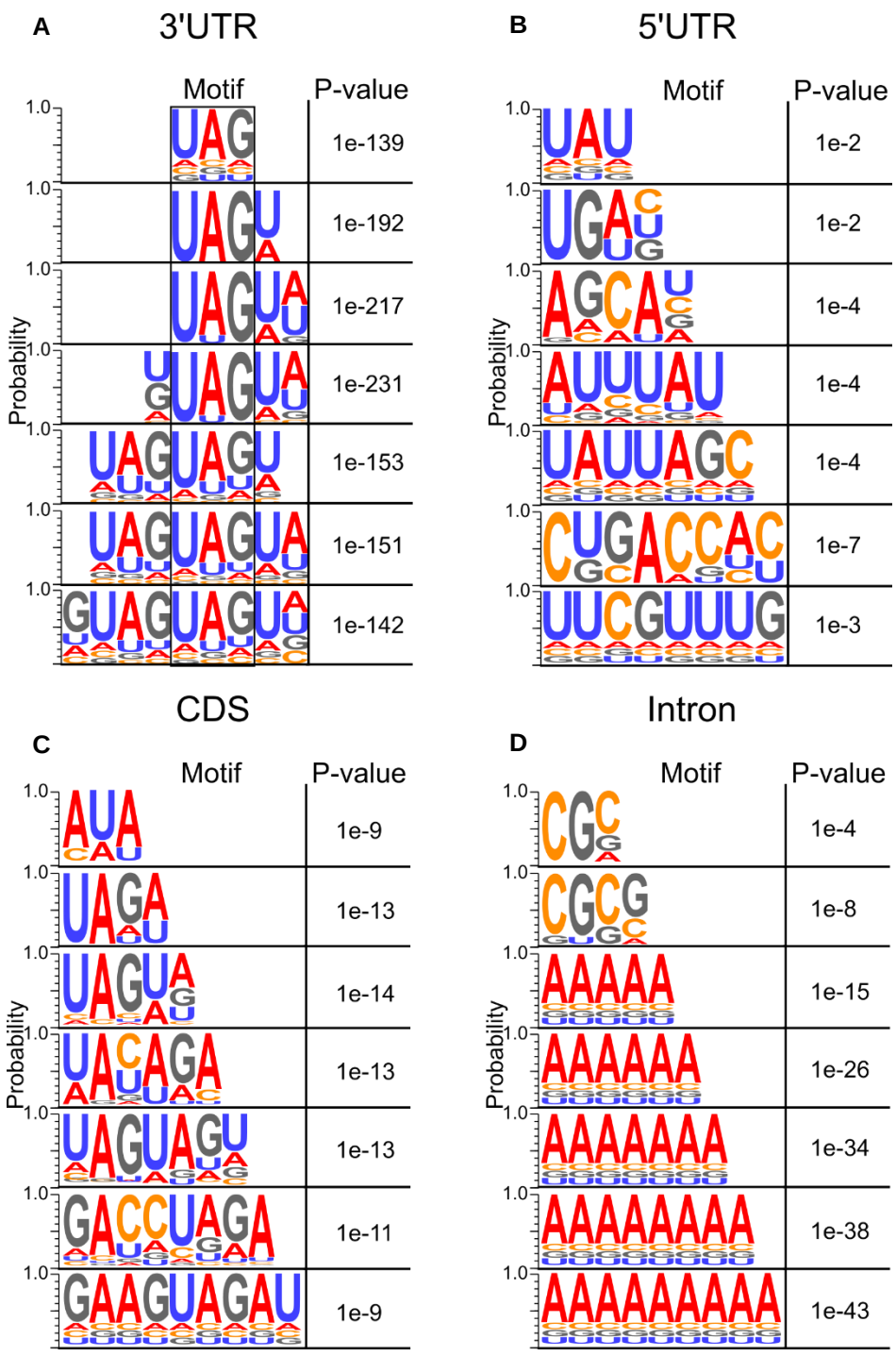
(A) Pie chart of the genomic locations of the aligned reads and filtered peaks before (left panel) and after (right panel) the filtering processes. (B) Metagene of exonic coverage along a scaled mRNA for aligned reads (left panel) and filtered peaks (right panel). Reads densities are normalized for library sizes. Peak and reads densities are averaged along all detectable transcripts based on RNA-seq (See methods).

2.3.5 Msi2 Motif Identification

Having identified a set of regions recognized by *Msi2*, I performed de novo motif searching to detect sequence-specific signatures of *Msi2* binding (Figure 9). The peak summit is presumed to be the direct binding site for *Msi2*. Thus, the search was restricted to 3-9mer motifs within 3'UTR, 5'UTR, intronic, and CDS peak summits. The summit of each peak was determined by the center of the region with the greatest read coverage and extended 25 nucleotides on either side. A group of motifs containing trinucleotide UAG sequences were enriched around the 3'UTR peak summits (Figure 9A). Interestingly, the larger motifs contain UAG repeats, indicating *Msi2* may require multiple UAGs for binding (see Chapter 2.3.6). Only peaks located in 3'UTRs generated a confidently enriched and coherent set of motifs, supporting the notion that *Msi2* primarily binds the 3'UTR of mRNA transcripts (Figure 9B-D).

The motifs that come out of the 5' UTR peaks are generally random and do not contain a similar motif across all of the motif sizes searched for (Figure 9B). The intronic motifs show a similar result to that of the 5' UTR motifs (Figure 9D). Additionally, as the size of the motif increases the more degenerate the motif becomes. These data imply that 5' UTR and intron signal may be a result of noise in the system. However due to the small number of peaks, the authenticity of these sites cannot be ruled out. CDS motifs show more consistency with the 3'UTR and to each other (Figure 9C). While they are not as consistent within themselves as the 3'UTR peak motifs, they do contain UAG motifs that are predicted to be functional *Msi2* binding sites. Due to the prevalence of a consistent motif and the large proportion of peaks aligned to the 3' UTRs, the targets will be defined as mRNA that contain a 3'UTR peak.

Figure 9: *Msi2* HITS-CLIP 3'UTR peaks are enriched for UAG core motif



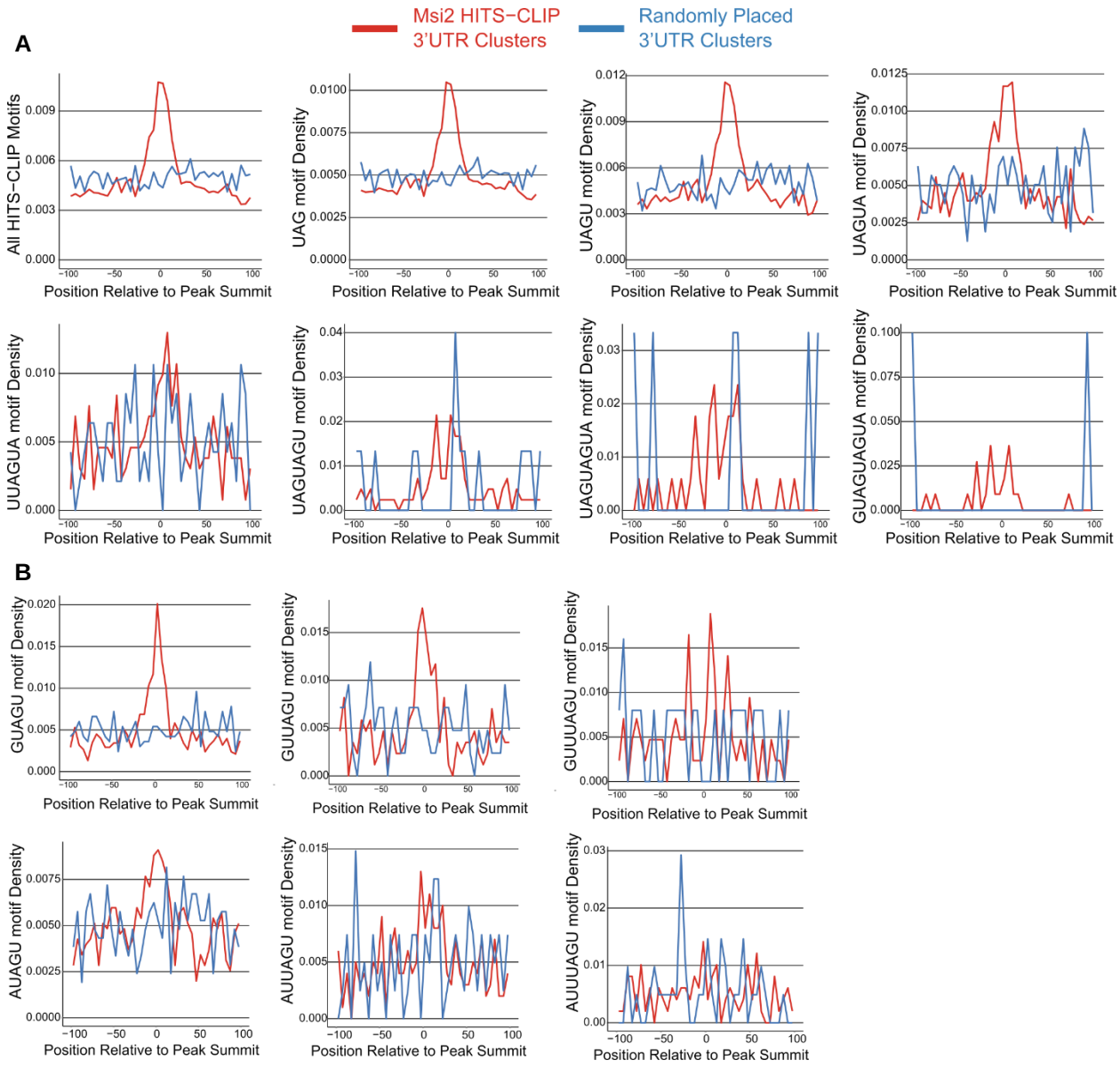
(A-D) De novo motif search for 3-9 mers from +/- 20 nucleotides surrounding *Msi2* HITS-CLIP peak summits in (A) 3'UTRs (B) 5'UTRs (C) Coding sequences and (D) Introns. The top motif identified for each N-mer search in 3'UTRs is displayed and positioned to highlight the shared UAG motif. P-values displayed are obtained from Homer software

2.3.6 Motif Enrichment at Msi2 Peaks

Since UAG can be found commonly by chance throughout the 3' UTR, I wanted to know if UAG motifs were enriched around the peak summit over a randomized set of peaks derived from 3' UTRs. The relative position of the individually identified motif or all identified 3-9mer motifs in aggregate showed enrichment around the summit of the HITS-CLIP peaks (Figure 10A). Each of the identified motifs up to 5-mer also showed an enrichment around the peak summit in contrast to the randomized peaks showing no enrichment. 6-9-mer motifs demonstrated modest localization to the peak summit over background, likely due to difficulty in finding exact motif matches which were required for this analysis.

Msi1 and *Msi2* have similar RNA-binding domains and are suspected to bind the same targets. Additionally, both have been shown to require UAG to bind their targets *in vitro*. It can be hypothesized that *Msi1* and *Msi2* may be able to bind the same motif, which may explain some redundancy observed between these two proteins. Thus, the positional distribution was assessed for the *Msi1* consensus motif, (G/A)U(n)AGU (n=1–3), identified by SELEX (Figure 10B) (Imai et al., 2001). GUAGU and GUUAGU were found to be enriched around the peak summits similar to the motifs identified for *Msi2*. GUUUAGU showed some level of enrichment but was greatly reduced as compared to the other uridine length variations, likely a result of requiring exact motif matches, as observed with the longer *Msi2* motifs. Strikingly, when the 5'-most G was replaced with an A there was no strong enrichment near the peak summit, implying that the A residue is not required for *Msi2* or potentially not commonly used. These data support the hypothesis that *Msi1* and *Msi2* motifs are interchangeable *in vivo*.

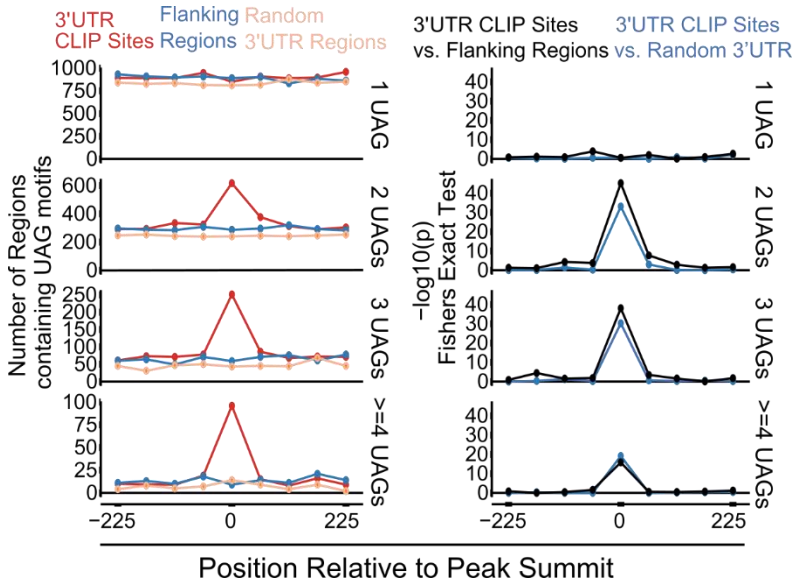
Figure 10: UAG containing motifs are enriched around 3'UTR CLIP peak summits



(A) Positional enrichment for the top 3-9mer motifs identified from HITS-CLIP was calculated in a window surrounding the 3'UTR HITS-CLIP peak summits. A background distribution was also calculated by shuffling the position of the searched window randomly within 3'UTR sequences. (B) Positional enrichment for the Msi1 motif predicted by SELEXC, G/AU$_{1\text{-}3}$AG. Otherwise as described in A.

Since the 7-9mer motifs identified consisted of repeated adjacent UAG sequences and because UAGs are enriched at the peak summit over background, it is likely that these regions are enriched for multiple UAGs. Indeed, previous studies demonstrated that the two RRMs of Msi proteins can independently bind UAG elements (Zearfoss et al., 2014). It is possible that the peak summits will contain multiple UAGs required for binding. A significant enrichment of two or more UAG motifs was observed surrounding the peak summit of the *Msi2* binding sites, in contrast to randomized 3'UTR regions or regions flanking the binding sites (Figure 11). Representative examples highlight the clustering of UAGs that can occur in these peak summits (Appendix 2). However, it is clear, based on the distribution of UAGs within these individual examples, that UAG alone is not sufficient for binding due to the large number of non-CLIP'ed UAGs and the lack of a UAG in some peaks, indicating that additional elements or features likely are required for *Msi2* recognition. Altogether, *Msi2*-HITS-CLIP reveals characteristics of *Msi2* binding and identifies a trinucleotide UAG core *Msi2* recognition motif.

Figure 11: Multiple UAGs are enriched in 3'UTR peak summits



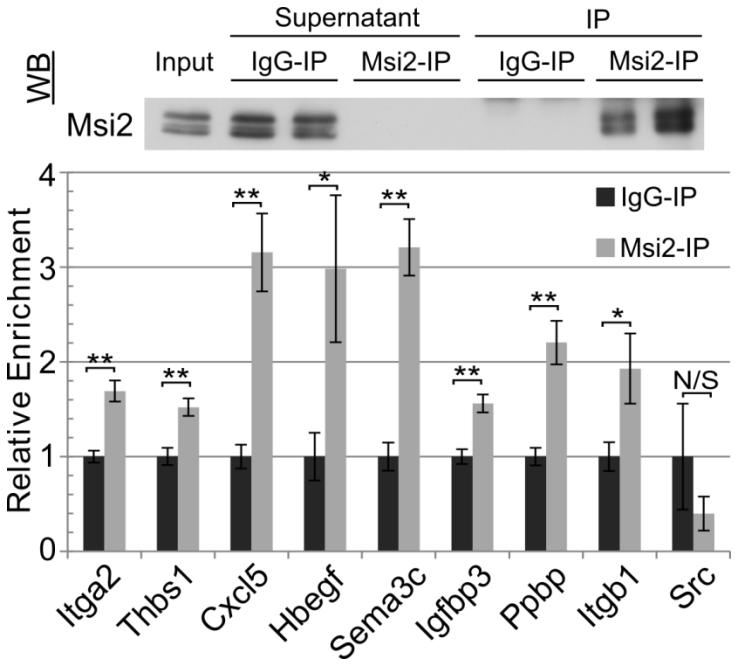
Number of UAG motif (1, 2, 3, ≥4) occurrences in a +/- 225-nucleotide window around the peak summit in 3'UTR HITS-CLIP sites, flanking region, or random 3'UTR background (left panel), with Fisher Exact Test showing motif enrichment in CLIP sites over the flanking region or 3'UTR background (right panel).

2.3.7 Msi2-RIP-qPCR Validation

RNA immunoprecipitation followed by qPCR (RIP-qPCR) was performed to confirm *Msi2* targeting identified from HITS-CLIP (Figure 12). *Msi2* was immunoprecipitated using a modified HITS-CLIP protocol using mild lysing and washing conditions and RNA was collected from the protein pellet. Several targets, *Itga2, Thbs1, Cxcl5, Sama3c, Igfbp3, Hbegf, Pbpb*, and *Itgb1*, and one non-target control, *Src*, were chosen for qPCR validation due to changes in RNA-seq and Ribo-seq data consistent with *Msi2* regulation (see Chapter 3). Western blotting confirmed depletion of *Msi2* in IP supernatant, with an enrichment for *Msi2* in the IP fraction, indicating enrichment for the targets should be observed over background. Indeed, the targets were enriched in the IP fraction over background (IgG control) using *Gapdh* and *Hprt* as

normalizing non-targeting mRNA controls (Figure 12). These data provide further validation that the targets detected in HITS-CLIP are *bona fide* targets of *Msi2*.

Figure 12: Msi2 HITS-CLIP binding is confirmed using RIP-qPCR



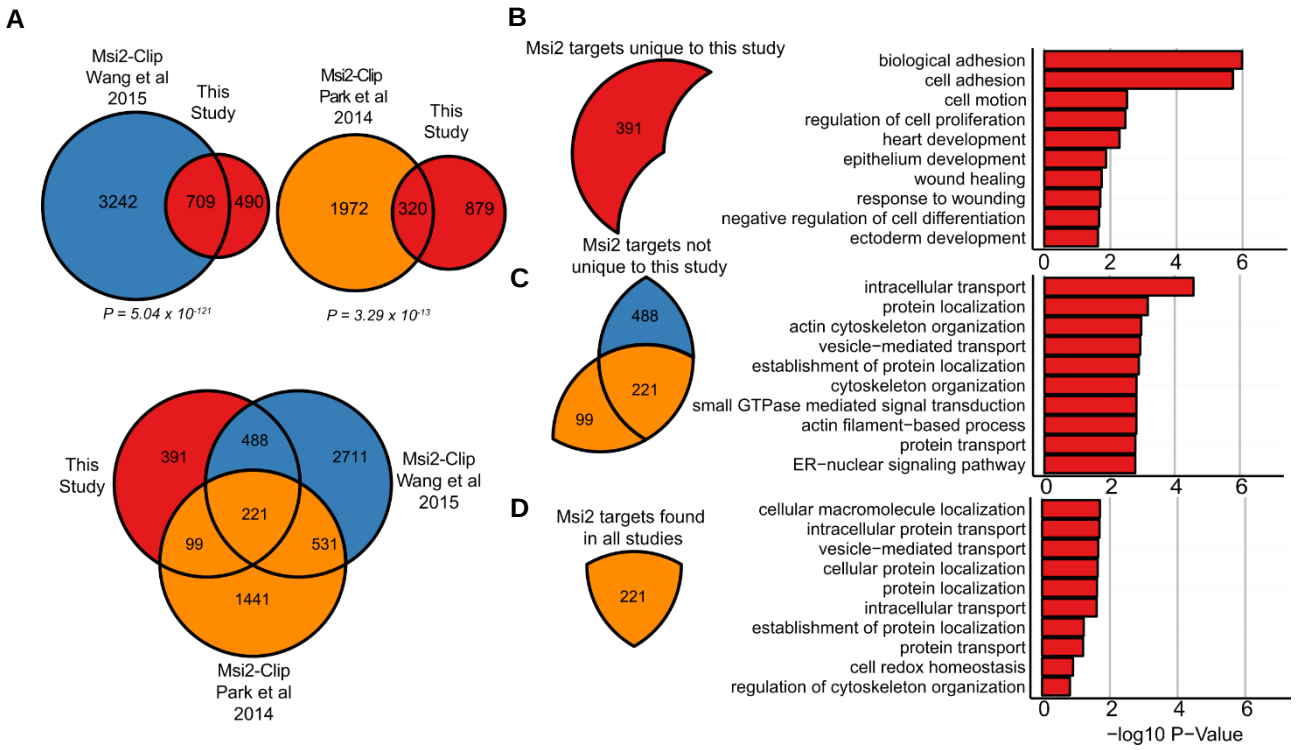
RNA-Immunoprecipitation of *Msi2* complexes with *Msi2* or control rabbit IgG antibody, followed by qPCR detection of the selected *Msi2* targets using Gapdh and Hprt as non-targeting controls. Relative Enrichment was calculated by enrichment over IgG control. Western blot validation of successful *Msi2* immunoprecipitation is shown. (** $p<0.01$; *$p<0.05$)

2.3.8 Comparison of Msi2-CLIP Datasets

Two other studies have reported *Msi2*-HITS-CLIP to varying degrees of success in hematopoietic and gastrointestinal systems (Park et al., 2014, 2015). In collaboration with Kent Riemondy, I wanted to compare the data obtained in this study to the data from those studies, in order to elucidate common or unique targets (Figure 13). The most noticeable difference between the datasets was the large number of genes identified in the other studies (Figure 13A). This is likely caused by differing filtering

methods used in those studies as compared to my own. Nevertheless, there is a significant degree of overlap between the datasets. Using GO term analysis on the different sets of genes, I found that genes unique to this study were enriched for terms associated with migration and adhesion (Figure 13B). In contrast, the gene sets that were not unique to this study or were common between all studies, showed enrichment for terms associated with common cellular processes such as intracellular transport and protein localization.

Figure 13: Keratinocyte specific Msi2 targets are enriched for migration associated terms



(A) Venn-diagrams comparing *Msi2* targets identified in *Msi2*-HITS CLIP from keratinocytes (this study), intestinal stem cells (GSE64388), and a leukemia cell line (GSE62115) (Done by Kent Riemondy). P-values were derived from the hypergeometric test (B) GO analysis of genes only found in keratinocytes. (C) GO analysis of genes shared between the intestine and the leukemia studies. (D) GO analysis of the common genes found in all three studies. The top ten enriched biological process pathways from each comparison are displayed with the annotated –log10 p-value.

CHAPTER 3

ANALYSIS OF *MSI2*-MEDIATED REGULATION OF TARGETS

3.1 Introduction

Many post-transcriptional regulators function by altering RNA stability or mRNA translation. Classic methods of measuring RNA abundance included radiolabeling or northern blotting. However, these techniques are generally limited to a small number of transcripts. The development of qPCR and microarrays allowed for larger numbers of transcript to be measured and quantified. These techniques work through hybridization to known sequences, thus, they can only work by measuring known RNA transcripts. RNA sequencing (RNA-seq) was developed in the early part of the 21st century and allows for transcriptome-wide RNA abundance measurements (Wang et al., 2009b). However, RNA-seq only measures steady state RNA levels and cannot directly determine translational changes or stability changes. For these measurements different techniques are required such as ribosome profiling or stability assays.

3.1.1 RNA Sequencing

RNA sequencing was developed to sample the entire transcriptome in cells (Wang et al., 2009b). Due to continued improvements to the technique over the years scientists can now quantitatively measure RNA abundance of many transcripts at once. Additionally, the computational analysis of the data has improved at an impressive rate, allowing many questions to be addressed that were previously time consuming and difficult. RNA-seq has been used in many instances to observe the regulatory effects of an RNA-binding protein on its targeted transcripts, such as the case for Argonaute and

miRNA-mediated regulation (Bracken et al., 2014). One technical limitation of RNA-seq is that the vast abundance of ribosomal RNA and transfer RNA can easily overwhelm the reads from other RNA species in a library (Peano et al., 2013). To avoid this, the libraries can be depleted of rRNA and tRNA, allowing for total transcriptome sequencing, or the RNA can be oligo dT-purified to remove and sequence only the polyadenylated RNA (i.e., mRNA). Each method has its own value. Poly-A selection primarily enriches for mRNAs, while ribosome depletion detects all RNAs. Regardless of the depletion method, RNA-seq cannot distinguish between transcript stability and transcription rates. It is the measurement of RNA steady-state levels. Since RNA-binding proteins can function in many different ways it is important to determine a possible regulatory mechanism for the protein of interest. In order to distinguish between all the possibilities, additional assays must be performed, such as adding polymerase inhibitors and watching transcript loss over time to measure RNA stability.

3.1.2 RNA Stability Measurements

Measuring RNA stability is classically done by removing transcription from the steady state equation using inhibitors of RNA polymerase II such as Actinomycin D or α-amanitin (Chen et al., 2008; Lindell et al., 1970). Transcripts of interest can be measured over time using qPCR for a small number of genes, microarrays for large groups of genes, or RNA-seq for a global view of all RNA changes. RNAs decay exponentially with time. To calculate a half-life, or the time needed for half the transcripts to be degraded, one needs to log transform the data and find the slope. The

slope is the decay constant that can be used to find the half-life as described previously (Chen et al., 2008). This way one can determine changes in RNA half-life and stability.

3.1.3 Determining mRNA Translation

Assessing the occupancy of ribosomes across a transcript has traditionally been performed by fractionating polysomes on sucrose gradients and detecting mRNA target(s) of interest through a variety of means. However, this approach can be time consuming and requires tests on multiple fractions. While this is conducive for looking at a few targets, it becomes increasingly difficult for probing transcriptome-wide changes in ribosome occupancy. Thus, ribosome sequencing (Ribo-seq) was developed as a way to measure ribosome occupancy across all transcripts (Ingolia, 2014; Ingolia et al., 2009). It has its roots from a proof of principle study done in 1969 (STEITZ, 1969). Steitz showed that ribosomes could be stalled and used to protect fragments of RNA from RNase degradation to produce ribosome protected fragments (RPFs). It wasn't until sequencing technology developed further that it became feasible to sequence these fragments and assess the ribosome occupancy transcriptome-wide. Because ribosome occupancy is closely tied to transcript levels, a metric called transcriptional efficiency (TE), or RPF per RNA transcript, which couples RNA-seq and Ribo-seq, was developed (Ingolia et al., 2009, 2011). This metric allows one to assess the change in translation independently of RNA abundance changes. This methodology has been used to untangle many questions in RNA biology, including the order of events in Argonaute-mediated post-transcriptional gene regulation (Meister, 2013).

3.1.4 Chapter Summary

The mechanism *Msi2* uses to regulate its target is generally thought to be translational control but recent evidence suggests that it could regulate RNA abundance (Katz et al., 2014; Sakakibara et al., 2001). In this Chapter I examine how *Msi2* regulates its targets, identified in Chapter 2. To do this, I first generated *Msi2* depleted cell-lines. Keratinocyte lines were generated by infection with lentiviral *Msi2* knockdown constructs or scrambled (negative control) sequences. Resulting cell lines were harvested for Ribo-seq and RNA-seq to assess ribosome occupancy or RNA abundance, respectively. These data were merged with the HITS-CLIP data from Chapter 2 to generate a list of targeted genes and their changes in ribosome occupancy and/or RNA abundance upon loss of *Msi2*. From these results I concluded that *Msi2* primarily alters target mRNA abundance rather than ribosome occupancy. Since *Msi2* is an RNA-binding protein localized to the cytoplasm and has no known nuclear role, the most likely explanation for this change is stability. I assessed the stability of a set of targets using Actinomycin D treatment followed by qPCR detection. Stability for these targets was increased in cell lines lacking *Msi2*, suggesting that modulation of RNA decay is the predominant mode of *Msi2* regulation. I further analyzed the sequencing data sets to identify processes under *Msi2* control. Proliferation and survival were among the most prominent processes. Surprisingly, migration was also represented in the data. Indeed, many of the affected genes are known to regulate migration and proliferation.

3.2 Materials and Methods

3.2.1 Western Blotting and qPCR

Western blotting was performed with 20-30 μg of protein lysate run on a 10% SDS-PAGE gel. Proteins were transferred to 0.2 μm PVDF and immune-probed for Musashi-2 and β-tubulin (as a loading control) (Appendix 10). Blots were incubated overnight with primary antibodies and detected using anti-rabbit HRP-conjugated secondary antibodies and Amersham ECL-Plus reagents (GE Healthcare Life Sciences). X-ray films were used to detect the signal, scanned and processed with ImageJ software. See Appendix 10 for antibody information. qPCR was performed on TRizol-extracted RNA using a SuperScript III RT kit (Life Technologies) and a BioRad CFX-384 machine. Relative expression was computed using ΔΔCq method normalized to *Hprt* and *Gapdh* values, with error bars denoting standard error of the mean. See Appendix 9 for qPCR primer information.

3.2.2 Ribosome Profiling and RNA-seq Library Preparation

Ribosome profiling was performed on scrambled shRNA and *Msi2* knockdown keratinocytes using the ART-Seq Ribosome profiling kit (Illumina). Briefly, cultured cells were grown to ~80% confluency on a 15 cm plate and treated with 50 μg/mL cyclohexamide for one minute before they were lysed, aliquoted, and digested with RNase. Ribosomes and associated RNA were isolated using illustra™ MicroSpin™ S-400 HR Columns (Illumina). RNA was extracted from the isolate and rRNA was depleted once using a Ribo-Zero Gold™ kit (Illumina). RNA fragments 28-32 nts long were isolated via denaturing PAGE, ligated to a 3' adapter, reverse transcribed,

circularized using CircLigase, depleted for rRNA a second time, and PCR-amplified following the provided protocol. RNA-seq was performed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs) with minor modifications using lysate matched to the ribosome profiling samples. Briefly, mRNA was isolated from total RNA using Dynabeads mRNA DIRECT Micro Purification Kit (Thermo Fisher) and fragmented for 15 minutes at 94° C. First strand synthesis, second strand synthesis, end repair, adapter ligation, and PCR were performed as described in the provided protocol. All PCR products were sequenced on an Illumina HiSeq 2000 using 1x100 sequencing.

3.2.3 Ribosome Profiling and RNA-seq Analysis

RNA-seq raw reads were first trimmed to remove 3' adaptor sequence (GATCGGAAGAGCACACGTCTGAACTCCAGTC, CutAdapt, default settings). Reads were then aligned to the mouse genome with Tophat with a supplied .gtf transcript annotation file (Illumina iGenomes, mm10) (settings = --bowtie1 --library-type fr-firststrand). Alignments uniquely overlapping coding sequences were counted with HTSeq Count (settings = -s reverse –t CDS). Differential expression was calculated with DESeq using default settings. Alignment statistics are shown in Appendix 5.

Ribosome profiling raw reads were first trimmed to remove 3' adaptor sequence (CutAdapt, default settings) and filtered to only retain sequences at least 20 nucleotides in length. Filtered reads were then aligned to remove contaminating ncRNAs by sequential Bowtie alignment (default settings) to an rRNA database (Illumina iGenomes, mm10), tRNA database (UCSC table browser, mm10), and ncRNA database

(Ensemble). Unaligned reads were then aligned to the mouse genome with Tophat with a supplied .gtf transcript annotation file (Illumina iGenomes, mm10) (settings = --bowtie1). Alignments uniquely overlapping coding sequences were counted with HTSeq Count (settings = -s yes –t CDS). Differential expression was next calculated with DESeq using default settings. Alignment statistics are shown in Appendix 6.

Transcripts with at least a BaseMean of 10 in the shScr libraries in both ribosome profiling and RNA-seq were used for downstream analysis. Translation efficiency was calculated as the ratio of the average reads per million for ribosome profiling and RNA-seq data. Cumulative fractions were computed using the ecdf() function in R. *Msi2* targets were defined as the subset of genes upregulated in both the RNA-seq and ribosome profiling datasets (FDR <0.05) and contain a high-confidence 3'UTR *Msi2* HITS-CLIP peak. Target genes were ranked based on degree of upregulation in the ribosome profiling data and the RNA-seq. The ranks for each dataset were summed, and targets genes were then ranked by the gene with the high summed rank, with ties in ranking decided by the gene with the largest number of total 3'UTR HITS-CLIP reads.

3.2.4 RNA-Stability Measurements

~300,000 shRNA-producing keratinocytes were plated into five 6cm dishes and allowed to grow in E-Low Media as previously described (Riemondy et al., 2015). Once cells reached ~80% confluency, medium was supplemented with 5 µg/mL Actinomycin D (Thermo Fisher). The zero time point was marked starting 5 minutes after Actinomycin D addition. Cellular RNA was harvested using TRizol (Thermo Fisher) at times 0, 2, 4, 6, and 8 hours and used in qPCR for targets (Appendix 9 for qPCR primer

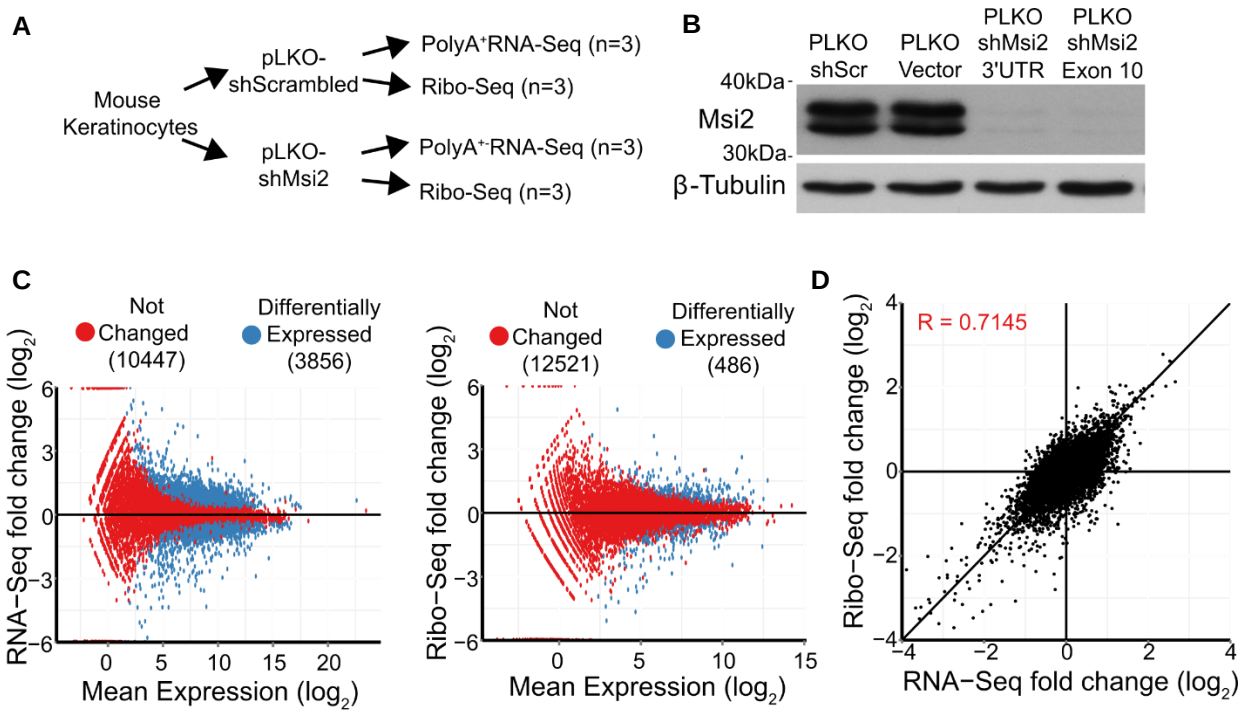
sequences). Relative expression was computed and normalized to the 0 hour time point for each target from four technical replicates using the ΔΔCq method normalized to *Hprt* and *Gapdh* values, with error bars denoting standard error of the mean. RNA half-lives were calculated from linear regression of log-transformed expression values for each target as described previously (Chen et al., 2008). ANCOVA analysis was performed on the resulting regression lines to assess statistical significance.

3.3 Results

3.3.1 RNA-seq and Ribo-seq Library

RNA-seq and Ribo-seq allow unparalleled insight into changes that occur upon loss of an RNA-binding protein. This information can be used to determine the transcript changes as well as the translational changes that occur transcriptome-wide. 1,199 mRNAs were identified with an *Msi2* HITS-CLIP peak located in the 3'UTR. To determine the mode by which *Msi2* regulates these targeted transcripts, knockdown cell lines were generated for RNA-seq and Ribo-seq (Figure 14A). *Msi2* was knocked down in keratinocyte cultures using one of two short hairpin RNAs (shRNA): 3'UTR-targeting or coding region-targeting. I conducted RNA-seq and Ribo-seq from the same biological samples (Figure 14A). Each shRNA efficiently reduced *Msi2* protein by an estimated 90% or more (Figure 14B). The 3'UTR-targeting shRNA reduction was more consistent across multiple derivations of the knockdown cell lines. Additionally, the 3'UTR-targeting shRNA allowed for easier add-back experiments without the need to mutate the CDS to relieve shRNA repression. Thus, all subsequent experiments utilized the 3'UTR-targeting shRNA (sh*Msi2*).

Figure 14: Changes in RNA-seq and Ribo-seq libraries correlate



(A) Flow-chart of experimental design. (B) Western blot against *Msi2* on keratinocytes infected with indicated shRNA lentiviral construct. (C) Log2 fold change plotted against Log2 mean expression for RNA-seq and Ribo-seq libraries. (D) Comparison of log2 fold-changes for RNA-Seq and Ribo-Seq. Pearson correlation coefficient displayed. Significantly changed genes with FDR < 0.05 plotted in blue. The number of genes in each category is indicated in parenthesis.

As mentioned previously, RNA-seq and Ribo-seq were performed on matched samples using poly-A selection in NEBNext Ultra Directional RNA Library Prep Kit for Illumina and RPF isolation from ART-Seq Ribosome profiling kit respectively. Three replicate pairs were sequenced (Figure 14A). Each library had ~20-50 million reads in RNA-seq, with roughly 90% being uniquely aligned to the genome and ~50% alignment to CDS (Appendix 5). In contrast, Ribo-seq obtained ~15-30 million reads per library with only ~5% aligning to CDS (Appendix 6). The majority of these reads were lost to tRNA alignments, indicating that there was insufficient nuclease digestion of the

samples. If this is true, the expectation would be that the 5' and 3' end of the reads aligning to tRNAs will fall to either side of the anticodon loop to produce a ~30 nucleotide fragment that would be isolated with the RPF. Indeed, this is what was observed (Appendix 7). In ribosome profiling, the alignment of the 5' end of the RPFs should move in multiples of three, or in phase with the movement of a ribosome, while reads from RNA-seq should not. This is a byproduct of the way ribosomes move along mRNAs while translating, and serves as a way to validate Ribo-seq. Indeed, the reads from my Ribo-seq show phasing as expected, demonstrating that I isolated RPFs (Appendix 7). Interestingly, the observed phasing is not as prominent as would be expected from appropriately RNase treated samples, providing further evidence for insufficient nuclease digestion. Unfortunately, the RNA-seq library could not be analyzed in a similar fashion due to the nature of the library construction.

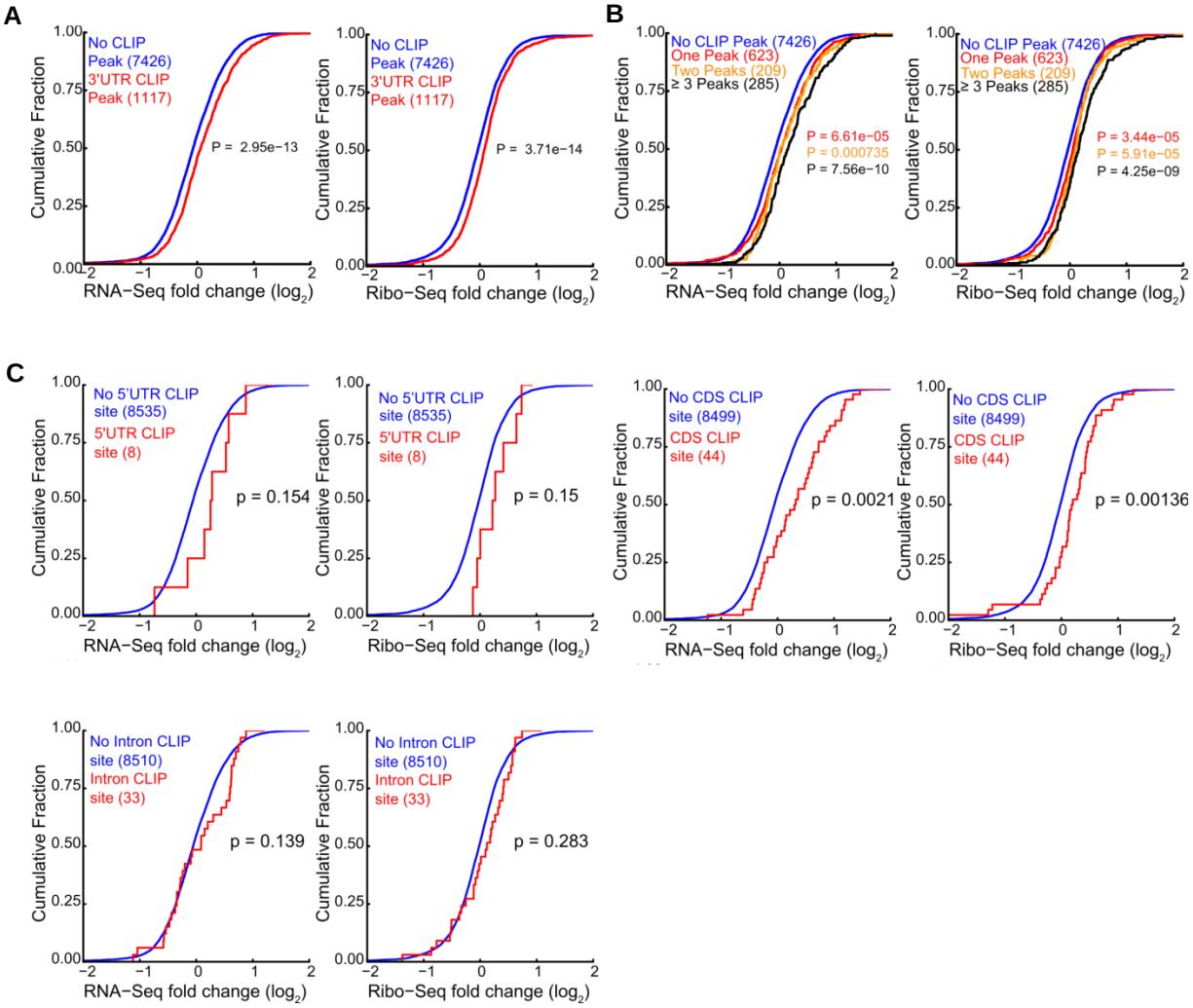
Despite the low alignment of the Ribo-seq libraries, I decided to proceed with differential gene expression analysis using DEseq. RNA-seq detected 3856 dysregulated genes in *Msi2* knockdown cells, whereas Ribo-seq detected 486 differentially expressed genes (FDR <0.05) (Figure 14C). This difference is likely due to the low reads coverage of the Ribo-seq data. Nevertheless, the $\log_2$-fold change between the datasets was highly correlated, as expected (Figure 14D). Transcripts with at least 10 reads aligned to the CDS per million in ribosome profiling and RNA-seq shScr libraries were used for downstream analysis. Using only CDS-aligned reads allows for more accurate comparison between the sequencing types.

3.3.2 Detected Msi2 Target Changes

Of the 1,199 target mRNAs identified through HITS-CLIP, 1,117 were detectable by both RNA-seq and Ribo-seq after filtering out genes with low coverage (minimum baseMean of 10 reads per million in shScr libraries). The next step was to determine how loss of *Msi2* affected these targets as a whole. In order to observe this effect, a cumulative distribution function (CDF) was used to compare the RNA-seq and Ribo-seq fold change of the 1,117 3'UTR-targeted mRNAs versus the non-targeted mRNAs (Figure 15). Cumulative distribution plots are interpreted by evaluating the direction and magnitude of any shift between targets and background, which should center close to 0 on the x-axis (log2-fold change) and at 0.50 on the y-axis (cumulative fraction). If the clip targets fold change shifts to the left of the background it means a reduction in RNA abundance or ribosome occupancy. Conversely, a shift to the right indicates an increase.

This method demonstrates that *Msi2*-bound mRNAs were more likely to be upregulated, or have a higher fold change as a whole (shifts to the right) upon reduction of *Msi2* levels (Figure 15A). The same is true for ribosome occupancy. One would expect that if these targets are regulated by *Msi2*, then the more *Msi2* binding in the 3'UTR, the greater the regulatory effect. To test this, CLIP targets were further subdivided as having one, two, or three or more *Msi2* binding sites. For both RNA-seq and Ribo-seq, the more *Msi2*-binding sites present in a target 3'UTR, the greater the extent to which the targets are dysregulated (Figure 15B).

Figure 15: Msi2 3'UTR CLIP targets change in RNA abundance and ribosome occupancy



(A) Cumulative distributions of changes in RNA-Seq (left panel) and ribosome Ribo-Seq (right panel) after *Msi2* knockdown. Genes with 3'UTR *Msi2*-HITS-CLIP peaks are plotted in red, genes without 3'UTR *Msi2*-HITS-CLIP peaks are plotted in blue. The number of genes in each category is indicated in parenthesis. (B) Cumulative distributions as in A, with genes with no (blue), one (red), two (yellow), or three or more (black) 3'UTR *Msi2*-HITS-CLIP peaks plotted separately. (C) Cumulative distributions as in A for those genes bound by *Msi2* (red) versus those not bound (blue) in the 5'UTR, coding sequence, and intron. For all panels, if the gene was bound by *Msi2* in multiple regions (5'UTR, CDS, 3'UTR, or intronic) these genes were classified as not bound to avoid conflating *Msi2* regulatory effects due to binding outside of the regions examined. Statistics are KS-Test one sided

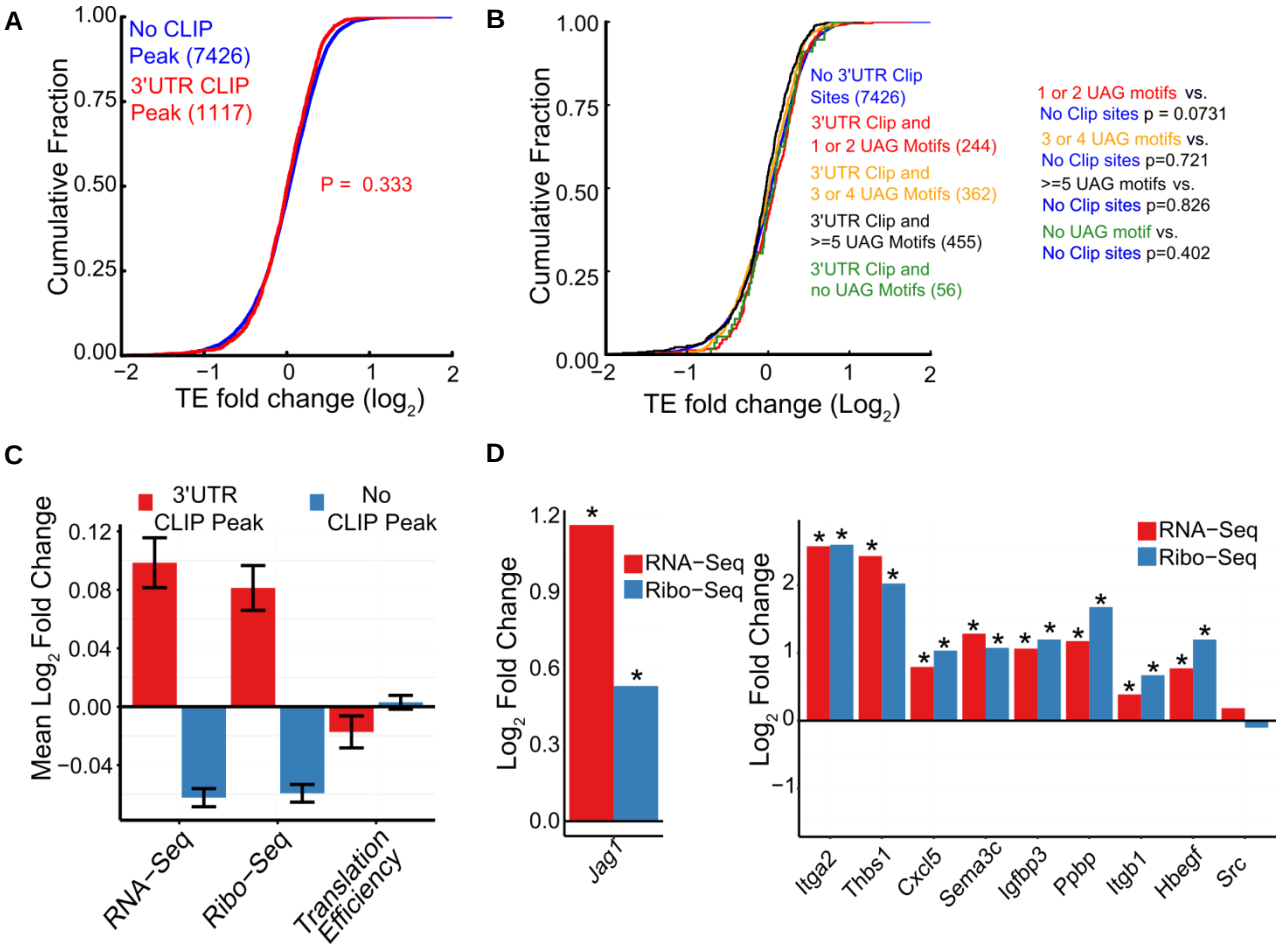
As mentioned in Chapter 2, *Msi2* binding was also detected in 5'UTR, coding, and intronic regions, albeit at a reduced levels. The number of genes containing 5'UTR, CDS, or intron binding events is very low after filtering (8, 44, and 33 genes respectively), compared to the 3'UTR binding events. CDF plots were generated using genes targeted by *Msi2* in these regions in order to test the regulatory effect of these binding events (Figure 15C). Targets with *Msi2* CDS binding events were more likely to be upregulated when *Msi2* was depleted, similarly to 3'UTR binding events. In contrast, 5'UTR and intron binding events did not demonstrate any discernable trend. However, due to low gene numbers and potential bias calculating significance in small datasets, these trends are not extremely reliable.

3.3.3 Translational Efficiency of Msi2 Targets

Previous studies report that *Msi2* primarily regulates translational efficiency (TE) (Gunter and McLaughlin, 2011; MacNicol et al., 2008; Sakakibara et al., 2001). My results from RNA-seq, however, indicate that *Msi2* may be regulating RNA abundance rather than translation efficiency. To test this, the fold changes of Ribo-seq and RNA-seq were compared in order to examine the correlation between the data, and to identify a population of transcripts that may be regulated primarily at the translational level (Figure 15D). As expected, the change in ribosome occupancy and mRNA abundances for *Msi2* targets was highly correlated, indicating that a change in RNA abundance results in a subsequent change in ribosome occupancy. There did not appear to be a dominant population that was primarily regulated at the translational level, indicated by

a population of genes shifted vertically on the graph, suggesting *Msi2* regulates target mRNA decay.

Figure 16: Msi2 target translation efficiency is not changed



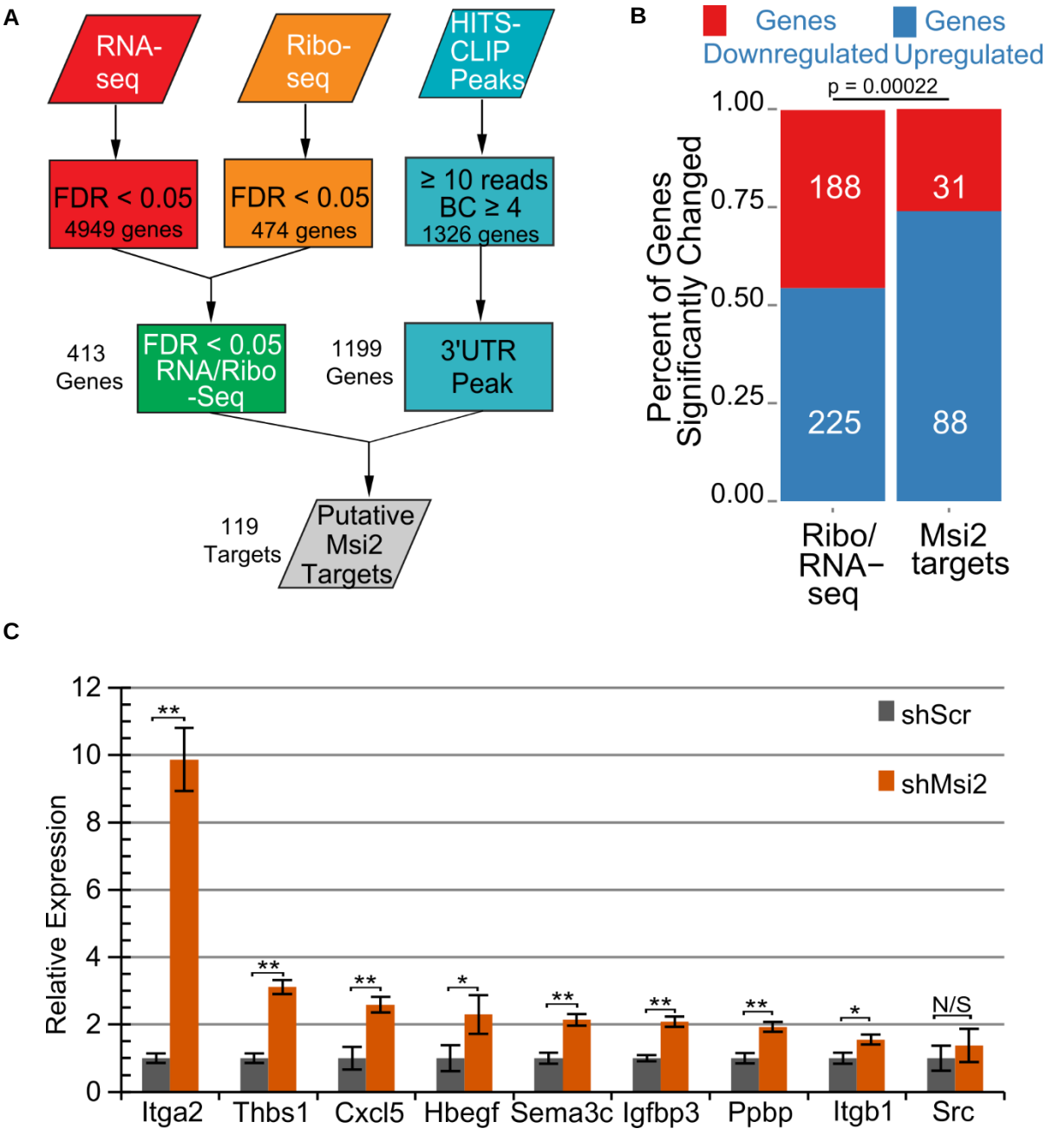
(A) Cumulative distributions of changes in translation efficiency for genes containing or not containing 3'UTR *Msi2*-HITS-CLIP peaks. (B) Cumulative distributions of Log2 Fold Changes in Translation Efficiency for genes that contain *Msi2* 3'UTR HITS-CLIP peaks, with 1-2 UAG motifs (Red), 3-4 UAG motifs (Orange), 5 or more UAG motifs (Black), no UAG motif (Green), or do not contain *Msi2* 3'UTR HITS-CLIP peaks. P-values were calculated by one-way KS tests. If a gene contained multiple peaks with differing numbers of UAGs, the gene was assigned to the highest UAG category that is found in the peaks. (C) Mean log2 fold-changes for RNA-Seq, Ribo-Seq and Translation Efficiency for genes containing or not containing 3'UTR *Msi2*-HITS-CLIP peaks. Standard error or the mean is displayed. (D) Log2 Ribo and RNA-seq Fold Changes for Jag1 and novel *Msi2* targets detected in this study and a negative control gene Src. Asterisk indicates FDR <0.05.

To directly examine the impact that loss of *Msi2* has on translation, translational efficiency (TE), the ratio of normalized ribosome protected fragments to normalized RNA-seq reads per transcript, was calculated. A CDF plot was used to assess any changes, as previously, and demonstrated no significant change in TE for *Msi2* targets over background (Figure 16A). Multiplicity of UAGs in the 3'UTR peaks, also, did not show any change in TE (Figure 16B). A bar chart comparing the mean log2-fold change between RNA-seq, Ribo-seq, and TE calculations highlight this observation (Figure 16C). This result was also evident when analyzing the previously identified *Msi2* target, Jag1, whose mRNA and ribosome occupancy were both significantly increased upon *Msi2* knockdown, and when analyzing a select set of targets identified in this study (Figure 16D).

3.3.4 High Confidence Targets

Msi2-HITS-CLIP, RNA-seq, and Ribo-seq measurements together provide a platform to identify high confidence targets regulated by *Msi2*. A list of regulated *Msi2* targets with *Msi2*-binding sites was generated as follows. Targets were filtered for FDR ≤ 0.05 in both Ribo-seq and RNA-seq as calculated by DEseq and cross-referenced with the high-confidence 3'UTR HITS-CLIP peaks (Figure 17A). Genes in this list were largely upregulated in the absence of *Msi2* (88 out of 119, 74%), consistent with *Msi2* being a negative regulator of gene expression (Figure 17B). These 88 genes were deemed high confidence *Msi2* targets (Appendix 8).

Figure 17: RNA-seq, ribo-seq, and HITS-CLIP allow identification of high confidence Msi2 targets



(A) Schematic depicting analysis method to extract high-confidence *Msi2* targets. Number of genes in each category are noted. (B) Proportion of genes upregulated in combined RNA-seq/Ribo-seq background data, or the identified high-confidence putative *Msi2* targets. P value was assessed with Chi-Squared test. (C) qPCR detection of the expression of selected *Msi2* targets in independently derived scrambled control and *Msi2* knockdown keratinocytes. (n= 4 biological replicates, ** p<0.01; *p<0.05).

To independently validate the observations from RNA-seq, a subset of identified targets was selected for confirmation of steady-state transcript levels by qPCR. Eight genes (*Itag2, Itgb1, Thbs1, Hbegf, Cxcl5, Ppbp, Sema3c, and Igfbp3)* were selected from the list of high confidence targets that exhibited mRNA abundance increases, increased ribosome occupancy, and 3'UTR *Msi2*-CLIP sites (Figure 12, 16D). The steady-state mRNA levels of all eight selected *Msi2* targets and a non-*Msi2*-target control, *Src*, were assessed using qPCR in independently generated *Msi2* knockdown cell lines and found to be upregulated to a similar extent as what was observed in RNA-seq (Figure 17C).

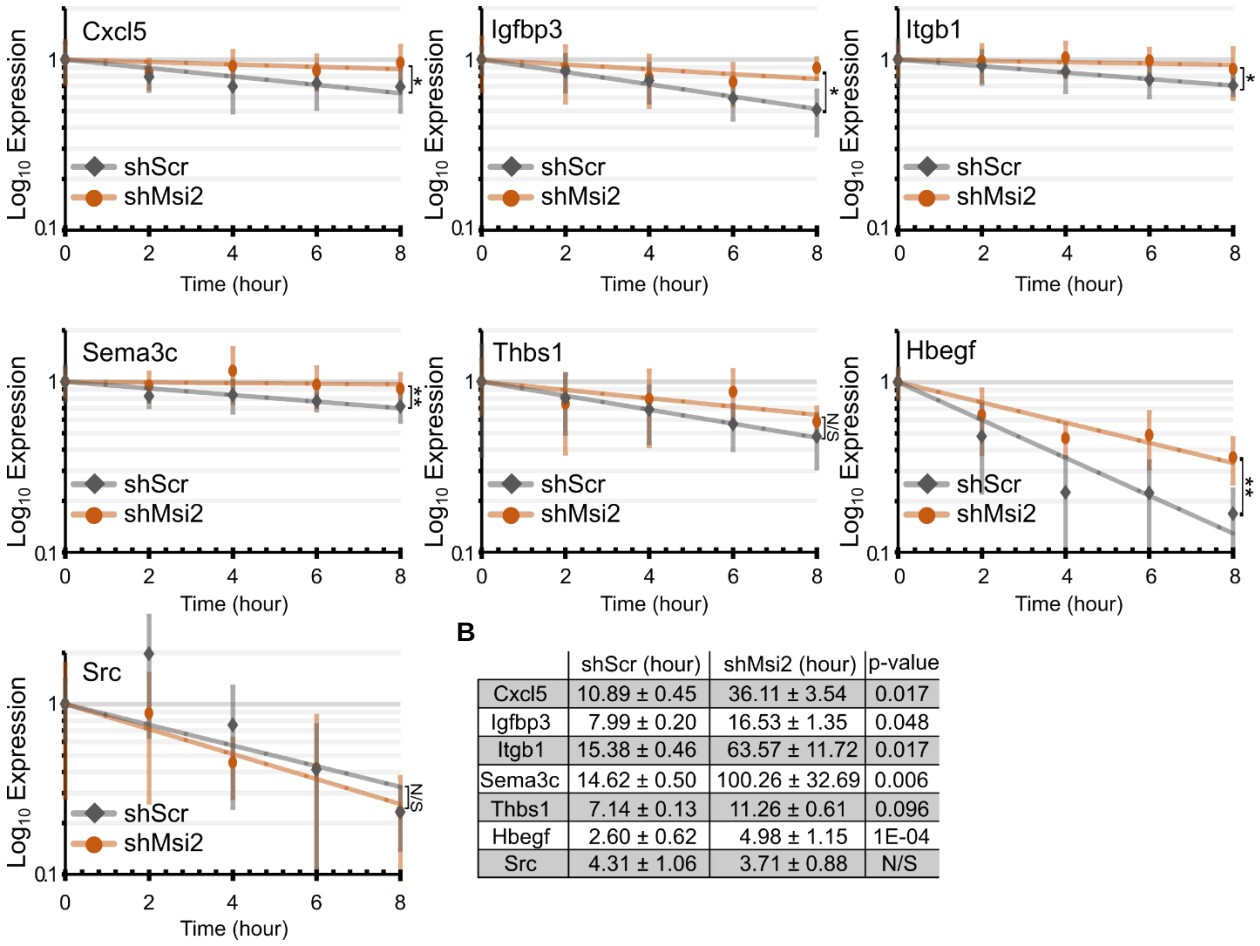
3.3.5 Msi2 Target Stability Changes

To determine whether *Msi2* affects the stability of target mRNAs, RNA decay curves for the eight targets validated previously and non-target *Src* control were assayed using an Actinomycin D time course in shScr and sh*Msi2* keratinocytes followed by qPCR. Subtle changes in two of the targets (*Itga2* and *Ppbp*) could not be accurately detected by qPCR and were discarded, leaving seven total. All target mRNAs except *Thbs1* were significantly more stable in sh*Msi2* cell lines, while *Src* transcript stability was not significantly changed (Figure 18A). While *Thbs1* did not achieve statistical significance, it also trended toward being more stable in *Msi2* knockdown samples. In support of this observation, RNA half-lives were calculated from the decay curves and found to be anywhere from 2- to 7-fold higher in *Msi2* knockdown cells over the scrambled control counterparts (Figure 18B). These

results demonstrate that *Msi2* likely functions to repress target genes at least in part by destabilizing the mRNA transcript.

Figure 18: Msi2 controls target stability

A



B

	shScr (hour)	shMsi2 (hour)	p-value
Cxcl5	10.89 ± 0.45	36.11 ± 3.54	0.017
Igfbp3	7.99 ± 0.20	16.53 ± 1.35	0.048
Itgb1	15.38 ± 0.46	63.57 ± 11.72	0.017
Sema3c	14.62 ± 0.50	100.26 ± 32.69	0.006
Thbs1	7.14 ± 0.13	11.26 ± 0.61	0.096
Hbegf	2.60 ± 0.62	4.98 ± 1.15	1E-04
Src	4.31 ± 1.06	3.71 ± 0.88	N/S

(A) RNA stability curves plotted using qPCR expression versus time of set of targets from previous qPCR validation and non-target control Src. ANCOVA analysis was used in determining statistical significance. Standard error mean is displayed for each time point. (** p<0.01; *p<0.05) (DB) RNA half-lives in hours calculated from the stability curves. ANCOVA p-values displayed. ** p< 0.01; *p<0.05) (ANCOVA). N.S. not significant.

Taken as a whole, these data support the hypothesis that *Msi2* represses target mRNA stability primarily through association with the 3'UTR and, to a lesser extent, the

CDS, similar to what has been reported for miRNAs. These data are consistent with *Msi2* acting as a repressive regulatory RNA-binding protein.

CHAPTER 4

CHARACTERIZATION OF *MSI2* PHENOTYPIC FUNCTIONS

4.1 Introduction

The list of high-confidence targets generated in Chapter 3 includes genes involved in proliferation and migration, with many having a known function in these processes in keratinocytes (Barrientos et al., 2008; Goshima et al., 2002; Huttenlocher and Horwitz, 2011; Kuo et al., 2011; Ridley et al., 2003; Yu et al., 2010). Proper control of these processes is critical for proper tissue maintenance and protection from disease (Arwert et al., 2012). Thus, cells employ a number of mechanisms to control these processes. *Msi2* is known to play a role in regulating cellular proliferation but has no previously known role in regulating migration (Park et al., 2015). However, *Msi1* does play a role in regulating migration in glioblastoma, leaving open the possibility that *Msi2* could do the same (Uren et al., 2015). Additionally, there is no reason to suspect *Msi2* couldn't regulate other processes in keratinocytes, which can be revealed through thorough analysis of the target genes. Therefore, I sought to determine the pathways *Msi2* regulates in keratinocytes using my datasets and cellular phenotypic assays.

4.1.1 Colony Formation and Growth Curve

Colony forming potential can be used to assess the ability of a cell type to form colonies. Cells at low density are plated and grown until colonies are large enough to be quantified after crystal violet staining. The intensity, size, and number of colonies can be measured and give insights into the processes being altered. More intensely stained colonies generally have higher cell density and may indicate smaller cells. The

size of the colony is an indication of growth speed, smaller colonies growing more slowly than larger ones. The number of colonies can give insight into defects in cell adhesion to the plate or to survive passaging. Colony formation is not a direct measure of proliferation, as survival and adhesion to the substrate also play a role in the ultimate size of a colony. However, it is a good approximation for the importance of a gene in one or more of these factors.

Growth curves can suffer from similar caveats as colony formation assays but provide the additional benefit of generating a doubling time by plotting numbers of cells versus time. Growth curves can be generated using one of several different readouts over time. In general, it involves counting the number of cells and plotting their increase over time.

4.1.2 Cell Cycle

The most direct measure of proliferation is cell cycle analysis using methods such as Fluorescence Activated Cell Sorting (FACS). Using this approach, the percent of cells in each phase of the cell cycle can be determined using a DNA marker to distinguish G1 (2N DNA) versus G2/M (4N DNA). S phase populations can be identified using incorporation of a labeled nucleotide into the DNA. Typically, this is done with 5-bromo-2'-deoxyuridine (BrdU) which mimics thymidine and can be detected with an antibody. However, newer chemistry technologies have been developed that allow quicker detection of the incorporated nucleotide. More proliferative cells will likely have a higher S/G2/M population and lower G1 population than a more slowly dividing cell population. However, this assay is best coupled with a readout of cell growth, as the

residency time in populations could be altered such that a more proliferative cell population has shorter S/G2/M phases, leading to a lower population in these phases.

Minor modifications to the way the cells are collected add the benefit of enabling the assessment of apoptosis in addition to growth. The presence of a sub-G1 population in this assay indicates the presence of apoptotic cells. Thus, depending on the questions being asked, cell cycle analysis can test two parameters influencing the growth of cells.

4.1.3 Cell Survival

In addition to the aforementioned (Chapter 4.1.2) test for apoptotic cells, cell death and survival can be tested by a number of different markers, such activated caspase-3. Another approach is to identify cells undergoing morphological changes associated with apoptosis, such as the classic blebbing phenotype. A fourth common apoptosis assay uses FACS for fluorophore-conjugated Annexin V, a marker of apoptotic cells, in combination with propidium iodide, a DNA stain that can only enter a cell with a disrupted plasma membrane, and thus should only mark apoptotic cells. The result is the separation and assessment of three populations: living, early apoptotic, and late apoptotic cells.

4.1.4 Cellular Migration

Cell migration can be measured using scratch assays, live cell imaging, or a variety of other assays. Scratch assays are a classic way of measuring migration of adherent cell cultures. There are a few variations but the general idea is that cells are grown to confluency and a scratch is made to remove cells from a region. This can be

done by physically scratching the cell monolayer or by removing a region that had been previously blocked. The cells are then monitored and their migration into the site is tracked. Generally, this technique is performed with a synchronous cell population or in the presence of low growth medium to remove confounding effects from proliferation. The distance the cells move per time point gives the relative velocity of the cells. This measurement can be compared across conditions. In some cases, cells cannot be grown in low growth medium, cannot be grown to confluency, or have different mobility dynamics as a sheet of cells. In these situations, another method for measuring cellular migration, live cell imaging, has recently emerged (Chapnick et al., 2013). In this method, cells are marked with a nuclear localized fluorescent reporter and tracked over time. Algorithms can then be used to measure the distance an individual cell travels per unit time to give a velocity. This method can be used to compare movement in varying conditions more reliably than a scratch assay and can more reliably disentangle migration from proliferation.

4.1.5 Chapter Summary

In this chapter I explore the phenotypic consequences of loss of *Msi2* in keratinocytes. As mentioned previously, *Msi2* targets genes involved in proliferation and migration. This observation was confirmed using Ingenuity Pathway Analysis, which also revealed an increase in survival-associated genes. Colony forming assays, growth curves, and cell cycle analysis were performed to assess proliferation in keratinocytes lacking *Msi2*, overexpressing *Msi2*, and containing an introduced rescuing *Msi2* gene. These data confirm that *Msi2* levels correlate with proliferation, previously

reported in the literature (Park et al., 2015). Additionally, cell cycle analysis demonstrated an increase in a sub-G1 population in *Msi2* knockout cells versus control cells. Therefore, I assessed apoptosis using Annexin V FACS with propidium iodide. I was able to confirm that loss of *Msi2* leads to an increase in apoptotic cells. Finally, I demonstrated that loss of *Msi2* increases migration in culture and that there is a loss of *Msi2* in the migrating cell population in wounded skin. In line with this observation, I found that the number of focal adhesions was increased in cells depleted for *Msi2*, suggesting a possible relationship between the number of focal adhesions and *Msi2*-controlled migration of keratinocytes.

4.2 Materials and Methods

4.2.1 IPA Analysis

Ingenuity Pathway Analysis was performed on datasets filtered for upregulated genes in RNA-seq, Ribo-seq, HITS-CLIP, and combined datasets with RNA/Ribo-seq and RNA/Ribo-seq HITS-CLIP targets as illustrated in Figure 4. Only experimentally observed relationships were considered. IPA build version 346717M was used for the analysis. Analysis references were set to Ingenuity Knowledge Base (Genes Only) with all cell types and mammals selected. Direct and indirect relationships were included. Data shown was taken directly from Summary of Analysis output in IPA.

4.2.2 Microscopy

Microscopy images for skin sections were taken using a Leica DM5500B microscope with a Hamamatsu C10600-10B camera and processed with the Leica image analysis suite, MetaMorph (MDS Analytical Technologies). The ImageJ/FIJI

software package was used to stitch together > 40 images using the Pairwise Stitching function to produce the skin wound images presented.

Microscopy for detecting focal adhesions was performed on a Nikon A1 laser scanning confocal microscope using a 100x/1.49 NA objective lens at the CU-Boulder Light Microscopy Core Facility. For each image, 9 optical slices of 0.125 microns were taken through the focal plane containing focal adhesions. Each Z-stack was converted into a 2D representation by taking the maximum intensity at each pixel. Presented micrographs were further processed to adjust brightness and contrast values. Identical microscope settings and image analysis parameters were used for all images presented. For quantifying focal adhesions, an imageJ/FIJI macro was constructed that counted focal adhesions in the following manner: To exclude partially imaged cells on the image boundary the following steps were performed. Maximum intensity projections of phalloidin staining were auto contrasted ("Enhance Contrast", "saturated=0.35"), then thresholded using the mean method (setAutoThreshold("Mean dark")). Thresholded images were converted to masks, and the binary operations Dilate, Close-, and Fill Holes were used to identify cellular regions. The Analyze Particles function was then applied to exclude regions that fell below a size threshold (size=1200-Infinity circularity=0.00-0.99). These masked regions were next overlaid on the Hoechst and vinculin images to exclude non-cellular regions (setPasteMode("AND")). Next, the number of cells within each image was calculated as follows. The DAPI maximum intensity projection was processed with CLAHE to enhance local contrast ("Enhance Local Contrast (CLAHE)", "blocksize=75 histogram=256 maximum=6 mask=*None* fast_less(accurate)") then auto contrasted ("Enhance Contrast", "saturated=0.35"), and

thresholded using the default method (setAutoThreshold("Default dark")). Nuclei counts were obtained by performing the Analyze Particles function ("size=25-Infinity circularity=0.00-1.00 show=Masks display summarize"). The nuclei counts were manually inspected after analysis and corrected when errors occurred. Lastly, the number of focal adhesions was calculated using methods from a previously published protocol for counting focal adhesions (Horzum et al., 2014). Briefly, the maximum intensity projections of vinculin staining underwent the following processing steps:

Background subtraction ("Subtract Background...", "rolling=50 sliding")

CLAHE ("Enhance Local Contrast (CLAHE)", "blocksize=19 histogram=256 maximum=6 mask=*None* fast_less(accurate)")

Exponentiation (“Exp")

Autocontrasting ("Enhance Contrast", "saturated=0.35")

Log3D filtering ("LoG 3D", "sigmax=5 sigmay=5")

Thresholding (setAutoThreshold("Default dark")

Focal Adhesion Counting ("Analyze Particles...", "size=2-2000 circularity=0.00-0.99 exclude")

The number of focal adhesions per image was then divided by the number of nuclei to obtain a focal adhesion count per cell per image count.

4.2.3 Flow Cytometry

Click-IT EdU cell proliferation assays were performed following the manufacturer’s instructions (Thermo Fisher). Keratinocytes were pulsed with 10 μM EdU for 1 hour prior to harvesting for EdU detection. DNA content was detected by

Hoechst Dye staining. For detecting apoptotic cells, keratinocytes were harvested via trypsinization, then washed in complete E-Low media, followed by a wash in ice cold PBS, and 1x Annexin-V-binding buffer (10 mM HEPES, 140 mM NaCl, 2.5 mM CaCl2, pH 7.4). Cells were then re-suspended in 100 μl of 1x Annexin-V-binding buffer. 5 μl of Alexa-Fluor 488 Annexin-V and 1 μl of 100 μg/ml propidium iodide solution was next added and the cells were incubated for 15 minutes at room temperature. 400 μl of 1x Annexin-V-binding buffer was added and the cells were then analyzed on a BD Cyan flow cytometer. Flow cytometry profiles were analyzed with FlowJo software (www.flowjo.com).

4.2.4 Quantifying Cellular Motility

Keratinocytes were transiently transfected with the pREX-H2B-mCherry plasmids and plated at a density of 400 cells/mm^2 on fibronectin- coated glass bottom 96-well plates (Matrical Bioscience, MGB096-1-2-LG-L). After 24 hours, time-lapse imaging was performed using an ImageXpress MicroXL imaging system (Molecular Devices), where an image of the mCherry channel (Excitation=562/40 nm, Emission= 641/75 nm) was acquired every 15 minutes for 4 hours for each condition in parallel. Time-lapse videos were analyzed as previously described using the Pathfinder program (Chapnick et al., 2013). Displaying of tracks was achieved using Pathfinder positional outputs for each cell, where each track for each cell was centered at coordinate (0, 0) using MATLAB in order to display normalized tracks. Data displayed represent one of two independent experiments done in triplicate, where each trial measured at least 200 cells. Error

displayed depicts the standard error of the mean for three trials. The p-value was calculated using t-test assuming equal variances in Excel.

4.3 Results

4.3.1 Identification of Msi2-regulated Pathways

With the HITS-CLIP, RNA-seq, and Ribo-seq data, I could identify processes and pathways controlled by *Msi2*. To identify a common group of dysregulated pathways in keratinocytes lacking *Msi2*, each parental dataset generated previously (Chapter 3.3.4) was analyzed using Ingenuity Pathway Analysis (IPA). Cell proliferation, survival, and cell cycle were among the top processes enriched in these datasets, consistent with the observed functions of *Msi2* in promoting cell growth (Park et al., 2015) (Figure 19A). Surprisingly, an enrichment for targets involved in cell movement also emerged. Indeed, when the high confidence targets (Chapter 3.3.4) were analyzed with IPA, cellular movement was identified as the most enriched process, followed by proliferation and survival (Figure 19B). In particular, KEGG pathway analysis highlighted genes involved in migration related processes such as focal adhesion, ECM-receptor interaction, and actin cytoskeleton (Figure 19C). Finding a common set of terms between the datasets strongly implies that *Msi2* regulates these processes.

Figure 19: Loss of Msi2 alters genes involved in proliferation, migration, and survival

A

RNA-seq Targets Upregulated
Molecular and Cellular Functions

Name	p-value	Genes
Cellular Assembly and Organization	7.80E-07 - 1.19E-36	451
Cellular Function and Maintenance	7.80E-07 - 1.19E-36	412
Cellular Growth and Proliferation	7.95E-07 - 3.83E-32	715
Cell Cycle	5.73E-07 - 6.53E-27	357
Cell Death and Survival	8.12E-07 - 4.86E-26	626

Ribo-seq Targets Upregulated
Molecular and Cellular Functions

Name	p-value	Genes
Cellular Movement	6.65E-06 - 1.48E-29	120
Cellular Growth and Proliferation	6.75E-06 - 2.51E-26	149
Cell Death and Survival	4.58E-06 - 9.91E-24	136
Cellular Assembly and Organization	7.08E-06 - 7.27E-20	92
Cellular Function and Maintenance	3.16E-06 - 7.27E-20	121

HITS-CLIP Targets (3'UTR peaks)
Molecular and Cellular Functions

Name	p-value	Genes
Cell-To-Cell Signaling and Interaction	7.06E-05 - 4.72E-19	95
Cellular Movement	8.21E-05 - 6.99E-19	94
Cellular Growth and Proliferation	8.27E-05 - 6.90E-18	130
Cell Death and Survival	8.30E-05 - 1.10E-17	119
Cell Morphology	5.92E-05 - 1.21E-13	96

RNA/Ribo-seq Targets Upregulated
Molecular and Cellular Functions

Name	p-value	Genes
Cellular Movement	5.51E-06 - 7.75E-31	111
Cellular Growth and Proliferation	4.55E-06 - 3.57E-28	135
Cell Death and Survival	5.83E-06 - 1.28E-23	122
Cellular Assembly and Organization	5.30E-06 - 3.21E-23	90
Cellular Function and Maintenance	3.50E-06 - 3.21E-23	108

B

High Confidence Targets Upregulated
Molecular and Cellular Functions

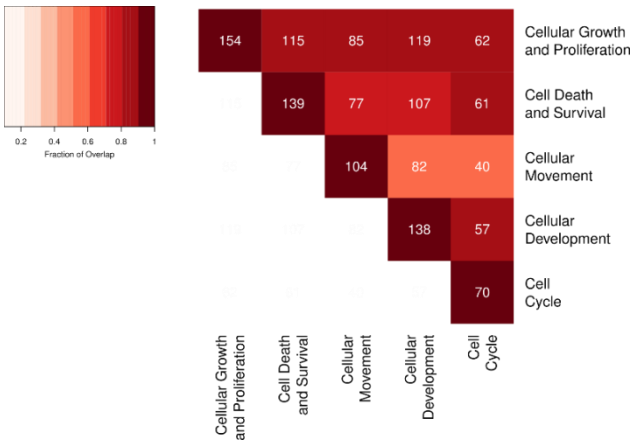
Name	p-value	Genes
Cellular Movement	1.12E-04 - 1.55E-20	51
Cellular Growth and Proliferation	1.19E-04 - 1.50E-16	61
Cell Death and Survival	1.27E-04 - 4.91E-14	57
Cellular Development	9.47E-05 - 1.94E-13	56
Cell Cycle	7.31E-05 - 2.69E-12	31

C

High Confidence Targets Upregulated
KEGG Pathway

Name	p-value	Genes
Focal adhesion	2.35E-08	COL4A1, MET, ITGB4, ITGA2, ACTN1 ITGB1, FLNB, FLNA,LAMB3, LAMA3, ITGAV, ITGA3, LAMC2, THBS1
ECM-receptor interaction	1.01E-07	LAMB3, COL4A1, LAMA3, ITGAV, ITGA2, LAMC2, ITGA3, THBS1, ITGB1, ITGB4
Regulation of actin cytoskeleton	9.35E-04	ITGAV, ITGB4, ACTN1, ITGA2, ITGA3, MYH9, ITGB1, TMSB4X, IQGAP1
p53 signaling pathway	1.09E-03	MDM2, PMAIP1, THBS1, CCNG1, IGFBP3, ATM
Adherens junction	7.56E-02	MET, CTNND1, ACTN1, IQGAP1
Cytokine-cytokine receptor interaction	1.08E-01	CXCL1, CXCL5, PPBP, MET, IL24, IL1A
Notch signaling pathway	1.67E-01	NOTCH2, DLL1, JAG1

D



(A) Ingenuity Pathway Analysis of parental and high confidence genesets. Colors coordinate to match those in Figure 17. The top terms for Molecular and Cellular functions are displayed. Only genes upregulated upon loss of *Msi2* were selected for analysis. (B) Ingenuity Pathway analysis of 88 high confidence *Msi2* targets. (C) KEGG Pathway enrichment of 88 high confidence *Msi2* targets. (D) Pairwise comparison of terms assigned each of the 88 genes to identify overlapping terms (Done by Kent Riemondy).

As mentioned previously, the list of *Msi2*-targets contains many genes that have functions in cell migration according to IPA, including *Flnb/a, Plec, Lama3*, and the eight targets selected for qPCR validation and RNA-stability measurements (Chapter 3.3.4 and 3.3.5), *Itgb1, Itga2, Cxcl5, Sema3c, Igfbp3, Hbegf*, and *Ppbp* (Appendix 8). Furthermore, consistent with Musashi's known functions in regulating the Notch signaling pathway (Imai et al., 2001), *Dll1, Jag1*, and *Notch2* are found on the list of high-confidence *Msi2* targets (Figure 19, Appendix 8). However, the previously identified *Msi2* targets *Numb, Pten* or *Cdkn1a* (Ito et al., 2010; Wang et al., 2015a) were not recovered in our datasets, although all three are detectable in keratinocytes. This may indicate that these targets are regulated by additional pathways that mask *Msi2* repression or that *Msi2* regulation of these targets may be cell type-specific.

From the IPA data I noticed that some genes overlapped multiple processes. To test the overlap between the terms assigned to each target, a pairwise comparison was performed by Kent Riemondy. It was found that many genes are associated with multiple processes such as proliferation and migration (Figure 19D). Indeed this overlap was one of the stronger overlapping set of terms in the data.

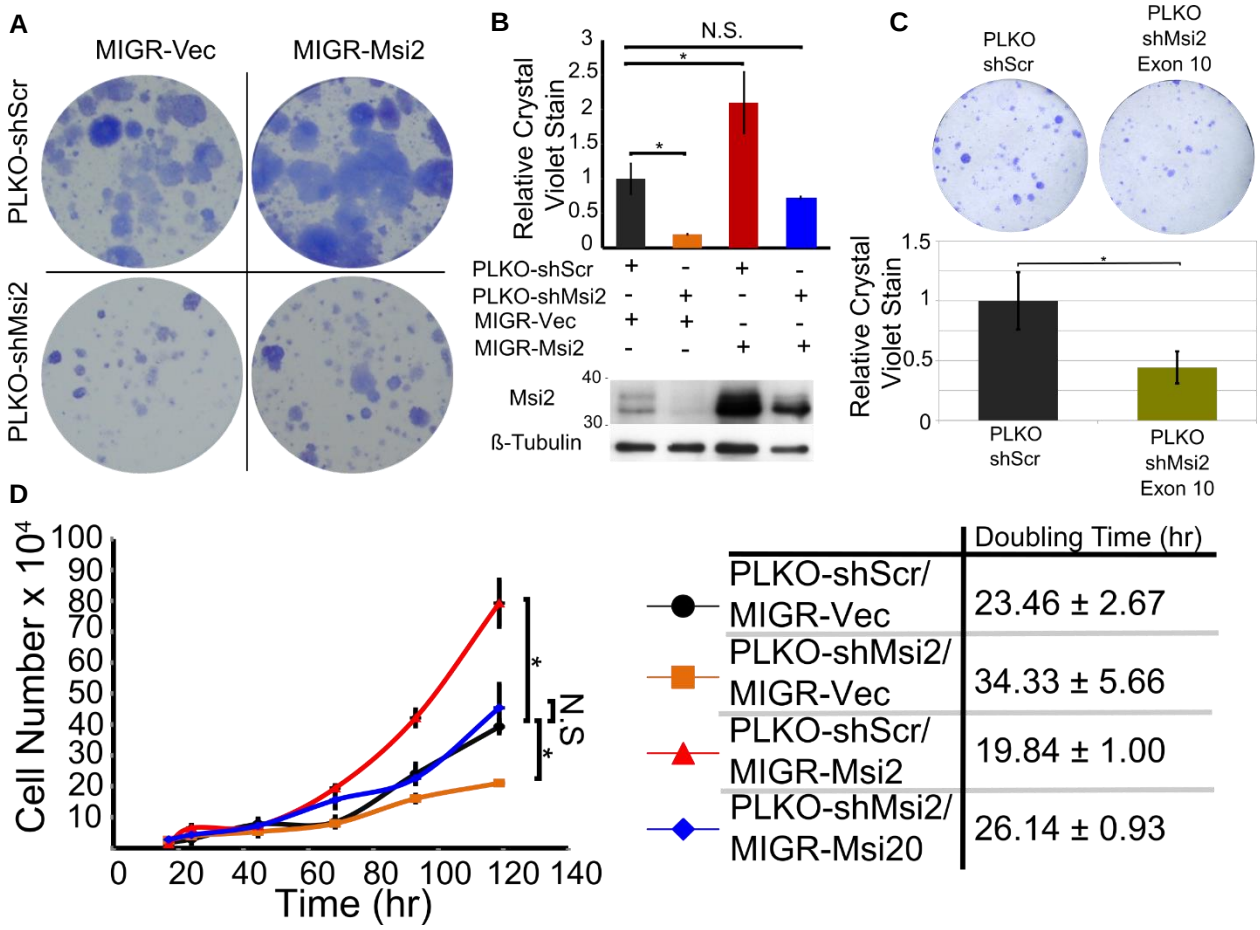
4.3.2 Msi2 Control of Cellular Proliferation

Msi2 has previously been shown to play a critical role in regulating cellular proliferation in many systems (Ito et al., 2010; Katz et al., 2014; Wang et al., 2015a). Interestingly, the list of high confidence targets contained genes involved in negative regulation of cell-cycle progression and pro-apoptotic factors including the DNA damage

checkpoint regulator *Atm*, and the pro-apoptotic inducer *Pmaip1*. Thus, I assessed the role of *Msi2* in regulating keratinocyte proliferation and survival. The ability for cells to form colonies and proliferate, though a multifactorial trait, provides insight into the importance of *Msi2* in cell proliferation, survival, etc. Four cell lines were generated using retroviral and lentiviral vectors, a control (pLKO-shScr, MIGR-vec), an *Msi2* knockdown (pLKO-sh*Msi2*, MIGR-vec), an overexpression (pLKO-shScr, MIGR-*Msi2*), and an *Msi2* rescued cell line (pLKO-sh*Msi2*, MIGR-*Msi2*) (Figure 20A). Keratinocytes with reduced *Msi2* demonstrated reduced colony proliferation capacity in contrast to *Msi2* overexpression, which resulted in enhanced colony proliferation capacity. Additionally, introduction of a shRNA resistant *Msi2* cDNA into the knockdown cells rescued the impaired colony proliferation capacity phenotype (Figure 20A). Western blot and crystal violet quantification of the colonies confirmed these findings (Figure 20B). As mentioned previously, a second shRNA was capable of reducing *Msi2* levels (Chapter 3.3.1). Colony forming analysis of cell lines made with this shRNA demonstrated similar results to the 3'UTR targeting shRNA (Figure 20C). To further characterize this phenotype and to help distinguish between adhesion defects, the doubling times of shScr, sh*Msi2*, overexpression and rescued keratinocytes were examined (Figure 20D). The first thing to note is the relatively equal numbers of cells starting around the 20 hour time point. This indicates that the ability of the cells to produce colonies is not in large part due to differences in the number of cells surviving plating or adhesion to the plate. Loss of *Msi2* increased doubling-time by ~50%. Whereas, overexpression of *Msi2* decreased doubling-time by ~15%. The rescue of

Msi2 in sh*Msi2* cell lines returned the doubling-time to near control levels, validating that the cell growth phenotype is a result of loss of *Msi2*, not shRNA off-targeting effects.

Figure 20: Msi2 promotes proliferation

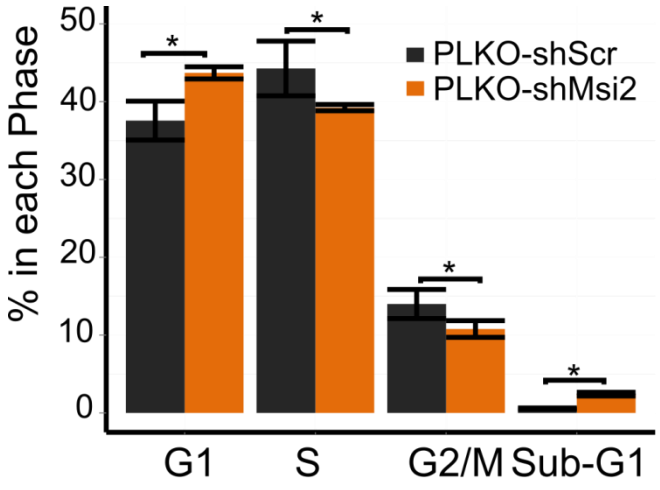


(A) Colony formation assay of keratinocytes with control (MIGR-Vec/shScr), *Msi2* overexpression (MIGR-*Msi2*/shScr), *Msi2* knockdown (MIGR-Vec/sh*Msi2*), or co-infected (MIGR-*Msi2*/sh*Msi2*) to rescue *Msi2* levels in the knockdown condition. Results are representative of two independent biological samples each assayed in triplicate. (B) Crystal violet quantification of colony forming and western blot of keratinocytes infected with the indicated lentiviral and retroviral constructs to knockdown or overexpress *Msi2*, respectively. Standard error of the mean displayed (C) Colony forming assay of keratinocytes with control (shScr) and *Msi2* knodown using CDS targeting shRNA (sh*Msi2* Exon 10). Crystal violet quantification as in B. Results are representative of 3 independent biological samples. (D) Growth curves of keratinocytes. Results are representative of n=2 independent biological replicated plated in duplicate. . * $p < 0.05$. (Student T-Test two-way). N.S.: not significant

4.3.3 Msi2 Control of Cell Cycle Progression

Since many of the *Msi2* targets are involved in cell cycle progression and control, cell cycle analysis by EdU incorporation was performed on shScr and sh*Msi2* keratinocytes. Loss of *Msi2* results in an increased G1 population and reduced S and G2/M populations, suggesting that *Msi2* regulates the G1-S transition (Figure 21). Additionally, the sub-G1 population, reflective of apoptotic or necrotic cells, was significantly increased in the *Msi2* knockdowns, indicating that the observed growth differences are caused by defects in both cell proliferation and survival (Figure 21).

Figure 21: Msi2 promotes cell cycle



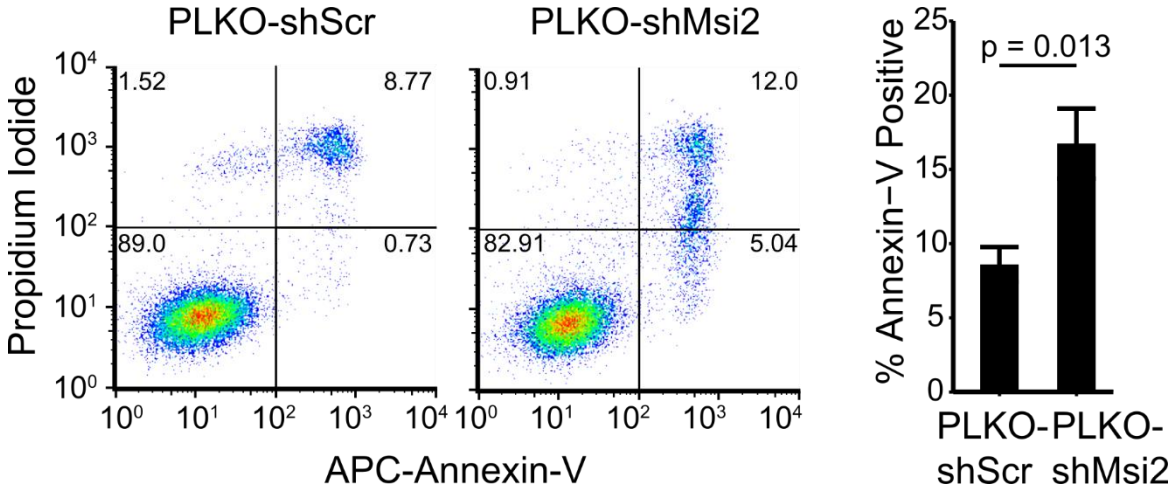
	PLKO-shScr	PLKO-shMsi2	
G1	37.57 ± 2.5	43.70 ± 0.78	*
S	44.27 ± 3.52	39.23 ± 0.40	*
G2/M	14.00 ± 1.87	10.79 ± 1.07	*
Sub-G1	0.56 ± 0.0	2.41 ± 0.24	*

Cell cycle analysis of EdU pulsed keratinocytes with shScr and sh*Msi2*. Representative chart of cell populations shown with quantification of n = 5 biological replicates. * $p < 0.05$. (Student T-Test two-way).

4.3.4 Msi2 Control of Survival

To measure apoptosis in *Msi2* knockdown cell lines, Annexin-V vs PI FACS was performed by Kent Riemondy. Annexin V marks apoptotic cell populations while PI distinguishes this population into early and late apoptotic cells. The knockdown cells showed ~2-fold increased levels of Annexin-V positive apoptotic cells when compared to control scrambled cells (Figure 22). Indeed, the late apoptotic (PI high, Annexin V positive) and early apoptotic (PI low, Annexin V positive) populations were increased. Together, these results provide evidence that *Msi2* regulates cell proliferation by controlling cell cycle progression and survival in keratinocytes, similar to observations in other non-stratified epithelial systems (Ito et al., 2010; Kharas et al., 2010; Wang et al., 2015a).

Figure 22: Msi2 promotes cell survival

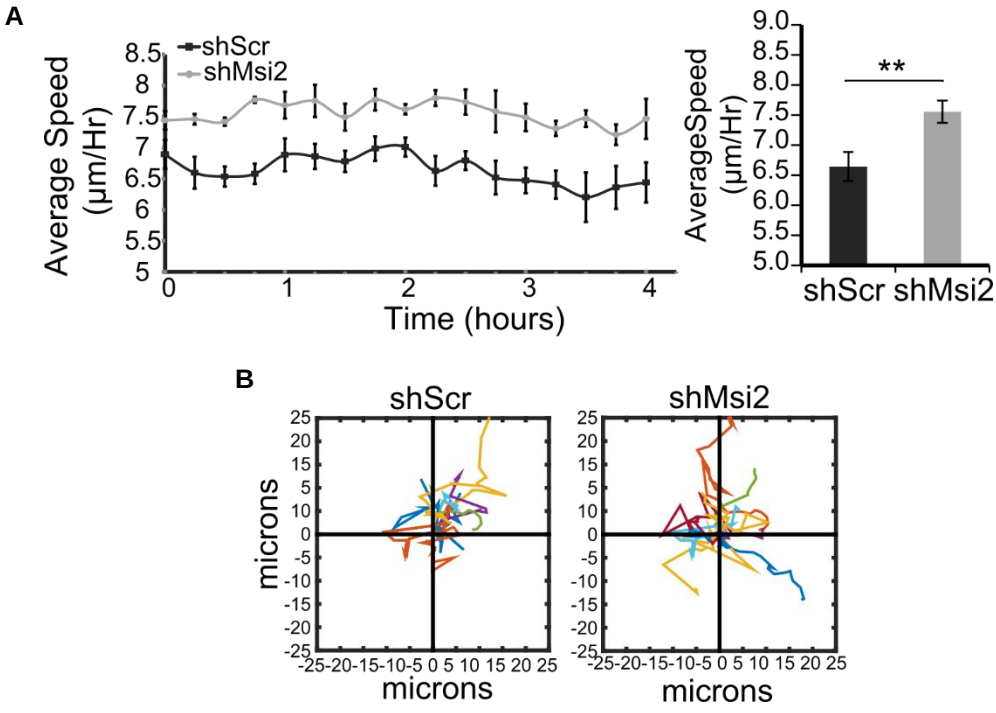


Propidium iodide and Annexin V flow cytometry analysis performed in triplicate to identify the population of apoptotic cells. Representative result displayed from n = 3 independent experiments (Student T-Test two-way) (Done by Kent Riemondy).

4.3.5 Msi2 Control of Migration

To investigate the role of *Msi2* in cell migration, individual cells were tracked using live-cell imaging and measuring cell migration velocity. This approach allowed me to investigate cellular migration differences without the confounding effects of cellular proliferation rates between sh*Msi2* and shScr cell lines. Consistent with the identification of numerous *Msi2* targets involved in cell migration, loss of *Msi2* enhanced migration velocity compared to scrambled control (Figure 22A). Migratory tracks of individual cells revealed no tendency for changed migratory patterns upon *Msi2* knockdown (Figure 23B).

Figure 23: Msi2 inhibits cellular migration

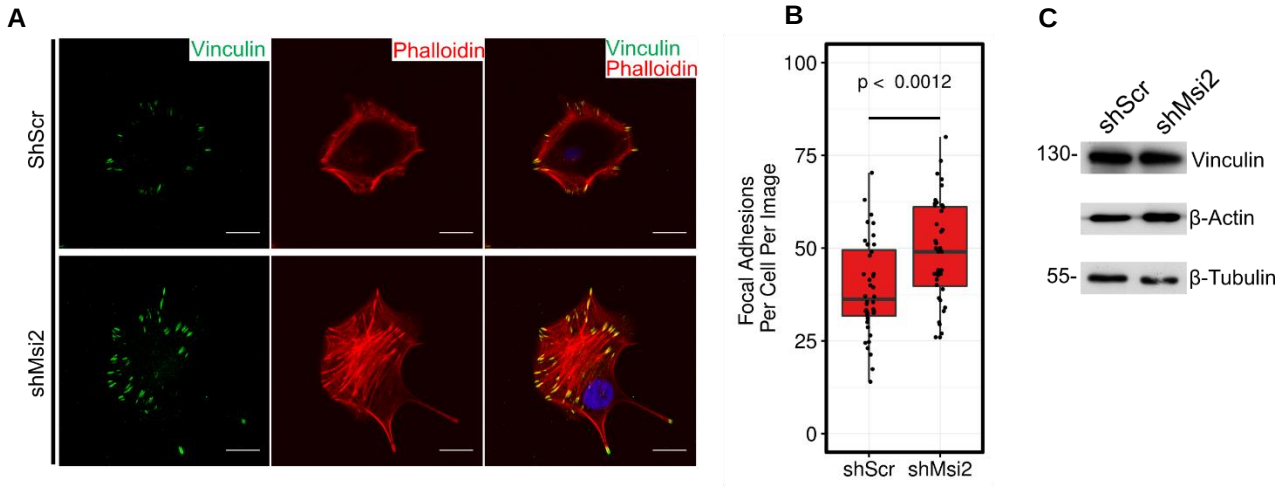


(A) Cellular migration speed is measured over time for the control and *Msi2* knockdown keratinocytes (right panel). Average migration speed for the control and *Msi2* knockdown keratinocytes (left panel). Data shown are representative of 2 independent experiments. (D) Example cell migration tracks for the control and *Msi2* knockdown keratinocytes measured in microns. * $p < 0.05$ ** $p < 0.01$ student two-way T-Test.

4.3.6 Number of Focal Adhesions is increased in shMsi2

The pathways that are most enriched for *Msi2* targeted genes are regulators of focal adhesion and integrin molecules (Figure 19C). Extensive studies have linked focal adhesion and integrins with cell migration (Case and Waterman, 2015; Huttenlocher and Horwitz, 2011; Ridley et al., 2003). To understand how *Msi2* target regulation affects focal adhesions, Kent Riemondy and I began by looking at perturbations in focal adhesion numbers and size in sh*Msi2* cell lines using immunofluorescence. Vinculin and Phalloidin mark FAs and actin, respectively, and provide a means by which focal adhesions can be visualized. A significant increase in FA numbers was observed in the *Msi2* knockdown cells (Figure 24A-B). The absolute amount of *Actin* and *Vinculin* did not change (Figure 24C). These data show that elevated expression of migration-associated genes in the absence of *Msi2* correlates with increased FA formation, likely leading to increased cell migration in these cells.

Figure 24: Loss of Msi2 results in increased numbers of focal adhesions

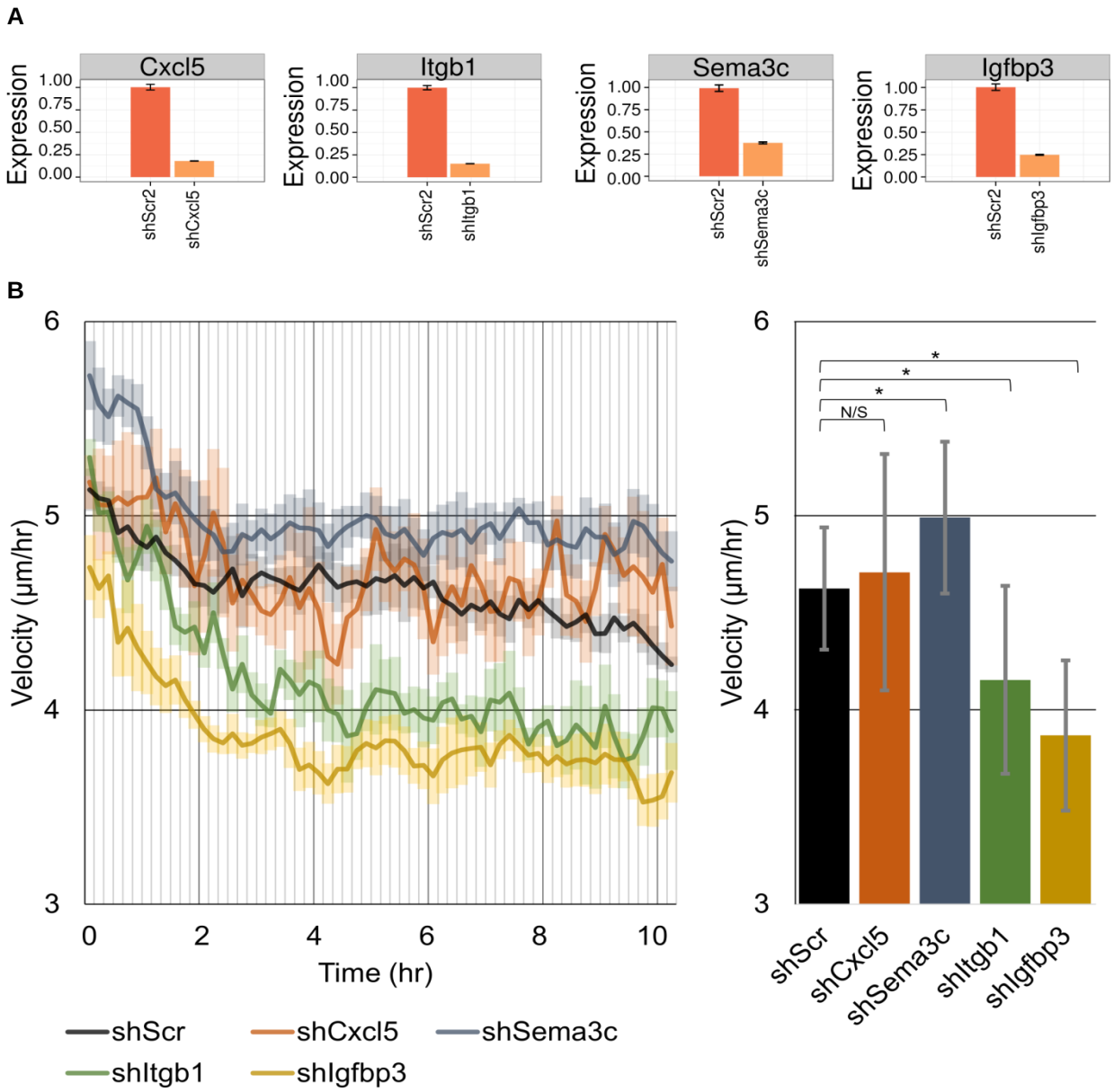


(A) Immunofluorescence of a focal adhesion marker, Vinculin (green), and staining for actin, Phalloidin (red), for shScr and sh*Msi2* keratinocytes. Scale bars = 20μm (B) Quantification of focal adhesion numbers per cell per image for 40 images for shScr and sh*Msi2* keratinocytes. (C) Western blot for Vinculin, beta-

Actin, and beta-Tubulin in shScr and sh*Msi2* cell lines. * student two-way T-Test. (Done by Kent Riemondy)

4.3.7 Knockdown of Msi2 Targets

Figure 25: Sum of Msi2 controlled genes contributes to migration control



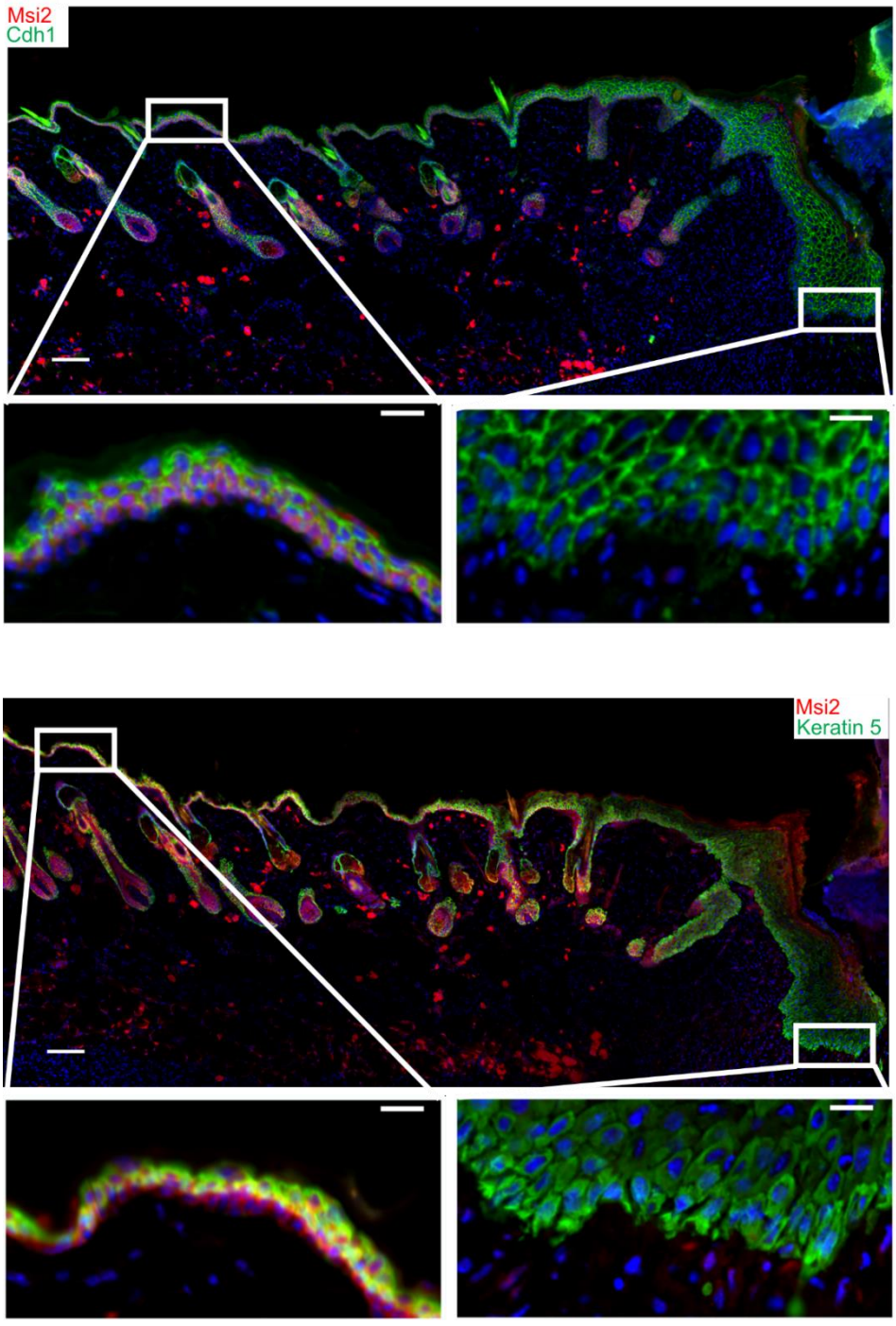
(A) qPCR validation of Cxcl5, Itgb1, Sema3c, and Igfbp3 knockdown via shRNA. (B) Average cell velocity over time for targets as determined by live cell imaging. (C) Average velocity of *Msi2* target knockdown cells versus scrambled shRNA. Error bars = Standard Error Mean. * p<0.05. n=5 technical replicates

To further investigate the link between *Msi2* and cellular migration, four validated targets, *Sema3c, Itgb1, Igfbp3,* and *Cxcl5* were knocked down in keratinocytes and measured for migration. Knockdowns were confirmed via qPCR (Figure 25A). The cells were expected to migrate slower than controls, the opposite phenotype of *Msi2* knockdown. Indeed, knockdown of *Itgb1* and *Igfbp3* demonstrated significantly decreased keratinocyte migration (Figure 25B). However, *Cxcl5* and *Sema3c* showed no change or increased migration, respectively. These results imply that *Msi2* regulation of a several target may influence cellular movement.

4.3.8 Msi2 Expression in Wounded Skin

To support a role for *Msi2* in regulating cell migration in a physiologically relevant condition, Kent Riemondy examined *Msi2* expression during wound healing in adult mouse skin. Wounded skin samples were stained with *Msi2*, keratin 5 (progenitor marker), and cadherin 1 (epithelial cell marker). In both the basal epidermal progenitors and the hair follicle stem cells, *Msi2* is readily detectable as demonstrated previously (Chapter 2) (Figure 26). However, within the epidermis at the leading edge of the wound, *Msi2* is nearly undetectable, yet returns gradually with increasing distance away from the wound site (Figure 26). This observation suggests an intriguing possibility that the downregulation of *Msi2* is a prerequisite step to promote cell migration in keratinocytes at the leading edge of the wound and that precise control of *Msi2* abundance may be required for proper wound healing processes.

Figure 26: Msi2 levels are reduced in wounded skin



Immunofluorescence of *Msi2* (red) and an epithelial cell marker, E-Cadherin (green), or *Msi2* (red) and a basal cell marker, Krt5 (green), in a 7-day old skin wound on mouse backskin. Asterisks (*) represent the granulation tissue at the wounded site. Nuclei are shown in blue. Scale bars = 100 μm (inset = 20 μm) (Done by Kent Riemondy)

CHAPTER 5

DISCUSSION AND FUTURE DIRECTION

Musashi RNA-binding proteins were originally identified in Drosophila as regulators of sensory organ development (Nakamura et al., 1994). They have since been shown to maintain stem cell populations and promote tumorigenesis in mammalian tissues (Ito et al., 2010; Katz et al., 2014; Kharas et al., 2010; Park et al., 2014; Sakakibara et al., 2002; Wang et al., 2015a). Due to these important functions, recent research has begun investigating Musashi proteins' RNA-binding properties and mechanisms of regulation (Li et al., 2015; Park et al., 2015). Current studies, however, have mainly focused on *Msi1*, leaving *Msi2* less well characterized (Imai et al., 2001). This is despite *Msi2* having been shown to play a critical role in maintaining stem cell dynamics in many tissues (Ito et al., 2010; Wang et al., 2015a). In stratified epithelia, *Msi2*'s functions, mechanisms, and targets remain unknown. Therefore, I studied the function of *Msi2* in the epithelial progenitor cells, keratinocytes. In this study, I identified *Msi2* targets, determined a possible mode of *Msi2* regulation, and characterized the pathways that *Msi2* regulates in skin keratinocytes. Here, I will discuss the implications of my findings, what I learned about Musashi biology, and how it relates to skin biology.

5.1 *Msi2* Target-binding Characteristics

5.1.1 Msi2 is the dominant Msi expressed in highly proliferative keratinocytes

To begin to understand the role *Msi2* plays in the skin and the populations expressing *Msi2*, I stained skin sections at differing developmental stages. These data demonstrated that *Msi2* is expressed in basal epidermal progenitor cells, bulge stem

cells, and the matrix in postnatal and adult animals. Epidermal progenitor cells and matrix cells are highly proliferative cell types that give rise to the stratified epithelium and the hair shaft respectively, while the bulge stem cells are slowly proliferating cells that generate new hair follicles during the hair cycle (Arwert et al., 2012; Blanpain and Fuchs, 2006; Schneider et al., 2009). It is not surprising to find highly proliferative basal progenitor cells expressing *Msi2*, given that *Msi2* regulates proliferation and maintains stem cell identity (Sutherland et al., 2013). Additionally, despite the relatively low proliferation rate of bulge stem cells, it is not surprising that these cells express *Msi2*, in light of the protein's aforementioned role in stem cells. It is, however, interesting to see expression in the transiently amplifying cell population of the matrix. This indicates that *Msi2* may have different functions in various cell populations within the same tissue. In keratinocytes *Msi2* likely regulates both stem cell maintenance and proliferation, while regulating only one of these processes in the matrix or bulge.

It is interesting that *Msi2* expression in the matrix is polarized to one side of the matrix. This opens the possibility that *Msi2* regulates a process that requires polarization in the matrix. Sonic hedgehog (*Shh*) expression and signaling has an opposite polarization to that of *Msi2* (Fan and Khavari, 1999; St-Jacques et al., 1998). Thus, given *Msi2*'s known function in regulating Notch signaling, *Msi2* may regulate Shh signaling in the matrix. However, since the epidermal progenitor cells are more amenable to cell culture and molecular approaches to studying RNA-binding proteins, I decided to study the role of *Msi2* in keratinocytes.

Based on the similarity of *Msi1* and *Msi2*, in addition to a previous study showing synergy between both proteins, I reasoned that *Msi1* might compensate for loss of *Msi2*

(Wuebben et al., 2012). To address this possibility, I assessed expression of *Msi1* in the skin. I first attempted *Msi1* immunostaining; however, I was unable to detect convincing levels of *Msi1* in skin or brain positive control samples. Fortunately, previous researchers had published RNA sequencing results from epidermis. Using these data, I showed that *Msi1* is not highly expressed in keratinocytes. Additionally, proteomics data from a previous study in the Yi lab did not indicate the presence of *Msi1* protein in skin (Wang et al., 2012). This indicates that *Msi2* is the dominant Msi species in keratinocytes and should not be subject to compensation by *Msi1*. Further support that compensation from *Msi1* would not pose a problem in my experiments came from my own sequencing data, which confirmed that *Msi2* is the dominant Msi protein, and that *Msi1* levels to not dramatically increase upon loss of *Msi2*.

Thus, based on expression pattern, *Msi2* could be a critical factor in controlling proliferation and stemness in skin cell populations, the loss of which would be predicted to lead to deterioration of skin function and hair follicle maintenance in the absence of *Msi1* compensation, which appears unlikely. Furthermore, it will be interesting to see if increasing the levels of *Msi2* in the epidermis, bulge, or matrix will lead to tumor formation.

5.1.2 Msi2 binds 3'UTRs at UAG-rich regions

HITS-CLIP is a method of isolating RNA-binding proteins and their direct targets (Darnell, 2012). I used HITS-CLIP to identify *Msi2*-bound RNAs and elucidate specific *Msi2*-binding sites *in vivo*. The autoradiogram generated during the HITS-CLIP protocol showed signal sensitivity to RNase concentration and was consistent with the size of

Msi2 crosslinked to an RNA, indicating that *Msi2*:RNA hybrids were isolated. The libraries, themselves, were of varying sizes ranging from short (<20 nucleotides) to long (>100 nucleotides). This provided a platform for confidently identifying *Msi2* targets and targeted regions using the longer reads and identifying peak summit regions using the shorter libraries.

Using these data, I found that 3'UTRs are the predominant region recognized by *Msi2*, consistent with the role of *Msi2* as a post-transcriptional regulator (Figure 8). The extent to which the data shows *Msi2* binding preference to 3'UTRs is dependent on how the data are filtered. Raw reads and unfiltered peaks showed approximately 30% alignment to 3'UTRs. Practically, there isn't a way to assign a filtering cutoff non-arbitrarily without testing many cutoffs simultaneously. Even then, it is difficult to determine which metric is most appropriate as a filter. Ultimately, I decided to filter the data by the absolute number of reads per peak and the number of libraries that have a read (biological complexity) and measure the change in percent peak alignment. The assumption was that nonspecific binding events will have low read coverage and/or biological complexity, and at some point the percent alignment to each region will stabilize and not vary to a significant degree. I found that, starting at 10 reads per peak, the percentage of peaks aligning to 3'UTRs stabilized at about 40-60%. Any more stringent filtering only decreased the number of peaks detected without dramatically altering the percentages. Thus, I decided to use this lowest filtering metric in order to identify more targets, and focused subsequent analysis on mRNA regions, 3'UTRs, 5'UTRs, introns, and CDSs in light of *Msi2*'s known role in gene regulation and the dominance of these collective regions in the data. I discarded the intergenic and

noncoding RNA annotations as nonspecific binding or alignment artifacts. However, due to the stabilization of percent alignment across multiple filtering stringencies, these sites could represent *bona fide Msi2* binding events and may indicate novel *Msi2* regulation (Appendix 2). It would be interesting to examine these possible noncoding RNA targets of *Msi2* and ask if there is any indication of regulation through ribo-minus RNA-seq.

One interesting side effect of sequencing longer libraries was the coverage of many of the peaks across the target's 3'UTR. In some cases, the entire annotated 3'UTR was covered. This made identification of direct *Msi2* binding sites, presumed to be peak summits, difficult. Fortunately, the shorter read libraries intensified the peak summits to the point that peak splitter software, coupled with peak intensity (point with highest coverage read density), could be employed to identify the summit center. The peak summit regions were defined as a 50-nucleotide region surrounding the peak summit, allowing me to assess *Msi2* binding motifs in peaks in different genomic regions.

De novo motif searching on these data revealed a potentially critical trinucleotide UAG core motif in the 3'UTR regions and, to a lesser extent, the CDS (Figure 9). While this core motif has been identified *in vitro*, this is the first time it has been detected from transcriptome-wide data *in vivo*, indicating that the target are real targets. Even though filtering was applied, since the 5'UTR and introns did not produce a coherent set of motifs, it is likely that those observed binding events are noise. However, the possibility remains that other structural elements are required, and that these observed sites represent true binding events. Indeed, it has been proposed that Msi proteins may

require the motif to reside at the terminal loop of an RNA hairpin. This could not only allow *Msi2* to bind the 5'UTR and intronic regions, but also may be required to provide specificity for binding a certain UAG in the correct structural context, since a UAG will occur every 64 nucleotides by chance. However, because of technical difficulties assigning a direct *Msi2*-binding site, I was not able to detect any preference for a hairpin in the peak summit. Interestingly, an enrichment was found for two or more closely positioned UAG motifs in *Msi2* 3'UTR peak summits (Figure 10, 11). A previous study suggests that each RNA-recognition motif in Msi can independently bind a UAG. The enrichment for UAGs in the peak summit may indicate that this could be happening *in vivo* and may be responsible for *Msi2* target recognition. These findings provide a link between the UAG motif that has been determined biochemically *in vitro* and the *Msi2*:mRNA binding events occurring *in vivo*.

Msi2-HITS-CLIP has been previously reported in the literature (Park et al., 2014, 2015; Wang et al., 2015a). In Chapter 2, I show that many of the targets in my data overlap with targets identified in hematopoietic and gastrointestinal systems. The non-overlapping targets may be a result of cell type-specific binding. Indeed, when my data were separated into unique and overlapping targets and run through GO term analysis, I found that my data contained terms associated with migration. The overlapping targets showed enrichment for basic cellular processes. This suggests that *Msi2* regulates a common set of targets involved in general cellular maintenance. This indicates that *Msi2* may regulate different processes, depending on the genes expressed in different cell types.

Through these experiments, I determined that *Msi2* preferentially interacts with 3'UTRs at regions enriched with UAG elements, and generated a list of potential targets. 3'UTR binding is common among post-transcriptional regulators, as it provides a platform upon which ribosomes will not knock the binding-proteins off, increasing their residency time, and gives them access to the poly-A tail and, through circularization, the mRNA cap for regulation of stability or translation.

5.2 Mechanism of *Msi2* Target Regulation

5.2.1 Msi2 functions by binding the 3'UTR and reducing transcript abundance

Since 3'UTRs are the most dominant region preferred by *Msi2*, and contain a coherent set of binding motifs, I restricted the list of targets to those with 3'UTR binding. With a list of targets in hand, I sought to determine the molecular mechanism by which *Msi2* regulates its targets. To accomplish this, I turned to RNA-seq and ribo-seq using *Msi2* depleted cells and measured mRNA abundance and ribosome occupancy, respectively. *Msi2* knockdown cells were generated using one of two shRNAs targeting the 3'UTR or CDS of mouse *Msi2*. Both shRNAs efficiently reduced *Msi2* levels. However, since I wanted to be able to show an overexpression and rescue of *Msi2* during phenotypic validation of my findings, I chose to use the 3'UTR-targeting shRNA. Libraries were filtered similarly to that of HITS-CLIP, with a requirement of at least 10 reads aligning to CDS per gene. All HTseq counting was performed on CDS in the RNA-seq data to make it as comparable to Ribo-seq as possible.

It had been generally accepted, from studies on *Msi1,* that *Msi2* functions by regulating translation rather than RNA abundance. However, indications from other

sequencing studies suggested that *Msi2* may play a role in modulating RNA abundance (Katz et al., 2014; Okano et al., 2002; Sakakibara et al., 2001). In one study the authors noted that one of their *Msi2* targets, *Jag1*, changed at the RNA level in addition to its ribosome occupancy suggesting *Msi2* may regulate transcript abundance (Katz et al., 2014). Indeed, through the RNA-seq data, I found that *Msi2* regulates mRNA abundance, with subsequent changes in ribosome occupancy as shown in the Ribo-seq data for HITS-CLIP identified targets. This indicated that the change in RNA abundance is responsible for the change in ribosome occupancy and confirmed through lack of translation efficiency changes (Figure 16) (Katz et al., 2014; Kawahara et al., 2008). Further confirmation of this binding comes from analysis of targets with multiple binding sites. The expectation is that *Msi2* will have a stronger regulatory effect if more *Msi2* binds the 3'UTR. Indeed, the more *Msi2* binds, the greater the regulatory effect on its targets becomes. 5'UTR and intron binding targets did not show significant regulation indicating that binding in these regions is noise in the system or is required for another process besides RNA abundance regulation. CDS binding showed a trend toward functional RNA regulation but because the number of genes was so small this it is hard to assess with confidence. Thus, *Msi2* predominantly or entirely regulates targets by binding the 3'UTR of the transcript.

The HITS-CLIP data together with RNA-seq and Ribo-seq provide a means to generate a list of high-confidence targets, bound and regulated by *Msi2*. By merging the data I found 119 targets, 88 of them upregulated when *Msi2* was knocked down. These 88 targets were defined as canonical *Msi2* targets. The downregulated genes may indicate that *Msi2* can decrease RNA abundance or that *Msi2* regulation is

overshadowed by perturbations in cellular processes. Because these issues deviate from the focus of the study, and due to time constraints, I did not explore these possibilities.

The aforementioned sequencing data were generated from a set of commonly derived knockdown cell lines grown under similar conditions. I wanted to validate that the changes observed in RNA-seq were a result of loss of *Msi2* and not of culturing conditions or the derivation of the cell line. Independently derived cell lines were isolated on different days, and 8 *Msi2* targets with 1 non-target from the list of 88 were validated through qPCR and demonstrated similar upregulation upon loss of *Msi2*, as shown through RNA-seq. These data indicate that *Msi2* primarily functions by binding 3'UTRs and regulating RNA abundance.

5.1.2 Msi2 likely regulates target stability

This change in RNA abundance could be explained by one of three possibilities: 1) *Msi2* regulates translation and the RNA abundance changes in response to lower translation of the target, 2) *Msi2* regulates RNA stability, or 3) *Msi2* regulates transcription or processing of its target transcripts. *Msi2* localization is predominantly cytoplasmic based on our staining, and the protein is not known to target factors involved in RNA processing or transcription, thus option 3 is likely not possible and was not explored. The remaining option is RNA stability change, either directly or indirectly. To test this, the eight high-confidence targets used in RNA-seq validation were used for stability measurements. Generally, to do this kind of experiment, cells are treated with an inhibitor of RNA polymerase, usually α-amanitin or Actinomycin D and the targets are

measured over time to generate a decay curve (Chen et al., 2008; Lindell et al., 1970). For my purposes, I used Actinomycin D. The data generated from this assay indicated that the stability of six of the targets, and possibly all targets, is increased in response to loss of *Msi2*, supporting a regulatory role for *Msi2* in modulating target stability (Figure 18). The change in the other two targets (*Itga2* and *Ppbp*) could not be detected with the primers used in qPCR and/or suffered from normalization problems, and were discarded.

This still does not answer how stability is regulated. It is possible that the primary function of *Msi2* is to regulate translation, with RNA stability changing as a result. Another possibility is that *Msi2* regulates RNA localization for autophagy or decay. This was not tested in this study but could be done while exploring the regulatory mechanisms of *Msi2*. Despite this, I can say with some confidence that *Msi2* has an effect on RNA stability for targets where *Msi2* binds the 3'UTR. Thus, the effect of *Msi2* regulation is similar to what is ascribed to miRNAs (Eichhorn et al., 2014), another prominent class of post-transcriptional regulators. It will be critical to understand how *Msi2* regulates the stability of its targets in order to gain a better understanding *Msi2* function.

5.3 *Msi2* Pathway Modulation

5.3.1 Msi2 regulates targets involved in proliferation, survival, and migration

The HITS-CLIP data together with RNA-seq and Ribo-seq data provide an unparalleled platform to identify pathways and processes regulated by *Msi2*. A list of high-confidence targets, bound and regulated by *Msi2*, was generated from the combine

set of parental (RNA-seq, Ribo-seq, HITS-CLIP) data. This list was enriched for targets upregulated upon knockdown of *Msi2*, further supporting *Msi2*'s role as a post-transcriptional regulator of gene expression. Some downregulated genes may be subject to other disrupted secondary regulation that masks the effects of *Msi2* de-repression. Other genes, for example *Perp*, are also bound by Argonaute and found to be downregulated in other datasets unrelated to *Msi2*, indicating that these genes may be false positives (Riemondy et al., 2015). Because of this observation and what is known about *Msi2*, I opted to only look at those genes that were upregulated upon loss of *Msi2* in each dataset and to dismiss the downregulated genes as secondary changes.

IPA on the parental and high-confidence datasets identified regulation of known processes like proliferation and survival, indicating that *Msi2* regulates proliferation in most cell types. Unexpectedly, I also identified genes associated with migration. Each dataset showed enrichment for similar terms, indicating that these processes are likely substantially perturbed in cells lacking *Msi2*. Many of the identified high-confidence genes share a functional role with cell proliferation and migration. Thus, *Msi2* could play a role in regulating these processes in keratinocytes by regulating a core set of targets that are involved in both.

Based on the expression data supplied to IPA, the program attempts to determine whether a pathway or process is enhanced or reduced by a protein. IPA determined that migration is likely to be enhanced when *Msi2* is knocked down, an assessment that turned out to be true, as discussed below. However, IPA also suggested that proliferation is likely to be enhanced which turns out to be false. This

implies that the identified targets may be regulating proliferation and migration differently in various cell populations depending on the transcriptional environment. Taking this idea further, it may indicate that *Msi2* can target a set of common genes yet have different phenotypic consequences.

KEGG analysis of the high-confidence targets revealed genes involved in focal adhesions, ECM-receptor interactions, Notch signaling, and the actin cytoskeleton (Figure 19, Appendix 8). The regulation of the Notch pathway has previously been observed in other studies, suggesting that *Msi2*-mediated regulation of the Notch pathway is more widespread than currently appreciated (Kaeda et al., 2014; Moghbeli et al., 2015; Rezza et al., 2010). However, the genes involved in Notch signaling, Notch2 and Dll1, have not previously been identified as *Msi2* targets. Thus, *Msi2* may regulate Notch signaling through multiple means depending on what factors are expressed in each cell type. Integrin signaling and focal adhesion formation were another top pathway identified in these data. It would be interesting to know if *Msi2* regulates this pathway in addition to Notch, since integrin signaling is critical for adhesion and survival in keratinocytes. To determine how widespread *Msi2* targeting of genes involved in migration is, datasets from this and other studies in hematopoietic and gastrointestinal systems were compared. The migration-related targets appeared to be specific to keratinocytes, implying *Msi2* can regulate different pathways depending on the transcriptional landscape (Figure 13). Additionally, there appeared to be a lack of a proliferation GO term in the common dataset. I interpreted this to mean that proliferation is controlled by commonly expressed and cell-type specific targets in order to regulate multiple pathways in different cell types. This is supported by the detection

of *Msi2* targets that have overlapping regulation in both migration and proliferation in keratinocytes.

5.3.2 Msi2 regulates proliferation and survival

Regulation of proliferation is a common theme in Msi biology (Sutherland et al., 2013). Indeed, I detected genes associated with proliferation in this study. Phenotypic characterization of *Msi2*-depleted cells revealed reduction in colony proliferation potential, converse to the enhancement of colony formation in *Msi2* overexpressing cells. Growth curve measurements demonstrated that the most likely reasons for these observed differences in colony forming capacity were proliferation or survival, since the starting number of cells was similar in each sample. This does not exclude the possibility that cells could have detached from the plate throughout growth of the colonies, but this is unlikely as the starting number of cells would be expected to vary across the samples as well. Cell cycle analysis demonstrated that there is a reduced S phase entrance and increased sub-G1 population, indicative of cell death as predicted from the genes identified as *Msi2* targets. Indeed, PI and Annexin V FACS demonstrated that the *Msi2* knock-down lines contain an increased population of apoptotic cells. Taken as a whole these data support my findings from the sequencing data that *Msi2* targets and regulates cellular proliferation and survival.

5.3.3 Msi2 regulates migration

IPA analyses identified many targets involved in the regulation of cell migration. Indeed, the eight mRNA targets selected for qPCR validation and RNA stability are involved in cell migration, as either structural adhesion molecules or regulators of cell

migration processes. Among *Msi2* targets, *Hbegf* is a growth factor that promotes migration in epithelial cells (Mine et al., 2005). *Itga2*, *Itgb1* are integrin subunits that can form heterodimeric transmembrane protein complexes that mediate cell-cell or cell-extracellular matrix attachment, and have stimulatory or suppressive effects on cell migration (Huttenlocher and Horwitz, 2011). Similarly, the glycoprotein *Thbs1* also mediates cell-to-cell or cell-matrix interactions through binding a multitude of receptors, and can act to simulate or inhibit migration depending on the cellular context (Streit et al., 2000). *Cxcl5* and *Ppbp* are members of the secreted chemokine CxC ligand family, which stimulate cell migration in addition to providing chemoattractant cues (Kuo et al., 2011; Yu et al., 2010). Additionally, *Sema3c*, a member of the semaphorin family, stimulates cell migration in diverse tissues (Goshima et al., 2002). Lastly, *Igfbp3*, an insulin growth factor binding protein, can stimulate or suppress cell migration depending on the cellular context (Chang et al., 2007). Taken together, *Msi2* may control cellular migration processes.

Thus, using live cell imaging, I demonstrated that the loss of *Msi2* significantly increased cell motility of keratinocytes (Figure 26). Further analysis looking at individual cell tracts demonstrated that the keratinocytes move as randomly as control cells. In an endogenous context it is likely that these cells would have a chemoattractant cue to determine which direction to move. Interestingly, *Msi1* has been found to control migration in glioblastomas (Uren et al., 2015; Vo et al., 2012). It is possible that this is a common pathway controlled by Msi proteins, but has been missed in *Msi2* until now. *Msi2* has been predominantly studied in hematopoietic systems, and only recently, in gastrointestinal systems (Park et al., 2014, 2015; Wang et al., 2015a). Migration

defects could be observed in the gastrointestinal system but have been missed. Another possibility is that *Msi2*-mediated regulation of migration is specific to certain cell types, among which are keratinocytes.

Knockdown of four *Msi2* targets associated with migration should demonstrate reduced migration. However, live cell imaging showed varying results with two *Msi2* targets, *Itgb1* and *Igfbp3*, showing reduced migration, one, *Cxcl5*, showing no change, and one, *Sema3c*, showing increased migration. Thus, these data imply that it is the sum of *Msi2* regulation that leads to regulation of migration.

These findings reveal that *Msi2* may have dual functions in promoting cell proliferation and inhibiting cell migration in the epidermis. It will be critical to investigate whether *Msi2*'s regulation of these processes is linked through the aforementioned signaling pathways.

5.3.4 Msi2 is lost in migrating keratinocytes

In keeping with the duality of *Msi2*'s function, the protein appears to be downregulated in the leading edge of wounded skin (Figure 26), suggesting that *Msi2* downregulation is an important prerequisite for stimulating cell migration in this cellular context. The duality of *Msi2*'s function is particularly interesting, as proliferation drops while migration increases. In most cases regulation of these processes is positively correlated, as is also true for *Msi1* (Vo et al., 2012). It is possible that these processes are inversely correlated to maintain basal localization of proliferative keratinocytes and prevent them from moving. A hallmark of many malignant cancers is their ability to proliferate and migrate (Lukong et al., 2008). Cells have multiple checkpoints to ensure

faithful execution of migration or proliferation. *Msi2* may play a checkpoint role in giving keratinocytes the option to proliferate without migration or to migrate without proliferating.

5.4 Concluding Remarks and Future Direction

Through this study, I have characterized *Msi2* binding, mode of regulation, and control of certain cellular processes using three powerful sequencing technologies that have not been used together before. I found *Msi2* to preferentially bind 3'UTRs of genes involved in proliferation, survival, and migration, at regions rich in UAG, and to regulate the stability of these target transcripts. Additionally, these processes were found to be dysregulated in *Msi2* depleted keratinocytes. These data have answered fundamental questions about *Msi2* biology. However, many questions remain. Most pressing, is how *Msi2* regulates transcript stability, directly or indirectly. This potential role sets *Msi2* apart from *Msi1* and may lead to new findings in Msi biology. Thus, it will be critical to explore the mechanism by which *Msi2* modulates stability. One direction would be to identify protein interactors that may mediate this process and what domains in *Msi2* mediate potential interactions. Additionally, it will be interesting to determine whether *Msi2* affects transcript localization.

Another question that remains unanswered is how *Msi2* recognizes its targets. I provide evidence that it is the presence of multiple UAGs that define *Msi2* binding, but not all targets have UAGs in their CLIP peak, and how these UAGs contribute to binding is still unknown. Mutational studies and luciferase assays would be useful in determining what elements are required for *Msi2* binding. Furthermore, it would be

useful to refine and complete secondary structural predictions in these regions once the exact binding sites of *Msi2* are elucidated.

The data presented in this thesis have interesting implications into *Msi2*'s function in the skin. First, it will be critical to understand the function of *Msi2* in wound healing. Furthermore, *Msi2* may function differently in the cell populations where it is expressed. Finally, as mentioned previously, *Msi2* has known oncogenic activity. It will be interesting to know whether *Msi2* plays a role in skin cancers. Future studies using both gain- and loss-of-function mouse models would allow someone to begin to answer some of these important questions.

Overall, my study has provided new insights into *Msi2* RNA target recognition, molecular mechanisms of target repression, and cellular functions in stratified epithelial progenitors. These findings provide a foundation to examine *Msi2* functions in normal development, tissue homeostasis, wound healing and tumorigenesis.

REFERENCES

Alani, R.M., Hasskarl, J., Grace, M., Hernandez, M.C., Israel, M.A., and Münger, K. (1999). Immortalization of primary human keratinocytes by the helix-loop-helix protein, Id-1. Proc. Natl. Acad. Sci. U. S. A. *96*, 9637–9641.

Andrés-Aguayo, L. de, Varas, F., Kallin, E.M., Infante, J.F., Wurst, W., Floss, T., and Graf, T. (2011). Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice. Blood *118*, 554–564.

Arumugam, K., Macnicol, M.C., and Macnicol, A.M. (2012a). Autoregulation of Musashi1 mRNA Translation During Xenopus Oocyte Maturation. Mol. Reprod. Dev. *79*.

Arumugam, K., MacNicol, M.C., Wang, Y., Cragle, C.E., Tackett, A.J., Hardy, L.L., and MacNicol, A.M. (2012b). Ringo/Cyclin-dependent Kinase and Mitogen-activated Protein Kinase Signaling Pathways Regulate the Activity of the Cell Fate Determinant Musashi to Promote Cell Cycle Re-entry in Xenopus Oocytes. J. Biol. Chem. *287*, 10639–10649.

Arwert, E.N., Hoste, E., and Watt, F.M. (2012). Epithelial stem cells, wound healing and cancer. Nat. Rev. Cancer *12*, 170–180.

Barbouti, A., Höglund, M., Johansson, B., Lassen, C., Nilsson, P.-G., Hagemeijer, A., Mitelman, F., and Fioretos, T. (2003). A novel gene, MSI2, encoding a putative RNA-binding protein is recurrently rearranged at disease progression of chronic myeloid leukemia and forms a fusion gene with HOXA9 as a result of the cryptic t(7;17)(p15;q23). Cancer Res. *63*, 1202–1206.

Barrientos, S., Stojadinovic, O., Golinko, M.S., Brem, H., and Tomic-Canic, M. (2008). Growth factors and cytokines in wound healing. Wound Repair Regen. *16*, 585–601.

Battelli, C., Nikopoulos, G.N., Mitchell, J.G., and Verdi, J.M. (2006). The RNA-binding protein Musashi-1 regulates neural development through the translational repression of p21WAF-1. Mol. Cell. Neurosci. *31*, 85–96.

Beronja, S., Janki, P., Heller, E., Lien, W.-H., Keyes, B.E., Oshimori, N., and Fuchs, E. (2013). RNAi screens in mice identify physiological regulators of oncogenic growth. Nature *501*, 185–190.

Blanpain, C., and Fuchs, E. (2006). Epidermal stem cells of the skin. Annu. Rev. Cell Dev. Biol. *22*, 339–373.

Blanpain, C., and Fuchs, E. (2009). Epidermal homeostasis: a balancing act of stem cells in the skin. Nat. Rev. Mol. Cell Biol. *10*, 207–217.

Bracken, C.P., Li, X., Wright, J.A., Lawrence, D.M., Pillman, K.A., Salmanidis, M., Anderson, M.A., Dredge, B.K., Gregory, P.A., Tsykin, A., et al. (2014). Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion. EMBO J. *33*, 2040–2056.

Bull, J.J., Müller-Röver, S., Patel, S. V, Chronnell, C.M., McKay, I.A., and Philpott, M.P. (2001). Contrasting localization of c-Myc with other Myc superfamily transcription factors in the human hair follicle and during the hair growth cycle. J. Invest. Dermatol. *116*, 617–622.

Byrne, C., Tainsky, M., and Fuchs, E. (1994). Programming gene expression in developing epidermis. Development *120*, 2369–2383.

Case, L.B., and Waterman, C.M. (2015). Integration of actin dynamics and cell adhesion by a three-dimensional, mechanosensitive molecular clutch. Nat. Cell Biol. *17*, 955–963.

Chang, K.-H., Chan-Ling, T., McFarland, E.L., Afzal, A., Pan, H., Baxter, L.C., Shaw, L.C., Caballero, S., Sengupta, N., Li Calzi, S., et al. (2007). IGF binding protein-3 regulates hematopoietic stem cell and endothelial precursor cell function during vascular development. Proc. Natl. Acad. Sci. U. S. A. *104*, 10595–10600.

Chapnick, D.A., Jacobsen, J., and Liu, X. (2013). The Development of a Novel High Throughput Computational Tool for Studying Individual and Collective Cellular Migration. PLoS One *8*, e82444.

Charlesworth, A., Wilczynska, A., Thampi, P., Cox, L.L., and MacNicol, A.M. (2006). Musashi regulates the temporal order of mRNA translation during Xenopus oocyte maturation. EMBO J. *25*, 2792–2801.

Chen, C.-Y.A., Ezzeddine, N., and Shyu, A.-B. (2008). Messenger RNA half-life measurements in mammalian cells. Methods Enzymol. *448*, 335–357.

Chen, K., Gao, Q., Zhang, W., Liu, Z., Cai, J., Liu, Y., Xu, J., Li, J., Yang, Y., and Xu, X. (2014). MSI1 regulates survival of hepatoma cell lines by activation of Wnt signaling pathway. Liver Int. n/a – n/a.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. Nature *460*, 479–486.

Choudhury, N.R., de Lima Alves, F., de Andrés-Aguayo, L., Graf, T., Cáceres, J.F., Rappsilber, J., and Michlewski, G. (2013). Tissue-specific control of brain-enriched miR-7 biogenesis. Genes Dev. *27*, 24–38.

Clancy, S., and Brown, W. (2008). Translation: DNA to mRNA to Protein.

Clingman, C.C., Deveau, L.M., Hay, S.A., Genga, R.M., Shandilya, S.M., Massi, F., and Ryder, S.P. (2014). Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. Elife *3*, e02848.

Cotsarelis, G., Sun, T.-T., and Lavker, R.M. (1990). Label-retaining cells reside in the bulge area of pilosebaceous unit: Implications for follicular stem cells, hair cycle, and skin carcinogenesis. Cell *61*, 1329–1337.

Cox, J.L., Wilder, P.J., Gilmore, J.M., Wuebben, E.L., Washburn, M.P., and Rizzino, A. (2013). The SOX2-Interactome in Brain Cancer Cells Identifies the Requirement of MSI2 and USP9X for the Growth of Brain Tumor Cells. PLoS One *8*.

Darnell, R. (2012). CLIP (cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. Cold Spring Harb. Protoc. *2012*, 1146–1160.

Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., and Xu, R.-M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev. *13*, 1102–1115.

Dobson, N.R., Zhou, Y.-X., Flint, N.C., and Armstrong, R.C. (2008). Musashi1 RNA-binding protein regulates oligodendrocyte lineage cell differentiation and survival. Glia *56*, 318–330.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.-H., Ghoshal, K., Villén, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol. Cell *56*, 104–115.

Eliseeva, I.A., Lyabin, D.N., and Ovchinnikov, L.P. (2013). Poly(A)-binding proteins: structure, domain organization, and activity regulation. Biochem. Biokhimiia *78*, 1377–1391.

ErLin, S., WenJie, W., LiNing, W., BingXin, L., MingDe, L., Yan, S., and RuiFa, H. (2015). Musashi-1 maintains blood-testis barrier structure during spermatogenesis and regulates stress granule formation upon heat stress. Mol. Biol. Cell mbc.E14–11 – 1497 – .

Fan, H., and Khavari, P.A. (1999). Sonic hedgehog opposes epithelial cell cycle arrest. J. Cell Biol. *147*, 71–76.

Fraser, B., Sobinoff, A., Hime, G., Siddall, N., Roman, S., and McLaughlin, E. (2008). The Role of the RNA-Binding Protein, Musashi-1, in Murine Spermatogonial Stem Cell Maintenance. Biol. Reprod. *78*.

Frye, M., Gardner, C., Li, E.R., Arnold, I., and Watt, F.M. (2003). Evidence that Myc activation depletes the epidermal stem cell compartment by modulating adhesive interactions with the local microenvironment. Development *130*, 2793–2808.

Gao, C., Han, C., Yu, Q., Guan, Y., Li, N., Zhou, J., Tian, Y., and Zhang, Y. Downregulation of Msi1 suppresses the growth of human colon cancer by targeting p21cip1. Int. J. Oncol.

Gebauer, F., and Hentze, M.W. (2004). Molecular mechanisms of translational control. Nat. Rev. Mol. Cell Biol. *5*, 827–835.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding

proteins. Nat. Rev. Genet. *15*, 829–845.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. *582*, 1977–1986.

Goshima, Y., Ito, T., Sasaki, Y., and Nakamura, F. (2002). Semaphorins as signals for cell repulsion and invasion. J. Clin. Invest. *109*, 993–998.

Gunter, K.M., and McLaughlin, E.A. (2011). Translational control in germ cell development: A role for the RNA-binding proteins Musashi-1 and Musashi-2. IUBMB Life *63*, 678–685.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Hadziselimovic, N., Vukojevic, V., Peter, F., Milnik, A., Fastenrath, M., Fenyves, B.G., Hieber, P., Demougin, P., Vogler, C., de Quervain, D.J.-F., et al. (2014). Forgetting Is Regulated via Musashi-Mediated Translational Control of the Arp2/3 Complex. Cell *156*, 1153–1166.

Han, Y., Ye, A., Zhang, Y., Cai, Z., Wang, W., Sun, L., Jiang, S., Wu, J., Yu, K., and Zhang, S. (2015). Musashi-2 Silencing Exerts Potent Activity against Acute Myeloid Leukemia and Enhances Chemosensitivity to Daunorubicin. PLoS One *10*, e0136484.

Heng, M.C.Y. (2011). Wound healing in adult skin: aiming for perfect regeneration. Int. J. Dermatol. *50*, 1058–1066.

Horzum, U., Ozdil, B., and Pesen-Okvur, D. (2014). Step-by-step quantitative analysis of focal adhesions. MethodsX *1*, 56–59.

Huelsken, J., Vogel, R., Erdmann, B., Cotsarelis, G., and Birchmeier, W. (2001). beta-Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. Cell *105*, 533–545.

Huntzinger, E., Kuzuoğlu-Öztürk, D., Braun, J.E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2012). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res.

Huttenlocher, A., and Horwitz, A.R. (2011). Integrins in cell migration. Cold Spring Harb. Perspect. Biol. *3*, a005074.

Iatsenko, I., Sinha, A., Rödelsperger, C., and Sommer, R.J. (2013). A new role for DCR-1/DICER in C. elegans innate immunity against the highly virulent bacterium Bacillus thuringiensis DB27. Infect. Immun.

Imai, T., Tokunaga, A., Yoshida, T., Hashimoto, M., Mikoshiba, K., Weinmaster, G., Nakafuku, M., and Okano, H. (2001). The neural RNA-binding protein Musashi1

translationally regulates mammalian numb gene expression by interacting with its mRNA. Mol. Cell. Biol. *21*, 3888–3900.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell *147*, 789–802.

Ito, T., Kwon, H.Y., Zimdahl, B., Congdon, K.L., Blum, J., Lento, W.E., Zhao, C., Lagoo, A., Gerrard, G., Foroni, L., et al. (2010). Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature *466*, 765–768.

Jackson, S.J., Zhang, Z., Feng, D., Flagg, M., O'Loughlin, E., Wang, D., Stokes, N., Fuchs, E., and Yi, R. (2013a). Rapid and widespread suppression of self-renewal by microRNA-203 during epidermal differentiation. Development *140*, 1882–1891.

Jackson, S.J., Zhang, Z., Feng, D., Flagg, M., O'Loughlin, E., Wang, D., Stokes, N., Fuchs, E., and Yi, R. (2013b). Rapid and widespread suppression of self-renewal by microRNA-203 during epidermal differentiation. Development *140*, 1882–1891.

Ji Eun Choi, Jun Sang Bae, Ju Hyung Lee, Kyu Yun Jang, Myoung Ja Chung, W.S.M. (2014). Musashi-1 expression and clinicopathological significance in young gastric cancer patients: A matched case-control study. Int. J. Oncol. *44*, 1185–1192.

Jia, J., Yao, P., Arif, A., and Fox, P.L. (2013). Regulation and dysregulation of 3'UTR-mediated translational control. Curr. Opin. Genet. Dev. *23*, 29–34.

Kaeda, J., Ringel, F., Oberender, C., Mills, K., Quintarelli, C., Pane, F., Koschmieder, S., Slany, R., Schwarzer, R., Saglio, G., et al. (2014). Upregulated MSI2 is associated with more aggressive leukemia. Leuk. Lymphoma 1–28.

Katz, Y., Li, F., Lambert, N.J., Sokol, E.S., Tam, W.L., Cheng, A.W., Airoldi, E.M., Lengner, C.J., Gupta, P.B., Yu, Z., et al. (2014). Musashi proteins are post-transcriptional regulators of the epithelial-luminal cell state. Elife *3*, e03915.

Kawahara, H., Imai, T., Imataka, H., Tsujimoto, M., Matsumoto, K., and Okano, H. (2008). Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. J. Cell Biol. *181*, 639–653.

Kawasaki, H., and Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature *431*, 211–217.

Kawase, S., Kuwako, K., Imai, T., Renault-Mihara, F., Yaguchi, K., Itohara, S., and Okano, H. (2014). Rfx Transcription Factors Control Musashi1 Transcription in Mouse Neural Stem/Progenitor Cells. Stem Cells Dev.

Kerner, P., Degnan, S.M., Marchand, L., Degnan, B.M., and Vervoort, M. (2011). Evolution of RNA-binding proteins in animals: insights from genome-wide analysis in the sponge Amphimedon queenslandica. Mol. Biol. Evol. *28*, 2289–2303.

Kharas, M.G., Lengner, C.J., Al-Shahrour, F., Bullinger, L., Ball, B., Zaidi, S., Morgan, K., Tam, W., Paktinat, M., Okabe, R., et al. (2010). Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nat. Med. *16*, 903–908.

Kim, D.H., Villeneuve, L.M., Morris, K. V, and Rossi, J.J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat. Struct. Mol. Biol. *13*, 793–797.

Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat. Methods *8*, 559–564.

Kong, J., and Lasko, P. (2012). Translational control in cellular and developmental processes. Nat. Rev. Genet. *13*, 383–394.

König, J., Zarnack, K., Luscombe, N.M., and Ule, J. (2012). Protein–RNA interactions: new genomic technologies and perspectives. Nat. Rev. Genet. *13*, 77–83.

Koster, M.I., and Roop, D.R. (2007). Mechanisms regulating epithelial stratification. Annu. Rev. Cell Dev. Biol. *23*, 93–113.

Kuang, R.-G., Kuang, Y., Luo, Q.-F., Zhou, C.-J., Ji, R., and Wang, J.-W. (2013). Expression and significance of Musashi-1 in gastric cancer and precancerous lesions. World J. Gastroenterol. *19*, 6637–6644.

Kuo, P.-L., Chen, Y.-H., Chen, T.-C., Shen, K.-H., and Hsu, Y.-L. (2011). CXCL5/ENA78 increased cell migration and epithelial-to-mesenchymal transition of hormone-independent prostate cancer by early growth response-1/snail signaling pathway. J. Cell. Physiol. *226*, 1224–1231.

Kuwako, K., Kakumoto, K., Imai, T., Igarashi, M., Hamakubo, T., Sakakibara, S., Tessier-Lavigne, M., Okano, H.J., and Okano, H. (2010). Neural RNA-binding protein Musashi1 controls midline crossing of precerebellar neurons through posttranscriptional regulation of Robo3/Rig-1 expression. Neuron *67*, 407–421.

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. Mol. Cell.

Lechler, T., and Fuchs, E. (2005). Asymmetric cell divisions promote stratification and

differentiation of mammalian skin. Nature *437*, 275–280.

Li, N., Yousefi, M., Nakauka-Ddamba, A., Li, F., Vandivier, L., Parada, K., Woo, D.-H., Wang, S., Naqvi, A.S., Rao, S., et al. (2015). The Msi Family of RNA-Binding Proteins Function Redundantly as Intestinal Oncoproteins. Cell Rep. *13*, 2440–2455.

Liefer, K.M., Koster, M.I., Wang, X.J., Yang, A., McKeon, F., and Roop, D.R. (2000). Down-regulation of p63 is required for epidermal UV-B-induced apoptosis. Cancer Res. *60*, 4016–4020.

Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G., and Rutter, W.J. (1970). Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. Science *170*, 447–449.

Liu, J., Rivas, F. V, Wohlschlegel, J., Yates, J.R. 3rd, Parker, R., and Hannon, G.J. (2005). A role for the P-body component GW182 in microRNA function. Nat. Cell Biol. *7*, 1261–1266.

Liu, X., Yang, W.-T., and Zheng, P.-S. (2014). Msi1 promotes tumor growth and cell proliferation by targeting cell cycle checkpoint proteins p21, p27 and p53 in cervical carcinomas. Oncotarget.

Lukong, K.E., Chang, K., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends Genet. *24*, 416–425.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Cáceres, J.F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. Nat. Struct. Mol. Biol. *19*, 760–766.

MacNicol, A.M., Wilczynska, A., and MacNicol, M.C. (2008). Function and regulation of the mammalian Musashi mRNA translational regulator. Biochem. Soc. Trans. *36*, 528–530.

Maris, C., Dominguez, C., and Allain, F.H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. *272*, 2118–2131.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. *14*, 447–459.

Mine, N., Iwamoto, R., and Mekada, E. (2005). HB-EGF promotes epithelial cell migration in eyelid development. Development *132*, 4317–4326.

Minuesa, G., Antczak, C., Shum, D., Radu, C., Bhinder, B., Li, Y., Djaballah, H., and Kharas, M.G. (2014). A 1536-Well Fluorescence Polarization Assay To Screen For Modulators Of The MUSASHI Family Of RNA-Binding Proteins. Comb. Chem. High

Throughput Screen.

Miyanoiri, Y., Kobayashi, H., Imai, T., Watanabe, M., Nagata, T., Uesugi, S., Okano, H., and Katahira, M. (2003). Origin of higher affinity to RNA of the N-terminal RNA-binding domain than that of the C-terminal one of a mouse neural protein, musashi1, as revealed by comparison of their structures, modes of interaction, surface electrostatic potentials, and backbone . J. Biol. Chem. *278*, 41309–41315.

Moghbeli, M., Forghanifard, M.M., Sadrizadeh, A., Mozaffari, H.M., Golmakani, E., and Abbaszadegan, M.R. (2015). Role of Msi1 and MAML1 in Regulation of Notch Signaling Pathway in Patients with Esophageal Squamous Cell Carcinoma. J. Gastrointest. Cancer.

Moore, M.A.S. (2010). A cancer fate in the hands of a samurai. Nat. Med. *16*, 963–965.

Mu, Q., Wang, Y., Chen, B., Qian, W., Meng, H., Tong, H., Chen, F., Ma, Q., Ni, W., Chen, S., et al. (2013). High expression of Musashi-2 indicates poor prognosis in adult B-cell acute lymphoblastic leukemia. Leuk. Res. *37*, 922–927.

Nagata, T., Kanno, R., Kurihara, Y., Uesugi, S., Imai, T., Sakakibara, S., Okano, H., and Katahira, M. (1999). Structure, backbone dynamics and interactions with RNA of the C-terminal RNA-binding domain of a mouse neural RNA-binding protein, Musashi1. J. Mol. Biol. *287*, 315–330.

Nakamura, M., Okano, H., Blendy, J.A., and Montell, C. (1994). Musashi, a neural RNA-binding protein required for Drosophila adult external sensory organ development. Neuron *13*, 67–81.

Nawy, T. (2014). Biochemistry: Where protein and RNAs meet. Nat. Methods *11*, 605–605.

Nikpour, P., Baygi, M.E., Steinhoff, C., Hader, C., Luca, A.C., Mowla, S.J., and Schulz, W.A. (2011). The RNA binding protein Musashi1 regulates apoptosis, gene expression and stress granule formation in urothelial carcinoma cells. J. Cell. Mol. Med. *15*, 1210–1224.

Nishimoto, Y., and Okano, H. (2010). New insight into cancer therapeutics: Induction of differentiation by regulating the Musashi/Numb/Notch pathway. Cell Res. *20*, 1083–1085.

Ohyama, T., Nagata, T., Tsuda, K., Kobayashi, N., Imai, T., Okano, H., Yamazaki, T., and Katahira, M. (2011). Structure of Musashi1 in a complex with target RNA: the role of aromatic stacking interactions. Nucleic Acids Res. *40*, 3218–3231.

Okano, H., Imai, T., and Okabe, M. (2002). Musashi: a translational regulator of cell fate. J. Cell Sci. *115*, 1355–1359.

Pare, J.M., Tahbaz, N., López-Orozco, J., LaPointe, P., Lasko, P., and Hobman, T.C.

(2009). Hsp90 Regulates the Function of Argonaute 2 and Its Recruitment to Stress Granules and P-Bodies. Mol. Biol. Cell *20*, 3273–3284.

Park, S.-M., Deering, R.P., Lu, Y., Tivnan, P., Lianoglou, S., Al-Shahrour, F., Ebert, B.L., Hacohen, N., Leslie, C., Daley, G.Q., et al. (2014). Musashi-2 controls cell fate, lineage bias, and TGF-β signaling in HSCs. J. Exp. Med. *211*, 71–87.

Park, S.-M., Gönen, M., Vu, L., Minuesa, G., Tivnan, P., Barlowe, T.S., Taggart, J., Lu, Y., Deering, R.P., Hacohen, N., et al. (2015). Musashi2 sustains the mixed-lineage leukemia-driven stem cell regulatory program. J. Clin. Invest. *125*, 1286–1298.

Peano, C., Pietrelli, A., Consolandi, C., Rossi, E., Petiti, L., Tagliabue, L., De Bellis, G., and Landini, P. (2013). An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. Microb. Inform. Exp. *3*, 1.

Quick-Cleveland, J., Jacob, J.P., Weitz, S.H., Shoffner, G., Senturia, R., and Guo, F. (2014). The DGCR8 RNA-Binding Heme Domain Recognizes Primary MicroRNAs by Clamping the Hairpin. Cell Rep. *7*, 1994–2005.

Rangarajan, A., Talora, C., Okuyama, R., Nicolas, M., Mammucari, C., Oh, H., Aster, J.C., Krishna, S., Metzger, D., Chambon, P., et al. (2001). Notch signaling is a direct determinant of keratinocyte growth arrest and entry into differentiation. EMBO J. *20*, 3427–3436.

Ravindran, G., and Devaraj, H. (2014). Prognostic significance of neural stem cell markers, Nestin and Musashi-1, in oral squamous cell carcinoma: expression pattern of Nestin in the precancerous stages of oral squamous epithelium. Clin. Oral Investig.

Rezza, A., Skah, S., Roche, C., Nadjar, J., Samarut, J., and Plateroti, M. (2010). The overexpression of the putative gut stem cell marker Musashi-1 induces tumorigenesis through Wnt and Notch activation. J. Cell Sci. *123*, 3256–3265.

Ridley, A.J., Schwartz, M.A., Burridge, K., Firtel, R.A., Ginsberg, M.H., Borisy, G., Parsons, J.T., and Horwitz, A.R. (2003). Cell migration: integrating signals from front to back. Science *302*, 1704–1709.

Riemondy, K., Wang, X., Torchia, E.C., Roop, D.R., and Yi, R. (2015). MicroRNA-203 represses selection and expansion of oncogenic Hras transformed tumor initiating cells. Elife *4*, e07004.

Sakakibara, S., Imai, T., Hamaguchi, K., Okabe, M., Aruga, J., Nakajima, K., Yasutomi, D., Nagata, T., Kurihara, Y., Uesugi, S., et al. (1996). Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. Dev. Biol. *176*, 230–242.

Sakakibara, S., Nakamura, Y., Satoh, H., and Okano, H. (2001). RNA-Binding Protein Musashi2: Developmentally Regulated Expression in Neural Precursor Cells and Subpopulations of Neurons in Mammalian CNS. J. Neurosci. *21*, 8091–8107.

Sakakibara, S., Nakamura, Y., Yoshida, T., Shibata, S., Koike, M., Takano, H., Ueda, S., Uchiyama, Y., Noda, T., and Okano, H. (2002). RNA-binding protein Musashi family: Roles for CNS stem cells and a subpopulation of ependymal cells revealed by targeted disruption and antisense ablation. Proc. Natl. Acad. Sci. *99*, 15194–15199.

Santoro, M.M., Gaudino, G., and Marchisio, P.C. (2003). The MSP receptor regulates alpha6beta4 and alpha3beta1 integrins via 14-3-3 proteins in keratinocyte migration. Dev. Cell *5*, 257–271.

Schneider, M.R., Schmidt-Ullrich, R., and Paus, R. (2009). The Hair Follicle as a Dynamic Miniorgan. Curr. Biol. *19*, R132–R142.

Sennett, R., and Rendl, M. (2012). Mesenchymal-epithelial interactions during hair follicle morphogenesis and cycling. Semin. Cell Dev. Biol. *23*, 917–927.

Siddall, N.A., McLaughlin, E.A., Marriner, N.L., and Hime, G.R. (2006). The RNA-binding protein Musashi is required intrinsically to maintain stem cell identity. Proc. Natl. Acad. Sci. U. S. A. *103*, 8402–8407.

Song, J.-J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat. Struct. Biol. *10*, 1026–1032.

STEITZ, J.A. (1969). Polypeptide Chain Initiation: Nucleotide Sequences of the Three Ribosomal Binding Sites in Bacteriophage R17 RNA. Nature *224*, 957–964.

St-Jacques, B., Dassule, H.R., Karavanova, I., Botchkarev, V.A., Li, J., Danielian, P.S., McMahon, J.A., Lewis, P.M., Paus, R., and McMahon, A.P. (1998). Sonic hedgehog signaling is essential for hair development. Curr. Biol. *8*, 1058–1069.

Streit, M., Velasco, P., Riccardi, L., Spencer, L., Brown, L.F., Janes, L., Lange-Asschenfeldt, B., Yano, K., Hawighorst, T., Iruela-Arispe, L., et al. (2000). Thrombospondin-1 suppresses wound healing and granulation tissue formation in the skin of transgenic mice. EMBO J. *19*, 3272–3282.

Sugiyama-Nakagiri, Y., Akiyama, M., Shibata, S., Okano, H., and Shimizu, H. (2006). Expression of RNA-Binding Protein Musashi in Hair Follicle Development and Hair Cycle Progression. Am. J. Pathol. *168*, 80–92.

Sutherland, J., Sobinoff, A., Gunter, K., Fraser, B., Pye, V., Bernstein, I., Boon, E., Siddall, N., De Andres, L., Hime, G., et al. (2015a). Knockout of RNA Binding Protein MSI2 Impairs Follicle Development in the Mouse Ovary: Characterization of MSI1 and MSI2 during Folliculogenesis. Biomolecules *5*, 1228–1244.

Sutherland, J.M., Siddall, N.A., Hime, G.R., and McLaughlin, E.A. (2000). RNA binding proteins in spermatogenesis: an in depth focus on the Musashi family. Asian J. Androl. *17*, 529–536.

Sutherland, J.M., McLaughlin, E.A., Hime, G.R., and Siddall, N.A. (2013). The Musashi family of RNA binding proteins: master regulators of multiple stem cell populations. Adv. Exp. Med. Biol. *786*, 233–245.

Sutherland, J.M., Fraser, B.A., Sobinoff, A.P., Pye, V.J., Davidson, T.-L., Siddall, N.A., Koopman, P., Hime, G.R., and McLaughlin, E.A. (2014). Developmental Expression of Musashi-1 and Musashi-2 RNA-Binding Proteins During Spermatogenesis: Analysis of the Deleterious Effects of Dysregulated Expression. Biol. Reprod. biolreprod.113.115261 – .

Sutherland, J.M., Sobinoff, A.P., Fraser, B.A., Redgrove, K.A., Davidson, T.-L., Siddall, N.A., Koopman, P., Hime, G.R., and McLaughlin, E.A. (2015b). RNA binding protein Musashi-1 directly targets Msi2 and Erh during early testis germ cell development and interacts with IPO5 upon translocation to the nucleus. FASEB J.

Taggart, J., Ho, T.-C., Amin, E., Xu, H., Barlowe, T.S., Perez, A.R., Durham, B.H., Tivnan, P., Okabe, R., Chow, A., et al. (2016). MSI2 is required for maintaining activated myelodysplastic syndrome stem cells. Nat. Commun. *7*, 10739.

Topisirovic, I., Svitkin, Y. V, Sonenberg, N., and Shatkin, A.J. (2011). Cap and cap-binding proteins in the control of gene expression. Wiley Interdiscip. Rev. RNA *2*, 277–298.

Uren, P.J., Vo, D.T., Rosa de Araujo, P., Pötschke, R., Burns, S.C., Bahrami-Samani, E., Qiao, M., de Sousa Abreu, R., Nakaya, H.I., Correa, B.R., et al. (2015). The RNA-binding protein Musashi1 is a central regulator of adhesion pathways in glioblastoma. Mol. Cell. Biol. *35*, 2965–2978.

Vaezi, A., Bauer, C., Vasioukhin, V., and Fuchs, E. (2002). Actin cable dynamics and Rho/Rock orchestrate a polarized cytoskeletal architecture in the early steps of assembling a stratified epithelium. Dev. Cell *3*, 367–381.

Vindry, C., Ngoc, L.V., Kruys, V., and Gueydan, C. (2014). RNA-binding protein-mediated post-transcriptional controls of gene expression: Integration of molecular mechanisms at the 3’ end of mRNAs? Biochem. Pharmacol.

Vo, D.T., Abdelmohsen, K., Martindale, J.L., Qiao, M., Tominaga, K., Burton, T.L., Gelfond, J.A.L., Brenner, A.J., Patel, V., Trageser, D., et al. (2012). The oncogenic RNA-binding protein Musashi1 is regulated by HuR via mRNA translation and stability in glioblastoma cells. Mol. Cancer Res. *10*, 143–155.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science (80-.). *297*, 1833–1837.

Wang, D., Zhang, Z., O’Loughlin, E., Lee, T., Houel, S., O’Carroll, D., Tarakhovsky, A., Ahn, N.G., and Yi, R. (2012). Quantitative functions of Argonaute proteins in mammalian development. Genes Dev. *26*, 693–704.

Wang, S., Li, N., Yousefi, M., Nakauka-Ddamba, A., Li, F., Parada, K., Rao, S., Minuesa, G., Katz, Y., Gregory, B.D., et al. (2015a). Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. Nat. Commun. *6*, 6517.

Wang, X., Xu, X., Ma, Z., Huo, Y., Xiao, Z., Li, Y., and Wang, Y. (2011). Dynamic mechanisms for pre-miRNA binding and export by Exportin-5. RNA *17*, 1511–1528.

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921–926.

Wang, Y., Jiang, C.-Q., and Fan, L.-F. (2015b). Correlation of Musashi-1, Lgr5, and pEGFR expressions in human small intestinal adenocarcinomas. Tumour Biol.

Wang, Z., Tollervey, J., Briese, M., Turner, D., and Ule, J. (2009a). CLIP: Construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. Methods *48*, 287–293.

Wang, Z., Gerstein, M., and Snyder, M. (2009b). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., 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.

Wuebben, E.L., Mallanna, S.K., Cox, J.L., and Rizzino, A. (2012). Musashi2 Is Required for the Self-Renewal and Pluripotency of Embryonic Stem Cells. PLoS One *7*.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.-M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468–474.

Ye, J., and Blelloch, R. (2014). Regulation of Pluripotency by RNA Binding Proteins. Cell Stem Cell *15*, 271–280.

Yi, R., O'Carroll, D., Pasolli, H.A., Zhang, Z., Dietrich, F.S., Tarakhovsky, A., and Fuchs, E. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. Nat Genet *38*, 356–362.

Yoda, A., Sawa, H., and Okano, H. (2000). MSI-1, a neural RNA-binding protein, is involved in male mating behaviour in Caenorhabditis elegans. Genes to Cells *5*, 885–895.

Younger, S.T., and Corey, D.R. (2011). Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. Nucleic Acids Res. *39*, 5682–5691.

Yu, M., Berk, R., and Kosir, M.A. (2010). CXCL7-Mediated Stimulation of Lymphangiogenic Factors VEGF-C, VEGF-D in Human Breast Cancer Cells. J. Oncol.

2010, 939407.

Zearfoss, N.R., Deveau, L.M., Clingman, C.C., Schmidt, E., Johnson, E.S., Massi, F., and Ryder, S.P. (2014). A Conserved Three-Nucleotide Core Motif Defines Musashi RNA-Binding Specificity. J. Biol. Chem. M114.597112 – .

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614.

Zhang, P., and Zhang, H. (2013). Autophagy modulates miRNA-mediated gene silencing and selectively degrades AIN-1/GW182 in C. elegans. EMBO Rep. *advance on*.

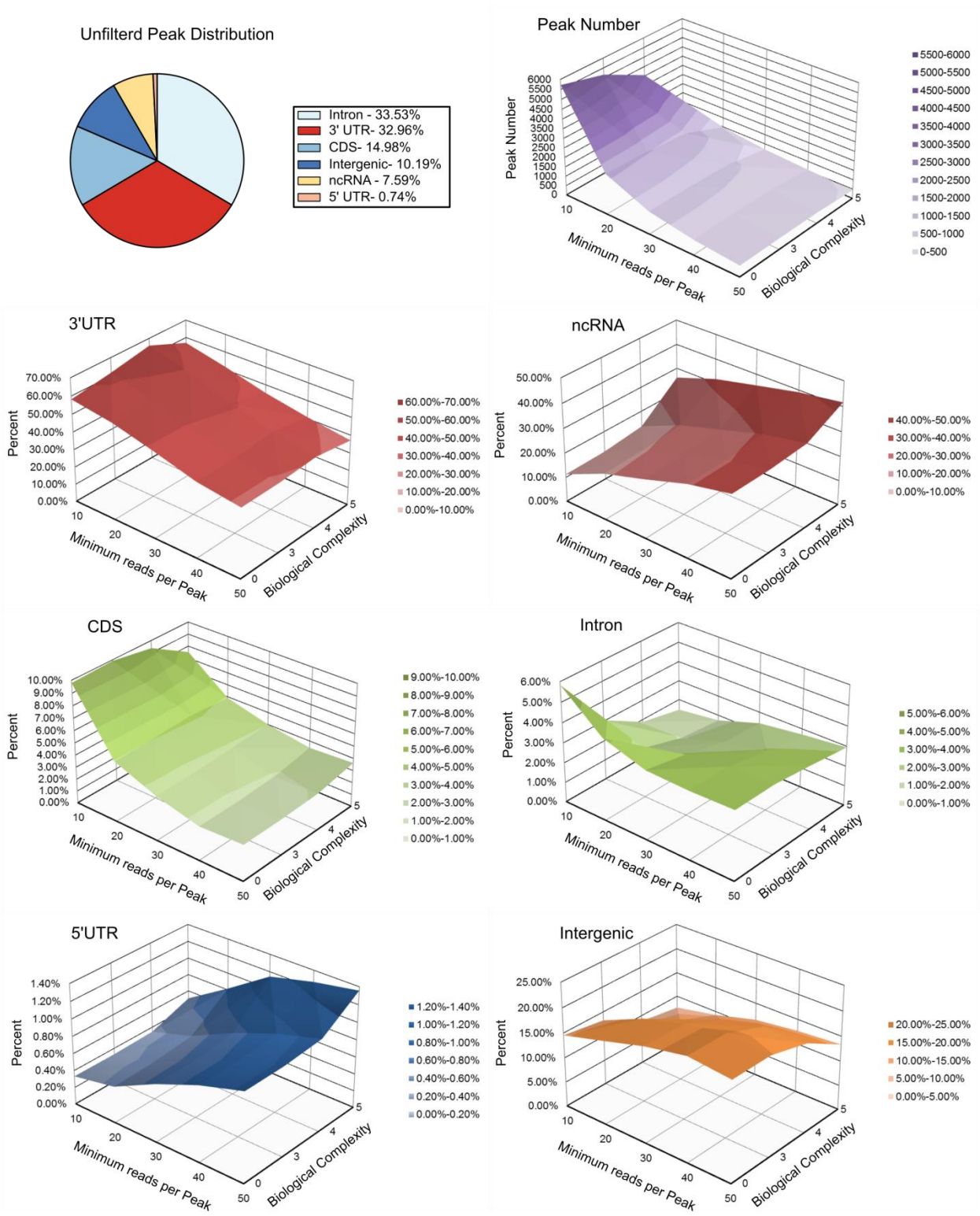
APPENDIX

Appendix 1: HITS-CLIP Mapping Statistics

Msi2 HITS-CLIP Mapping Statistics

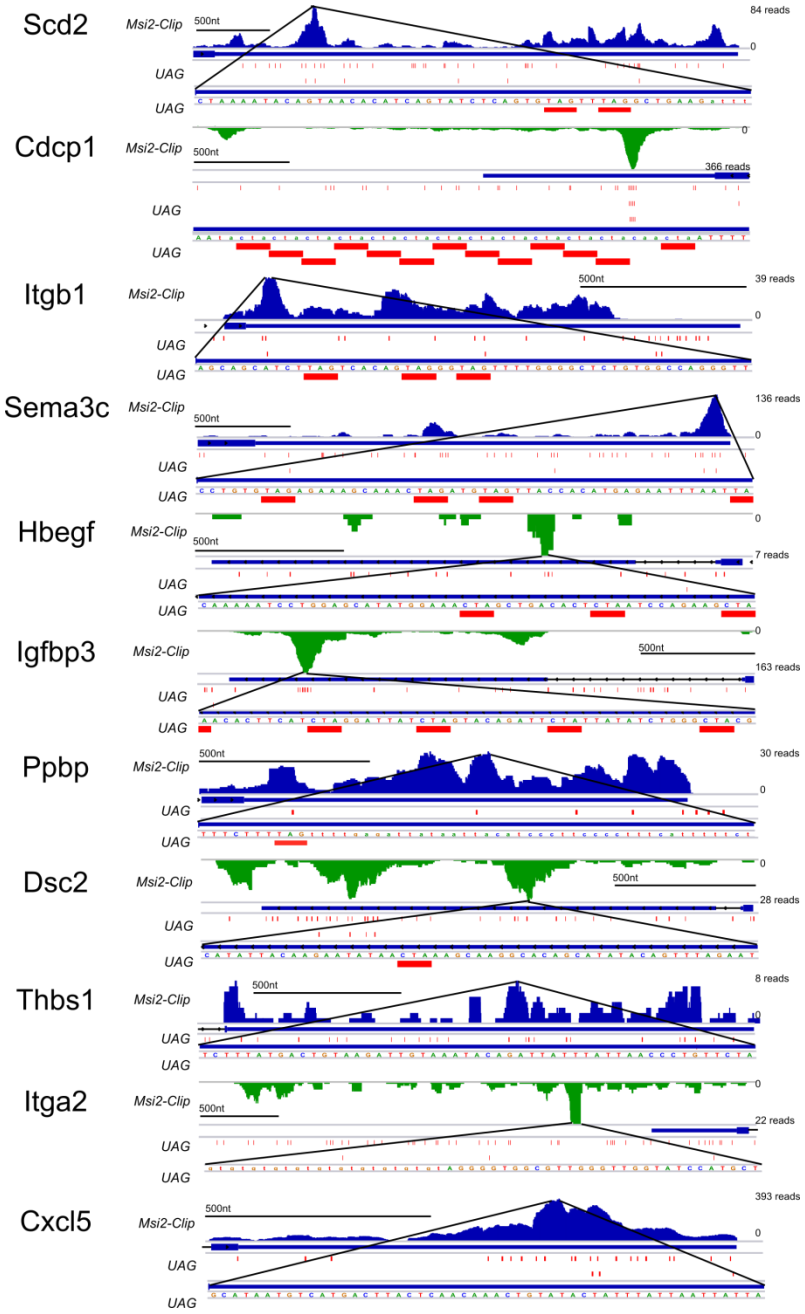
	Keratinocytes					
	rep. 1	rep. 2	rep. 3	rep. 4	rep.5	Total
Raw Reads	7,694,905	8,130,804	4,646,423	7,886,026	5,696,326	34,054,484
Trimmed and Not sequence duplicate reads	451,785	1,188,587	1,091,490	1,556,367	1,145,519	5,433,748
Reads uniquely aligned ≥ 20 nt	58,738	738,408	857,567	1,202,388	851,987	3,709,088
Reads with unique alignment positions						358,074
Unfiltered Peaks						126,250
Filtered Peaks						4,051
5'UTR Peaks						14
CDS Peaks						380
3'UTR Peaks						2,671
Intronic Peaks						76
ncRNA Peaks						548
Intergenic Peaks						359

Appendix 2: Analysis of HITS-CLIP filters



Colors in surface plot denote similar scales in y-axis and legend range.

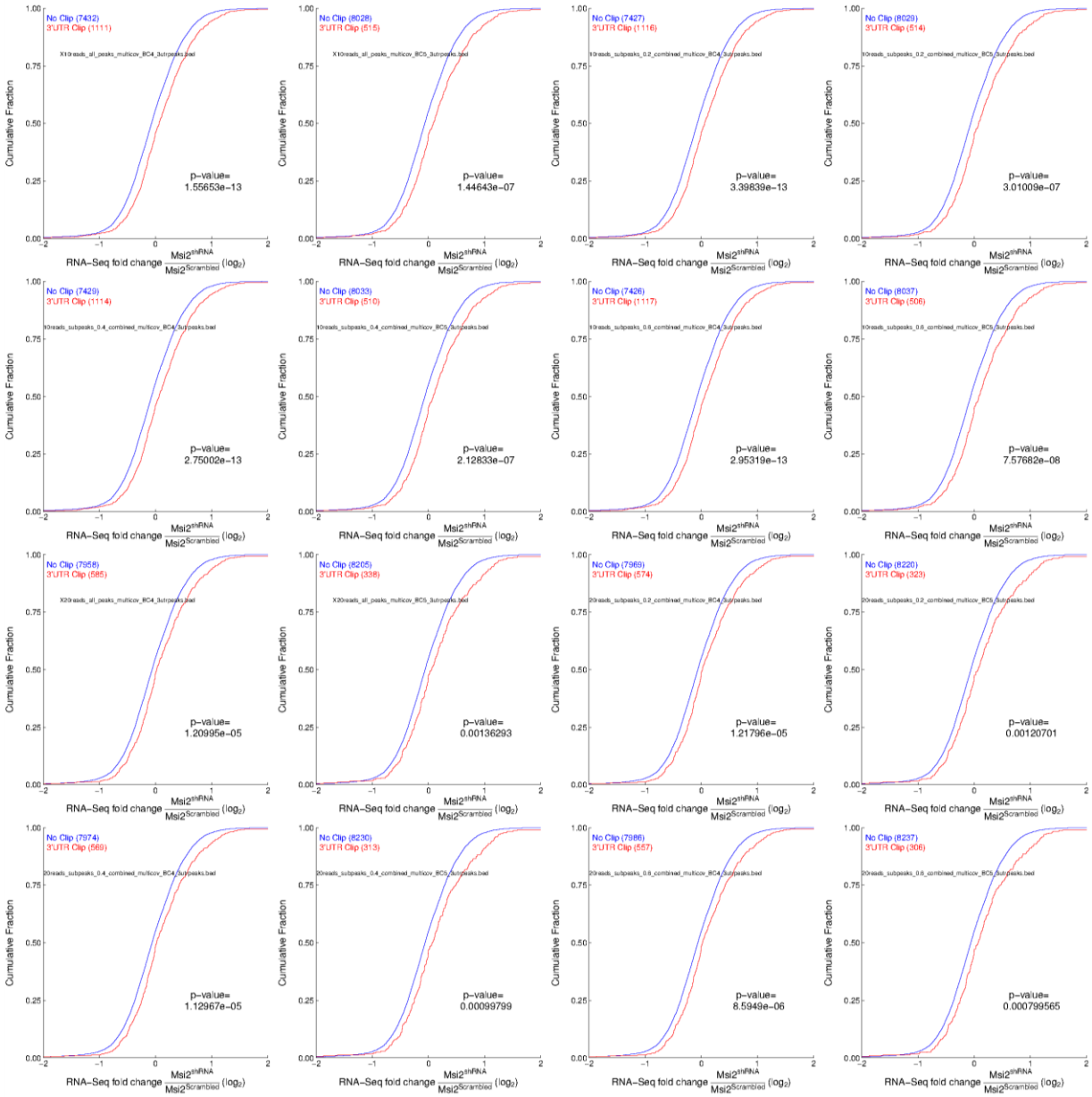
Appendix 3: Additional HITS-CLIP gene tracts

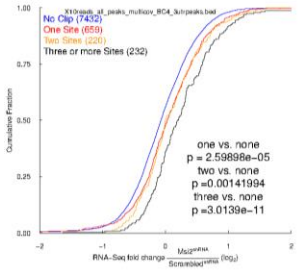


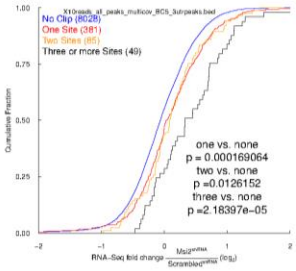
Gene tracts of a set of *Msi2* targets from HITS-CLIP data. Targets were selected as a representative set of genes and includes genes validated in this study as *Msi2* targets. Peak summit inset displayed below tracts. UAG core motifs and the reverse complement motif CUA are marked in red across the 3'UTR of targets and in the inset of the peak summits. Blue marks positive strand reads. Green marks negative strand reads. Scale bar = 500 nucleotides. Y-axis is peak height

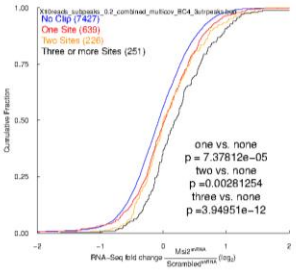
Appendix 4: Analysis of HITS-CLIP cut offs in RNA-seq and Ribo-seq data

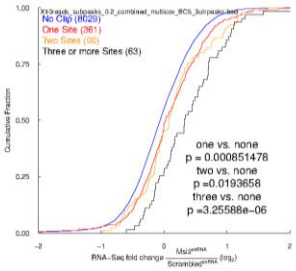
RNA-seq

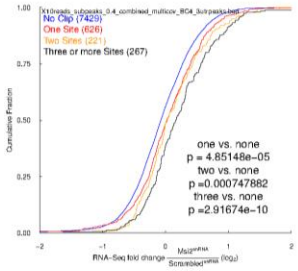


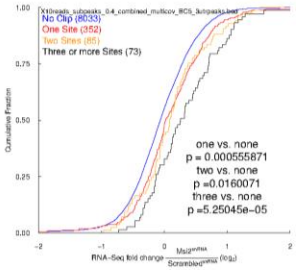

X10reads_all_peaks_multicov_BC4_3utrpeaks.bed
No Clip (7432)
One Site (659)
Two Sites (226)
Three or more Sites (232)
one vs. none
p = 2.59898e−05
two vs. none
p =0.00141994
three vs. none
p =3.0139e−11
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

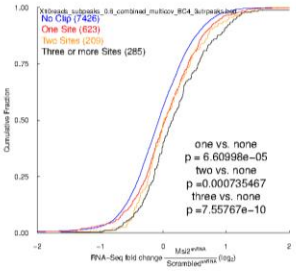

X10reads_all_peaks_multicov_BC5_3utrpeaks.bed
No Clip (8028)
One Site (381)
Two Sites (85)
Three or more Sites (49)
one vs. none
p = 0.000169064
two vs. none
p =0.0126152
three vs. none
p =2.18397e−05
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

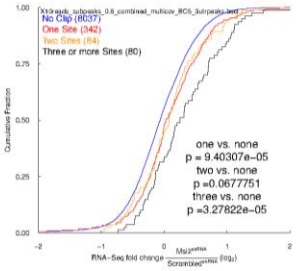

X10reads_subpeaks_0.2_combined_multicov_BC4_3utrpeaks.bed
No Clip (7427)
One Site (639)
Two Sites (226)
Three or more Sites (251)
one vs. none
p = 7.37812e−05
two vs. none
p =0.00281254
three vs. none
p =3.94951e−12
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

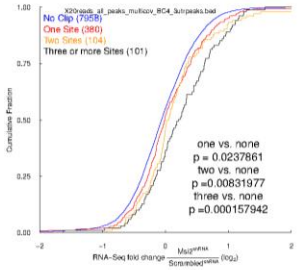

X10reads_subpeaks_0.2_combined_multicov_BC5_3utrpeaks.bed
No Clip (8029)
One Site (361)
Two Sites (80)
Three or more Sites (63)
one vs. none
p = 0.000851478
two vs. none
p =0.0193658
three vs. none
p =3.25588e−06
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

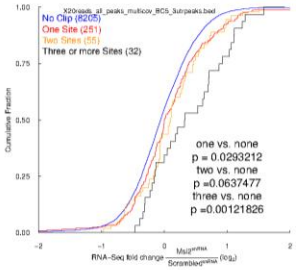

X10reads_subpeaks_0.4_combined_multicov_BC4_3utrpeaks.bed
No Clip (7429)
One Site (626)
Two Sites (221)
Three or more Sites (267)
one vs. none
p = 4.85148e−05
two vs. none
p =0.00074788
three vs. none
p =2.91674e−10
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

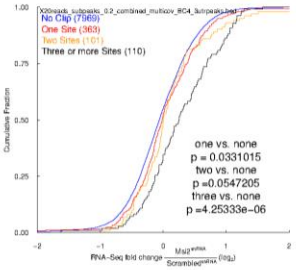

X10reads_subpeaks_0.4_combined_multicov_BC5_3utrpeaks.bed
No Clip (8033)
One Site (352)
Two Sites (85)
Three or more Sites (73)
one vs. none
p = 0.000555871
two vs. none
p =0.0160071
three vs. none
p =5.25045e−05
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

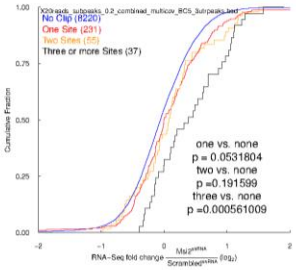

X10reads_subpeaks_0.6_combined_multicov_BC4_3utrpeaks.bed
No Clip (7426)
One Site (623)
Two Sites (209)
Three or more Sites (285)
one vs. none
p = 6.60998e−05
two vs. none
p =0.000735467
three vs. none
p =7.55767e−10
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

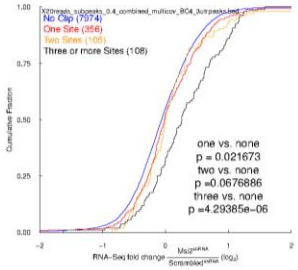

X10reads_subpeaks_0.6_combined_multicov_BC5_3utrpeaks.bed
No Clip (8037)
One Site (342)
Two Sites (84)
Three or more Sites (80)
one vs. none
p = 9.40307e−05
two vs. none
p =0.0677751
three vs. none
p =3.27822e−05
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

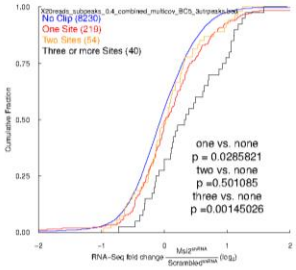

X20reads_all_peaks_multicov_BC4_3utrpeaks.bed
No Clip (7958)
One Site (380)
Two Sites (104)
Three or more Sites (101)
one vs. none
p = 0.0237861
two vs. none
p =0.00831977
three vs. none
p =0.000157942
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

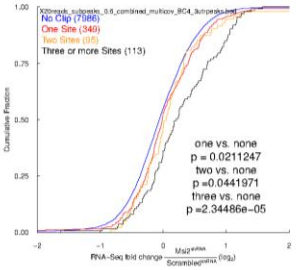

X20reads_all_peaks_multicov_BC5_3utrpeaks.bed
No Clip (8205)
One Site (251)
Two Sites (55)
Three or more Sites (32)
one vs. none
p = 0.0293212
two vs. none
p =0.0637477
three vs. none
p =0.00121826
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

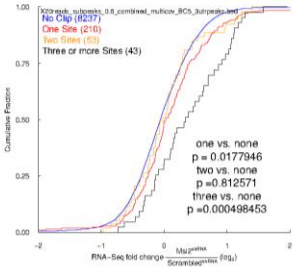

X20reads_subpeaks_0.2_combined_multicov_BC4_3utrpeaks.bed
No Clip (7969)
One Site (363)
Two Sites (101)
Three or more Sites (110)
one vs. none
p = 0.0331015
two vs. none
p =0.0547205
three vs. none
p =4.25333e−06
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)


X20reads_subpeaks_0.2_combined_multicov_BC5_3utrpeaks.bed
No Clip (8220)
One Site (231)
Two Sites (55)
Three or more Sites (37)
one vs. none
p = 0.0531804
two vs. none
p =0.191599
three vs. none
p =0.000561009
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

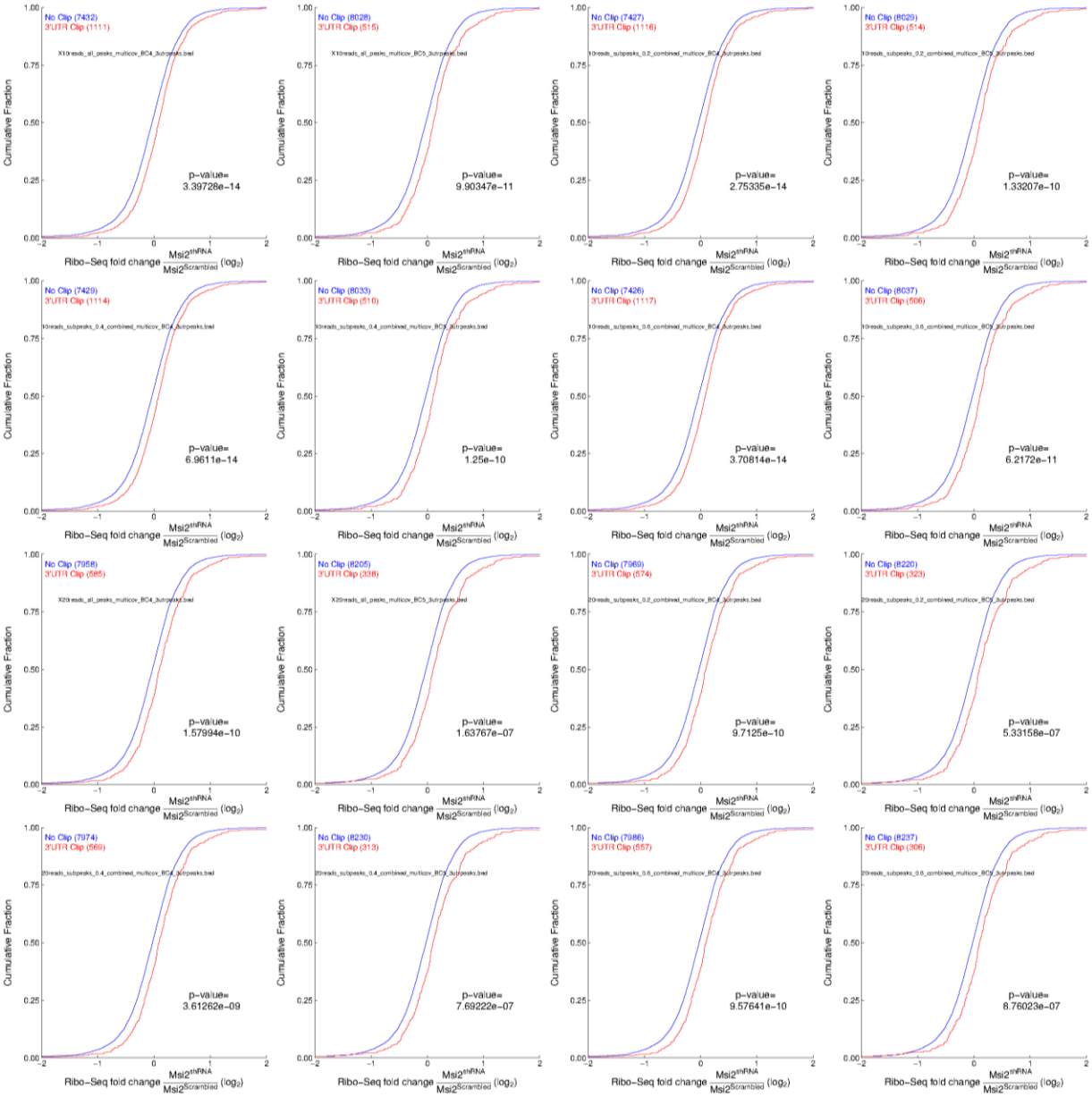

X20reads_subpeaks_0.4_combined_multicov_BC4_3utrpeaks.bed
No Clip (7974)
One Site (356)
Two Sites (105)
Three or more Sites (108)
one vs. none
p = 0.021673
two vs. none
p =0.0676886
three vs. none
p =4.29385e−06
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

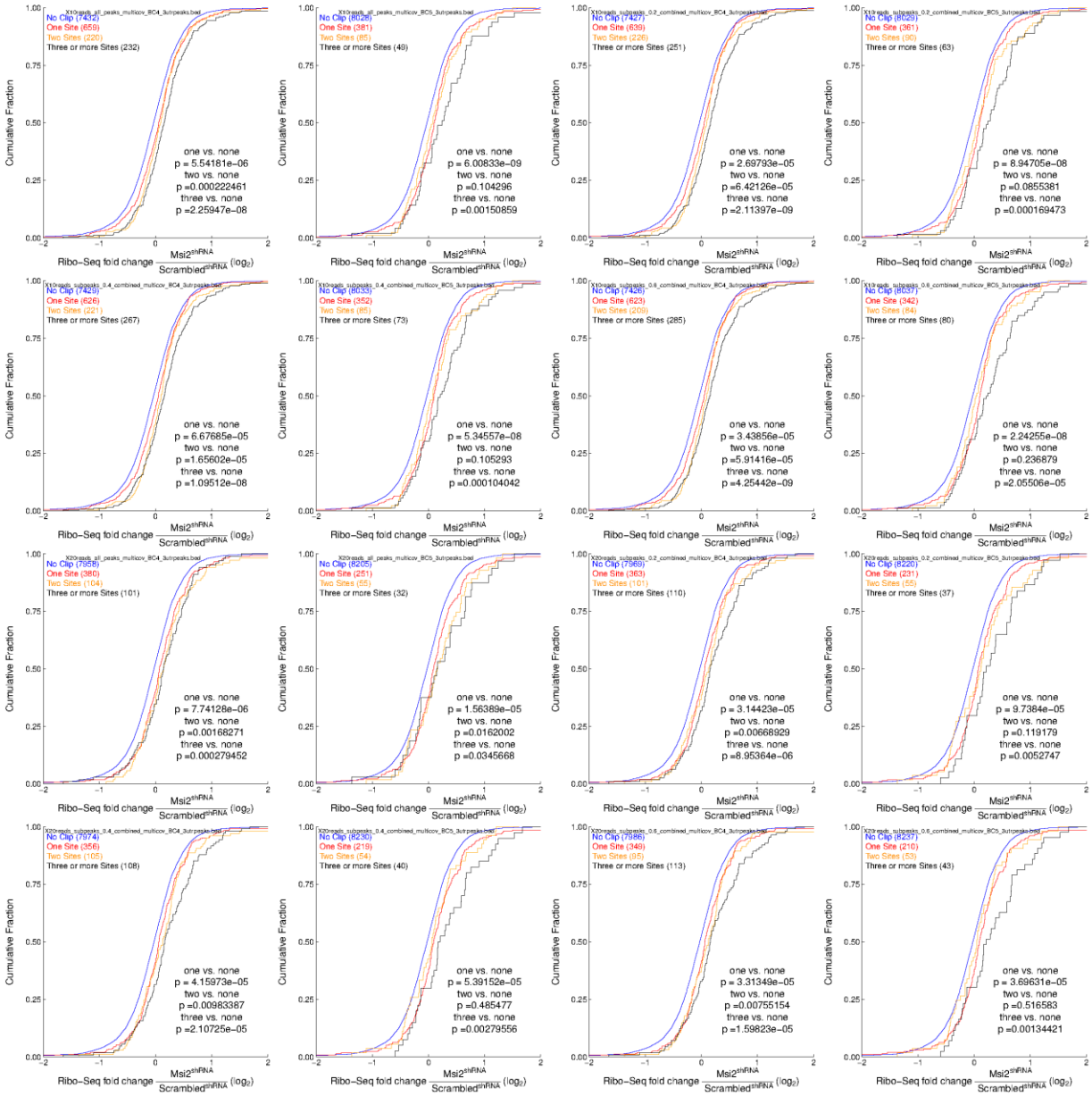

X20reads_subpeaks_0.4_combined_multicov_BC5_3utrpeaks.bed
No Clip (8230)
One Site (219)
Two Sites (54)
Three or more Sites (40)
one vs. none
p = 0.0285821
two vs. none
p =0.501085
three vs. none
p =0.00145026
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)


X20reads_subpeaks_0.6_combined_multicov_BC4_3utrpeaks.bed
No Clip (7886)
One Site (340)
Two Sites (95)
Three or more Sites (113)
one vs. none
p = 0.0211247
two vs. none
p =0.0441971
three vs. none
p =2.34486e−05
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)


X20reads_subpeaks_0.6_combined_multicov_BC5_3utrpeaks.bed
No Clip (8237)
One Site (210)
Two Sites (53)
Three or more Sites (43)
one vs. none
p = 0.0177946
two vs. none
p =0.812571
three vs. none
p =0.000498453
Cumulative Fraction
RNA−Seq fold change Msi2mRNA / ScrambledmRNA (log2)

Ribo-seq


No Clip (7432)
3'UTR Clip (1111)
p−value=
3.39728e−14
No Clip (8028)
3'UTR Clip (515)
p−value=
9.90347e−11
No Clip (7427)
3'UTR Clip (1116)
p−value=
2.75335e−14
No Clip (8029)
3'UTR Clip (514)
p−value=
1.33207e−10
No Clip (7429)
3'UTR Clip (1114)
p−value=
6.9611e−14
No Clip (8033)
3'UTR Clip (510)
p−value=
1.25e−10
No Clip (7426)
3'UTR Clip (1117)
p−value=
3.70814e−14
No Clip (8037)
3'UTR Clip (506)
p−value=
6.2172e−11
No Clip (7958)
3'UTR Clip (585)
p−value=
1.57994e−10
No Clip (8205)
3'UTR Clip (338)
p−value=
1.63767e−07
No Clip (7969)
3'UTR Clip (574)
p−value=
9.7125e−10
No Clip (8220)
3'UTR Clip (323)
p−value=
5.33158e−07
No Clip (7974)
3'UTR Clip (569)
p−value=
3.61262e−09
No Clip (8230)
3'UTR Clip (313)
p−value=
7.69222e−07
No Clip (7986)
3'UTR Clip (557)
p−value=
9.57641e−10
No Clip (8237)
3'UTR Clip (306)
p−value=
8.76023e−07
Cumulative Fraction
Ribo−Seq fold change Msi2shRNA / Msi2Scrambled (log2)


No Clip (7432)
One Site (659)
Two Sites (220)
Three or more Sites (232)
one vs. none
p = 5.54181e−06
two vs. none
p =0.000222461
three vs. none
p =2.25947e−08
No Clip (8029)
One Site (381)
Two Sites (85)
Three or more Sites (49)
one vs. none
p = 6.00833e−09
two vs. none
p =0.104296
three vs. none
p =0.00150859
No Clip (7427)
One Site (639)
Two Sites (226)
Three or more Sites (251)
one vs. none
p = 2.69793e−05
two vs. none
p =6.42126e−05
three vs. none
p =2.11397e−09
No Clip (8029)
One Site (361)
Two Sites (90)
Three or more Sites (63)
one vs. none
p = 8.94705e−08
two vs. none
p =0.0855381
three vs. none
p =0.000169473
No Clip (7429)
One Site (626)
Two Sites (221)
Three or more Sites (267)
one vs. none
p = 6.67685e−05
two vs. none
p =1.65602e−05
three vs. none
p =1.09512e−08
No Clip (8033)
One Site (352)
Two Sites (85)
Three or more Sites (73)
one vs. none
p = 5.34557e−08
two vs. none
p =0.105293
three vs. none
p =0.000104042
No Clip (7426)
One Site (623)
Two Sites (209)
Three or more Sites (285)
one vs. none
p = 3.43856e−05
two vs. none
p =5.91416e−05
three vs. none
p =4.25442e−09
No Clip (8037)
One Site (342)
Two Sites (64)
Three or more Sites (80)
one vs. none
p = 2.24255e−08
two vs. none
p =0.236879
three vs. none
p =2.05506e−05
No Clip (7958)
One Site (380)
Two Sites (104)
Three or more Sites (101)
one vs. none
p = 7.74128e−06
two vs. none
p =0.00168271
three vs. none
p =0.000279452
No Clip (8205)
One Site (251)
Two Sites (55)
Three or more Sites (32)
one vs. none
p = 1.56389e−05
two vs. none
p =0.0162002
three vs. none
p =0.0345668
No Clip (7969)
One Site (363)
Two Sites (101)
Three or more Sites (110)
one vs. none
p = 3.14423e−05
two vs. none
p =0.00668929
three vs. none
p =8.95364e−06
No Clip (8220)
One Site (231)
Two Sites (55)
Three or more Sites (37)
one vs. none
p = 9.7384e−05
two vs. none
p =0.119179
three vs. none
p =0.0052747
No Clip (7974)
One Site (356)
Two Sites (105)
Three or more Sites (108)
one vs. none
p = 4.15973e−05
two vs. none
p =0.00983387
three vs. none
p =2.10725e−05
No Clip (8236)
One Site (219)
Two Sites (54)
Three or more Sites (40)
one vs. none
p = 5.39152e−05
two vs. none
p =0.485477
three vs. none
p =0.00279556
No Clip (7986)
One Site (340)
Two Sites (95)
Three or more Sites (113)
one vs. none
p = 3.31349e−05
two vs. none
p =0.00755154
three vs. none
p =1.59823e−05
No Clip (8237)
One Site (210)
Two Sites (53)
Three or more Sites (43)
one vs. none
p = 3.69631e−05
two vs. none
p =0.516583
three vs. none
p =0.00134421
Cumulative Fraction
Ribo−Seq fold change Msi2shRNA / ScrambledshRNA (log2)

Appendix 5: RNA-seq Mapping Statistics

RNA-Seq Mapping Statistics

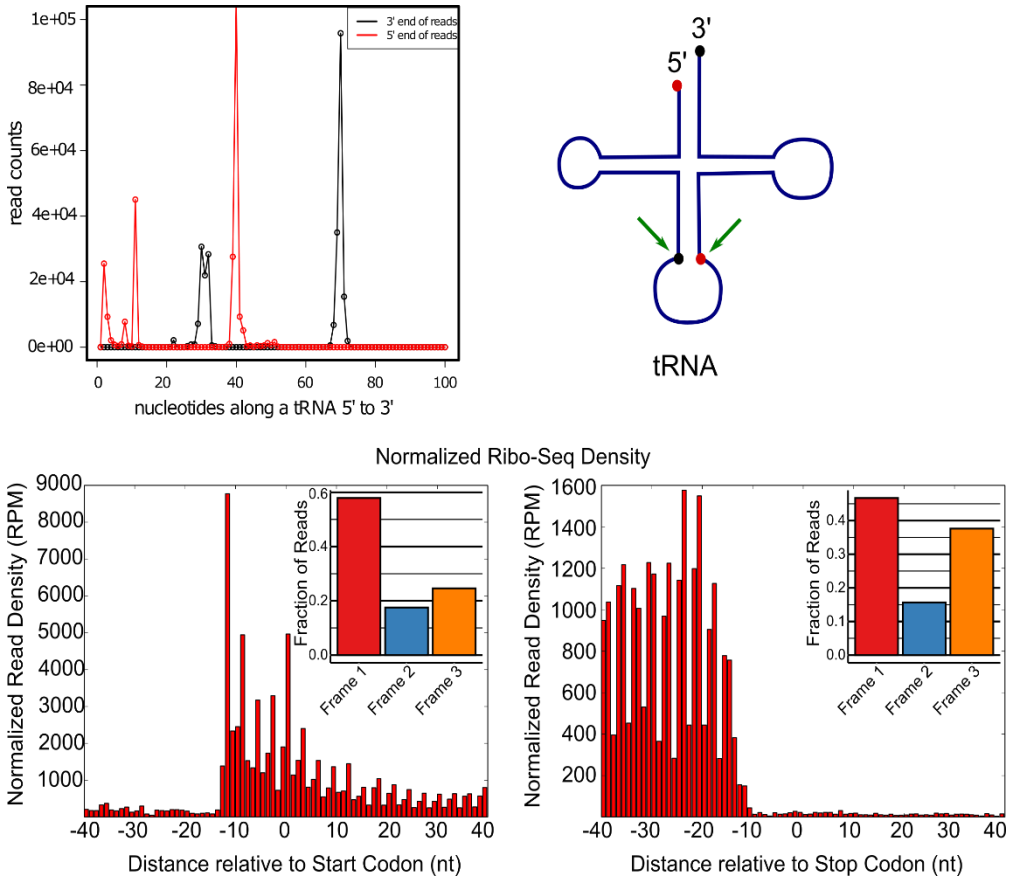
	Scrambled shRNA			Msi2 shRNA		
	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3
Raw Reads	34,915,136	46,588,226	21,659,727	18,678,045	25,786,006	56,068,489
Uniquely Mapped	31,838,134	42,016,475	19,181,729	16,317,461	21,791,747	48,595,619
Unique Alignment (%)	91.19	90.19	88.56	87.36	84.51	86.67
Unique Alignment to exons (%)	84.67%	83.11%	79.97%	76.77%	71.76%	75.37%
Unique Alignment to CDS (%)	56.93%	53.52%	52.13%	49.37%	46.95%	47.77%

Appendix 6: Ribo-seq Mapping Statistics

Ribo-Seq Mapping Statistics

	Scrambled shRNA			Msi2 shRNA		
	rep. 1	rep. 2	rep. 3	rep. 1	rep. 2	rep. 3
Raw Reads	21,330,185	30,971,905	14,418,386	38,448,389	21,734,079	25,236,826
Reads Remaining after Trimming	20,782,378	29,549,952	14,306,681	38,320,475	21,692,332	25,150,111
Reads Mapped to rRNA	19,010,247	27,374,135	13,943,666	37,382,044	21,237,040	24,360,146
Reads Mapped to tRNA	12,178,359	20,926,648	10,688,405	28,808,082	17,371,276	19,633,843
Reads Mapped to ncRNA	7,436,835	14,298,292	4,431,590	12,183,993	7,644,116	7,199,619
Reads Uniquely Mapped to CDS	2,196,405	2,959,625	702,602	2,069,440	512,256	998,028
Unique CDS Alignment (%)	10.30%	9.56%	4.87%	5.38%	2.36%	3.95%

Appendix 7: Ribosome profiling read phasing and tRNA contamination.



Metagene across tRNA showing 5' and 3' alignment of the reads contaminating Ribo-seq library (top left). Diagram of where RNase cleavage likely occurred (green arrows) to generate contaminating reads (top right). Metagene of Ribo-seq reads density across the start codon or termination codon for coding regions (bottom). Reads overlap was only assessed for the 5' nucleotide of the read. 29 nucleotide long reads from all Ribo-seq libraries were used to assess phasing. Bar chart displays fraction of reads with 5' nucleotide aligning to frame 1, frame 2, or frame 3 of the gene coding sequence displayed.

Appendix 8: List of High Confidence Targets

Gene id	Log2 Rnaseq	Log2 Riboseq	Log2 TE change	Final rank	HITS-CLIP reads	
Itga2	2.57	2.60	1.02	1	115	log2FC
Flnb	2.52	2.62	1.07	2	42	4
Sptbn1	2.11	2.08	0.98	3	87	3
Thbs1	2.43	2.02	0.76	4	83	2
Akap12	1.81	1.91	1.07	5	12	1
Plec	1.45	2.03	1.49	6	33	0
Specc1	1.61	1.88	1.20	7	29	-1
Vcan	2.10	1.79	0.81	8	10	-2
Flna	1.42	1.51	1.06	9	29	-3
Cpa4	1.51	1.32	0.88	10	82	-4
Pglyrp3	1.44	1.45	1.01	11	22	
Cd200	1.40	1.49	1.07	12	119	
Ptgs2	1.58	1.24	0.79	13	233	
Ahnak	1.24	1.53	1.23	14	165	
Ppbp	1.17	1.68	1.42	15	180	
Lama3	1.38	1.21	0.89	16	146	
Rab32	1.13	1.87	1.67	17	11	
Iqgap1	1.19	1.31	1.09	18	33	
Phldb2	1.24	1.20	0.97	19	217	
Dll1	1.33	1.15	0.88	20	13	
Fst	1.17	1.19	1.02	21	288	
Sema3c	1.28	1.08	0.87	22	337	
Plod2	1.26	1.11	0.90	23	13	
Cxcl3	1.02	1.34	1.24	24	92	
Igfbp3	1.06	1.20	1.10	25	320	
Il1a	1.33	0.97	0.78	26	20	
Rnf145	1.13	1.12	0.99	27	29	
Ctgf	1.23	1.04	0.88	28	26	
Sptan1	1.02	1.19	1.12	29	59	
Utrn	1.07	1.12	1.04	30	21	
Hnrnph1	0.93	1.28	1.27	31	15	
Notch2	1.33	0.80	0.69	32	12	
Serpinb2	1.12	0.95	0.89	33	121	
Macf1	0.74	1.60	1.81	34	24	
Tmsb4x	0.93	1.18	1.18	35	28	

Gprc5a	0.99	1.04	1.03	36	33	
Cxcl1	1.02	0.98	0.98	37	139	
Ccbe1	1.26	0.72	0.69	38	21	
Hbegf	0.77	1.20	1.34	39	12	
Tes	0.60	1.31	1.64	40	60	
Akap13	1.15	0.76	0.76	41	10	
Lamb3	0.94	0.95	1.01	42	37	
Tfrc	0.98	0.90	0.95	43	303	
Met	0.96	0.91	0.96	44	128	
Ccng1	0.98	0.87	0.93	45	97	
Fat1	0.94	0.92	0.99	46	70	
Svil	0.78	1.11	1.26	47	15	
Slc12a2	1.03	0.76	0.83	48	332	
Atm	0.88	0.95	1.05	49	11	
Flrt2	1.11	0.66	0.73	50	157	
Axl	0.72	1.07	1.28	51	82	
Cxcl5	0.79	1.03	1.18	52	564	
Slc4a7	1.11	0.64	0.72	53	93	
Mdm2	0.81	0.96	1.11	54	37	
Itgav	1.10	0.64	0.72	55	330	
Pmaip1	0.69	1.07	1.30	56	53	
Son	0.71	1.05	1.26	57	44	
Tmem176b	0.55	1.18	1.55	58	11	
Txnrd1	0.87	0.84	0.98	59	32	
Fbln2	0.87	0.82	0.97	60	10	
Cenpf	0.67	1.04	1.29	61	63	
Jag1	1.16	0.53	0.64	62	105	
Prpf8	0.82	0.80	0.98	63	21	
Ranbp2	0.77	0.88	1.08	64	72	
Sf3b3	0.85	0.78	0.95	65	14	
Plaur	0.67	0.91	1.18	66	24	
Ctnnd1	0.96	0.58	0.77	67	181	
Wnk1	0.90	0.62	0.82	68	169	
Dhx9	0.94	0.57	0.77	69	154	
Atp13a3	0.71	0.78	1.05	70	253	
Usp34	0.60	0.79	1.14	71	13	
Itga3	0.85	0.60	0.84	72	74	
Dock9	0.82	0.61	0.87	73	13	
Pabpc1	0.85	0.56	0.82	74	29	
Pxdn	0.80	0.60	0.87	75	57	

Kpnb1	0.54	0.79	1.18	76	43	
Col4a1	0.59	0.67	1.06	77	34	
Myh9	0.56	0.67	1.08	78	235	
Lima1	0.50	0.70	1.15	79	126	
Il24	0.58	0.63	1.04	80	104	
Lamc2	0.54	0.64	1.08	81	120	
Itgb1	0.39	0.67	1.22	82	170	
Emp1	0.49	0.59	1.07	83	464	
Itgb4	0.56	0.55	0.99	84	35	
Myof	0.34	0.57	1.18	85	52	
Mcam	0.46	0.54	1.05	86	11	
Prl8a9	0.40	0.53	1.09	87	25	
Actn1	0.43	0.52	1.07	88	41	
Dmkn	-0.55	-0.54	1.01	89	38	
Aars	-0.68	-0.47	1.16	90	10	
Slc2a1	-0.61	-0.54	1.05	91	22	
Csnk1a1	-0.71	-0.53	1.13	92	163	
Ppp1cb	-0.55	-0.66	0.93	93	161	
Perp	-0.73	-0.54	1.14	94	561	
Nt5e	-0.64	-0.75	0.93	95	104	
Rab11a	-0.91	-0.64	1.21	96	21	
Fgfbp1	-0.71	-0.87	0.90	97	28	
Gipc1	-1.07	-0.62	1.36	98	12	
Smarcd2	-0.81	-0.88	0.95	99	46	
Krt16	-1.02	-0.70	1.25	100	13	
Sqstm1	-0.94	-0.88	1.04	101	62	
Atf4	-0.59	-1.25	0.63	102	28	
Gsta4	-1.31	-0.65	1.58	103	15	
Dsc2	-0.97	-1.10	0.91	104	325	
Prdx5	-1.15	-0.89	1.20	105	14	
H1f0	-0.92	-1.15	0.85	106	11	
Ly6a	-1.40	-0.84	1.48	107	23	
Clic4	-0.99	-1.15	0.89	108	22	
Adi1	-1.27	-0.99	1.21	109	12	
Psat1	-1.32	-0.99	1.26	110	11	
Slc7a5	-1.36	-1.10	1.19	111	16	
Pof1b	-1.05	-1.33	0.83	112	10	
Dbi	-1.36	-1.14	1.17	113	16	
Spink5	-1.63	-1.65	0.98	114	187	
Ly6d	-1.87	-1.55	1.24	115	34	

Ly6g6c	-2.43	-2.04	1.31	116	10	
Lgals7	-2.90	-2.45	1.36	117	59	
Stfa3	-3.10	-2.52	1.49	118	49	
BC100530	-4.33	-4.66	0.80	119	34	

Appendix 9: Primer and Adapter Sequences

Oligo name	Sequence	Usage
RP1	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA	library PCR
RT Primer	GCCTTGGCACCCGAGAATTCCA	Reverse Transcription and library PCR
5' adaptor	GUUCAGAGUUCUACAGUCCGACGAUCNN	linker ligation (Note RNA)
3' Adaptor	NNTGGAATTCTCGGGTGCCAAGG	linker ligation (5' preadenylated)
qpcr-ppbp-F	CTCAGACCTACATCGTCCTGC	qPCR
qpcr-ppbp-R	AGCGCAACAAGGATCAGGC	qPCR
qpcrCxcl5-F	TGCGTTGTGTTTGCTTAACCG	qPCR
qpcrCxcl5-R	CTTCCACCGTAGGGCACTG	qPCR
qpcrSema3cF	ATGGCATTCCGGGCGATTT	qPCR
qpcrSema3cR	GGTTTTGGTTTCTCGAAGCTCA	qPCR
qpcrItgb1-F	TGGTCAGCAACGCATATCTGG	qPCR
qpcrItgb1-R	GATCCACAAACCGCAACCT	qPCR
qpcrItga2-F	CGATACACATAACCCTCAGCTC	qPCR
qpcrItga2-R	CTGCCTATGATAACCCCTGTC	qPCR
qpcrIgfbp3-F	TCTAAGCGGGAGACAGAATACG	qPCR
qpcrIgfbp3-R	CTCTGGGACTCAGCACATTGA	qPCR
qpcrThbs1-F	CATCCAGAGCATCTTCACCAG	qPCR
qpcrThbs1-R	CAGCCTTTGTTCCTGAGAATG	qPCR
qpcrHbegf-F	CGGGGAGTGCAGATACCTG	qPCR
qpcrHbegf-R	TTCTCCACTGGTAGAGTCAGC	qPCR
qpcrSrc-F	GAACCCGAGAGGGACCTTC	qPCR
qpcrSrc-R	GAGGCAGTAGGCACCTTTTGT	qPCR

Appendix 10: Antibody Usage

Primary Antibody	Usage	Dilutions	Cat. Number	Company
K5	IF	1:2000	Sig-3475	Covance
Msi2	IF	1:200	ab76148	Abcam
Msi2	WB	1:1000	ab76148	Abcam
Msi2	IP/CLIP	5µg	ab76148	Abcam
β-tubulin	WB	1:5000 in 5% BSA	2146	Cell Signaling
β Actin	WB	1:5000	3700	Cell Signaling
β4-integrin	IF	1:2000	553745	BD Biosciences
Cdh1	IF	1:200	NA	gift from E. Fuchs
Vinculin	IF	1:200	V9131	Sigma-Aldrich
Vinculin	WB	1:1000	V9131	Sigma-Aldrich
Phalloidin	NA	1:50	A22287	ThermoFisher

Appendix 11: shRNA sequences

Name	shRNA number	sequence	Targeted region
PLKO-shScrCtrl	SHC002	scrambled sequence	Scrambled
PLKO-sh*Msi2* 3'UTR	*msi2* shRNA TRCN0000071973	CCGGCCCAGCTTAATATCT AGTTAACTCGAGTTAACTAG ATATTAAGCTGGGTTTTTG	3UTR
PLKO-sh*Msi2* CDS	*msi2* shRNA TRCN0000071974	CCGGCCCAACTTTGTGGCA ACCTATCTCGAGATAGGTT GCCACAAAGTTGGGTTTTT G	CDS
PLKO-Cxcl5	TRCN0000331476	CCGGTCCCAAATTGATCGC TAATTTCTCGAGAAATTAGC GATCAATTTGGGATTTTTG	CDS
PLKO-Itgb1	TRCN0000313028	CCGGTGTTGACAGTTTCCA ATTAAACTCGAGTTTAATTG GAAACTGTCAACATTTTTG	3UTR
PLKO-Sema3c	TRCN0000067391	CCGGCGATGCTCTTTCAAC CCGAATCTCGAGATTCGGG TTGAAAGAGCATCGTTTTTG	CDS
PLKO-Igfbp3	TRCN0000287884	CCGGGCCAAGATGGATGTC ATCAAACTCGAGTTTGATGA CATCCATCTTGGCTTTTTG	CDS