A POTENTIAL ROLE FOR LONG NON-CODING RNAs IN REGULATING ESCAPE DOMAIN BOUNDARIES ON THE INACTIVE X CHROMOSOME

A Thesis in
Biomedical Sciences
by
Heather Marie Young

© 2015 Heather Marie Young

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2015
The thesis of Heather Marie Young was reviewed and approved\* by the following:

Laura Carrel
Associate Professor of Biochemistry and Molecular Biology
Thesis Advisor

Sarah Bronson
Associate Professor of Cellular and Molecular Physiology

Ralph L. Keil
Associate Professor of Biochemistry and Molecular Biology
Chair of Biomedical Sciences Graduate Program

Gregory S. Yochum
Assistant Professor of Biochemistry and Molecular Biology

\*Signatures on file in the Graduate School
Abstract

In mammalian females, dosage compensation occurs by transcriptionally silencing one X chromosome. While most genes are silenced some genes, such as Kdm5c, escape inactivation. At least four escapees in mouse are flanked by long, non-coding RNAs (lncRNAs) that also escape X inactivation, raising the possibility that lncRNAs may be involved in regulating escape domains on the inactive X. X-linked bacterial artificial chromosome (BAC) transgene approaches have shown that Kdm5c ectopically escapes inactivation, suggesting that sequences contained within the BAC contribute to Kdm5c escape. Truncated transgenes lacking sequences distal of Kdm5c, including the lncRNA escapee AK148627 adjacent to Kdm5c, leads to an escape domain expansion that extends into normally inactivated loci. These results indicate that sequences distal to Kdm5c are necessary for X inactivation boundary regulation. Since lncRNAs can be cis-acting regulators within the genome, our goal is to determine whether the lncRNA juxtaposed to Kdm5c has a functional role in regulating escape, specifically in silencing adjacent genes.

AK148627 maps 2.6 kb downstream from Kdm5c and is transcribed in the same orientation. The close juxtaposition prompted us to test whether they are two distinct transcripts. While robust AK148627 transcription can be detected from both undifferentiated (prior to X-chromosome inactivation (pre-XCI)) and somatic (post-XCI) cells, transcription is not apparent immediately upstream of the annotated transcription start. Transcription extending from annotated Kdm5c exons to AK148627 is not detectable. Sequences upstream of the annotated lncRNA transcription start site show weak promoter activity by luciferase evaluation. Altogether, these data argue that AK148627 is indeed a separate transcript from Kdm5c.

To evaluate a possible cis-regulatory function for AK148627, a luciferase reporter vector containing AK148627 and its regulatory sequences was used. Inclusion of AK148627 sequences decreased luciferase reporter activity suggesting that expression of the lncRNA from these sequences represses adjacent gene expression. Additionally, vectors containing a 5’ truncated lncRNA lacking the transcription start site did not reduce luciferase levels, nor could AK148627 transcripts be detected. Interestingly, AK148627 is expressed from the luciferase vector in HEK293 cells in which AK148627 is not conserved in the human genome. This suggests that the RNA transcript itself is important for the suppression seen. Importantly, AK148627 is expressed pre-XCI at levels that are roughly three times higher than post-XCI levels in mouse ES cells supporting a regulatory role at critical time points during inactivation. These results suggest that the RNA transcript is expressed at a biologically relevant time point in pre-XCI cells and are consistent with an X-chromosome inactivation regulatory role as predicted by previous transgene studies that require the lncRNA to silence adjacent genes.
# Table of Contents

List of Figures..............................................................................vi  
List of Tables.............................................................................vii  
List of Abbreviations.................................................................viii  
Acknowledgements....................................................................ix  

## Chapter 1: Literature Review

1. X-chromosome inactivation.....................................................1  
   1.1 Murine imprinted and random inactivation.............................1  
   1.1.2 X-inactivation center (XIC)..............................................1  
   1.1.3 Counting X chromosomes and random choice.................2  
   1.1.4 Xist spreading.........................................................2  
   1.1.5 Maintenance of the inactive X.......................................3  
1.2 Genes that escape inactivation............................................3  
   1.2.1 Differences in escape vs. inactivated genes....................4  
   1.2.2 Models for escape gene mechanisms.............................5  
1.3 lncRNA characterization....................................................5  
   1.3.1 Regulatory lncRNAs....................................................6  
   1.3.2 Ddx3x locus..............................................................6  
1.4 Kdm5c escape domain.......................................................7  

## Chapter 2: Materials and Methods

2. Cell lines and tissue culture conditions..................................10  
   2.1.1 Mouse embryonic stem cell culture conditions and differentiation...10  
   2.1.2 HEK293 cells..............................................................10  
   2.1.3 Mouse tissues.............................................................10  
2.2 Plasmid DNA preparation and transfection..........................10  
   2.2.1 Plasmid construction..................................................10  
   2.2.2 Bacterial transformation..............................................10  
   2.2.3 Plasmid mini-prep purification......................................11  
2.3 Genomic AK148627 transcriptional characterization............11  
   2.3.1 Transcriptional cDNA analysis.....................................11  
   2.3.2 Determining AK148627 promoter activity......................11  
2.4 Functional role assessment of AK148627..........................12  
   2.4.1 Repressor/enhancer evaluation....................................12  
   2.4.2 Expression from luciferase vector constructs...................12  
   2.4.3 Evaluation of AK148627 regulatory DNA sequences and RNA transcript...12  
   2.4.4 Levels of endogenous AK148627 expression during X inactivation....12  
2.5 CRISPR knockout at endogenous AK148627 locus.............13  
   2.5.1 Targeting genomic DNA for CRISPR guide RNA design....13  
   2.5.2 CRISPR vector cloning..............................................13  
   2.5.3 Bacterial transformation...........................................13  
   2.5.4 Plasmid mini-prep purification....................................13  

## Chapter 3: Results

3. AK148627 is a separate transcript from Kdm5c.....................16  
   3.1.1 Transcription termination sequences............................16  
   3.1.2 Transcription orientation..........................................17  
   3.1.3 AK148627 promoter activity.....................................18  
   3.2 AK148627 suppresses luciferase expression.......................20  
   3.2.1 AK148627 expression from luciferase vector..................20  
   3.2.2 AK148627 reduces luciferase expression......................21  
   3.3 AK148627 expression in pre- and post-XCI cells................21  

## Chapter 4: Discussion

4. Model for AK148627 regulation of gene expression on the X chromosome....25
4.2 CRISPR knockout to delete endogenous AK148627

References
List of Figures

Figure 1: X inactivation is controlled by genes and IncRNAs within the X-inactivation center……2

Figure 2: Chromosome-wide epigenetic modifications lead to transcriptional silencing on the inactive X chromosome………………………………………………………………………3

Figure 3: There are differences between silent and escape domains………………………………4

Figure 4: Poor conservation of IncRNA at Ddx3x locus and escape domain expansion………6

Figure 5: Delimiting sequences necessary for transgenic Kdm5c escape…………………………8

Figure 6: Transcription downstream of the Kdm5c locus…………………………………………17

Figure 7: AK148627 transcription is not due to read-through from Kdm5c………………………18

Figure 8: Sequences adjacent to AK148627 function as a weak promoter element……………19

Figure 9: AK148627 is expressed from basic luciferase vector……………………………………20

Figure 10: AK148627 represses luciferase expression………………………………………………21

Figure 11: AK148627 is more highly expressed prior to the onset of X inactivation……………22

Figure 12: AK148627 expression is lower than Kdm5c levels during X inactivation……………23

Figure 13: CRISPR knockout of AK148627…………………………………………………………27
List of Tables

Table 1: Primers.................................................................................................14
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor</td>
</tr>
<tr>
<td>Ddx3x</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 3 gene</td>
</tr>
<tr>
<td>ES</td>
<td>Embryonic stem (cells)</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescent in situ hybridization</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>Histone 3 lysine 4 dimethylation</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Histone 3 lysine 4 trimethylation</td>
</tr>
<tr>
<td>H3K9me2</td>
<td>Histone 3 lysine 9 dimethylation</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>Histone 3 lysine 9 trimethylation</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Histone 3 lysine 27 trimethylation</td>
</tr>
<tr>
<td>Kdm5c</td>
<td>Lysine (K)-specific demethylase 5C gene</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobase (1000 bases)</td>
</tr>
<tr>
<td>IncRNA</td>
<td>Long non-coding RNA</td>
</tr>
<tr>
<td>LTR</td>
<td>Long terminal repeats</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase (10^6 bases)</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb repressive complex 2</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>SATB1</td>
<td>Special AT-rich sequence binding protein 1</td>
</tr>
<tr>
<td>SINE</td>
<td>Short interspersed element</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>Tsix</td>
<td>Inactive X specific transcript, antisense gene</td>
</tr>
<tr>
<td>Xa</td>
<td>Active X chromosome</td>
</tr>
<tr>
<td>XCI</td>
<td>X-chromosome inactivation</td>
</tr>
<tr>
<td>Xi</td>
<td>Inactive X chromosome</td>
</tr>
<tr>
<td>XIC</td>
<td>X-inactivation center</td>
</tr>
<tr>
<td>Xist</td>
<td>X-inactivation specific transcript</td>
</tr>
</tbody>
</table>
Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Dr. Laura Carrel for the continuous support during my thesis research. Her patience, motivation, enthusiasm, and immense knowledge has helped to mold my brain into a mature scientist. I could not have imagined a mentor who would have helped me progress as much as Laura.

I would also like to thank my thesis committee members Drs. Sarah Bronson, Ralph Keil and Gregory Yochum for their encouragement, insightful comments and questions that helped my research project excel. Their dedication to my education is something for which I am grateful.

Dr. Lindsay Horvath is a former member of the Carrel lab who made me feel welcome as a rotation student and was someone who would eventually take me under her wing to teach me technical skills and how to begin to think logically in a different mindset in regards to science. Her patience and understanding was greatly appreciated. Sherri Rennoll is a graduate student in Dr. Greg Yochum’s lab whom has helped me throughout my graduate career. She has taught me not only technique, but also valuable troubleshooting skills as well as how to think critically about data. Her dedication to science and to helping others is outstanding. I cannot thank her enough for all her time. My friends and colleagues at Hershey have given me a spectacular support system and I am very blessed to be part of a team that is so compassionate about science and is always there in times of need whether that be personal or for a casual science conversation.

My undergraduate professors helped to make my graduate school dream a reality. I am grateful that I had the opportunity to perform undergraduate research and get my roots planted in science and research. Thank you for helping me realize my potential.

Lastly, I would like to thank my parents who not only encouraged me to dream big, but made me believe that I could do anything I set my mind to. I want to thank my dad for those nights where I struggled with homework and he helped me every step of the way to make sure that I continued to excel at school. I want to thank my mom for being a support system and never letting me get down on myself because I had always done the best I could. I could not have asked for better parents and I am very grateful for their continued love and support no matter where my path in life leads.
CHAPTER 1
LITERATURE REVIEW
1.1 X-chromosome inactivation

Mammalian females have two X chromosomes while males have one X and one Y chromosome. The two X chromosomes in females are distinguishable cytogenetically as one appears similar to an autosome, while the other appears as a highly condensed Barr body near the nuclear membrane (Ohno et al. 1959, Barr et al. 1949). In 1961, Mary Lyon observed a mottled coat pattern in female mice heterozygous for X-linked coat color that was not present in males. She hypothesized that females randomly inactivate one X chromosome to account for the coat differences (Lyon 1962). Today we know that female eutherian mammals undergo X-chromosome inactivation (XCI) to equalize X-linked expression levels to that of males. The goal of X inactivation is to transcriptionally silence one X chromosome (Xi), while the other X (Xa) continues to be expressed.

1.1.1 Murine imprinted and random XCI

In mice, there are two forms of X inactivation that occur, in which the first stage is imprinted X inactivation. The pre-implantation two-to-four-cell stage embryo undergoes imprinted inactivation. Here, the paternal X chromosome is inactivated while the maternal X chromosome remains active (Takagi & Sasaki 1975, West et al. 1977). Cells of the trophoectoderm and primitive endoderm keep the paternal X inactivated, while cells that contribute to the inner cell mass reactivate the paternal X chromosome (Mak et al. 2004, Okamoto et al. 2004). Cells that are part of the inner cell mass undergo random inactivation around the time of implantation after imprinted inactivation has occurred. During this process, each cell inactivates either the maternal X chromosome or the paternal X chromosome (Takagi et al. 1982, Mak et al. 2004). Regardless of the choice, the inactive X is clonally maintained through subsequent cellular divisions and results in a general mosaic in which a subset of cells will have the maternal X chromosome inactivated and a subset will have the paternal X chromosome inactivated (Lyon 1962).

1.1.2 X-inactivation center (XIC)

The X-inactivation center (XIC) is a region of the X chromosome that is both necessary and sufficient to induce silencing (Brockdorff et al. 1991, Lee et al. 1999, Lee et al. 1996, Heard et al. 1999, Migeon et al. 1999). The XIC is roughly 700 kb in size and contains both positive and negative regulators of inactivation (Figure 1). The most notable regulator of the entire X inactivation process is the IncRNA Xist. Xist is monoallelically expressed from the inactive X and will ultimately spread across the chromosome from which it is transcribed (in cis) to silence most genes (Penny et al. 1996, Brockdorff et al. 1991, Brown et al. 1991). Xist is positively regulated by other IncRNAs within the XIC. The IncRNA Jpx is upstream of Xist and enhances Xist expression while another IncRNA, Ftx, prevents Xist CpG island methylation (Tian et al. 2010, Froberg et al. 2013, Chureau et al. 2011). Tsix is a RNA transcript that is transcribed antisense to Xist and is upregulated on the active X (Lee et al. 1999). Tsix expression is regulated by enhancers such as Xite and Linx (Ogawa et al. 2003, Nora et al. 2012).


Figure 1. X inactivation is controlled by genes and lncRNAs within the X-inactivation center. A cartoon depicting the XIC is shown. Tsix is downregulated on the inactive X and therefore is shown as a red arrow. Positive regulators of Tsix are also shown in red. Xist, which is upregulated on the inactive X, is shown as a green arrow along with all the positive regulators of Xist. The X-pairing region (Xpr) is shown as a gray bar. Other arrows are indicative of additional genes in that region.

1.1.3 Counting X chromosomes and random choice

For X inactivation to occur, the right number of X chromosomes must be present. For mice, a general rule of “n-1” is followed to determine whether inactivation will occur. X chromosomes are found within close proximity to one another within the nucleus at the onset of inactivation and this localization is dependent upon the X-pairing region (Xpr) believed to be located within the XIC as the precise location of this region is unknown (Bacher et al. 2006, Augui et al. 2007). Alternative mechanisms (besides the Xpr) for how this process is mediated have been proposed, including genes such as Rlim that act in a dosage-dependent manner to initiate imprinted inactivation (Jonkers et al. 2009).

The X chromosome targeted for inactivation is marked by upregulated expression of Xist (Nesterova et al. 2003). Theoretically, maternal and paternal X chromosomes are inactivated at a 50/50 ratio. Mutations within the XIC can skew this ratio to preferentially inactivate one chromosome (Percec et al. 2002). It is unclear how monoallelic upregulation of both Tsix and Xist occurs, however it is hypothesized that separation after pairing can lead to monoallelic expression of both Tsix and Xist through positive regulators in cis (Masui et al. 2011). Prior to X inactivation, Tsix and Xist are expressed at low levels from both X chromosomes (Okamoto et al. 2011). Unlike Xist, Tsix will not be persistently upregulated after the active X choice has been made (Lee et al. 1999).

1.1.4 Xist spreading

The underlying mechanisms for how Xist coats the entire X chromosome are not fully understood. It is believed that as Xist is transcribed it is tethered to the locus by the YY1 protein, which dually interacts with both the Xist RNA and genomic DNA (Jeon et al. 2011). Other proteins, such as hnRNPU and SATB1, are also believed to help tether Xist to the inactive X (Hasegawa et al. 2010, Agrelo et al. 2009) (Figure 2). There are two main hypotheses driving the field regarding how Xist may spread. The first hypothesis involves a topological association in which Xist exploits three-dimensional chromosome structure and spreads to domains within close proximity to the XIC (Engreitz et al. 2013). Alternatively, another hypothesis proposes Xist
spreads to gene rich regions based on sequence elements such as long-interspersed elements (LINEs) before spreading to adjacent gene poor regions (Simon et al. 2013).

Figure 2. Chromosome-wide epigenetic modifications lead to transcriptional silencing on the inactive X chromosome. As Xist is upregulated on the future inactive X, it is tethered to its transcriptional locus by YY1 (Jeon 2011). As it spreads to coat the chromosome, other anchoring proteins such as SATB1 and hnRNPU are recruited (Hasegawa 2010, Agrelo 2009). It is believed that the Xist transcript recruits PRC2 to modify histone residues to transcriptionally silence the chromosome in cis.

1.1.5 Maintenance of the inactive X

Xist accumulation creates a nuclear compartment in which transcription factors, such as RNA polymerase II (RNA pol II), are excluded to prevent gene transcription (Chaumeil et al. 2006). Additionally, Xist can interact with polycomb recruiting complex 2 (PRC2) to epigenetically modify the inactive X via the Xist RepA domain, as a deletion of RepA abolishes repressive heterochromatic marks on the inactive X (Zao et al. 2008). X inactivation depletes euchromatic histone modifications including histone 3 lysine 4 di- and trimethylation (H3K4me2, H3K4me3) prior to X inactivation and this process leads to a deposition of histones which are enriched for repressive heterochromatic marks such as histone 3 lysine 27 trimethylation and histone 3 lysine 9 methylation (H3K27me3, H3K9me) after gene silencing (Heard et al. 2001, Okamoto et al. 2002, O’Neill et al. 2008, Chadwick et al. 2004). Enrichment for both PRC2 and H3K27me3 are not sufficient for silencing and PRC2 enrichment is not required for post X inactivated cells to maintain their silent state (Plath et al. 2003, Silva et al. 2003). The role of Xist in X inactivation demonstrates how a lncRNA can regulate gene expression in cis through recruitment of protein complexes containing histone-modifying enzymes.

1.2 Genes that escape inactivation

While the ultimate goal of X inactivation is to transcriptionally silence one X chromosome, some genes escape inactivation. Females who inherit only one X chromosome have Turner syndrome, which is characterized by underdeveloped secondary sex characteristics (Zinn et al. 1993). Females that have an additional X chromosome display a mild
phenotype that includes cognitive delays and sex organ abnormalities (Tartaglia et al. 2010). It can be surmised that while only one X chromosome is completely active, genes that escape inactivation from the inactive X are critical for normal female development. Escape genes on the inactive X are expressed at lower levels in comparison to their expression levels on the active X chromosome (e.g. (Yang et al. 2010)). The variability of escape can be tissue specific or species specific. For example, roughly 15% of genes on the human inactive X chromosome escape inactivation while approximately 3% of genes on the mouse inactive X escape (Carrel and Willard 2005). Escape gene organization also differs between species. The 15% of human escape genes on the inactive X are clustered together while the 3% of escape genes on the mouse inactive X escape are unclustered or dispersed (Carrel and Willard 2005, Yang et al. 2010). Human escape clusters are not evenly distributed across the chromosome, but rather seem to be influenced by X-chromosome evolution (Carrel and Willard 2005). This suggests that escape genes have not yet acquired sequences necessary to silence expression on the inactive X (Jegalian et al. 1998). An alternative hypothesis is that escape genes have Y homologues so dosage compensation may not be necessary to achieve equal expression between the sexes (Carrel and Willard 2005, Yang et al. 2010).

1.2.1 Differences in escape vs. inactivated genes

The epigenetics of escape and silenced genes vary (Figure 3). Genes silenced during inactivation have histone modifications associated with repressed gene transcription, such as H3K27me3 and H3K9 methylation (Boggs et al. 2002, Brinkman et al. 2006, Calabrese et al. 2012). In contrast, escape genes have modifications associated with active transcription as well as euchromatic marks, such as H3K4me3, at the promoter (Rougeulle et al. 2003, Khalil et al. 2007, Pinter et al. 2012). Additionally, Xist accumulation is reduced at escape genes by as much as 50% (Engreitz et al. 2013). Escape genes also tend to lie more exterior on the inactive X than silenced genes, which may facilitate the long-range interactions (i.e. two regions of the chromosome within close spatial proximity allowing for co-regulation) that are seen between escapees (Clemson et al. 2006, Splinter et al. 2011). Replication of escape genes also occurs earlier than replication of silenced genes (Hansen et al. 1996).

Figure 3. There are differences between silent and escape domains. Characteristics of silent and escape genes on the inactive X are shown. Red and grey circles are heterochromatic histone modifications while green and blue circles are euchromatic histone modifications. White circles are representative of an unmethylated promoter, while black circles are representative of a methylated promoter.
There are differences in element composition between genes that escape inactivation and genes that are silenced. Long interspersed nuclear elements (LINEs) are highly enriched on the X chromosome (Korenberg et al. 1988, Boyle et al. 1990) but are depleted near escape genes in humans, suggesting they may be important for gene silencing. In mouse, there is little evidence to support this is the case (Carrel et al. 2006, Wang et al. 2006, Ke et al. 2003). It has been hypothesized that LINEs serve as a way-station to anchor Xist and facilitate spread and silencing (Gartler et al. 1983, Lyon 1998). Other sequences such as long terminal repeats (LTRs) and short interspersed elements (SINEs) have been suggested to regulate escape, but there is little evidence for a functional role for these sequences. While LTRs are depleted at escape genes in mouse, SINEs have been shown to be increased at escape genes (Tsuchiya et al. 2004, Wang et al. 2006).

1.2.2 Models for escape gene mechanisms

Based on the altered distribution of LINEs, SINEs, and LTRs, it has been hypothesized that such elements may act as of a way-station that facilitates the spread of Xist over DNA regions (Gartler et al. 1983, Lyon et al. 1998). Genomic architecture has also been proposed as recent studies have shown the importance of inactive X conformation at the onset of inactivation (Engreitz et al. 2013). This hypothesis is hard to evaluate as topological domains are not conserved between cells and escape expression levels are highly variable from tissue to tissue (Splinter et al. 2011, Talebizadeh et al. 2006, Calabrese et al. 2012, Reinious et al. 2012). The insulator hypothesis is another theory within the field. Here, an insulator protein, such as the zinc-finger protein-coding CCCTC-binding factor gene (CTCF), flanks each side of the escape gene to create an exterior pocketed region of the chromosome that is more accessible to transcription factors and RNA pol II (Berletch, Yang et al. 2011). CTCF frequently binds to boundaries throughout the genome (Cuddapah, Jothi et al. 2009) and this factor also mediates the formation of chromatin loop structures (Murrell et al. 2004, Zhao et al. 2008). The few female mouse ES cell datasets available to analyze escape loci limits the information that can be gained to either support or refute this hypothesis. Lastly, the role of long non-coding RNAs (lncRNAs) in regulating escape gene boundaries has been hypothesized and this thesis will focus specifically on this aspect.

1.3 lncRNA characterization

In mammals, lncRNAs are expressed roughly 10-fold less than protein-coding genes (Guttman et al. 2010, Ravasi et al. 2006). Since lncRNAs are expressed at such low levels in comparison to protein-coding genes, it is possible that some lncRNAs are non-functional and simply reflect transcriptional noise in the genome (Struhl et al. 2007). It is likely that while some lncRNAs are randomly transcribed and are noise, other lncRNA have been shown to regulate expression of specific target genes (see section 1.3.1).

In general, lncRNAs are transcripts longer than 200 nucleotides that lack open-reading frames, are poorly conserved among mammals, are associated with actively transcribed histone modifications, are characterized by DNaseI hypersensitivity sites, are associated with RNA pol II, and can be spliced and polyadenylated (Hung and Chang et al. 2010, Wu et al. 2010). Many lncRNAs associate with chromatin regulatory proteins such as PRC2, suggesting they have a functional role in regulating gene expression (Khalil et al. 2009, Kozoiel et al. 2010).
1.3.1 Regulatory IncRNAs

A small subset of IncRNAs expressed in mouse ES cells maintains ES pluripotent states and repress lineage specific programs (Guttman et al. 2011). While a majority of IncRNAs act in trans, some may act in cis to regulate expression of adjacent genes (Guttman et al. 2011, Ponjavic et al. 2009, Orom et al. 2010). Many IncRNAs have been found to play a role in regulating autosomal gene expression. One example, HOX antisense intergenic RNA (HOTAIR), is a trans-acting IncRNA, that represses HOXD expression during development (Rinn et al. 2007). In contrast, the IncRNA Air acts in cis to silence four genes at the Igf2r locus by both promoting Air transcription at the locus as well as recruiting the H3K9 methyltransferase G9a to promoters of adjacent genes (Nagano et al. 2008, Seidl et al. 2006). HOXA transcript at the distal tip (HOTTIP) activates HOXA in cis by recruiting histone-modifying enzymes (Wang et al. 2011). A new class of IncRNAs, called IncRNA activators (IncRNA-a) is emerging (Orom et al. 2010, Wang et al. 2011, Lai, Orom et al. 2013). These IncRNA-a act on nearby genes to enhance transcription and increase expression levels. Additional evidence to support a functional role for IncRNAs has come from recent human cell line studies that have found several IncRNAs to be enhancers of genes critical for regulating development and differentiation (Orom et al. 2010).

1.3.2 Ddx3x locus

The DEAD box helicase 3 gene (Ddx3x) is a highly conserved locus in mammals (Figure 4A). Adjacent to Ddx3x is a IncRNA, 2010308F09Rik, that escapes inactivation (Lopes et al. 2011). This IncRNA is conserved in mouse and rat, but not in other mammals. In mouse and rat, Ddx3x and 2010308F09Rik are the only two escape genes in the escape domain, but in mammals lacking 2010308F09Rik, the escape domain includes neighboring genes (Figure 4B).

![Figure 4. Poor conservation of IncRNA at Ddx3x locus and escape domain expansion. (A) A representation of a 1.2 Mb segment of the inactive X chromosome. Arrows demarcate the direction of transcription with yellow indicating genes that are silenced and blue indicating genes that escape (B) Chart depicting inactive X chromosome loci conservation amongst several species. Genes that are silenced are yellow blocks while genes that escape are indicated by blue blocks. Stars indicate where the IncRNA is conserved.](image-url)
1.4 *Kdm5c* escape domain

*Kdm5c* is a H3K4 histone demethylase that removes methyl groups from di- and trimethylated H3K4 (Iwase et al. 2007, Klose et al. 2007) and is a highly conserved gene in mammals (Aguinik et al. 1994, Xu et al. 2002, Lee et al. 2007). *Kdm5c* is important for normal development as mutations in this gene are implicated in X-linked mental retardation cases (Jensen et al. 2005, Tzschach et al. 2006). *Kdm5c* was the first escape gene identified that escapes inactivation in both mouse and human (Aguinik et al. 1994). Like previously described escape genes, *Kdm5c* is enriched for H3K4me3 and lacks H3K27me3 (Jegalian et al. 1998, Splinter et al. 2011, Pinter et al. 2012). Expression levels of *Kdm5c* in mouse from the inactive X are roughly 30% to that of the active X, but fluctuate depending on tissue type (Carrel et al. 1996, Sheardown et al. 1996). The adjacent IncRNA, *AK148627*, which is 2.6 kb downstream from *Kdm5c* also escapes inactivation (Lopes et al. 2011). This IncRNA is not conserved at the human *Kdm5c* locus, which may suggest a species specific functional role (Reinius et al. 2010, Lopes et al. 2011). Studies that localized CTCF binding sites using chromatin immunoprecipitation (ChIP)-based methodologies have identified a CTCF binding site upstream of *Kdm5c* (Filippova et al. 2005).

The *Kdm5c* escape domain has been evaluated to delimit the sequences necessary for escape. Bacterial artificial chromosomes (BACs) harboring a region of the X chromosome containing a gene which frequently escape inactivation (*Kdm5c*), were randomly integrated into female mouse ES cells (Figure 5A). After differentiation, four full-length, single copy transgene lines were identified in which the BAC had integrated onto the inactive X (Figure 5B). Genes at the site of integration were properly silenced, however transgenic *Kdm5c* escapes inactivation. Therefore, sequences within the BACs were sufficient for *Kdm5c* escape at an ectopic location (Li and Carrel 2008). By aligning the sequences of overlapping BACs, the region responsible for escape of *Kdm5c* from inactivation was narrowed to 112 kb.
Figure 5. Delimiting sequences necessary for transgenic Kdm5c escape. (A) The escape domain at Kdm5c including the adjacent lncRNA AK148627 are depicted by blue arrows with neighboring silenced genes in yellow. The red dotted box shows the overlap between the two transgenes. (B) Representative images of RNA FISH analysis conducted on cell lines with stable BAC integrations. Orange and purple boxes are representative of the transgenes in (A).

Recent studies further delimited this escape domain using cell lines that contained single-copy transgenes, however sequences distal to Kdm5c on the BACs were truncated. Truncation of the 3' end of Kdm5c and adjacent lncRNA AK148627 in a BAC that was integrated at a normally inactivated Tex11 locus, led to expansion of the escape domain (Horvath et al. 2013). This study also found that truncation of the 3' end of Kdm5c caused an extension of the escape domain to chromosomal genes as far as 300 kb downstream from the BAC integration site.

It is hypothesized that AK148627 acts in cis to silence genes downstream of Kdm5c on the inactive X chromosome. Work for this thesis focused on AK148627 and its potential role in regulating the distal boundary of the Kdm5c escape domain. While AK1486247 is not the only sequence deleted from the transgene, it is certainly the most interesting candidate given data that supports the role of lncRNAs in regulating gene expression. My thesis project tested whether AK148627 1) was expressed as an independent transcript in mouse ES cells, 2) was a unique transcript controlled by its own promoter elements, and 3) was expressed during critical time periods during the X inactivation process.
CHAPTER 2

MATERIALS AND METHODS
2.1 Cell lines and tissue culture conditions

2.1.1 Mouse embryonic stem cell culture conditions and differentiation

Undifferentiated ES cell lines were maintained at 37°C in 10% CO₂ and grown on 0.2% gelatin (Sigma) coated plates with mouse embryonic fibroblast (MEF) feeders in Dulbecco’s minimal Eagle medium (DMEM, all reagents from Invitrogen unless otherwise noted) supplemented with 15% fetal bovine serum (Gemini Bio-Products), 2 mM glutamine, 1 mM sodium pyruvate, 100 μM non-essential amino acids, 100 μM 2-mercaptoethanol, antibiotics, and 1000 U/ml leukemia inhibitory factor (LIF, Chemicon). To induce differentiation, ES cells were transferred to untreated petri dishes and grown for three days in differentiation media lacking LIF, DMEM supplemented with 10% fetal bovine serum, 2 mM glutamine, 100 μM 2-mercaptoethanol and antibiotics, at 37°C in 10% CO₂. The resultant embryoid bodies were pelleted by gravity and transferred to treated tissue culture dishes to allow for differentiation over the next seven days. In total, cells were differentiated for ten days.

2.1.2 HEK293 cells

HEK293 cells were maintained at 37°C in 5% CO₂ in DMEM supplemented with 10% fetal bovine serum, and 2 mM L-glutamine. Cells were passaged at a density of 1:10.

2.1.3 Mouse tissues

Tissue samples used for expression analysis were previously collected and are from (C57B6 x M. Castaneous) F1 animals (Carrel et al. 1996). Two individual (one male one female) newborn RNA samples were used for RT-PCR analysis. The female mouse ES cell line EL16 (M.castaneus x 129) was previously provided by En Li (Novartis Institute), and subclones used for XCI studies. Subclone SA13 was isolated and stably retained two Xs for >40 doublings (>94% cells are 40,XX).

2.2 Plasmid DNA preparation and transfection

2.2.1 Plasmid construction

Vectors were cut with the appropriate restriction enzymes for 4 hours using 1 μg of DNA in a 50 μl reaction for vector. PCR products tagged with restriction enzyme sequences were purified after amplification using GenScript QuickClean II Purification Kit (L00419) with an elution volume of 30 μl. The 30 μl elution would then be used to cut the PCR product in a 50 μl enzyme reaction for 4 hours. Enzymatic reactions were deactivated for 20 minutes at 65°C. Vectors were dephosphorylated to prevent self-ligation and phosphatase was deactivated at 65°C for 15 minutes. Ligations were set up in a total reaction volume of 10 μl with molecular ratios of vector:insert of 1:2 and 1:4 using Invitrogen’s T4 DNA ligase system. Ligations were incubated overnight at 16°C according to the manufacturer’s guidelines (NEB).

2.2.2 Bacterial transformation

Plasmids were electroporated into electrocompetent DH5α E.coli cells using the BioRad Gene Pulser Xcell Electroporation System. Approximately 100 ng of plasmid was added to 25 μl of electrocompetent cells and electroporated using a 1 mm chilled cuvette. Electroporation
conditions were as follows: 1.8kV, 25 μF and 200 Ω. Only those cells with pulse times ranging between 4.0 msec and 5.0 msec were resuspended in 450 µl of SOC media. 300 µl of resuspension was plated onto appropriate antibiotic treated LB plates.

2.2.3 Plasmid mini-prep purification

Colonies were individually selected and grown in 5 ml cultures overnight. The following day, plasmids were isolated using Promega’s Wizard Plus SV Miniprep DNA Purification System (A1460) with an elution volume of 45 µl. Plasmids were screened for insert orientation and overall integrity by restriction enzyme digest and PCR analysis.

2.3 Genomic AK148627 transcriptional characterization

2.3.1 Transcriptional cDNA analysis

Mouse ES cells were harvested at approximately 80% confluency from 60mm tissue culture dishes at days 0, 3, 7 and 10 during cellular differentiation and RNA was isolated using Zymogen’s RNA isolation kit (R1054) with an elution volume of 35 µl. RT-PCR was performed using M-MLV reverse transcriptase (Invitrogen) with modified conditions suggested by the manufacturer to allow for amplification of prep with lower RNA yield. Briefly, 500 ng of RNA was treated with DNaseI (all reagents are from Invitrogen unless otherwise noted) to eliminate genomic DNA. Half the volume was placed into a +RT tube containing 10 µM random hexamers (NEB) and 1 mM dNTPs while the remaining half was transferred into a –RT tube lacking the random hexamers. After a 5 minutes incubation at 65°C tubes were placed on ice for 5 minutes. +RT and –RT reactions were placed into a tube containing 10 mM DTT, 1x first strand buffer, 20 units RNAseOUT and M-MLV (+RT reaction only). Tubes were incubated at room temperature for 15 minutes before being placed at 37°C for 1 hour. Complete deactivation was achieved by incubating tubes at 70°C for 15 minutes.

In addition to mouse ES cells, two neonate mouse RNA samples (2115, 2116) previously collected were analyzed as a post X inactivation control under suggested manufacturer’s protocol. Strand-specific RT-PCR was also performed using HY009 in place of random hexamers with NL213 as a reaction control (Table 1).

2.3.2 Determining AK148627 promoter activity

Given that promoter elements typically reside 5’ to the gene whose expression it controls, we amplified a 1.1 kb (containing the transcription start sequence) and 917 bp region (lacking the transcription start sequence) immediately upstream of AK148627 from SA13 genomic DNA and cloned these fragments into the pGL3-basic vector (Promega, supplied by Dr. Greg Yochum) utilizing the KpnI (NEB) restriction site upstream of the firefly luciferase gene (Table 1-primers include KpnI enzyme sequence, Figure 11). PCR amplification conditions used 1.5 mM salt and cycling parameters were set at 95°C for 5 min, 95°C for 30 sec, 57°C for 30 sec, 72°C for 1 min, and 72°C for 10 min with a total of 35 cycles. HEK293 cells were seeded at 50,000 cells per well in a 24-well plate the day prior to transfection. The day of transfection, 1 µg of DNA containing 2 ng of a renilla reporter plasmid, 100 ng of the desired luciferase construct and up to 1 µg with a promoterless pBlueScript plasmid was transfected using a mixture of 0.3 M calcium chloride and 2x HBS solutions. Cells were transfected for 6 hours at 37°C, rinsed with 1x PBS and then incubated in fresh growth media for an additional 18-24 hours. Cells were
lysed using 1x passive lysis buffer (Biotium reagents supplied by Dr. Greg Yochum). Fluorescence was measured with 25 µl of lysed sample and 50 µl of firefly and renilla working reagents using a Glomax 20/20 luminometer with GLOMAX software. Values were normalized using the standard firefly:renilla ratio comparison.

2.4 Functional role assessment of AK148627

2.4.1 Repressor/enhancer evaluation

The gene encoding AK148627, along with 1.1 kb of upstream regulatory sequences, was amplified by PCR using primer sequences listed in Table 1 and SA13 genomic DNA as the template. PCR cycling conditions were performed at 1.5 mM MgCl₂ and 95°C for 5 min, 95°C for 30 sec, 57°C for 30 sec, 72°C for 2:30 min and 72°C for 10 min. The total 3.7 kb amplicon was cloned into the BamHI (NEB) site of the pGL3-promoter vector in which the firefly luciferase gene was driven by a mammalian SV40 promoter. To confirm the results seen with the pGL3-promoter vector, the same region was cloned into a thymidine kinase (TK) promoter-driven firefly luciferase vector TK-luciferase. HEK293 cells were transfected, lysed and analyzed as previously described above.

2.4.2 Expression from luciferase vector constructs

Confirmation of the transcription status of the full length 3.7 kb region (lncRNA including upstream sequences) was also tested by cloning into the BamHI site of the pGL3-basic vector. HEK293 were transfected and 24 hours post transfection, cells were harvested and RNA was isolated using Zymogen’s RNA isolation kit. Random hexamer RT-PCR amplification was performed to generate cDNA under modified conditions previously described. As AK148627 is not conserved in human cells, any expression of the RNA transcript was from the pGL3-basic vector construct.

2.4.3 Evaluation of AK148627 regulatory DNA sequences and RNA transcript

A 2.5 kb truncated AK148627 lacking the transcription start sequence and upstream regulatory region was cloned into the BamHI site of the pGL3-promoter vector. HEK293 cells were transfected, lysed and analyzed as previously described. The truncated lncRNA was also cloned into the pGL3-basic vector and transfected into HEK293 cells for cDNA analysis. Random hexamer RT-PCR was performed as previously described to detect any AK148627 transcription from the truncated construct.

2.4.4 Levels of endogenous AK148627 expression during X inactivation

Female SA13 mouse ES cells were harvested at day 0, 3, 7 and 10 of the cellular differentiation protocol previously described. RNA was isolated using Zymogen’s RNA isolation kit and RT-PCR was performed under modified conditions previously described. cDNA was diluted 1:50 and evaluated by qRT-PCR using primers listed in Table 1. Optimization of primers allowed for amplification at 1.5 mM MgCl₂, 57°C annealing for 30 seconds and a 30 second extension at 72°C. The program was run on a BioRad MyiQ iCycler and analyzed with BioRad iQ5 software.
2.5 CRISPR knockout at endogenous AK148627 locus.

2.5.1 Targeting genomic DNA for CRISPR guide RNA design

Protopspacer adjacent motif (NGG) elements both upstream and downstream of AK148627 were identified by sequence analysis until 4 PAMs were identified. Sequences 22 nucleotides in length directly adjacent (5’ orientation to NGG) were run through several programs designed to trace the sequence uniqueness in the murine genome. Those sequences that did not have mismatches (particularly within the last 10 base pairs) were used for guide sequences. Four pairs of guide RNAs were selected in total for cloning.

2.5.2 CRISPR vector cloning

Oligos were resuspended in annealing buffer (10 mM Tris, pH 7.5-8.0, 50 mM NaCL, 1 mM EDTA). 1 µl of equimolar concentrations (100µM) of each primer per guide RNA pair (2 ul total), 1 µl 10x ligation buffer (NEB), 6.5 µl dH2O and 0.5 µl T4 PNK (NEB) were mixed and heated to 95°C for 5 minutes. Mixture was allowed to cool to room temperature. Annealed and phosphorylated primers were diluted 1:250 before ligation into PX462 vector (fourth generation CRISPR vector from Addgene containing puromycin gene and fused guide/tracrRNA).

Oligo duplexes containing a 5’-AAAC tag were subcloned into the BbsI site of the vector. PX462 was digested with BbsI (Thermo FastDigest) using a single-step digestion/ligation set up. 10 µl liguations were set up using Invitrogen’s T4 DNA ligase system. 100 ng of PX462, 2 µl of annealed primer oligos, 2 µl FastDigest buffer, 1 mM DTT, 1 mM ATP, 1 µl BbsI, 0.5 µl T4 DNA ligase were mixed to of final volume of 20 µl with dH2O. Ligations were performed in a thermocycler under the following conditions: 37°C for 5 minutes, 23°C for 5 minutes in which these two steps were cycled six times. This was repeated for each pair of oligos.

2.5.3 Bacterial transformation

Plasmids were electroporated into electrocompetent DH5α E.coli cells using the BioRad Gene Pulser Xcell Electroporation System. 2 µl of ligation mixture was added to 25 µl of electrocompetent cells and electroporated using a 1 mm chilled cuvette. Electroporation conditions were as follows: 1.8kV, 25 µF and 200 Ω. Only those cells whose pulse times ranging between 4.0 msec and 5.0 msec were resuspended in 450 µl of SOC media. 300 µl of resuspension was plated onto 50 µg/mL ampicillin treated LB plates.

2.5.4 Plasmid mini-prep purification

Colonies were individually selected and grown in 5 ml cultures overnight. The following day, plasmids were isolated via standard protocol using Promega’s Wizard Plus SV Miniprep DNA Purification System (A1460) with an elution volume of 45 µl. Plasmids were screened for insert by restriction enzyme digest.
Table 1. Primers

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-&gt;3’)</th>
<th>Size</th>
<th>Genomic Location</th>
<th>Vector components</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Detecting Transcription across the Kdm5c/AK148627 locus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL212</td>
<td>GCCTCACAGAGAGAGCCATC</td>
<td>218 bp cDNA; 484 bp gDNA</td>
<td>Kdm5c</td>
<td></td>
</tr>
<tr>
<td>NL213*</td>
<td>GGTCACTGTCCCCATCTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs29295789F</td>
<td>ATTCCTCTCTGGGCCAAGCT</td>
<td>600 bp w/ HY009</td>
<td>Spans AK148627 TSS</td>
<td></td>
</tr>
<tr>
<td>HY008</td>
<td>CCACCCAGAGCCTGAAAGAG</td>
<td>250 bp w/ HY009</td>
<td>AK148627</td>
<td></td>
</tr>
<tr>
<td>HY009*</td>
<td>TCGATC-GGTACC- CCTGTGTCAGACCATATACTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs29295787F</td>
<td>GAACCAGCAAGCTGTCATAGC</td>
<td>264 bp</td>
<td>AK148627</td>
<td></td>
</tr>
<tr>
<td>rs29295787R</td>
<td>GCAAAGTTCCTCACATGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Assessing AK148627 upstream regulatory sequences for promoter activity**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-&gt;3’)</th>
<th>Size</th>
<th>Genomic Location</th>
<th>Vector components</th>
</tr>
</thead>
<tbody>
<tr>
<td>HY005</td>
<td>CCATCA-GGTACC- AGTCCAGAAGCTAGGACCA</td>
<td>917 bp w/ HY007</td>
<td>1 kb upstream sequence of AK148627 (-) TSS</td>
<td>Cloned into pGL3- basic (promoterless vector)</td>
</tr>
<tr>
<td>HY006</td>
<td>TCGATC-GGTACC- ATGCTCTGATCTGCTGCTT</td>
<td>1145 bp</td>
<td>1 kb upstream sequence of AK148627 (+) TSS</td>
<td></td>
</tr>
<tr>
<td>HY007</td>
<td>TCGATC-GGTACC- AGAGTTGGGCGCAAGAGGAGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Delimiting sequences necessary for repressive lncRNA function**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-&gt;3’)</th>
<th>Size</th>
<th>Genomic Location</th>
<th>Vector components</th>
</tr>
</thead>
<tbody>
<tr>
<td>HY014</td>
<td>CCATCA-GGTACC- AGTCCAGAAGCTAGGACCA</td>
<td>3.7 kb w/ HY015</td>
<td>AK148627 w/ 1 kb upstream regulatory sequences and 500 bp downstream</td>
<td>Cloned into pGL3-basic and pGL3-SV40 promoter vector</td>
</tr>
<tr>
<td>HY015</td>
<td>TCGATC-GGTACC- GAAGAGAAGGCGCTCGTTGATG</td>
<td>2455 bp</td>
<td>AK148627 (-) 1 kb upstream regulatory sequences</td>
<td></td>
</tr>
<tr>
<td>HY028</td>
<td>CCATCA-GGTACC- CCACCCAGAGCCTGAAAGAG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**qRT-PCR analysis of AK148627**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-&gt;3’)</th>
<th>Size</th>
<th>Genomic Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>HY016</td>
<td>CCACCGAAAACATGACAGGG</td>
<td>104 bp</td>
<td>AK148627</td>
</tr>
<tr>
<td>HY017</td>
<td>AACAATAGCTCTGGTGGCACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY020</td>
<td>AGCATTTGCAATAGGGCTAGAGTC</td>
<td>108 bp</td>
<td>Xist</td>
</tr>
<tr>
<td>HY021</td>
<td>AGGGGTTCTAATGTCAGACGACAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY024</td>
<td>CAGTTTCTCCATCCCACAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY025</td>
<td>CCTGCGACTTTCAACAGCAAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CRISPR oligo sequences**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-&gt;3’)</th>
<th>Size</th>
<th>Genomic Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>HY030</td>
<td>CACCCCCCAGCTGAAGTAGAGCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY038</td>
<td>AAAACTGTACTTCTGAAGGCTGG</td>
<td>Targets PAMs upstream of AK148627</td>
<td></td>
</tr>
<tr>
<td>HY032</td>
<td>CACGCTATGGTCAGTTGCAAGAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY039</td>
<td>AAACCTTCTGAAGACTGACATACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY034</td>
<td>CACCTGTCCGGTAAATTATGTGC</td>
<td>Targets PAMs downstream of AK148627</td>
<td></td>
</tr>
<tr>
<td>HY040</td>
<td>AAAAGCACAATTTCAGTGAGACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY036</td>
<td>CACCCAAATTATTTGTGATAGTCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HY041</td>
<td>AAATGCATCAAACAATATTGTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Primers used for strand-specific RT-PCR.

Restriction sites are underlined.
CHAPTER 3
RESULTS
Previous work in the Carrel lab (Li an Carrel 2008, Horvath et al. 2013) delimited the sequences necessary for Kdm5c to escape inactivation and these studies led to the hypothesis that AK148627 suppresses neighboring gene expression. This chapter focuses on characterizing the transcription of AK148627 which is a central tenant of this hypothesis.

3.1 AK148627 is a separate transcript from Kdm5c

Current RNA-seq data from the UCSC Genome Browser (http://genome.ucsc.edu/) suggests that AK148627 is transcribed in the same direction as Kdm5c; however current methods for identifying IncRNAs can have flaws that make orientation hard to identify (Carninci et al. 1996). Due to the close proximity of AK148627 to the 3’ end of Kdm5c (Figure 6A), it is necessary to confirm the orientation of AK148627 transcription.

3.1.1 Transcription termination sequences

Mammalian transcription is terminated by an “AATAAA” motif at the 3’ end (Proodfoot et al. 1991). Unspliced expressed sequence tag (EST) data supports a consensus stop sequence for Kdm5c. Correspondingly, ESTs terminate at the AATAAA motif. Despite the lack of data supporting that AK148627 is spliced to Kdm5c, a small number of ESTs can be detected in the intergenic region between Kdm5c and AK148627. Sequences at the 3’ end of Kdm5c and sequences in the intergenic region were visually screened for a transcription termination motif. As denoted by red hexagons, there is a transcription termination motif that aligns with the EST consensus stop for Kdm5c (Figure 6B). Additionally, there is an alternative stop sequence in the intergenic region that has limited EST data to support this as a frequent termination site. This raises the possibility that AK148627 is transcribed due to read-through transcription from the Kdm5c locus, prompting us to investigate the transcripts expressed at the AK148627 locus.
Random-primer reverse transcription using two, post-XCI mouse RNA samples (m1 and m2) and female mouse ES cells were used to assess transcription at the 5’ end of AK148627. As a control, an antisense primer for Kdm5c was used to detect sense, strand-specific transcription of Kdm5c to validate reaction conditions (Figure 7A). This primer could then be used to detect mature mRNA as it spans intron 22 in genomic DNA. Primers anchored 150 bp downstream of the AK148627 transcription start site and primers anchored around the annotated transcription start site were used to detect transcripts near the 5’ end of AK148627. While robust transcription of both AK148627 and Kdm5c were detected, no transcription spanning the annotated AK148627 transcription start site was detected (Figure 7B). Additionally, a strand-specific primer was used to detect transcription orientation of AK148627 in respect to Kdm5c in RNA sample m2. These data indicate that AK148627 is transcribed in the same orientation as Kdm5c. Altogether, these data support that AK148627 is a separate transcript from Kdm5c in pre and post-XCI cells (Figure 7B and 7C).
Figure 7. AK148627 transcription is not due to read-through from Kdm5c. (A) Diagram of the Kdm5c locus. Blue arrows indicate transcription orientation. Black arrows are representative of primers used for PCR to interrogate transcript expression in qRT-PCR reactions. Red arrows indicate primers used in single-strand qRT-PCR reactions. (B) Image of PCR-amplified cDNAs that were subjected to agarose gel electrophoresis. Indicated are the primer pairs used that are represented in (A). M1 and M2 are independent murine somatic cell RNA samples. (ss) is representative of single-strand qRT-PCR analysis. (C) Image of amplified mouse ES cDNAs subject to agarose gel electrophoresis. Primer pairs are numbered similar to (B) and are representative of pairs demonstrated in (A).

3.1.3 AK148627 promoter activity

To test for promoter activity at the AK148627 locus, two luciferase constructs were generated. A 1145 bp region, that includes the AK148627 transcription start site and approximately 1 kb of upstream sequences (designated + TSS), in addition to a 917 bp region that retains upstream sequences but truncates the putative AK148627 transcription start (designated –TSS) were cloned into a promoterless firefly luciferase vector (Figure 8A and 8B). These constructs were expressed in HEK293 cells and firefly luciferase levels were compared to the viral SV40 promoter vector. After normalization to the SV40 promoter values, the +TSS region immediately upstream of AK148627 contains weak promoter activity in an orientation dependent manner in HEK293 (Figure 8C). Together with the RT-PCR results above, these data suggest that AK148627 is a separate and distinct transcript from Kdm5c.
Figure 8. Sequences adjacent to AK148627 function as a weak promoter element. (A) and (B) are diagrams depicting the luciferase reporter constructs used to evaluate the potential promoter function of sequences upstream of AK148627. In (A), a 1145 bp region containing the AK148627 TSS was cloned upstream of the luciferase gene in the pGL3-basic plasmid. In (B), a 917 bp fragment lacking the AK148627 TSS was cloned upstream of luciferase. (C) Luciferase levels detected in HEK293 cells transfected with the indicated reporter constructs. Dual blue bars are representative of two independent constructs tested in technical triplicates. Statistical significance was determined using a student's T-test. Error bars represent standard deviation.
3.2 AK148627 suppresses luciferase expression

3.2.1 AK148627 expression from luciferase vector

To determine if AK148627 is expressed from the vector, we cloned sequences into the pGL3-basic vector. One construct contained a 3.7 kb fragment that included all regulatory sequences and the lncRNA, while another vector contained a 2.5 kb fragment in which the lncRNA transcription start site was truncated. Constructs were expressed in HEK293 where AK148627 is not conserved in humans. RT-PCR was performed on collected RNA samples from the transfected HEK293 cells. As a control, an antisense primer for UBE1 was used to detect sense, strand-specific transcription of UBE1 to validate reaction conditions. This primer pair could then be used to show that mature mRNA was transcribed by spanning an intron as in a similar method described in section 3.1.2 of this thesis. Since AK148627 is not conserved in humans, mouse genomic DNA (SA13 gDNA) was needed for a positive control for the PCR reaction conditions when evaluating AK148627 expression from the basic vector in transfected HEK293 cells. As expected AK148627 is only expressed from vectors in which the transcription start is present (Figure 9). Expression is not seen upon truncation of the transcription start. These results additionally support our conclusion that AK148627 is a separate transcript from Kdm5c.

Figure 9. AK148627 is expressed from basic luciferase vector. (A) Diagram depicting the luciferase constructs used to evaluate the role of AK148627 in regulating gene transcription. Fragments of the indicated lengths were subcloned downstream of luciferase in the pGL3-basic plasmid. (B) Schematic of lncRNA locus. Black arrows demarcate primers pairs (Table 1). Image of amplified cDNAs generated from transfected HEK293 cells were subject to agarose gel electrophoresis. Indicated are the primer pairs used that are represented above.
3.2.2 AK148627 reduces luciferase expression

Cloning the entire 3.7 kb IncRNA into a SV40 promoter-driven luciferase vector, there is a reduction in firefly luciferase expression relative to the SV40 vector alone. To further segment the 3.7 kb region for potential repressor activity, the 1145 bp upstream region (+TSS), as well as the 2455 bp truncated IncRNA (-TSS) were cloned and analyzed (Figure 10A). Reduction in luciferase expression is only seen with the 3.7 kb IncRNA segment in both HEK293 and mouse ES cells (Figures 10B and 10C). These data indicate that the IncRNA RNA transcript is necessary for repression of luciferase expression.

![Diagram](image)

**Figure 10. AK148627 represses luciferase expression.** (A) Diagram depicting the luciferase constructs used to evaluate the role of AK148627 in regulating gene transcription. Fragments of the indicated lengths were subcloned downstream of luciferase in the pGL3-promoter plasmid. (B) Luciferase levels detected in HEK293 cells transfected with the indicated reporter plasmids. (C) Luciferase levels detected in mouse ES cells transfected with the indicated reporter plasmid. Error bars depict standard deviation. A student's T-test was used to evaluate statistical significance. Dual blue bars are representative of two independent constructs tested. Experiments performed in technical triplicates.

3.3 AK148627 expression in pre- and post-XCI cells

Hypothesizing that AK148627 regulates the distal escape boundary at Kdm5c during X inactivation requires it to be expressed in pre-XCI ES cells if this is biologically relevant. Prior to XCI in a mouse, Xist is expressed at low level from both X chromosomes and around the time of
implantation when random XCI occurs, Xist is upregulated from the future inactive X (Takagi et al. 1982, Nesterova et al. 2003). *Xist* levels from the inactive X continue to rise during the roughly 10-day process. To address whether *AK148627* is expressed during X inactivation, mouse ES cells were differentiated over ten days and harvested at days 0, 3, 7 and 10 of the differentiation process. Female mouse ES cells are fully differentiated at day 10 when cultured with differentiation media lacking LIF and *Xist* levels are lower in day 0 pre-XCI compared to fully differentiated cells (Lee et al 1999, Simon et al. 2013). We find that *Xist* expression is roughly 30-fold greater in day 10 cells than in day 0 cells, confirming that our mouse ES cells have indeed undergone X inactivation, and more importantly confirm that RNA isolated from day 0 cells is representative of a pre-XCI state (Figure 11A). *AK148627* expression is approximately 3-fold higher prior to X inactivation in undifferentiated cells in comparison to differentiated cells (Figure 11B). Relative *AK148627* expression levels were also compared to *Kdm5c* levels (Figure 12). In fitting with IncRNA properties, *AK148627* is expressed at a lower level than *Kdm5c* throughout X inactivation. These data support that *AK148627* is expressed at critical times during the onset of X inactivation.

![Figure 11. AK148627 is more highly expressed prior to the onset of X inactivation.](image)

(A) qRT-PCR analysis of cDNAs synthesized from RNAs isolated from mouse ES cells on the indicated days of induction of X inactivation. M1 and M2 are RNA samples isolated from differentiated mouse somatic cell lines as a control. The data are normalized to levels seen at day 0. (B) qRT-PCR analysis of *AK148627* levels normalized to M1 and M2. In (A) and (B) levels are also normalized to *Gapdh* expression. Both graphs are representative of a biological replicate performed in technical triplicate. Error bars represent standard deviation.
Figure 12. *AK148627* expression is lower than *Kdm5c* levels during X inactivation. qRT-PCR analysis of cDNAs synthesized from RNAs isolated from mouse ES cells on the indicated days during X inactivation. 2116 and B119 are murine somatic samples as a control. Dark grey bars are representative of *Kdm5c* levels while light grey bars are representative of *AK148627*. Samples were performed in technical triplicate. Error bars are standard deviation.
CHAPTER 4
DISCUSSION
4.1 Model for AK148627 regulation of gene expression on the X chromosome

It can be hypothesized from the 80 kb BAC sequence deletion that AK148627 is not acting as an activating-lncRNA (Orom et al. 2010) to direct escape. If the lncRNA were an activating-lncRNA, it would be expected that Kdm5c would not escape inactivation when AK148627 is deleted. Previous studies in our lab demonstrated that upon an 80 kb BAC truncation, genes downstream from the Tex11 transgene integration site now inappropriately escape inactivation. For this reason, we hypothesized that AK148627 may play a role in inactivating genes downstream of Kdm5c. It is unlikely that the lncRNA silences downstream genes through transcriptional interference. The nearest downstream gene to AK148627 is roughly 65 kb downstream, which surpasses the distance demonstrated in transcriptional interference (Nagano et al. 2008). lncRNAs can silence loci in cis by recruiting chromatin-modifying complexes (Nagano et al. 2008, Pandey et al. 2008). An immunoprecipitation assay using an antibody against the PRC2 catalytic subunit EZH2 has shown that the AK148627 transcript can be pulled down (co-IP) (Zhao, Ohsumi et al. 2010), suggesting that AK148627 may recruit EZH2 to the locus to silence adjacent genes. However, these experiments can be non-specific (Angers et al. 2002) and further evidence of the AK148627/EZH2 interaction using a more specific assay is necessary. A definitive experiment that looks at endogenous AK148627 recruitment of EZH2 to the transcriptional locus is optimal, however not trivial, but would most likely involve high-specificity pull-down assays using a probe for AK148627 and assessing for EZH2 interactions.

Our analysis using reporter assays suggest that the entire AK148627 gene is important for the repressive function seen upon luciferase expression levels (Figure 10). This conclusion must be taken cautiously as there are several limitations with using plasmids to report the activity of a lncRNA. It is important to remember that plasmid sequences may not recapitulate endogenous regulation because they lack genomic context and chromatin properties of the endogenous locus. Not only does a plasmid lack epigenetic regulation, it also lacks 3D chromatin conformations that can influence gene expression. A plasmid is a circular piece of double-stranded DNA that can be supercoiled, but lacks structure that may facilitate or inhibit interactions that can effect gene expression. While the luciferase results support our hypothesis, it is important to recognize the limitations of the model system.

Previous studies used transgenes to evaluate escape at an ectopic location on the X chromosome (Li and Carrel 2008, Horvath et al. 2013). By inserting a large piece of DNA into an ectopic location within the genome, the genomic architecture of that locus may be altered, influencing gene expression. Multiple transgenes have mimicked escape genes not only by escaping inactivation at an ectopic location, but transgenes have been able to recapitulate endogenous inactive X expression levels (Li and Carrel 2008). Other factors, such as chromosome location, may also have a larger influence on escape than transgene composition (Ciavatta et al. 2006). To definitively determine whether AK148627 regulates the Kd5mc distal escape domain boundary, a gene knockout is ideal.

4.2 CRISPR knockout to delete endogenous AK148627

The CRISPR/Cas9 technology is a specific way to target genes of interest for deletion at an endogenous locus (Horvath et al. 2010). The Cas9 nickase has been shown to be efficient at deleting sequences up to 6 kb within the mouse genome (Ran et al. 2013). This system relies
on several important factors in order to get efficient targeting of the enzyme to the appropriate location within the genome. Protospacer adjacent motifs (PAM) are necessary to recruit the Cas9 nuclease to the genomic DNA. Such motifs are often annotated by the trinucleotide “NGG” sequence. In order to begin to target AK148627 for knockout using the CRISPR/Cas9 system, PAMs need to be recognized both upstream and downstream of AK148627 in order to cut both upstream and downstream of the gene of interest that will be deleted upon repair by non-homologous end joining. Importantly, the PAMs must be on different strands of the DNA (one antisense, one sense). The 5’ PAM is roughly 50 base pairs upstream of the annotated transcription start site of AK148627 to delete the transcription start site. Sequences directly 5’ to the PAM are the sequences that will be targeted by guide RNAs. These guide RNAs are 22 nucleotides in length and will be the exact same sequence as the target sequence in order to base pair with the opposite strand of DNA to allow for the Cas9 enzyme to nick the DNA. A cartoon illustrating these concepts is outlined in Figure 13A. When the guide RNA sequence is cloned into a CRISPR vector, it is often cloned next to a tracrRNA sequence so that the transcription of the guide RNA will include this tracrRNA sequence as a fusion transcript. The tracrRNA is important to provide secondary structure of the RNA molecule such that the guide RNA sequence is available for complementary base pairing with its intended genomic target. The guide RNAs are screened prior to cloning to ensure that sequences chosen are unique to the genome to avoid off-target effects. Due to the size of AK148627 and deletions of considerable size efficiently seen with the Cas9 nickase system, we chose to utilize a nickase system to try and knockout our IncRNA of interest at the endogenous location in female mouse ES cells.

We utilized PAM sequences both upstream and downstream of AK148627 that had unique sequences that could be used for guide RNA cloning. After confirming vector integrity post guide RNA cloning, female mouse ES cells were transfected with CRISPR constructs. Colonies that remained after antibiotic selection were pooled together for preliminary analysis to evaluate whether our transfection setup was successful. Primers were anchored both upstream and downstream of AK148627 outside of the potential deleted region and the pooled colonies were screened via PCR (Figure 13B & C). An unedited locus would produce an amplicon of 2.6 kb, while a deletion of AK148627 would produce an amplicon of 1 kb. Figure 13D suggests that in a subset of pooled cells, AK148627 was deleted upon CRISPR transfection.
Figure 13. CRISPR knockout of AK148627. (A) Schematic for CRISPR/Cas9 nickase system. Cas9 enzyme is represented by the orange shape, guide RNAs are represented as the bold green line. The fused tracrRNA is depicted as a red squiggly line. (B,C) Location of the PCR primers used to detect a possible deletion in ES cells transfected with the CRISPR plasmid. (D) Image of PCR products generated from genomic DNA prepared from ES cells that were untransfected or transfected with the AK148627-specific CRISPR plasmid and subjected to agarose gel electrophoresis.

While amplification of the 1 kb fragment is not robust, there are some factors to take into consideration. It is unlikely that both the active and inactive X chromosome AK148627 locus were targeted on the first transfection attempt because of the frequency to which genome editing occurs. Ideally both AK148627 loci should be targeted. This could account for the faint 1 kb band that is present, but the fact that deletions occurred could also explain the intensity reduction of the 2.6 kb product as both the (-) CRISPR and (+) CRISPR were amplified with 40 PCR cycles.
The next steps would be to target both AK148627 alleles and determine whether a deletion at the endogenous locus reveals similar results as to the truncated transgene studies previously published (Horvath et al. 2013). To determine whether the deletion of AK148627 leads to adjacent downstream gene expression at the Kdm5c locus, qRT-PCR can be used to detect changes in gene expression from neighboring genes such as Tspyl2, Gpr173, etc. By comparing levels of expression from downstream genes in unedited female mouse ES cells to AK148627 knockout mouse ES cells, we should be able to determine whether downstream genes are appropriately silenced. We can use Xist as a marker for undifferentiation and Kdm5c as a positive control for a gene that we know escapes inactivation. While escape levels of expression of Tspyl2 and Gpr173 may not be as robust at Kdm5c, we would expect to see elevated expression levels if AK148627 is necessary for the silencing of adjacent genes. Ultimately, we could hypothesize that as AK148627 is transcribed from the locus, EZH2 is recruited to repress neighboring gene expression via H3K27me3. With neighboring downstream genes being roughly 25 kb away, it could be possible that the genomic architecture of the locus could facilitate the close proximity necessary to bring EZH2 into range to repress downstream genes. We could utilize DNA/RNA FISH to demonstrate that AK148627 acts at its transcriptional locus. The EZH2 data previously published would help to put a working model together (Zhao et al. 2008). It is very likely for various reasons that potential functional domains of AK148627 (i.e. RepA domain of Xist) have yet to be identified. Further studies could investigate the IncRNA for functional domains necessary to recruit EZH2. Ultimately, the big unanswered question in the field of X inactivation that remains is how elements can act in one fashion on the inactive X that are not seen on the active X.

The IncRNA is not the only sequence that was deleted from the transgene and it is possible that upon AK148627 deletion, genes downstream of Kdm5c are still properly silenced. This would not necessarily suggest that AK148627 is not involved, but rather AK148627 is not necessary for silencing. As mentioned previously, AK148627 is only 2.6 kb of an 80 kb region of interest. The lack of ChIP-seq data for female mouse ES cells also limits our understanding of what chromatin modifying or structural proteins may bind within this 80 kb region. We can utilize datasets from somatic tissues and cells to get an idea of the epigenetics of the locus by looking for signatures of IncRNAs, such as RNA pol II, DNaseI hypersensitivity and various histone modifications that are indicative of euchromatin (Hung and Chang et al. 2010, Wu et al. 2010). These approaches can be extended to the 80kb region that was deleted to fully evaluate boundary function.

Alternatively if qRT-PCR results do not suggest a functional role for AK148627, there are other possible scenarios. First, it has been shown that ncRNAs nearby protein-coding genes are expressed due to proximity (Ponjavic et al. 2009). It could be that the open chromatin conformation of escapee Kdm5c allows RNA pol II access to AK148627 for transcription. If this were the case, it would be expected that there would be a lesser agreement amongst EST data of a consensus transcription start site. While AK148627 is not conserved between mouse and human, there is still valuable information to be learned. While the mechanism of silencing may greatly vary, important proteins such as EZH2 may be identified that are largely conserved.

This thesis has demonstrated several important concepts in addressing the hypothesis that AK148627 acts in cis to silence genes downstream of Kdm5c on the inactive X chromosome. AK148627 is driven by a weak promoter and can be designated as a separate
transcript from the juxtaposed \textit{Kdm5c} transcript. There is a significant reduction in luciferase expression upon cloning the entire AK148627 sequence into the luciferase vector. AK148627 is expressed from this vector only in its entirety as a 5' truncation does not facilitate transcription from the vector. Additionally, AK148627 is appropriately expressed at critical time points during X inactivation and is elevated prior to XCI. Altogether this thesis supports a potential role for the lncRNA, \textit{AK148627}, in silencing adjacent genes downstream of the escapee \textit{Kdm5c} in murine X inactivation.
REFERENCES


