UNIVERSITY OF CALIFORNIA
RIVERSIDE

Developmental *Xist* Induction is Mediated by Enhanced Splicing

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Cell, Molecular, and Developmental Biology

by

Cheryl Anne Stork

March 2019

Dissertation Committee:
Dr. Sika Zheng, Chairperson
Dr. Martin Garcia-Castro
Dr. Weifeng Gu

The Dissertation of Cheryl Anne Stork is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

First and foremost I would like to thank the Zheng Lab. Dr. Zheng, thank you for the advice and mentorship on experiments and being a good scientist over the years, it has been invaluable. My fellow lab mates, Jeff Li and John Vuong. Jeff thanks for the help with the single cell experiments, for all the jokes, and just for being a soundboard on some of the tougher days in lab. John, my cell culture lab buddy, thanks for all the discussion and late night science chats. I would like to thank Dr. Volkan Ergin, you have been a tremendous resource as we optimized protocols and also a great source of mentorship throughout the years. I would like to thank Dr. Min Zhang, the only one who has been in lab with me since the early days, I would like to thank you for all the advice and help you have given me over the years from working with me on primary cultures or talking about papers or working on animals. It's been a pleasure working with you all these years. I would also like to thank Dr. Lin Lin, you have been such an amazing presence in the lab with your positivity and willing to help. Thank you for your help with the protein work for the NAR paper, for letting me watch your IUE experiments, and just always willing to answer any random questions.

I would also like to thank my committee members over the years, Dr. Martin Garcia-Castro and Dr. Weifeng Gu for all of your guidance and experimental questions that would get me to think outside the box of my project. I would like to give a special thanks to Dr. Bachant for taking the time to listen to me and be an advisor when graduate school was getting especially difficult, our

talks were imperative in keeping me motivated during the especially difficult patches. I would also like to thank Dr. Monica Carson and Dr. Djurdica Coss for their advice on my career and research.

I would especially like to thank Kathy Redd, you have been there since I entered the program and have let me randomly come into your office countless of times. Thanks for all of the help with navigating CMDB and the countless amount of advice you have given me over the years. I would also like to thank the Dynamic Genome group, working with you guys have given me many opportunities to work on my teaching skills and revitalized my passion for mentorship and teaching the youth.

I would like to thank all of the friends I have made in graduate school. More than anything you guys knew how arduous the past few years have been but kept me going and kept reminding me that I had what it takes to finish this. Thanks for laughs, the rants, the beers, and just being there. I would like to thank a few people in particular. Lee Doss, my trench buddy from day one, thanks for always being there. My kindred spirit and best friend Jason, you have been there through all of the trials of graduate school since we met, thanks for keeping me grounded and reminding me to keep going when I wanted to quit, and thanks for reminding to not be all work and no play and try to go out and have fun once in awhile. To Erin Sternberg, Stephanie Thurmound, and Kris Dias, thanks for all the cheers and beers over the years.

Lastly I would like to thank my parents. Without you guys I would simply not be here. There were many times that you guys didn't completely understand what I was doing but you just let me do what I needed to do to follow my passions and continue my education. I wouldn't have been able to make it this far without all of the support you both have given me throughout the years.

The text of this dissertation and figures in part is a reprint of the material as it appears in Nucleic Acids Research, 2018. The co-author Sika Zheng listed in that publication directed and supervised the research which forms the basis for this Chapter 2, 3, and 4 of this dissertation. The other authors Zhelin Li prepared the cDNA from single cells that I used for the restriction fragment length polymorphism analysis presented in figure 12 and 13 and Dr. Lin Lin prepared protein lysates and performed the western blot analysis of PTBP1 expression presented in figure 5 and 9.

The citation is as followed:

Stork, C., Li, Z., Lin, L., & Zheng, S. (2018). Developmental Xist induction is mediated by enhanced splicing. *Nucleic Acids Research*, *426*, 857–12. http://doi.org/10.1093/nar/gky1198

DEDICATION

I dedicate this dissertation to my parents, without their understanding and support this wouldn't have been possible.

ABSTRACT OF THE DISSERTATION

Developmental *Xist* Induction is Mediated by Enhanced Splicing

by

Cheryl Anne Stork

Doctor of Philosophy, Graduate Program in Cell, Molecular, and Developmental Biology
University of California, Riverside, March 2019
Dr. Sika Zheng, Chairperson

Dosage compensation in female placental animals is accomplished by silencing one of the two X-chromosomes. X-chromosome inactivation (XCI) is initiated early in development by preferentially upregulating the lncRNA *Xist* expression from the future inactive X-chromosome. The work presented in this dissertation aims to understand the transcriptional regulation of *Xist* during XCI induction

Xist has long been identified as being both required and sufficient to induce X-chromosome inactivation when expressed *in cis* from the future inactive X-chromosome. *Xist* has 8 exons, is alternatively spliced, 3' polyadenylated, and 5' capped similar to protein coding genes but is retained in the nucleus. Multiple studies have identified that one of the earliest events of XCI is upregulation of the *Xist* transcript. However post-transcriptional regulation of *Xist* hasn't been well

characterized. The role of splicing of lncRNAs is currently unknown. Splicing of *Xist* has possibly been overlooked historically because *Xist* was first cloned as spliced RNA and the spliced form is sufficient to induce XCI therefore most studies have focused on the spliced form.

In Chapter 2 I show that differentiation induces splicing and that this is post-transcriptionally regulated after upregulation of *Xist* during the initiation of XCI. New technologies have allowed identification of a significant amount of potential binding partners, suggesting more regulation of *Xist* exists than was previously known. In Chapter 3, I show a new role for known splice regulator PTBP1 in regulating splicing of *Xist.* My data shows that PTBP1 is required for efficient splicing of *Xist* and suggests that this interaction is indirect and may involve additional trans acting factors. In Chapter 4, I show that splicing efficiency may be associated with the choice of which X-chromosome to silence during induction of XCI.

In summary the work presented in this dissertation suggests that splicing of *Xist* is an essential step of *Xist* induction from the inactive X-chromosome during XCI. This work suggests that inefficient splicing of *Xist* prevents expression of *Xist,* providing potential insight to how ES cells choose which chromosome to silence and potential binding partners involved in this choice.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 1: Introduction

Introduction to lncRNAs

Since the structure of DNA was discovered in the 1950s, there has been an extensive search to understand the roles and function of our genome. Upon completion of the human genome in 2001, it has since been found that the human genome contains over 3 billion base pairs and is estimated to encode for ~22,000-25,000 protein coding genes, however the actual number in humans is currently unknown. Since the 1970s, much of this noncoding DNA has been labeled as "junk DNA" (Ohno, 1972) due to the overwhelming presence of transposons, pseudogenes, and simple tandem repeats. In the 1970s hints begun to show that more of the DNA is transcribed than can be solely attributed to protein coding genes. With the development of whole genome technologies in the 1990s and 2000s, work done by the Encyclopedia of DNA elements (ENCODE) project suggests that 75% of the human genome is transcribed into processed or primary transcripts (Djebali, 2012).

LncRNAs are transcripts that are greater that 200nt, are transcribed by RNA Pol II, spliced, 3' polyadenylated, 5' capped, and often lack observable open reading frames. Some lncRNAs have cryptic ORFs in their sequence, however do not code for proteins. Some characteristics that have differentiated lncRNAs from mRNAs are that on average they are shorter than mRNA transcripts, have larger exons, expressed at low levels, are tissue and temporally regulated, typically retained in the nucleus, and less conserved among different species (Darrien et al., 2012; Cabili et al., 2011).

Originally identified by shotgun sequencing and microarrays, there was an ongoing question on the functionality of lncRNAs. Work by Guttman et al using chromatin immunoprecipitation followed by parallel sequencing (ChIP-seq) identified about 1,675 intergenic domains marked by H3K4me3 (marks promoters) and H3K36me3 (marks exons) domains that didn't overlap with current annotated genes. These transcripts were associated with RNA-binding complexes, immune surveillance, ESC pluripotency, neuronal processes, morphogenesis, gametogenesis, and muscle development (Guttman et al., 2009). Many of these lncRNAs promoters showed strong enrichment of RNA Pol II binding in ESCs, suggesting they are important for embryonic development.

Majority of lncRNAs display tissue specific (TS) expression patterns. A comparative study of 11 tissue types showed that lncRNAs are primarily found in a subset of tissue types including the brain, liver, heart, skeletal muscle, and retina. The richest source of lncRNAs however was found in the brain and was involved in development, neuronal maintenance, and function (Wu et al., 2013; Mattick, 2011). Other lncRNAs are ubiquitously expressed in multiple cell types and involved in cellular function and viability. Ubiquitously expressed lncRNAs are shorter than TS lncRNAs, highly expressed, and display more sequence conservation (Jiang, Li, and Zhao, 2016).

LncRNAs are classified into multiple types 1) intergenic, 2) intronic, 3) bidirectional 4) sense and 5) antisense transcripts depending on their location relative to protein coding genes. Intergenic lncRNAs do not overlap protein

coding genes while sense and antisense transcripts overlap the sense or antisense transcript of a gene respectively. Although much is known about expression and location of lncRNAs, the function of these RNA transcripts are still being discovered.

Transcriptional regulation of long noncoding RNAs

LncRNAs have many similarities to mRNAs. However studies have shown that although lncRNAs are lowly expressed they are highly regulated and have very specific expression profiles that depend on the cell type, developmental stage, tissue type, disease state, and even subcellular compartment they are found in. Currently very little is known about the how long noncoding RNAs are transcribed however some similarities and differences compared to mRNAs have been identified. Overall, lncRNA transcription and post-transcription mechanisms are very similar to mRNAs, however they are regulated differentially at their biogenesis, maturation, and degradation to account for the observed specificity in in lncRNA function and expression.

At the level of chromatin, lncRNA transcription initiation is marked by increased enrichment of H3K27ac and H3K4me3 marks at the mapped transcription start site (TSS) similar to mRNAs. Presence of H3K4me3 is typically sufficient to recruit the RNA Pol II to the TSS, however in lncRNAs repression of RNA Pol II binding by chromatin remodelers Swr1, Isw2, Rsc, and Ino80 provide an additional step of regulation (Marquardt et al., 2014).

LncRNAs undergo divergent transcription to produce either the sense or antisense RNA transcripts. A well-studied example of divergent transcription is the *Xist* and *Tsix* sense/antisense pair. Divergent transcription is enhanced by H3K56ac marks and the presence of nucleosome remodelers SWI/SNF and CAF-1 involved in nucleosome positioning promoting transcription in coding direction (Marquardt et al., 2014). LncRNAs can produce functional transcripts in sense/antisense pairs therefore it is important that there are additional regulators to assure the correct transcript is being expressed. This is absent in mRNAs, because transcription elongation is only productive in the sense direction. To ensure efficient termination of antisense transcripts, polyadenylation sites are found to be enriched in the antisense direction and U1 snRNP binding sites are enriched in the sense direction to promote efficient splicing (Alamada et al,, 2013 ; Quinn and Chang, H. 2016) These observations were further supported by work done by Krchňákova *et al.* that shows that strength of 5' splice sites, which recognize the U1 snRNP, determines the splicing efficiency of lncRNAs (Krchňákova et al., 2018)..

In unique cases, some lncRNAs undergo alternative processing different from mRNAs. LncRNAs *MALAT1* and *NEAT1* undergo 3' cleavage by RNase P to produce tRNA like small noncoding RNAs that are transported to the cytoplasm to perform their functions. Some intronic lariats formed by splicing of lncRNAs will disbranch and circularize forming ciRNAs. In some cases the introns of lncRNAs will produce sno-lncRNAs which are lncRNAs flanked by two

sno-RNAs. The variance in transcriptional products suggest that lncRNA elongation and splicing are involved in their observed specificity in different cell types and time points in development. Currently the exact mechanism of splicing of lncRNAs is still a mystery and identifying the interacting proteins involved would provide insight on how splicing of lncRNAs promotes functional specificity in the cell.

Regulation of biological processes in the cell often requires direct action of the lncRNA itself. Therefore mechanisms are in place in the cell to degrade lncRNAs to ensure proper expression occurs when necessary. Mutating UPF1 involved in the nonsense mediated decay pathway (NMD), was shown to increase the global expression of lncRNAs in *Arabidopsis Thaliana* suggesting that lncRNAs are targets of NMD. Poly(A)-binding protein nuclear 1 (PABPN1) promotes poly(A) dependent degradation by the nuclear exosome. PABPN1 deficiency results in the accumulation of lncRNAs suggesting that lncRNAs are also targeted for the nuclear exosome (Quinn and Chang, 2016).

Overall lncRNAs are highly regulated by various additional factors to ensure that they are expressed in the right place and the right time and degraded when their function is completed. The presence of these fail-safe mechanisms further supports the belief that lncRNAs are important in genomic regulation during development. These additional steps serve as stop-points to ensure that each step is met at the right time and place in cells and tissues.

LncRNAs have diverse roles in genomic regulation

Only a few lncRNAs have been well characterized to date, however trends in their functions are starting to emerge. Although there is very little sequence conservation between species, common modes of action have begun to appear. During development lncRNA mediated regulation of transcription has fallen into one of four observed categories 1) signals, 2) decoys, 3) scaffolds, and 4) guides.

LncRNA transcription is responsive to developmental cues and diverse stimuli unique to the cellular environment and factors present suggesting significant transcriptional control is involved. In this sense, lncRNA transcription can serve as molecular signal for a specific cell type or developmental phase. LncRNAs *Kcnq1* and *Air* accumulate at promoter region of the paternal X chromosome allele in the mouse placenta, signaling silencing *in cis* of the paternal X-chromosome (Mohammed et al., 2009; Wang and Chang 2011). *Xist* similarly is only expressed from the inactive X-chromosome during imprinted XCI and in the epiblast to ensure that random XCI doesn't occur at inappropriate times in cells. Another example, lncRNA-ROR expression is positively regulated by binding of pluripotency factors Oct4, Sox2, and Nanog. Differentiation depletes Oct-4 and lncRNA-ROR to promote differentiation (Loewer et al., 2010).

LncRNAs can also regulate transcripts negatively or positively by acting *in trans* as decoys for target proteins and prevent them from binding their gene

targets. Hung *et al*. found that DNA damage induces the expression of the lncRNA PANDA to favor cell cycle arrest (Hung et al., 2011). When expressed, PANDA binding to NF-YA is responsible for activating apoptosis (Puvvula et al., 2014).

LncRNAs that possess different binding sites and domains for effector molecules can act as scaffolds by recruiting effector proteins to initiate function at the correct time and place. Loss of expression of these lncRNAs would effect the proper location of these effector molecules. For example lncRNA binding of PRC1 and PRC2 to the lncRNA ANRIL is required to silence the CDKN2A/B locus *in cis (*Kong et al., 2018). Currently the mechanisms of these interactions are not well understood.

Guide lncRNAs act both *in cis* and *in trans* similarly to scaffolds. Guide lncRNAs bind to RNA binding proteins then directs them to their specific target. Well-known *Xist* is an example of a *cis* acting guide lncRNA responsible for directing RNA binding proteins responsible for the associated chromatin modifications of the X-chromosome during XCI induction (Lee, 2010; Plath et al., 2002; Wutz et al., 2002).

As technologies become more advanced the ability to identify lncRNAs increases. However much needs to be studied regarding the functions of these lncRNAs. Currently functions have only been assigned to a small subset of identified lncRNAs however trends such as their involvement in epigenetic regulation of their targets have emerged. Future projects should focus on finding

the interacting proteins and regulators of these lncRNAs to gain additional insight to their functional roles in the genome.

X-chromosome Inactivation: How females keep the balance during early embryonic development

Males and females express comparable X-linked gene expression from their X-chromosomes despite the additional X chromosome found in female eutherian animals. Dosage compensation is accomplished in females through the epigenetic silencing of one of the X-chromosomes in females by X-chromosome inactivation (XCI). XCI was first postulated in Lyon's 1961 seminal Nature paper where she hypothesized that each cell has one active X-chromosome and one inactive X-chromosome. Her hypothesis provided no new data in her 1961 paper but was built upon the identification of the bar body in the 1940s by Barr and Bertram and Ohno's observation that each cell has one heteropyknotic and condensed X-chromosome contrasting to the uncondensed X-chromosome that was similar in character as autosomes (Lyon,1961; Ohno, 1959; Barr and Bertram 1949).

Most studies on XCI have been done in mice that undergo two phases of XCI, imprinted and random. Between the 2 to 8 cell stage in mice, the paternally derived X-chromosome is inactivated in each of the cells of the pre-implantation embryo. Cells of the trophoectoderm will maintain the paternal imprint and form the extraembryonic tissues such as the placenta. Cells that are fated to become the blastocyst will reactivate the paternal X-chromosome resulting in two active

X-chromosomes in the blastocyst. Upon differentiation and cell lineage specification, females will compensate for the additional X-chromosome by silencing one of the X-chromosomes through random XCI mediated by the lncRNA *Xist*. The progeny of these initial cells of the epiblast maintain this silencing throughout cell lineage specification. All cells in the body will have one of the two X-chromosomes silenced.

The occurrence of both random and imprinted XCI is unique to mice. Marsupials only undergo imprinted XCI of the paternal chromosome. Other eutherians such as rabbits and humans only under go random XCI. In humans, both X-chromosomes remain active until blastocyst formation, however expression is dampened on both active X-chromosomes prior to implantation (Sahakyan et al., 2017). X-chromosome dampening (XCD) has only been previously observed in *C.elegans* hermaphrodites (Strome et al., 2014). Currently it is unknown if lncRNAs play a role in XCD in either species.

Murine embryonic stem cells (ES cells) are derived from the epiblast of the post-implantation embryo and house two active X-chromosomes. Upon differentiation, ES cells are able to recapitulate the process of random XCI *in vitro* and serve as useful model for studying XCI.

Early studies of XCI have identified the lncRNA *Xist* as both necessary and sufficient for inducing XCI. When ES cells are induced to differentiate, the amounts of X-chromosomes to autosomes are counted and an X-chromosome is stochastically chosen to be silenced. The mechanism of counting and choice of

XCI are currently unknown. The inactive X-chromosome will upregulate transcription of the lncRNA *Xist* from the previously identified X-inactivation center (*Xic*). Following expression, *Xist* is polyadenylated and spliced, but retained in the nucleus. In the nucleus, the spliced form acts as both a guide and decoy lncRNA to bring complexes to the targeted X-chromosome. *Xist* expression is tightly regulated to differentiation status to ensure that XCI isn't induced prematurely. Pluripotency factors Nanog, Oct4, and Sox2 bind to intron 1 of *Xist* and repress XCI in undifferentiated cells. Upon differentiation, reduction of *Nanog, Oct-4,* and *Sox2* gene expression has been associated with activation of *Xist* expression *in cis* on the inactive X-chromosome (Navarro et al., 2008).

Once activated and transcriptionally processed, stable spliced *Xist* RNA attaches and spreads across the X-chromosome through interactions with CIZ1 and HnRNPU (Riding-Figueroa et al., 2017; Sunwoo et al., 2017). *Xist* coating recruits repressive complex PRC1 and 2 and various chromatin modifiers to the inactive X-chromosome. These trans acting factors then deposit H2AK119ub and H3K27me3 marks to the target X-linked genes preventing their expression.

Advancements in deep sequencing techniques have led to the influx of identification of RNA binding proteins that bind to *Xist* (Chu et al., 2015; Mchugh et al., 2015). Currently the functions of these proteins are of growing interest in the field but much work still needs to be done.

X-chromosome Inactivation and health and disease*.*

Random silencing of one of the X-chromosomes in females results in a mosaic of X-linked gene expression. This functional hemizyous characteristic of X-chromosomes predisposes females to effects of heterozygous mutations in X-linked genes when XCI is skewed. If a disease causing mutation is on the X-chromosome that remains preferentially active, the phenotype can be more severe. On the contrary, if the mutation is on the preferentially inactive X-chromosome the phenotype can be less severe. In cases where XCI is random, the cells that express the wild type gene can lessen or masks the effects of the mutant gene.

Rett syndrome is a neurodevelopment disorder that exclusively effects females and is caused by a mutation in the X-linked MECP2 gene. The severity of the disease is tightly associated with the amount of XCI skewing that occurs. For example if the active X-chromosome in most of the cells carries the mutated version of *Mecp2*, the disease will be more severe. If the mutated *Mecp2* resides on the inactive X-chromosome in most cells, the phenotype will be less severe. In men, a mutated MECP2 gene results in death shortly after birth because they do not have a backup copy of the gene like females (Archer et al., 2007).

XCI is protective in X-chromosome aneuploidies. Extra X-chromosomes are largely compensated in males by XCI. Often cells will silence all but one X-chromosomes in X-chromosome aneuploidies. Deleterious effects by X-chromosome aneuploidies are typically caused by changes in expression of

"escape" genes. During XCI, subsets of genes are able to be expressed from both the active and inactive X-chromosome and escape XCI. In Klinefelter's (XXY) and trisomy X (XXX) males will have an increased expression of these escape genes. Male breast cancer has been linked to increases in X-linked gene expression in Klinefelter's patients (Rudas et al., 2000). Turner syndrome patients (XO) will have a dosage reduction of these escape genes (Tartaglia et al., 2010; Zinn and Ross, 2001).

Mutations in escape genes have been associated with cancer. Mutations in escape gene KDM6A have been associated with medulloblastoma, prostate cancer, and renal carcinomas (Deng et al., 2014). Many tumor suppressor and oncogenes have been identified as escape genes of XCI. Cells that carry an active oncogene or inactive tumor suppressor can be more prone to cancers developing in those tissues. Therefore differences in escape from XCI and XCI skewing can contribute to the phenotypic variances we see in X-linked gene based diseases.

Establishing XCI: Functions of the X-inactivation center, Xist, and interacting proteins during XCI induction

The X inactivation center (*Xic*) is identified as the minimum locus required for XCI and houses various lncRNAs including *Xist* and *Tsix* (Brown et al., 1991a; Brown et al., 1991b). Although much has been studied about the function of the *Xic*, the entirety of the full *Xic* is currently unknown. In eukaryotes, transcription is influenced by 3D genomic environment. Techniques such as HI-C incorporates

chemical crosslinking with chromatin, fragmentation, DNA ligation and hi-throughput DNA sequencing to identify neighboring loci genome wide. Using Hi-C, Nora *et al.* observed that the *Xic* comprises of two topologically associated domains (TAD) that are key to spreading of the lncRNA *Xist.* Genes on the *Xic* are classified as XCI activators and repressors. The respective promoters of *Xist* and *Tsix* are housed in different TADs. Activators of *Xist* which promotes XCI are housed in one TAD and activators of *Tsix* that promote repression of *Xist* and XCI are in a separate TAD suggesting the 3D architecture of the *Xic* is important in regulatory function (Nora et al., 2012).

Xist has 8 exons and is processed similar to mRNAs but retained in the nucleus. It is induced by differentiation and cell lineage specification in the epiblast cells of the blastocyst. Much of the transcriptional regulation of *Xist* itself and how it regulates XCI is currently still being studied. However, current studies propose the following model of *Xist* mediated induction of XCI. Prior to differentiation pluripotency genes Oct4/ Nanog/ Sox2 and CTCF bind to intron 1 and the regulatory elements of *Xist* inhibiting its expression. Upon differentiation, depletion of these pluripotency genes promotes expression of *Xist* (Navarro et al., 2008). Upon differentiation depletion of *Nanog* upregulates expression of the E3 ubiquitin ligase Rnf12 which then inhibits expression of REX1, a known repressor of *Xist* expression. *Jpx* and *Ftx* are present 10-100kb distal of the *Xic* (Chureau et al., 2002). *Jpx* is induced upon differentiation and removes CTCF from the promoter of *Xist* (Tian et al., 2010). Lack of *Ftx,* represses *Xist*

expression by ~20% but isn't necessary for XCI (Furlan et al., 2018). The combined actions of activators and repressors suggest the need of precise *Xist* expression patterns during induction of XCI.

Spliced *Xist* induces the silencing of X-linked genes through interactions of various RNA binding proteins and complexes with the tandem repeat elements found in the *Xist* transcript and the X-chromosome. Therefore a considerable amount of interest in the field has been in identifying the proteins that interact with these repeat regions and their effect on XCI. *Xist* houses 6 known tandem repeats A thru F. Loss of the A-repeat is associated with loss of the *Xist* mediated silencing but doesn't affect *Xist* localization. Proteomic and genetic screening has identified that the A-repeat binds to chromatin modifiers SPEN, RBM15, and WTAP. Further studies have found SPEN recruits the NCoR-HDAC3 complex responsible for deacetylating and opening the chromatin on the X-chromosome. RBM15 and WTAP are involved in depositing N^6-methyladenosine (m^6A) along *Xist* required for its function (Moindrot et al., 2015). Interactions of m^6A marks and the m^6A binding protein YTHDC1 to mediate *Xist* mediated gene silencing through an unknown mechanism (Patil et al., 2016).

Originally it was believed that the A-repeat was involved in recruitment of the polycomb repressive complexes 1 and 2 to the X-chromosome, however recent studies have observed that to not be the case. Work done by Pintacuda *et al.* identified that hnRNPK interactions with PRC1 and 2 at the B-repeat are responsible for initiating polycomb repressive complex recruitment to *Xist*

involved in *Xist* splicing. PRC1 and PRC2 catalyzes the H2AK119ub1 and H3K27me3 along the X-chromosome respectively (Pintacuda et al., 2017).

The E-repeat is found in exon 7 of *Xist* and has been associated with SET histone methyltransferase complex subunit ASH2L however loss of ASH2L resulted in upregulation of escape genes without effecting non escape genes. Sunwoo *et al.* found that loss of ASH2L didn't abolish enriched signal of *Xist* RNA on the Xi. Multiple groups have identified CIZ1 anchoring protein as also binding to the Xi, however this was shown in somatic cells and differentiated cell types (Ridings-Figueroa et al., 2017; Sunwoo et al., 2017). These data suggest that ASH2L may be important in regulation of escape genes during the maintenance phase of XCI but not necessarily XCI induction.

These studies have provided much needed insight on the mechanism of *Xist* mediated XCI and provided a useful approach to understanding lncRNA function in gene regulation. Identification of many RBPs that are direct and indirect binding partners of *Xist* have provided hints on the mechanism of *Xist* mediated silencing however many questions remain. Currently the function and nature of the interaction of these newly identified RBPs and *Xist* is currently unknown.

Transcriptional Regulation of Xist expression

LncRNAs are often processed similarly to mRNA transcripts of protein coding genes. Many lncRNAs are spliced, 5'capped, and 3' polyadenylated however very little is known on how splicing is accomplished. *Xist* is spliced very similarly

to mRNA transcripts however it is never exported out of the nucleus. Studies have observed that *Xist* is primarily unspliced in undifferentiated cells and that the spliced version of *Xist* is sufficient for *Xist* mediated XCI. However how *Xist* is spliced is unknown. *Xist* transcription produces different isoforms depending on promoter used, alternative polyadenylation sites, and alternative splicing (Ma et al., 2005; Memili et al., 2001; Johnston et al., 1998). Alternative splicing of intron 7 results in a long and short isoform of *Xist.* The long isoform is the dominant transcript involved in random XCI. However targeting cells to express primarily the short isoform showed no difference is ability to induce XCI (Yue and Ogawa, 2017).

Many groups have identified the splicing regulator PTBP1 as binding to the E-repeat in exon 7 of *Xist* (Vuong et al., 2016; McHugh et al., 2015; Chu et al., 2015). PTBP1 is involved in splicing, polyadenylation, and stability of mRNA transcripts suggesting a potential role in splicing of the *Xist*. Interaction between the A-repeat and splicing factor *ASF/SF2* is necessary for spliced RNA accumulation suggesting splicing efficiency of *Xist* might be one of the factors involved in X-chromosome choice to induce XCI (M. Royce-Tolland et al., 2010)

Conclusions

LncRNAs have varied functions in regulating gene expression. These studies have revealed that lncRNAs overall are important in every step of development. Currently the field has gained immense knowledge of how lncRNAs function on

their targets, however very little is known on how lncRNAs themselves are regulated. Understanding how lncRNAs are transcriptionally regulated can elucidate how they accomplish significant specificity. The mature *Xist* transcript is involved in XCI however most studies haven't studied how splicing of *Xist* is accomplished. It is well known that different mRNA transcripts have different affinities to their regulatory proteins; therefore it is possible a similar process is required for lncRNAs. Understanding how splicing of lncRNAs is controlled could provide insight into how the tissue and temporally specificity of lncRNA function arose.

To that end, this dissertation will focus on splicing regulation of *Xist* in ES cells. ES cells recapitulate XCI *in vitro* and will allow us to monitor *Xist* transcript regulation. In chapter two I will investigate the role of differentiation induced splicing and provide insight into its temporal regulation. In chapter three, we explore the effects of known splicing regulator PTBP1 on *Xist* splicing and XCI. One of the ongoing questions of XCI is how the cell knows which X-chromosome to silence. In chapter four I will focus on the potential role splicing may have in allelic choice of X-chromosome to silence using *Mus musculus 129SvJ / Mus castaneous (CAS)* ES cells known to display skewed X-chromosome inactivation.

References

Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* (2013). doi:10.1038/nature12349

Archer, H. *et al.* Correlation between clinical severity in patients with Rett syndrome with a p.R168X or p.T158M MECP2 mutation, and the direction and degree of skewing of X-chromosome inactivation. *J. Med. Genet.* (2007). doi:10.1136/jmg.2006.045260

Barr, M. L. & Bertram, E. G. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis [2]. *Nature* (1949). doi:10.1038/163676a0

Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* (1991). doi:10.1038/349038a0

Brown, C. J. *et al.* Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* (1991). doi:10.1038/349082a0

Cabili, M. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* (2011). doi:10.1101/gad.17446611

Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* (2015). doi:10.1016/j.cell.2015.03.025

Chureau, C. *et al.* Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum. Mol. Genet.* (2011). doi:10.1093/hmg/ddq516

Chureau, C. *et al.* Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res.* (2002). doi:10.1101/gr.GR-1529R

Deng, X., Berletch, J. B., Nguyen, D. K. & Disteche, C. M. X chromosome regulation: Diverse patterns in development, tissues and disease. *Nature Reviews Genetics* (2014). doi:10.1038/nrg3687

Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* (2012). doi:10.1101/gr.132159.111

Furlan, G. *et al.* The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.03.024

Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* (2009). doi:10.1038/nature07672

Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. in *Nature Genetics* (2011). doi:10.1038/ng.848

Jiang, C. *et al.* Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget* (2016). doi:10.18632/oncotarget.6859

Johnston, C. M. *et al.* Developmentally regulated Xist promoter switch mediates initiation of X inactivation. *Cell* (1998). doi:10.1016/S0092-8674(00)81739-0

Kong, Y., Hsieh, C. H. & Alonso, L. C. ANRIL: A lncRNA at the CDKN2A/B locus with roles in cancer and metabolic disease. *Front. Endocrinol. (Lausanne).* **9,** 1–13 (2018).

Kornienko, A. E., Guenzl, P. M., Barlow, D. P. & Pauler, F. M. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* **11,** 1 (2013).

Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: Past, present, and future. *Genetics* (2013). doi:10.1534/genetics.112.146704

Lee, J. T. The X as model for RNA's niche in epigenomic regulation. *Cold Spring Harbor perspectives in biology* (2010). doi:10.1101/cshperspect.a003749

Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* (2010). doi:10.1038/ng.710

Lyon, M. F. Gene action in the X-chromosome of the mouse (mus musculus L.). *Nature* (1961). doi:10.1038/190372a0

Ma, M. & Strauss, W. M. Analysis of the Xist RNA isoforms suggests two distinctly different forms of regulation. *Mamm. Genome* (2005). doi:10.1007/s00335-004-2464-3

Marquardt, S. *et al.* A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* (2014). doi:10.1016/j.cell.2014.04.036

Mattick, J. S. The central role of RNA in human development and cognition. *FEBS Letters* (2011). doi:10.1016/j.febslet.2011.05.001

McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* (2015). doi:10.1038/nature14443

Memili, E., Hong, Y. K., Kim, D. H., Ontiveros, S. D. & Strauss, W. M. Murine Xist RNA isoforms are different at their 3′ ends: A role for differential polyadenylation. *Gene* (2001). doi:10.1016/S0378-1119(01)00353-5

Mohammad, F., Mondal, T. & Kanduri, C. Epigenetics of imprinted long noncoding RNAs. *Epigenetics : official journal of the DNA Methylation Society* (2009). doi:10.4161/epi.4.5.9242

Moindrot, B. *et al.* A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Rep.* (2015). doi:10.1016/j.celrep.2015.06.053

Navarro, P. *et al.* Molecular coupling of Xist regulation and pluripotency. *Science (80-.).* (2008). doi:10.1126/science.1160952

Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* (2012). doi:10.1038/nature11049

Ohno, S. So much “junk” DNA in our genome. *Brookhaven Symp. Biol.* (1972). doi:citeulike-article-id:3483106

Patil, D. P. *et al.* M6 A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* (2016). doi:10.1038/nature19342

Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell* **68,** 955–969.e10 (2017).

Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. *Xist* RNA and the Mechanism of X Chromosome Inactivation. *Annu. Rev. Genet.* (2002). doi:10.1146/annurev.genet.36.042902.092433

Puvvula, P. K. *et al.* Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat. Commun.* (2014). doi:10.1038/ncomms6323

Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics* (2016). doi:10.1038/nrg.2015.10

Ridings-Figueroa, R. *et al.* The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* (2017). doi:10.1101/gad.295907.117

Royce-Tolland, M. E. *et al.* The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nat. Struct. Mol. Biol.* **17,** 948–954 (2010).

Rudas, M. *et al.* Karyotypic findings in two cases of male breast cancer. *Cancer Genet. Cytogenet.* (2000). doi:10.1016/S0165-4608(00)00254-5

Sahakyan, A. *et al.* Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* (2017). doi:10.1016/j.stem.2016.10.006

Schlackow, M. *et al.* Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell* **65,** 25–38 (2017).

Sigova, A. A. *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci.* **110,** 2876–2881 (2013).

Strome, S., Kelly, W. G., Ercan, S. & Lieb, J. D. Regulation of the X chromosomes in Caenorhabditis elegans. *Cold Spring Harb. Perspect. Biol.* (2014). doi:10.1101/cshperspect.a018366

Sun, S. *et al.* XJpx RNA activates xist by evicting CTCF. *Cell* (2013). doi:10.1016/j.cell.2013.05.028

Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl. Acad. Sci.* (2017). doi:10.1073/pnas.1711206114

Tartaglia, N., Cordeiro, L., Howell, S., Wilson, R. & Janusz, J. The spectrum of the behavioral phenotype in boys and adolescents 47,XXY (Klinefelter syndrome). *Pediatric endocrinology reviews : PER* (2010). doi:10.1002/nbm.3066.Non-invasive

Thakur, P. K. *et al.* Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5? splice-site sequences due to weak interactions with SR proteins. *Work. Aging Retire.* (2018). doi:10.11821/dlxb201802008

Thakur, P. K. *et al.* Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5? splice-site sequences due to weak interactions with SR proteins. *Work. Aging Retire.* (2018). doi:10.11821/dlxb201802008

Thomas, C. A. The Genetic Organization of Chromosomes. *Annu. Rev. Genet.* (1971). doi:10.1146/annurev.ge.05.120171.001321

Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, Jpx, Is a molecular switch for X chromosome inactivation. *Cell* (2010). doi:10.1016/j.cell.2010.09.049

Ulitsky, I. & Bartel, D. P. XLincRNAs: Genomics, evolution, and mechanisms. *Cell* **154,** 26–46 (2013).

Ulitsky, I. *et al.* XDivergent transcription: A driving force for new gene origination? *Cell* **155,** 990–996 (2013).

Vuong, J. K. *et al.* PTBP1 and PTBP2 Serve Both Specific and Redundant Functions in Neuronal Pre-mRNA Splicing. *Cell Rep.* (2016). doi:10.1016/j.energy.2018.10.173

Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* (2011). doi:10.1016/j.molcel.2011.08.018

Wu, P. *et al.* Roles of long noncoding RNAs in brain development, functional diversification and neurodegenerative diseases. *Brain Research Bulletin* (2013). doi:10.1016/j.brainresbull.2013.06.001

Wu, X. & Sharp, P. A. XDivergent transcription: A driving force for new gene origination? *Cell* (2013). doi:10.1016/j.cell.2013.10.048

Wu, X. & Sharp, P. A. XDivergent transcription: A driving force for new gene origination? *Cell* **155,** 990–996 (2013).

Wu, X. & Sharp, P. A. XDivergent transcription: A driving force for new gene origination? *Cell* **155,** 990–996 (2013).

Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* (2002). doi:10.1038/ng820

Yue, M. *et al.* Xist RNA repeat E is essential for ASH2L recruitment to the inactive X and regulates histone modifications and escape gene expression. *PLoS Genet.* (2017). doi:10.1371/journal.pgen.1006890

Yue, M. & Ogawa, Y. CRISPR/Cas9-mediated modulation of splicing efficiency reveals short splicing isoform of Xist RNA is sufficient to induce X-chromosome inactivation. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx1227

Zinn, A. R. & Ross, J. L. Molecular analysis of genes on Xp controlling Turner syndrome and premature ovarian failure (POF). *Seminars in Reproductive Medicine* (2001). doi:10.1055/s-2001-15394

Chapter 2: Xist splicing and differentiation

Abstract

X-inactive specific transcript (*Xist*) is a long noncoding RNA that is essential to the process of X-chromosome inactivation. Random XCI is mediated by the upregulation of female specific *Xist* from the future inactive X-chromosome (Xi) upon differentiation induction. XCI cannot be induced in the absence of *Xist*. *Xist* is processed similarly to protein coding gene transcripts, however never leaves the nucleus. Previous studies have primarily used the spliced *Xist* to induce XCI suggesting a potential role of splicing in XCI. Currently the mechanism of splicing of lncRNAs remains a mystery. We found that *Xist* RNA is primarily unspliced in naÏve ES cells. Upon differentiation *Xist* splicing becomes efficient across all exons independent of transcription. Splicing of *Xist* happens post-transcriptionally suggesting one of the checkpoints of XCI induction is splicing of *Xist* once expression of *Xist* is upregulated to a certain level. These findings suggest differentiation induced splicing is one of the checkpoints required to induce XCI.

Introduction

During early development, female placental animals compensate for the genetic copy number differences of X-linked genes between males (XY) and females (XX), by silencing one of the X-chromosomes in females through the process of X-chromosome inactivation (XCI) (Lyon, 1961). In mice this chromosome-wide silencing process is a two-phase process, that is cell lineage specific, of imprinted and random XCI (Hyunh and Lee, 2003; Okamoto et al., 2004). For either random or imprinted XCI to occur, the 17-kb long non-coding RNA (lncRNA) X-inactive specific transcript *Xist* is expressed *in cis* from the targeted X-chromosome (Borsani et al., 1991; Brockdorff et al., 1991; Brown et al., 1991). Accumulation of the *Xist* on the future inactive X-chromosome has been correlated with the initiation of XCI on that same chromosome. Studies in human and mouse has named a unique *cis*-acting region of the X-chromosome name the X-inactivation center (*Xic*) as the main controller of the expression of *Xist* and induction of XCI (Augui et al., 2011).

Immediately upon formation of the zygote, gene expression is bi-allelic and expressed from both of the X-chromosomes. At the two-to-four cell stage, *Xist* is expressed exclusively from the paternal X-chromosome (Xp) (Hyunh and Lee, 2003; Okamoto et al., 2004). Upon expression, *Xist* coats the Xp *in cis* from the 4-cell stage onward and progressively initiates imprinted XCI (iXCI) exclusively on the paternal X-chromosome (Xp) (Maharens et al., 1997; Kay et al.,1994). Upon *Xist* coating, the paternal X-chromosome is silenced with the

removal of methylated H3K4 and acetylated H3K9 activation marks and the depositing of repressive H3K27me3 by polycomb complexes (Mager et al., 2003; Wang et al., 2001). This imprinted XCI is then maintained in cells fated to become the trophoectoderm and the primitive endoderm.

Upon implantation of the zygote, X-reactivation of the Xp occurs in cells of the inner cell mass of the blastocyst (Mak et al., 2004; Okamoto et al., 2004). It is believed that the initial step of X-chromosome reactivation is the repression of *Xist* on the Xp followed by the loss of these Xp localized repressive complexes. Further silencing is accomplished by the loss of repressive histone marks along the Xp and reactivation of paternal X-linked genes. The second reactivation event, random XCI, occurs upon implantation (Chaumeil et al., 2006; Clemson et al., 2006). After implantation and when epiblast cells begin primordial germ cell lineage specification, random XCI takes place. In epiblast cells both X-chromosomes are active and the maternal (Xm) or paternal (Xp) X-chromosome are stochastically inactivated (Monkhorst et al., 2008). Once random XCI occurs, it is maintained throughout life.

Random XCI is initiated by the upregulation of *Xist* from the future inactive X-chromosome (Xi). In both humans and mice, overwhelming evidence suggests that *Xist* RNA coats the Xi *in cis* to recruit epigenetic silencing factors involved in silencing the X-chromosome (Chaumeil et al., 2002; de Napoles et al., 2004; Fang et al., 2004; Plath et al., 2004). The consensus view is that *Xist* induction is necessary for both random and imprinted XCI during development.

X-inactivation occurs during the earliest differentiation steps in embryonic development with very little amount of cells available at this stage. Embryonic stem cells, pluripotent stem cells derived from the inner cell mass, are a useful *in vitro* system for studying X-inactivation. Undifferentiated mouse embryonic stem (ES) cells can be maintained in an undifferentiated state and differentiation can be induced easily *in vitro*. Undifferentiated ES cells exhibit two active X-chromosomes while they are pluripotent. Mimicking embryonic development, females ES cells recapitulate random XCI when stimulated to differentiate (Leahy et al., 1999; Monk M, 1981).

Xist expression is one of the earliest known events of X-chromosome inactivation. Differentiated ES cells initiate expression of *Xist* RNA from the future Xi chromosome followed by further steps to fully establish XCI (Sun et L., 2006; Panning et al., 1996). This initiation step involves counting and choice of the X-chromosome to be silenced however how this choice is made is currently unknown. Using the embryoid body method, *Xist* spreads and coats the entire Xi *in cis* as early as 48 hours after induction of differentiation. After 72 hours, continued silencing doesn't require *Xist* and many characteristics of X-inactivation have been acquired (Wutz et. al., 2000). Much later during differentiation, chromatin is further modified as XCI becomes fully established and irreversible. Maintenance of XCI after this point appears to be independent of *Xist* RNA expression. Many studies have suggested various transcriptional controls of the initial *Xist* upregulation (Mak et al., 2004; Okamoto et al., 2004).

However, although *Xist* is processed and spliced, the contribution of post-transcriptional regulation is still not well characterized.

The current model of random XCI is that before XCI initiates, unstable *Xist* RNA is expressed and an unknown blocking factor prevents *Xist* upregulation and its interaction with the Xi *in cis*. Upon differentiation, *Xist* becomes upregulated through stabilization, transcriptional upregulation, and/or release of this blocking factor (Panning and Jaenisch, 1998). Stabilized *Xist* then coats the Xi *in cis* followed by recruitment of chromatin modifications of the Xi (Avner and Heard, 2001). Currently how this choice is made and what initiates this choice is unknown. Additionally, mature *Xist* is spliced, however the role of splicing of *Xist* if any is currently unknown.

After transcription, most lncRNAs are processed very similar to protein-coding genes. Most lncRNAs are 5'capped, polyadenylated, and undergo splicing, however are not always transported to the cytoplasm. Why lncRNAs are spliced is currently unknown. Messenger RNA splicing usually allows one transcript to make multiple variants and promotes nuclear export to the cytoplasm but neither apply to *Xist* RNA that is retained in the nucleus (Jonkers et al., 2008).

In UCSC annotation, *Xist* consists of 8 exons. Retaining and deleting of the last intron creates the long and short isoforms respectively. The long isoform is the major isoform and the short isoform is expressed lowly and only in a couple differentiated tissues (Ma and Strauss, 2005). Dominant expression of the short

isoform using CRISPR-Cas9 was not distinguishable from the long isoform at activating XCI (Yue and Ogawa, 2017). Therefore it is unclear what role, if any, introns have in *Xist* regulation. Notably human and mouse *Xist* is only 47% similar but share exon-intron conservation suggesting selection pressure to maintain *Xist* splicing (Hendrich, Brown, and Willard, 1993). We asked whether *Xist* RNA splicing could be a regulatory checkpoint of *Xist* biogenesis. Surprisingly we found differentiation drastically increased *Xist* RNA splicing efficiency in C57/BL6 ES cells (or BL6 ES cells) or F1 2-1 ES cells (a hybrid of *Mus musculus 129SvJaeJ* and *Mus musculus castaneus EiJ*, which has been widely used to study allelic expression of *Xist* and XCI).

Results

ES differentiation induces Xist splicing

After induction of *Xist* expression, *Xist* is 5'- end capped, 3'-polyadenylated, and spliced to produce the functional mature *Xist* transcript. Barbara Panning's group showed that alternative splicing factor 2 / splicing factor 2 (ASF/SF2) binds to the A-repeat of *Xist* and loss of the A-repeat correlated with less processed *Xist* in male ES cells. It has been shown that XCI fails to initiate on the X-chromosome that lacks the A-repeat suggesting that splicing may be a regulatory step for *Xis*t induction and function (Royce-Tolland et al., 2010). To investigate the events of initial *Xist* induction, RT-qPCR quality primers were designed to measure total, spliced, and unspliced *Xist* transcripts. For total *Xist*, primers were designed to target within the exon. By targeting the exon, the primers are not able to distinguish between spliced and unspliced transcripts. For unspliced transcripts, the primer pair anneals to a specific splice site hybridizing to the intron and exon respectively. In spliced transcripts these intron primers will fail to bind, allowing preferential amplification of only unspliced *Xist* transcripts. For the spliced transcripts, forward and reverse primers anneal to separate exons separated by a large intron. We used multiple qualified pairs to increase the robustness of the analysis.

To investigate the initial events of *Xist* inactivation, we used strand-specific RT-qPCR to measure the expression of spliced and unspliced transcripts after 24 hours of retinoic acid induced differentiation. Using exon 1 and exon 4

targeted primers, I observed that total *Xist* transcripts were increased after 24 hours of differentiation of female wild type F1 2-1 ES cell in monolayer culture (Figure 1A). This observation is in agreement with the increase of *Xist* expression upon initiation of XCI. Using primers targeting various exon junctions spanning *Xist,* I observed a significant and consistent increase in the spliced *Xist* transcript across exon 1 - exon 2, exon 3 - exon 4, and exon 5 - exon 6 (Figure 1A). This increase in spliced mature *Xist* transcripts could be attributed to both splicing and RNA decay. In contrast, I observed a modest but significant decrease in unspliced transcripts. As transcription continues during expression of *Xist*, this modest decrease could be attributed to both continuing transcription of new *Xist* transcripts and splicing of already made transcript (Figure 1A).

An increase in spliced *Xist* could be attributed to the increase of total *Xist* transcripts upon differentiation induction in addition to splicing. To mask the transcription effect and solely look at splicing, I normalized the amount of spliced transcripts to the total *Xist* transcripts before and after differentiation After normalization to the total, the increase of spliced *Xist* was ~2-fold (Figure 1B) while the amount of unspliced transcripts was reduced ~3-fold (Figure 1C).

Enhanced splicing could explain for the increase in spliced transcript and the decreased in unspliced transcript. To directly test whether splicing is induced by differentiation, primers flanking intron 2 (the smallest intron of *Xist*) to simultaneously detect both unspliced and spliced transcripts (Figure 1D). Intron 2 is only 148nt, so both the unspliced and spliced transcripts can be detected. The

detection of both transcripts allows us to calculate the splicing ratio (spliced/unspliced) as a measurement of splicing efficiency. I found that differentiation induced a significant increase in splicing efficiency after 24hrs of differentiation in female F1 2-1 ES cells (Figure 1D).

To determine if differentiation induced splicing occurs across other introns, I focused on intron 1. Due to the length of intron 1 (2794nt), we were unable to find an effective primer flanking the exon 1 - exon 2 of *Xist.* Therefore we designed forward primers to bind to exon 1 and upstream of the intron 1 - exon 2 boundary along with a reverse primer to anneal to exon 2. By multiplex PCR using these primers we are able to amplify both the spliced (74nt) and unspliced (127nt) *Xist* transcripts. I found that in undifferentiated conditions *Xist* is primarily in the unspliced form. However I observed enhanced splicing consistently after 24 hours differentiation (Figure 1E).

I next wanted to determine if differentiation induced splicing was regulated by individual introns or was synchronized across all introns. Long introns can be processed co-transcriptionally and stepwise before exons are joined.(reviewed in Georgomanolis, Sofiadis, and Papantonis, 2016). Therefore I wanted to see if each flanking exons are spliced concurrently or does splicing occur at different rates. If splicing occurs across different introns at different times, splicing would lead to the formation of many isoforms whose introns were not fully removed. In contrast, coordinated splicing would result in either fully unspliced or fully spliced transcripts. It is also possible that some transcripts would have some introns but

not all removed. To determine which of these possibilities was the case for *Xist*, we used primers flanking multiple exons to detect *Xist* transcripts.

My results were consistent to what was seen across individual introns. In both undifferentiated and differentiated conditions, transcripts were either fully unspliced or fully spliced suggesting that splicing of introns is synchronized (Figure 1F). Because PCR efficiency is skewed to smaller PCR products in long-range PCR reactions, the smaller spliced PCR product (562nt) is over represented in our analysis.

An increase in spliced transcript could be the result of two possibilities. The first possibility is that splicing efficiency increases upon differentiation. However another possibility is that once the transcript is spliced, it becomes more stable. If the unspliced transcripts are unstable and have higher rates of degradation, that could skew the spliced/unspliced ratio of transcripts in favor the spliced population. To test for the stability of the *Xist* transcript, I measured the half-life of the spliced and unspliced transcript after treatment with 5ug/ml actinomycin D. My data showed that there were no significant differences in the stability of both the spliced and unspliced *Xist* after differentiation (Figure 2A-B). This data supports the first possibility that the increase in splicing is due to the increased splicing efficiency once differentiation was induced. Our data suggests that prior to differentiation, a population of *Xist* transcript is expressed however remains unspliced.

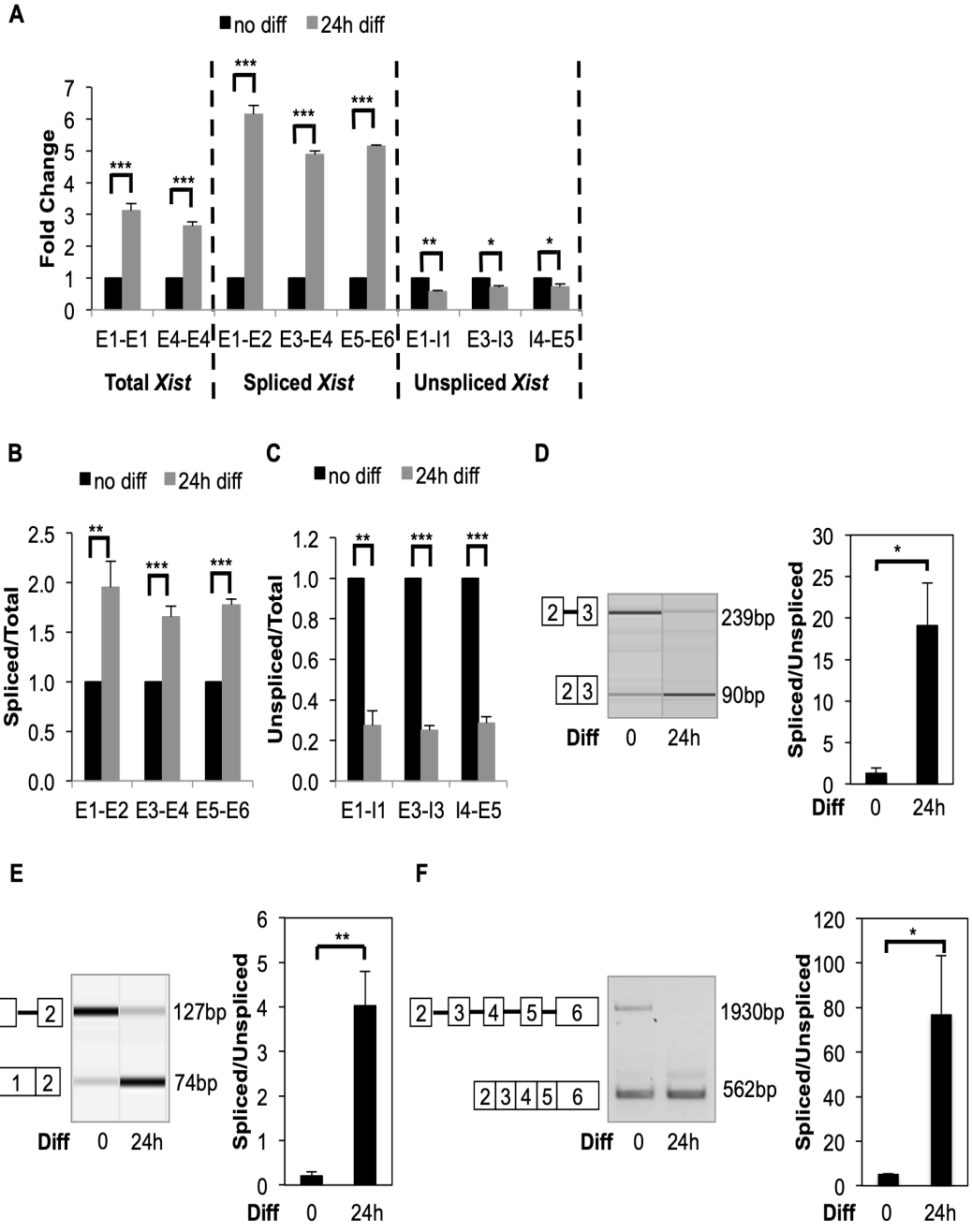
A
no diff
24h diff
Fold Change
E1-E1
E4-E4
E1-E2
E3-E4
E5-E6
E1-I1
E3-I3
I4-E5
Total Xist
Spliced Xist
Unspliced Xist
B
Spliced/Total
C
Unspliced/Total
D
239bp
90bp
Diff
0
24h
Spliced/Unspliced
E
127bp
74bp
F
1930bp
562bp

Figure 1. Differentiation induces *Xist* splicing (A) RT-qPCR analysis of total, spliced, and unspliced Xist transcripts. Adherent F1 2-1 cells were differentiated by LIF withdrawal and addition of RA. (B) Ratios of spliced/total Xist transcripts quantified by RT-qPCR. (C) Ratios of unspliced/total Xist transcripts quantified by RT-qPCR. (D) Splicing assay examines intron 2 splicing before and 24 h after differentiation. Representative virtual gel image of intron 2 splicing (left panel). Quantified intron 2 splicing ratio (right panel) (E) Splicing assay examines intron 1 splicing before and 24 h after differentiation. Representative virtual gel image of intron 1 splicing (left panel). Quantified intron 1 splicing ratio (right panel). (F) Splicing assay examines splicing from exon 2 to exon 6 before and 24 h after differentiation. Representative gel image (left panel). PCR cycles were 33 and 29 for undifferentiated and differentiated cells respectively. For differentiated cells, PCR amplification for 33 cycles did not produce the unspliced product and the spliced product was over saturated, so amplification of 29 cycles were used for the presentation. Quantified splicing ratio (right panel). Error bars indicate S.E.M. of three biological replicates. *P-value<0.05; **P-value<0.01; ***P-value<0.001 Student's t-test (one-tailed, unpaired).

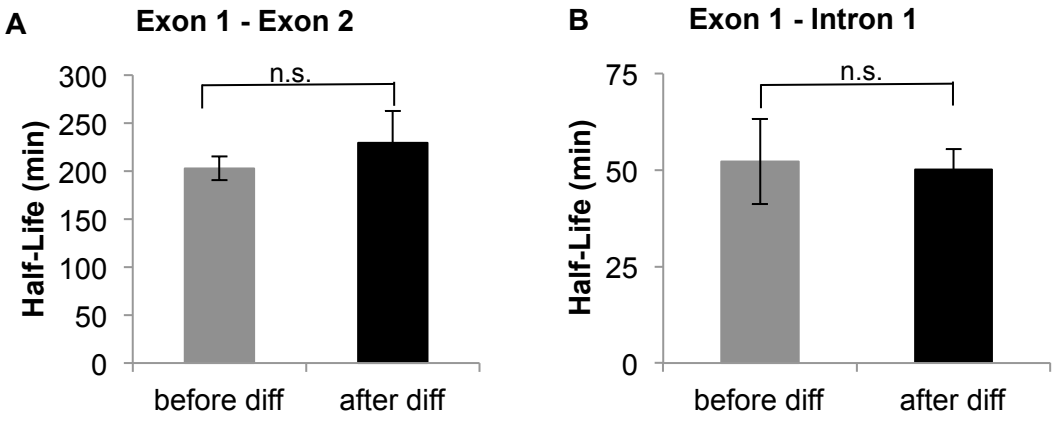

Figure 2. Differentiation has no effect of *Xist* RNA stability RNA half-lives of A) spliced and B) unspliced *Xist* transcripts before differentiation and 24 hours after differentiation. Error bars indicate S.E.M of three biological replicates. ns, p-value>0.05.

Upon differentiation of the ES cells, *Xist* continues to be expressed at a higher rate on the Xi as splicing of the transcript increases. Once transcription of *Xist* ramps ups during the early stages of random XCI and splicing of the transcript becomes more efficient on the Xi, *Xist* expression is repressed on the active X-chromosome. It is known that skewing of X-inactivation occurs in endogenous conditions; therefore it is possible that splicing efficiency differences between different alleles of *Xist* could be one of the regulatory steps of *X* chromosome inactivation induction.

Enhanced Splicing is independent of Xist transcription

As mentioned previously, prior to differentiation *Xist* is transcribed at low levels from both X-chromosomes, however is predominantly unspliced. Therefore one of the unanswered questions regarding *Xist* is whether these initial transcripts are being spliced or if only newly synthesized *Xist* induced by differentiation are being spliced. This process can be tested by inhibiting transcription of *Xist* at the initial induction of splicing and determining the splicing rates of the transcript. To do this I first needed to identify at what point is splicing initiated after differentiation is induced. Using RT-qPCR I monitored the splicing ratio of Exon 1- Exon 2 and Exon 2 - Exon 3 each hour post differentiation induction to determine when splicing occurs. My data shows that splicing is low between 1 and 6 hours post differentiation with retinoic acid and ramps up between 7 and 8 hours post induction (Figure 3A). After 7 and 8 hrs of induction, splicing steadily

increased until 22 hours post induction. This data suggests that enhancement of splicing must occur between 7 and 8 hours of differentiation.

To determine if this increase is splicing was processing of already transcribed *Xist,* we inhibited transcription with actinomycin D at 6 hrs post differentiation induction with retinoic acid (before splicing enhancement) and analyzed splicing at 8 hours and 10 hours of differentiation (Figure 3B). If scenario 1 is correct and only newly transcribed *Xist* is spliced (co-transcription), then actinomycin D should prevent the splicing of *Xist.* If scenario 2 is correct, and already made *Xist* is spliced, actinomycin D shouldn't have an effect on splicing ratios of *Xist* (post-transcription). In scenario 2, we should see an increase in spliced transcripts after actinomycin D inhibition since already present nascent transcripts are still being processed.

I found that the amount of spliced transcripts was unaffected by actinomycin D, however the amount of unspliced transcripts diminished at 8 hours and 10 hours when compared to before actinomycin D treatment (Figure 3C and D). These data suggests scenario 2 that pre-existing unspliced *Xist* becomes spliced once splicing enhancement occurs at 7 hours post differentiation.

To quantitatively measure the magnitude of expression changes of unspliced and spliced *Xist* RNA individually, I performed RT-qPCR .

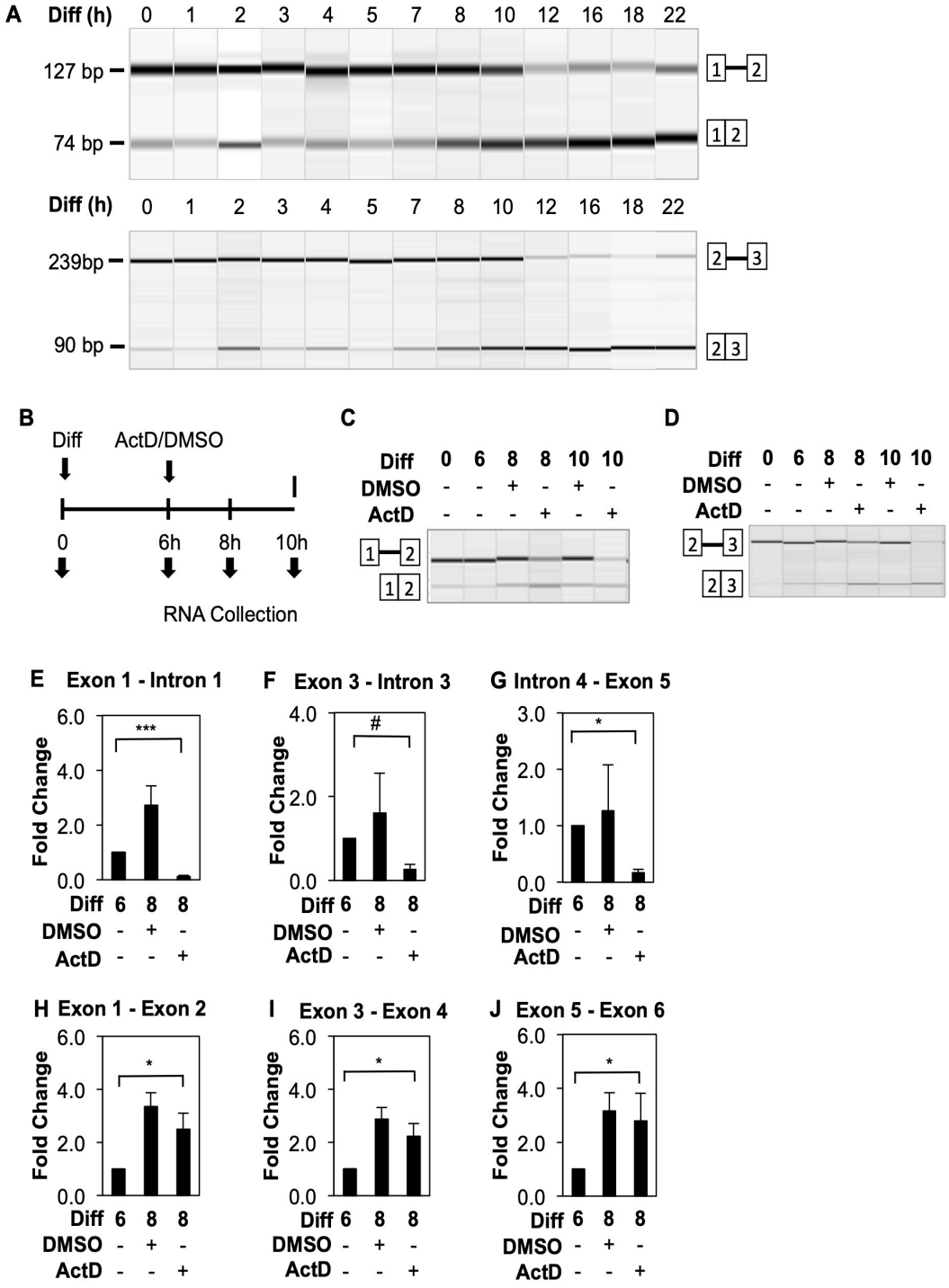
A
Diff (h) 0 1 2 3 4 5 7 8 10 12 16 18 22
127 bp
74 bp
1 2
Diff (h) 0 1 2 3 4 5 7 8 10 12 16 18 22
239bp
90 bp
2 3
B
Diff
ActD/DMSO
0
6h
8h
10h
RNA Collection
C
Diff 0 6 8 8 10 10
DMSO - - + - + -
ActD - - - + - +
D
Diff 0 6 8 8 10 10
DMSO - - + - + -
ActD - - - + - +
E Exon 1 - Intron 1
F Exon 3 - Intron 3
G Intron 4 - Exon 5
H Exon 1 - Exon 2
I Exon 3 - Exon 4
J Exon 5 - Exon 6
Fold Change
Diff 6 8 8
DMSO - + -
ActD - - +

Figure 3. Enhanced *Xist* splicing is independent of *Xist* transcription
(A) Representative RT-PCR virtual gel images of intron 1 splicing (*top* panel) and intron 2 splicing (*bottom* panel) in adherent F1 2-1 cells between 0-22 h post differentiation (RA induction and LIF withdrawal). (B) Schematic diagram of experimental procedures. Cells were differentiated for 6 hours before treatment with 5ug/ml Actinomycin D or control DMSO. RNA was collected before differentiation (time 0), at 6 hours post differentiation induction (time 6), at 2, and 4 hours post treatment (time 8 and time 10). (C) Representative RT-PCR virtual gel image of intron 1 splicing following the aforementioned experimental procedures. (D) Representative RT-PCR virtual gel image of intron 2 splicing following the aforementioned experimental procedures. (E-G) RT-qPCR analysis of unspliced *Xist* transcripts using primers detecting various introns. (H-J) RT-qPCR of spliced *Xist* transcripts using primers for various exon-exon junctions. Error bars indicate S.E.M of three biological replicates. #, p-value=0.06; *, p-value<0.05 ; ***, p-value<0.001 Students t-test (1-tailed, unpaired).

Compared to 6 h post differentiation, unspliced *Xist* stayed relatively flat after DMSO treatment, however decreased significantly (~4-fold) after 2h of treatment with actinomycin D (Figure 3 E-G). In contrast, spliced transcripts remained upregulated even after transcription inhibition across multiple exons (Figure 3H-J).

The above results indicate that *Xist* splicing occurs post-transcriptionally, because most previously transcribed unspliced transcripts have completed their transcription and have been spliced while there appears to be no change in splicing after transcription is inhibited by actinomycin D. Additionally supporting post-transcriptional splicing of *Xist* was data concerning splicing of the upstream intron 1 and downstream introns 3, -4, and -5 being consistent with one another (Figure 3 E-J). One of the additional steps of post-transcriptional processing is polyadenylation of the 3'end. To further test if splicing is post-transcriptional, I looked at the splicing changes of polyadenylated (poly(A)+) RNA fractions. I found in undifferentiated cells most of polyadenylated *Xist* was unspliced, however splicing of these poly(A)+ *Xist* transcripts was enhanced upon differentiation (Figure 4A-B).

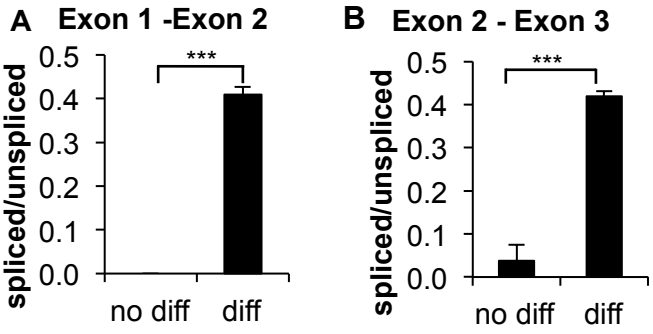

Figure 4. Differentiation induces splicing of poly(A)+ *Xist* transcripts (A) Intron 1 splicing ratio of polyadenylated *Xist* transcripts. (B) Intron 2 splicing ratio of polyadenylated *Xist* transcripts. Error bars indicate S.E.M of three biological replicates. ***, p-value<0.001 Students t-test (1-tailed, unpaired).

Discussion

In this chapter we examined the early steps of X-chromosome Inactivation by focusing on the lncRNA *Xist* during differentiation of female ES cells. The goal of this chapter was to first identify if differentiation effects splicing of *Xist.* Although it is known that *Xist* is spliced during XCI, *Xist* splicing regulation hasn't been studied in depth. Differentiation of ES cells mimics the steps of random XCI during embryonic development and serves as a model of monitoring *Xist* regulation *in vitro*. Using strand-specific RT-qPCR we are able to monitor *Xist* total, unpsliced, and spliced transcripts separately and monitor their levels in naive and differentiated conditions.

Our measurement of the total *Xist* transcripts using exon specific primers in naive WT ES cells showed that in naive conditions *Xist* is expressed at lower levels and is primarily unspliced. During early stages of differentiation with retinoic acid, we see a 2-3 fold increase in total *Xist*. Additionally, measurements across multiple exons of *Xist* shows that when normalized to total *Xist*, spliced transcripts are significantly increased in early differentiation and unspliced transcripts are significantly decreased. The observation that the *Xist* transcript is primarily only spliced once differentiation occurs suggest the presence of splicing elements that are sensitive to differentiation status. Downregulation of pluripotent factors (i.e. Nanog, Oct 3/4, Sox2) have been shown as a requirement of X-chromosome inactivation initiation (Sousa et al., 2018; Minkovsky, 2012; Navarro et al., 2008). There is a possibility that these

pluripotency factors have a role in regulating yet to be identified factors involved in the inhibition and activation of splicing. Ciaudo *et al.* found that loss of mRNA degradation regulators *Eif1* and *Rent1* resulted in inhibition of differentiation induced *Xist* upregulation and inhibition of XCI (Ciaudo et al., 2006). Upon deeper analysis they found that unspliced levels *Xist* remained similar to WT in *Eif1* and *Rent1* however differentiated induced spliced *Xist* was reduced further suggesting a connection between splicing and XCI.

Since the identification of *Xist,* much has been studied regarding its expression, activation, and how it coats the X-chromosome. Although it is known that *Xist* is 5'-capped, 3' polyadenylated, and undergoes alternative splicing, splicing is often overlooked. Primarily splicing of protein-coding transcripts is necessary for export out of the nucleus, however *Xist* is retained in the nucleus. Splicing of lncRNA *Blustr* is required for activation of neighboring gene *Sfmtb2* (Engreitz et. al.). Possible reasons very few studies on splicing are done is because i) *Xist* was identified as spliced cDNA ii) spliced *Xist* is sufficient to induce XCI iii) methods for *Xist* detection have been designed to detect spliced *Xist* and iv) *Xist* splicing thought to occur rapidly and passively.

PCR efficiency is skewed to smaller amplicons due to ease of amplification, therefore when measuring spliced and unspliced transcripts, consideration should be taken to ensure that the amplicons are close in size. In cases of large differences in amplicon sizes, there is an underlying bias to amplify the smaller amplicon. To measure splicing accurately we examined

intron 2 which yielded amplicon sizes 239 and 90 bp. (Figure 1D). For analyzing intron 1, I used multiplex PCR to amplify similarly sized amplicon (unspliced 127bp vs spliced 74bp (Figure 1E). The splicing across intron 2- intron 5 (exon 2 - exon 6 , Figure 1F) demonstrates this bias. When comparing the splice ratios after differentiation, we see an increase across exons, however the increase is significantly higher across exon 2 to exon 6, possibly due to bias of amplifying the 562nt spliced amplicon as compared to the much larger 1930nt unspliced amplicon.

Splicing can occur individually per intron or coordinated across all of our introns. In the scenario that splicing occurs individually there could possibly be isoforms that aren't fully spliced. Our data shows that *Xist* is either completely spliced or unspliced. When looking across various exon regions, only a completely spliced or the completely unspliced isoform is amplified. This suggests that splicing occurs coordinately across exons.

In our data we observed that the unspliced fraction of *Xist* decreased after differentiation. This could be due to two possibilities: i) unspliced transcripts are more easily degraded or ii) differentiation induces splicing of the present transcripts. Consistent with scenario 1, stability of spliced transcripts on the Xi and instability on the Xa has been suggested as one of the essential steps of *Xist* continued expression and induction of XCI. (Panning, Dausman, Jaenisch, 1997). In ES cells treated with actinomycin D, we saw no differences in stability between unspliced and spliced before and after differentiation transcripts (Figure

2A-B). Supporting scenario 2, our data suggests that differentiation induced splicing of present and new transcripts results in the increase in splicing ratio.

Prior to differentiation *Xist* is expressed at low levels from both active X-chromosomes in the inner cell mass of the epiblast. Splicing can occur concurrently with transcription (co-transcription) of after transcription has been completed (post-transcription). One question in the field is whether differentiation induces splicing of *Xist* or only newly transcribed *Xist* is spliced. Inhibition of *Eif1* and *Rent1* only reduced spliced Xist in differentiated cells while unspliced *Xist* levels remained the same as WT (Ciaudo et al., 2006). Cells with reduced levels of differentiation induced spliced *Xist* could not activate XCI. Inhibition of transcription with actinomycin D resulted in no change in splicing efficiency in WT ES cells. In the same ES cells unspliced transcripts were significantly decreased compared to DMSO treated controls. Our data suggests that splicing occurs post-transcriptionally when most *Xist* has completed transcription. This is further supported by our data that found spliced polyadenylated (poly(A)+) transcripts were enhanced in differentiated conditions.

Although *Xist* splicing has not been studied extensively, the role of splicing has been suggested in other publications. Studies using a doxycycline inducible transgene expressing exclusively spliced *Xist* were sufficient to induce XCI in undifferentiated ES cells as long as induction was done in the first 48 hours (Wutz and Jaenisch, 2000). Other studies done by Royce-Tolland *et al*. have shown that loss of the A-repeat prevents activation of XCI in the mutated X-

chromosome The A-repeat in exon 1 of *Xist* binds to alternative splicing factor 2 / splicing factor 2 (ASF2/SF2) (Royce-Tolland et al., 2010). Loss of the A-repeat results in a significant decrease in spliced transcripts in the mutated X-chromosome, however levels of unspliced *Xist* are comparable to WT. They also showed that RNA interference inhibition of ASF2/SF2 was sufficient to inhibit splicing of *Xist.* In both of these studies only changes in the splicing efficiency of *Xist* resulted in inhibition of XCI, further supporting *Xist* splicing as a regulatory step in XCI during the earliest steps of differentiation.

Taken together our results in WT cells suggest that in undifferentiated cells, *Xist* is expressed at low levels and remains unspliced. Once differentiation is induced, *Xist* expression ramps up initially however splicing at this point is still low. As *Xist* expression increases it begins to recruit splicing factors to the transcript. Once *Xist* is at a critical concentration at around 7-8h post induction (Figure 3A) and enough splicing factors are recruited to the transcript, splicing ramps up. As *Xist* is spliced it begins to recruit various binding factors to initiate silencing. The absence of splicing changes in undifferentiated cells suggests both differentiation and splicing of *Xist* are some of the earliest steps in XCI induction.

***Table* 1.1 *Xist* Splicing Primer Sequences**

Region	species	Forward Primer	Reverse Primer	Size (unspliced)	Size (Spliced)
Xist E2_E3	mouse	TGCTGGAGAGA GCCCAAA	TTCTGCTGAG ATGTAATGTA GCTGT	239	90
Xist E1_E2	mouse	CTACTGCTCCT CCGTTACATCA	AGGAGCACA AAACAGACTC CA	2868	74
Xist I1-E2	mouse	ACCAATTGAAA ACGCTGACA	AGGAGCACA AAACAGACTC CA	127	
Xist E2_E6	mouse	TGCTGGAGAGA GCCCAAA	AAGACCCAG TTTTCTGTGC TG	1930	562

Materials and Methods

Cell culture

Feeder independent WT female F1 2-1 mouse ES cells were expanded on 0.1% gelatinized tissue culture plates in 2i culture media containing 50% Neurobasal (Gibco cat. no. 21103049) and 50% DMEM/F12 (Gibco cat. no. 11320082) supplemented with 1% B27+RA (Gibco cat. no. 17504044), 1% N2 Supplement (R&D System cat.no. AR009), 1% Glutamax, 7.4 mM B27 Fraction V (Gibco cat. no. 15260037), 1% Penicillin-Streptomycin (GE Healthcare Life Sciences cat. no. SV30010), 3 μM CHIR99021 (Sigma cat.no. SML1046), 1 μM PD0325901 (Selleckchem cat. no. S1036), 150 μM 1-thioglycerol (Sigma cat. no. M6145), and 1000 U/ml mLIF (Gemini Bio-Products cat. no. 400-495). To assist in initial attachment, 2i medium was supplemented with 2% ESC grade FBS for the first 24 hours then switched to serum free. To prevent emerging of colonies and spontaneous differentiation, cells were passaged every two days.

Differentiation of embryonic stem cells

For monolayer differentiation, cells were plated at 200-300 x 10^3 per 6-well and were grown in LIF conditions for 24hrs and then LIF was removed and cells were grown in differentiation media for 24 hours (DMEM supplemented with 15% ESC FBS, 1% Nucleosides, 1% L-glutamine, 0.1mM beta mercaptoethanol, and 1 μM retinoic acid (Sigma cat. no. R2625)). RNA was collected at 24 hours post retinoic acid addition. To minimize cell death, media was replenished every 2 days.

Actinomycin D Treatments

Cells were plated at 100 x 10^3 per 6-well for 24 hours (h) prior to actinomycin D treatment.. For transcription inhibition prior to the inflection point of enhanced splicing, cells were differentiated for 6 h in media lacking leukemia inhibitory factor (LIF) and supplemented with 1μM retinoic acid. After 6 h of differentiation, cells were treated with 5 μg/ml actinomycin D for 4 h added directly to growth media. Total RNA was collected from undifferentiated cells (0h), after 6 h of differentiation, and after 2 h or 4 h of actinomycin D post treatment (8h and 10h).

RNA Isolation, cDNA synthesis, and RT-qPCR

Trizol isolation of RNA was performed following manufacturer's protocol (Life Technologies). Isolated RNA was treated with 4 units of Turbo DNase (Ambion) at 37 °C for 35 minutes to degrade any remaining DNA. RNA was purified using Phenol-Chloroform (pH 4.5, VWR cat. no. 97064-744). RNA concentrations were measured using Nanodrop 2000c (Thermo Fisher). First-strand cDNA was generated using 1 μg RNA and 200 units of M-MLV reverse transcriptase (Promega cat. no. M1705) with a RT primer in a 20 μl reaction following the manufacturer's protocol. For *Xist* specific analysis, the RT primer was a *Xist* gene specific reverse primer (E6-E7 reverse) targeting exon 7 (50μM). For X-linked and differentiation genes random hexamers (30μM) were used. For polyadenylated (poly(A)+) RNA, oligodT primers (50μM) were used for cDNA generation. The RT-qPCRs were conducted using a Quantstudio 6 Flex Real-Time PCR instrument, 2x Syber Master Mix (Life Technologies) following

manufacturer's protocol. Three technical replicates of all samples were performed for each biological replicate. An additional no-template control was included in every experiment. Outliers were defined by a Ct variance higher than 0.5. Analysis was done with Quanstudio 6 software and Microsoft Excel. Statistics test were based on $\Delta\Delta$Ct values.

Splicing assay

RT-PCRs were performed under standard conditions for Taq polymerase from New England Biolabs (cat. no. M0267E). For amplification of spliced regions of *Xist* spanning exons 1-2, exons 2-3, and exons 2-6, exon specific primers for corresponding exons were used Table 1. The PCRs were performed at a Tm of 60C for 29-35 cycles. PCR products greater than 1kb were resolved in 1.5% TAE Agarose gel. Smaller products were resolved with capillary electrophoresis (QIAxcel Advanced System) or on a 2% TBE Agarose gel. Relative quantities of spliced and unspliced products were measured on the QIAxcel software (Qiagen) or ImageLab (Bio-Rad) and the splicing ratio was calculated as the molar ratio of spliced to unspliced PCR products. The electropherograms of capillary electrophoresis were converted to virtual gel images by QIAxcel for presentation.

References

Avner, P. & Heard, E. X-chromosome inactivation: Counting, choice and initiation. Nature Reviews Genetics (2001). doi:10.1038/35047580

Borsani, G. et al. Characterization of a murine gene expressed from the inactive X chromosome. Nature (1991). doi:10.1038/351325a0

Brockdorff, N. et al. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature (1991). doi:10.1038/351329a0

Brown, C. J. et al. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature (1991). doi:10.1038/349038a0

Chaumeil, J., Okamoto, I., Guggiari, M. & Heard, E. Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. Cytogenet. Genome Res. (2002). doi:10.1159/000071577

Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. (2006). doi:10.1101/gad.380906

Ciaudo, C. et al. Nuclear mRNA degradation pathway(s) are implicated in Xist regulation and X chromosome inactivation. PLoS Genet. (2006). doi:10.1371/journal.pgen.0020094

Clemson, C. M., Hall, L. L., Byron, M., McNeil, J. & Lawrence, J. B. The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. Proc. Natl. Acad. Sci. (2006). doi:10.1073/pnas.0601069103

de Napoles, M. et al. Polycomb group proteins ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev. Cell 7, 663–676 (2004).

Fang, J., Chen, T., Chadwick, B., Li, E. & Zhang, Y. Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. J. Biol. Chem. (2004). doi:10.1074/jbc.C400493200

Georgomanolis, T., Sofiadis, K. & Papantonis, A. Cutting a long intron short: Recursive splicing and its implications. Front. Physiol. (2016). doi:10.3389/fphys.2016.00598

Hendrich, B. D., Brown, C. J. & Willard, H. F. Evolutionary conservation of possible functional domains of the human and murine Xist genes. Hum. Mol. Genet. (1993). doi:10.1093/hmg/2.6.663

Huynh, K. O. & Lee, J. T. Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature (2003). doi:10.1038/nature02222
Jonkers, I. et al. Xist RNA Is Confined to the Nuclear Territory of the Silenced X Chromosome throughout the Cell Cycle. Mol. Cell. Biol. (2008). doi:10.1128/MCB.02269-07

Kay, G. F., Barton, S. C., Surani, M. A. & Rastan, S. Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development. Cell (1994). doi:10.1016/0092-8674(94)90049-3

Leaky, A., Weixiong, J., Kuhnert, F. & Stuhlmann, H. Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. J. Exp. Zool. (1999). doi:10.1002/(SICI)1097-010X(19990615)284:1<67::AID-JEZ10>3.0.CO;2-O

Lyon, M. F. Gene action in the X-chromosome of the mouse (mus musculus L.). Nature (1961). doi:10.1038/190372a0
Ma, M. & Strauss, W. M. Analysis of the Xist RNA isoforms suggests two distinctly different forms of regulation. Mamm. Genome (2005). doi:10.1007/s00335-004-2464-3

Mager, J., Montgomery, N. D., De Villena, F. P. M. & Magnuson, T. Genome imprinting regulated by the mouse Polycomb group protein Eed. Nat. Genet. (2003). doi:10.1038/ng1125

Mak, W. et al. Reactivation of the Paternal X Chromosome in Early Mouse Embryos. Science (80-.). (2004). doi:10.1126/science.1092674

Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. (1997). doi:10.1101/gad.11.2.156

Minkovsky, A. mammalian X chromosome. 30, 48–54 (2012).
MONK, M. A Stem-line Model for Cellular and Chromosomal Differentiation in Early Mouse Development. Differentiation 19, 71–76 (1981).

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X Inactivation Counting and Choice Is a Stochastic Process: Evidence for Involvement of an X-Linked Activator. Cell (2008). doi:10.1016/j.cell.2007.12.036

Navarro, P. et al. Molecular coupling of Xist regulation and pluripotency. Science (80-.). (2008). doi:10.1126/science.1160952

Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic Dynamics of Imprinted X Inactivation during Early Mouse Development. Science (80-.). (2004). doi:10.1126/science.1092727

Panning, B., Dausman, J. & Jaenisch, R. X chromosome inactivation is mediated by Xist RNA stabilization. Cell (1997). doi:10.1016/S0092-8674(00)80355-4
Panning, B. & Jaenisch, R. DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev. (1996). doi:10.1101/gad.10.16.1991

Panning, B. & Jaenisch, R. RNA and the epigenetic regulation of X chromosome inactivation. Cell (1998). doi:10.1016/S0092-8674(00)81155-1
Plath, K. et al. Developmentally regulated alterations in polycomb repressive complex 1 proteins on the inactive X chromosome. J. Cell Biol. (2004). doi:10.1083/jcb.200409026

Royce-Tolland, M. E. et al. The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. Nat. Struct. Mol. Biol. (2010). doi:10.1038/nsmb.1877

Sousa, E. J. et al. Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. Cell Stem Cell (2018). doi:10.1016/j.stem.2018.05.001

Sun, B. K., Deaton, A. M. & Lee, J. T. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol. Cell 21, 617–628 (2006).

Wang, J. et al. Imprinted X inactivation maintained by a mouse Polycomb group gene. Nat. Genet. (2001). doi:10.1038/ng574
Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell (2000). doi:10.1016/S1097-2765(00)80248-8

Yue, M. & Ogawa, Y. CRISPR/Cas9-mediated modulation of splicing efficiency reveals short splicing isoform of Xist RNA is sufficient to induce X-chromosome inactivation. Nucleic Acids Res. (2017). doi:10.1093/nar/gkx1227

Chapter 3: Regulation of *Xist* splicing by RNA binding protein Polypyrimidine Tract Binding Protein 1

Abstract

LncRNAs are being recognized as important in the regulation of developmental processes. Many lncRNAs are processed similar to protein coding genes in order to perform their function, however the precise mechanism of splicing of lncRNAs is still unknown. Recent studies have identified many potential RNA binding partners of *Xist*, however the nature and importance of these interactions are currently being studied. Our group and others found that polypyrimidine tract binding protein 1 (PTBP1) binds to exon 7 of *Xist*.RNA. PTBP1 is a known splicing regulator of protein coding genes and is expressed early in development. We found that in female ES cells that lacked PTBP1 fail to significantly upregulated mature *Xist* because a defect in *Xist* splicing. Our data suggests that PTBP1 is important in regulating splicing of *Xist* at the onset of *Xist* dependent XCI and that this interaction is indirect and may involve other transacting factors.

.

Introduction

New technologies have allowed for unprecedented identification of new genes. Full genome sequencing surprisingly has led to the discovery of many long noncoding RNAs but not many new protein genes. Although their functions are still being discovered, studies in cell lines and mouse models demonstrated lncRNAs as regulators of gene expression and silencing (reviewed by Rinn and Chang, 2012). LncRNAs are shown to regulate transcription of related genes by i) acting as decoys to prevent proteins from access to the DNA ii) as a scaffold to bring multiple proteins together to form complexes or iii) guides that target gene regulation in allele specific fashion (Spitale, Tsai, and Chang, 2011; Kino et al.,2010; Rinn et al, 2007).

Xist is an example of a scaffolding lncRNA. After differentiation, *Xist* expression is upregulated and then is spliced. Spliced *Xist* then acts as a molecular scaffold for multiple protein factors and protein complexes (Smola et al., 2016; Chu et al., 2015; McHugh et al., 2015). RNA binding proteins identified have been associated with global structure of *Xist* and temporally regulated by cell-induced changes. However, the nature of these protein-RNA interactions are debated and not fully understood in the field. The accepted belief is that *Xist* function primarily depends on the primary *Xist* transcript (Borsani et al., 1991; Brockdorff et al., 1991; Brown et al., 1991). *Xist* is processed similarly to protein coding genes, however the steps and factors involved haven't been studied in depth. However studies into the structure and the nature of these protein RNA-

interactions with *Xist* suggest both the primary sequence and the structure are important for silencing the X-chromosome (Smola et al., 2016; Lu et al., 2016; Chen et al., 2016; Fang et al., 2015).

Xist undergoes splicing and polyadenylation, however the general role of intron splicing of *Xist* is unknown. In protein coding genes, processing of mRNA is required for nuclear export to the cytoplasm while alternative splicing allows for multiple functional variants from an individual primary transcript. Neither applied to *Xist, which* is retained in the nucleus and only has two known isoforms that differ on the 3' end. *Xist* has 8 exons and retention or splicing of intron 7 determines the expression of the long or short isoform of *Xist* (Ma and Strauss, 2005). The long isoform is the primary isoform expressed in ES cells during differentiation and involved in XCI, however the short isoform is expressed in other tissues. Work by Yue *et al*. has shown that the short or long isoform of *Xist* are indistinguishable in their activation of *Xist* (Yue and Ogawa, 2017). Therefore the question remains of what is the role of intron splicing is in *Xist*.

Interestingly, human and mouse *Xist* have 47% similarity, however their exon-intron structure and key repeat regions are conserved suggesting functional importance (Nesterova et al., 2001). In recent years many RNA binding proteins have been identified as binding to key repeat regions of *Xist* (Smola et al., 2016; Chu et al., 2015; Minajigi et al., 2015; Mchugh et al., 2015). Identified RBPs that interact with *Xist* are temporally regulated, involved in the 3-D chromatin structure, gene silencing on the X-chromosome, and in the attachment of *Xist*

RNA to the X-chromosome during XCI, however the function of many identified RBPs in XCI regulation is still unknown (reviewed by da Rocha and Heard, 2017). Different subsets of RBPs have been shown to associate with *Xist* in undifferentiated and differentiated conditions, suggesting that RBPs may functionally regulate *Xist* during induction of XCI.

Xist is involved in multiple steps of X-chromosome inactivation mediated by protein interactions with key repeat regions in the Xist primary sequence. There are 7 identified repeat regions A thru F in *Xist* . The A-repeat is the most conserved and is essential for XCI mediated transcriptional repression. XCI fails to be induced by *Xist* lacking the A-repeat. Repeats BCD are associated with PRC1/2 in depositing histone methylation marks during XCI. Multiple groups have identified PTBP1 as binding to the E-repeat of *Xist* (Vuong et al, 2016; Chu et al., 2015; McHugh et al., 2015)

The E-repeat is in exon 7 of *Xist,* is highly conserved in eutherian mammals, and is required for stable localization of *Xist* on the XCI . The E-repeat is associated with recruitment of AS2HL chromatin modifying enzyme in regulating histone modifiers and escape gene expression, however it's importance in *Xist* regulation is currently unknown (Yue et al., 2017). PTBP1 is a known regulator of the export, splicing, stability, and export of mRNAs of protein coding genes. PTBP1 is temporally regulated and is involved in regulation of genes during differentiation. PTBP1 has also been identified in a forward genetic screen as impairing XCI. Considering PTBP1 is known to regulate

splicing and stability of precursor transcripts, in this chapter I tested whether PTBP1 is a regulator of differentiation induced splicing of *Xist.*

Results

Our lab and other labs have discovered that PTBP1 binds to *Xist* in the E-repeat of exon 7 (Vuong et al., 2016; Chu et al., 2015; McHugh et al., 2015). To test whether PTBP1 influences exon splicing, our lab generated female *Ptbp1-/-* ES cells from a $Ptbp1^{loxp}$ allele via homologous recombination. The $Ptbp1^{loxp}$ allele contains two loxp sites flanking exon 2 of *Ptbp1* (Figure 5A). Deletion of exon 2 in *Ptbp1* eliminates functional PTBP1 proteins (Boutz et al., 2007; Makeyev et al., 2007; Spellman et al. 2007). The female *Ptbp1-/-* ES cell line was validated by genotyping PCR (Figure 5B) and confirmed with western blot using an antibody that recognizes the N terminus of PTBP1 (Figure 5C). In another PTB, PTBP3, the deletion of the equivalent exon to exon 2 results in a N-terminally truncated protein isoforms. To test whether deletion of exon 2 resulted in N-terminally truncated PTBP1, we used an antibody reactive to the C-terminus of PTBP1. We observed a smaller and less abundant protein (Figure. 5D), which might correspond to a N-terminally truncated PTBP1. To test loss of PTBP1 function we looked at splicing of *Ptbp2*. PTBP1 represses exon 10 inclusion in *Ptbp2* thus targeting *Ptbp2* for nonsense mediated decay. In *Ptbp1-/-* ES cells we observed only *Ptbp2* transcripts with exon 10 retained (Figure 5E-F).

Xist RNA splicing is impaired in Ptbp1 mutant cells

PTB proteins house both a nuclear localization signal (NLS) and nuclear export signal (NES) and shuffle between the cytoplasm and the nucleus (Garcia-Blanco, Jamison, and Sharp,1989 ; Ramanelli et al 1997; Li and Yen, 2002).

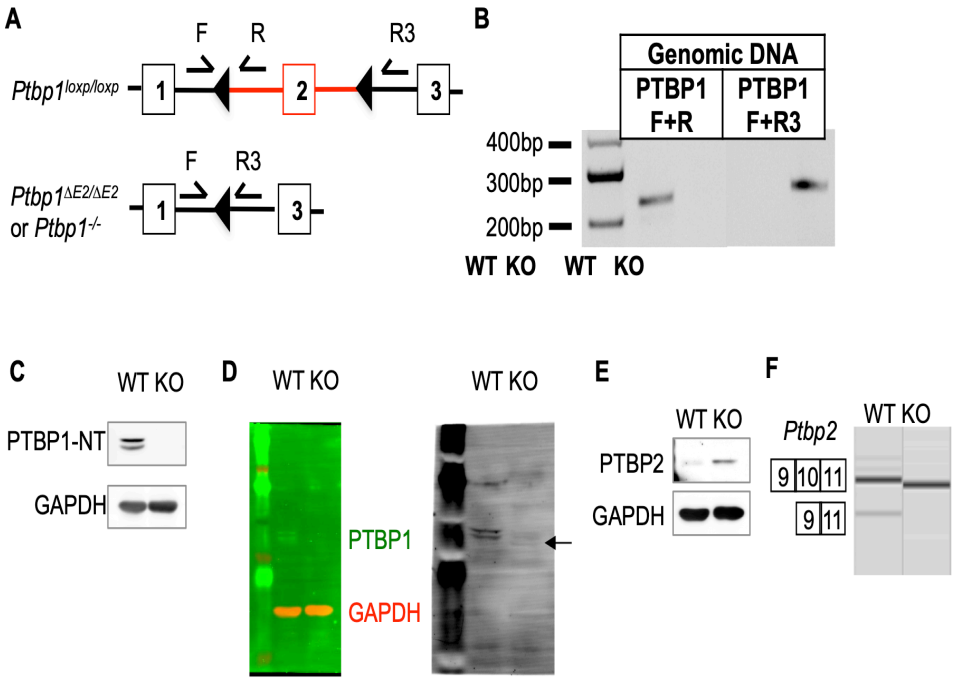

Figure 5. Characterization of *Ptbp1* mutant ES cells *A*) Schematic of *Ptbp1* gene structure in *Ptbp1*$^{loxp/loxp}$ and *Ptbp1*-/- (or *Ptbp*$^{\Delta E2/\Delta E2}$)cells, and the genotyping primers used. Close triangles represent loxp sites. *Ptbp1*$^{loxp/loxp}$ and *Ptbp1*-/- are herein referred to as WT and KO, respectively. (B) Representative gel shows genotyping of WT and *Ptbp1-/-* ES cells using different genotyping primer pairs. (C) Western blot confirmation of PTBP1 knockout using an anti-PTBP1-NT antibody. (D) Western blot of *Ptbp1-/-* ES cells versus WT control using PTBP1 C-terminus antibody. Left: overlapping blots of PTBP1 (in green) and GAPDH (in red). Right: over-exposed PTBP1 Western (high PMT) shows a smaller and substantially less abundant protein (arrow) in the *Ptbp1-/-* ES cell sample. (E) Western blot showing increase in PTBP2 expression in *Ptbp1-/-* ES cells. (F) Representative virtual gel image showing increased *Ptbp2* exon 10 splicing in *Ptbp1-/-* ES cells versus WT control.

Studies from other groups have show that PTB proteins are involved in splicing, polyadenylation, mRNA stability, and translation initiation.

How PTB proteins function depends on their localization and interactions with other binding factors. (Perez et al., 1997; Kamath et al., 2001; Li and Yen, 2002). I wanted to determine if PTBP1 had any functional roles in splicing of *Xist*. Cells become insensitive to *Xist* inactivation as early as 48 hrs post differentiation (Wutz and Jaenisch, 2000). WT BL6 showed a significant increase in spliced exon 1 - exon 2 and exon 2-exon 3 when differentiated in monolayer growth conditions for 24 hours (Figure 6A-B). *Ptbp1-/-* cells displayed lower splicing when compared to WT cells (Figure 6A-B). *Ptbp1* regulation can be exon specific so next I wanted to test if the splicing defect was only present in individual exon splicing events or across all exons. To distinguish these possibilities, I used primers flanking multiple exons (2-6 and 3-7). Our results show that loss of *Ptbp1* resulted in less splicing across multiple exons when compared to WT BL6 ES cells (Figure 6C-D).

PTBP1 has cellular functions in stability of RNAs as well as splicing. Studies show PTBP1 expressed at the earliest in the epiblast cells (Suckale et al., 2011). An alternative explanation for the decrease of spliced *Xist* transcripts in *Ptbp1-/-* cells is a decrease of the spliced transcript in female *Ptbp1-/-* ES cells. In this scenario, even if splicing can proceed, loss of spliced transcripts would attenuate and differentiation enhanced splicing ratio in *Ptbp1-/-* ES Cells.

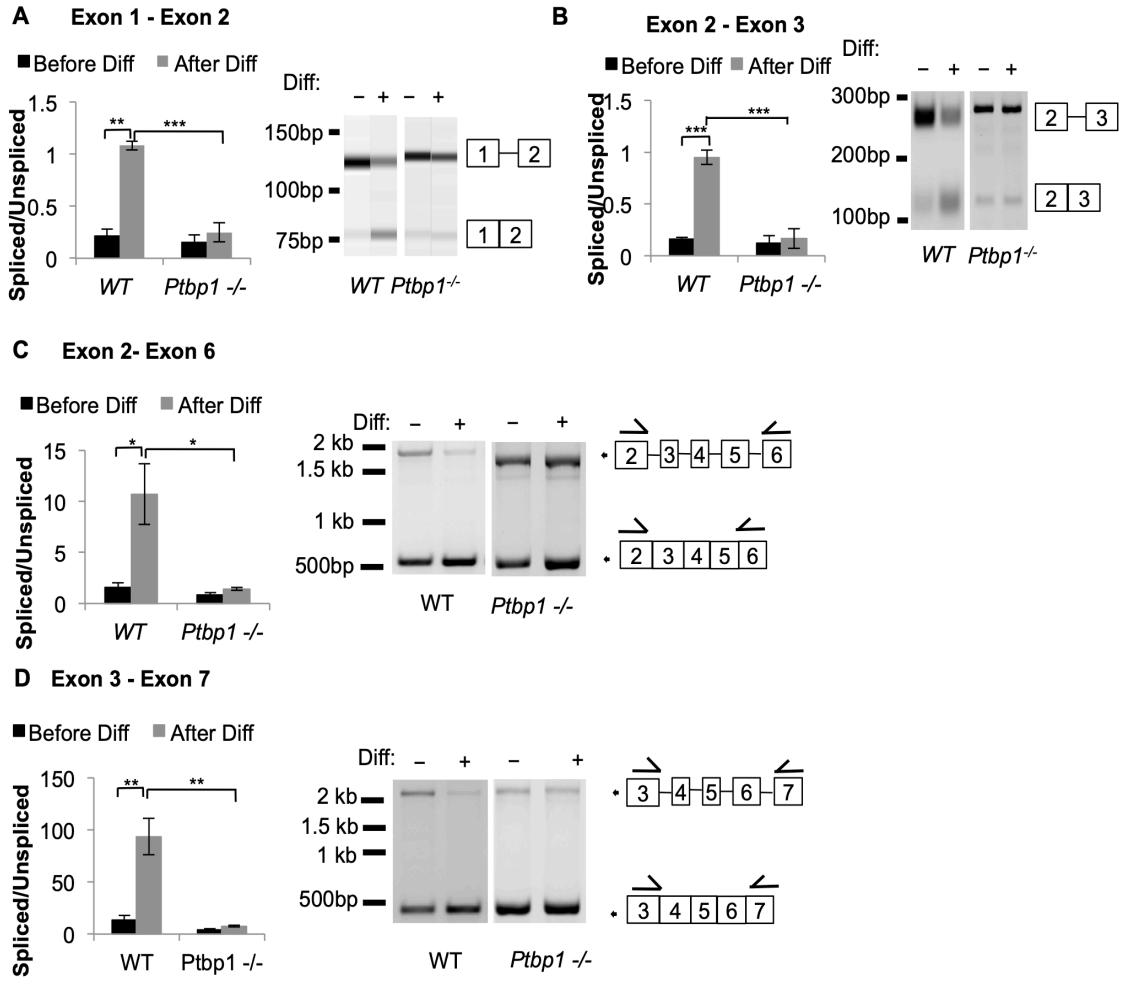

Figure 6. *Xist* RNA splicing is impaired in monolayer *Ptbp1* mutant ES cells
Representative virtual gel image (*left* panel) and quantified splicing ratio (*right* panel) for (A) intron 1, (B) intron 2 , (C) intron 2 thru intron 5, and (D) inron 3 to intron 6 in WT and *Ptbp1-/-* cells upon differentiation using the 24 hour monolayer method. Error bars indicate S.E.M of three biological replicates. #, p-value=0.06; *, p-value<0.05 ; ***, p-value<0.001 Students t-test (1-tailed, unpaired).

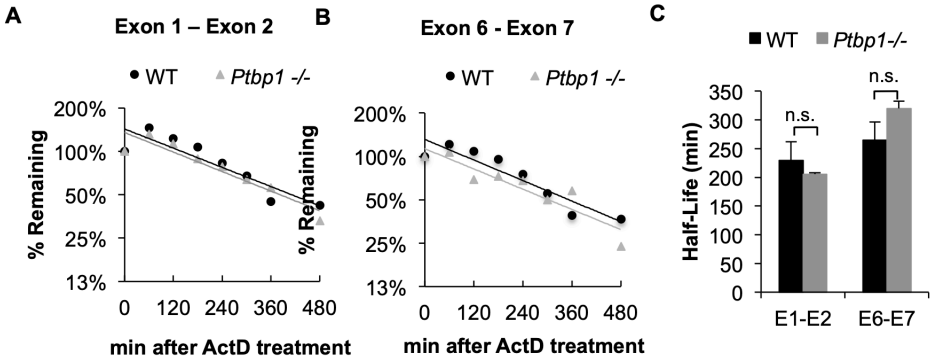

Figure 7. *Xist* RNA stability is not impaired in monolayer *Ptbp1* mutant ES cells (A-C) The half-lives of spliced *Xist* RNA do not change before or 24 hours after differentiation. Percent remaining spliced *Xist* transcripts after actinomycin-D treatment are measured using primer pairs targeting A) Exon 1 to Exon 2 and (B) Exon 6 to Exon 7. Error bars indicate S.E.M of three biological replicates. ns; Students t-test (1-tailed, unpaired).

However I observed no change in stability of WT Bl6 and *Ptbp1 -/-* ES cells treated with actinomycin D before or after differentiation (Figure 7A-C).

PTBP1 and PTBP2 have similar functions with PTBP2 expression upregulated later during differentiation when PTBP1 levels go down. *Xist* mediated X-inactivation is irreversible as early as 48 hours. Embryoid body (EB) differentiation of ES cells in suspension culture is the favored model to mimic later phases of differentiation. I next asked if *Ptbp1-/-* defects in splicing were maintained at later stages of differentiation. EB differentiation for 2 to 4 days in WT Bl6 ES cells dramatically enhanced *Xist* splicing, however *Ptbp1-/-* cells displayed lower splicing ratios than WT cells(Figure 8A-B).

Since loss of PTBP1 showed a splicing defect in *Ptbp1-/-* ES cells and enhanced splicing in WT ES cells, I wanted to determine if *Xist* splicing levels correlated with PTBP1 protein levels during EB differentiation. I observed an increase in PTBP1 in undifferentiated and 24 hours after differentiated, however PTBP1 levels steadily decreased from day 2 through day 8 (Figure 9A-B). *Xist* splicing efficiency is therefore not correlative with PTBP1 protein expression. PTBP1 protein is high in undifferentiated cells, which is supported by work in other groups that showed PTBP1 is expressed in the epiblast.

One of the early activators of *Xist* induction is induced by downregulation of pluripotency factors during differentiation of ES cells. One possibility for this observed reduction of *Xist* transcripts, was that *Ptbp1-/-* ES cells differentiated more slowly than WT cells.

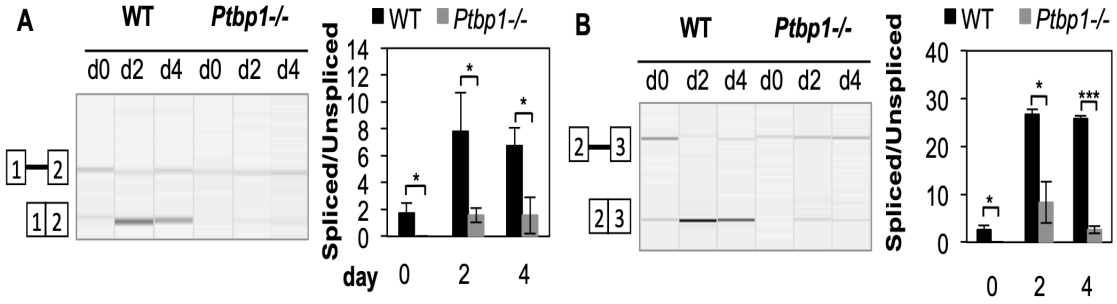

Figure 8. *Xist* RNA splicing is impaired in *Ptbp1* mutant embryoid bodies
Representative virtual gel image (*left* panel) and quantified splicing ratio (*right* panel) for (A) intron 1, (B) intron 2 in WT and *Ptbp1-/-* embryoid bodies, Error bars indicate S.E.M of three biological replicates. #, p-value=0.06; *, p-value<0.05 ; ***, p-value<0.001 Students t-test (1-tailed, unpaired).

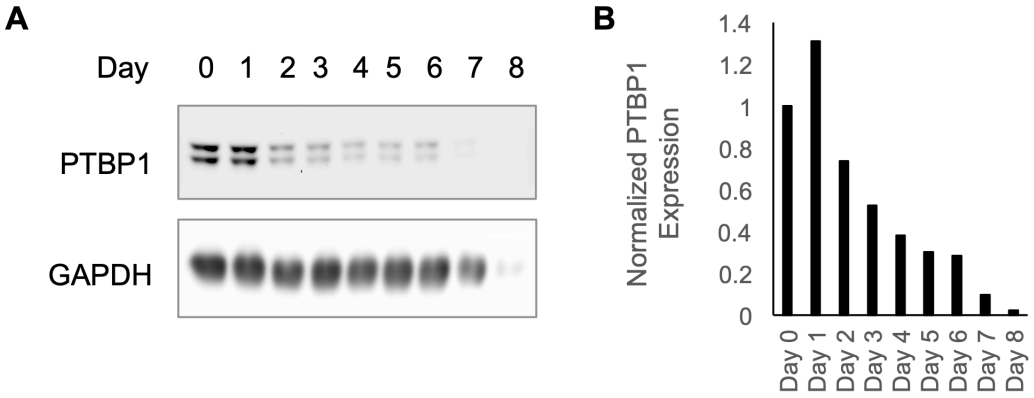

Figure 9. Expression of PTBP1 in wildtype embryoid bodies (A)Western blot analysis of PTBP1 levels during embryoid body differentiation in WT cells. (B) Quantification of PTBP1 normalized to GAPDH.

To test this I measured the expression of pluoripotency marker *Nanog*, epiblast marker *Fgf*, and neuroectoderm marker *Zfp521* (Figure 10A). In WT ES cells our data showed that *Nanog* expression decreased at day 2 and then slightly reactivated at day 4 as previously noted by *Trott et al.* (Trott et al., 2013). Increases in *Fgf5* are seen as early as day 2 and *Zfp521* at day 4, consistent with sequential steps of cell lineage specification. *Ptbp1-/-* ES cells showed similar levels of *Nanog , Fgf5,* and *Zfp521* in undifferentiated cells (day 0). Differentiated *Ptbp1-/-* EBs displayed decreased levels of *Nanog*, and increased levels of *Fgf5* and *Zfp521* compared to WT suggesting earlier exit from differentiation (Figure 10A-C). Differentiation enhances XCI induction, if PTBP1 regulated *Xist* splicing through its effect on differentiation, we should of observed higher levels of splicing in *Ptbp1-/-* cells. Our data suggests that PTBP1 regulation of *Xist* splicing is permissive and not instructive. PTBP1 levels do not appear to instruct *Xist* splicing levels, however PTPB1 needs to be present at an initial certain level to enhance *Xist* expression and for splicing to ramp up. Additionally, I observed no morphological defects in *Ptbp1-/-* ES cells during differentiation (Figure 10D). PTPB1 effects on *Xist* splicing appear to be independent of its effect on cellular differentiation (Figure10E)

Proper induction and maintenance of random XCI is responsible for silencing of X-linked genes. *Ptbp1-/-* ES cells display significantly impairs splicing of *Xist* as early as 24-48 hours post differentiation prior to when XCI becomes irreversible. Therefore I asked whether *Ptbp1-/-* resulted in de-

repression of X-linked gene targets of XCI. I observed that *Ptbp1-/-* cells showed modest increases in X-linked genes Glucose-6-phosphate 1-dehydrogenase (*G6pdx)* and Lysosome-associated membrane protein 2 (*Lamp2)*, however showed a modest decrease in Phosphoglycerate Kinase 1 (*Pgk1)* (Figure 10F-H). One possibility is that low levels of spliced *Xist* present in absence of *Ptbp1* could accumulate over time and can induce some levels of XCI. However the assays used are probably not sensitive enough to detect XCI. Another possibility is that there are other factors independent of *Xist* involved in X-linked gene re-activation as shown by Yang *et al* (Yang et al., 2016).

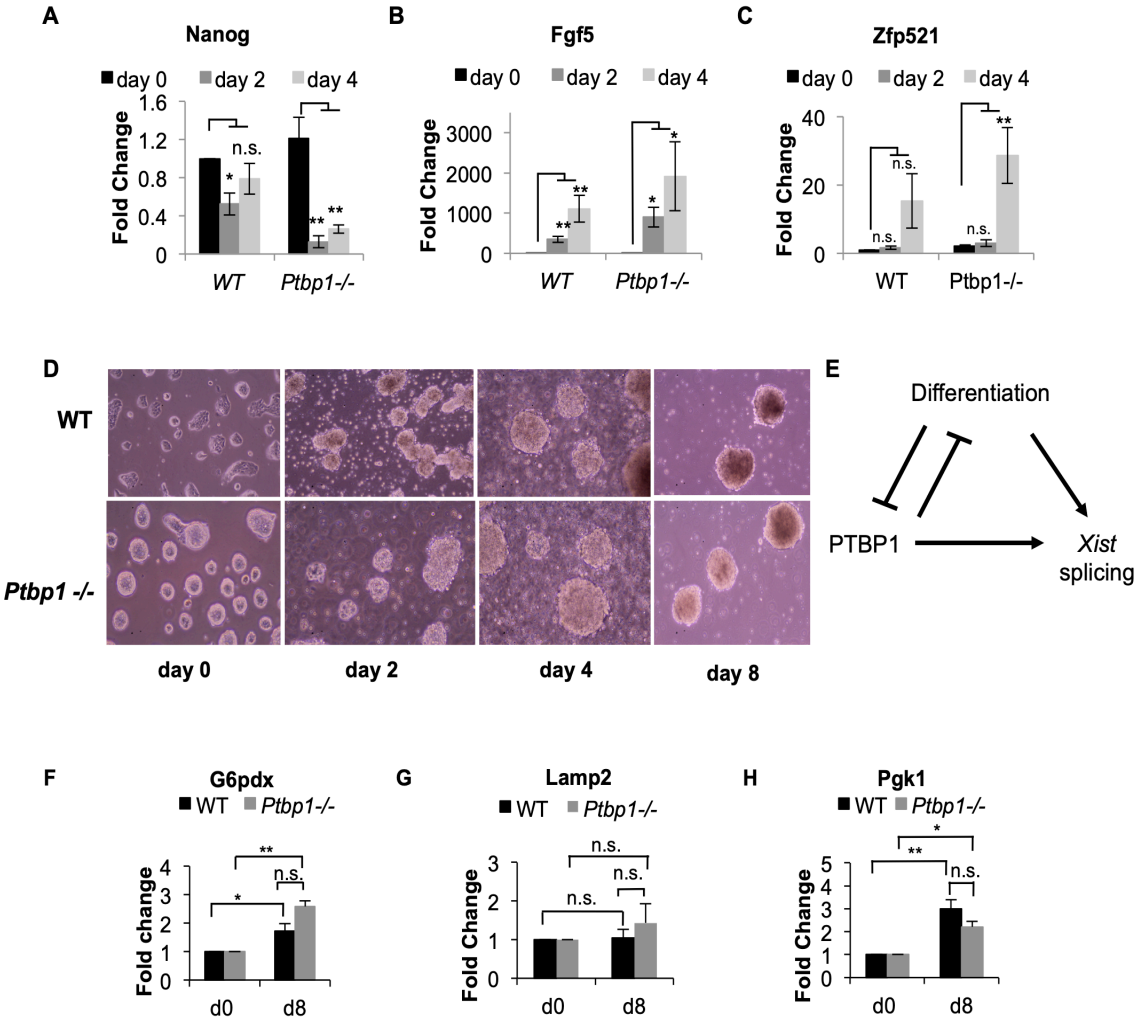

Figure 10. **The effects of mutant *Ptbp1* on differentiation and XCI target X-linked genes** (A-C) RT-qPCR analyzes expression levels of (A) pluripotency gene *Nanog*, (B) epiblast marker *Fgf5*, and (C) neuroectoderm marker *Zfp521 in* WT and *Ptbp1-/-* embryoid bodies normalized to respective undifferentiated controls (day 0). (D) Embryoid body formation in undifferentiated WT and *Ptbp1-/-* ES cells (day 0) and at day 2 and 4 during differentiation. (E) A schematic summarizing the relationship between differentiation (F-H) RT-qPCR analyzes expression levels of X-linked genes (F) *G6pdx,* (G) *Lamp2,* and (H) *Pgk1.* WT and *Ptbp1-/-* embryoid bodies normalized to respective undifferentiated controls (day 0). *, p-value<0.05 ; **, p-value<0.01; ***, p-value<0.001 Students t-test (1-tailed, unpaired).

Discussion

In this chapter I examined the role of polypyrimidine tract binding protein 1 (PTBP1) had in regulating the lncRNA *Xist* . For many years, groups have studied the factors involved in *Xist* regulation and transcriptional silencing of the X-chromosome. Studies so far focused the genes involved in regulating the activation of *Xist* upon initiation of differentiation and cell lineage specification. Recently, studies have focused on identifying what binding partners are involved in *Xist* mediated silencing of the X-chromosome during XCI induction. RNA binding proteins have long been shown to be involved in the fate of all transcripts including lncRNAs in the cell.

Multiple groups recently identified many RBPs that interact with *Xist* are involved in RNA modeling pathways, chromatin modifications, splicing, and the nuclear matrix. One RBP, PTBP1 was shown by multiple groups as binding to the E-repeat in exon 7 of *Xist* (Vuong et al., 2016; Chu et al., 2015; McHugh et al., 2015). PTBP1 is a known regulator of splicing, polyadenylation, and transport of mRNAs. By focusing on PTBP1, we found that loss of PTBP1 in *Ptbp1-/-* ES cells resulted in earlier exit from pluripotency and significantly impaired (but not completely inhibit) splicing of *Xist.* Our data suggests a model in which PTBP1 serves a permissive role in *Xist* splicing rather than instructive and that this role may be indirect in differentiated ES cells.

In the previous chapter we showed that differentiation induces splicing of *Xist*. ES cells lacking *Rent1 or Eif1* only reduced differentiation induced spliced

Xist, however unspliced transcripts levels remain the same as WT. The cells are also unable to upregulate *Xist* and fail to induce XCI suggesting a connection between splicing and XCI activation (Ciaudo et al.,2006). Therefore we set out to identify if PTBP1 is an early regulator of *Xist* splicing upon differentiation induction.

We observed significantly impaired splicing of *Xist* in differentiated *Ptbp1-/-* ES cells in both monolayer and embryoid body models compared to WT Bl6 cells. In both WT and *Ptbp1-/-* ES cells we see primarily unspliced transcripts in undifferentiated cells. In WT BL6 ES cells, *Xist* is spliced after 1 day of differentiation and maintained at similar levels of spliced transcripts during extended differentiation. However, *Ptpb1-/-* ES cells display significant impairment of splicing as early as 1 day remains impaired throughout differentiation. There were no observed differences in stability of WT Bl6 and *Ptpb1-/-* ES cells suggesting PTBP1's regulation of *Xist* is based on changes in splicing efficiency. This data was intriguing since it suggests that splicing is regulated very early in differentiation and then maintained, and this initial splicing regulation is regulated in part by PTBP1. Our data was consistent with studies by *Wutz et al.* that suggested that the initiation window of when *Xist* can initiate repression is limited to the 32 hours post differentiation before X-inactivation becomes irreversible at 72 hours post differentiation (Wutz and Jaenisch, 2000).

PTBP1 is expressed in undifferentiated ES cells at 24 hours of differentiation but starts to decrease at 48 hours. We observed splicing

impairment in *Ptbp1-/-* ES cells is beyond 48 hours. This data suggests that some level of PTBP1 is required at initiation of differentiation to induce splicing of *Xist,* but splicing isn't correlated to the amount of PTBP1 present. This suggests that PTBP1 is more permissive instead of instructive in *Xist* splicing at the initiation of differentiation. This is consistent with work done by Chu *et al.* that found different proteins interact with *Xist* in a developmentally related manner in undifferentiated and differentiated cells (Chu et al., 2015).

Xist induction is tightly correlated with pluripotency status. Exit from pluripotency is associated with higher levels of *Xist* and induction of random XCI. However in *Ptbp1-/-* embryoid bodies we observe enhanced differentiation. This data suggests that PTBP1 may have separate roles in differentiation and splicing and may involve interactions with other factors. *Xist* splicing isn't mediated by differentiation inhibition by PTBP1. Additionally differentiation induced *Xist* splicing isn't mediated by differentiation induced downregulation of PTBP1. This further suggests that PTBP1's regulation of *Xist* may be indirect and require binding to other factors that promote splicing of the transcript.

Loss of spliced *Xist* has been associated with inefficient induction XCI. Impaired splicing of *Xist* observed in *Ptbp1-/-* ES cells would suggest impairment of silencing of X-linked genes. However our data only saw modest increases in two of the XCI target genes studied (*G6px, Lamp2)*. One possibility is since *Ptbp1-/-* didn't completely inhibit splicing; that a small amount of spliced transcript is sufficient to silence XCI targeted genes. Another possibility is that expression

of these X-linked genes is still relatively low at this point in differentiation, thus masking any changes in silencing since we do not see much a change in expression in WT ES cells between day 0 and day 8. Another possibility is this assay may not be sensitive enough to accurately monitor silencing of X-linked genes.

How lncRNAs are transcriptionally being regulated is still a growing field. It is becoming increasingly apparent that RNA binding proteins are heavily involved in transcriptional regulation of lncRNAs and this is developmentally regulated however the mechanisms and interacting proteins are still being studied. *Xist* is one of the most studied lncRNAs and its function in XCI is well understood, however the mechanisms that regulate its expression and function are still a mystery.

LncRNAs have less efficient splicing due to lower amount of functional splicing enhancers, lower binding to splicing regulator U2AF, lacking of cooperative splice sites for binding to SR proteins, and weak polypyrimidine tracts (Krchnakova et al., 2018). PTBP1 binds to the E-repeat of *Xist* in exon 7, downstream of intron 6. PTBP1 is more commonly known for repressing splicing of its target exons by blocking binding U2AF binding resulting in alternative exon exclusion, our results suggest the opposite for *Xist* (Sauliere et al., 2006). One possibility is that PTBP1 binding to the E-repeat serves as a landing pad for SR proteins or other factors that promote splicing. Alternatively PTBP1 expression controls additional splicing factors that interact with *Xist*. Since PTBP1 regulation

appears to be indirect, future studies will need to be done to identify what binding partners bind to PTBP1 when it's bound to *Xist* in differentiated ES cells.

Materials and Methods

Cell culture

Feeder dependent female WT and *Ptbp1*$^{-/-}$ BL6 mouse ES cells were expanded on inactive male murine embryonic fibroblast (MEF) feeders on 0.1% gelatinized tissue culture plates in DMEM media (Gibco cat. no.10313039) supplemented with 15% ESC grade FBS (Gibco cat. no.10439024), 1% Nucleosides (EMD Millipore cat. no. ES008D), 1% Glutamax (Gibco cat. no.35050061), 0.1mM beta mercaptoethanol (Acros Organics cat. no. 125472500), and 1000 U/ml mLIF (EMD Millipore cat.no. ESG1106). To prevent emerging of colonies and spontaneous differentiation, media were changed every day and cells were passaged every two days on male feeders. Feeder independent female WT and *Ptbp1*$^{-/-}$ BL6 mouse ES cells, and WT female F1 2-1 mouse ES cells were expanded on 0.1% gelatinized tissue culture plates in 2i culture media containing 50% Neurobasal (Gibco cat. no. 21103049) and 50% DMEM/F12 (Gibco cat. no. 11320082) supplemented with 1% B27+RA (Gibco cat. no. 17504044), 1% N2 Supplement (R&D System cat.no. AR009), 1% Glutamax, 7.4 mM B27 Fraction V (Gibco cat. no. 15260037), 1% Penicillin-Streptomycin (GE Healthcare Life Sciences cat. no. SV30010), 3 μM CHIR99021 (Sigma cat.no. SML1046), 1 μM PD0325901 (Selleckchem cat. no. S1036), 150 μM 1-thioglycerol (Sigma cat. no. M6145), and 1000 U/ml mLIF (Gemini Bio-Products cat. no. 400-495). To assist in initial attachment, 2i medium was supplemented with 2% ESC grade FBS for

the first 24 hours then switched to serum free. To prevent emerging of colonies and spontaneous differentiation, cells were passaged every two days.

Differentiation of embryonic stem cells

For monolayer differentiation, cells were plated at 200-300 x 10^3 per 6-well and were grown in LIF conditions for 24hrs and then LIF was removed and cells were grown in differentiation media for 24 hours (DMEM supplemented with 15% ESC FBS, 1% Nucleosides, 1% L-glutamine, 0.1mM beta mercaptoethanol, and 1 μM retinoic acid (Sigma cat. no. R2625). RNA was collected at 24 hours post retinoic acid addition. To generate embryoid bodies, feeder independent WT ES F1 2-1, WT C57B6, and *Ptbp1-/-* ES cells were plated at a density of 2.5 x 10^6 cells on 0.5% Agarose coated polystyrene dishes (Grenier Bio-One cat. no. 633102) and maintained in DMEM supplemented with 10% FBS, 1% Glutamax, 1% Non-Essential Amino Acids (GE Healthcare, cat. no. SH3023801), 1% Sodium Pyruvate (Lonza cat. no. 13-115E), and 0.1mM betamercaptoethanol without mLIF for up to 8 days. For western blots, WT and *Ptbp1-/-* mESCs were plated at 300 x 10^3 on 0.5% agarose coated 6-well dishes (VWR, cat. no. 10062-892) to generate embryoid bodies. To minimize cell death, media was replenished every 2 days.

Actinomycin D Treatments

Cells were plated at 100 x 10^3 per 6-well for 24 hours (h) prior to actinomycin D treatment. For stability experiments, cells were treated with 5 μg/ml actinomycin-D (Sigma, cat. no. A1410) or an equal volume of DMSO vehicle (Sigma, cat. no.

D2438) in growth media for 8 h. RNA was collected at 0, 2, 4, 6, and 8 h post treatment. Expression of *Xist* was calculated using the ΔΔCT method before and after treatment. Percent *Xist* remaining was calculated as the ratio of *Xist* transcripts at the time of collection relative to time 0.

Generation of Ptbp1$^{-/-}$ ES cells

Mouse *Ptbp1$^{loxp/loxp}$* ES cells were generated following a previously published protocol (10). In brief, healthy E3.5 blastocysts were collected from *Ptbp1$^{loxp/loxp}$* intercrosses and individually cultured in KSOM medium (DMEM high glucose with 15% (vol/vol) KO serum, 2 mM GlutaMAX,1 mM sodium pyruvate, 0.1 mM MEM NEAA, 0.1 mM β-Mercaptoethanol, 103 U/ml of LIF, 1 μM PD0325901 and 3 μM CHIR99021. After overnight culture, blastocysts were transferred to defined serum-free medium (1:1 mixture of DMEM-F12/N2 and Neurobasal/B27, 1 mM GlutaMAX, 0.5 mM sodium pyruvate, 0.1 mM MEM NEAA, 0.1 mM β-Mercaptoethanol) plus 2i/LIF to allow attachment to the feeder layer. After outgrowth from blastocysts, cells were disaggregated and cultured in the serum-free medium plus 2i/LIF together with MEF. Cells were monitored daily for colony formation. Cells exhibiting differentiation or atypical ES cell morphology were discarded and the healthy ES cell lines were maintained in the serum-free medium plus 2i/LIF with MEF. *Ptbp1$^{-/-}$* mES cells were generated via expressing Cre recombinase in *Ptbp1$^{loxp/loxp}$* ES cells. Ptbp1 homozygous knockout ES cell clones were identified by PCR genotyping and confirmed by PTBP1 Western blot.

RNA Isolation, cDNA synthesis, and RT-qPCR

Trizol isolation of RNA was performed following manufacturer's protocol (Life Technologies). Isolated RNA was treated with 4 units of Turbo DNase (Ambion) at 37 °C for 35 minutes to degrade any remaining DNA. RNA was purified using Phenol-Chloroform (pH 4.5, VWR cat. no. 97064-744). RNA concentrations were measured using Nanodrop 2000c (Thermo Fisher). First-strand cDNA was generated using 1 µg RNA and 200 units of M-MLV reverse transcriptase (Promega cat. no. M1705) with a RT primer in a 20 µl reaction following the manufacturer's protocol. For *Xist* specific analysis, the RT primer was a *Xist* gene specific reverse primer (E6-E7 reverse) targeting exon 7 (50µM). For X-linked and differentiation genes random hexamers (30µM) were used. For polyadenylated (poly(A)+) RNA, oligodT primers (50µM) were used for cDNA generation. The RT-qPCRs were conducted using a Quantstudio 6 Flex Real-Time PCR instrument, 2x Syber Master Mix (Life Technologies) following manufacturer's protocol. Three technical replicates of all samples were performed for each biological replicate. An additional no-template control was included in every experiment. Outliers were defined by a Ct variance higher than 0.5. Analysis was done with Quanstudio 6 software and Microsoft Excel. Statistics test were based on $\Delta\Delta$Ct values.

Splicing assay

RT-PCRs were performed under standard conditions for Taq polymerase from New England Biolabs (cat. no. M0267E). For amplification of spliced regions of

Xist spanning exons 1-2, exons 2-3, and exons 2-6, exon specific primers for corresponding exons were used (Supplemental Table 1). The PCRs were performed at a Tm of 60C for 29-35 cycles. PCR products greater than 1kb were resolved in 1.5% TAE Agarose gel. Smaller products were resolved with capillary electrophoresis (QIAxcel Advanced System) or on a 2% TBE Agarose gel. Relative quantities of spliced and unspliced products were measured on the QIAxcel software (Qiagen) or ImageLab (Bio-Rad) and the splicing ratio was calculated as the molar ratio of spliced to unspliced PCR products. The electropherograms of capillary electrophoresis were converted to virtual gel images by QIAxcel for presentation.

Immunoblotting

For primary antibodies, we used antibodies to PTBP1-NT (1:1,000), PTBP1-CT (1:200), PTBP2 (1:2,000), GAPDH (1:1,000, Ambion, AM4300). For secondary antibodies we used Alexa Fluor 488 donkey anti rabbit (Invirogen, A21206) and AF 647 donkey anti mouse (Invirogen, A31571).

References

Borsani, G. *et al.* Characterization of a murine gene expressed from the inactive X chromosome. *Nature* (1991). doi:10.1038/351325a0

Boutz, P. L. *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* (2007). doi:10.1101/gad.1558107

Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* (1991). doi:10.1038/351329a0

Chen, C. K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science (80-.).* (2016). doi:10.1126/science.aae0047

Chen, C. K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science (80-.).* (2016). doi:10.1126/science.aae0047

Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* (2015). doi:10.1016/j.cell.2015.03.025

Da Rocha, S. T. & Heard, E. Novel players in X inactivation: Insights into Xist-mediated gene silencing and chromosome conformation. *Nature Structural and Molecular Biology* (2017). doi:10.1038/nsmb.3370

Fang, R., Moss, W. N., Rutenberg-Schoenberg, M. & Simon, M. D. Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS Genet.* (2015). doi:10.1371/journal.pgen.1005668

Garcia-Blanco, M. A., Jamison, S. F. & Sharp, P. A. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev.* (1989). doi:10.1101/gad.3.12a.1874

Kamath, R. V., Leary, D. J. & Huang, S. Nucleocytoplasmic Shuttling of Polypyrimidine Tract-binding Protein Is Uncoupled from RNA Export. *Mol. Biol. Cell* (2001). doi:10.1091/mbc.12.12.3808

Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* (2010). doi:10.1126/scisignal.2000568

Krchňáková ,Z. *et al.* Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5? splice-site sequences due to weak interactions with SR proteins. *Work. Aging Retire.* (2018). doi:10.11821/dlxb201802008

Li, B. & Benedict Yen, T. S. Characterization of the nuclear export signal of polypyrimidine tract-binding protein. *J. Biol. Chem.* (2002). doi:10.1074/jbc.M109686200

Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* (2016). doi:10.1016/j.cell.2016.04.028

Ma, M. & Strauss, W. M. Analysis of the Xist RNA isoforms suggests two distinctly different forms of regulation. *Mamm. Genome* (2005). doi:10.1007/s00335-004-2464-3

Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. *Mol. Cell* (2007). doi:10.1016/j.molcel.2007.07.015

McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* (2015). doi:10.1038/nature14443

Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science (80-.).* (2015). doi:10.1126/science.aab2276

Moindrot, B. *et al.* A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Rep.* (2015). doi:10.1016/j.celrep.2015.06.053

Nesterova, T. B. *et al.* Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Research* (2001). doi:10.1101/gr.174901

Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* (2012). doi:10.1016/j.scienta.2012.12.026

Sauliere, J., Sureau, A., Expert-Bezancon, A. & Marie, J. The Polypyrimidine Tract Binding Protein (PTB) Represses Splicing of Exon 6B from the -Tropomyosin Pre-mRNA by Directly Interfering with the Binding of the U2AF65 Subunit. *Mol. Cell. Biol.* (2006). doi:10.1128/MCB.00893-06

Smola, M. J. *et al.* SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the *Xist* lncRNA in living cells. *Proc. Natl. Acad. Sci.* (2016). doi:10.1073/pnas.1600008113

Smola, M. J. *et al.* SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the *Xist* lncRNA in living cells. *Proc. Natl. Acad. Sci.* (2016). doi:10.1073/pnas.1600008113

Spellman, R., Llorian, M. & Smith, C. W. J. Crossregulation and Functional Redundancy between the Splicing Regulator PTB and Its Paralogs nPTB and ROD1. *Mol. Cell* (2007). doi:10.1016/j.molcel.2007.06.016

Spitale, R. C., Tsai, M. C. & Chang, H. Y. RNA templating the epigenome: Long noncoding RNAs as molecular scaffolds. *Epigenetics* (2011). doi:10.4161/epi.6.5.15221

Suckale, J. *et al.* PTBP1 is required for embryonic development before gastrulation. *PLoS One* (2011). doi:10.1371/journal.pone.0016992

Trott, J. & Martinez Arias, A. Single cell lineage analysis of mouse embryonic stem cells at the exit from pluripotency. *Biol. Open* (2013). doi:10.1242/bio.20135934

Vuong, J. K. *et al.* PTBP1 and PTBP2 Serve Both Specific and Redundant Functions in Neuronal Pre-mRNA Splicing. *Cell Rep.* (2016). doi:10.1016/j.energy.2018.10.173

Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* (2000). doi:10.1016/S1097-2765(00)80248-8

Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* (2000). doi:10.1016/S1097-2765(00)80248-8

Yang, L., Kirby, J. E., Sunwoo, H. & Lee, J. T. Female mice lacking Xist RNA show partial dosage compensation and survive to term. *Genes Dev.* (2016). doi:10.1101/gad.281162.116

Yue, M. *et al.* Xist RNA repeat E is essential for ASH2L recruitment to the inactive X and regulates histone modifications and escape gene expression. *PLoS Genet.* (2017). doi:10.1371/journal.pgen.1006890

Yue, M. & Ogawa, Y. CRISPR/Cas9-mediated modulation of splicing efficiency reveals short splicing isoform of Xist RNA is sufficient to induce X-chromosome inactivation. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx1227

Chapter 4: Allele specific silencing of the X-chromosome and X-chromosome Inactivation.

Abstract

During early development in female placental animals, one of the X-chromosomes are silenced through the epigenetic process of X-chromosome inactivation (XCI). Random XCI occurs post-implantation in the epiblast and includes the counting and choice of which X-chromosome to inactivate prior to increasing expression of *Xist* from the future inactivate chromosome. The decision of which chromosome to inactivate is normally random. However, XCI is skewed in hybrid mice as determined by the alleles of the X-controlling element (*Xce*) locus. Cattanach over 40 years ago identified the Xce locus as the main determinant of XCI skewing however the mechanism is currently unknown (Cattanach and Isaacson, 1967). The Xce locus is unknown, however identified candidate regions overlap with X-inactivation center that includes the *Xist* gene (Thorvaldsen et al., 2012; Chadwick et al., 2006; Simmler et al., 1993; Cattanach and Papworth, 1981). The four known *Xce* alleles are defined based in order of which Xce will remain active during XCI: $Xce^a < Xce^b < Xce^c < Xce^d$ (Rastan, 1982). Here we use F1 2-1 ES cells that are a *Mus musculus 129SvJaeJ (129) /Mus custaneous (CAS)* heterozygote. The 129 allele has an Xce^a and the CAS allele has a Xce^c with XCI skewed to inactivate the 129 allele. Prior to differentiation both $Xist^{129}$ and $Xist^{CAS}$ RNA are unspliced. Differentiation of $Xist^{129}$ exhibits higher splicing efficiency than $Xist^{CAS}$, contributing to it being preferentially silenced during XCI. Additionally, our single cell analysis shows that allelic choice of Xist splicing is linked to which X-chromosome is silenced.

Introduction

Dosage compensation in females is accomplished by the epigenetic silencing of one of the X-chromosomes in females during early development. Once established in the embryo, the inactivated X-chromosome is maintained throughout development.

In female placental animals, one of the X-chromosomes is silenced through the process of X-chromosome inactivation. This occurs in two phases, imprinted and random XCI. Imprinted XCI is mediated by expression of *Xist* from paternal chromosome during the 2 to 8 cell stage of development. Cells that develop into the extraembryonic tissues and the trophoectoderm will maintain imprinted XCI. The cells that will undergo differentiation and cell-lineage specification into the embryo undergo X-chromosome reactivation to remove the epigenetic marks on the paternal X-chromosome in the epiblast cells of the blastocyst. Differentiation of the epiblast induces the second phase of silencing, called random X-chromosome Inactivation. During random XCI, either the paternal or the maternal X-chromosome will be silenced. Random XCI is thought to be a multi-step process of first counting of the X-chromosomes, stochastic choice of which X-chromosome to inactivate, expression of *Xist* from inactive X-chromosome, followed by initiation, spreading, and maintenance of the inactive state on the inactive X-chromosome.

Theoretically, silencing of the Xp or Xm in an individual should be found at equal ratios in cells, however this is not the case in cells derived from

heterogenic backgrounds. In mice where both X-chromosomes are from homogenic backgrounds, the ratio of inactive Xp to Xm is closer to 50:50. However in heterozygous mice, skewing of XCI is observed (Calaway et al., 2013). Extent of skewing of XCI has been dependent on the alleles of the X controlling element (Xce) identified by Cattanach (Cattanach and Isaacson, 1967). To date four alleles of the Xce has been described based on how often a specific X-chromsome allele remains active. Currently 4 main Xce alleles have been identified according to likelihood of remaining active: $Xce^a < Xce^b < Xce^c < Xce^d$ (Cattanach, 1991; Johnston and Cattanach, 1981; West and Chapman, 1978; Cattanach and Williams, 1972; Cattanach, Pollard, and Perez, 1969; Cattanach and Isaacson, 1967). *Mus musculus 129 SvJaeJ* (129) mice have a Xce^e allele versus the Xce^c allele found in *Mus castaneous (CAS)* and F1 2-1 ES cells have shown XCI skewing, with the X-chromosme derived from 129 mice being silenced ~70% of the time.

How the Xce functions in XCI skewing is not well understood however the locus candidate region has been identified as overlapping the *Xic* region that expresses *Xist* and *Tsix*, suggesting a potential regulation of the *Xic* by the Xce (Chadwick et al., 2006). Currently there are three favored models of X-chromosome silencing 1) The blocking factor mode consisting of a blocking factor expressed on autosomes that can prevent XCI on the active X-chromosome, 2) The two factor model in which autosomes express a blocking factor that will competently bind to a trans-acting competent factor from the X-

chromosome and the non bound factor will inactivate the X-chromosome 3) The sensing and choice model : after cells differentiate the Xpr region on the two X-chromosomes pair and the *Tsix/Xite* regions determine which X-chromosome to silent (reviewed by Starmer and Magnuson, 2009). The stochastic model suggested by Monkhorst and colleagues proposes that a *trans-acting* factor is produced from autosomes to activate *Tsix* and while an unknown X-linked gene produces a *trans-acting* factor that promotes *Xist* (Monkhorst et al., 2008). If the concentration of *Xist* expressed is sufficient, XCI will occur. The stochastic model suggests that binding sites on each of the X-chromosomes will determine which X-chromosome will silence. For example if one of the X-chromosome is more sensitive to the *Xist* promoting factor, it will preferentially be silenced.

Each of the current proposed models suggest that a trans-acting factor is involved in the choice of the X-chromosome to silence, however the mechanism of this regulation is still unknown. To understand skewing and gain more insight on potential mechanisms of how the choice of which X-chromosome to silence is made, we focused our studies the hybrid F1 2-1 ES cells derived from a cross of *Mus musculus 129SvJaeJ* and *Mus musculus castaneous* EiJ mouse strains. These cells are the favored models for studying XCI because the F1 2-1 ES cell line preferentially silences the X-chromosome derived from the 129 allele when differentiation is induced.

If trans-acting factors are required for inactivating the X-chromosome as the previous models suggest, there is still the remaining question of what

mechanism promotes the inactivation of one X-chromosome versus the other and how this relates to known existence of the Xce role in X-inactivation skewing.

In the previous two chapters I observed splicing as an important regulatory checkpoint for XCI induction. Allele specific splicing has been observed in mRNA, considering *Xist* is processed very similar mechanism to mRNA splicing, splicing may be one of the factors that determine which X-chromosome is silenced based on which allele is more efficient at splicing of the *Xist* gene (Li et al., 2012). We found that both $Xist^{129}$ and $Xist^{CAS}$ are unspliced in female F1 2-1 undifferentiated ES cells. Upon differentiation $Xist^{129}$ exhibits higher splicing efficiency than $Xist^{CAS}$, likely playing a role in the preferential inactivation of the X^{129} chromosome. Analysis of X-linked genes, we found that the choice of which X-chromosome to silence is linked to which X-chromosome is more efficiently spliced.

Results

Allelic difference in Xist RNA splicing efficiency is associated with skewed Xist induction

Differentiation induced X-inactivation is random in mice, however has been found to be skewed in certain female inbred lines. Female mice carrying two X-chromosomes from the same genetic background display random XCI. However in mice heterozygous for the X-chromosomes from different genetic backgrounds, skewed XCI is observed. Female hybrid mice will preferentially silence one X-chromosome over the other and it has been associated with the strength of the X chromosome controlling element allele (Calaway, 2013). Strains with weaker Xce, will be preferentially silenced in hybrid mice, however the mechanism is currently unknown.

For example, in the *129/Castaneous* hybrid mice, the 129 line has been shown to carry the weakest allele Xce^{a} compared to the stronger Xce^{c} found in Castaneous mice. This stochastic allelic difference at the X-chromosome is associated with the preferential expression of *Xist* from the X^{129} rather than from the X^{CAS}. The mechanism of skewed XCI is currently unclear, however it likely involves different events between the *Xic* and the Xce during the choice and initiation step of XCI. Sequencing of all the exon-intron boundaries of *Xist* using genomic DNA from *M. musculus 129SvJaeJ* and *M. musculus castaneus EiJ*, I observed many deletions and SNPs. between the two alleles. I observed in Chapter 2 and 3 that splicing appears to be an early regulator of *Xist* regulation.

We hypothesize that nonrandom *Xist* induction was associated with allelic differences in RNA splicing.

In order to measure both unspliced and spliced *Xist* variants derived from the 129 and CAS alleles, a polymorphic site located in an exon close to the 5' or 3' splice site and compatible with a splicing assay is needed. We cloned and sequenced the exon-intron boundaries spanning the entire *Xist* genomic sequence using genomic DNA from *M. musculus 129SvJaeJ* and *M. musculus castaneous* wildtype mice. We found a single nucleotide polymorphism (SNP) located in exon 3 close to the 3' splice site which generated a Tfi1 restriction enzyme recognition site only in the CAS allele (Figure 11A and B).

To validate the new SNP, we PCR amplified the sequences flanking the polymorphic site in both *M. musculus 129SvJaeJ* and *M. musculus castaneus EiJ* genomic DNA. Primers that annealed to Exon 2 (E2-F2) and Exon 3 (E2E3-R) produced an uncut product of 281nt. Primers that annealed to Intron 1 (I1E2-F) and Exon 3 (E2E3-R) yielded products that were 408 nt in length (Figure 11C). Amplified PCR products from the *CAS* allele were sensitive to Tfi1 restriction enzyme digest and were completely digested resulting in faster migrating bands (232 bp and 49 bp for the E2F2 amplicon and 359 bp and 49 bp for the I1E2-F amplicon). The *129* allele was resistant to Tfi1 restriction digestion resulting in slower migrating uncut amplicons (408 bp for E2F2 amplicon and 281 bp for I1E2-F amplicon). These results validated the use of Tfi1 digestion of RT-PCR products to distinguish $Xist^{129}$ and $Xist^{CAS}$ RNA in our splicing assay.

To examine the splicing efficiency of *Xist*^{CAS} AND *Xist*^{129} RNA, we preformed RT-PCR followed by restriction enzyme fragment length polymorphism (RFLP) analysis for to quantify each allele's expression based on the Tfi1 site. In undifferentiated F1 2-1 embryoid bodies, we observed primarily unpsliced forms of both the *Xist*^{CAS} and *Xist*^{129} transcript, consistent with our previous data that *Xist* is not efficiently spliced in undifferentiated cell conditions. Analysis of the E2F2 amplicon shows a steady decrease of the unspliced transcript originating from the *Xist*^{129} transcript that correlates with the steady increase of the spliced *Xist*^{129} transcript at day 4 and day 8. We also observed similar results in the *Xist*^{CAS} transcript, however the majority of the spliced transcript originated from the *Xist*^{129} (Figure 11D). Although we see both *Xist*^{CAS} and *Xist*^{129} RNA display increased splicing efficiency, the *Xist*^{129} splicing ratio (spliced/unspliced) was significantly higher than that of *Xist*^{CAS} (Figure 11E). This data suggested more efficient splicing is associated with preferential expression of *Xist*^{129} RNA.

To further test the different splicing efficiencies between *Xist*^{CAS} and *Xist*^{129,} we directly measured the premature unspliced *Xist* RNA. I1E2-F and E2E3-R primers only amplified the premature unspliced *Xist* RNA. Similar to the results seen in with E2-F2 and E2E3-R primers, precursor *Xist*^{129} was consistently lower than *Xist*^{CAS}. Quantification of the *Xist*^{129} / *Xist*^{CAS} ratio before and after differentiation also showed that *Xist*^{129} precursor RNA was consistently lower than *Xist*^{CAS} precursor RNA. Additionally the *Xist*^{129}/*Xist*^{CAS} ratio decreased

with differentiation, showing that *Xist*^{129} is spliced more efficiently

Allelic specific splicing of Xist is linked to the choice of which X-chromosome to silence

To test whether allele specific splicing of *Xist* was related to the choice of which to which chromosome is silenced during X-inactivation, we compared allelic splicing of *Xist* to allelic expression of X-linked genes. X-linked genes to use in the analysis had to follow two conditions 1) had an allelic specific restriction enzyme site and 2) was a target of X-chromosome inactivation. Following these two conditions, we performed Restriction fragment length polymorphism analysis on XCI target genes *G6pdx* and *Ogt*. Since XCI results in a population of cells that will either have the X^{129} or the X^{CAS} silenced, we needed to isolate single cells that each had one X-chromosome silenced to determine the splicing patterns of each cell. We differentiated F1 2-1 hybrid cells and isolated single cells from embryoid bodies by fluorescent activated cell sorting (FACS). Using Smart-Seq, we isolated cDNA from single intact cells and performed restriction fragment length polymorphism analysis to determine allele specific X-linked gene expression (Figure 12A).

X-inactivation is suggested to be a sequential process (reviewed in Goto and Monk, 1998; Takagi et al., 1982; Monk and Harper, 1979).

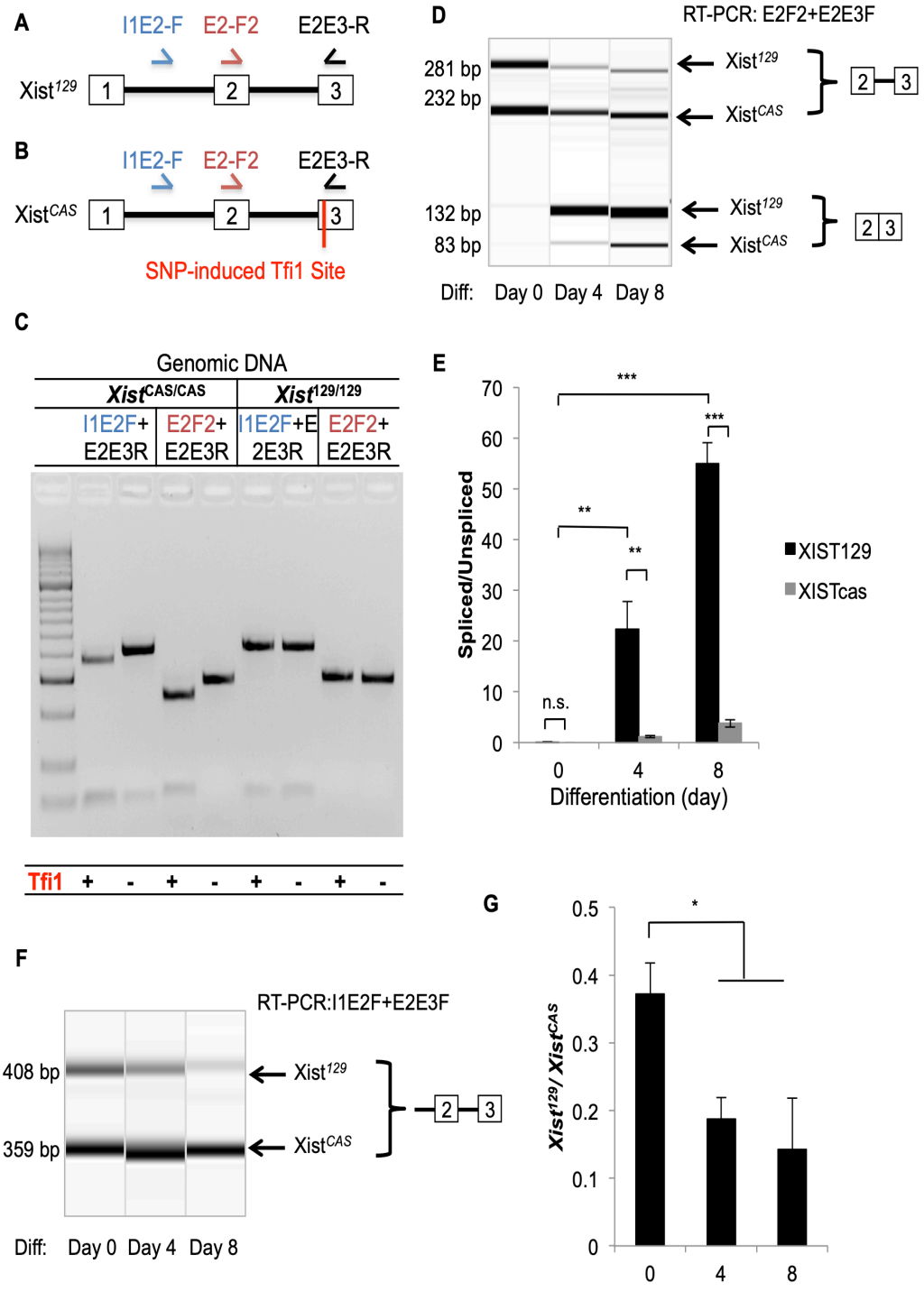
A
I1E2-F
E2-F2
E2E3-R
Xist129
1
2
3
B
I1E2-F
E2-F2
E2E3-R
XistCAS
1
2
3
SNP-induced Tfi1 Site
C
Genomic DNA
XistCAS/CAS
Xist129/129
I1E2F+ E2E3R
E2F2+ E2E3R
I1E2F+E 2E3R
E2F2+ E2E3R
Tfi1 + - + - + - + -
D
RT-PCR: E2F2+E2E3F
281 bp
232 bp
132 bp
83 bp
Xist129
XistCAS
Xist129
XistCAS
2 3
2 3
Diff: Day 0 Day 4 Day 8
E
Spliced/Unspliced
70
60
50
40
30
20
10
0
***
***
**
**
n.s.
0
4
8
Differentiation (day)
XIST129
XISTcas
F
RT-PCR:I1E2F+E2E3F
408 bp
359 bp
Xist129
XistCAS
2 3
Diff: Day 0 Day 4 Day 8
G
Xist129/ XistCAS
0.5
0.4
0.3
0.2
0.1
0
*
0
4
8

Figure 11. Allelic difference in *Xist* splicing is associated with non-random *Xist* induction (A,B) Schematics of the primer locations and allele-specific restriction enzyme digestion for distinguishing the unspliced and spliced *Xist* transcripts from the $Xist^{129}$ and $Xist^{CAS}$ alleles. SNP induced Tfi1 cut site is present in exon 3 of the $Xist^{CAS}$ allele. (C) A representative gel showingTfi1 specifically cuts exon 3 of $Xist^{CAS/CAS}$ genomic DNA but not $Xist^{129/129}$ DNA. Genomic DNA was amplified using the I1E2F or the E2F2 forward primer and the E2E3R primer before Tfi1 treatment. (D) Splicing assay detects allele-specific expression of unspliced and spliced transcripts from the $Xist^{129}$ and $Xist^{CAS}$ alleles before and after differentiation. (E) Quantification of intron 2 splicing ratios of $Xist^{CAS}$ and $Xist^{129}$ transcripts, respectively. (F) A representative RT-PCR virtual gel image shows the expression of unspliced $Xist^{CAS}$ and unspliced $Xist^{129}$ transcripts using I1E2-F and E2E3-R primers. (G) Ratios of unspliced $Xist^{129}$ transcript relative to unspliced $Xist^{CAS}$ transcript upon differentiation. Error bars indicate S.E.M of 3 biological replicates

In female ES cells three patterns of *Xist* expression have been observed: 1) low level bi-allelic expression of *Xist* from both active X-chromosomes characteristic of undifferentiated ES cells, 2) differential bi-allelic expression of *Xist* that changes over time, and 3) mono-allelic expression of *Xist* from one X-chromosome (Panning, Dausmann, and Jaenisch, 1997). In ES cells low levels of bi-allelic expression of *Xist* was characteristic of undifferentiated cells. Cells induced to embryoid bodies lose this bi-allelic expression of *Xist* and begin to display differential bi-allelic expression of *Xist*, suggesting an intermediate phase in which the selected Xi begins to be more efficient at expressing *Xist* and silencing X-linked genes, how this is regulated is currently unknown. Cells don't begin to display mono-allelic expression of *Xist* until day 6 after *Xist* dependent XCI has already irreversible and XCI is fully established.

In our data we found some single cells expressed *Xist* from both X-chromosomes (Figure 12B-D) and some did not (Figure 13A-B). Some of the cells expressed unprocessed *Xist* from both X-chromosomes. This appears to follow with the intermediate stage as described by Panning's group. In those cells although both transcripts are expressed, only the $Xist^{129}$ appears to be spliced. This data suggests that these cells likely fall into category two and fall into the intermediate stage of *Xist* inactivation. The cells that only express *Xist* from one chromosome only express the $Xist^{129}$, consistent with the skewing that has been previously seen in F1 2-1 ES cells. These EBs were collected from 6-8 days and based on previous groups work likely represent the cells that have

enter the *irreversible* phase of XCI when cells have committed to inactivating one of the X-chromosomes (Wutz and Jaenish, 2000).

In F1 2-1 cells there is a strong skewing of XCI that preferentially silences the X^{129} allele. In mature cells, assessing allelic choice cannot be fairly assessed because at this time point the XCI is established and the choice step of XCI has been made. In intermediate cells, bi-allelic transcription is still occurring and XCI is *Xist* dependent. Using a time point when transcription and processing are still occurring will allow for more accurate analysis of splicing of each allele.

We found that all intermediate cells that predominantly spliced $Xist^{129}$ also expressed X-linked genes from the X^{CAS} chromosome (Figure 12B-D). In 12 out of 13 cells (all except #5) spliced $Xist^{CAS}$ is not detectable, suggesting a strong selection for allelic specific splicing of $Xist^{129}$. Our RFLP analysis of these cells also showed strong expression of X-linked genes *G6pdx* and *Ogt* from the X^{CAS} chromosome suggesting silencing of the same genes on the X^{129} allele (Figure 12C-D). Cell #5 had a very weak spliced $Xist^{CAS}$ and appears to express a faint band for *G6pdx* and *Ogt* from both X^{129} and X^{CAS} alleles. Cell #5 expressed significantly more *Ogt* from the X^{CAS} allele consistent with XCI induction on the X^{129}. It is likely that Cell #5 represents a cell that is early in the intermediate phase of XCI and is on its way to splice exclusively $Xist^{129}$ and silence the X^{129} exclusively.

The probability of observing 12-13 cells exclusively spliced $Xist^{129}$ and inhibit the X^{129} is very small, and greater than what can be guessed by chance.

In all of the spliced cells the spliced *Xist* and the X-linked gene originated from opposing X-chromosomes (Figure 13A-B). Our observations suggest that allele-specific splicing is intrinsically linked to the choice of the inactive X-chromosome. Additionally since the choice is made by the mature phase, our data suggests that splicing occurs during the intermediate phase or during the phase when XCI is both *Xist* dependent and reversible. This is also consistent with our data that *Xist* splicing is required for *Xist* activity to induce XCI.

Based on data, I propose that during the undifferentiated phase, *Xist* is expressed from both alleles, however it is lowly expressed. As differentiation is induced and cells enter the "intermediate" phase and still expressed *Xist* from both alleles, *Xist* more efficiently spliced on the X^{129}. As the intermediate phase proceeds, transcription is increased from the X^{129} chromosome as is splicing due to more efficient binding to the splicing machinery potentially assisted by the presence of PTBP1. Spliced *Xist* is needed to recruit the repressive machinery to the X-chromosome. While this is happening on the X^{129} chromosome, transcription and silencing is inhibited on the X^{CAS} chromosome due to inefficient levels of spliced *Xist* to silence the $X^{CAS,}$ during the critical window of *Xist* dependent XCI resulting in silencing of the $X^{129.}$

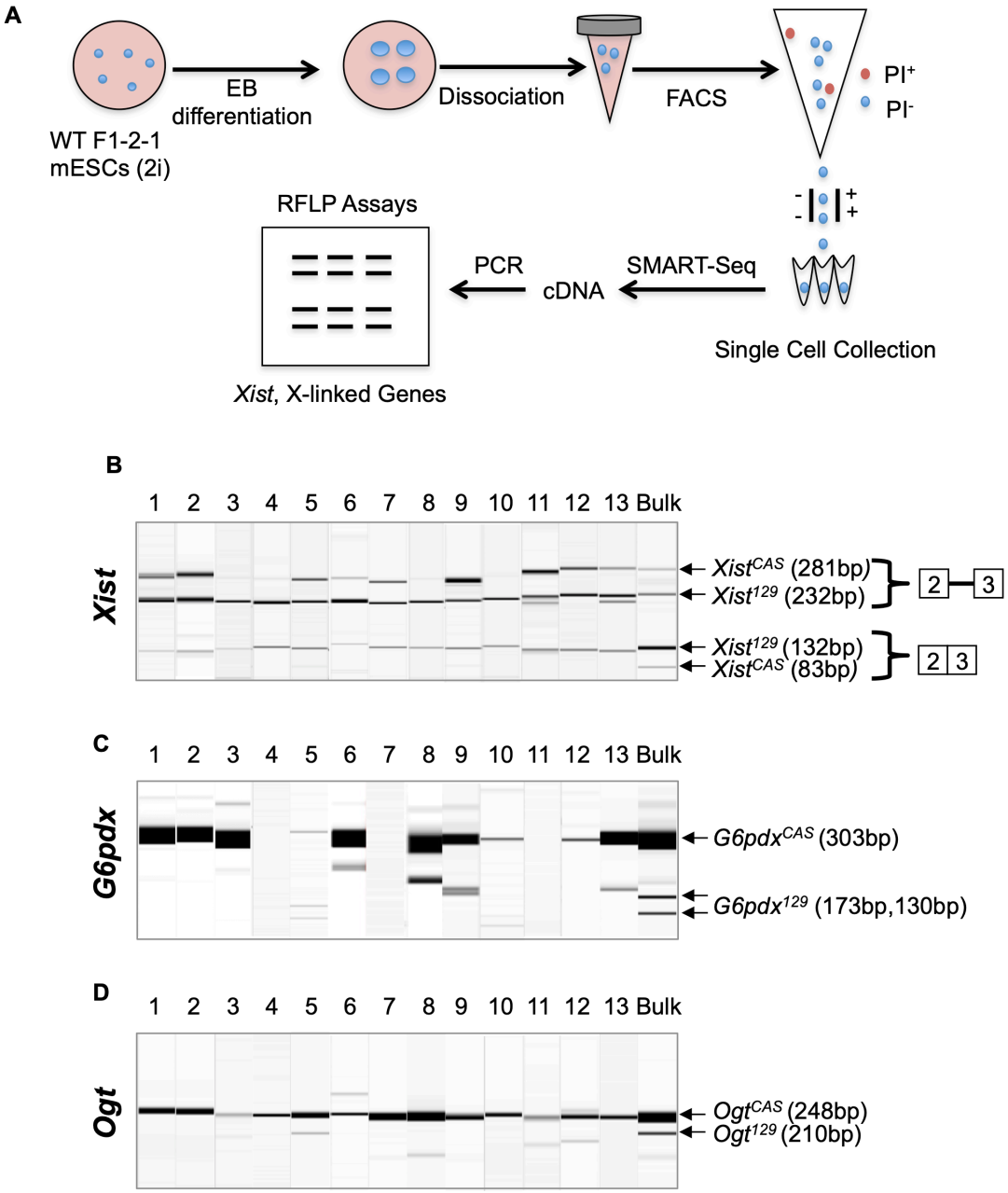
A
WT F1-2-1
mESCs (2i)
EB
differentiation
Dissociation
FACS
PI+
PI-
Single Cell Collection
SMART-Seq
cDNA
PCR
RFLP Assays
Xist, X-linked Genes
B
1 2 3 4 5 6 7 8 9 10 11 12 13 Bulk
Xist
Xist^CAS (281bp)
Xist^129 (232bp)
2 3
Xist^129 (132bp)
Xist^CAS (83bp)
2 3
C
1 2 3 4 5 6 7 8 9 10 11 12 13 Bulk
G6pdx
G6pdx^CAS (303bp)
G6pdx^129 (173bp,130bp)
D
1 2 3 4 5 6 7 8 9 10 11 12 13 Bulk
Ogt
Ogt^CAS (248bp)
Ogt^129 (210bp)

Figure 12. Allele specific *Xist* splicing is linked to the choice of inactive X chromosome (A,B) Schematics of the primer locations and allele-specific restriction enzyme digestion for distinguishing the unspliced and spliced *Xist* transcripts from the $Xist^{129}$ and $Xist^{CAS}$ alleles. SNP induced Tfi1 cut site is present in exon 3 of the $Xist^{CAS}$ allele. (C) A representative gel showingTfi1 specifically cuts exon 3 of $Xist^{CAS/CAS}$ genomic DNA but not $Xist^{129/129}$ DNA. Genomic DNA was amplified using the I1E2F or the E2F2 forward primer and the E2E3R primer before Tfi1 treatment. (D) Splicing assay detects allele-specific expression of unspliced and spliced transcripts from the $Xist^{129}$ and $Xist^{CAS}$ alleles before and after differentiation. (E) Quantification of intron 2 splicing ratios of $Xist^{CAS}$ and $Xist^{129}$ transcripts, respectively. (F) A representative RT-PCR virtual gel image shows the expression of unspliced $Xist^{CAS}$ and unspliced $Xist^{129}$ transcripts using I1E2-F and E2E3-R primers. (G) Ratios of unspliced $Xist^{129}$ transcript relative to unspliced $Xist^{CAS}$ transcript upon differentiation. Error bars indicate S.E.M of 3 biological replicates. ns, p-value>0.05; *, p-value < 0.05; **, p-value< 0.01; ***, p-value < 0.001.

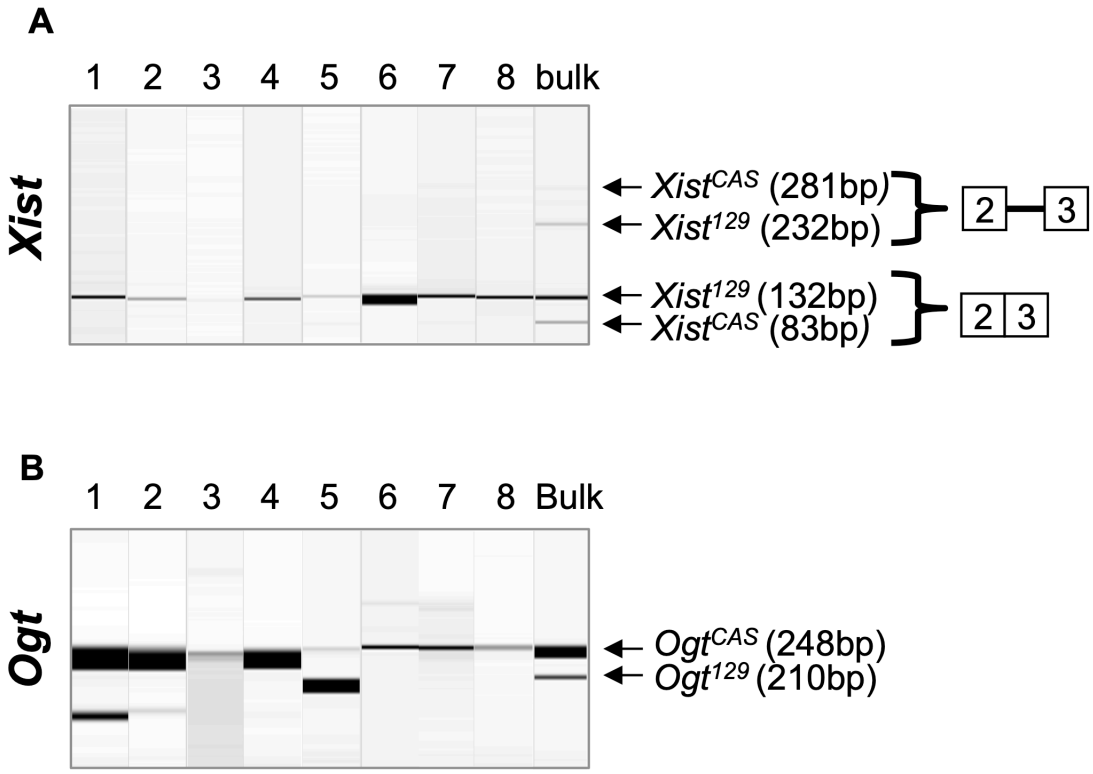

Figure 13. Spliced *Xist* and X-linked genes originate from different X chromosomes in mature embryoid bodies A) Restriction fragment length polymorphism analysis combined with splicing assay detects allelic specific expression of unspliced and spliced *Xist* transcripts from the X^{129} and X^{CAS} chromosomes in mature cells. F1 2-1 cells were differentiated using the embryoid body method for 8 days (with added retinoic acid from day 4). (B) Restriction fragment length polymorphism analysis detects allelic expression of X-linked genes *Ogt* from the X^{129} and X^{CAS} chromosomes. *G6pdx* expression was not detectable in these samples.

Discussion

In Chapter 2 we showed that differentiation induces post-transcriptional splicing of *Xist* during the early phases of XCI induction and in Chapter 3 we showed that PTBP1 is one of the RBPs involved in this process. In this chapter we found that during skewed XCI, $Xist^{129}$ is preferentially spliced during the intermediate phase of XCI and found only $Xist^{129}$ is fully spliced when XCI is established. *Xist* is expressed from both X-chromosomes, however differentiation induced splicing is more efficient on the *Xist* transcript derived from the future inactive X-chromosome suggesting splicing is a crucial regulatory step in deciding which X-chromosome to silence. This data suggests that *Xist* expressed from the future Xi may be more efficient at recruiting transacting factors involved in splicing of the *Xist* transcript and the silencing of *Xist* activators on the active X-chromosome. This results is more spliced *Xist* being produced by future inactive X and its eventual silencing. The data suggests that splicing efficiency may be a crucial mechanism involved in regulating *Xist* dependent X-chromosome inactivation.

In this chapter, we have examined whether splicing played a role in allelic choice of which X-chromosome to silence during induction of X-chromosome inactivation. Using a combination of RT-PCR from individual ES cells and restriction fragment length polymorphism analysis (RFLP), our studies in hybrid X^{129}/X^{CAS} ES cells showed a link between differentiation induced splicing and skewed XCI. In our hybrid F1 2-1 cells $Xist^{129}$ and $Xist^{CAS}$ were differentially spliced suggesting an existence to splicing factors or elements that appear to be

sensitive to differentiation conditions since both remain primarily unspliced in undifferentiated ES cells. Based on our results in Chapter 3, these splicing factors may interact with PTBP1 since loss of PTBP1 resulted in significant reduction of splicing.

XCI skewing has been attributed to the preferential silencing of the X-chromosome that has the weakest X controlling element however the mechanism of this regulation is still a mystery (Cattanach, 1991; Johnston and Cattanach, 1981; West and Chapman, 1978; Cattanach and Williams, 1972; Cattanach, Pollard, and Perez, 1969; Cattanach and Isaacson, 1967). There is an intrinsic bias to express *Xist* from the X^{129} during XCI induction due the presence of the Xce^{a} allele, which is the weakest of the 4 identified X controlling element alleles.

Although it has been proven in other studies that hybrid F1 2-1 ES cells skew inactivation in favor of silencing the X^{129}, the mechanism of this silencing hasn't been studied in depth. Our data is the first to show differential splicing between the $Xist^{129}$ and $Xist^{CAS}$ as one of the potential mechanisms of the choice of which X-chromosome to silence. Some studies have suggested that structural variances within segmented duplications present in the candidate Xce locus may be the cause of these variances in Xce allele strength due to potential differences in binding sites for transacting factors (Calaway et al., 2013;, Yalcin et al., 2011). Our work expands upon previous work and suggests that these transacting factors may be involved in splicing and that $Xist^{129}$ may be more efficient in recruiting such factors during the intermediate phase of XCI.

One caveat to this data however is that during single cell collection we were not able to isolate cells that expressed *Xist*^{CAS}. Previous studies showed that XCI preferentially silences the X^{129} about 75% of the time in hybrid X^{129}/X^{CAS} cells, however in our bulk cell analysis the ratio of splice *Xist*^{129}/ *Xist*^{CAS} was closer to10:1 (Calaway, 2013). Therefore in our cells, capturing the *Xist*^{CAS} is difficult prior to any additional selection steps. We used three stringent gates to isolate viable single cells using FACS, resulting in only about 1% being selected (refer to methods). If inactivating X^{CAS} had any changes on the morphology, viability, or size of the cells, these may have been selected against during FACS. Embryoid bodies are made from suspension cultures resulting in a heterogeneous population of cell aggregates of different shapes and sizes, which may influence differentiation of individual cells. Differentiation of ES cells is largely directed by the microenvironment that the cells are exposed to, and the cells position in embryoid bodies plays a role in the differentiation status and cell fate of those cells (Karp et al., 2011; Hamazaki et al., 2004). Our cells were isolated from embryoid bodies, therefore this variance in allelic splicing status we see in our data may be partly due to this inherent variance found in ES cells isolated from EBs. For example, the cells that appear to be in the intermediate phase of XCI induction may have been less differentiated or in an earlier cell state when differentiation was induced than cells that have completed splicing of *Xist* by the time of our cell collection.

Splicing of lncRNAs are intrinsically less efficient than protein coding genes, splicing is highly dependent on 5' splice site strength and thymidine rich polypyrimidine tracts due to lncRNA's inability to efficiently recruit the splicing machinery (Krchňáková et al., 2018). SR protein binding to lncRNAs is less efficient than in protein coding genes. It is possible that the Xce^{a} has more binding sites for trans acting SR Proteins and those proteins interact more efficiently with certain factors that can bind to $Xist^{129}$ more efficiently than $Xist^{CAS}$.

Deletion of the A-repeat on $Xist^{129}$ in exon 1 of *Xist* results in depletion of spliced *Xist* and failure to induce XCI on the X^{129}. Depletion of ASF/SF2, which binds to the A-repeat, also results in failure to spliced *Xist.* Mutant cells lacking the A-repeat expressed similar levels of the primary transcript to the WT. These results are consistent with spliced *Xist* being needed for functional XCI (Royce Tolland et al., 2010).

In both the Royce-Tolland study and ours, *Xist* is either completely spliced or completely unspliced. Binding of RBPs involved in splicing regularly effects the splicing of the exon it is bound to. However both the Royce-Tolland study and our study show that loss of splicing effects all of exons. Our data is consistent with the notion that *Xist* splicing might be essential for *Xist* to form the high order structures required for interactions with the X-chromosome. This would be consistent with the need for *Xist* to be completely spliced.

The observation that both unspliced $Xist^{129}$ and $Xist^{CAS}$ are expressed in the intermediate phase, however only spliced $Xist^{129}$ is observed in the mature

phase suggests that splicing of *Xist* on the Xi results in inhibition of *Xist* expression on the Xa. These data is consistent with splicing efficiency contributing to allelic choice of which X-chromosome to silence during induction of XCI. Determining the source of splicing repression on the Xa and activation on the Xi and the binding partners involved could reveal novel splicing regulatory mechanisms of *Xist.*

Both the importance and mechanisms of long noncoding RNA splicing is still a mystery but appears to be primary sequence dependent to interact with RBPs. The work in this chapter suggest a role of splicing for the lncRNA *Xist,* however may provide insight in the importance of splicing of lncRNAs to their function and specificity that may be applied to the studies of other lncRNAs.

Materials and Methods

Differentiation of embryonic stem cells

To generate embryoid bodies, feeder independent WT ES F1 2-1, WT C57B6, and *Ptbp1-/-* ES cells were plated at a density of 2.5 x 10^6 cells on 0.5% Agarose coated polystyrene dishes (Grenier Bio-One cat. no. 633102) and maintained in DMEM supplemented with 10% FBS, 1% Glutamax, 1% Non-Essential Amino Acids (GE Healthcare, cat. no. SH3023801), 1% Sodium Pyruvate (Lonza cat. no. 13-115E), and 0.1mM betamercaptoethanol without mLIF for up to 8 days. To promote embryoid body differentiation, 2 μM retinoic acid was added to the media of embryoid bodies on days 4 thru 8 and replenished every two days. For western blots,. To minimize cell death, media was replenished every 2 days

RNA Isolation, cDNA synthesis, and RT-qPCR

Trizol isolation of RNA was performed following manufacturer's protocol (Life Technologies). Isolated RNA was treated with 4 units of Turbo DNase (Ambion) at 37 °C for 35 minutes to degrade any remaining DNA. RNA was purified using Phenol-Chloroform (pH 4.5, VWR cat. no. 97064-744). RNA concentrations were measured using Nanodrop 2000c (Thermo Fisher). First-strand cDNA was generated using 1 μg RNA and 200 units of M-MLV reverse transcriptase (Promega cat. no. M1705) with a RT primer in a 20 μl reaction following the manufacturer's protocol. For *Xist* specific analysis, the RT primer was a *Xist* gene specific reverse primer (E6-E7 reverse) targeting exon 7 (50μM). For X-linked and differentiation genes random hexamers (30μM) were used. For

polyadenylated (poly(A)+) RNA, oligodT primers (50μM) were used for cDNA generation. The RT-qPCRs were conducted using a Quantstudio 6 Flex Real-Time PCR instrument, 2x Syber Master Mix (Life Technologies) following manufacturer's protocol. Three technical replicates of all samples were performed for each biological replicate. An additional no-template control was included in every experiment. Outliers were defined by a Ct variance higher than 0.5. Analysis was done with Quanstudio 6 software and Microsoft Excel. Statistics test were based on ΔΔCt values.

Allele specific splicing assay and restriction fragment length polymorphism analysis

PCR reactions were performed under standard conditions for Taq polymerase (New England Biolabs) on RT products derived from WT F1 2-1 $X^{CAS}X^{129}$ ES cells. For *Xist* 30-35 cycles was used. For X-linked genes *G6pdx* and *Ogt*, 35-40 cycles were used. PCR products were digested with 2.5 units of Tfi1 (New England Biolabs cat. no. R0546L) that specifically cut exon 3 on the $Xist^{CAS}$ allele in a 25 μl reaction for 2 hours at 65°C. For single cell studies, PCR products were digested with 2.0 units of restriction enzymes overnight at 37°C. For analysis of X-linked genes, PCR products were digested with Ava1 (NEB, cat. no. R0152S) or ScrF1 (NEB, cat. no. R0110S) to cut *Ogt* or *G6pdx* respectively (Lin et. al. 2007, Minkovsky and Plath 2013, 18076287). To remove restriction enzymes, samples were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter cat. no. A63881) following the manufacturer's protocol. Eluted

PCR products were diluted 1:3 -1:5 when necessary before analysis on the Qiagen QIAxcel Capillary Electrophoresis system (Qiagen) using the QIAxcel DNA Screening Kit (Qiagen, cat. no. 929004) following the manufacturer's protocol. The QIAxcel Alignment Marker QX 15-600bp (Qiagen, cat. no. 929530) and the Qiaxcel 25-500bp size marker (Qiagen, cat. no. 929560) were used in the analysis. Allelic ratios were calculated as a ratio of X^{129} and X^{CAS} allele specific *Xis*t as determined by digestion with Tfi1. Splicing ratios were calculated as the ratio of spliced and unspliced $Xist^{129}$ and $Xist^{CAS}$ specific PCR products. For confirmation of allele specific digestion by Tfi1, genomic DNA collected from the brain of wildtype *mus musculus castaneus* EIJ and *mus musculus domesticus* 129S4/SvJaeJ mice (The Jackson Laboratory) were used.

Single Cell Sorting and cDNA preparation

WT F1 2-1 ES cells were differentiated into embryoid bodies for 6 days. Cells were washed two times with 1x PBS and stained with propidium iodide (PI) (1:1500) (BD Pharmigen, cat. no. 556547) to identify apoptotic cells. We used three stringent gating steps to select single cells: (1) negative staining of propidium iodide as an indicator of healthy cells; (2) typical FSC and SSC settings to filter out cell debris and cell chunks; (3) linearity between SSC height and SSC area to ensure single cell (singlet in sorting). Single PI-negative WT F1 2-1 cells from embryoid bodies were then sorted into PCR strips using a MoFlo Astrios EQ (Beckman Coulter) according to manufacturers' protocol. Prior to cell sorting, quality calibrations (QC) were performed according to manufacturer’s

recommendation. The cDNA was prepared using SMART-Seq v4 Ultra Low Input Kit for 12 reactions (Clontech) according to manufacturer's protocol.

References

Calaway, J. D. *et al.* Genetic Architecture of Skewed X Inactivation in the Laboratory Mouse. *PLoS Genet.* (2013). doi:10.1371/journal.pgen.1003853

Cattanach BM, Pollard CE, Perez JN (1969) Controlling elements in the mouse X-chromosome. I. Interaction with the X-linked genes. Genet Res 14: 223–235

Cattanach BM, Williams CE (1972) Evidence of non-random X chromosome activity in the mouse. Genet Res 19: 229–240.

Cattanach, B. M. & Papworth, D. Controlling elements in the mouse: V. Linkage tests with X-linked genes. *Genet. Res.* (1981). doi:10.1017/S0016672300020401

Cattanach BMRC Identification of the Mus spretus Xce allele. (1991) Mouse Genome 89: 565–566.

Chadwick, L. H., Pertz, L. M., Broman, K. W., Bartolomei, M. S. & Willard, H. F. Genetic control of X chromosome inactivation in mice: Definition of the Xce candidate interval. *Genetics* (2006). doi:10.1534/genetics.105.054882

Hamazaki, T. Aggregation of embryonic stem cells induces Nanog repression and primitive endoderm differentiation. *J. Cell Sci.* **117,** 5681–5686 (2004).

Johnston, P. G. & Cattanach, B. M. Controlling elements in the mouse: IV. Evidence of non-random X-inactivation. *Genet. Res.* (1981). doi:10.1017/S0016672300020127

Karp, J. M. *et al.* Controlling size, shape and homogeneity of embryoid bodies using poly(ethylene glycol) microwells. *Lab Chip* **7,** 786–794 (2007).

Krchňáková ,Z. *et al.* Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5? splice-site sequences due to weak interactions with SR proteins. *Work. Aging Retire.* (2018). doi:10.11821/dlxb201802008

Li, G. *et al.* Identification of allele-specific alternative mRNA processing via transcriptome sequencing. *Nucleic Acids Res.* (2012). doi:10.1093/nar/gks280
Monk, M. & Gene, T. H. E. X. Regulation of X-Chromosome Inactivation in Development in Mice and Humans. *Microbiol. Mol. Biol. Rev.* (1998). doi:10.4314/ahs.v13i4.17

Monk, M. & Harper, M. I. Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos [18]. *Nature* (1979). doi:10.1038/281311a0

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X Inactivation Counting and Choice Is a Stochastic Process: Evidence for Involvement of an X-Linked Activator. *Cell* (2008). doi:10.1016/j.cell.2007.12.036

Panning, B., Dausman, J. & Jaenisch, R. X chromosome inactivation is mediated by Xist RNA stabilization. *Cell* **90,** 907–916 (1997).

Rastan, S. Primary non-random X-inactivation caused by controlling elements in the mouse demonstrated at the cellular level. 139–147 (2019).

Royce-Tolland, M. E. *et al.* The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nat. Struct. Mol. Biol.* **17,** 948–954 (2010).

Simmler, M. C., Cattanach, B. M., Rasberry, C., Rougeulle, C. & Avner, P. Mapping the murine Xce locus with (CA)n repeats. *Mamm. Genome* (1993). doi:10.1007/BF00364788

Starmer, J. & Magnuson, T. A new model for random X chromosome inactivation. *Development* (2009). doi:10.1242/dev.025908

Takagi, N., Sugawara, O. & Sasaki, M. Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. *Chromosoma* (1982). doi:10.1007/BF00294971

Thorvaldsen, J. L., Krapp, C., Willard, H. F. & Bartolomei, M. S. Nonrandom X chromosome inactivation is influenced by multiple regions on the murine X chromosome. *Genetics* (2012). doi:10.1534/genetics.112.144477

West, J. D. & Chapman, V. M. Variation for X chromosome expression in mice detected by electrophoresis of phosphoglycerate kinase. *Genet. Res.* (1978). doi:10.1017/S0016672300018565

Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* (2000). doi:10.1016/S1097-2765(00)80248-8

Yalcin, B. *et al.* Sequence-based characterization of structural variation in the mouse genome. *Nature* **477,** 326–329 (2011).

Conclusions and Future Outlooks

Introduction

LncRNAs have shown to be very important in regulation of gene expression; they are regulated in every tissue and at specific times. The stringent regulation of lncRNAs suggests an evolutionary conserved mechanism to regulate gene expression. For how important lncRNAs appear to be, very little is known on how they are transcribed and processed. Many long noncoding RNAs appear to be spliced and processed in order to be functional, however very little is known regarding how splicing occurs. Advances in genomic technologies have provided more insight in how these RNA transcripts work but have also led to additional questions. *Xist*'s role in X-chromosome inactivation is one of the most understood functions of an lncRNA, however very little is known how *Xist* itself is regulated. The past few years have been focused on understanding each of the steps of XCI function, *Xist* transcriptional control, *Xist* spreading and localization, and repression. Recent works by many groups have identified known RNA binding proteins that interact with *Xist* (Vuong et al., 2016; Chu et al., 2015; Mchugh et al., 2015). Some of the proteins shed light on how *Xist* interacts with the X-chromosome, however some suggest roles in *Xist* transcript regulation. The identification of splicing factors PTBP1 and ASF2 to the E-repeat and the A-repeat respectively suggested splicing might be a method of transcriptional regulation of *Xist* during XCI, a field that hasn't been studied in depth.

Conclusions and Future Directions

In this thesis, we investigated how *Xist* is transcriptionally regulated during differentiation. The results from these studies reveal that while splicing is not the only regulation involved in *Xist* mediated XCI, splicing acts as a fail-safe mechanism to ensure that XCI is only induced on the future inactivate X-chromosome. Prior to differentiation, *Xist* is lowly expressed and unspliced, and therefore XCI is not induced and both X-chromosomes remain active. Upon differentiation and downregulation of pluripotency factors, *Xist* expression is upregulated. However once *Xist* is expressed, it must determine which X-chromosome to silence. Our data suggests, that one of the X-chromosomes is more efficient at splicing of *Xist* and the accumulation of spliced *Xist* on the Xi may inhibit further expression of *Xist* from the active X-chromosome leading to this choice. The need of *trans-acting* factors to splice lncRNAs has been suggested by other groups. Our data suggests that PTBP1 may be one of the *trans-acting* factors involved in splicing of *Xist* . However since PTBP1 has only been identified to bind to the E-repeat, a question as to how the other introns are spliced as the same time remains. It is possible that PTBP1 serves as a scaffolding protein required to bring other splicing factors to the *Xist.* This mechanism of action suggests that splicing of *Xist* could be mediated by how *Xist* efficiently binds to these factors. The 129 and CAS alleles of *Xist* are very different in their sequence and have the greatest skewing during XCI. It is

possible that *Xist*^{129} can more efficiently bind to PTBP1 or additional factors and therefore be spliced more efficiently.

Therefore my data proposes the following model where *Xist* is lowly expressed by both X-chromosomes, upon differentiation *Xist* recruits additional factors to mediate it's splicing. Once *Xist* levels reach a certain critical concentration, they recruit PTBP1 and additional factors. *Xist*^{129} is more efficient at binding to PTBP1 and other transacting factors involved in it's splicing and therefore is spliced more. As the spliced population of *Xist* from the X^{129} increases, *Xist* splicing and transcription is silenced on the future active X-chromosome.

Future studies will need to identify what binding partners to PTBP1 exist during XCI and if they also have binding sites with *Xist.* Once these binding partners are identified then experiments will need to focus on seeing if different X-chromosome alleles have differential binding. Although our data doesn't answer all questions it does provide insight into a not well-studied field.

References

Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* (2015). doi:10.1016/j.cell.2015.03.025

McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* (2015). doi:10.1038/nature14443

Vuong, J. K. *et al.* PTBP1 and PTBP2 Serve Both Specific and Redundant Functions in Neuronal Pre-mRNA Splicing. *Cell Rep.* (2016). doi:10.1016/j.energy.2018.10.173