The Pennsylvania State University

The Graduate School

College of Medicine

IDENTIFICATION OF A CTCF-INDEPENDENT INSULATOR THAT DELIMITS X-INACTIVATED FROM EXPRESSED TRANSCRIPTS WITHIN THE *UBA1* GENE

A Dissertation in

Biochemistry and Molecular Biology

by

Katie E. Prothero

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2010

The dissertation of Katie Eva Prothero was reviewed and approved* by the following:

Laura Carrel
Associate Professor of Biochemistry and Molecular Biology
Thesis Advisor
Chair of Committee

Ira Ropson
Associate Professor of Biochemistry and Molecular Biology

Jiyue Zhu
Associate Professor of Cellular and Molecular Physiology

Judith S. Bond
Distinguished Professor of Biochemistry and Molecular Biology
Head of the Department of Biochemistry and Molecular Biology

*Signatures are on file in the Graduate School

ABSTRACT

In female mammals, X chromosome inactivation (XCI) transcriptionally silences most genes on one X. Nonetheless, in humans, 15% of genes escape XCI and are expressed from both active and inactive Xs. Many of these escape genes are clustered, suggesting their expression is coordinately regulated. An attractive model is that these clustered escape genes are flanked by regulatory sequences such as insulators. To identify sequences that control XCI, I focused on a cluster of human escape genes that includes the Ubiquitin Activating Enzyme E1 gene, *UBA1*. *UBA1* has a distinctive gene structure with five alternative promoters and 5' untranslated exons that each splice to a single exon 2. Inactive X expression is also unique; two upstream exons are X-inactivated whereas two downstream exons assayed escape XCI. The 2.1 kb between X-inactivated and expressed exons is the smallest inactive X expression boundary and is a tractable region to identify sequences that regulate inactive X expression. To test the prediction that escape domains are regulated by chromatin insulators, eight overlapping constructs within the *UBA1* boundary were tested; a 330 bp sequence demonstrated insulator function. Despite similar gene structure and high sequence conservation, all mouse *Uba1* transcripts are X-inactivated and the locus lacks insulator activity. The *UBA1* insulator is novel; while the insulator-binding protein CTCF has been identified at one XCI expression boundary, CTCF binds the *UBA1* locus downstream of the insulator and appears unrelated to inactive X regulation. That the insulator is functional in XCI gene regulation is suggested by epigenetic modifications that specifically demarcate this sequence on the inactive X. Chromatin immunoprecipitation demonstrated that the *UBA1* insulator segregates heterochromatin from the euchromatic escape domain. Additionally, DNA methylation differences distinguish the human insulator; this sequence is hypomethylated on the inactive X, whereas the active X is heavily methylated, as is the orthologous sequence that lacks boundary activity on both Xs in mouse. The identification of a novel insulator at the *UBA1* expression boundary strengthens the model that escape domains are regulated by chromatin boundaries and pinpoints a region that can be further dissected to identify novel factors that bind and regulate XCI escape.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIAITONS

3C	Chromosome conformation capture
4C	Circularized chromosome conformation capture
Alu	Short interspersed nuclear element, Alu family (primate specific)
α-MEM	Alpha minimum essential media
Amp^R	Ampicillin resistance gene
B1	Short interspersed nuclear element, B1family (rodent specific)
bp	Base pair
cDNA	Complementary DNA
cen	Centromere
^oC	Degrees Celsius
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
CpG	CpG dinucleotides
C_t	Cycle threshold
Ctcf/CTCF	CCCTC-binding factor
Ddx3x/DDX3X	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked gene
Ddx3y/DDX3Y	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked gene
DMD	Differentially methylated domain
DMEM	Dulbecco's modified eagle medium
DNA	Deoxyribonucleic acid
Dnmt1/DNMT1	DNA (cytosine-5)-methyltransferase 1 gene

Dnmt3b/DNMT3b	DNA (cytosine-5)-methyltransferase 3 beta gene
DNMTase	DNA methyltransferases
EBF	Early B-cell factor
Eed	Embryonic ectoderm development gene
Eif2s3x/EIF2S3	Eukaryotic translation initiation factor 2, subunit 3 X-linked gene
EMSA	Electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA elements
ES	Embryonic stem
EST	Expressed sequence tag
Ezh1	Enhancer of zeste homolog 2 gene
FBS	Fetal bovine serum
FISH	Fluorescence *in situ* hybridization
G9a	Euchromatic histone-lysine N-methyltransferase 2
H19	H19, imprinted maternally expressed transcript gene
H2A	Histone 2A
H2AK119	Histone 2A lysine 119
H3	Histone 3
H3Ac	Histone 3 acetylation
H3K20	Histone 3 lysine 20
H3K27me3	Histone 3 lysine 27 trimethylation
H3K4	Histone 3 lysine 4
H3K9	Histone 3 lysine 9
H4	Histone 4
H4K20	Histone 4 lysine 20

HMTase	Histone methyltransferase
HP1	Heterochromatin protein 1
Hprt/HPRT	Hypoxanthine guanine phosphoribosyl transferase gene
Ifg2r/IGF2	Insulin-like growth factor 2 receptor gene
Jarid1c/JARID1C	Jumonji, AT rich interactive domain 1C gene
Jpx/Enox	Non-protein coding RNA 183 gene
kb	Kilobase
KRAB	Krüppel associated box family of proteins
L1	Long interspersed nuclear element subclass 1
LINE	Long interspersed nuclear element
LINE-1	Long interspersed nuclear element subclass 1
LTR	Long terminal repeat
MAR	Matrix attachment region
Mb	Megabase
Mecp2/MECP2	Methyl CpG binding protein 2 gene
μF	Microfarad
μg	Microgram
mg	Milligram
MIR	Mammalian interspersed repeat
μL	Microliter
mL	Milliliter
μM	Micromolar
mM	Millimolar

mya	Million years ago
Nanog	Nanog homeobox protein
Neo^R	Neomycin reporter gene
Oct4	Octamer 4 gene or POU class 5 homeobox 1
OFD1	Oral-facial-digital syndrome 1 gene
OMIM	Online Mendelian Inheritance in Man
PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
Pctk1/PCTK1	PCTAIRE cdc2 protein kinase 1 gene
PLXNA3	Plexin A3 gene
pmol	Picomole
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PR-Set7	Histone lysine N methyltransferase PR-set7
qRT-PCR	Quantitative real-time polymerase chain reaction
Rbm10/RBM10	RNA binding motif protein 10 gene
Ring1b	Ring finger protein 1b gene
RNA	Ribonucleic acid
Rnf12	E3 ubiquitin ligase encoded by the RL1M (ring finger protein, L1M domain interacting) gene
RT-PCR	Reverse transcriptase-polymerase chain reaction
SAR	Scaffold attachment region
SATB1	Special AT-rich sequence binding protein 1
SINE	Short interspersed nuclear element

siRNA	Small interfering RNA
SmcHD1	Structural maintenance of chromosomes flexible hinge domain containing 1 protein
SNP	Single nucleotide polymorphism
Sox2	SRY-box containing gene 2
SRY	Sex determining region of the Y chromosome gene
STS	Steroid sulfatase (microsomal), isosome S gene
Su(Hw)	Suppressor of hairy-wing
Suv39h1	Suppressor of variegation 3-9 homology 1 (histone-lysine N-methyltransferase Suv39h1)
Suv39h2	Suppressor of variegation 3-9 homology 2 (histone-lysine N-methyltransferase Suv39h2)
Suz12	Suppressor of zeste 12 homolog gene
Tsix/TSIX	Inactive X specific transcript, antisense gene
TSS	Transcription start site
UBA1/Uba1	Ubiquitin activating enzyme E1 gene
USF1	Upstream transcription factor 1
Usp11/USP11	Ubiquitin specific peptidase 11 gene
UTR	Untranslated region
Utx	Ubiquitously transcribed tetratricopeptide repeat, X linked gene
V	Volt
vEZF1	Vascular endothelial zinc finger 1
X	X chromosome
Xa	Active X
XAR	X added region

Xce	X controlling element
XCI	X chromosome inactivation
XCR	X conserved region
Xi	Inactive X
Xic/XIC	X inactivation center
XIST/Xist	Inactive X specific transcript gene
Xite/XITE	X inactivation intergenic transcription element
Xpr	X pairing region
Y	Y chromosome
Yy1	Yy1 transcription factor
ZNF157	Zinc finger protein 157 gene
ZNF41	Zinc finger protein 41 gene

ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible without the assistance, guidance and support of a large number of people who have assisted me throughout my studies at The Pennsylvania State University College of Medicine. First and foremost, I would like to thank my advisor, Dr. Laura Carrel. I truly appreciate both the countless conversations in her office and her infectious enthusiasm for research. Her expectation for quality work and her ability to make me think critically have driven me to become a better and more competent scientist.

Further, the completion of this project would not have gone as smoothly had it also not been for the professional and collegial attitude that existed in the lab and the hard work of the members of the lab. I would like to thank both current and former members: our two laboratory technicians Sarah Arnold-Croop and Melanie Moon, and the graduate students Jill Stahl, Lindsay Horvath, and Dr. Nan Li for being both good colleagues and good friends throughout this process. Our discussions throughout my studies were both constructive and entertaining and kept me energized through the good times and bad.

I also offer my sincere thanks to the members of my thesis committee: Dr. Sergei Grigoryev, Dr. Ira Ropson, and Dr. Jiyue Zhu for advice, encouragement and support throughout my graduate studies. I would also like to acknowledge Dr. Gregory Yochum. His knowledge and guidance allowed me to finish a critical experiment towards completion of my thesis research, and for that I am grateful. I am immensely appreciative of the generosity of Dr. Hunt Willard (Duke), Dr. Gary Felsenfeld (NIH), and Dr. Sarah Bronson for providing cell lines and plasmids, without which I would not have been able to undertake my research.

Finally, I would like to thank my family for their unending support, limitless encouragement, and their selfless willingness to help me in any way that they could. Knowing that there was always a sympathetic ear, an appreciation and respect for my efforts, and a wealth of unconditional love helped to get me to this point in my life.

Chapter 1

X CHROMOSOME INACTIVATION

The regulation of gene dosage throughout the genome is essential for normal development in all types of organisms. In humans, whole chromosome duplications or deletions result in severe phenotypes such as Down Syndrome, a result of three copies of chromosome 21, and Turner Syndrome, resulting from the loss of one X chromosome. Even the duplication or deletion of small genomic regions can cause detrimental phenotypes. Potocki-Lupski syndrome and Smith-Magenis syndrome result from the duplication and deletion in chromosome 17p11.2, respectively and cause a variety of distinct phenotypic anomalies including skeletal deformities and mental impairment (reviewed in (Berg et al. 2010)). From these examples it is clear that perturbing the normal dosage of genes can have severe phenotypic consequences. However, even in cases of chromosomal aneuploidy a surprisingly large proportion of genes are dosage compensated by transcriptional regulation such that normal and affected individuals have similar transcript expression levels (Ait Yahya-Graison et al. 2007). Therefore it is not surprising that many organisms can not only tolerate the dosage imbalance of some genes, but have evolved sex determination systems that rely on them.

Mammals utilize an XY system of sex determination in which males are XY and females are XX. Female mammals have evolved a form of X chromosome dosage compensation, known as X chromosome inactivation (XCI), that involves chromosome-wide transcriptional silencing of the majority of genes on one X chromosome. The process of X chromosome inactivation is an extraordinary example of the complex mechanisms involved in dosage compensation and long-

range gene regulation. As a result of XCI the inactive X chromosome in females is easily distinguished from its active counterpart by distinct chromatin and DNA modifications as well as differences in nuclear organization and replication timing. Together, these distinctions, termed epigenetic modifications, serve to effectively silence most genes on the inactive X. However, some genes are biallelically expressed from both the active and inactive X chromosomes (Carrel et al. 2005; Yang et al. 2010), despite XCI silencing. These genes are said to escape XCI and offer a unique opportunity to study the mechanism of X chromosome gene regulation to add to our understanding of how XCI silences the majority of genes along the chromosome. To this end, the study presented in the following chapters explores the mechanism by which one cluster of genes on the human X chromosome escapes silencing. However, in order to fully appreciate the complexities of XCI gene regulation a brief review of the evolutionary history of the X chromosome and the XCI process itself is necessary.

Therian Sex Chromosome Evolution

Therian mammals (marsupials and placentals) utilize an XY system of sex determination. With few exceptions, as a result of a single sex-determining gene, *SRY*, on the Y chromosome that initiates testis development, males are XY (Sinclair et al. 1990; Haqq et al. 1994). In contrast, females are typically XX. Intriguingly, the absence of *SRY* on the X is only one of many features that distinguish the X and Y chromosomes. In humans, the ~160 megabase (Mb) X chromosome comprises 5% of the genome and includes some 1,400 transcripts (RefSeq RNA entries from hg18, build 36.1 available at the UCSC Genome Browser http://genome.ucsc.edu). In contrast, the ~60 Mb Y chromosome is relatively gene poor and includes only 195 annotated RefSeq RNAs. Further, while transcript totals for both the X and Y chromosomes include multicopy genes that are expressed only during spermatogenesis, the majority of genes on the X do not have sex-specific functions and lack Y homologs (Skaletsky et al. 2003; Ross et al. 2005;

Mueller et al. 2008). These differences in the size and gene content of the sex chromosomes extend to other eutherian mammals as well. The X chromosome represents ~3-5% of the haploid genome and is highly conserved in gene content from marsupials to humans (Watson et al. 1990; Spencer et al. 1991; Mikkelsen et al. 2007). The Y chromosome varies greatly in size and gene content such that only a few genes, required for spermatogenesis, are conserved between species (reviewed in (Waters et al. 2007)). The genetic diversity between the sex chromosomes is likely a result of their rich evolutionary history.

Because both monotremes and therians utilize an XY sex determination system, for many years the two systems were thought to share a common ancestor. The platypus, one of only a few extant species of monotremes, utilizes a sex chromosome complex composed of multiple X and Y chromosomes (reviewed in (Waters et al. 2007)). However, comparative mapping of X-linked platypus sequences revealed significant orthology to the chicken Z chromosome, and surprisingly no homology to the human X (Veyrunes et al. 2008). Closer analysis of 128 markers spaced along the human X revealed that much of the long arm of the X is homologous to platypus chromosome 6 (Veyrunes et al. 2008). This evidence suggests the therian XY sex determination system arose independently after the divergence of monotremes ~165 million years ago (mya) from an identical pair of autosomes (Ohno 1967; Veyrunes et al. 2008). Successive chromosomal rearrangements suppressed recombination and allowed the proto-X and -Y chromosomes to evolve independently. Following the suppression of recombination, the Y chromosome rapidly accumulated deleterious mutations independently within the mammalian lineages (Graves et al. 2006; Wilson et al. 2009) resulting in large variations in genetic content between species. In contrast, the X chromosome could recombine in females such that deleterious mutations could be eliminated allowing gene dosage and function to be conserved (Wilson et al. 2009). Therefore, ancestral sequences are largely preserved along the X chromosome (Ohno 1967; Wilson et al. 2009). Outside of two small XY identical pseudoautosomal regions (PARs) that allow the

chromosomes to pair and recombine during male meiosis, only 16 functional single-copy human XY pairs (gametologs) remain (Ross et al. 2005).

Analysis of the human X chromosome suggests that a series of stepwise events led to the divergence of the proto-X and -Y eutherian chromosomes. Sequence comparisons and gene mapping studies of the marsupial X chromosome demonstrate that the 76 Mb chromosome retains considerable homology to the long arm of the human X (Watson et al. 1990; Mikkelsen et al. 2007). This region is highly conserved between therian species and is termed the XCR, or X conserved region. However, X-linkage is not conserved for all genes on the human X and much of the short arm shares homology with autosomal loci in marsupials (Spencer et al. 1991; Mikkelsen et al. 2007). Therefore, the short arm of the human X is likely an evolutionarily recent addition to the sex chromosomes that was added sometime after the marsupial-eutherian divergence ~150 mya. The release of the completed human X chromosome sequence has enabled comparison of human X-linked sequences with more distantly related species to determine their autosomal origins. One such analysis revealed that the XCR shares significant homology with chicken chromosome 4. Further, the short arm of the human X, termed the X added region, or XAR, shares homology with chicken chromosome 1 (Ross et al. 2005). This further supports the conclusion that the XAR is a secondary acquisition, likely the result of a translocation from an autosome. A subsequent comparison of the sequence divergence between 19 human XY gametologs identified four individual strata on the X each representing a distinct evolutionary event that suppressed recombination in one stratum and facilitated XY divergence (Lahn et al. 1999). More recently, comprehensive analysis of the completed XY sequences revised this view, splitting the most recently diverged strata into two, suggesting that at least five major events suppressed XY pairing (Ross et al. 2005). From these studies it is clear that the evolution of the human X chromosome was not a continuous process, but instead was punctuated by multiple

events that acted as a barrier to recombination between large portions of the X and Y chromosomes (Lahn et al. 1999; Ross et al. 2005).

With the development of the XY sex determination system and genetically distinct sex chromosomes, the need for mechanisms to compensate for the X-linked gene dosage imbalance between males and females emerged. The degradation and subsequent loss of genes on the Y left males monosomic for most genes on the X. In contrast, females are disomic for all X-linked genes. To further complicate matters, since the X and Y are derived from an equivalent pair of autosomes, all X-linked genes, lacking a Y-gametolog, that were previously disomic on the autosomes are now monosomic in XY males. To counter this imbalance in gene dosage between XY males and XX females, two mechanisms of dosage compensation evolved in early therians. First, *X chromosome inactivation* silences the majority of genes on one X chromosome in females, leading to functional monosomy for most genes in both sexes. Second, *X chromosome upregulation* increases gene expression on the single active X chromosome in females equalizing gene dosage relative to autosomes (reviewed in (Prothero et al. 2009). Because of these two regulatory mechanisms the X-linked genetic content is largely conserved between eutherian species (Ohno 1967).

X Chromosome Inactivation

X chromosome inactivation (XCI) is the process in which female eutherians silence the majority of genes on one X chromosome to equalize gene dosage between XX females and XY males. XCI, also called Lyonization, was first hypothesized by Mary Lyon in 1961. She proposed that one X chromosome in females, of either parental origin, was silenced early in development (Lyon 1961). This hypothesis was based on three observations in female mice: 1) those having a single X chromosome (a 39,XO karyotype) develop normally (Welshons et al. 1959), 2) cells from multiple tissues contain a heterochromatic X chromosome (Ohno et al. 1960)

and 3) those heterozygous for X-linked mutations affecting coat color exhibit a mosaic phenotype, visible as different colored patches reflecting the color of the parental allele that is expressed from the active X. This hypothesis was quickly extended to include all mammals and explained the viability and phenotypes of individuals with X chromosome aneuploidies (Lyon 1962). Nearly 50 years have passed since the original tenets of XCI were proposed and today they remain true with only the rare exception for all eutherian species. The process of XCI ensures that both males and females are functionally monosomic for most genes on the X chromosome.

Random XCI, as postulated by Lyon, accounts for the X-linked expression patterns in adult eutherian females and is responsible for equalizing gene dosage between XY males and XX females. A second form of XCI, imprinted XCI, selectively silences the paternally derived X in the extraembryonic tissues of eutherians (reviewed in (Heard et al. 2006)) and in somatic tissues of marsupials (Sharman 1971). In eutherians, both imprinted and random XCI share many aspects of the gene silencing mechanism. Both processes result in an inactive X nuclear territory with a gene-rich outer rim and a gene poor inner core (Chaumeil et al. 2006; Clemson et al. 2006; Namekawa et al. 2010) and both require the non-coding RNA *Xist,* the inactive X specific transcript, for gene silencing (both of these features are addressed in more detail below) (Penny et al. 1996; Marahrens et al. 1997; Namekawa et al. 2010). However, initial events at the onset of silencing differ significantly between the two forms of XCI; imprinted XCI is parentally controlled while random XCI requires a counting and choice mechanism that silences either the maternally or paternally derived allele (Namekawa et al. 2010). Therefore, although the two processes are similar in their mechanism of gene silencing, random XCI is the predominant regulatory mechanism affecting gene expression on the X chromosome in adult tissues and is the focus for the remainder of this review.

For the most part, random XCI reflects events that initiate early in development and results in the transcriptional silencing of one of the two X chromosomes in all cells of XX females. Much of our current understanding of the XCI process has come from studies in mouse using knockouts or an embryonic stem (ES) cell model system that upon *in vitro* differentiation recapitulates the XCI process. These experiments have identified a single X-linked non-coding RNA, termed the inactive X specific transcript, or *Xist*, that is essential for random XCI (Penny et al. 1996; Marahrens et al. 1997). *Xist* was first postulated to play a regulatory role in XCI upon the discovery that it was expressed only from the inactive X in both mouse and human (Borsani et al. 1991 ; Brockdorff et al. 1991; Brown 1991). The *Xist* gene has since been mapped to the region of the X chromosome termed the X inactivation center, or *Xic*, which contains multiple elements that influence *Xist* expression (Figure 1-1) (reviewed in (Payer et al. 2008). Of these elements, *Xist* expression is directly controlled by transcription of the antisense non-coding RNA *Tsix* across the *Xist* promoter (Figure 1-1) (Luikenhuis et al. 2001; Shibata et al. 2004; Ohhata et al. 2008). Studies of the remaining regulatory elements within the murine *Xic* reveal that disruption of *Xist* or *Tsix* regulation results in abnormal XCI as summarized in Table 1-1.

Recently, studies of XCI in humans suggest that some aspects of XCI regulation may not be conserved between species. Similar to mouse, *XIST*, the human ortholog of murine *Xist*, is

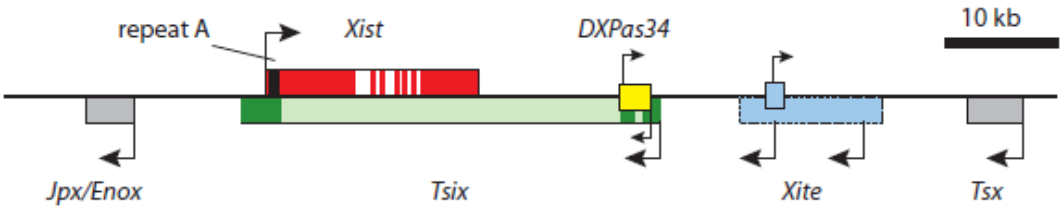


Figure 1-1. The mouse X inactivation center (*Xic*). A schematic of the known elements and regions within the *Xic* thought to affect counting and choice during XCI (see Table 1-1). Arrows indicate the transcriptional direction for each element. Elements of unknown size have a dotted line as a border. This image is adapted from (Payer et al. 2008).

Table 1-1. Elements of the *Xic* that regulate XCI identified using transgene or deletion experiments in mouse.

***Xic* Element**	**Type**	**Function**	**Experiment**	**Reference**
Xist	Non-coding RNA	Silencing	Knockout prevents XCI	(Penny et al. 1996; Marahrens et al. 1997)
			Expression from ectopic transgenes induces long-range inactivation in *cis* on autosomes and the X	(Lee et al. 1997; Wutz et al. 2000; Wutz et al. 2002)
			Deletion of one allele in XX ES cells* induces nonrandom XCI† of normal allele	(Penny et al. 1996)
Tsix	Non-coding RNA	Antisense regulator of *Xist*	Deletion of one allele induces nonrandom XCI† of mutant allele	(Lee et al. 1999)
			Knockout results in aberrant XCI‡ only in XX cells	(Lee 2002; Lee 2005)
			Expression of *Tsix* from both Xs prevents *Xist* accumulation and silencing	(Stavropoulos et al. 2001)
Repeat A	Repetitive domain within *Xist*	Silencing domain of *Xist*	Deletion prevents silencing	(Wutz et al. 2002)
DXPas34	Repetitive domain within *Tsix*	Enhancer of *Tsix*	Deletion of one allele results in skewed XCI§	(Stavropoulos et al. 2005)
Xite	Non-coding RNA	Enhancer of *Tsix*	Deletion of one allele results in skewed XCI§	(Ogawa et al. 2003)

*Murine Embryonic stem (ES) cells recapitulate the *in vivo* XCI process upon differentiation
†Nonrandom XCI - selection of either the normal or mutant X in all cells
‡Aberrant XCI - X chromosome are chaotically inactivated leading to an inactive X in XY or XO cells or two inactive Xs in XX cells
§Skewed XCI – increased probability that a particular X becomes the inactive X

required for random XCI (Chow et al. 2007). However, studies of human *TSIX* demonstrate that antisense transcription across the promoter cannot regulate the expression of human *XIST* since *TSIX* in humans is truncated and does not fully overlap the *XIST* gene (Migeon et al. 2001). Despite this apparent difference in regulation, upon XCI the *Xist/XIST* gene is expressed exclusively from the inactive X in both species (Borsani et al. 1991; Brockdorff et al. 1991; Brown et al. 1991) and produces a long (17 kb in mouse; 19 kb in human) untranslated RNA (Hong et al. 2000) that coats the chromosome in *cis* (Brown et al. 1992; Clemson et al. 1996). Therefore, variation between species may simply represent differences in the regulatory process that initiates *Xist/XIST* expression, while the overall process of XCI and *Xist*-mediated silencing appears to be more highly conserved.

Recently, much effort has focused on the events that lead to the upregulation of *Xist* on the future inactive X chromosome in mouse. Since the overall XCI process is likely conserved, many aspects of the process are expected to be shared by other eutherian species including humans. Therefore, understanding how the process of XCI effectively inactivates the majority of genes on one mouse X chromosome requires a synopsis of the events preceding gene silencing. These events are generally divided into sequential stages: counting and choice, initiation, spreading, and maintenance. The basic underlying mechanisms are presumed to be conserved between species.

Counting And Choice

In the initial counting step of XCI, cells are thought to measure the X:autosome ratio to ensure that the appropriate number of Xs are inactivated. Only diploid cells that have more than one X (or *Xic*) initiate the XCI process, following the n-1 rule, such that all supernumerary X chromosomes are inactivated. After this counting step, cells that initiate XCI choose which of the X chromosomes will be inactivated. In normal cells, all X chromosomes have an independent

probability of being chosen as the inactive X chromosome (Monkhorst et al. 2008; Monkhorst et al. 2009). Yet, this "choice" of which X chromosome is inactivated depends on the regulation of *Xist* in an allele-specific manner. Selection of the inactive X requires *Xist* expression and in mouse is dependent on the relative level of *Tsix* transcription across *Xist* (Nesterova et al. 2003). *Tsix* transcription itself is influenced by at least two enhancers *Xite* and DXPas34 (Figure 1-1) (Ogawa et al. 2003; Stavropoulos et al. 2005). Therefore the expression of *Xist* on the future inactive X is ultimately determined by the regulation of *Tsix*. Nevertheless, the mechanism by which the two X chromosomes communicate in *trans* to achieve mutually exclusive choice of which chromosome to inactivate is only beginning to emerge.

Recent clues as to how a cell accurately assesses the X chromosome content and communicates choice have come from exploring the intra- and inter-chromosome interactions during XCI. At the onset of XCI, fluorescence *in situ* hybridization (FISH) based measurements place the two *Xics* in close proximity suggesting that the two chromosomes transiently pair (Bacher et al. 2006; Xu et al. 2006). This pairing likely relies on *Xite*, *Tsix* and at least one additional X-linked region termed the X pairing region, or *Xpr*, located 200 kb upstream of *Xist* (Augui et al. 2007; Xu et al. 2007). In addition, the *trans*-acting transcription factors Ctcf and Oct4 are essential for this pairing (Xu et al. 2007; Donohoe et al. 2009). Exactly how this transient pairing elicits the exchange of information necessary for the mutually exclusive selection of the inactive X remains a mystery, but likely involves the close proximity of factors involved in the counting and choice mechanisms. The recent discovery of a potent activator of XCI, *Rnf12*, an E3 ubiquitin ligase, that initiates *Xist* expression in a dose-dependent manner (Jonkers et al. 2009) marks a major step towards understanding how cells count the X chromosomes. XCI activators that induce *Xist* expression have been postulated to explain XCI initiation in XX females and the lack of XCI in XY males, but until recently, none had been identified (reviewed in (Navarro et al. 2010)). However, whether additional XCI-activators are

necessary, and how cells measure the X:autosome ratio is still unclear. In addition to inter-chromosomal *Xic* pairing, intra-chromosomal interactions at the *Xic* may also regulate *Xist* expression and the selection of the inactive X. Using chromosome conformation capture (3C), two higher order chromatin domains were identified at the *Xic* immediately prior to the upregulation of *Xist* (Tsai et al. 2008). One domain places *Xist* in contact with *Jpx/Enox*, another non-coding gene that is located within the *Xic* (Figure 1-1) that had not previously been implicated in XCI. A second domain allows *Tsix* to interact with *Xite*. Although the exact nature and function of these two chromatin domains is unclear, both are regulated in a developmentally specific manner suggesting their importance (Tsai et al. 2008). On the inactive X, the *Xite-Tsix* domain dissolves after the onset of XCI while the *Xist-Jpx* domain persists. In contrast, the *Xist-Jpx* domain disappears on the future active X chromosome. This developmental specificity of the two domains suggests that their regulation may influence XCI. Even in the absence of physical interaction, these domains could influence the relative expression levels of *Tsix* and *Xist* resulting in the upregulation of *Xist* on one X chromosome and the initiation of XCI. However, even with the identification of these inter- and intra-chromosome interactions that likely coordinate counting and choice, the mechanisms underlying these processes are still poorly understood. Further, these experiments highlight the importance of appreciating the three-dimensional positioning of inactive X sequences within the nucleus at all stages of the XCI process.

Initiation

The selection of the future inactive X and the initiation of XCI is marked by the sudden monoallelic disappearance of *Tsix* (Lee et al. 1999) and the accumulation of *Xist* in *cis* (Kay et al. 1993; Panning et al. 1997; Sheardown et al. 1997). This contrasts the expression of these genes prior to XCI initiation, where all Xs are active (Monk et al. 1979) and both *Tsix* and *Xis*t are expressed at low levels (Debrand et al. 1999; Lee et al. 1999). Once the future inactive X is

chosen, *Tsix* is rapidly downregulated (Lee et al. 1999) which coincides with the disappearance of the *Tsix-Xite* chromatin domain from the inactive X chromosome (Tsai et al. 2008). This down regulation of *Tsix* enables the transcriptional induction of the *Xist* allele, resulting in the upregulation of *Xist* on the inactive X (Sun et al. 2006). *Trans*-acting factors have also been implicated in regulating the *Tsix* downregulation/*Xist* upregulation switch on the future inactive X, including the transcription factors Ctcf, Yy1, Oct4 and Nanog (Donohoe et al. 2007; Navarro et al. 2008; Donohoe et al. 2009). However, all of these *trans*-acting factors function as XCI inhibitors and repress *Xist* expression. The only known activator of *Xist* is the recently identified XCI activator, *Rnf12* (Jonkers et al. 2009). But, whether *Rnf12* activates *Xist* directly or indirectly by the suppression of *Tsix* is not known. A role for *Jpx/Enox*, the noncoding gene within the *Xic*, as an activator of *Xist* has been postulated from the observed *Jpx/Enox/Xist* interaction within the chromatin domain that persists on the inactive X through XCI initiation (Tsai et al. 2008). Although numerous studies have now identified both *trans*- and *cis*-acting activators and inhibitors of the XCI process, the mechanisms that regulate the *Tsix* downregulation/*Xist* upregulation switch on the inactive X are still being resolved.

Spreading

Once the cascade of counting and choice and initiation steps induce *Xist* upregulation on the inactive X, *Xist* RNA accumulates in *cis* and coats the inactive X in both mouse and human (Brown et al. 1992; Clemson et al. 1996). This results in the *Xist*-mediated silencing of the majority of genes on the inactive X chromosome. However, this *Xist*-mediated silencing is not a unique property of X-linked sequences. Ectopic *Xist* RNA, expressed from multi-copy transgenes, can localize to autosomal sequences and induce silencing (Lee et al. 1996; Herzing et al. 1997; Wutz et al. 2000), albeit to a lesser extent. Therefore, *Xist* localization and silencing does not require sequences that are specific to the X chromosome. Yet, the spreading of *Xi*st

RNA and gene silencing are less efficient on autosomes than on the X chromosome. This suggests that the X chromosome may be enriched in sequences that promote spreading. In fact certain repetitive sequences have been identified that are enriched on the X chromosome relative to autosomes and have been hypothesized to propagate the XCI silencing signal along the X (see Chapter 1, *Genomic Influences On XCI Propagation And Escape Gene Regulation* for details) (Korenberg et al. 1988; Boyle et al. 1990; Lyon 1998). Although the underlying mechanisms regulating the spread of *Xist*-mediated silencing are not fully understood, the sequences within *Xist* that are responsible for localization to the inactive X and gene silencing have been identified. A series of deletions within the *Xist* gene demonstrates that the spreading and localization of *Xist* along the chromosome are mediated by functionally redundant sequences that act cooperatively and are distributed throughout the gene (Wutz et al. 2002). In contrast, deletion of a single conserved repetitive sequence at the 5' end of *Xist*, termed repeat A, eliminates silencing (Wutz et al. 2002).

Until recently, the mechanisms underlying the *Xist*-mediated silencing of X-linked sequences were largely unknown. However, SATB1, a nuclear protein which regulates chromatin structure, has been identified as a silencing factor for *Xist*, implicating a class of chromatin organizing proteins in the process of XCI (Agrelo et al. 2009). The requirement of SATB1 for *Xist* silencing was identified using a lymphodema model in which ectopic *Xist* expression results in XCI that is equivalent to the XCI in ES cells (Agrelo et al. 2009). Further analysis in mouse ES cells confirmed that SATB1 is developmentally regulated and is expressed during XCI initiation (Agrelo et al. 2009). Intriguingly, knockdown of SATB1, by siRNA in ES cells, reduced the efficiency of gene silencing, but did not eliminate *Xis*t-mediated silencing completely, suggesting that other factors may function in a similar manner (Agrelo et al. 2009). While the precise role that SATB1 has in XCI gene silencing is not known, it is likely to be involved in the nuclear organization of the inactive X. That SATB1 could couple *Xist* silencing

to nuclear organization is an intriguing possibility and underscores the importance of nuclear organization in XCI.

Maintenance

The final stage of XCI is characterized by maintenance of the *Xis*t-mediated silencing induced in the early stages of XCI. While the gene silencing early in XCI is reversible, this silencing becomes irreversible during the maintenance phase (Csankovszki et al. 1999; Wutz et al. 2000). Silencing is achieved through numerous epigenetic modifications, including chromatin and DNA alterations, isolation within a defined nuclear territory, and late DNA replication, that all contribute the inactive state of the chromosome (and are discussed in more detail in the following section, *Epigenetic Modifications Of The Inactive X Chromosome*). Many of these epigenetic features that are associated with the inactive X appear to be redundant and the removal of individual features does not induce global reactivation of the chromosome (Csankovszki et al. 2001). Even the conditional knockout of *Xist*, after the initial stages of XCI where it is essential for the initial gene silencing, does not induce the global reactivation of genes on the inactive X during XCI maintenance (Csankovszki et al. 1999; Csankovszki et al. 2001). Yet, *Xist* continues to associate with the inactive X and likely contributes to the inactive state, as the sporadic reactivation of individual genes was observed in cells with *Xist* deletions (Csankovszki et al. 1999; Csankovszki et al. 2001). Additionally, the inactive X is sequestered to a *Xist* RNA domain at the periphery of the nucleus that excludes transcriptional machinery further ensuring that the genes on the inactive X are silenced (Chaumeil et al. 2006; Clemson et al. 2006). Combined, the many epigenetic modifications that associate with the X during the maintenance stage of XCI ensure that the silent state of the inactive X is propagated through all subsequent cell divisions and remains stable throughout the lifetime of the organism.

Epigenetic Modifications Of The Inactive X Chromosome

The term epigenetics is used to describe the stable and heritable alterations in gene expression that occur without changing the primary DNA sequence. This epigenetic control of expression results from the activities of histone modifying enzymes, chromatin-remodeling complexes and DNA methylation enzymes that incorporate DNA methylation and histone modifications which influence expression. Additionally, these epigenetic modifications coordinate the organization and compaction of the genome into discrete chromatin domains (reviewed in (Grewal et al. 2007)). In general, chromatin is divided into two types; euchromatin that is less condensed and relatively accessible for transcription factors and heterochromatin that is highly condensed and largely inaccessible. Differences in the compaction of these two forms of chromatin are due to the incorporation of reversible post-translational modifications of the individual amino acid residues within the histone proteins that control the condensation of chromatin fibers, including acetylation, methylation, phosphorylation and ubiquitination (reviewed in (Peterson et al. 2004; Grewal et al. 2007)). Euchromatin, throughout the genome, is characterized by histone modifications that reduce chromatin condensation, such as acetylation (Shogren-Knaak et al. 2006). Therefore, it is not surprising that many expressed regions of the genome are associated with euchromatin. In contrast, heterochromatin is usually associated with silenced regions and contains hypoacetylated histones as well as other modifications, usually associated with gene silencing, that result in condensation of the chromatin within the domain (reviewed in (Dillon et al. 2002; Grewal et al. 2007)). One of the most prominent examples of heterochromatin in the genome is the inactive X chromosome.

The inactive X chromosome is easily distinguishable from the active X by widespread epigenetic modifications. Such alterations initially led to the identification of the inactive X as the darkly-staining, condensed Barr body at the periphery of interphase nuclei in female cells

(Barr et al. 1949). Today, it is widely accepted that *Xist* RNA forms a nuclear compartment that largely excludes RNA polymerase II and other transcription factors (Chaumeil et al. 2006; Clemson et al. 2006). DNA within this compartment is non-randomly distributed; non-coding and repetitive sequences are internalized, whereas coding sequences remain at the periphery (Dietzel et al. 1999; Chaumeil et al. 2006; Clemson et al. 2006), but are reported to move inward upon silencing (Chaumeil et al. 2006). In addition to this spatial distinction between active and inactive chromosomes, temporal and physical differences also distinguish the two Xs.

The inactive X is distinguished from the active X by asynchronous replication timing. Early work, measuring thymidine-H^3 incorporation established that one X chromosome is late replicating while replication of the second X is indistinguishable from that of autosomes (Taylor 1960; Gilbert et al. 1962; Morishima et al. 1962). More recently, fluorescence *in situ* hybridization (FISH) and allele-specific analysis of individual genes confirmed that X-inactivated alleles replicate later in S-phase than do active alleles (Boggs et al. 1994; Hansen et al. 1996; Xiong et al. 1998). Therefore, in contrast to autosomal alleles that exhibit synchronous replication, the X chromosomes are asynchronously replicated. The late replication of the inactive X likely promotes silencing, but the mechanism linking replication timing and transcriptional inhibition at X-liked loci is unknown. Intriguingly, genes that are reactivated post XCI using the demethylation agent 5-azacytidine replicate early and are nearly synchronous with replication on the active X (Hansen et al. 1997). Therefore, this suggests that late replication is important for proper XCI silencing and that the transcriptionally non-permissive heterochromatic state of the inactive X may also suppress its own replication.

Concurrent with spatial restriction and late replication, the inactive X chromatin undergoes significant remodeling along the length of the chromosome (Figure 1-2). One of the earliest chromatin changes on the inactive X, occurring simultaneously with or immediately after *Xist* upregulation, is the loss of histone modifications associated with active chromatin, including

histone 3 lysine 4 (H3K4) methylation and histone 3 lysine 9 (H3K9) acetylation (Figure 1-2) (Heard et al. 2001; O'Neill et al. 2008). Subsequently, the inactive X loses many of the remaining euchromatic modifications including histone 4 (H4) acetylation (Keohane et al. 1999; Heard et al. 2001). However, whether the initial loss of histone modifications that are typically associated with euchromatin is caused by or is simply a consequence of *Xist* silencing has not been established. As the inactive X loses its euchromatic modifications these are replaced by numerous modifications associated with heterochromatin.

Components of the polycomb repressive complexes, PRC1 and PRC2, are recruited in *cis*, concurrent with or immediately after *Xist* upregulation on the inactive X (Figure 1-2) (Plath et

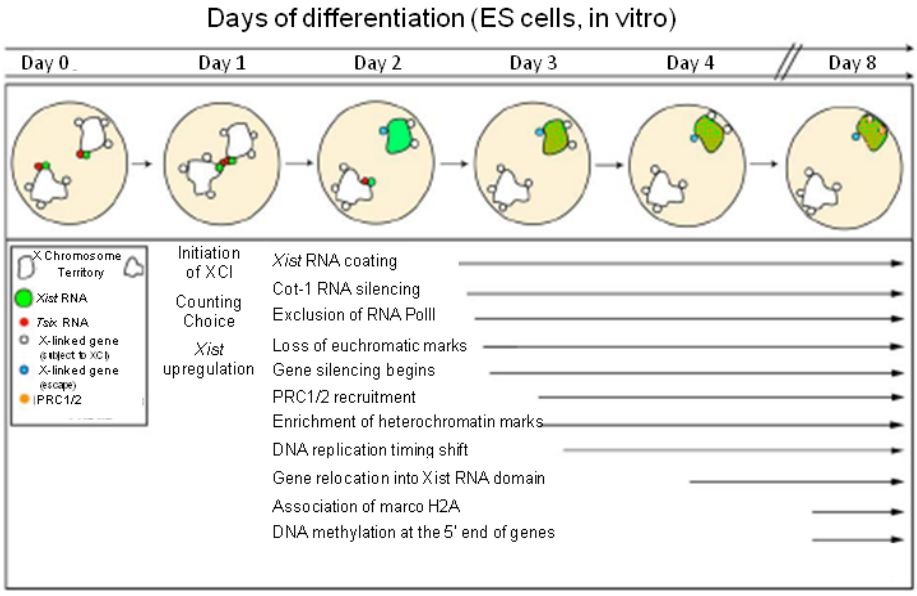


Figure 1-2. Kinetics of the epigenetic changes associated with XCI during mouse ES cell differentiation, a system that closely recapitulates *in vivo* XCI events. Random XCI involves a series of events that results first in *Xist* upregulation (Day 2) and then in gene silencing and stable propagation of this silenced state through epigenetic marks (Day 3+). *Xic trans*-interactions take place just prior to the monoallelic upregulation of *Xist* (Day 1) during XCI and are believed to facilitate counting, choice and initiation. The subsequent kinetics of chromatin changes as well as the relocation of genes into the transcriptionally inactivated *Xist* RNA compartment are shown. Only genes that escape XCI, and the *Xist* gene itself, remain outside or at the periphery of the *Xist* RNA compartment. The silencing of repetitive elements early in XCI is designated as Cot-1 RNA silencing. This image is adapted from (Okamoto et al. 2009).

al. 2003; Silva et al. 2003; de Napoles et al. 2004; Plath et al. 2004). Composed of the Ezh2, Eed, and Suz12 proteins, the PRC2 complex is responsible for the rapid enrichment of trimethylated histone 3 lysine 27 (H3K27me3) that follows *Xist* upregulation (Plath et al. 2003; Silva et al. 2003). This intriguing developmental timing of PRC2 recruitment, early in XCI, supports a role for the complex in XCI initiation, yet mice lacking Eed undergo normal XCI in the absence of H3K27me3 (Kalantry et al. 2006). This suggests that the PRC2 complex is dispensable for random XCI. Further, an *Xist* mutant deleted for the Repeat A sequences, necessary for silencing, was able to recruit PRC2 to the X and methylate H3K27 suggesting that this chromatin modification is not sufficient for silencing (Plath et al. 2003; Kohlmaier et al. 2004). Similar to PRC2**,** PRC1 is recruited and transiently associates with the inactive X (de Napoles et al. 2004; Fang et al. 2004; Plath et al. 2004; Schoeftner et al. 2006). The PRC1 protein, Ring1b, an E3 ubiquitin ligase, mono ubiquitinates histone 2A lysine 119 (H2AK119) at the initiation of XCI (de Napoles et al. 2004). Like the PRC2 mediated H3K27 methylation, PRC1 successfully ubiquitinates H2AK119 in mutants lacking the silencing function of *Xist* (Schoeftner et al. 2006). Further, in the absence of ubiquitinated H2A, *Xist* effectively initiates silencing (Leeb et al. 2007). That both PRC2 and PRC1 are recruited to the inactive X at the same developmental time point, yet are dispensable for silencing, suggests the chromatin changes made by the two complexes may be functionally redundant. Two additional chromatin changes that occur at this early time point are histone 3 lysine 9 (H3K9) methylation and histone 4 lysine 20 (H4K20) monomethylation (Heard et al. 2001; Mermoud et al. 2002; Kohlmaier et al. 2004). The histone methyltransferases (HMTases) responsible for methylating H3K9 and H3K20 on the inactive X have not been identified, although candidate HMTases are known. Three of the eight HMTases, G9a, Suv39h1 and Suv39h2, that have been identified to methylate H3K9 in mouse and human (Volkel et al. 2007; Wu et al. 2010), have been examined relative to XCI. Knockout of G9a in mouse did not disrupt XCI suggesting that this HMTase is not involved in XCI silencing,

although H3K9 methylation on the inactive X was not examined (Ohhata et al. 2004). Further, the double knockout of Suv39h1 and Suv39h2 in mouse abolished H3K9 trimethylation in constitutive heterochromatin, but not that of the inactive X chromosome (Peters et al. 2002). Combined these studies suggest that additional HMTases are responsible for the observed H3K9 trimethylation on the inactive X. Additionally, one candidate HMTase that may methylate H3K20 on the inactive X is PR-Set7 that is responsible for this methylation in other species (Nishioka et al. 2002; Karachentsev et al. 2005).

A second wave of epigenetic remodeling follows *Xist* accumulation and gene silencing and includes incorporation of the histone H2A variant, macro H2A (Costanzi et al. 1998; Mermoud et al. 1999; Rasmussen et al. 2000), DNA methylation at the promoters of X-linked genes (Mohandas et al. 1981; Keohane et al. 1996; Gilbert et al. 1999; Hellman et al. 2007), and, at least in humans, the enrichment of the heterochromatin protein HP1 (Figure 1-2) (Chadwick et al. 2003). The conditional knockout of *Xist* after the establishment of XCI resulted in the absence of macro H2A from the inactive X, but did not affect silencing (Csankovszki et al. 1999). This suggests that similar to many of the early chromatin changes associated with the initiation of XCI, macro H2A is not required for silencing. Even today, the functional role of macro H2A enrichment on the inactive X has not been defined, and no direct association with *Xist* has been identified, even though its localization to the inactive X requires *Xist*. An additional, widely-studied modification that is incorporated on the inactive X late in the XCI process is DNA methylation. The role of the DNA methylation at promoters in gene silencing has been clearly defined throughout the genome and on the inactive X (reviewed in (Klose et al. 2006)). In mammals, methylation is restricted to cytosine residues that are adjacent to a guanine, referred to as CpG dinucleotides. CpGs are largely depleted from the genome with the exception of short genomic regions called CpG islands, usually associated with promoters. Notably, CpG island methylation is known to interfere with transcription factor binding and as such is associated with

transcriptionally silenced regions (Klose et al. 2006). As expected the inactive X is heavily methylated at the promoters of most inactivated genes (Jegalian et al. 1998; Gilbert et al. 1999; Zeschnigk et al. 2009; Yasukochi et al. 2010). Early studies demonstrated that the absence of DNA methylation on the inactive X, induced by treatment with the demethylation agent 5-azacytidine, led cells to reactivate some genes on chromosome (Mohandas et al. 1981; Graves 1982). The hypothesis that DNA methylation was essential for maintaining silence on the inactive X was confirmed in subsequent studies. Mice lacking the DNA methyltransferase (DMNTase), DNMT1, are impaired in their ability to methylate DNA and are unable to stably maintain XCI silencing as measured by the sporadic reactivation of an X-linked transgene (Sado et al. 2000). Studies in human fibroblasts lacking the DMNTase DMNT3b show similar reactivation of X-linked genes after XCI (Hansen et al. 2000). Although the loss of DNA methylation correlates with higher rates of sporadic reactivation, global reactivation of all X-linked genes does not occur. Recently, an additional *trans*-acting factor, SmcHD1, was demonstrated to be essential for promoter methylation the inactive X, and maintenance of XCI (Blewitt et al. 2008). Female SmcHD1 deficient mice die in mid-gestation and lack methylation at multiple genes that were erroneously expressed from the inactive X (Blewitt et al. 2008). Surprisingly, these embryos successfully initiate XCI, recruit PRC2 and incorporate H3K27me3 on the inactive X, but they do not effectively maintain gene silencing (Blewitt et al. 2008). Notably, SmcHD1 is the first regulatory factor that has been identified to be essential for the maintenance of XCI. Although poorly characterized, SmcHD1 contains a C-terminal domain that has been implicated in facilitating chromatin organization. Therefore, it is an intriguing possibility to consider that, in addition to DNA methylation, SmcHD1 could play a functional role in maintaining the higher order structure of the inactive X within the *Xist* nuclear compartment.

Work from many laboratories continues to reveal the complex nature of inactive X heterochromatin. Despite extensive studies, no one individual epigenetic feature has proven critical for the establishment or maintenance of random XCI (Csankovszki et al. 2001; Hernandez-Munoz et al. 2005; Kalantry et al. 2006; Blewitt et al. 2008) with the exception of *Xist*, and even then, only during a narrow window of development (Csankovszki et al. 1999; Wutz et al. 2000). These many epigenetic layers create a redundant network of modifications to ensure that the inactive X does not reactivate if one modification is compromised (Csankovszki et al. 2001). One prediction of this epigenetic redundancy is that these chromatin modifications should colocalize on the inactive X, but at least for some modifications this does not appear to be the case. Immunofluorescent staining of human chromosomes showed that the inactive X is characterized by at least two distinct forms of heterochromatin organized into non-overlapping bands of H3K9 and H3K27 trimethylation (Chadwick et al. 2004). Further, several additional epigenetic modifications on the inactive X have been observed to spatially segregate (Duthie et al. 1999; Valley et al. 2006; Chadwick 2007; Murakami et al. 2009). Recently, efforts have been made to understand how specific chromatin modifications are distributed across the X at the sequence level. Initial studies have focused on sequences at the *Xic* (Rougeulle et al. 2004; Marks et al. 2009) or on individual X-linked loci (Valley et al. 2006; Mietton et al. 2009). Intriguingly, the limited studies that have been done suggest that the heterochromatin on the X is not quite so neatly arranged into blocks of non-overlapping modifications as the cytological evidence suggested. By immunofluorescence, macro H2A appears to be non-randomly distributed and is associated specifically with regions of H3K27 trimethylation banding (Valley et al. 2006; Chadwick 2007), yet chromatin immunoprecipitation (ChIP) suggests a more uniform distribution (Mietton et al. 2009). While this difference may be due to sensitivity differences between the techniques, at a minimum, the proposal that the inactive X is organized into non-overlapping blocks of heterochromatin, characterized by specific features, is an

oversimplification. Allele-specific ChIP experiments suggest that the majority of loci are enriched for specific heterochromatin modifications, macro H2A or H3K9me3, while lacking others, consistent with the idea of non-overlapping types of heterochromatin (Valley et al. 2006). However, some loci are enriched for chromatin modifications thought to occur separately in the two types of heterochromatin (Valley et al. 2006). This subset of loci may represent areas of overlap of the two heterochromatin types found on the inactive X, or the chromatin modifications within these types of heterochromatin may not be as specific to one domain or the other as once thought. These studies highlight the importance of additional high-resolution analysis to fully reveal the location and combination of functional marks that comprise the inactive X histone code. Currently, large-scale genome-wide approaches using array-based techniques or massively parallel sequencing strategies are underway to decipher the complex chromatin code that directs regulatory sequences throughout the genome (Heintzman et al. 2009). However, expanding these studies to include the X chromosome presents a complication as the two transcriptionally distinct X chromosomes in a female cell each carry their own complement of epigenetically-encoded regulatory instructions. Approaches that distinguish the active from inactive allele are optimal and will be instrumental in characterizing the complex epigenetic landscape of the X.

The heterochromatic state of the inactive X is unprecedented in scale. No other chromosome in the genome is covered in such a large region of heterochromatin. Yet, how this heterochromatic state forms on the inactive X is still under investigation. A possible mechanism that could facilitate the formation of heterochromatin on the inactive X comes from one of the properties of heterochromatin. Heterochromatin is known to propagate itself and spread into adjacent sequences effectively silencing nearby genes (reviewed in (Grewal et al. 2007). This process may account for much of the heterochromatin formation on the inactive X chromosome; however in some regions this spreading must be limited to prevent inappropriate silencing. One euchromatic region on the inactive X encompasses the *Xist* genomic locus, allowing its

expression from the inactive X (Marks et al. 2009), and must therefore prevent heterochromatin spreading and erroneous silencing. Additionally, genes that escape XCI (discussed in detail below Chapter 1, *Genes That Escape X Chromosome Inactivation*) will also need to limit the spread of nearby heterochromatin. By examining these regions on the X chromosome that successfully block the spread of heterochromatin, we can gain insight into how heterochromatin spreads and is maintained on the inactive X chromosome.

Genes That Escape X Chromosome Inactivation

If XCI is a chromosome wide silencing event, then all X-specific genes would be inactivated and unable to contribute to physiological functions. Consequently, individuals with a physical loss of one X chromosome would be expected to be phenotypically normal. However, in humans, most females with a karyotype of 45,X die *in utero*, and those that survive exhibit phenotypes characteristic of Turner syndrome, including short stature, ovarian failure, and cardiovascular anomalies (reviewed in (Bondy et al. 2009)). These phenotypic abnormalities of females monosomic for the X, combined with the anomalies observed in individuals with multiple X chromosomes (reviewed in (Tartaglia et al. 2010)), suggested that the inactive X is at least partially functional. These observations led to the proposal that some genes are expressed from the inactive X (Lyon 1962).

The existence of genes that escape XCI silencing such that they are biallelically expressed from both X chromosomes is now well established. Early approaches identified escape genes based on the ability to distinguish the allelic origin of protein expression from the two X chromosomes at specific loci (Shapiro et al. 1979). These methods were refined with the advent of new techniques that allow the transcriptional activity of X-linked genes to be measured directly. Recently, large scale analysis analyzed expression of 624 X-linked genes by directly measured expression from the inactive X using a combination of qualitative assessment in mouse-

human somatic cell hybrid cells that retain a single human X, and quantitative analysis of the relative level expression from each X chromosome by monitoring expression at transcribed polymorphisms in fibroblasts (Carrel et al. 2005). This dual-faceted analysis of X-linked genes established that 15% of genes escape XCI, although many are expressed at reduced levels on the inactive X compared to the active X (Carrel et al. 2005). An additional 10% of human genes demonstrate a variable pattern of XCI expression, escaping XCI in some cell lines yet are X inactivated in others (Carrel et al. 2005). Escape genes that are expressed at higher levels in females compared to males have also been identified indirectly using microarrays (Sudbrak et al. 2001; Craig et al. 2004; Talebizadeh et al. 2006; Johnston et al. 2008). However, microarray analyses report lower estimates of the total number of escape genes on the human X. This is likely due to the low level of inactive X expression for most escape genes resulting in minimal male/female expression differences that are not detected by microarray approaches. Therefore, despite the differences in the total number of genes identified, the two complimentary approaches are largely consistent in their ability to identify highly expressed escape genes.

Human escape genes are located along the length of the X chromosome, but are non-randomly distributed (Figure 1-3) (Carrel et al. 2005). The most recent evolutionary stratum of the X chromosome has the highest density of escape genes, while long arm of the chromosome corresponding to the XCR and hence the oldest stratum has the lowest density (Carrel et al. 2005). This non-random distribution coincides with the location of many of the remaining XY gametologs (Ross et al. 2005), and supports the prediction that the acquisition of X inactivation is dependent on the decay of Y-linked gametologs (Jegalian et al. 1998). The high proportion of escape genes in the recent evolutionary strata that lack Y-gametologs may represent an intermediate stage in this process, in which the genes have not yet acquired the necessary features to undergo silencing since they remain expressed upon XCI. Closer inspection of the location of escape genes along the chromosome reveals that many human escape genes are found in clusters

that contain at least one gene with Y homology (Carrel et al. 2005). The clustering of multiple genes that escape XCI on the human X is intriguing and suggests that X-linked genes are organized into domains that are controlled by regional mechanisms (Miller et al. 1995; Carrel et al. 2005).

In contrast to human, only a handful of mouse genes escape XCI. A recent study using high-throughput RNA sequencing identified nearly 400 X-linked genes with expressed

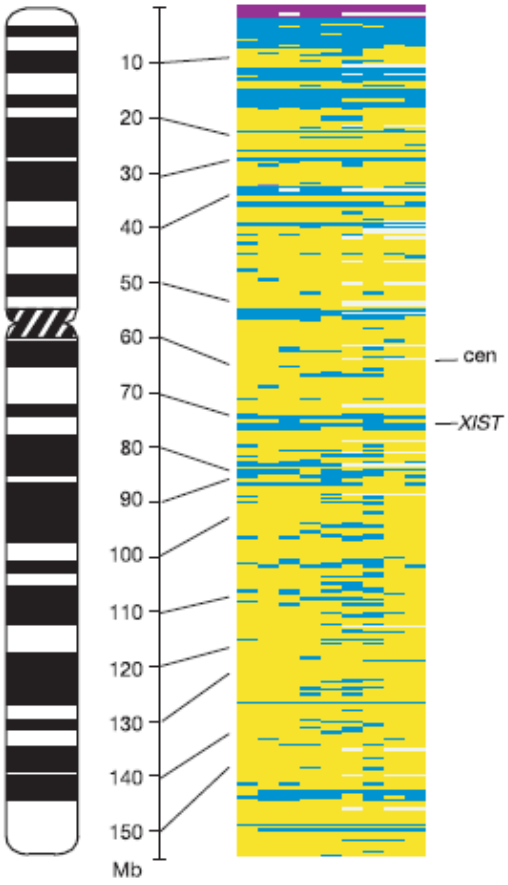


Figure 1-3. X inactivation profile of the human X chromosome. The XCI status of 624 genes on the human X chromosome was determined in nine mouse-human somatic cell hybrids containing an inactive human X. The expression of each gene from each of the hybrids is displayed linearly and its chromosomal location is shown relative to the X chromosome schematic. Expression from the inactive X is shown: expressed genes are blue, X inactivated genes are yellow, pseudoautosomal genes are purple, and untested hybrids are white. The positions of the centromere (CEN) and *XIST* are shown. This image was adapted from (Carrel et al. 2005).

polymorphisms such that allelic origin of the transcript could be identified (Yang et al. 2010). Of these, thirteen (~3%) demonstrated significant expression from the inactive X (Yang et al. 2010). These thirteen escape genes are randomly distributed along the chromosome (Yang et al. 2010). Further, most of the mouse escape genes have functional Y-gametologs and they are distributed as singletons along the chromosome, unlike the clustered human escape genes (Yang et al. 2010). This suggests that clustering is not a common feature of escape shared between human and mouse. However, despite differences in the number and organization of escape genes, that eight of the thirteen mouse genes also escape XCI in humans suggests that the escape of at least some genes is conserved between species.

Consistent with the hypothesis that the escape of some genes is conserved, several XY gametologs have been identified that escape XCI in multiple eutherian species (Basrur et al. 2004; Carrel et al. 2005; Yen et al. 2007; Yang et al. 2010). This suggests that escape is a shared mechanism between species that equalizes gene dosage in XY gametologs where two functional copies are present in both males and females. Furthermore, the rodent-human somatic cell hybrid model system faithfully maintains the XCI status of all human X-linked genes despite a rodent genetic background (Carrel et al. 1996). Combined, these studies suggest that the mechanisms underlying escape gene regulation are likely conserved between species.

Since escape from XCI has been postulated to equalize gene dosage of XY gametologs in XY males and XX females, it is surprising to find that escape does not necessarily equate to equal expression. The vast majority of human escape genes lack a Y homolog and the additional expression from the inactive X for some genes results in higher overall expression in females compared to males (Talebizadeh et al. 2006; Johnston et al. 2008). Even in the case of XY gametologs, expression levels between males and females are not always equivalent. In mouse, at least five XY gametologs display sexually dimorphic expression in brain; females have higher expression that is not compensated by the Y-linked allele in males (Xu et al. 2002). Additional

studies using an elegant mouse system that can distinguish between phenotypic sex and sex chromosome content contributions have since confirmed that the higher expression in female brain, for at least three XY gametologs, *Jarid1c*, *E1f2s3x* and *Utx*, is directly related to expression from the inactive X chromosome (Xu et al. 2006; Xu et al. 2008; Xu et al. 2008). Further, studies of XY gene pairs revealed that many X and Y gametologs vary widely in their expression levels and tissue distribution between the sexes (Yan et al. 2005; Wilson et al. 2009), confirming that sexually dimorphic expression is not limited to brain. Yet the fact remains, despite differential expression from both X chromosomes, the majority of X-linked escape genes are expected to be largely dosage compensated. In humans, only 5% of X-linked genes have increased expression in females (Johnston et al. 2008), compared to the 15% that escape XCI and are biallelically expressed (Carrel et al. 2005). Low expression from many X-linked escape genes may contribute, but for at least some genes post-transcriptional and translational regulation mechanisms ensure dosage equivalence. While expression levels of mouse *Eif2s3x* are sexually dimorphic; females have higher expression, protein levels are not (Xu et al. 2006). Furthermore, the XY gametolog pair, *Ddx3x* and *Ddx3y*, is widely expressed in many tissues, yet translation of *Ddx3y* is restricted to the male germline (Ditton et al. 2004). Whether these novel post-transcriptional and translational regulation mechanisms represent gene-specific or chromosome-wide strategies to equalize X-linked gene dosage remains to be seen. Nevertheless, post-transcriptional and translational strategies are unlikely to specifically target transcripts produced from the inactive X given that females which are heterozygous for a mutant steroid sulfatase, *STS*, allele, a gene known to escape XCI, successfully produce STS protein in all cells irrespective of whether the normal allele is on the active or inactive X (Shapiro et al. 1979). Today, it is clear that escape genes utilize many strategies to ensure at least nearly equivalent dosage between the sexes. Although the variation of X-linked gene expression between tissues

and individuals suggests absolute dosage compensation is not a strict requirement for all X-linked genes.

The mechanisms that allow some genes to escape XCI silencing such that they are expressed from the inactive X, in spite of the largely heterochromatic environment of the chromosome, is not well understood. Epigenetic features in escape domains are likely to play a large role in this process and are discussed in more detail in the next section (Chapter 1, *Epigenetic Features Of Escape Genes*). But, how euchromatic epigenetic modifications are targeted to escape domains or conversely, how these domains are protected from heterochromatic modification is not clear. Therefore, the examination of escape domains and the mechanism by which they resist silencing will add to our understanding of how XCI silencing spreads along the length of the inactive X.

Epigenetic Features Of Escape Genes

As described above (Chapter 1, *Epigenetic Modifications Of The Inactive X Chromosome*), much of the inactive X is characterized by spatial, temporal, and physical epigenetic changes that distinguish it from the active X. Not surprisingly, genes that escape XCI are similarly distinguished from X inactivated genes. Escape genes are characterized by epigenetic features that typify active transcription and euchromatin throughout the genome including the active X chromosome. Much of the inactive X chromosome is sequestered within the *Xist* nuclear compartment in interphase cells, but genes that escape XCI genes remain outside (Chaumeil et al. 2006; Clemson et al. 2006). Consistent with this idea, a recent study using a combination of RNA *in situ* hybridization and ChIP demonstrated that two escape genes in mouse do not associate with *Xist* RNA, whereas *Xist* is detected on at least one X inactivated gene (Murakami et al. 2009). This evidence suggests that nuclear localization outside of the *Xist* domain is necessary for escape. However, recent studies of autosomal genes that demonstrate

localization exterior to chromosome territories is not sufficient to upregulate gene expression (Morey et al. 2009). Therefore, additional investigation is needed to determine the role that nuclear localization plays in escape gene regulation. Genes that escape XCI have been shown to replicate synchronously with their active X counterparts, significantly earlier in S phase than X inactivated genes (Boggs et al. 1994). While these spatial and temporal distinctions are based on observations of a relatively small number of genes with opposing XCI status, they likely contribute to expression and highlight the differences between escape and silent loci on the inactive X chromosome.

In contrast to the temporal and spatial distinctions that have been observed for relatively few genes, chromosome-wide analyses have identified a host of physical attributes that distinguish inactivated genes from those that escape XCI. Recently, several chromosome wide studies have investigated the relationship between XCI and CpG methylation at the promoters of X-linked genes. As expected, genes that escape inactivation lack promoter methylation (Weber et al. 2007; Zeschnigk et al. 2009; Yasukochi et al. 2010). This suggests that promoter methylation contributes to escape. However, some X inactivated genes also lack promoter methylation (Zeschnigk et al. 2009; Yasukochi et al. 2010) suggesting that additional signals are necessary to ensure proper silencing of these X inactivated genes. This fine-tuning of expression regulation likely comes from specific histone modifications of escape genes and regions. Unlike the majority of the inactive X, escape genes retain many histone modifications usually associated with active chromatin including histone H3/H4 acetylation (Boggs et al. 1996; Gilbert et al. 1999; Goto et al. 2002; Brinkman et al. 2006). Escape genes can also be distinguished from active genes by the enrichment of H3K4 dimethylation both at the promoter and throughout the gene body similar to biallelically expressed autosomal genes (Rougeulle et al. 2003). In contrast to escape genes, X inactivated genes are enriched for H3K4 dimethylation only at the promoter (Rougeulle et al. 2003). Not surprisingly, H3/H4 acetylation and H3K4 methylation, generally

thought to be active marks do not exclusively associate with transcription on the inactive X, similar to the autosomal loci that were simultaneously queried (Brinkman et al. 2006; Valley et al. 2006). Further, the presumptive inactive marks, H3K27me3 or H3K9 methylation do not specifically associate with silenced regions (Brinkman et al. 2006; Valley et al. 2006; Marks et al. 2009). While these studies highlight the importance of epigenetic modifications on gene regulation, they also emphasize the complexities of regulating expression on the inactive X chromosome. However, understanding the mechanisms that regulate escape domains, specifically the mechanisms that prevent the encroachment of surrounding heterochromatin into these domains is crucial to understanding how XCI silencing spreads along the X chromosome.

Genomic Influences On XCI Propagation And Escape Gene Regulation

Although a picture of the epigenetic landscape on the inactive X has emerged, how XCI spreads in *cis* and maintains silence along the 160 Mb X chromosome is still poorly understood. Understanding escape gene regulation, specifically how they avoid the *Xist*-mediated silencing that successfully silences the majority of genes on the X, may provide clues towards understanding the process as a whole. Moreover, models proposed to explain XCI silencing must incorporate escape gene regulation and account for species differences in escape gene organization.

Repetitive Element--Mediated Silencing Models Of X Chromosome Inactivation

Initial clues to the XCI process came from X;autosome translocations and *Xist/Xic* transgene experiments in which autosomal sequences were at least partially inactivated (Rastan 1983; Lee et al. 1997; White et al. 1998). These early studies establish that sequences unique to the X chromosome are clearly not required for XCI silencing, but could be enriched or organized on the X in a specialized way to promote stable, chromosome-wide silencing upon XCI. Such

observations led Stanley Gartler and Arthur Riggs to postulate that specific X sequences, or "way stations", propagate XCI (Figure 1-4 A, model i) (Gartler et al. 1983). On the X chromosome these sequences would allow the inactivation signal to spread along the length of the chromosome by serving as docking sites for *Xist* RNA or heterochromatin proteins. Autosomal sequences that lack the necessary enrichment or proper organization of these "way stations" would be unable to effectively propagate the XCI signal resulting in inefficient silencing. In fact, the repetitive element LINE-1 or L1, a long interspersed nuclear element, has been proposed for such a function (Lyon 1998). Chromosome-wide analysis of the distribution of L1s and motifs within L1s supports this hypothesis as these sequences are enriched relative to autosomes. Additionally the model predicts that escape genes lack these elements such that they are ineffectively silenced, and in fact, L1s are depleted within escape regions of the X (Bailey et al. 2000; Ross et al. 2005; Carrel et al. 2006; Wang et al. 2006). Nevertheless, the enrichment and distribution of L1 elements on the X could, in part, be due to the unique biology of sex chromosome evolution. The apparent enrichment of L1s on the X chromosome may simply reflect the selective loss of deleterious L1 elements from the autosomes, causing the X to appear such that it has accumulated L1s (Boissinot et al. 2001). Although the deletion of L1s from autosomes may explain their low L1 density as compared to the X chromosome, but this does not eliminate the possibility that the enrichment of L1s on the X has a functional role in XCI.

In addition to L1s, a number of other repeats and sequence features have been identified whose distribution correlates with escape or X inactivated genes. Among the sequences identified, escape genes are enriched for Alu repetitive elements (Wang et al. 2006) and $(GATA)_n$ simple repeats (McNeil et al. 2006), but are depleted for LTRs (long terminal repeat elements) (Tsuchiya et al. 2004) and MIRs (mammalian interspersed repeat elements) (Ke et al. 2003; Wang et al. 2006). Similar to the "way station" model, that requires specific sequences to propagate XCI silencing, sequences enriched in escape domains could be necessary for escape

gene regulation. In fact, two independent studies successfully used a subset of the escape gene genomic features, including motifs within L1s, to correctly predict XCI status for a large proportion of genes on the X (Carrel et al. 2006; Wang et al. 2006). The high XCI prediction rate in each of these studies, despite using different classifier sequences, strongly supports a role for local genomic sequence environment in regulating escape. One study supports the L1 repeat hypothesis as specific motifs within L1s showed high XCI prediction capability even in regions of relatively low L1 density (Carrel et al. 2006).

If the hypothesis that L1s function as "way stations" to seed heterochromatin and propagate XCI silencing is true, one prediction is that epigenetic modifications associated with XCI may localize to L1s just prior to accumulation elsewhere on the X. This hypothesis was recently tested by utilizing chromatin immunoprecipitation sequencing techniques, or ChIP-seq, to measure the distribution of histone modifications along the X during XCI initiation in mouse (Marks et al. 2009). Trimethylated H3K27 (H3K27me3), one of the earliest histone modifications associated with XCI silencing, did not spread linearly away from the *Xic*, nor was it found to be enriched at L1s or any other genomic feature previously implicated in XCI regulation when compared to the rest of the X chromosome at any point during the differentiation of ES cells (Marks et al. 2009). These data potentially eliminate L1 elements and the "way station" model as the system responsible for XCI silencing. However, the possibility remains that either L1s and nearby sequences are not epigenetically distinguished for enough time to be adequately discriminated from more distant sequences using this approach or that L1s are not marked by H3K27me3 in a developmentally specific manner. Additionally, since computational approaches suggest that specific L1 motifs may serve to promote the XCI signal (Carrel et al. 2006), this subset of L1s may not be distinguishable from other L1 sequences using the ChIP-seq approach. However, even if L1s do not act as "way stations", they may still play a functional role in the XCI process.

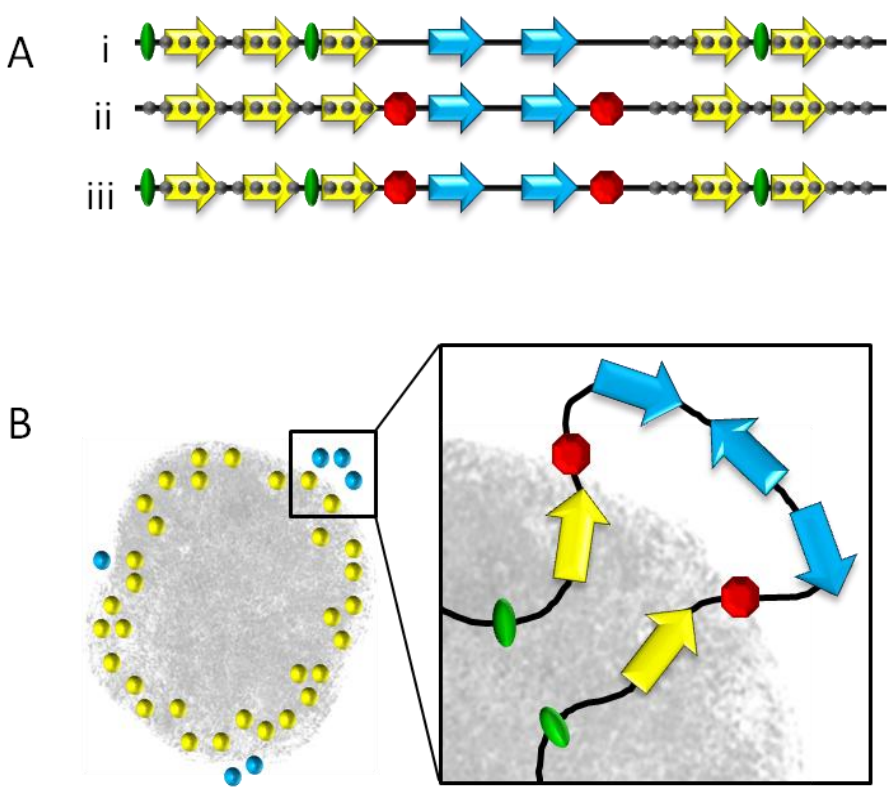


Figure 1-4. Models of X chromosome inactivation escape gene regulation. (A) Genomic sequences or boundary elements may regulate inactive X expression and these differ in their proximity to escape gene domains. **i**) Inactive X heterochromatin (gray circles) is propagated from "way stations" (green ovals) and encompasses X inactivated genes (yellow). In this model, escape genes (blue) reside too far from "way stations" to be inactivated. Although not shown, a variation of this model is that specific escape sequences could lie too far from inactivated genes to impart an effect (Carrel et al. 2006; McNeil et al. 2006). Alternatively, the placement of L1s or other repetitive sequence elements (green) may also induce the formation of heterochromatin leading to gene silencing (Chow et al. 2010). Nevertheless, regulatory elements that rely on distance to distinguish genes cannot account for the close juxtaposition of some escape genes and inactivated genes (Tsuchiya et al. 2004; Carrel et al. 2005), but could explain large Mb sized domains with large transition regions elsewhere on the X. **ii**) Sequences such as chromatin insulators, DNA boundary elements or barriers (red octagon) flank coordinately regulated genes and protect them from silencing. This model, or at least with respect to the CTCF protein (Filippova et al. 2005), also cannot fully explain escape gene expression (Ciavatta et al. 2006). **iii**) Incorporation of both models; "way stations" propagate XCI which is prevented from reaching escape genes by boundary elements. (B) Three dimensional organization of X-linked genes can further affect escape gene expression (Chaumeil et al. 2006). A cross-section of the inactive X is enlarged. Both active (blue) and inactive (yellow) genes lie at the periphery of the *XIST*-delineated inactive X territory (gray). Non-genic and repeat sequences (black line) reside largely within the *XIST* compartment. Exterior positioning of escape genes is facilitated by boundary elements (red) and/or excessive distance from way stations (green) or L1 repeats. This figure was adapted from (Prothero et al. 2009).

The correlation between L1 density and the efficiency of silencing was originally used to satisfy the "way station" model, but L1 distribution along the X also supports a repeat-induced silencing model in which L1s function to nucleate heterochromatin, independent of additional factors, along the inactive X chromosome (Figure 1-4 A, model i) (Chow et al. 2010). The finding that L1 sequences are internalized within the *Xist* compartment (Chaumeil et al. 2006; Clemson et al. 2006), combined with the discovery that some L1s are expressed from the inactive X in an *Xist* dependent manner (Chow et al. 2010), support this alternative role for L1s in XCI silencing. The inactive X is sequestered within a unique *Xist* nuclear compartment and the three-dimensional positioning of genes is proposed to influence XCI status, as at least one escape gene remains outside this compartment (Figure 1-4 B) (Chaumeil et al. 2006). A direct role in facilitating the compaction of inactive X sequences is suggested by the recent kinetic studies correlating L1 density with gene silencing (Chow et al. 2010). One of the first steps in heterochromatin formation, prior to gene silencing, is the appearance of LINE and SINE elements within the *Xist* compartment, whereas regions of low LINE density remain outside (Chow et al. 2010). Intriguingly the movement of expressed L1s into the *Xist* nuclear compartment also correlates with movement of a gene to the interior of the compartment upon its silencing, late in XCI (Chow et al. 2010). Combined these findings support an alternative function for L1s in XCI silencing, in which L1 enrichment serves a dual purpose. Silent L1s may propagate XCI silencing by serving as docking sites for *Xist* RNA or heterochromatin proteins leading to the formation of the *Xist* nuclear compartment, while active L1s facilitate localized gene silencing in regions that remain outside the *Xist* compartment (Chow et al. 2010). In this model, escape genes would remain at the periphery either because they lie too far from expressed L1s to be effectively silenced, or because they are in some way immune to the L1-mediated silencing. However, in light of recent evidence demonstrating that localization alone is insufficient to regulate gene expression (Morey et al. 2009), further investigation into the role of L1s and heterochromatin formation during the

XCI process is necessary. Further, as described above (Chapter 1, *Epigenetic Modifications Of The Inactive X Chromosome*), non-overlapping domains of heterochromatin, characterized by their histone modifications, indicate that the inactive X is not a uniform mass of heterochromatin. Therefore, it is likely that multiple elements are necessary to promote silencing throughout the different regions of the inactive X. A reasonable hypothesis if L1 sequences function in XCI, is that L1s would be expected to interact with the silencing signal or heterochromatin proteins; however, to date, no physical interaction between *Xist* RNA or heterochromatin proteins has been identified. Moreover, no evidence has definitively established a role for L1s in XCI silencing, underscoring the importance of further characterizing the role of L1s in the XCI process.

A limitation to the L1 silencing models, either the "way station" model that propagates the XCI signal or a repeat-induced silencing model that nucleates heterochromatin formation, is that they cannot account for the regulation of closely juxtaposed X inactivated and expressed genes (Tsuchiya et al. 2004; Carrel et al. 2005). Both the "way station" and repeat-induced silencing models predict that escape genes lie too far from such sequences to be effectively silenced (Figure 1-4, model i), yet genes of opposite XCI status in close proximity are found on the human X chromosome (Carrel et al. 2005). Analysis of L1 density throughout one escape domain on the human X demonstrates that some escape genes are located in regions with high L1 density (Tsuchiya et al. 2004). Such regions cannot be accounted for in either the "way station" or repeat-induced silencing model. Additionally, the repeat-induced silencing model can account for some closely juxtaposed genes by active L1 facilitated silencing, since not all closely juxtaposed genes, with differing XCI expression patterns, are located near an intact L1 element. Moreover, L1-mediated silencing models fail to explain how escape domains are protected from the encroachment of nearby heterochromatin. Therefore, despite the mounting evidence that supports a functional role for L1s in XCI silencing, additional elements are necessary to explain all of the facets of escape gene regulation.

Boundary Element Model Of X Chromosome Inactivation

Appealing candidates to regulate expression at closely juxtaposed loci are sequences such as are DNA boundary elements that could block the spread of heterochromatin into escape domains (Figure 1-4 A, model ii) (Gaszner et al. 2006). DNA boundary elements or chromatin insulators represent a diverse group of regulatory elements that share little more than the common ability to successfully delimit distinct chromosomal regions of gene expression. At least some insulators in *S. cerevisiae* are DNA sequences that are capable of recruiting protein complexes that prevent nucleosome assembly thereby disrupting the contiguous array of nucleosomes required for the spread of heterochromatin (reviewed in (Bi et al. 2001)). An alternative mechanism of insulator action has been identified in *S. cerevisiae*, *D. melanogaster* and vertebrates, where DNA sequences recruit chromatin modifying enzymes that actively prevent the spread of heterochromatin, known as barrier function (Figure 1-5 A) (Oki et al. 2005; Zofall et al. 2006; Huang et al. 2007). In addition to preventing the spread of heterochromatin and protecting domains from position effects, many eukaryotic sequences have been identified that block an enhancer from interacting with a promoter when located within the intervening sequence (Figure 1-5 B) (Bell et al. 1999; Majumder et al. 2003; Filippova et al. 2005). One of the best examples of a vertebrate insulator sequence that is capable both of blocking an enhancer and protecting a domain from position effects is the well characterized chicken β-globin insulator. This insulator lies between a region of heterochromatin and the developmentally regulated β-globin gene cluster (Bell et al. 1999). Recent characterization has revealed that these activities are separate functions and are due to the binding of specific transcription factors to the insulator (Recillas-Targa et al. 2002; Dickson et al. 2010). The enhancer blocking ability is due to binding of the multifunctional transcription factor CTCF, or CCCTC-binding factor. Although CTCF binding does not affect the barrier function it likely contributes to the nuclear organization of the region (Recillas-Targa

et al. 2002; Yusufzai et al. 2004). In addition, the transcription factor, USF1, binds the insulator and recruits specific-chromatin modifying enzymes which aid barrier function by actively maintaining the open conformation of the euchromatic domain (West et al. 2004; Huang et al. 2007). The binding of a third transcription factor, vEZF1, restricts the spread of DNA methylation and protects the β−globin locus from promoter methylation (Dickson et al. 2010). Combined, these insulator proteins collectively result in the robust barrier function of the chicken β-globin insulator sequence. Intriguingly, many insulator sequences have been identified to function through protein-protein interactions to form chromatin loops, whereby distant DNA sequences are brought in close proximity, or by binding additional proteins that cluster in the nucleus to form insulator bodies (reviewed in (Bushey et al. 2008; Raab et al. 2010)). Additional

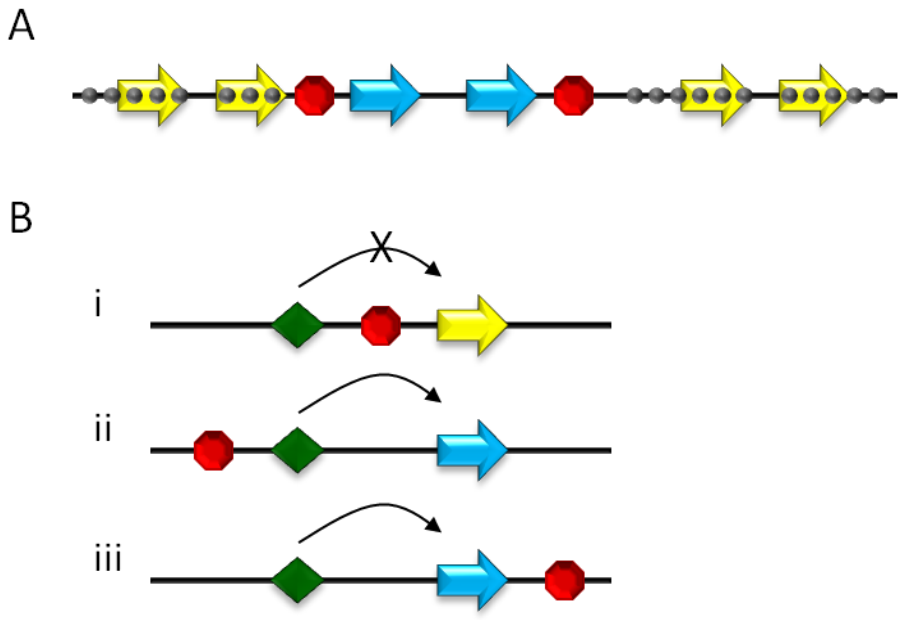


Figure 1-5. Mechanisms of chromatin insulators. (A) Barrier function. Insulators (red octagons) flank euchromatin domains. By recruiting factors that disrupt nuclosome positioning, or actively remodel chromatin, insulators prevent the spread of heterochromatin (gray circles) and the silencing of the genes within the euchromatic region (blue). (B) Enhancer-blocking function. **i)** Insulators positioned between an enhancer (green) and promoter block productive interaction and prevent gene expression (yellow). **ii-iii)** Insulators positioned outside of the enhancer/promoter region do not affect enhancer/promoter communication.

sequences that are capable of organizing chromatin into similar structures have also been implicated in barrier function, such as scaffold/matrix attachment proteins (S/MARs) (Gombert et al. 2003; Goetze et al. 2005). Many chromatin insulators were originally identified by their ability to function in transgene assays, and most have been confirmed at the endogenous loci (reviewed in (Raab et al. 2010)). Intriguingly, an insulator sequence with enhancer-blocking ability was identified within a boundary between X inactivated and expressed transcripts (Filippova et al. 2005). This finding supports a role for insulators in the regulation of escape domain boundaries on the inactive X.

The ability of insulators to function on the inactive X to protect escape domains from silencing by protecting them from the spread of heterochromatin is an interesting prospect. In both mouse and human, several expression boundaries between X inactivated and expressed genes bind the insulator protein CTCF (Filippova et al. 2005). As mentioned previously, CTCF is a multifunctional protein that has widespread roles in transcription regulation throughout the genome, including a role as a chromatin insulator (Phillips et al. 2009). A role for inactive X regulation by CTCF at escape domains, as opposed to gene-specific regulation, is supported by the comparative analysis of CTCF binding at the *Jarid1c*/*JARID1C* locus in mouse and human (Filippova et al. 2005). The *Jarid1c* locus escapes X inactivation in both species, but whereas human *JARID1C* is embedded within a multigene escape domain, mouse *Jarid1c* is surrounded by X inactivated genes (Tsuchiya et al. 2004). Intriguingly, only mouse *Jarid1c*, located adjacent to X inactivated sequences, binds CTCF (Filippova et al. 2005). *S*equence comparison of *Jarid1c/JARID1C* reveals a highly degree of similarity between the species despite differences in XCI status (Tsuchiya et al. 2004). Further, gene function is conserved between species suggesting that CTCF binding is not related to a gene-specific function. Analysis also revealed that CTCF sites within mouse *Jarid1c* are unmethylated throughout development (Filippova et al. 2005). Since methylation of CpGs within the CTCF binding site has been shown to prevent

binding (Pant et al. 2004), the unmethylated sites within mouse *Jarid1c* can serve as a platform for CTCF to bind at the onset of XCI. Analysis of another orthologous escape gene pair, *Eif2S3x* and *EIF2S3* revealed CTCF binding in both species at a site adjacent to an X inactivated gene (Filippova et al. 2005). Combined, the comparative analysis at these two escape domains demonstrates that CTCF binds escape domains at the boundary between X inactivated and expressed genes. Further, these data support the idea that a conserved mechanism regulates the expression of escape genes on the inactive X chromosome.

The evidence presented above supports a role for CTCF in XCI escape gene regulation, but whether CTCF is required to establish escape domains is unknown. An XCI gene regulation model that places insulator sequences flanking escape domains is enticing as it can easily explain opposite XCI states of closely juxtaposed genes. Repositioning CTCF binding sites between species could explain differences in escape domain size between species. However, at a minimum, XCI gene regulation cannot be quite that simple. CTCF binding sites alone are not sufficient to establish an escape domain (Ciavatta et al. 2006). A transgene carrying a GFP reporter flanked by insulators that contained CTCF binding sites was silenced upon XCI, although the insulators successfully blocked position effects from silencing the transgene on the active X chromosome (Ciavatta et al. 2006). Given that CTCF has many functions, binds many locations, and that the distribution of binding sites on the X is not consistent with a role solely in escape gene regulation (Kim et al. 2007; Xie et al. 2007), it is not completely surprising that CTCF alone could not establish an escape domain. In fact, a recent genome-wide study of CTCF binding sites identified over 300 sites on the X chromosome (Kim et al. 2007). Additional factors that could serve to direct CTCF to escape domain boundaries and specify a role in XCI regulation have yet to be identified. Moreover, other sequence elements may bind XCI regulatory factors and insulate escape domains in a manner similar to that predicted for CTCF. One candidate sequence is matrix attachment regions, or MARs, that have been identified at other genomic loci

to insulate sequences from position effects (Gombert et al. 2003). But, similar to CTCF, an X-linked transgene flanked by MARs was inactivated upon XCI (Chong et al. 2002). Additional analysis of sequences within XCI expression boundaries is necessary to identify additional regulatory elements that may play a functional role in XCI, and to define the role of CTCF in escape gene regulation.

In humans, many escape genes are found in large clusters, whereas others abut X inactivated genes. In contrast, the few escape genes in mouse are singletons surrounded by X inactivated regions. Therefore, the regulatory pressures on individual escape genes may vary, and result in different forms of regulation in a gene-specific and potentially species-specific manner. As a result, escape domains and escape genes may incorporate aspects of multiple models of escape regulation (Figure 1-4 A, model iii). However, as a human inactive X chromosome retains escape clusters in mouse-human somatic cell hybrids that have a mouse genetic background, the mechanisms regulating escape gene maintenance and expression are likely shared between species. In addition, it is possible that escape genes incorporate gene-specific regulatory mechanisms. However, no evidence of regulatory elements that promote the expression of individual genes irrespective of the transcriptional activity in adjacent regions has been found (Luoh et al. 1995). Further, despite the organizational differences in escape genes between mouse and human, the L1 depletion within many escape domains and the placement of CTCF-dependent insulators specifically at XCI expression boundaries in mouse and human (Tsuchiya et al. 2004; Filippova et al. 2005) suggest that escape genes are coordinately controlled at the level of chromosomal domains. To complicate matters further, the role of gene positioning, within and outside of the *Xist* nuclear compartment adds an additional layer of complexity to escape gene regulation (Figure 1-4 B). In order to determine the specific elements responsible for escape gene regulation additional studies are needed to examine the role of specific sequences within the *Jarid1c* and *Eif2s3x* domains.

A newly established murine ES cell transgene system that allows manipulation of X-linked sequences near one escape gene will be a powerful tool in understanding the sequence requirements necessary for escape in mouse. This system recapitulates XCI upon differentiation of ES cells, and was recently utilized to address the role of localized genomic sequence at one XCI expression boundary (Li et al. 2008). In this system a transgene, randomly integrated onto the X chromosome, successfully recapitulates endogenous expression patterns of genes with opposite XCI expression status (Li et al. 2008). Consequently, despite multiple integration sites, local sequences contained within the X-linked transgene successful direct the escape of at least one gene (Li et al. 2008). Whether this escape is due to the previously identified CTCF-dependent insulator (Filippova et al. 2005) or the local genomic environment contained within the transgene has not been resolved. Yet this model clearly shows that expression within one escape domain is directly dependent on genomic sequences contained within the transgene. Further dissection of this escape domain, and the contributions of individual sequences elements needs to be addressed. However, as escape domains may be regulated by different regulatory mechanisms depending on the local genomic environment, expanding our studies to incorporate additional escape domains will be vital to our understanding of the complexities of escape gene regulation across the X chromosome. Neither CTCF nor L1 can account for the escape of all genes on the inactive X. L1 density/expression cannot discriminate some closely juxtaposed genes of opposite XCI status, and CTCF binding alone does not induce the formation of an escape domain (Ciavatta et al. 2006). What additional factors may be involved in XCI regulation remains to be seen, but the analysis of additional domains will add to our understanding of XCI escape. Additionally, no study has fully analyzed an expression boundary between X inactivated and expressed transcripts to search for additional regulatory elements that may protect escape genes from erroneous silencing by the spreading of heterochromatin. Therefore, further analysis of escape domains and the nearby sequences is necessary.

Biological Significance Of X Chromosome Inactivation

The mammalian X chromosome encodes ~1,400 genes, spanning nearly 160 Mb. Not surprisingly, the mutation of a number of these genes results in detrimental phenotypes in affected individuals. To date 375 phenotypes show a distinctive X-linked inheritance pattern and the underlying X-linked locus has been identified for 239 of these (Online Mendelian Inheritance in Man, OMIM - http://www.ncbi.nlm.nih.gov/Omim). In fact the X chromosome is associated with a disproportionately large number of disease conditions for its size, although this may be due, in part to the ease of identifying X-linked disorders since recessive mutations are directly revealed in XY males. A second possibility is that the X chromosome contains a disproportionate number of genes that contribute to particular traits. This second possibility is supported by the finding that 20% of all mental disability traits map to the X chromosome (OMIM) where a disproportionately large number of genes, coding for essential brain functions, are located (Zechner et al. 2001; Inlow et al. 2004). However, this description of X-linked diseases fails to account for why the severity of disease associated phenotypes can vary widely between affected females (Dobyns et al. 2004; Orstavik 2009). A complicating factor in the penetrance of X-linked diseases in females is the process of XCI.

X chromosome inactivation transcriptionally silences the majority of genes on one of the two X chromosomes in eutherian females. This process is essential, and failure of cells to undergo XCI leads to embryonic lethality (Rastan et al. 1985). As a result of XCI, females are mosaics of cells with either the maternally- or paternally-derived X chromosome inactivated. In most females XCI is a random process leading to nearly equal populations of cells having the maternally or paternally derived X as the inactive allele (Amos-Landgraf et al. 2006). Yet, females with X-linked diseases can exhibit highly skewed XCI in which a large proportion of cells have active X chromosomes derived from a single parental allele (Orstavik 2009).

Differences in XCI skewing between individuals are frequently proposed to explain phenotype spectrums in female carriers of X-linked mutations (Orstavik 2009). One example is that a strong association between XCI skewing and the severity of phenotype has been observed in patients with Rett Syndrome, a neurodevelopmental disorder caused by mutation of the methyl-CpG binding protein 2 gene, *MECP2* (Van Esch et al. 2005). A mouse model that recapitulates the phenotypes associated with Rett Syndrome confirms that XCI skewing has a profound effect on the severity of the disease; the more skewed XCI such that the normal *Mecp2* allele is on the active X, the less severe the phenotype (Young et al. 2004). However, not all X-linked diseases show such a strong correlation between XCI skewing and phenotype variation. In the case of Duchenne Muscular Dystrophy, a disease that is the result of a mutant X-linked *DMD* gene, multiple studies of symptomatic and non-symptomatic carriers demonstrate a correlation between XCI skewing and disease presentation (Burn et al. 1986; Pegoraro et al. 1994; Pegoraro et al. 1995; Yoshioka et al. 1998). However, a similar study identified symptomatic carriers that did not have skewed XCI (Matthews et al. 1995). These examples highlight the need for more rigorous analysis of the relationship between XCI skewing and X-linked disease phenotypes, and suggest that the severity of a given phenotype is not dependent on XCI skewing alone.

Escape from XCI may also influence the severity of phenotypes associated with X-linked disorders. Steroid sulfatase deficiency or X-linked ichthyosis is caused by a mutant steroid sulfatase gene (*STS*). Since the STS gene escapes XCI, females that are heterozygotes for a mutant *STS* allele are phenotypically normal or display only a mild phenotype as all cells produce at least some normal STS protein (Shapiro et al. 1979). In addition to single gene disorders, escape genes have also been hypothesized to contribute to phenotypes associated with X chromosome aneuploidies. Many of the characteristic phenotypes of Turner Syndrome (45, XO), Kleinfelter Syndrome (47,XXX), and multiple X syndrome (47,XXXX; 48,XXXXX; 49,XXXXXX) are thought to be due to improper dosage of PAR genes (Rao et al. 1997; Ross et

al. 2000), but not all of the phenotypic traits can be explained by PAR genes alone. Additionally, X aneuploidy mouse models exhibit a relatively normal phenotype (Ashworth et al. 1991; Lue et al. 2001) which may be due in part to the scarcity of escape genes on the mouse inactive X. Another example of a syndrome associated with a gene that escapes XCI is X-linked infantile Spinal Muscular Atrophy that is characterized by low muscle tone and has been linked to mutations in the ubiquitin activating enzyme E1 gene, *UBA1* (Ramser et al. 2008), although the role of escape in disease presentation has not been analyzed. Further, genes that exhibit variable patterns of escape should impact the severity of X-linked diseases as well. Partial escape of the *OFD1* gene, responsible for X-linked dominant disorder oral facial digital type 1 syndrome, has been hypothesized to contribute to variation (Morleo et al. 2008). Combined these studies show that XCI is a critical process essential for proper gene dosage and normal development.

Dissertation Overview

As our understanding of the process of X chromosome inactivation has evolved, so too has the complexity of mechanisms proposed to explain gene regulation on the inactive X chromosome. Today, several enticing mechanisms of escape gene regulation have been proposed, and are supported by increasingly large bodies of evidence, as described previously (Chapter 1, *Genomic Influences On XCI Propagation And Escape Gene Regulation*). Yet, none of these models that are proposed to explain XCI gene regulation have been functionally tested. This is due to the fact that no developmental model system exists that allows for the manipulation of human sequences to directly test their function in the XCI process. Further, none of the proposed elements involved in XCI gene regulation, including local sequence composition, *cis*-regulatory elements and/or chromatin modifications, predict escape genes with complete accuracy. This suggests that either multiple regulatory systems are used to regulate genes on the inactive X, or novel regulatory mechanisms are at work.

The distribution of certain repetitive sequence elements appears to correlate with human escape genes: L1s are depleted at escape genes (Bailey et al. 2000; Carrel et al. 2006; Wang et al. 2006), while Alu sequences are enriched (Wang et al. 2006). However, it is difficult to imagine how the depletion or enrichment of repetitive sequences can account for the inactive X expression differences at nearby loci. Recent, evidence suggests *Xist*-dependent transcription of L1s may aid in the effective silencing of genes near expressed genes (Chow et al. 2010), yet their presence near human escape domains has not been examined. Models that position barrier elements, such as insulators, at the boundaries of escape domains are appealing and can account for differences in the X inactivation states of nearby genes. To date, boundary elements have only been confirmed near one escape gene in mouse (Filippova et al. 2005). Without a functional system to confirm their role in XCI, the identification of such elements at additional escape domains would further strengthen the hypothesis that boundary elements regulate escape. Further, the organization of escape genes along the human X chromosome suggests that escape is regulated at the level of chromosomal domains and both repetitive sequences and barrier elements could account for this. Additionally, escape domains that are located near silenced genes will require tightly controlled regulation to ensure nearby heterochromatin does not encroach on and silence genes within the escape domain. Therefore, to ensure proper regulation of the escape domains on the inactive X, we **hypothesize that *cis*-acting regulatory elements located at the boundaries of these domains regulate expression and protect genes within these domains from improper silencing**.

To expand and clarify our understanding of the mechanisms regulating gene expression on the inactive X, I chose to analyze sequences at the cluster of escape genes in Xp11.23 with the smallest boundary between X inactivated and expressed transcripts. In this cluster, the ubiquitin activating enzyme E1, *UBA1* (formerly designated as *UBE1* and *A1S9T*), PCTAIRE cdc2 kinase homologue, *PCTK1*, and the ubiquitin specific peptidase 11, *USP11* genes escape X inactivation

(Figure 1-6) (Carrel et al. 1996; Carrel et al. 2005; Goto et al. 2009), whereas the upstream RNA binding motif gene, *RBM10* (previously designated *DXS8237E*)**,** is subject to inactivation (Coleman et al. 1996) (Figure 1-6). Intriguingly, unlike *UBA1* and *PCTK1* that escape X inactivation in all tested cell lines, *USP11* escapes inactivation in only a subset of tested lines (Carrel et al. 2005; Goto et al. 2009). This variable expression pattern of *USP11* was not addressed in the current study. Instead, I focused on the 5' boundary of the escape domain. Unlike other escape domains that have large boundaries between X inactivated and expressed genes that have been previously studied, the small size of the human *RBM10*/*UBA1* expression boundary allows all sequences within the boundary to be closely examined. Further, mouse *Uba1* is X inactivated (Disteche et al. 1992; Yang et al. 2010) enabling comparative analysis between species to identify regulatory elements specific to the human escape domain.

The small size of the expression boundary separating expressed human *UBA1* transcripts from upstream silenced transcripts suggests the regulatory elements responsible are acting at or near this boundary. However, it is difficult to imagine how the depletion or enrichment of

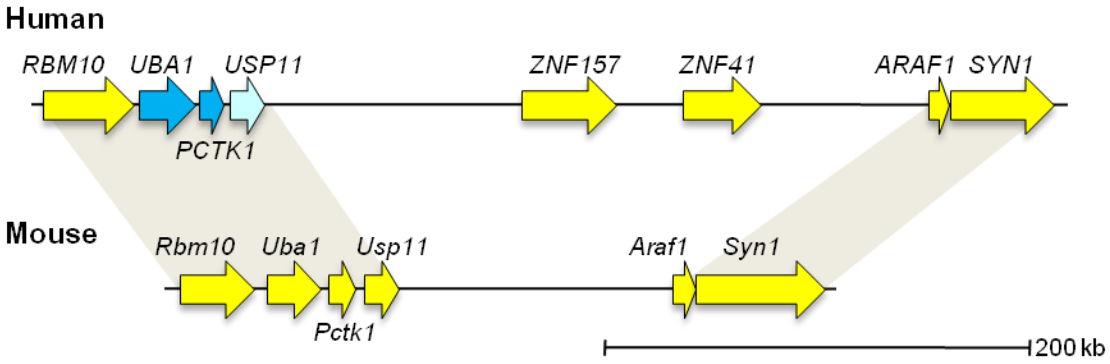


Figure 1-6. Genomic region surrounding the human Xp11.23 escape domain and the orthologous region in mouse. X inactivation status of genes in the region is shown; silenced genes are yellow, and expressed genes are blue. *USP11* has variable expression (light blue), escaping X inactivation in a subset of cell lines. The human locus contains two additional genes within the region, *ZNF157* and *ZNF41*. This difference likely is a result of a human specific expansion of the Krüppel-associated box (KRAB) family of proteins to which *ZNF157* and *ZNF41* belong (Venter et al. 2001).

repetitive sequences can account for the inactive X expression differences at such closely juxtaposed loci. Further, the presence of active L1s that may facilitate silencing near the *UBA1* locus has not been examined. Although boundary elements at the human *UBA1* escape boundary have been postulated (Goto et al. 2009), to date, the region has not been assayed for insulator function. In the chapters that follow, I sought to identify *cis*-acting regulatory elements at the human *UBA1* escape domain. To this end, three complementary approaches were utilized to identify and distinguish gene-specific elements from those specifying the human *UBA1* escape domain:

(1) Comparative analysis of the *UBA1* locus between human and mouse identified candidate sequence elements lacking in mouse *Uba1*. Further, expression analysis of human and mouse *UBA1* refined the location of the escape boundary between X inactivated and expressed transcripts to a small region within the 5' untranslated region of the human *UBA1* gene itself.

(2) Functional analysis of sequences near the *UBA1* expression boundary between X inactivated and expressed transcripts identified a human-specific chromatin insulator that localized to a region within the *UBA1* escape boundary.

(3) Analysis of epigenetic modifications at the human and mouse *UBA1* locus identified distinct methylation and histone modification profiles that delimit the identified human insulator sequence and distinguish the human inactive X from the active X chromosome.

Combined, these approaches successfully identify sequences at the human *UBA1* locus capable of regulating inactive X expression. The significance of the experimental results (Chapter 3) and their impact on the prevailing models of gene regulation on the inactive X are discussed in detail in Chapter 4, along with future experiments that will refine the role of insulators on the inactive X chromosome.

Chapter 2

MATERIALS AND METHODS

Cell Line And Culture Conditions

Most mouse-human somatic cell hybrids utilized in our laboratory require complementation of a murine *Uba1* temperature sensitive mutation to select for active or inactive human X chromosomes (Brown et al. 1989). Importantly, none of these hybrids were used for the present experiments, as this selection requires high levels of *UBA1* that may influence normal transcriptional control. Instead, previously characterized active X hybrids, Aha11aB1 and A23-1Ac5 (Brown et al. 1997; Carrel et al. 2005) were selected for *HPRT* activity. Inactive X hybrids were initially identified cytogenetically to carry a human X, but lacked *HPRT* activity. Subsequently, RT-PCR for multiple X loci confirmed the presence of a normal human inactive X. Two of these inactive X hybrids, LT23-1E2Buv5Cl26-7A2 and L23-4B have been previously described (Brown et al. 1997; Carrel et al. 2005). Two additional hybrids used for these studies, A48-5E-aza4A3-S2 and A62-1AS1, were obtained from H. Willard. These lines had been established by fusing the mouse cell line A9 (Littlefield 1966) to human fibroblasts GM2859 and GM7151 respectively and screened as described above. Somatic cell hybrids retaining an active human X chromosome were maintained in alpha-minimal essential media (α-MEM) supplemented with 10% fetal bovine serum (FBS), 2mM glutamine, HAT (Invitrogen 21060-017) and antibiotics. Somatic cell hybrids retaining an inactive X were maintained in α-MEM supplemented with 10% FBS, 2mM glutamine, and antibiotics.

The human fibroblast lines 77 and 78 were provided by H. Willard and previously characterized (Clarke et al. 1992; Kirchgessner et al. 1995). Additional primary human fibroblast

lines GM02621, GM03322, GM01695, and GM00135, were obtained from the NIGMS Human Genetic Mutant Cell Repository; IMR90 and WI38 were obtained from ATCC. Human cell lines 77, 78, GM02621, GM03322, GM01695, and GM00135 were maintained in α-MEM supplemented with 20% FBS, 2mM glutamine, non-essential amino acids and antibiotics. Cell lines IMR90 and WI38 were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS, 2 mM glutamine, 1X non-essential amino acids, sodium pyruvate, and antibiotics. K562, a human erythroidleukemia cell line, was used to carry out our enhancer-blocking assay and was obtained from ATCC. K562 cells were maintained in RPMI 1640 medium supplemented with 10% FBS, 2 mM glutamine, and antibiotics.

Mouse fibroblasts were established from tissue outgrowths from adult C57Bl/6 mice provided by S. Bronson (Penn State College of Medicine). The non-randomly inactivated, mouse B119 cell line, an early passage primary fibroblast cell line derived from a (T16H × CAST) F1 female newborn mouse (Carrel et al. 1996) was provided by H. Willard. The parental T16H line carries a balanced X:16 translocation, such that F1 progeny from the T16H x *M. castaneous* cross, will inactivate the normal *M. castaneous* X chromosome in all cells. All mouse fibroblasts were maintained in α-MEM supplemented with 10% FBS, 2 mM glutamine and antibiotics.

Expression Assays

RNA was isolated using Trizol (Invitrogen) according to manufacturer's recommendations. 5 μg of RNA was used for each reverse transcription reaction, which were performed as described (Stahl et al. 2010). Expression of specific human *UBA1* splice variants was analyzed in somatic cell hybrids containing either an active or inactive X chromosome. cDNA from hybrid cell lines was amplified in the presence of 10% DMSO. Primer sequences for exon-specific amplification are shown in Table 2-1. Expression from the inactive X in primary cells was determined using the quantitative allele-specific primer extension assay, qSNaPshot,

and performed in triplicate as described using the ABI PRISM Q-Snapshot Multiplex Kit (Carrel et al. 2005; Stahl et al. 2010). Both the non-allele specific PCR and allele-specific PCR was performed using a 55^{o}C annealing temperature for all SNPs tested. Informative, primary non-randomly inactivated, female human fibroblast cell lines were tested for expression at SNPs rs2070169, rs41310655 and rs5945431. Mouse cell line B119 was assayed using SNP rs29651331. Inactive X expression was determined by first normalizing cDNA samples to genomic DNA to account for fluorescent nucleotide incorporation and detection bias, followed by calculating the allele ratio as a percentage of the minor allele compared to the major allele. Primer sequences for the allele-specific expression assays are shown in Table 2-2.

Table 2-1. Primers used in exon-specific amplification of human *UBA1* splice variants.

Gene	Exon	Exon-Specific Primers*		Annealing Temperature (^{o}C)	Number of Cycles
UBA1	Exon 1a	UBE1-47 A1S9T-11	F:GTGGCTTCAGCTCATCTTTG R:TCTTGGCCATTCCGTTGGTT	58	30
UBA1	Exon 1b	UBE1-45 A1S9T-11	F:TGAAGGTTCTCGGCTTGTGA R:TCTTGGCCATTCCGTTGGTT	55	40
UBA1	Exon 1d	UBE1-49 A1S9T-11	F:GCAGCGGCGATTCTAGGC R:TCTTGGCCATTCCGTTGGTT	58	35
UBA1	Exon 1e	UBE1-39 A1S9T-11	F:TGTCATTCCTTTCACTCCAGGA R:TCTTGGCCATTCCGTTGGTT	55	35

*All primers in this table were designed by Laura Carrel.

Table 2-2. Primers used in the qSNaPshot allele-specific expression assays.

Species	SNP	Gene	Non-specific Primers		Allele-specific Primer	
Human	rs41310655	*UBA1*	UBA1 snap 2 F1 UBA1 snap 2 R1	F:GCGCAAACGAGTTCATTGAT R:TTTGCTAGACGCCAGGTTTT	UBA1 snap 2 F2	F:GCTTAATTCATGAGGAGC
Human	rs2070169	*UBA1*	UBE1 snap 1F* UBE1 snap 1R*	F:CCCATCATGCAGTGGCTATAC R:TACTTCTGCTTGCCCAGCTT	UBE1 snap 3R*	R:TTGCCCGTCATAACGGTTCTGG
Human	rs5945431	*PLXNA3*	HSSEX snap 1F* HSSEX snap 1R*	F:ATTCTCTCTGAGCACCTGGA R:ACAACACTGAGGACGTTTGG	HSSEX snap 2R*	R:AGTAGGCTCAGGAGGAGGAG
Mouse	rs29651331	*Uba1*	Ube1x SNAP 1 F1 Ube1x SNAP 1 R1	F:AGGGCATGATCCAACTCAAT R:ATGCCTCCACGGATGTAGTC	Ube1x SNAP 1 R2	R:CAGATACTAAAGGTATAAGG

* Denotes primers designed by Gabrielle Nickel.

Sequence Analysis

The 5' untranslated region of the *UBA1* locus was compared between human and mouse using two sequence comparison tools. *UBA1* sequences 5' of the translation start site were aligned using the MUSCLE alignment tool (Edgar 2004). MUSCLE alignments were used to determine the percent identity between human and mouse *UBA1* sequences. Sequence similarity was visualized using a nucleic acid dot plot (http://www.vivo.colostate.edu/molkit/dnadot/) with parameters: 19 bp window, 3 bp mismatch. All sequences used throughout this study were obtained from the UCSC genome browser (http://genome.ucsc.edu); human sequences were from the February 2009 assembly (hg19), and mouse sequences were from the July 2007 assembly (mm9). Repetitive sequences were identified using the RepeatMasker track in the UCSC genome browser. L1 and Alu repeat density was determined throughout the 4 Mb flanking the *UBA1* locus using a sliding-window analysis (50 kb window; 5 kb slide). Full length, intact L1 repetitive elements were identified using the L1Xplorer analysis program (http://line1.bioapps.biozentrum.uni-wuerzburg.de/l1xplorer.php) (Penzkofer et al. 2005).

Enhancer Blocking Assays

The pJC13-I plasmid obtained from G. Felsenfeld has been previously described (Chung et al. 1993). An uninsulated control plasmid, pJC13-NI, was made by removal of the chicken β-globin HS4 insulators located between the enhancer and neomycin reporter gene by *KpnI* restriction digestion and self-ligation of the vector (Figure 2-1). Candidate *UBA1* insulator sequences were PCR amplified using primers engineered with either *KpnI* or *NdeI* restriction site tails to facilitate direct cloning into the pJC13-NI vector (Table 2-3). Test sequences were inserted into *KpnI* or *NdeI* restriction sites within the pJC13-NI vector to position sequences 5' or 3' of the enhancer, respectively. Plasmid integrity and orientation was confirmed by multiple

restriction digests and/or DNA sequencing. Two independent clones were assayed for each construct, excluding mouse *Uba1* Fragment A. The mouse fragment A contains a large simple TC rich repetitive sequence that is difficult to amplify. For this fragment we assayed one independent clone of the constructs containing the insert 5' or 3' of the enhancer. We sequenced the entire mouse *Uba1* Fragment A insert to confirm all sequences, excluding the repetitive region, were intact in both constructs. The enhancer blocking assay was performed in at least duplicate for each independent construct generated for each candidate insulator sequence, as previously described with minimal modifications (Chung et al. 1993). Briefly, 42.5 pmol of *AatII* linearized vector DNA was mixed with 10^7 mid log phase K562 cells in serum–free RPMI media. The cells were electroporated at 200V and 960 μF using a Bio Rad Gene PuslerXcell System. After a 24-

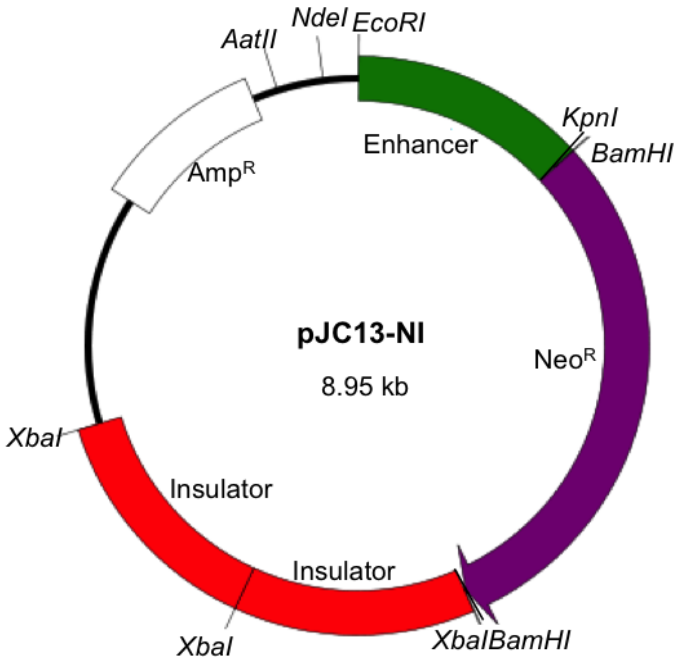


Figure 2-1. Schematic of the pJC13-NI vector. Candidate insulator sequences were cloned upstream (*NdeI*) or downstream (*KpnI*) of the enhancer (green) to test their ability to block activation of the Neomycin reporter gene (Neo^R-purple). Tandem copies of the chicken β-globin insulator (red) block downstream position effects. The plasmid was linearized prior to transfection into using the unique *AatII* restriction site. Ampicillin resistance (Amp^R) allows selection for the plasmid in bacterial constructs.

hour recovery in medium, the cells were plated in soft agar with 750 μg/mL G418. Neomycin resistant colonies were counted after 3 weeks of selection. All test constructs were normalized to the pJC13-NI control. Statistical significance of the reduction in colony formation was evaluated using the Student's t-test.

Table 2-3. Primers designed to amplify candidate insulator sequences at the *UBA1* locus.

Gene	Fragment	Enhancer Blocking Primers*		Product Size (bp)	Annealing Temperature (°C)	Number of Cycles	Permanent Stock (BCC#)
Human *UBA1*	1	KEP043 InsKpnR†	F:gcatctggtaccTTCCCTTGCAAGTACCAGAG‡ R:tagcgtggtaccCCAGCAACCAGAAAGCAGAT	1674	58	35	280;287
		InsNdeIF† InsNdeIR†	F:gcatctcatatgTTCCCTTGCAAGTACCAGAG‡ R:tagcgtcatatgCCAGCAACCAGAAAGCAGAT				310,314
Human *UBA1*	2	KEP043 KEP013	F:gcatctggtaccTTCCCTTGCAAGTACCAGAG‡ R:gggcccggtaccACTCAGCCACTAGAGAAACC	610	55	35	367,372
		InsNdeIF† UBE1-32†	F:gcatctcatatgTTCCCTTGCAAGTACCAGAG‡ R:tagcgtcatatgACTCAGCCACTAGAGAAACC				389,393
Human *UBA1*	3	KEP035 InsKpnR†	F:gggcccggtaccGGTTTCTCTAGTGGCTGAGT R:tagcgtggtaccCCAGCAACCAGAAAGCAGAT	1076	55	35	404,408,434
		KEP036 InsNdeIR†	F:gggccccatatgGGTTTCTCTAGTGGCTGAGT R:tagcgtcatatgCCAGCAACCAGAAAGCAGAT				409,410
Human *UBA1*	4	KEP035 KEP030	F:gggcccggtaccGGTTTCTCTAGTGGCTGAGT R:gcgcgcggtaccCTGCCTTGGTCCAATCAAAC	551	55	35	448,450,451, 454
Human *UBA1*	5	KEP031 InsKpnR†	F:ggcccggtaccAGTTTGGGTTTGATTGGACCA R:tagcgtggtaccCCAGCAACCAGAAAGCAGAT	552	57	35	455,456,457
Human *UBA1*	6	KEP029 KEP030	F:gggcccggtaccCCTTGTGCCAGCAATGTTCT R:gcgcgcggtaccCTGCCTTGGTCCAATCAAAC	435	55	35	422,421,425

*All primers contain tails (lowercase) to aid cloning: a short filler sequence to ensure efficient cutting, followed by a *KpnI* or *NdeI* site.
†Denotes primers designed by Melanie Moon.
‡ An internal KpnI restriction site was removed from the primer by replacing the G with the underlined base.

Table 2-3 (cont.). Primers designed to amplify candidate insulator sequences at the *UBA1* locus.

Gene	Fragment		Enhancer Blocking Primers*	Product Size (bp)	Annealing Temperature (°C)	Number of Cycles	Permanent Stock (BCC#)
Human *UBA1*	7	KEP054 KEP032	F:gggcccggtaccCAGGCGTCATCTGATTTTCA R:gcgcgcggtaccACACTCTCCCAGGTCCTGAA	421	55	35	481,482,483, 484
		KEP049 KEP055	F:gggccccatatgCAGGCGTCATCTGATTTTCA R:gcgcgccatatgACACTCTCCCAGGTCCTGAA				490,491
Human *UBA1*	8	KEP031 KEP032	F:gggcccggtaccAGTTTGGGTTTGATTGGACCA R:gcgcgcggtaccACACTCTCCCAGGTCCTGAA	334	57	35	440,441,428
		KEP037 KEP038	F:gggccccatatgAGTTTGGGTTTGATTGGACCA R:gcgcgccatatgACACTCTCCCAGGTCCTGAA				471,472,473, 477, 478
Human *UBA1*	9	InsCpGF† InsCpGR†	F:gcatctggtaccTCAGGAAGCATGGCCTTAAC R:tagcgtggtaccTCCGACTCCAAGGTCAGATT	697	54	35	300,302,295, 296
		CpGInsNdeIF† CpGInsNdeIR†	F:gcatctcatatgTCAGGAAGCATGGCCTTAAC R:tagcgtcatatgTCCGACTCCAAGGTCAGATT				326,334,325, 331
Mouse *Uba1*	A	KEP026 KEP012	F:gggcccggtaccCGCAGTTAGCTTTCCCTCAC R:gcgcgcggtaccTACTTGGCAGGCTGACACAG	761	55	35	375
		KEP018 KEP053	F:gggccccatatgCGCAGTTAGCTTTCCCTCAC R:gcgcgccatatgCACAGTACTTGGCAGGCTGA				523
Mouse *Uba1*	B	KEP024 KEP010	F:gggcccggtaccCTGTGTCAGCCTGCCAAGTA R:gcgcgcggtaccCTGGGAATCAAGGAGACCAA	1607	55	35	379,403
		KEP021 KEP052	F:gggccccatatgCTGTGTCAGCCTGCCAAGTA R:gcgcgcgcatatgCTGGGAATCAAGGAGACCAA				479,474,495, 516, 525
Mouse *Uba1*	C	KEP025 KEP017	F:gggcccggtaccCTTGGTGGAGCTGTTCTCCT R:gcgcgcggtaccGCTGAGATAGGGGTCTCACG	587	55	35	373,380
		KEP047 KEP048	F:gggccccatatgCTTGGTGGAGCTGTTCTCCT R:gcgcgccatatgGCTGAGATAGGGGTCTCACG				504,505,507

*All primers contain tails (lowercase) to aid cloning: a short filler sequence to ensure efficient cutting, followed by a *KpnI* or *NdeI* site.
†Denotes primers designed by Melanie Moon.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed in human lines 78, GM02621, IMR90, and mouse fibroblasts. ChIP assays containing 7.5 x 10^6 cells were performed as described (Nelson et al. 2006) with minor modifications. Briefly, cells were crosslinked with 1% formaldehyde for 8 minutes, washed with PBS and harvested. Chromatin lysates were sonicated to an average size of ~150-300 bp using a Misonex XL-2000 probe sonicator (5 x 20 sec, 6-W pulses with 60-sec rest intervals on ice). Lysates were cleared by centrifugation and incubated with primary antibody overnight at 4^{o}C (CTCF – Abcam ab70303; Acetyl H3 - Millipore 17-615; trimethyl-H3(K27) - Millipore 17-622; and normal Rabbit IgG – Millipore 12-370). An aliquot of soluble chromatin was reserved as the input fraction for quantitation. Immunocomplexes were captured by Protein A Agarose/Salmon Sperm DNA (Millipore 16-157), washed, and precipitated DNA fragments were isolated as described (Nelson et al. 2006). ChIP DNA was quantitated using the ABI 7900HT real-time PCR System in the Penn State College of Medicine Functional Genomics Core Facility. Technical triplicates of each ChIP DNA sample were quantitated by comparison to known concentrations of the input fraction. Samples (1x Power Sybr Green Master Mix (ABI), 0.1 μM of each primer, and 2 μL DNA) were amplified using recommended PCR Conditions: 10 min at 95^{o}C, followed by 50 cycles of 95^{o}C for 15 sec, 60^{0}C for 1 min. Dissociation curves were calculated for each sample to ensure C_t values reflected a single amplified product. Enrichment values for all antibodies were calculated by comparison to known amounts of input DNA. CTCF enrichment is reported as the fold increase over an experimentally identified, negative binding control region that lies upstream of the *UBA1* locus. The CTCF enrichment value for the negative control region was calculated as the mean enrichment for triplicates of two adjacent primer sets in the region just 3' of the *RBM10* gene that lacked CTCF binding in our study and others (McDaniell et al. 2010). Whereas CTCF enrichment was reported

as the fold increase over a control region, the *UBA1* locus lacked a suitable region devoid of acetyl H3 and trimethyl-H3K27. Therefore, histone enrichment is reported as the calculated enrichment value by comparison to input DNA. All assays were performed in triplicate, with the exception of cell line GM02621 in which all assays were performed in duplicate. All primers used in ChIP assays are shown in Table 2-4.

Allele-specific analysis of purified DNA from the ChIP assays was performed using the quantitative allele-specific primer extension assay, qSNAPshot, similar to that of the *UBA1*/*Uba1* expression analysis. Histone and CTCF enrichment was measured on the inactive and active Xs utilizing the *UBA1* SNP rs41310655 in cell line GM02621.

DNA Methylation Analysis

Bisulfite conversion of genomic DNA, in which all unmethylated cytosine residues are deaminated to form uracil upon treatment with bisulfite, was performed using the EZ DNA Methylation Kit (Zymo Research). Amplification primers (Table 2-5) were designed using Pyrosequencing Assay Design Software v.1.0 (Qiagen) or MethPrimer software (http://www.urogene.org/methprimer/index1.html) to amplify the coding strand of bisulfite converted DNA. For most regions tested, PCR products were subcloned into the pGEM-T vector (Promega) and at least eight individual strands were sequenced using the M13F-20 and M13R primers. Alternatively, bisulfite modified DNA was PCR amplified and analyzed directly by pyrosequencing (Qiagen PyroMark Q24) following addition of a biotin tag that involved a second PCR reaction with a labeled universal PCR primer (5'-GGGACACCGCTGATCGTTTA3') (PMID: 12866414). Multiple nested sequencing primers were necessary to sequence biotinylated PCR products in their entirety. Percent methylation was calculated by dividing the number of methylated strands by the total number of sequenced strands for each individual CpG site or directly by pyrosequencing.

Figure 2-4. Primers used for chromatin immunoprecipitation of human and mouse cells.

Gene	ChIP Primers		Relative Location*	Chromosomal Location of PCR Product
Human *H19*	Human H19-h1 F[†]	F:CCCATCTTGCTGACCTCAC		chr11:2,024,182-2,024,201
	Human H19-h1 R	R:AGACCTGGGACGTTTCTGTG		
Human *H19*	Human H19-NC F[‡]	F:GAGCTCTAAGGGAGGCTCCAG		chr11:2,027,399-2,027,418
	Human H19-NC R	R:CATCATGGTGTCCTCACAGG		
Human *UBA1*	Hu UBA1 ChIP Set 3 F[§]	F:GGCAGGGAAGGACAGAGTGT	-4155	chrX:47,046,152-47,046,171
	Hu UBA1 ChIP Set 3 R[§]	R:CAGCCCAATTTTCCAACAAA	-4028	
Human *UBA1*	Hu UBA1 ChIP Set 4 F[§]	F:ATCTGGAGACAATGCCCTCA	-3617	chrX:47,046,721-47,046,740
	Hu UBA1 ChIP Set 4 R[§]	R:ATGGTTCTCAAAGGGCATCC	-3459	
Human *UBA1*	Hu UBA1 ChIP Set 5 F	F:GGCTGCCCTAGTCTCAGGAT	-2632	chrX:47,047,674-47,047,693
	Hu UBA1 ChIP Set 5 R	R:CTCCTGCCAGTTGAGCTTTG	-2506	
Human *UBA1*	Hu UBA1 ChIP Set 6 F	F:ATGTCTTGATTGGGGTGCTG	-2020	chrX:47,048,292-47,048,311
	Hu UBA1 ChIP Set 6R	R:GGTTTGCTGTGGTGGTGATT	-1888	
Human *UBA1*	Hu UBA1 ChIP Set 7 F	F:CATACATCGCTGGAGGCAAT	-1050	chrX:47,049,254-47,049,273
	Hu UBA1 ChIP Set 7 R	R:GCATAGTGGGGTGCATCTGT	-926	
Human *UBA1*	Hu UBA1 ChIP Set 8 F	F:GCCCAAGGAAGAATTTCCAG	-366	chrX:47,049,949-47,049,968
	Hu UBA1 ChIP Set 8 R	R:GTGGGAAGAGGAAACGGAAC	-231	
Human *UBA1*	Hu UBA1 ChIP Set 9 F	F:GCCCCTCTAGCAAAGCATCT	-28	chrx:47,050,272-47,050,291
	Hu UBA1 ChIP Set 9 R	R:AACGGAAGAGCAAACCACCT	+92	
Human *UBA1*	UBA1 SNAP2 F	F:GCGCAAACGAGTTCATTGAT	+904	chrX:47,051,260-47,051,279
	UBA1 SNAP2 R	R:TTTGCTAGACGCCAGGTTTT	+1080	
Human *UBA1*	Hu UBA1 ChIP Set 11 F	F:AAATCCCATCGATCCTTTCC	+1152	chrX:47,051,529-47,051,548
	Hu UBA1 ChIP Set 11 R	R:TAGCCTTCCTGTGTCCCAGT	+1349	
Human *UBA1*	UBE1 CTCF2 ChIP F	F:TGAGTTTGGGTTTGATTGGA	+2151	chrX:47,052,509-47,052,528
	UBE1 CTCF2 ChIP R	R:ATAGCCAAGGCAAAGTGAGC	+2329	

*The location relative to *UBA1* exon 1a transcription start site (human-ChrX:47,050,199; mouse-ChrX:20,235, 547).
[†]Primers designed for a CTCF-binding region at the *H19* locus (Burke et al. 2005).
[‡] Primers designed for a region that does not bind CTCF at the *H19* locus (Burke et al. 2005).
[§]Primers used as the negative control in human CTCF ChIP experiments.

Figure 2-4 (cont.). Primers used for chromatin immunoprecipitation of human and mouse cells.

Gene	ChIP Primers		Relative Location*	Chromosomal Location of PCR Product
Human *UBA1*	Hu UBA1 ChIP Set 13 F	F:GTACACGTCCCCTGGGTTTT	+2241	chrX:47,052,537-47,052,556
	Hu UBA1 ChIP Set 13 R	R:CTGCCCAGCAAATAGGGATT	+2329	
Human *UBA1*	Hu UBA1 ChIP Set 14 F	F:AGCATGGCCTTAACGGTTTT	+2753	chrX:47,053,061-47,053,080
	Hu UBA1 ChIP Set 14 R	R:TGGTGGACATTGTGGCAGTA	+2881	
Human *UBA1*	UBE1 CTCF1 ChIP F	F:ATGAATCGCAACCGAGTAGG	+2908	chrX:47,053,267-47,053,285
	UBE1 CTCF1 ChIP R	R:CCGTTGTGGTTGTGGTTGT	+3086	
Human *UBA1*	Hu UBA1 ChIP Set 15 F	F:GCCCCAGAATCTCCTTTGAG	+3614	chrX:47,053,937-47,053,956
	Hu UBA1 ChIP Set 15 R	R:GATGGGGGTCTGAAGGAAAA	+3757	
Human *UBA1*	Hu UBA1 ChIP Set 16 F	F:CTCAGCATCTTCCCCATCAG	+4054	chrX:47,054,376-47,054,395
	Hu UBA1 ChIP Set 16 R	R:GCTAGAAGCCAACGGGATCT	+4196	
Human *UBA1*	Hu UBA1 ChIP Set 17 F	F:CAGAGAAACCCAACCCCATT	+4836	chrX:47,055,164-47,055,183
	Hu UBA1 ChIP Set 17 R	R:GGCCCTGTGTCAATCTCTGA	+4984	
Human *UBA1*	Hu UBA1 ChIP Set 18 F	F:TGCTAACCTGCCTCCTCTCA	+5567	chrX:47,055,921-47,055,940
	Hu UBA1 ChIP Set 18 R	R:AGGCCCCCACCATAACATAC	+5741	
Mouse *Uba1*	Igf2r ChIP F[†]	F:TATCGGCCCTCGTGTAGTTC		chr17:12,934,371-12,934,390
	Igf2r ChIP R	R:GAGGATTCCACGCGTTAGAG		
Mouse *Uba1*	M Uba1 ChIP Set 8 F[‖]	F:CTTGCTCCCTGTGCCTGTAG	-3499	chrX:20,232,175-20,232,194
	M Uba1 ChIP Set 8 R[‖]	R:GGGTCCCATTCTGGACTGAT	-3353	
Mouse *Uba1*	M Uba1 ChIP Set 9 F[‖]	F:GACCCCTTCACAGGGATTGT	-2536	chrX:20,233,147-20,233,166
	M Uba1 ChIP Set 9 R[‖]	R:CTGCGACCCTTTAATGCAGA	-2381	
Mouse *Uba1*	M Uba1 ChIP Set 11 F	F:ATGCCAACCTTAAGGCGTTT	-1672	chrX:20,234,023-20,234,044
	M Uba1 ChIP Set 11 R	R:TTGAGCACCTATTCAGTGCAGA	-1503	
Mouse *Uba1*	M Uba1 ChIP Set 12 F	F:ACATGGGCATGGTGGTACAT	-707	chrX:20,234,935-20,234,955
	M Uba1 ChIP Set 12 R	R:TGCAACTAGGCTGACTGAACA	-592	

*The location relative to *UBA1* exon 1a transcription start site (human-ChrX:47,050,199; mouse-ChrX:20,235, 547).
[†]Primers designed for a CTCF-binding region at the *H19* locus (Burke et al. 2005).
[‖]Primers used as the negative control in mouse CTCF ChIP experiments.

Figure 2-4 (cont.). Primers used for chromatin immunoprecipitation of human and mouse cells.

Gene	ChIP Primers		Relative Location*	Chromosomal Location of PCR Product
Mouse *Uba1*	M Uba1 ChIP Set 13 F M Uba1 ChIP Set 13 R	F:GGGGTCGTACCAAAACCTGT R:TCGATCGTGCAAAAAGAACC	-145 +28	chrX:20,235,556-20,235,575
Mouse *Uba1*	M Uba1 ChIP Set 15 F M Uba1 ChIP Set 15 R	F:CAGAGAAGATGCCGCAGGTA R:TCCTGCATACCCACTTTTCG	+1029 +1155	chrX:20,236,683-20,236,702
Mouse *Uba1*	M Uba1 ChIP Set 104 F M Uba1 ChIP Set 104 R	F:GTTTCCTGAGGCTTGCTGAA R:CCTCCTGCCTCACTAAGCAC	+2246 +2354	chrX:20,237,882-20,237,901
Mouse *Uba1*	M Uba1 ChIP Set 17 F M Uba1 ChIP Set 17 R	F:TTGGGGAAGTGCTTTATCCA R:CATGACCACTTAGGGGCGTA	+2960 +3100	chrX:20,238,628-20,238,647
Mouse *Uba1*	Ube1x CTCF2 ChIP F Ube1x CTCF2 ChIP R	F:GTTAACCAGGCTTGCCTCAG R:GCAGGAGGGTCTGAGTTTGA	+3598 +3811	chrX:20,239,339-20,239,358
Mouse *Uba1*	M Uba1 ChIP Set 106 F M Uba1 ChIP Set 106 R	F:TGCAGTAGCCCTGTTTCTTG R:AGCTGGGAATCAAGGAGACC	+3993 +4127	chrX:20,239,655-20,239,674
Mouse *Uba1*	Ube1x CTCF1 ChIP F Ube1x CTCF1 ChIP R	F:CAAACTGGGCAGGGAAAATA R:GGCCAGCTAGGATCACTGTC	+4399 +4604	chrX:20,240,132-20,240,151
Mouse *Uba1*	M Uba1 ChIP Set 20 F M Uba1 ChIP Set 20 R	F:CCTGAGCCATGCATTTCCTA R:GGGGCAGGTCTTCAAATAGC	+4997 +5111	chrX:20,240,639-20,240,658
Mouse *Uba1*	M Uba1 ChIP Set 110 F M Uba1 ChIP Set 110 R	F:CCTCAGAATCTGCCCTCGTC R:AGCCTGTTGGGGAAAAAGGT	+5739 +5863	chrX:20,241,391-20,241,410
Mouse *Uba1*	M Uba1 ChIP Set 109 F M Uba1 ChIP Set 109 R	F:CAGCAGCTGTGGTACCTGTT R:CGAAGAGGCAAGGAGAGGTC	+6151 +6252	chrX:20,241,780-20,241,799
Mouse *Uba1*	M Uba1 ChIP Set 22 F M Uba1 ChIP Set 22 R	F:TCCTTGCCTCTTCGCACTTA R:GGGGGAATGTAGACCCTGAA	+6239 +6407	chrX:20,241,935-20,241,954
Mouse *Uba1*	M Uba1 ChIP Set 111 F M Uba1 ChIP Set 111 R	F:AGGCCTGCTAAGGTCTCATT R:TGAGAAAGGCATCAGAGGAAGT	+6652 +6737	chrX:20,242,263-20,242,284

*The location relative to *UBA1* exon 1a transcription start site (human-ChrX:47,050,199; mouse-ChrX:20,235, 547).

Table 2-5. Primers used in the methylation analysis of the human and mouse *UBA1* locus.

Gene	Methylation Primer*		Chromosomal Primer Sequence†	Annealing Temperature, # of Cycles (°C, #)
Human *UBA1*	UBE1 meth 1a CpG F	F:GGTAATTTTATAGTTTAGGTTTATATTTTT (-211)	GCAATTCCACAGCTCAGGCTTACATTTT	55 °C, 37
	UBE1 meth 1a CpG R	R:TTAAAATTAAAAAAACACCCTCCTC (+219)	GGAGGAGGGTGCTTCCCTAATCCCAAC	
Human *UBA1*	UBE1 meth 11	F:GTATAGGGGGTGAGTTGGAGTAT (+142)	GCACAGGGGGTGAGCTGGAGCAC	55 °C, 40
	UBE1 meth 12	R:CTCCTCTCCCTAAAAACATCTAAAC (+647)	GCCCAGATGTTTCTAGGGAGAGGAG	
Human *UBA1*	UBE1 meth 13	F:TTAGGTTTTTGTGATTTTTGGTTTT (+567)	TCAGGCCTTTGTGACCTCTGGTTTT	55 °C, 40
	UBE1 meth 14	R:AAAAAACCTCATCCCTTTACTAAAC (+1095)	GTCTAGCAAAGGGATGAGGCCTCCT	
Human *UBA1*	UBE1 meth 15	F:TTTAGTAAAGGGATGAGGTTTTTTATATT (+1072)	TCTAGCAAAGGGATGAGGCCTCCTACATT	55 °C, 40
	UBE1 meth 8	R:ACTTAAACCACTACTTCCCCATC (+1624)	GATGGGGAAGCAGTGGCCCAAGT	
Human *UBA1*	UBE1 meth 5	F:AAAATTTAATAGTGTTTTGTGTTAGTAATG (+1731)	AAAATTTAACAGTGCCTTGTGCCAGCAATG	55 °C, 37
	UBE1 meth 6	R:AATCAAAAATTAAAAACCAACCTAACC (+2022)	GGCCAGGCTGGTCTCCAATTCCTGACC	
Human *UBA1*	UBE1 meth 1	F:GGTTGGTTTTTAATTTTTGATTTTAG (+2001)	GGCTGGTCTCCAATTCCTGACCTCAG	55 °C, 37
	UBE1 meth 2	R:CAAAAAAAATTCCACCCAATTTATA (+2418)	CATAAATTGGGTGGAACCTTTCTTG	
Human *UBA1*	UBE1 meth 9	F:TTATTATTTAGGATTTGGGAGAGTGT (+2461)	TCATTATTCAGGACCTGGGAGAGTGT	53 °C, 40
	UBE1 meth 10	R:AAACAAAACCTTAAACCCCAC (+3005)	GTGGGGCCCAAGGTTTTGTTC	
Human *UBA1*	UBE1 meth 1d CpG F	F:TAGGTGGTGTTGTTTTATTTTTAAT (+2880)	CAGGTGGTGCTGTTCCACTTCTAAC	55 °C, 40
	UBE1 meth 1d CpG R	R:CTCTAAAAACACATCCTAAAAATCC (+3466)	GGACCCCCAGGATGTGCTTTTAGAGA	
Mouse *Uba1*	mUbe1x CpG 1a F	F:ATATTTTAGGGGTTGATGTTTATTA (-225)	ATATTCTAGGGGCTGATGCCCACCA	55 °C, 37
	mUbe1x CpG 1a R	R:AAAACACCCTCTTCCAAATACTATC (+162)	GACAGCATTTGGAAGAGGGTGCTTC	
Mouse *Uba1*	mUbe1x meth 7	F:TTATTGTTTTAGTGGGATTATGGTG (+707)	CTACTGCTTTAGTGGGACTATGGTG	55 °C, 40
	mUbe1x meth 8	R:CCTAATTCAACCATTCACTATTCAAC (+1313)	GCTGAACAGTGAATGGCTGAACTAGG	
Mouse *Uba1*	mUbe1x meth 3	F:TTGAATAGTGAATGGTTGAATTAGG (+1289)	CTGAACAGTGAATGGCTGAACTAGG	55 °C, 40
	mUbe1x meth 4	R:AAATATTCTTTAATACATCCAAAAAAAA (+1504)	TTCTTCTTGGATGTATTAAAGAACATTC	
Mouse *Uba1*	mUbe1x meth 1	F:GGTTTGTTGAAAGAATTAGGGATTT (+2255)	GGCTTGCTGAAAGAATTAGGGACCT	55 °C, 40
	mUbe1x meth 2	R:CCCAAATCAAAAAATAAAAATTAAAAATAA (+2726)	TTATTCTTAATCCTCATTTTCTGATCTGGG	

*All primers were designed to amplify the coding strand of bisulfate-converted DNA and the 5' end of the primer (#) is given relative to the *UBA1* TSS of exon 1a.
†Sequences shown are equivalent to the primers designed for bisulfate-converted DNA, but give the *UBA1* sequence prior to conversion.

Table 2-5 (cont.). Primers used in the methylation analysis of the human and mouse *UBA1* locus.

Gene		Primer Sequence*	Chromosomal Primer Sequence†	Annealing Temp., # of Cycles (°C, #)
Mouse *Uba1*	mUbe1x meth 5	F:TGATTTGGGAATATTTAAGTTTTTAAAT (+2718)	TGATCTGGGAACATTCAAGTCCTCAAAT	55 °C, 40
	mUbe1x meth 6	R:CACCCAAAATAACCCTACCTAAC (+3209)	GCTAGGCAGGGCCACTTTGGGTG	
Mouse *Uba1*	mUbe1x meth 9	F:GTGGGGGTTGTTTATTAAATTTTAT (+3207)	GTGGGGGCTGCCCATCAAACCTCAT	53 °C, 40
	mUbe1x meth 10	R:AATTCCAAAACAATCAAAACTACAC (+3774)	AATTCCAAAACAATCAAAACTACAC	
Mouse *Uba1*	mUbe1x CpG 1d F	F:AGAAGTTTTAAATTGGGTAGGGAAA (+4391)	AGAAGTCCCAAACTGGGCAGGGAAA	55 °C, 40
	mUbe1x CpG 1d R	R:AACTAACAAAAAACCAACAAAAACC (+4786)	GGTCCCTGCTGGCTCTCTGCCAGCC	
Mouse *Uba1*	Pyro M Ube1x Set 3 F	F:TAATGGGTATGTTTGGAGGT (+187)	TAATGGGCATGTTTGGAGGT	59 °C, 35
	Pyro M Ube1x Set 3 R‡	R:CTACCCTCCAACAAAATACTCAT (+406)	ATGAGTACTCTGTTGGAGGGCAG	
	Pyro M Ube1x Set 3a S§	F1:GGTATGTTTGGAGGTTG (+192)	GGCATGTTTGGAGGTTG	
	Pyro M Ube1x Set 3b S§	F2:GGTGTTTATTTTTGAGGAGA (+232)	GGTGTCTACTTTTGAGGAGA	
	Pyro M Ube1x Set 3c S§	F3:ATTTAGGAGTTTGTTTTATG (+286)	ATTTAGGAGTCTGTTTTATG	
Mouse *Uba1*	Pyro M Ube1x Set 4 F	F:GAGTATTTTGTTGGAGGGTAGAAT (+386)	GAGTACTCTGTTGGAGGGCAGAAT	57 °C, 40
	Pyro M Ube1x Set 4 R‡	R:AAAAACCTAACAAACAATCTTCCT (+664)	AGGAAGACTGTCTGCCAGGCTCTT	
	Pyro M Ube1x Set 4 S§	F1:GGAGAGAAGTTTGAAGTTAT (+427)	GGAGAGAAGTTTGAAGTTAC	
	Pyro M Ube1x Set 4c S§	F2:TGTGATTTTTGATGGG (+542)	TGTGACCTCTGATGGG	
	Pyro M Ube1x Set 4d S§	F3:GGGAGAGGAGTTTTTTAAT (+615)	GGGAGAGGAGCCTCTCAAC	
Mouse *Uba1*	Pyro M Ube1x Set 5 F	F:TTAGGGGTTTAGATTTTGGGAGAG (+598)	TCAGGGGCCCAGACCCTGGGAGAG	59 °C, 35
	Pyro M Ube1x Set 5 R‡	R:CCCCCTAAAAAACAATTCTCTTCT (+782)	AGAAGAGAATTGCTCTCTAGGGGG	
	Pyro M Ube1x Set 5 S§	F1:GTTAGGTTTTTTTGTGTAG (+654)	GCCAGGCTCTTCTGTGCAG	
	Pyro M Ube1x Set 5b S§	F2:TTGTTTTAGTGGGATTATG (+710)	CTGCTTTAGTGGGACTATG	
Mouse *Uba1*	Pyro M Ube1x Set 6 F	F:TTGGGTATTAGATGGAGTAATAAGGTAGGT(+4017)	CTGGGTATTAGATGGAGCAATAAGGCAGGC	59 °C, 35
	Pyro M Ube1x Set 6 R‡	R:AAAAAAAACCTCAAAACCATTCTTC (+4268)	GAAGAATGGCCTTGAGGTCCTTCCT	
	Pyro M Ube1x Set 6a S§	F1:TTGTTTTATTTTATTGAGTT (+4055)	TTGTTCCATCTCACTGAGTC	
	Pyro M Ube1x Set 6b S§	F2:AGTAAATTTATAATTTTGTT (+4207)	AGTAAACCCATAACCTTGCT	

*All primers were designed to amplify the coding strand of bisulfate-converted DNA and the 5' end of the primer (#) is given relative to the *UBA1* TSS of exon 1a.
†Sequences shown are equivalent to the primers designed for bisulfate-converted DNA, but give the *UBA1* sequence prior to conversion.
‡Denotes primers in which the universal tail (PMID: 12866414) has been added to facilitate amplification of biotinylated products.
§Denotes primers designed to sequence biotinylated PCR products on the PyroMark Q24 Pyrosequencing machine.

Chapter 3

RESULTS

All of the experiments described in this chapter were designed and performed by me, excluding the expression analysis of human *UBA1* isoforms (Figure 3-2) in somatic cell hybrid cells that was designed by Laura Carrel and performed by Matt Blome.

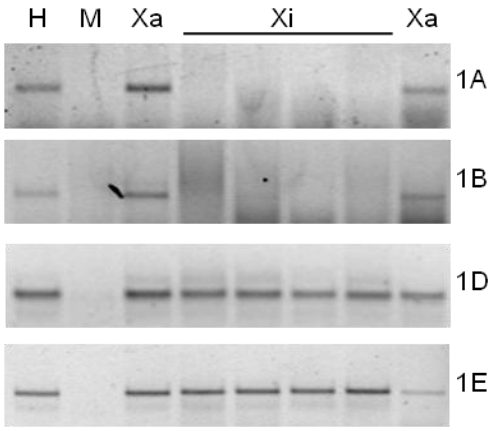
Alternative 5' Exons Of Human UBA1 Show Different X Inactivation Patterns

To identify elements involved in gene expression on the inactive X, we initially focused on the region between the X inactivated gene *RBM10*, and the downstream escape gene *UBA1* in Xp11.23. Analysis of human ESTs using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) identified multiple *UBA1* untranslated exons that each splice to exon 2 containing the ATG start codon (e.g. BG703038, BI461515, AW501935, BG036961, and AW178197) (Figure 3-1). The two major splice variants, based on EST abundance, are exons 1a and 1d, and are annotated in genome databases within *UBA1* transcript variant 1 (NM_153280.2) and 2 (NM_003334.3). Exons 1a and 1d have also been described previously as Ex1 and Ex1' (Goto et al. 2009).

To refine the location and size of the expression boundary between X inactivated and expressed transcripts, the X inactivation status of individual *UBA1* isoforms was analyzed by RT-

Figure 3-1. 5' end of the human *UBA1* locus from exon 1a through exon 5. The untranslated exons (black) splice to exon 2 containing the ATG start codon and the downstream translated exons (gray). Arrows indicate the direction of transcription.

PCR in mouse-human somatic cell hybrids retaining a single human active or inactive X chromosome (Carrel et al. 1996) (Figure 3-2). Hybrid cells retain only one human X chromosome, and enable the direct measurement of a gene's X inactivation status without having to distinguish inactive from active X transcripts. Further, hybrids recapitulate the XCI expression status of the vast majority of X-linked genes in normal human cells (Carrel et al. 2005). Therefore the XCI status of all human X-linked genes expressed in hybrid cells can be determined. To ensure amplification from spliced transcripts, cDNA from hybrid cell lines was amplified with human-specific primers, one exon specific primer and a primer spanning exons 2 and 3. Transcripts for four of the five splice variants were detected in active X hybrid lines; absence of exon 1c transcripts indicated a more tissue-restricted expression pattern. Only expression from the two downstream-most exons, 1d and 1e, was observed in four independent hybrid lines containing an inactive X, indicating that transcripts initiating at these two exons escape X inactivation. Therefore the transition from X inactivated to transcripts that escape lies in the region between exons 1b and 1d.



Figue 3-2. Expression of human *UBA1* isoforms in mouse-human somatic cell hybrids. Negative image of ethidium bromide-stained RT-PCR products for *UBA1* transcripts. Human (H) and Mouse (M) control lanes establish that primers specifically amplify a human product. Expression for each variant was tested in two independent active (Xa) and four independent inactive (Xi) X hybrids. Exon 1c was not expressed in the somatic cell hybrids and therefore could not be assayed. Data provided by Matt Blome.

To confirm and extend these data, allelic expression was tested in primary cell lines that are non-randomly X inactivated (Carrel et al. 1999). Unlike normal fibroblasts which randomly inactivate one of the two Xs, leading to a mixed population of cells with inactive Xs derived from both parental alleles, populations of non-randomly inactivated cells have inactive Xs derived from a single parental allele. By comparison of transcript abundance between the two alleles utilizing informative SNPs, the expression of genes on the inactive X can be measured. Monoallelic expression of SNP rs5945431, that measures transcript abundance of the X inactivated *PLXNA3* locus establishes that the population of inactive Xs is derived from a single parental allele in three informative lines and validates the assay (SNP A, Fig 3-3). Two SNPs within human *UBA1* were assayed to confirm the X inactivation expression differences of the human splice variants observed in hybrid cells. Monoallelic expression of SNP rs41310655, that

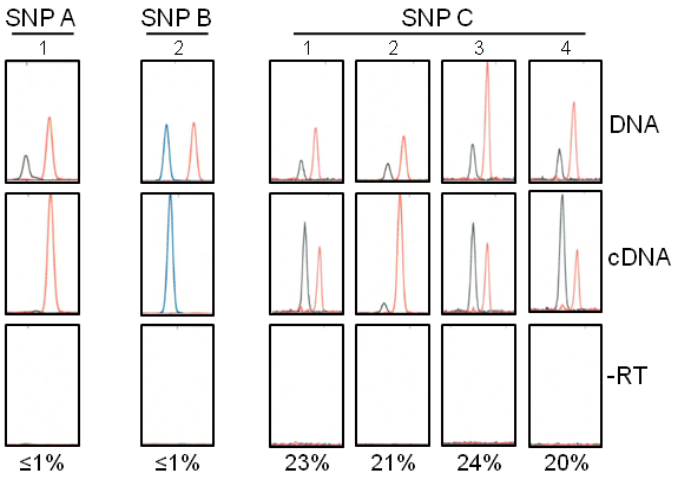


Figure 3-3. X inactivation status of select loci on the human X chromosome. Allele specific expression for *PLXNA3* (SNP A, rs5945431) and *UBA1* (SNP B, rs41310655; and SNP C, rs2070169) in primary non-randomly inactivated human female cell lines. The level of Xi expression relative to Xa expression is indicated at the bottom, and was calculated as the mean of the allele ratios in cDNA samples after normalization to genomic DNA in at least three replicates. Cell lines tested: (1) GM01695, (2) GM02621, (3) GM03322, and (4) GM00135. SNP A was also assayed in the informative lines 3 and 4 with similar results to cell line 1 (data not shown). The difference in the peak height ratio for SNP C in cell line 2 as compared to the other lines is a result of the G allele (black) corresponding to the inactive X in cell line 2, whereas the G allele is the active X in the other three cell lines.

measures primary transcripts from exons 1a and 1b, verifies both are subject to X inactivation (SNP B, Fig 3-3). Yet, biallelic expression of a downstream SNP, rs2070169, included in all *UBA1* transcripts indicated some *UBA1* transcripts escape inactivation (SNP C, Fig 3-3). That both SNPs show different X inactivation patterns in the doubly-informative line GM02621 confirms that individual *UBA1* splice variants have different X inactivation states. Further, reduced expression from the inactive X relative to active X levels (Xi/Xa=22% ± 2%) is consistent with inactive X expression from only a subset of alternatively spliced exons and previous data indicating most escape genes are only partially expressed from the inactive X chromosome (Carrel et al. 2005). Taken together, these data reveal novel *UBA1* exon-specific X inactivation patterns that narrow the boundary between *UBA1* transcripts subject to X inactivation and those that escape to a 2.1 kb region between *UBA1* exons 1b and 1d. This inactive X-specific expression boundary, within human *UBA1*, is the smallest such boundary that has been identified on the either the mouse or human X chromosome and is therefore ideal to dissect sequences regulating inactivation X expression.

Conserved Gene Structure, But Not X Inactivation Status Of Mouse Uba1

Similar to human *UBA1*, the orthologous gene in mouse, *Uba1*, initiates transcription at multiple sites and has four alternative untranslated exons 5' of the coding sequence (Figure 3-4). EST analysis suggests *Uba1* lacks the 1e splice variant and accordingly this region poorly aligns with the human locus. Both exons 1a and 1d share greater than 70% identity between mouse and human and by EST prevalence are the most abundant. Human exon 1b shares 59% identity with the orthologous mouse sequence; however the mouse *Uba1* exon 1b, as defined by EST data, is shorter than the human exon. Mouse ESTs define an additional alternative exon between exons 1b and 1d, although this exon 1c does not align to human exon 1c.

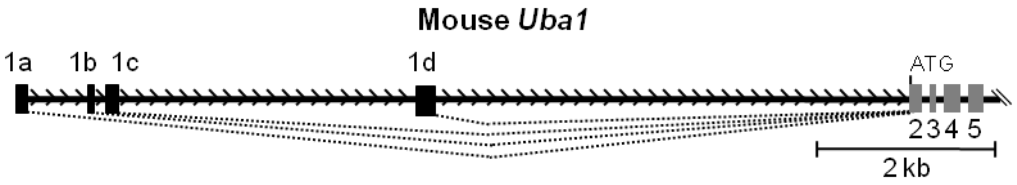


Figure 3-4. 5' end of the mouse *Uba1* locus from exon 1a through exon 5. The untranslated exons (black) splice to exon 2 containing the ATG start codon and the downstream translated exons (gray). Arrows indicate the direction of transcription.

To confirm previous reports that mouse *Uba1* is X inactivated (Disteche et al. 1992; Yang et al. 2010), we assayed an exon 9 SNP, rs29651331, in the mouse B119 cell line. Similar to the non-randomly X inactivated cell lines utilized in the human expression studies, the B119 line is also non-randomly inactivated, such that the *M. castaneous* X chromosome is inactivated in all cells. Therefore, to verify the parental origin of the alleles, genomic DNA from *M. castaneous* and C3H mouse strains was compared to the B119 line. Despite similar gene structure at the *UBA1* locus between human and mouse, monoallelic expression of SNP rs29651331 confirmed all splice variants expressed in the mouse fibroblast line B119 are subject to X inactivation (Figure 3-5). Furthermore, a SNP within exon 1d, analyzed as part of a recent global X inactivation survey in mouse, directly demonstrates that transcripts initiating at this exon are X inactivated (Yang et al. 2010). These data clearly establish that mouse *Uba1*, unlike the human locus, is X inactivated, and therefore suggests comparative analysis between mouse and human should differentiate *UBA1* sequences regulating the complex alternative splicing at this locus from those that specify the human escape domain.

Sequence Analysis Identifies Sequences Unique To The Human UBA1 Escape Domain

Much of the region between exon 1a and 1d is highly similar between human and mouse (Figure 3-6). The exons corresponding to the major splice variants as defined by ESTs, exon 1a and exon 1d, are highly conserved between species. The X inactivated sequences of human

UBA1 are highly similar to mouse, and despite their X inactivation differences, human exon 1d is 84% identical to that of mouse. High sequence identity (>76%) extends 500 bp upstream of the exon 1d transcription start site, suggesting that most promoter elements at exon 1d are likely conserved as well. While the exons and adjacent sequences are similar, human sequences within the X inactivation boundary between inactivated exon 1b and escape exon 1d align poorly with mouse. In humans, much of this boundary is repetitive, including a small AT-rich domain and short-interspersed repetitive elements (SINEs). The poorly aligning mouse sequence is also highly repetitive, comprised of a large (~800 bp) CT-rich repeat and SINE elements. In addition to repetitive sequences, the human *UBA1* expression boundary contains a 250 bp stretch of single-copy sequence (Fig. 3-6). The placement of this human-specific single-copy sequence within the *UBA1* expression boundary makes this sequence an attractive candidate for regulating the human-specific escape domain.

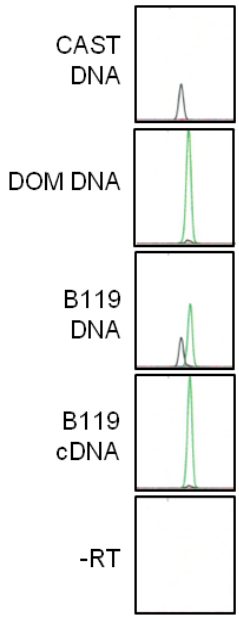


Figure 3-5. Allele-specific expression analysis of mouse *Uba1*. Allelic expression at the exonic SNP rs29651331 in non-randomly X inactivated B119 cell line. The B119 cell line contains an inactive M. castaneous allele and active C3H derived allele. M. castaneous (CAST) and C3H (DOM) genomic DNA were used to mark the parental origin of the SNP in the B119 cell line. B119 DNA confirms the presence of both alleles.

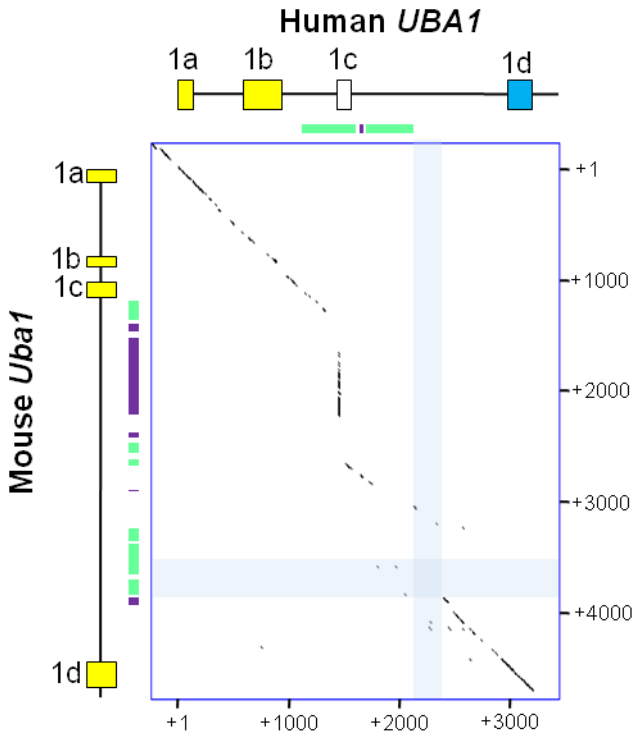


Figure 3-6. Nucleic acid dot plot sequence comparison of human *UBA1* (x-axis) and mouse *Uba1* (y-axis) (19 bp window, 3 mismatch). Human (ChrX:47,049,988-47,053,664) and mouse (ChrX:20,235,322-20,240,333) sequences for the 5' untranslated region of *UBA1* are shown from exon 1a through exon 1d. The X inactivation state of each exon in shown (silenced – yellow; expressed – blue). Repeat elements, SINE (green) and low-complexity (purple) repeats, in *UBA1/Uba1* are identified. The 250 bp of human, single copy sequence that aligns poorly with mouse is highlighted in blue.

Repetitive Elements Are Unlikely To Explain Inactive X Expression At The UBA1 Locus

Bioinformatic analyses of the human inactive X have identified sequence elements that correlate with silenced regions and those that correlate with escape domains (Carrel et al. 2006; Wang et al. 2006) on the human inactive X. L1 sequences are enriched on the human X chromosome (29%) when compared to the genome as a whole (17%) (Ross et al. 2005). Notably, the L1 content is not distributed evenly along the X chromosome, and instead many X inactivated regions are enriched for L1 elements as compared to escape regions (Bailey et al. 2000; Carrel et al. 2006; Wang et al. 2006). To determine whether L1 elements could be regulating expression at the *UBA1* locus, repetitive sequences within the human escape domain were compared to

sequences in mouse to identify differences that may account for the X inactivation of mouse *Uba1*. There is a marked localized decrease in L1 density in the ~1.25 Mb region encompassing the human *UBA1* escape domain compared to surrounding sequences despite that the entire 4 Mb region is lower than the X chromosomal average (Figure 3-7). This reduction in L1 density begins upstream of *UBA1*, within the X inactivated *RBM10* gene, and the density remains low throughout the *UBA1* 5' UTR and downstream escape genes, similar to the reduced L1 density at other human escape domains (Bailey et al. 2000; Carrel et al. 2006; Wang et al. 2006), suggesting low L1 density may contribute to escape. However, reduced L1 density alone does not determine escape as sequences surrounding mouse *Uba1* are also reduced in L1s (Figure 3-7 B). Intriguingly, the L1 content of the human locus is comprised entirely of interrupted L1s, while the mouse sequence contains two full-length, intact L1s within in the 4 Mb region surrounding *Uba1* (Figure 3-7 B). The presence of the these large L1 elements, located near the adjacent X inactivated genes, contributes to the large peaks of L1 density in these regions and the increased variability in L1 density throughout the 4 Mb region in mouse, and may influence silencing of

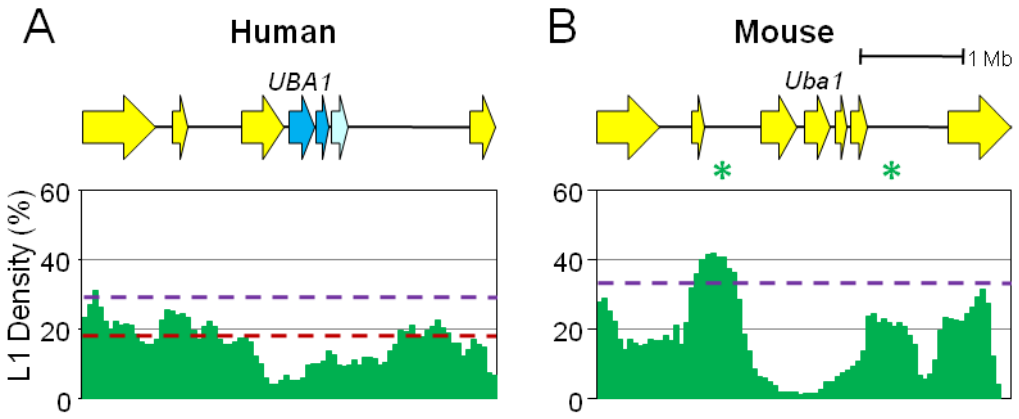


Figure 3-7. L1 repetitive element distribution in human and mouse in the 4 Mb region surrounding the *UBA1*/*Uba1* locus. (A) Human (ChrX:46,850,199–47,250,198). (B) Mouse (ChrX:20,035,547-20,435,546). L1 repeat density calculated using a sliding window analysis (50,000 bp bin, 5000 bp slide) within the 4 Mb window (green) is shown relative to the genome wide average (red dotted line) and X chromosome average (purple dotted line) (Ross et al. 2005). Full length, intact L1 elements are shown (asterisks).

mouse *Uba1* and the surrounding genes. Active L1 elements are reported to facilitate silencing of at least one gene the mouse X chromosome (Chow et al. 2010), and could be acting in a similar manner at the mouse *Uba1* locus.

Unlike L1s, Alu repeats are primate specific SINEs enriched at human escape domains (Wang et al. 2006). To investigate whether Alu sequences could regulate expression at the human *UBA1* locus, the density of Alu repeats and their murine equivalent, B1s, was analyzed in the 4 Mb surrounding *UBA1* similar to the L1s. The Alu repeat density throughout the analyzed region is considerably higher than the B1 density (Figure 3-8). However, this discrepancy in B1 density as compared to Alu density is a characteristic of the X chromosome as a whole (B1s<1%, Alus~8%). Intriguingly, both Alu and B1 repeats are more abundant in the 4 Mb window surrounding the *UBA1* locus than their respective chromosome averages. Further, the B1 density is relatively consistent throughout the throughout the region, albeit higher than the chromosomal average (Figure 3-8 B), while the region immediately surrounding human *UBA1* and *RBM10* is depleted of Alu elements compared to the surrounding genes (Figure 3-8 A). This relatively low

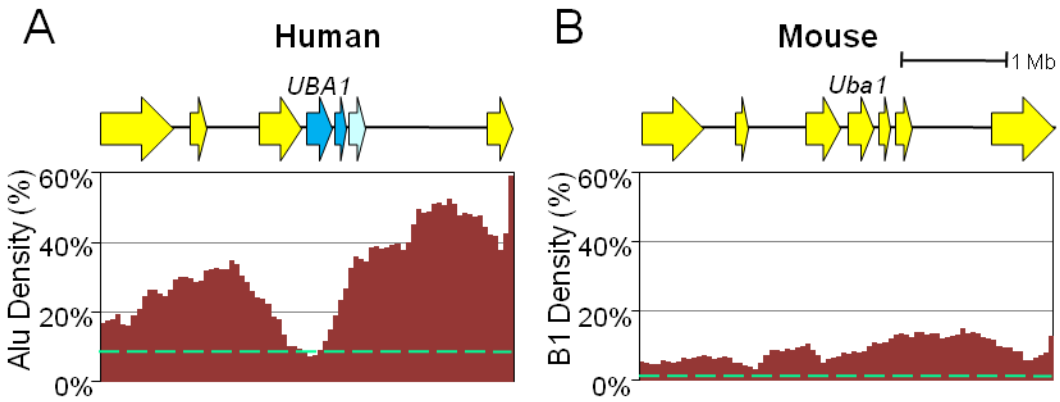


Figure 3-8. SINE repetitive element distribution in human and mouse in the 4 Mb region surrounding the *UBA1*/*Uba1* locus. (A) Human Alu repeat distribution (ChrX: 46,850,199–47,250,198). (B) Mouse B1 repeat distribution (ChrX:20,035,547-20,435,546). Repeat density (red) was calculated using a sliding window analysis (50,000 bp bin; 5,000 bp slide) and is shown relative to the average X chromosome density (green dotted line).

density of Alu repeats at human *UBA1* is similar to the chromosomal average and contradictory to previous reports of Alu enrichment at escape domains (Wang et al. 2006) suggesting Alus are dispensable for escape. From these analyses of the repetitive content near the human *UBA1* escape domain, it is clear that neither L1 nor Alu distribution has the specificity required to determine the X inactivation status of individual *UBA1* isoforms.

Identification Of An Insulator At The Human UBA1 Expression Boundary

Since human *UBA1* escape cannot be regulated by repetitive sequences alone, we explored whether another prominent model of inactive X gene regulation that positions CTCF-bound insulators at escape gene boundaries (Filippova et al. 2005) could account for the inactive X expression of *UBA1*. To this end, sequences at the *UBA1* locus were tested for insulator activity using the well-established enhancer blocking assay (Chung et al. 1993; Bell et al. 1999). This assay identifies insulators as sequences that impede enhancer function specifically when placed between the enhancer and reporter gene. In particular, candidate sequences are inserted into a vector containing an enhancer-driven, neomycin reporter gene. The effect of candidate insulator sequences on neomycin activity was measured by counting G418-resistant colonies after the stable integration of the enhancer-blocking construct into mammalian cells.

We initially tested the CpG island surrounding *UBA1*/*Uba1* exon 1d, as an insulator within a CpG island is proposed to explain X inactivation escape of *Jarid1c* in mouse (Filippova et al. 2005). Enhancer blocking was detected for both human and mouse CpG island fragments but was seen regardless of fragment insertion location, either upstream or downstream of the enhancer (fragments 9,C; Fig. 3-9 A-B). Such position-independent activity is not characteristic of an insulator. Rather, this activity may instead specify a repressor, and was not completely unexpected as this CpG island sequence includes the exon 1d promoter and adjacent sequences.

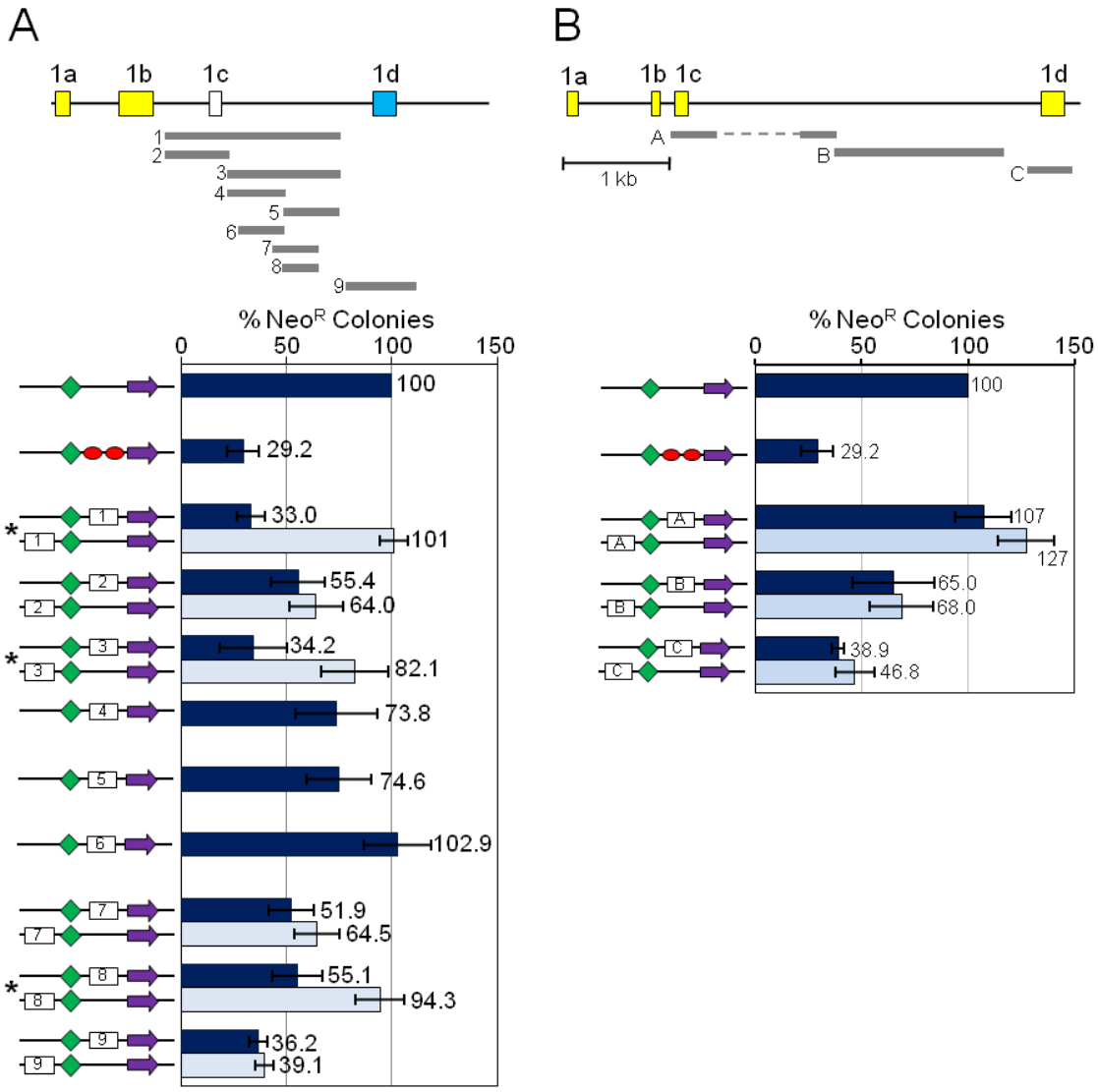


Figure 3-9. Enhancer blocking activity at the *UBA1*/*Uba1* locus. (A) Human. (B) Mouse. *UBA1* exons are drawn to scale and fragments assayed are indicated (numbered gray bars). The assay measures neomycin gene (purple arrow) expression that requires activation by the enhancer (green diamond). This interaction can be disrupted by insulators (e.g. 1.2 kb chicken β-globin insulator (red oval)) in a position-dependent manner. Potential insulator sequences (numbered rectangles) were assayed 5' and 3' of the enhancer. Fragments that were not assayed 3' of the enhancer did not show a statistically significant decrease in colony formation as compared to the uninsulated control. Error bars indicate standard deviation of the mean, calculated from multiple transfections of at least two independently derived test sequences. The dotted line in fragment A indicates a CT-rich simple repeat that was not cloned. *Denotes constructs with a statistically significant decrease in colony formation between fragments placed 3' and 5' of the enhancer, $p<0.005$.

Intriguingly, the enhancer-blocking assay did identify strong position-dependent insulator activity at the human *UBA1* locus using a 1.7 kb fragment that spans nearly the entire region

between exons 1b and 1d (fragment 1, Fig 3-9 A). This activity could be narrowed to a 1.1 kb fragment (fragment 3, Fig. 3-9 A) with a strong response quite similar to the well-characterized 1.2 kb chicken β-globin insulator (Fig. 3-9 A, (Bell et al. 1999; Filippova et al. 2005)). Furthermore, using a series of overlapping *UBA1* fragments, we mapped the majority of position-dependent enhancer-blocking activity to a 334 bp fragment (fragment 8, Fig. 3-9 A). Notably, this fragment corresponds to the single-copy sequence that fails to align with mouse *Uba1* sequences. Although this small 334 bp insulator fragment is less efficient than larger segments, a 250 bp insulator segment within the chicken β-globin insulator responds quite similarly (Bell et al. 1999) and suggests that clustered insulator fragments or additional sequences are required for optimal enhancer-blocking activity (Chung et al. 1997; Bell et al. 1999). Insulation at the *UBA1* locus appears to be human-specific, as all mouse *Uba1* sequences tested failed to detect position-dependent enhancer-blocking activity, including fragment B which aligns to the human insulator fragment 3 (Fig. 3-9 B). The striking placement of a human-specific insulator sequence in the transition region between X inactivated and expressed transcripts of *UBA1* is consistent with a role in proper inactive X gene regulation.

CTCF Binding Is Not Inactive X Specific

A strong candidate to confer enhancer-blocking activity at *UBA1* is CTCF. The major vertebrate insulator protein, CTCF confers the enhancer-blocking ability of the chicken β-globin insulator (Bell et al. 1999) and binds the insulator fragment 5’ of *Jarid1c* (Filippova et al. 2005). Furthermore, a recent study identified some 13,000 CTCF binding sites genome-wide, and localized binding at one site to an ~2 kb region at the 5’ end of the human *UBA1* locus (Kim et al. 2007). A conserved CTCF consensus site within the 1d CpG island of *UBA1* likely contributes to the observed binding; however due to the large size of the chromatin fragments used in the study, contributions from additional nearby sites could not be entirely excluded. To refine the location

of CTCF binding within the 5' UTR of *UBA1*, chromatin immunoprecipitation (ChIP) of primary cells was performed. The close proximity of the insulator sequence to exon 1d (516 bp), required that chromatin was sheared to an average size of ~150-300 bp in order to distinguish binding at the consensus site from nearby sequences (Figure 3-10). Because chromatin fragments were quite small, high-resolution analysis of the region was possible, and unlike previous studies (Kim et al. 2007), binding at the insulator and other sequences near the exon 1d consensus site were independently assessed. CTCF binding at the insulator sequence was distinguished from binding at exon 1d by measuring CTCF binding at closely spaced regions throughout the 5' UTR. Since chromatin was randomly sheared prior to immunoprecipitation, CTCF binding at either the insulator or CpG island should be specifically enriched at these loci, although some larger fragments may be detected when adjacent loci are tested, resulting in a narrow Gaussian distribution of CTCF enrichment, centered at the binding site. On the other hand, binding at multiple sites is predicted to result in a broad distribution of binding encompassing all sites, and centered between them. Further, CTCF binding specific to the inactive X can inferred from enrichment differences between male lines, containing a single active X and that of females with both an active and inactive X.

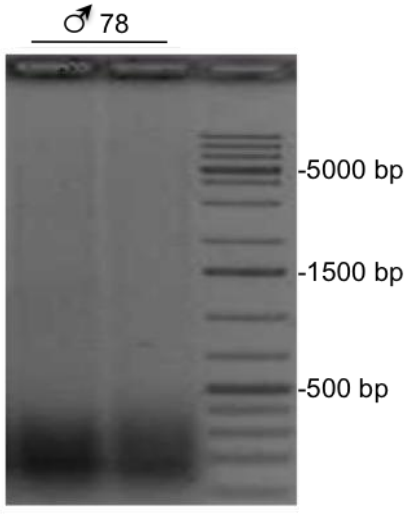


Figure 3-10. Example of chromatin fragment size used in ChIP experiments. Negative image of ethidium-bromide stained human male 78 DNA fragments purified immediately following formaldehyde crosslinking, sonication to shear the chromatin, and reversal of crosslinking to separate protein from DNA.

CTCF enrichment at human *UBA1* is remarkably similar in male and two female primary fibroblast lines; a large narrow peak surrounds exon 1d (Figure 3-11 A). The placement of the peak centered on exon 1d is consistent with CTCF binding the previously identified 20-mer consensus site within *UBA1* exon 1d (Kim et al. 2007), and with recent ChIP-seq experiments (ENCODE consortium project (McDaniell et al. 2010)). Importantly, when the CTCF binding at *UBA1* was compared to other known autosomal CTCF binding sites (Burke et al. 2005), enrichment at *UBA1* was found to be greater than or equal to the enrichment at a known binding site at *H19* (Figure 3-12 A). Furthermore, binding at *UBA1* is significantly greater than the *UBA1* negative control locus and a region near *H19* that does not bind CTCF (Figure 3-12 A), supporting the conclusion that CTCF robustly binds *UBA1* at the consensus site in exon 1d. That

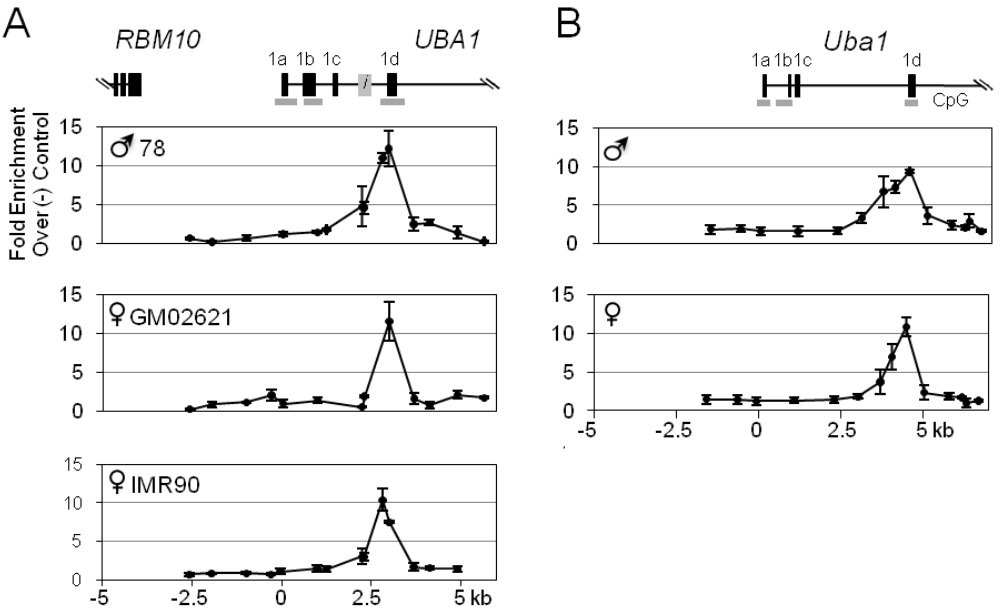


Figure 3-11. CTCF binding at the *UBA1* locus. (A) Human (Chrx:47,045,199-47,056,199) (B) Mouse (ChrX:20,230,530-20,242,530). CTCF enrichment is shown as the fold enrichment over an experimentally determined negative binding control region upstream of *UBA1*. The above results represent data from one of the triplicate experiments performed for each cell line. Error bars represent the standard deviation of the mean enrichment for the technical triplicates in the experiment. The *UBA1*/*Uba1* exons (black) and the insulator (gray) are drawn to scale and show the location of CTCF binding within the 5' UTR.

binding was similar between male and female cell lines, suggests CTCF binds the *UBA1* expressed exon 1d on both the active and inactive X chromosomes.

To evaluate the species specificity of CTCF binding at *UBA1*, we assayed similarly prepared chromatin from male and female mouse fibroblasts. CTCF enrichment in mouse was similar to that of human *UBA1*; a single peak surrounds the *Uba1* exon 1d in male and female cells (Figure 3-11 B). Further, comparison of CTCF binding at *Uba1* to a known binding site near *Igf2r* (Figure 3-12 B), confirmed CTCF robustly binds mouse *Uba1* similar to that seen at human *UBA1* and suggests that CTCF binds the *UBA1* consensus site at exon 1d in a species-independent manner. Furthermore, differences in the relative enrichment levels of CTCF between species and between males and females, are consistent with a lack of CTCF at the X

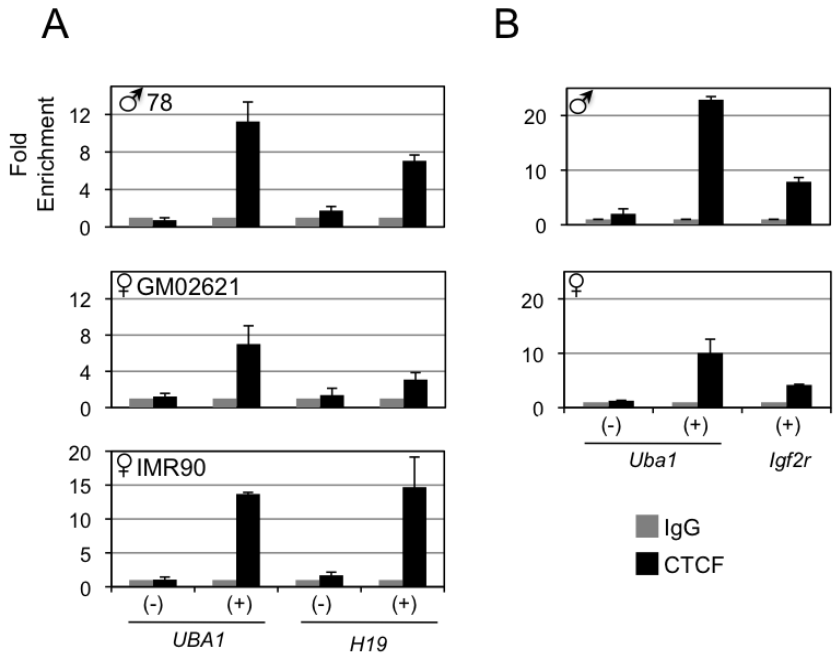


Figure 3-12. Comparison of CTCF binding at *UBA1* locus to control loci. (A) Human. (B) Mouse. Enrichment value at each site is shown as fold enrichment over normal rabbit IgG for one of the triplicate experiments performed in each cell line. Error bars indicate standard deviation of the mean for technical triplicates. *UBA1* loci were experimentally determined: positive (+) binding loci correspond to the exon 1d consensus site, negative (-) loci correspond to the negative binding region used to normalize *UBA1* CTCF data. Human *H19* (+) and (-) CTCF binding sites, and *Igf2r* (+) binding sites were previously identified (Burke et al. 2005; McNeil et al. 2006).

inactivated 1d exon in female mouse and with variation in the total chromatin assayed in each cell line. Therefore, CTCF likely binds mouse *Uba1* only at the expressed exon 1d on the active X chromosome.

The small size of the chromatin fragments utilized in this assay allowed CTCF binding at sequences near the *UBA1* exon 1d consensus site to be assessed. Intriguingly, the presence of a single narrow peak of CTCF enrichment was observed at human *UBA1*. The insulator and other sequences near the exon 1d consensus site were markedly reduced in CTCF enrichment. This combined with the fact that peak enrichment was centered on the consensus site supports the conclusion that CTCF binds only the exon 1d consensus site within the 5' UTR of human *UBA1*. Furthermore, this data demonstrates that the large localized region of CTCF binding at the 5' UTR of *UBA1* reported previously (Kim et al. 2007) results from binding only at the exon 1d consensus site in the large chromatin fragments used in the study. The lack of CTCF binding at the insulator sequence establishes that the *UBA1* insulator sequence is one of a few insulators that exhibit enhancer-blocking ability without directly binding CTCF (Kim et al. 2003; Magdinier et al. 2004). Furthermore, that CTCF fails to bind the human *UBA1* insulator, and binds *UBA1* in a sex- and species-independent manner, suggests binding is unrelated to the establishment of the human escape domain. This, instead, supports a gene-specific regulatory role for CTCF.

Histone Modifications At UBA1 Delimit The Insulator Sequence

Distinct histone modifications are known to demarcate active from inactive chromatin throughout the genome (Kouzarides 2007). To determine whether histone modifications distinguish the closely juxtaposed X inactivated and expressed sequences at human *UBA1*, high resolution mapping of acetylated histone H3 (H3Ac) and trimethylated histone H3 lysine 27 (H3K27me3) enrichment was performed. Similar to CTCF, ChIP was used to evaluate histone modifications across the *UBA1* expression boundary. Furthermore, this analysis of the *UBA1* 5'

UTR will characterize the histone modifications that define the *UBA1* inactive X expression boundary.

To definitively position specific histone modifications on the active or inactive X chromosomes, allele-specific ChIP was performed utilizing SNP rs41310655 in *UBA1* that maps between exons 1b and 1c. Allele-specific expression studies demonstrated that the G allele is the active X in the informative cell line GM02621, while the T allele is the inactive X (Figure 3-3). By measuring the abundance of each allele in the immunoprecipitated GM02621 chromatin as compared to genomic DNA, the enrichment on the active and inactive X was determined in a manner similar to that of the expression assays. Importantly, this approach established that the relative representation of each allele in input chromatin is identical to genomic DNA (Figure 3-13). Furthermore, as expected, greater than 90% of the H3Ac enrichment near exon 1b was from the active X, while H3K27me3 enrichment was primarily from the inactive X (Figure 3-13). These data establish that the lack of a modification at a particular site is not due to underrepresentation of chromatin from the active or inactive X chromosomes and validate the ChIP methods used to map histone modifications across the *UBA1* expression boundary.

H3Ac, known to mark active chromatin domains (Kouzarides 2007), is heavily enriched at two sites in the 5' UTR of *UBA1* on the active X in males; one large H3Ac peak surrounds exon 1d, and a second broad peak encompasses exons 1a and 1b (Figure 3-14 A). Perhaps not

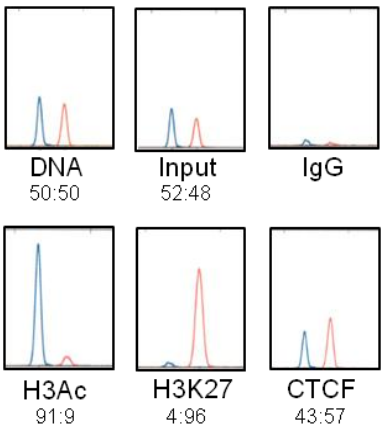


Figure 3-13. Allele-specific ChIP analysis of GM02621 DNA at human *UBA1* (SNP 41310655). The ratio of Xa:Xi enrichment is indicated below each panel and was calculated by comparing the allele ratio in each ChIP sample after normalization to genomic DNA in at least three replicates.

surprisingly, the H3Ac binding corresponds to the expressed exons and with EST abundance for transcripts originating at these exons. Enrichment of H3Ac at *UBA1* in females is similar to that of males; two prominent peaks correspond to the expressed exons (Figure 3-14 A). Intriguingly, female IMR90 shows increased H3Ac binding at exon 1a-1b compared to female GM02621 and male 78 (Figure 3-14 A). However, as *UBA1* escapes inactivation in female GM02621 which lacks this H3Ac increase, this additional H3Ac binding is not required for escape, and therefore likely results from differences in alternative exon usage between cell lines. Similar to human *UBA1*, H3Ac enrichment in both male and female mouse was most prominent at exon 1d, although binding at exons 1a-1b was evident as well (Figure 3-14 B). The similar binding profiles at the *UBA1* locus between species suggest H3Ac enrichment at the expressed exons is a

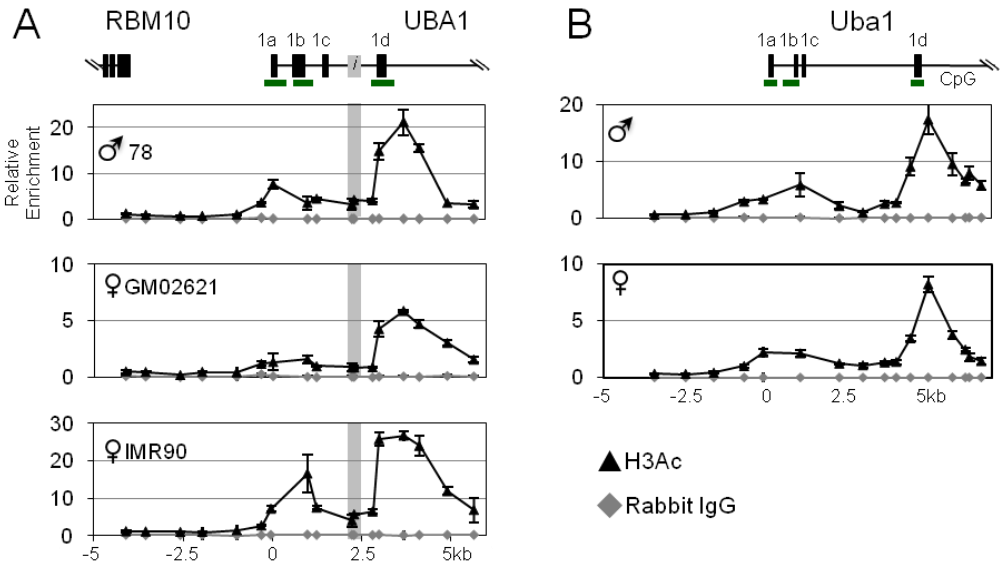


Figure 3-14. Histone H3Ac enrichment at the *UBA1* locus. (A) Human (Chrx:47,045,199-47,056,199). (B) Mouse (ChrX:20,230,530-20,242,530). Relative enrichment is shown as the calculated enrichment compared to input DNA. The relative enrichment of normal rabbit IgG is shown for comparison. The above results represent data from one of the triplicate experiments performed for each cell line. Error bars represent the standard deviation of the mean enrichment for the technical triplicates in the experiment. The *UBA1*/*Uba1* exons (black), CpG islands (green) and the insulator (gray) are drawn to scale.

gene-specific characteristic of the *UBA1* locus. That H3Ac is absent between exons 1b and 1d in humans combined with the similar binding profile of mouse *Uba1* suggests H3Ac does not distinguish the human inactive X expression boundary, and instead distinguishes expressed regions of the *UBA1* locus.

H3K27me3, known to mark inactive domains (Kouzarides 2007), was largely absent throughout the entire human *UBA1* region tested in male (Figure 3-15). This result is consistent with *UBA1* expression from the active X. In contrast, H3K27me3 enrichment was quite prominent at X inactivated sequences in female humans, yet the mark was absent at *UBA1* sequences that are expressed from both Xs (Figure 3-15). That H3K27me3 associates specifically with inactive *UBA1* sequences is consistent with previous reports that H3K27 marks X

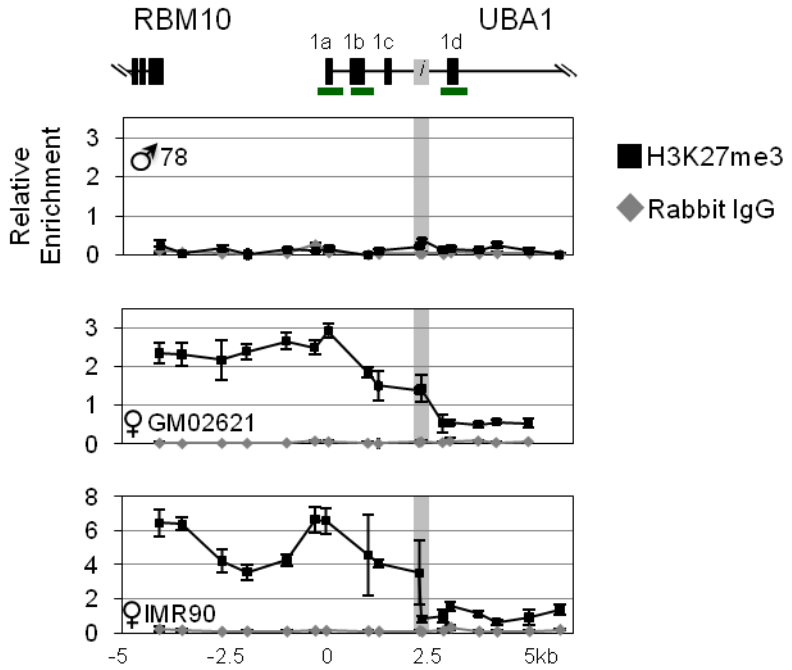


Figure 3-15. Histone H3K27me3 enrichment at human *UBA1* (Chrx:47,045,199-47,056,199). Relative enrichment is shown as the calculated enrichment compared to input DNA. The relative enrichment of normal rabbit IgG is shown for comparison. The above results represent data from one of the triplicate experiments performed for each cell line. Error bars represent the standard deviation of the mean enrichment for the technical triplicates in the experiment. The *UBA1*/*Uba1* exons (black), CpG islands (green) and the insulator (gray) are drawn to scale.

inactivated sequences (Brinkman et al. 2006). Intriguingly, H3K27me3 was absent at all male and female mouse *Uba1* sequences (data not shown). Despite this difference in the H3K27me3 distribution between species at the locus, the placement of the human *UBA1* insulator sequence at the boundary of H3K27me3 enrichment in females is intriguing and further supports a role for this sequence as a candidate inactive X regulatory element. Combined, these data position the insulator at the boundary of active and inactive chromatin demarcated by histone modifications. This intriguing placement is consistent with a functional role for the insulator as a barrier between active and inactive chromatin domains on the inactive X.

Identification Of A Hypomethylated Region Specific To The Human UBA1 Insulator Sequence

Methylation of cytosine residues to form 5-methylcytosine normally occurs at cytosines that are adjacent to guanine residues termed CpGs. One of the most well-known forms of DNA methylation occurs at gene promoters and is associated with heterochromatin and inhibition of gene expression (reviewed in (Klose et al. 2006)). DNA methylation has been implicated in the regulation of diverse cellular processes (reviewed in (Bernstein et al. 2007)), including the regulation of imprinted regions by the preferential methylation of one parentally derived allele. This raises the possibility that a differentially methylated domain (DMD) at *UBA1*, similar to that observed at the *H19/IGF2* imprinted locus, could allow regulatory factors to distinguish the active from inactive X based on methylation differences. The presence of DMDs at the *UBA1* locus has been reported (Goto et al. 2009); however, these inactive X hypermethylated/active X hypomethylated domains are located some distance from the expression boundary and the human-specific insulator sequence identified in this study. To determine whether a DMD coincides with the human *UBA1* insulator we evaluated the methylation state of the 154 CpGs across 3.7 kb, surrounding the *UBA1* expression boundary in primary human cell lines. The methylation state

of CpGs was determined by deamination of non-methylated cytosines to form uracil by bisulfate treatment of the DNA (Hayatsu et al. 1970; Shapiro et al. 1974). Sequencing of treated DNA subsequently distinguished methylated from non-methylated residues. Methylation on the inactive X was inferred from differences between males, having a single active X, and females, having both an active and inactive X. Further, methylation on only one X in females is not expected to exceed 50%, due to representation of both active and inactive Xs.

Methylation at the human *UBA1* CpG islands for exons 1a, 1b, and 1d parallel the known expression patterns. Males have little to no methylation throughout the CpG islands and exonic regions of exons 1a, 1b, and 1d (Figure 3-16 A), as predicted for the 5' ends of expressed genes on the active X chromosome. In contrast, four female lines were more methylated at the X inactivated exons 1a and 1b, though the percent methylation did not exceed 50% suggesting this methylation is on only the inactive X (Figure 3-16 A). Further, females lack methylation at the inactive X expressed exon 1d, similar to males. To more completely evaluate the role of methylation at the *UBA1* locus methylation analysis was extended beyond exon boundaries. Intriguingly, the methylation pattern of the inter-exonic region of the *UBA1* inactive X expression boundary, between exons 1b and 1d, has a distinctive signature on the human active and inactive Xs. Non-CpGs island sites across the region are heavily methylated on the active X in males, yet are only partially methylated in female fibroblasts (Figure 3-16 A). This pattern is especially evident at the insulator sequence although there is some variation in the size of the partially methylated region between the individual fibroblast lines tested (Figure 3-16 A; 3-17). These data support the recent finding that gene bodies are hypermethylated on the active X in both males and females (Hellman et al. 2007). The placement of the DMD on the inactive X and within the expression boundary of *UBA1* is intriguing and suggests hypomethylation of the inactive X within *UBA1* may have a functional role in specifying the inactive X expression boundary.

To evaluate whether the identified DMD at the insulator is specific for the human locus, we measured methylation at 102 of 103 total CpG sites at the orthologous region in mouse *Uba1*. Similar to human, mouse *Uba1* CpG island methylation patterns closely predict the known expression profiles. *Uba1* exons are unmethylated on the active X in males, while the X inactivated exons (1a, 1b and 1d) are hemimethylated in female (Figure 3-16 B). Intriguingly, the entire inter-exonic region between exon 1b and 1d is heavily methylated in both male and female lines (Figure 3-16 B). This hypermethylation of the inter-exonic region in mouse suggests the DMD in human *UBA1* is unique to the inactive X expression boundary. That this DMD demarcates the *UBA1* insulator on the human inactive X supports a functional role for this sequence, perhaps as a platform allowing methylation-sensitive factors to identify and bind specifically to the inactive X.

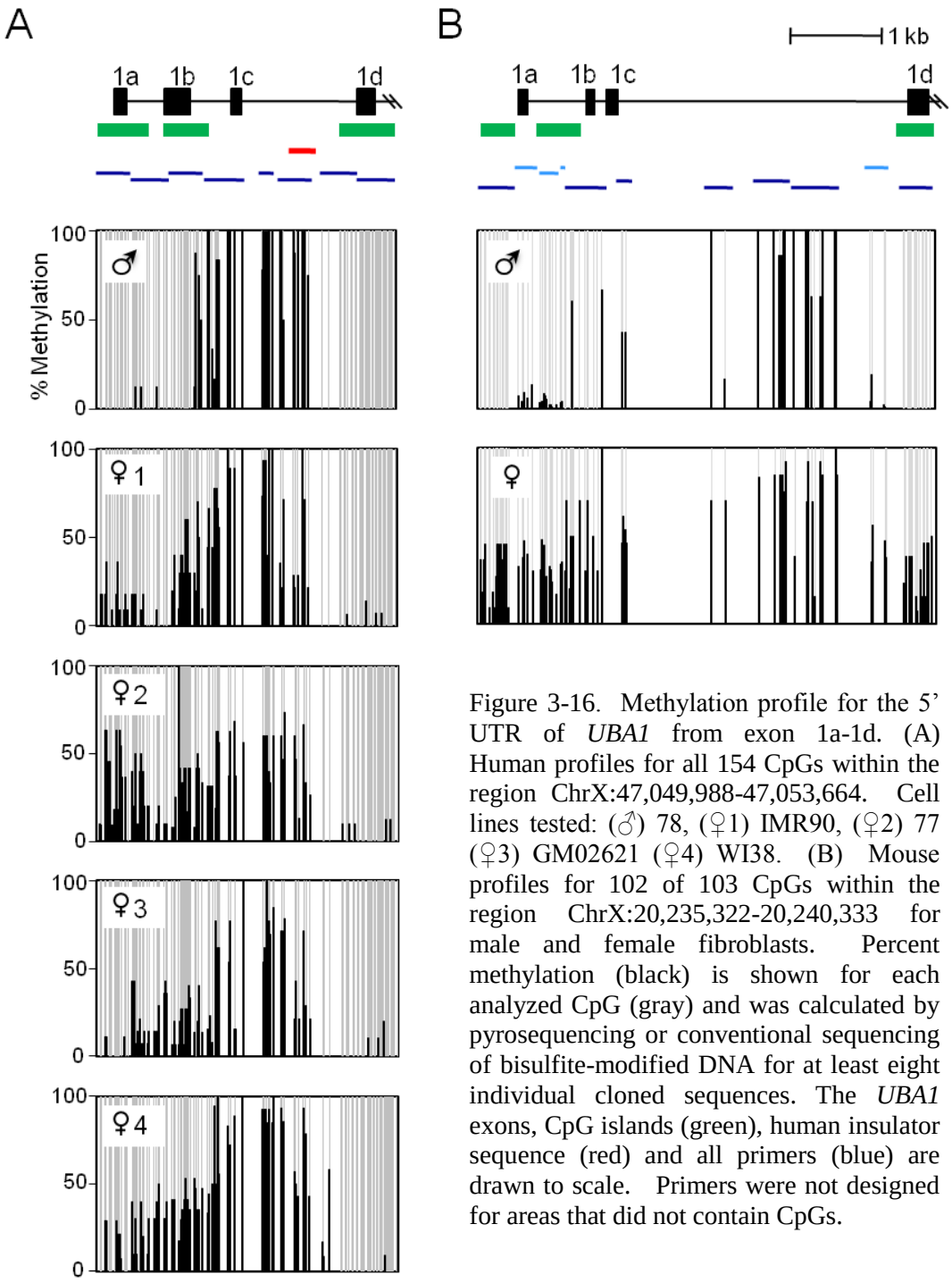


Figure 3-16. Methylation profile for the 5' UTR of *UBA1* from exon 1a-1d. (A) Human profiles for all 154 CpGs within the region ChrX:47,049,988-47,053,664. Cell lines tested: (♂) 78, (♀1) IMR90, (♀2) 77 (♀3) GM02621 (♀4) WI38. (B) Mouse profiles for 102 of 103 CpGs within the region ChrX:20,235,322-20,240,333 for male and female fibroblasts. Percent methylation (black) is shown for each analyzed CpG (gray) and was calculated by pyrosequencing or conventional sequencing of bisulfite-modified DNA for at least eight individual cloned sequences. The *UBA1* exons, CpG islands (green), human insulator sequence (red) and all primers (blue) are drawn to scale. Primers were not designed for areas that did not contain CpGs.

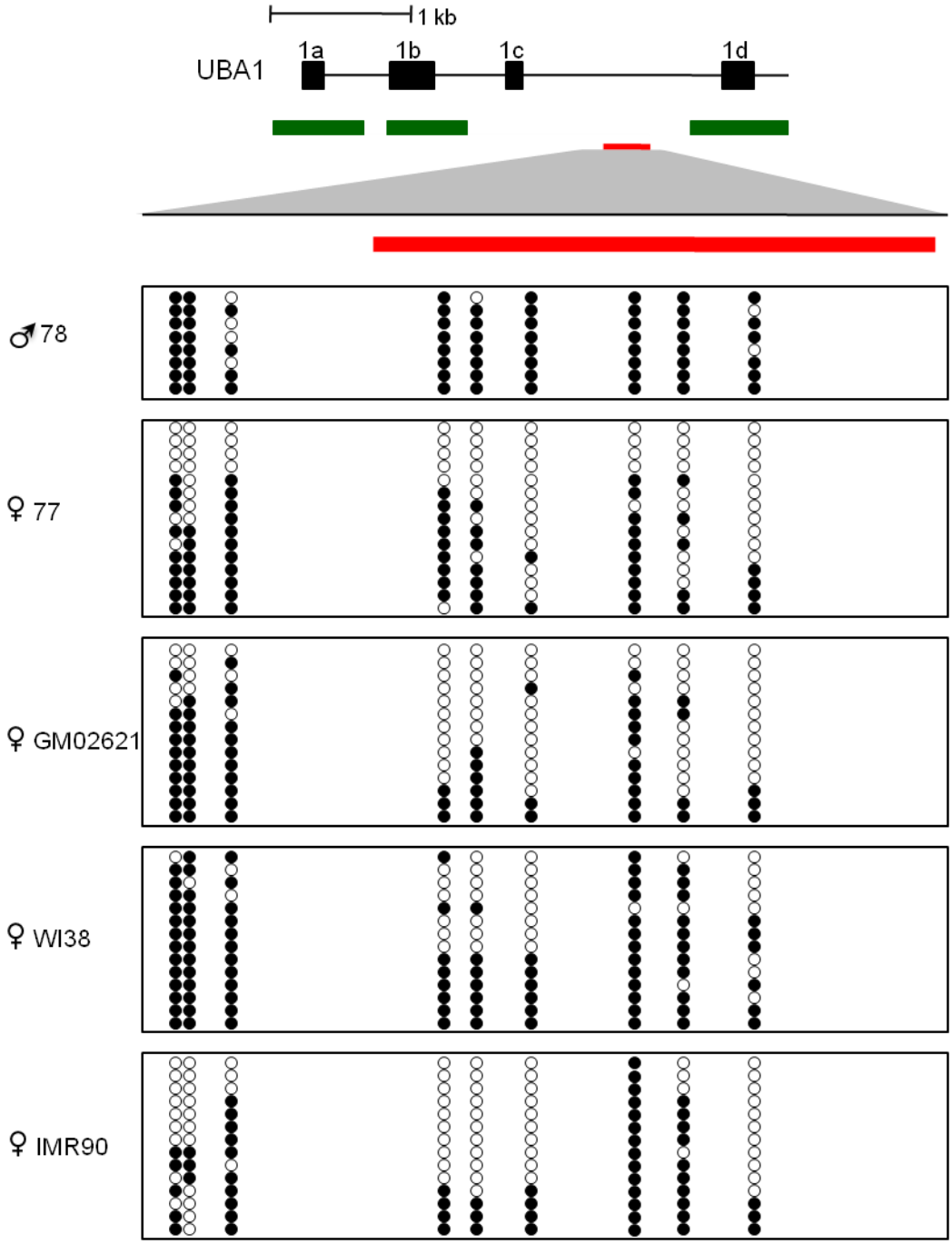


Figure 3-17. Methylation pattern of individual cloned sequences of all CpGs near the human *UBA1* insulator. Exons, CpG islands (green), and the insulator (red) are shown for comparision. The 475 bp region (ChrX:47,052,217-47,052,691) surrounding the insulator (red) is enlarged and scaled accordingly. Methylation (black) at all CpG sites (circles) within the enlarged region for male and females lines is shown.

Chapter 4

DISCUSSION AND FUTURE DIRECTIONS

In the studies presented in the previous chapter we characterized the 5' expression boundary of the *UBA1* escape domain on the inactive X chromosome in order to identify elements that regulate XCI escape domains. We predicted that the regulation of XCI escape domains occurs at the boundary between X inactivated and expressed genes to prevent heterochromatin spreading and erroneous silencing of escape genes. In the course of analyzing the *UBA1/RBM10* expression boundary, we identified novel human *UBA1* exon-specific X inactivation patterns that refined this expression boundary to a region within the 5' UTR of *UBA1* (Figure 3-2 and Fig 4-1). In support of our hypothesis that escape is regulated at the boundaries of escape domains, we identified an insulator sequence, capable of enhancer-blocking activity, that is specific to the human escape domain and is positioned between X inactivated and expressed domains as defined by expression analysis and epigenetic modifications. Importantly, comparisons of the human *UBA1* expression boundary and the *UBA1* escape domain to the orthologous X inactivated region in mouse allowed us to address the validity of the two prominent models of XCI gene regulation at the *UBA1* locus (Figure 1-4, model i&ii). Local L1 density and/or specific L1 elements (Lyon 1998; Chow et al. 2010) and the presence of chromatin insulators (Filippova et al. 2005) have been postulated to influence XCI gene regulation. However, none of these elements have been functionally confirmed to regulate XCI gene expression. Therefore, the requirement of specific sequence elements for XCI gene regulation can be addressed by determining whether these elements are conserved at the *UBA1* escape locus.

.

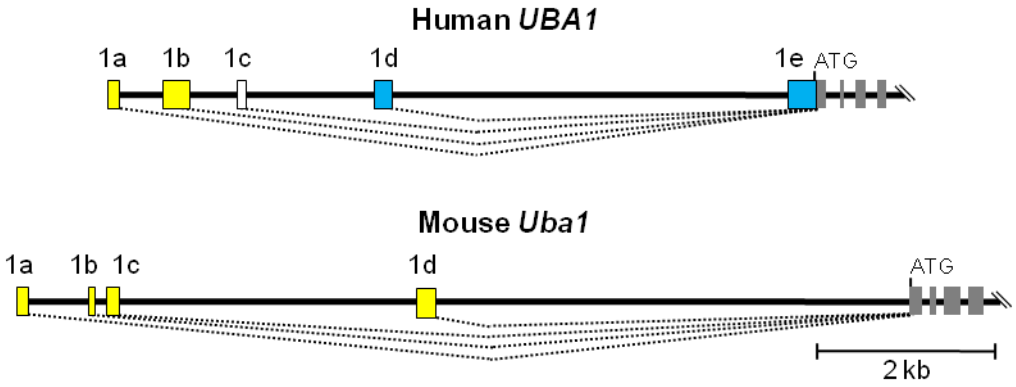


Figure 4-1. Expression of the 5' untranslated exons from the inactive X of the *UBA1* locus. The two most upstream exons are silenced (yellow), while the two downstream exons escape (blue) in human *UBA1*. Human *UBA1* exon 1c was not expressed from either X in the hybrid cells and therefore could not be analyzed (white). All untranslated exons are silenced in mouse *Uba1*.

Repetitive Sequence Content And XCI Escape

Since the initial observations of L1 enrichment on the X chromosome relative to autosomes using cytological methods (Korenberg et al. 1988; Boyle et al. 1990) and the discovery that silencing of autosomal sequences by ectopic *Xic/Xist* sequences is more efficient in regions of high L1 density (Rastan 1983; Lee et al. 1997; White et al. 1998), L1 elements have been postulated to promote the spread of the XCI silencing (Lyon 1998). However, the mechanism(s) by which L1s act, either as "way stations" that propagate the XCI signal or through the repeat-induced formation of heterochromatin, are just beginning to be queried (Lyon 1998; Chow et al. 2010). To date, no physical interaction between L1 elements and *Xist* RNA has been found that directly links L1s to the *Xist*-mediated silencing. However, a novel architectural role for L1 sequences that may influence silencing was proposed with the finding that these repeats are spatially restricted, early in XCI, within the Barr body; regions of high L1 density are found within the *Xist* compartment and are efficiently silenced, while regions of low density remain outside (Chow et al. 2010). Despite the mounting circumstantial evidence that supports a role for L1 density in XCI gene regulation, without functional confirmation, the possibility remains that

the distribution of L1s is simply a remnant of the unique evolution of the X chromosome (Boissinot et al. 2001).

To address the role of L1 density on XCI regulation at the *UBA1* locus, we compared L1 density within the human escape domain to the orthologous X inactivated sequences in mouse. Here we showed that the human *UBA1* escape domain is indeed depleted of L1s compared to the surrounding region (Figure 3-7). However, low density L1 is not restricted to the escape domain, but spreads into the upstream X inactivated gene, *RBM10*. These data are consistent with previous hypotheses that low L1 density creates an escape prone region (Chow et al. 2010), yet L1 distribution at *UBA1* does not have the specificity required to determine the X inactivation status of individual *UBA1* isoforms. Intriguingly, the X inactivated mouse *Uba1* locus is also strikingly deplete of L1s. So then why does mouse *Uba1* not escape X inactivation? That answer may come from the distribution of intact L1 elements near the *UBA1* region. One model of L1 facilitated silencing suggests that full-length active L1 elements can silence genes by driving antisense transcription across X inactivated genes that lie adjacent to escape domains (Chow et al. 2010). Extension of this model to the mouse *Uba1* locus predicts that the two full-length L1s located nearby may facilitate silencing of the region (Figure 3-7). LINE-driven antisense transcription, at the onset of XCI, across the mouse *Uba1* region could efficiently silence the region despite the overall low L1 density. The low L1 density at human *UBA1*, combined with a lack of full-length L1s is predicted to prevent efficient silencing in this region resulting in the *UBA1* escape domain. Further, the low L1 density of the human *UBA1* region coupled with its positioning outside of the *Xist* nuclear compartment (Chaumeil et al. 2006; Murakami et al. 2009) supports the proposed role of L1s in the nuclear localization of X inactivated domains to the interior of the *Xist* compartment (Chow et al. 2010). Therefore, we conclude that the local L1 environment may contribute to XCI gene regulation at the *UBA1/Uba1* locus, but lacks the specificity required to regulate expression of specific human *UBA1* isoforms.

In addition to L1s, the distribution of other repetitive sequences has been correlated with X inactivated or expressed regions on the inactive X (Bailey et al. 2000; Ke et al. 2003; McNeil et al. 2006; Wang et al. 2006). One such sequence is the Alu repetitive element that is consistently enriched within regions on the human X that escape XCI (Wang et al. 2006). While no functional evidence has been found that supports a role for Alus in XCI gene regulation, we examined their distribution to determine whether the observed Alu enrichment at escape genes extends to the *UBA1* escape domain. We found that both fragmented and full-length Alu repeats are indeed enriched throughout much of the region surrounding *UBA1*, although the *UBA1* gene itself is strikingly sparse (Figure 3-8). Although relatively low compared to the surrounding sequences, the density of Alu sequences at *UBA1* is similar to the X chromosome average as a whole. Genomic sequences surrounding mouse *Uba1* are enriched in B1s, the murine Alu equivalent. That human *UBA1* escapes inactivation, yet contains relatively few Alu sequences, suggests enrichment of Alu repeats is not required for escape. However, one full-length Alu repeat was identified within the expression boundary between X inactivated and expressed *UBA1* transcripts. Could this Alu, positioned within the human *UBA1* escape boundary, influence escape? Potentially, although two additional full-length Alus are located upstream of the X inactivated *UBA1* exon 1a suggesting Alu placement at the boundary is not a unique feature of the region. Without functional evidence that Alu elements, positioned at the expression boundaries of escape genes, regulate expression we are left to rely on similarities between escape domains to identify sequences involved in XCI gene regulation. As human *UBA1* lacks the Alu enrichment found at other escape loci these elements are likely a dispensable feature of XCI escape.

The XCI regulation of the *UBA1/Uba1* locus cannot fully be explained by the distribution of two prominent repetitive sequences that have previously been reported to correlate with XCI gene expression. Neither L1 nor Alu repeats are enriched in such a manner to suggest that they are able to effectively regulate the XCI expression of human *UBA1* isoforms in such close

proximity to X inactivated sequences. Therefore, additional repetitive elements or other regulatory sequences must be required at the human *UBA1* locus to prevent erroneous silencing.

Boundary Elements And XCI Escape

Chromatin insulators are one type of DNA boundary element that is capable of establishing boundaries between transcriptionally permissive and transcriptionally repressive chromatin domains. As such, insulators are appealing candidates to regulate XCI gene expression, particularly at closely juxtaposed loci of opposite XCI status, by preventing the spread of heterochromatin and gene silencing into escape domains (Figure 1-4 A model ii). Further, chromatin insulators have been identified in multiple species and they regulate a variety of genes (reviewed in (West et al. 2002)). While the mechanisms by which insulators act can vary, the overall effect is to protect domains from the inappropriate action of enhancers and/or silencers (enhancer-blocking function) or from position effects caused by proximity to inactive condensed chromatin (barrier function) (reviewed in (West et al. 2002)). That insulators may function similarly on the inactive X chromosome to protect escape domains from silencing is an enticing prospect, supported by the identification of a chromatin insulator within the expression boundary adjacent to the mouse escapee *Jarid1c* (Filippova et al. 2005).

The *Jarid1c* insulator and many other chromatin insulators, including the human *UBA1* insulator identified in this study, have been identified using functional transgene assays. Although there are several limitations to using these assays, such as the enhancer blocking assay, to identify insulator sequences (reviewed in (Maston et al. 2006)), they remain a widely used, and a largely accurate tool for analyzing insulator function. Perhaps the largest caveat is that the *in vitro* activity of a given sequence, in the context of a reporter transgene or plasmid, may not recapitulate events at the endogenous loci. In one such case, several independent studies established the enhancer-blocking ability of one insulator element that binds the suppressor of

hairy-wing, or Su(Hw), insulator protein in *D. melanogaster* (Golovnin et al. 2003; Parnell et al. 2003; Soshnev et al. 2008). However, further analysis revealed that this particular Su(Hw) sequence element did not function as an insulator as demonstrated in the reporter assays, but, instead, as a transcriptional activator at the endogenous locus (Soshnev et al. 2008). While this example emphasizes the importance of sequence context in determining the role of specific regulatory elements, the fact remains that the large majority of sequences identified as insulators using *in vitro* assays have been subsequently confirmed to function similarly at the endogenous locus (reviewed in (Raab et al. 2010)). The developmental timing of XCI makes identifying bona fide insulators in the context of the inactive X difficult since no model system has been developed that allows the manipulation of human X-linked sequences at their endogenous location in the context of XCI silencing. Functional transgene based insulator assays are therefore an important first step in identifying sequences with unusual regulatory properties. Our identification of a chromatin insulator that is specific to the human *UBA1* escape domain using a functional transgene assay does not conclusively establish that in the context of the X chromosome this sequence functions as an insulator. However, our assay demonstrates that the *UBA1* insulator sequence is capable of blocking an enhancer in a position-dependent manner when integrated into the human genome. Further, that the vast majority of insulators that have been identified using functional transgene or plasmid based assays and later analyzed for insulator function at the endogenous loci function in a similar manner at both (reviewed in (Raab et al. 2010)). Therefore, we conclude that the *UBA1* insulator sequence, with enhancer-blocking ability, likely functions as an insulator at the endogenous locus as well.

Four additional lines of evidence argue that the *UBA1* insulator may play an important role in the regulation of the *UBA1* escape domain. First, only human *UBA1* sequences are functional in the enhancer-blocking assay. Mouse *Uba1* sequences, that are X inactivated upon XCI, lack enhancer-blocking ability. Second, the human *UBA1* insulator is located within a

region that aligns poorly with mouse *Uba1* and therefore is positioned within a region that is not required for the shared gene-specific functions of *UBA1* between mouse and human. Third, epigenetic modifications delimit the insulator sequence at endogenous *UBA1* on the inactive X chromosome. Histone modifications that are characteristic of heterochromatin are enriched upstream of the insulator (Figure 3-15), while modifications that characterize euchromatin are enriched downstream of the insulator (Figure 3-14). Further, a differentially methylated domain surrounds the insulator sequence; the active X is hypermethylated and the inactive X is hypomethylated (Figure 3-16 and Fig 3-17). Lastly, the *UBA1* insulator is positioned within the 5' expression boundary of the *UBA1* escape domain placing it at an ideal location to protect the boundary of the escape domain from the encroachment of the adjacent heterochromatin consistent with the boundary model of XCI gene regulation. While this evidence strongly supports a function for the insulator in the regulation of the *UBA1* escape domain, the underlying mechanism(s) remains to be seen.

Following predictions of the boundary model of XCI gene regulation, chromatin insulators are expected to flank escape domains to prevent heterochromatin spreading. This prediction is true for at least two escape domains that have been analyzed. A CTCF-dependent insulator was identified at the XCI expression boundary of *Jarid1c* in mouse (Filippova et al. 2005) and our studies have identified an insulator at the human *UBA1* expression boundary. That now two chromatin insulators are positioned at XCI expression boundaries suggests their placement is not random, and strongly supports a role for insulators in the regulation of escape domains. Additionally, binding of the CTCF insulator protein was also demonstrated within the expression boundary of mouse *Eif2s3x* and human *EIF2S3* (Filippova et al. 2005) in further support of the model. An extension of the boundary model predicts that domains lacking insulators may be unable to escape as they would be subject to the silencing associated with

heterochromatin spread. This is observed for the orthologous *Uba1* region in mouse that lacks an insulator sequence and is subject to XCI.

Although both the *Jarid1c* (Filippova et al. 2005) and our *UBA1* insulators were identified by their enhancer-blocking ability, the underlying regulatory mechanisms at the endogenous loci is still unknown. The placement of the *UBA1* insulator at the boundary of a euchromatic and heterochromatic domain is reminiscent of the location of the chicken β-globin insulator which sits between a region of heterochromatin and the developmentally regulated β-globin gene cluster (Bell et al. 1999). By actively recruiting histone acetyltransferases to the region, the chicken β-globin insulator prevents heterochromatin from spreading and silencing the β-globin genes (Recillas-Targa et al. 2002; Huang et al. 2007). In fact, the chicken β-globin insulator sequence itself is enriched for acetylated H3/H4 (Huang et al. 2007). Here we show that human *UBA1*, unlike the chicken β−globin insulator, contains a region of H3Ac immediately downstream of the insulator sequence. Comparisons of male and female H3Ac profiles clearly establish that much of the enrichment at the region is due to active X contributions, but we expect that the inactive X is also is enriched for H3Ac at the expressed *UBA1* exon 1d and our data are consistent with this expectation. The difference in AcH3 enrichment at the *UBA1* insulator as compared to the β-globin insulator is intriguing and supports two possible regulatory mechanisms for the insulator. Either the *UBA1* insulator functions to recruit H3Ac only during XCI initiation when the boundary of the escape domain is first established or the *UBA1* insulator uses an alternative mechanism to maintain the domain that does not lead to AcH3 enrichment at the insulator sequence. If the *UBA1* insulator functions only during the establishment of the escape domain, it is not surprising that the sequence is not enriched for AcH3 in the post-XCI cell lines that were analyzed in this study. Additional data supporting a role for the *UBA1* insulator in the establishment of the escape domain, but not its maintenance, come from genome-wide studies of

DNaseI hypersensitivity. In contrast to many of known insulator sequences that are hypersensitive to DNaseI treatment (Chung et al. 1997; Antes et al. 2001; Komura et al. 2007; Blackledge et al. 2009) the *UBA1* insulator sequence is not hypersensitive in the post-XCI cell lines analyzed (Birney et al. 2007; Boyle et al. 2008). The lack of AcH3 and DNaseI hypersensitivity at the *UBA1* insulator suggests that this insulator sequence may not function in post-XCI cell lines and instead may be required only for the establishment of escape domains upon XCI initiation. However, that the *UBA1* insulator is epigenetically distinguished by methylation differences on the active and inactive X in post-XCI cells supports a functional role in maintaining the boundaries of the escape domains for the sequence. Further analysis is necessary to determine underlying mechanisms of *UBA1* insulator sequence and whether this sequence functions during the XCI initiation and in the maintenance of the silenced state of the inactive X.

The barrier model of XCI gene regulation predicts that chromatin insulators, capable of regulating XCI escape domains, would exhibit barrier function. However, it is possible that enhancer-blocking insulators could also regulate XCI. At least some insulators that function in an enhancer-blocking capacity have been shown to interact with other insulators or tether chromatin to structural elements forming ordered chromatin structures (reviewed in (Gaszner et al. 2006)). Therefore, the *UBA1* insulator may function as an enhancer-blocking insulator in functional assays, but could potentially form highly-ordered structures that would segregate the heterochromatin and euchromatin domains at *UBA1* thereby establishing the active chromatin state of the escape domain. Whether the *UBA1* insulator continues to organize the chromatin of the escape domain after its establishment, in the absence of DNaseI hypersensitivity, remains to be seen.

Unlike the vast majority of chromatin insulators which are capable of blocking an enhancer in a CTCF-dependent manner (reviewed in (Phillips et al. 2009)), the *UBA1* insulator

does not bind CTCF. In fact, CTCF binds outside of the expression boundary in a sex-independent manner at a conserved consensus site within the human *UBA1* exon 1d (Figure 3-11 and (McDaniell et al. 2010)). That CTCF binding is similar in mouse *Uba1*, which lacks insulator sequences; it is likely that CTCF serves an alternative function at the *UBA1* locus. This is not surprising as CTCF is a ubiquitously expressed, multifunctional protein with known roles as a transcriptional activator, repressor and insulator (reviewed in (Phillips et al. 2009)). Indeed, the X chromosome alone contains over 300 experimentally identified binding sites, and although some of these sites do correlate with inactive X expression boundaries, others occur within X inactivated and escape domains alike (Kim et al. 2007). The large number and diverse locations of CTCF binding sites suggest CTCF serves multiple functions on the X chromosome. Intriguingly, many autosomal CTCF binding sites segregate alternative promoters within a single gene, and may specify proper promoter usage within these genes (Kim et al. 2007). Given this observation, it is plausible that binding of CTCF at *UBA1* could regulate proper usage of the *UBA1* alternative promoters in a gene-specific manner and would explain the similar binding that occurs in mouse and human. However, functional analysis is necessary to determine whether CTCF regulates promoter usage at the *UBA1* locus.

In light of our identification of a CTCF binding site near the *UBA1* expression boundary that lacks the binding specificity required of XCI gene regulators (i.e. inactive X specific binding), we must now question the function of other similarly placed CTCF binding sites. For example, CTCF binding has been observed near the *Eif2s3x/EIF2S3* escape genes in mouse and human and the placement of these sites is consistent with a role in escape gene regulation as a chromatin insulator (Filippova et al. 2005). However, this assumption, based on location alone, may be premature and functional analysis of the CTCF binding is necessary to exclude a gene-specific regulatory role. However, it is also important to note that even at human *UBA1* where CTCF is likely binding both the active and inactive X chromosomes, an additional inactive X

specific function for CTCF at the *UBA1* locus cannot be completely excluded. One possibility is that the *UBA1* insulator, functioning in an inactive X specific manner, could interact with CTCF in *cis* and form an ordered chromatin structure that regulates the XCI escape domain. By acting in combination with the downstream CTCF binding site, the *UBA1* insulator could recruit chromatin remodeling complexes downstream of the insulator sequence which would explain the downstream AcH3 enrichment. To date, there is no experimental evidence that supports an interaction between the *UBA1* insulator and the downstream CTCF binding site, and any relationship will need to be addressed.

An alternative hypothesis, based on the observed lack of acetyl H3 and DNaseI hypersensitivity, is that the *UBA1* insulator sequence, functions as a human-specific enhancer at the endogenous locus. As an enhancer, the *UBA1* insulator may activate the downstream *UBA1* promoters allowing the expression of specific *UBA1* isoforms from the inactive X chromosome in female humans. However, there is no experimental evidence that supports an interaction between the *UBA1* insulator and downstream sequences. Further, the enhancer-associated protein, p300, does not bind the *UBA1* insulator sequence (ENCODE Consortium data (Birney et al. 2007)) suggesting that this sequence does not function as an enhancer. Instead, we hypothesize that the *UBA1* insulator, functions as an insulator at the endogenous locus to successfully prevent the improper silencing of *UBA1* isoforms from the inactive X chromosome upon X inactivation.

It is also important to note that while the CTCF binding data for the primary human cell lines presented here completely agrees with recent CTCF ChIP-seq experiments in a large number of both primary and immortalized human cell lines (ENCODE Consortium data (Birney et al. 2007) (McDaniell et al. 2010)), a recent examination of CTCF binding in hybrid cells by Yuji Goto reported binding within in the *UBA1* expression boundary, quite close to exon 1b, exclusively on human inactive X (Goto et al. 2009). However, this study did not analyze binding at the downstream *UBA1* consensus site. Neither our experiments, nor any of the available ChIP-

seq data identified differential CTCF binding at an upstream *UBA1* site. Further, our allele-specific ChIP analysis (Figure 3-13) measured CTCF binding quite close to this reported site and established that the very low level of CTCF enrichment at this site is due to both active and inactive X contributions. Whether a human X chromosome in a hybrid cell is regulated differently or why there may be tissue- or cell line-specific differences in CTCF binding at this upstream locus remains unknown. But this clearly shows that the site is not responsible for boundary establishment in the cells tested in our study. Another difference between our study and the recently published study by Goto is that the differentially methylated domain that we identified in our female lines at the *UBA1* insulator sequence was not differentially methylated in hybrid cells (Goto et al. 2009). Although we observed hypomethylated CpGs within the *UBA1* insulator sequence in multiple female primary cell lines, there was variation between individuals in the overall methylation state of the expression boundary (Figure 3-16 and Fig 3-17). The lack of methylation in hybrids at the insulator sequence observed by Goto could be an extreme of the variation that we observed. Although a hypomethylated domain would be required for the binding of methylation-sensitive regulatory elements within the region, DNA methylation is not the only epigenetic difference between the active and inactive X chromosome at *UBA1*. Other features, such as the association with *Xist* and/or chromatin modifications may be essential for XCI regulatory factors to distinguish the two Xs. Alternatively, these methylation differences may suggest plasticity in XCI expression boundaries.

Our identification of an insulator at the XCI expression boundary of human *UBA1* and the finding that it is located between heterochromatin and euchromatin domain*s* strongly supports the conclusion that insulators are required for the escape of at least some domains. However, unlike the CTCF-dependent insulator located near mouse *Jarid1c*, the *UBA1* insulator does not bind CTCF. Although these insulators bind different factors, that at least two XCI expression boundaries and possibly a third, based on CTCF binding, suggest they are not located at XCI

expression boundaries by fortuitous chance. Yet these insulators do not bind the same insulator protein and therefore may not regulate expression in the same way. It was recently postulated that insulators evolved from a class of promoters that bind specific transcription factors to mediate a variety of forms of insulation (Raab et al. 2010). Thus, despite binding different proteins, CTCF-dependent and CTCF-independent insulators may still regulate escape domains in a similar manner on the inactive X. Many insulators function by forming highly ordered chromatin loops either by protein-protein interactions or by binding additional proteins that cluster in the nucleus to form insulator bodies (reviewed in (Bushey et al. 2008; Raab et al. 2010)). Whether multiple insulator types direct chromatin remodeling that sequesters escape domains outside of the *Xist* nuclear compartment and allows expression of these domains is an intriguing hypothesis and one that will need further investigation. Importantly, as insulators have now been identified at XCI expression boundaries in both mouse and human they likely represent a conserved mechanism between species to regulate escape domains. However, without functional conformation of the *UBA1* insulator, the possibility remains that this sequence does not regulate XCI gene expression. Further, other sequence elements that were not identified in our analysis of the *UBA1* expression boundary may function either in conjunction with or independently of this insulator sequence to regulate the boundary of the *UBA1* escape domain. Yet, our evidence supporting the inactive X specific function of the *UBA1* insulator at the XCI expression boundary, combined with the location of the *Jarid1c* insulator (Filippova et al. 2005) strongly supports a role for insulators to regulate escape domains. But, whether insulators are a universal regulatory mechanism of XCI escape or only regulate select domains remains to be seen.

Revised Model of XCI

The overall goal of the study of escape gene regulation presented here was to identify *cis*-acting regulatory elements at the boundary of escape domains that regulate XCI gene expression. In addition to identifying a novel CTCF-independent insulator involved in XCI, we also tested the two prominent models of gene-regulation on the inactive X: L1 repeat and CTCF-dependent insulator-facilitated escape at the *UBA1* locus. This analysis has added new insight to the mechanisms of XCI gene regulation. In summary, we propose a new model of XCI gene regulation where L1 repetitive elements and chromatin insulators function together to effectively silence the majority of genes on the inactive X, while at the same time protecting escape genes from erroneous silencing (Figure 4-2).

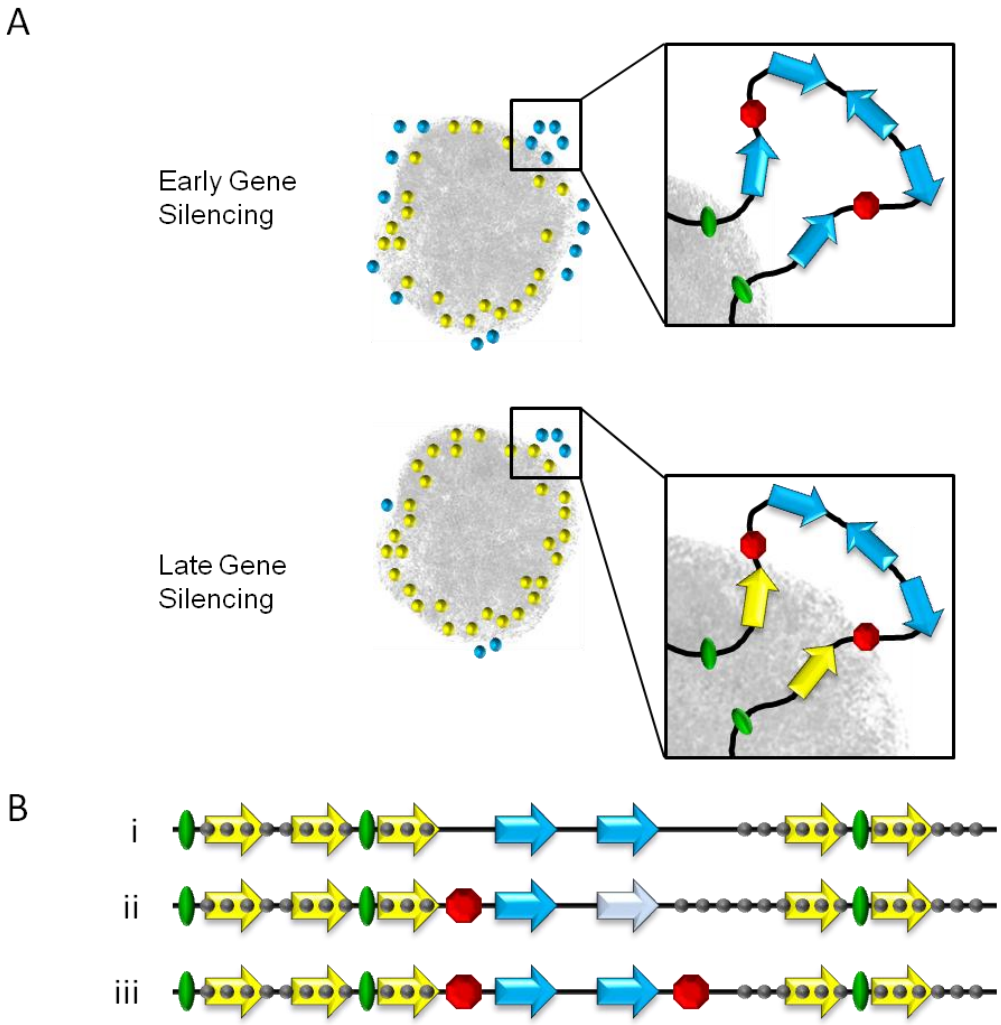


Figure 4-2. Revised model of XCI escape gene regulation. **(A)** Early in XCI regions with high L1 density (green oval) move into the *Xist* nuclear compartment (gray) either by propagation of XCI silencing by "way stations" (green ovals) or by repeat-induced heterochromatin formation. Genes that move inward are silenced (yellow), while genes associated with regions of low L1 density or lie far from "way stations" remain at the periphery of the *Xist* compartment and are expressed (blue). After this initial reorganization LINE driven antisense transription silences additional genes in these escape prone regions, while chromatin insulators (red octagon) protect escape domains from silencing. **(B)** Coordinated regulation of gene expression on the inactive X. **i)** Inactive X heterochromatin (gray circles) spreads along the X chromosome from "way stations" (green ovals) that propagate silencing, or after repeat-induced heterochromatin condensation and encompasses X inactivated genes (yellow). Some escape genes (blue) reside too far from "way stations" (green) or in regions with few repeats and are not silenced. **ii)** Chromatin insulators (red octagon) located on one side of an escape boundary effectively prevent the spread of heterochromatin from one side of the domain. Heterochromatin can spread into the domain from the uninsulated side and may lead to variable escape (light blue) of genes located at the edge of the escape domain. **iii)** Chromatin insulators flank escape domains preventing the spread of heterochromatin into the escape domain.

At the onset of XCI, the upregulation of *Xist* on the inactive X chromosome initiates silencing of the majority of genes on the X chromosome. As the *Xist*-mediated silencing signal spreads along the chromosome, regions of high L1 density condense and relocate to the interior of the *Xist* nuclear compartment (Chow et al. 2010). Whether this is a result of propagation of the silencing signal by "way stations" that recruit remodeling complexes to induce gene silencing or is caused directly by repeat-induced condensation of heterochromatin is not known. Upon this initial silencing, genes within regions of high L1 density move inward and are effectively silenced (Chaumeil et al. 2006; Chow et al. 2010). However, regions of low L1 density remain outside and create escape prone regions that are silenced late in the XCI process (Figure 4-2 A). At the *UBA1* locus, the escape prone region, as defined by low L1 density, consists of the X inactivated gene *RBM10* and the entire escape domain consisting of *UBA1*, *PCTK1* and *USP11*. This region is therefore expected to remain outside of the *Xist* compartment during this early silencing phase. To effectively silence genes within these escape prone regions, such as *RBM10*, additional regulatory factors are necessary that will assist in relocating these genes to the interior of the *Xist* nuclear compartment.

After the initiation of gene silencing, full length, intact L1s are expressed in an *Xist*-dependent manner and drive antisense transcription across genes that require silencing (Chow et al. 2010). These LINE-driven antisense transcripts remain localized to escape prone genes as they move into the *Xist* compartment and are silenced. Insulators, either CTCF-dependent or CTCF-independent, located within these escape prone regions protect certain genes from silencing (Figure 4-2 A). In humans the *UBA1* escape prone region lacks full length L1s and therefore may be expected to demonstrate variable silencing between individuals. However, the presence of an insulator sequence within the *UBA1* escape prone region prevents the spread of heterochromatin into the escape domain, such that only *RBM10* and the upstream-most *UBA1*

exons are silenced by the spread of heterochromatin into the escape prone region. In mouse, the orthologous region is X inactivated. This is likely a result of two intact L1 elements combined with a lack of insulator sequences. LINE-driven antisense transcription from the two L1 elements in the murine escape prone region could effectively silence the entire region in mouse.

In conclusion, localized L1 density, expressed LINE elements, and chromatin insulators all coordinately regulate escape on the inactive X chromosome (Figure 4-3 B, model iii). L1 repetitive sequences facilitate silencing, while chromatin insulators prevent silencing in certain regions. This model does not require active L1s or insulators for genes to escape silencing. Distance from L1 dense regions alone may be sufficient for some genes to escape without additional regulation (Figure 4-2 B model i). Likewise, escape domains may not require chromatin insulators to flank genes, although this may lead to variable silencing of some genes (Figure 4-2 B, model ii). Failure to spread the silencing signal effectively along the length of the chromosome may explain variable escape of some genes between individuals on the inactive X chromosome (Carrel et al. 2005).

Future Directions

From the studies of gene expression on the inactive X, including the analysis of the *UBA1* locus presented here, it is clear that XCI gene regulation is a complex process, involving multiple genomic and epigenetic mechanisms to insure proper gene silencing. The identification of an insulator at the human *UBA1* expression boundary not only supports a role for insulators in XCI regulation, but also raises the question of how widespread insulator sequences are on the X chromosome. Do insulators regulate the boundaries of all escape domains, or are they found at only select domains? Further, the *UBA1* insulator is one of only a handful of CTCF-independent insulators capable of enhancer-blocking activity in vertebrates (Kim et al. 2003; Magdinier et al. 2004) and relatively little is known about this subset of insulators. Therefore, the identification of

the *UBA1* insulator offers a unique opportunity to explore both XCI gene regulation by chromatin insulator sequences and regulation by CTCF-independent insulators in general.

The logical next experiment to characterize the insulator function would be to perform a set of deletion experiments of the insulator sequence at endogenous *UBA1* to determine its role in XCI gene regulation. However, there is no functional model system that has been successfully used to test the function of human sequences in the developmental context of XCI. Human ES cells may eventually provide a feasible model system to analyze sequences in their native context, but currently, the process of XCI has not been thoroughly evaluated in these cells to ensure that they recapitulate the *in vivo* process. Deletion experiments in somatic cell hybrids or adult human fibroblasts would establish whether the *UBA1* insulator is required for maintenance of the escape domain, but would not address whether this sequence is required for its establishment as we hypothesize. Transgene experiments in the mouse ES cell model system may be able to establish a requirement for the *UBA1* insulator sequence for both the establishment and maintenance of the escape domain, but again the ability of human sequences to escape in a mouse genetic background has not been tested. Therefore, the experiments proposed below do not use deletions to directly address the function of the *UBA1* insulator. Instead, experiments that indirectly assess the insulator function are proposed and will expand our understanding of insulators as well as provide important insights into XCI gene regulation.

Determination Of The Insulator Function(s) Of The *UBA1* Insulator

Despite the limitations described earlier in this chapter, functional reporter assays remain an accurate measure of endogenous insulator function at the majority of insulator sequences. We identified the *UBA1* insulator sequence by its ability to block an enhancer from activating a reporter gene. However, many insulators are capable of both blocking enhancer-promoter interactions and preventing the spread of heterochromatin, although these activities can be

separable functions (Recillas-Targa et al. 2002). The placement of the *UBA1* insulator between chromatin domains defined by the heterochromatic histone modification H3K27me3 and the euchromatic mark H3Ac is suggestive of an insulator that is capable of blocking the spread of heterochromatin. Insulator sequences positioned at XCI expression boundaries, including the boundary within *UBA1*, could prevent the spread of heterochromatin into the escape domains and avert erroneous silencing. However, we only tested the insulator for the ability to block enhancer-promoter interaction using an enhancer-blocking assay. Several reporter assays have been successfully used to test the ability of candidate insulator sequences to block the spread of heterochromatin. In one such assay, candidate insulator sequences are inserted into a vector such that they flank a reporter gene that is driven by a weak promoter/enhancer. Upon transfection, reporter activity is monitored over time to determine whether the reporter is silenced by position effects. Candidate sequences that successfully protect the reporter from position-effect silencing are identified as insulator sequences capable of blocking the spread of heterochromatin. Constructs similar to those used in my enhancer-blocking assays could be used for such an experiment with some modifications. Enhancer-blocking constructs, designed to test a candidate sequence's ability to block the enhancer in a position-independent manner, resemble constructs that test for barrier function; candidate insulator sequences are upstream of the enhancer/reporter while the chicken β-globin insulator is located downstream. Yet these enhancer-blocking constructs are not ideal to use to study barrier function. Using a reporter that confers antibiotic resistance to a cell will give those cells expressing the reporter a growth advantage under selection. Therefore, to minimize this selective growth advantage, and unlike in our enhancer-blocking assays, cells will need to be grown without selection prior to testing for barrier function. Additionally, at a minimum, a second selectable marker located outside of the flanking insulators is necessary to select for cells that successfully integrate the transgene.

Determining whether the *UBA1* insulator and/or the surrounding sequences are capable of preventing the spread of heterochromatin will provide valuable insight into the possible regulatory mechanisms used at the endogenous *UBA1* locus. As enhancer-blocking and barrier functions can be independent functions of nearby sequences at some chromatin insulators (Recillas-Targa et al. 2002), it will be necessary to test all *UBA1* sequences within the expression boundary as well as the *UBA1* exon 1d CpG island, which contains the CTCF binding site, for barrier function. Even though CTCF binds *UBA1*on both the active and inactive X and therefore is unlikely to participate in inactive X specific regulation, there remains the possibility that this CTCF binding has a role in *UBA1* escape. A genome-wide survey of CTCF binding sites identified many sites that were located between active and repressive chromatin domains suggesting that CTCF may regulate the barrier function of some insulators (Cuddapah et al. 2009) even though to date no such function for CTCF has been demonstrated. Characterization of the CTCF binding site within *UBA1* will be important to eliminate CTCF as a regulator of *UBA1* escape. By testing all *UBA1* expression boundary sequences for barrier function, any additional sequences that may contribute to insulator function will be identified. Determining whether the *UBA1* insulator can function in the additional capacity and prevent heterochromatin spread is important for understanding how insulators may function in XCI gene regulation. It is expected that insulators, capable of regulating XCI escape domains, would prevent the spread of heterochromatin and therefore exhibit barrier function in the reporter assays. However, outside of the context of XCI, there is a chance that the *UBA1* insulator may not demonstrate this barrier function. Yet, even insulators that are only capable of enhancer-blocking could potentially regulate XCI. Many enhancer-blocking insulators appear to function such that they interact with other insulators or tether chromatin to structural elements within the nucleus forming chromatin loop domains that separate enhancers from promoters (reviewed in (Gaszner et al. 2006)). The

formation of similar chromatin structures by the *UBA1* enhancer-blocking insulator could regulate the XCI escape domain.

Characterization Of The Insulator Sequence At The Endogenous *UBA1* Locus

Insulator proteins are known to act in several distinct ways to achieve a similar end result. In addition to recruiting epigenetic remodeling complexes, such as histone methyltransferases and histone acetyltransferases (Huang et al. 2007), many insulators are capable of playing a role in nuclear organization by interacting with each other through protein-protein interactions to form chromatin loops or by binding additional proteins that cluster in the nucleus to form insulator bodies resulting in the formation organized chromatin structures (reviewed in (Bushey et al. 2008; Raab et al. 2010)). If the identified *UBA1* insulator is functional in its endogenous sequence context, then we expect that the sequence should share at least some properties of known insulators.

The importance of three-dimensional architecture in XCI suggests that X-linked insulator sequences may contribute to the nuclear organization of the inactive X. At the initiation of XCI, the sequences of the *Xic* are highly ordered and form two chromatin loops that are postulated to regulate XCI (Tsai et al. 2008). That X sequences may also be highly organized during the maintenance of XCI is an intriguing possibility that may explain aspects of XCI gene regulation. Escape domains may be positioned at the periphery of the *Xist* nuclear compartment by insulator interactions allowing them to be expressed. Therefore, to discover whether sequences at the *UBA1* escape domain are indeed highly ordered, circularized chromatin conformation capture (4C) followed by high-throughput sequencing can be used to identify DNA sequences that specifically interact with each other. The 4C technique identifies interacting regions by first crosslinking chromatin, then restriction digesting the interacting regions and finally self-ligating the crosslinked fragments. The frequency with which two restriction fragments are ligated is an

indication of how often they interact. Sequencing can then be used to identify sequences that interact with the *UBA1* insulator. Subsequent FISH analysis in interphase nuclei will confirm the colocalization of *UBA1* sequences with the identified interacting regions. If the *UBA1* insulator binds other insulator sequences to form ordered chromatin loops, then *UBA1* sequences will be preferentially ligated to a select few regions in the genome using 4C. However, if *UBA1* sequences are not ordered, then the insulator will interact with a large number of different chromatin regions. Not only will these studies determine whether the *UBA1* insulator forms distinct chromatin loops to regulate XCI escape, but they will also shed light on whether escape domains in general form highly ordered structures.

Identification Of Protein(s) That Bind The *UBA1* Insulator

Insulators are postulated to have evolved from a class of promoters that bind specific transcription factors to mediate a variety of forms of insulation (Raab et al. 2010). This is not surprising given that all of the known *trans*-acting factors that bind vertebrate insulator sequences are transcription factors, including CTCF, USF1 and VEZF1, three known transcription factors that function independently to ensure proper regulation at the chicken β-globin locus (Bell et al. 1999; West et al. 2004; Huang et al. 2007; Dickson et al. 2010). To date, no protein has been identified to specifically bind the known CTCF-independent insulators. However, recent genome-wide ChIP-seq experiments associated with the Encyclopedia of DNA Elements, or ENCODE, project have identified candidate *trans*-acting factors (Birney et al. 2007). The goal of the ENCODE project is to identify all of the functional elements within the human genome. To this end, the ENCODE researchers at the HudsonAlpha Institute for Biotechnology used ChIP-seq to identify binding sites of many transcription factors and the results from these experiments are publically available on the UCSC Genome browser (http://genome.ucsc.edu) (Birney et al. 2007). These studies failed to identify binding at the *UBA1* insulator sequence of

any of the known transcription factors that have previously been demonstrated to bind insulators including CTCF and USF1. Notably, of the transcription factors that have been looked at, only the Early B-cell factor, or EBF, family of transcription factors bind at the *UBA1* insulator sequence. Similar to CTCF, >500 binding sites for EBF were identified along the X chromosome suggesting that this protein may function in multiple regulatory capacities, although to date no XCI or X-linked gene regulatory function has been identified.

The EBF family of transcription factors consists of four members, EBF1-4, and all bind a conserved core consensus site 5'-CCNNGG-3'(Liberg et al. 2002). Therefore, it is not surprising that the antibody used in the ENCODE study did not distinguish between individual EBF proteins. As the name implies, the EBF family is necessary for early B-cell development (Lin et al. 1995), but new evidence has revealed additional roles outside of B-cells. At least some EBF proteins are ubiquitously expressed in adult tissues, whereas others exhibit tissue specific and/or are restricted to particular developmental stages (Liberg et al. 2002). However, to date, no experimental evidence indicates that EBF is involved in XCI gene regulation. Therefore, to confirm that EBF transcription factors interact with the *UBA1* sequence, an electrophoretic mobility shift assay (EMSA) will demonstrate the specificity of the interaction *in vitro*. If EBF regulates the *UBA1* escape domain through its binding to the insulator, then we expect that it would be specifically targeted to the inactive X. A comparison of EBF binding at the *UBA1* insulator between male and female fibroblasts is necessary, as the ChIP-seq studies only assayed a single female cell line. Males, having only a single active X, are expected to lack binding if EBF is specific for the inactive X. While human *UBA1* contains a distinct 5'- CCNNGG-3' binding site, mouse *Uba1* lacks this site therefore should not bind EBF. Thus, the absence of EBF at mouse *Uba1* will need to be confirmed.

If EBF binding is specific for the human inactive X, then additional studies are necessary to determine whether EBF is responsible for the activity of the *UBA1* insulator. Since endogenous

insulator function is difficult to assay in the context of XCI we can indirectly test the role of EBF on the *UBA1* insulator using the enhancer-blocking reporter assay in two ways. First, *UBA1* insulator sequences lacking the 5'-CCNNGG-3' EBF binding site can be assayed for insulator activity. Constructs lacking the binding site should no longer be able to insulate the reporter if EBF binding is necessary for insulator function. Second, the EBF knockdown using siRNA in cells that have stably integrated the *UBA1* insulator construct should induce normal reporter activity if EBF is responsible for insulator function. The knockdown of individual EBF proteins would also serve to identify which family member binds *UBA1* and confers insulator activity. If these studies of EBF do not establish a role for this transcription factor in XCI regulation, or regulation of the *UBA1* insulator, then we will need to search for other potential binding factors. Transfac or PATCH (http://www.gene-regulation.com), two computational programs that are designed to identify transcription factor binding sites in a given DNA sequence, can be used to identify potential *trans*-acting binding proteins. Candidate transcription factors could be subsequently assayed for binding at *UBA1* using ChIP. However, given that multiple promoter and enhancer elements are located within the 5' UTR of *UBA1* these programs return a large number of possible factors that bind within the region that would need to be assayed. Therefore, we chose to examine a promising candidate regulatory factor that by ChIP has been identified to bind *UBA1*.

References

Agrelo, R., A. Souabni, et al. (2009). *SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells*. Dev Cell **16**(4):p. 507-16.

Ait Yahya-Graison, E., J. Aubert, et al. (2007). *Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes*. Am J Hum Genet **81**(3):p. 475-91.

Amos-Landgraf, J. M., A. Cottle, et al. (2006). *X chromosome-inactivation patterns of 1,005 phenotypically unaffected females*. Am J Hum Genet **79**(3):p. 493-9.

Antes, T. J., S. J. Namciu, et al. (2001). *The 5' boundary of the human apolipoprotein B chromatin domain in intestinal cells*. Biochemistry **40**(23):p. 6731-42.

Ashworth, A., S. Rastan, et al. (1991). *X-chromosome inactivation may explain the difference in viability of XO humans and mice*. Nature **351**(6325):p. 406-8.

Augui, S., G. J. Filion, et al. (2007). *Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic*. Science **318**(5856):p. 1632-6.

Bacher, C. P., M. Guggiari, et al. (2006). *Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation*. Nat Cell Biol **8**(3):p. 293-9.

Bailey, J. A., L. Carrel, et al. (2000). *Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis*. Proc Natl Acad Sci U S A **97**(12):p. 6634-9.

Barr, M. L. and E. G. Bertram (1949). *A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis*. Nature **163**(4148):p. 676.

Basrur, P. K., A. Farazmand, et al. (2004). *Expression pattern of X-linked genes in sex chromosome aneuploid bovine cells*. Chromosome Res **12**(3):p. 263-73.

Bell, A. C., A. G. West, et al. (1999). *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators*. Cell **98**(3):p. 387-96.

Berg, J. S., L. Potocki, et al. (2010). *Common recurrent microduplication syndromes: diagnosis and management in clinical practice*. Am J Med Genet A **152A**(5):p. 1066-78.

Bernstein, B. E., A. Meissner, et al. (2007). *The mammalian epigenome*. Cell **128**(4):p. 669-81.

Bi, X. and J. R. Broach (2001). *Chromosomal boundaries in S. cerevisiae*. Curr Opin Genet Dev **11**(2):p. 199-204.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*. Nature **447**(7146):p. 799-816.

Blackledge, N. P., C. J. Ott, et al. (2009). *An insulator element 3' to the CFTR gene binds CTCF and reveals an active chromatin hub in primary cells*. Nucleic Acids Res:p.

Blewitt, M. E., A. V. Gendrel, et al. (2008). *SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation*. Nat Genet **40**(5):p. 663-9.

Boggs, B. A. and A. C. Chinault (1994). *Analysis of replication timing properties of human X-chromosomal loci by fluorescence in situ hybridization*. Proc Natl Acad Sci U S A **91**(13):p. 6083-7.

Boggs, B. A., B. Connors, et al. (1996). *Reduced levels of histone H3 acetylation on the inactive X chromosome in human females*. Chromosoma **105**(5):p. 303-9.

Boissinot, S., A. Entezam, et al. (2001). *Selection against deleterious LINE-1-containing loci in the human lineage*. Mol Biol Evol **18**(6):p. 926-35.

Bondy, C. A. and C. Cheng (2009). *Monosomy for the X chromosome*. Chromosome Res **17**(5):p. 649-58.

Borsani, G., R. Tonlorenzi, et al. (1991). *Characterization of a murine gene expressed from the inactive X chromosome*. Nature **351**(6324):p. 325-9.

Boyle, A. L., S. G. Ballard, et al. (1990). *Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization*. Proc Natl Acad Sci U S A **87**(19):p. 7757-61.

Boyle, A. P., S. Davis, et al. (2008). *High-resolution mapping and characterization of open chromatin across the genome*. Cell **132**(2):p. 311-22.

Brinkman, A. B., T. Roelofsen, et al. (2006). *Histone modification patterns associated with the human X chromosome*. EMBO Rep **7**(6):p. 628-34.

Brockdorff, N., A. Ashworth, et al. (1991). *Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome*. Nature **351**(6324):p. 329-31.

Brown, C. J., A. Ballabio, et al. (1991). *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome*. Nature **349**(6304):p. 38-44.

Brown, C. J., L. Carrel, et al. (1997). *Expression of genes from the human active and inactive X chromosomes*. Am J Hum Genet **60**(6):p. 1333-43.

Brown, C. J., B. D. Hendrich, et al. (1992). *The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus*. Cell **71**(3):p. 527-42.

Brown, C. J. and H. F. Willard (1989). *Noninactivation of a selectable human X-linked gene that complements a murine temperature-sensitive cell cycle defect*. Am J Hum Genet **45**(4):p. 592-8.

Brown, S. D. (1991). *XIST and the mapping of the X chromosome inactivation centre*. Bioessays **13**(11):p. 607-12.

Burke, L. J., R. Zhang, et al. (2005). *CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin*. Embo J **24**(18):p. 3291-300.

Burn, J., S. Povey, et al. (1986). *Duchenne muscular dystrophy in one of monozygotic twin girls*. J Med Genet **23**(6):p. 494-500.

Bushey, A. M., E. R. Dorman, et al. (2008). *Chromatin insulators: regulatory mechanisms and epigenetic inheritance*. Mol Cell **32**(1):p. 1-9.

Carrel, L., C. M. Clemson, et al. (1996). *X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse*. Hum Mol Genet **5**(3):p. 391-401.

Carrel, L., P. A. Hunt, et al. (1996). *Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene*. Hum Mol Genet **5**(9):p. 1361-6.

Carrel, L., C. Park, et al. (2006). *Genomic environment predicts expression patterns on the human inactive X chromosome*. PLoS Genet **2**(9):p. e151.

Carrel, L. and H. F. Willard (1999). *Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others*. Proc Natl Acad Sci U S A **96**(13):p. 7364-9.

Carrel, L. and H. F. Willard (2005). *X-inactivation profile reveals extensive variability in X-linked gene expression in females*. Nature **434**(7031):p. 400-4.

Chadwick, B. P. (2007). *Variation in Xi chromatin organization and correlation of the H3K27me3 chromatin territories to transcribed sequences by microarray analysis*. Chromosoma **116**(2):p. 147-57.

Chadwick, B. P. and H. F. Willard (2003). *Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome*. Hum Mol Genet **12**(17):p. 2167-78.

Chadwick, B. P. and H. F. Willard (2004). *Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome*. Proc Natl Acad Sci U S A **101**(50):p. 17450-5.

Chaumeil, J., P. Le Baccon, et al. (2006). *A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced*. Genes Dev **20**(16):p. 2223-37.

Chong, S., J. Kontaraki, et al. (2002). *A Functional chromatin domain does not resist X chromosome inactivation: silencing of cLys correlates with methylation of a dual promoter-replication origin*. Mol Cell Biol **22**(13):p. 4667-76.

Chow, J. and E. Heard (2009). *X inactivation and the complexities of silencing a sex chromosome*. Curr Opin Cell Biol **21**(3):p. 359-66.

Chow, J. C., C. Ciaudo, et al. (2010). *LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation*. Cell **141**(6):p. 956-69.

Chow, J. C., L. L. Hall, et al. (2007). *Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible*. Proc Natl Acad Sci U S A **104**(24):p. 10104-9.

Chung, J. H., A. C. Bell, et al. (1997). *Characterization of the chicken beta-globin insulator*. Proc Natl Acad Sci U S A **94**(2):p. 575-80.

Chung, J. H., M. Whiteley, et al. (1993). *A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila*. Cell **74**(3):p. 505-14.

Ciavatta, D., S. Kalantry, et al. (2006). *A DNA insulator prevents repression of a targeted X-linked transgene but not its random or imprinted X inactivation*. Proc Natl Acad Sci U S A **103**(26):p. 9958-63.

Clarke, J. T., P. J. Wilson, et al. (1992). *Characterization of a deletion at Xq27-q28 associated with unbalanced inactivation of the nonmutant X chromosome*. Am J Hum Genet **51**(2):p. 316-22.

Clemson, C. M., L. L. Hall, et al. (2006). *The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences*. Proc Natl Acad Sci U S A **103**(20):p. 7688-93.

Clemson, C. M., J. A. McNeil, et al. (1996). *XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure*. J Cell Biol **132**(3):p. 259-75.

Coleman, M. P., H. J. Ambrose, et al. (1996). *A novel gene, DXS8237E, lies within 20 kb upstream of UBE1 in Xp11.23 and has a different X inactivation status*. Genomics **31**(1):p. 135-8.

Costanzi, C. and J. R. Pehrson (1998). *Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals*. Nature **393**(6685):p. 599-601.

Craig, I. W., J. Mill, et al. (2004). *Application of microarrays to the analysis of the inactivation status of human X-linked genes expressed in lymphocytes*. Eur J Hum Genet **12**(8):p. 639-46.

Csankovszki, G., A. Nagy, et al. (2001). *Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation*. J Cell Biol **153**(4):p. 773-84.

Csankovszki, G., B. Panning, et al. (1999). *Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation*. Nat Genet **22**(4):p. 323-4.

Cuddapah, S., R. Jothi, et al. (2009). *Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains*. Genome Res **19**(1):p. 24-32.

de Napoles, M., J. E. Mermoud, et al. (2004). *Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation*. Dev Cell **7**(5):p. 663-76.

Debrand, E., C. Chureau, et al. (1999). *Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression*. Mol Cell Biol **19**(12):p. 8513-25.

Dickson, J., H. Gowher, et al. (2010). *VEZF1 elements mediate protection from DNA methylation*. PLoS Genet **6**(1):p. e1000804.

Dietzel, S., K. Schiebel, et al. (1999). *The 3D positioning of ANT2 and ANT3 genes within female X chromosome territories correlates with gene activity*. Exp Cell Res **252**(2):p. 363-75.

Dillon, N. and R. Festenstein (2002). *Unravelling heterochromatin: competition between positive and negative factors regulates accessibility*. Trends Genet **18**(5):p. 252-8.

Disteche, C. M., E. Zacksenhaus, et al. (1992). *Mapping and expression of the ubiquitin-activating enzyme E1 (Ube1) gene in the mouse*. Mamm Genome **3**(3):p. 156-61.

Ditton, H. J., J. Zimmer, et al. (2004). *The AZFa gene DBY (DDX3Y) is widely transcribed but the protein is limited to the male germ cells by translation control*. Hum Mol Genet **13**(19):p. 2333-41.

Dobyns, W. B., A. Filauro, et al. (2004). *Inheritance of most X-linked traits is not dominant or recessive, just X-linked*. Am J Med Genet A **129A**(2):p. 136-43.

Donohoe, M. E., S. S. Silva, et al. (2009). *The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting*. Nature **460**(7251):p. 128-32.

Donohoe, M. E., L. F. Zhang, et al. (2007). *Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch*. Mol Cell **25**(1):p. 43-56.

Duthie, S. M., T. B. Nesterova, et al. (1999). *Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis*. Hum Mol Genet **8**(2):p. 195-204.

Edgar, R. C. (2004). *MUSCLE: a multiple sequence alignment method with reduced time and space complexity*. BMC Bioinformatics **5**:p. 113.

Fang, J., T. Chen, et al. (2004). *Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation*. J Biol Chem **279**(51):p. 52812-5.

Filippova, G. N., M. K. Cheng, et al. (2005). *Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development*. Dev Cell **8**(1):p. 31-42.

Gartler, S. M. and A. D. Riggs (1983). *Mammalian X-chromosome inactivation*. Annu Rev Genet **17**:p. 155-90.

Gaszner, M. and G. Felsenfeld (2006). *Insulators: exploiting transcriptional and epigenetic mechanisms*. Nat Rev Genet **7**(9):p. 703-13.

Gilbert, C. W., S. Muldal, et al. (1962). *Time-sequence of human chromosome duplication*. Nature **195**:p. 869-73.

Gilbert, S. L. and P. A. Sharp (1999). *Promoter-specific hypoacetylation of X-inactivated genes*. Proc Natl Acad Sci U S A **96**(24):p. 13825-30.

Goetze, S., A. Baer, et al. (2005). *Performance of genomic bordering elements at predefined genomic loci*. Mol Cell Biol **25**(6):p. 2260-72.

Golovnin, A., I. Biryukova, et al. (2003). *An endogenous Su(Hw) insulator separates the yellow gene from the Achaete-scute gene complex in Drosophila*. Development **130**(14):p. 3249-58.

Gombert, W. M., S. D. Farris, et al. (2003). *The c-myc insulator element and matrix attachment regions define the c-myc chromosomal domain*. Mol Cell Biol **23**(24):p. 9338-48.

Goto, Y., M. Gomez, et al. (2002). *Differential patterns of histone methylation and acetylation distinguish active and repressed alleles at X-linked genes*. Cytogenet Genome Res **99**(1-4):p. 66-74.

Goto, Y. and H. Kimura (2009). *Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene*. Nucleic Acids Res **37**(22):p. 7416-28.

Graves, J. A. (1982). *5-azacytidine-induced re-expression of alleles on the inactive X chromosome in a hybrid mouse cell line*. Exp Cell Res **141**(1):p. 99-105.

Graves, J. A., E. Koina, et al. (2006). *How the gene content of human sex chromosomes evolved*. Curr Opin Genet Dev **16**(3):p. 219-24.

Grewal, S. I. and S. Jia (2007). *Heterochromatin revisited*. Nat Rev Genet **8**(1):p. 35-46.

Hansen, L., K. C. Arden, et al. (1997). *Chromosomal mapping and mutational analysis of the coding region of the glycogen synthase kinase-3alpha and beta isoforms in patients with NIDDM*. Diabetologia **40**(8):p. 940-6.

Hansen, R. S., T. K. Canfield, et al. (1996). *Role of late replication timing in the silencing of X-linked genes*. Hum Mol Genet **5**(9):p. 1345-53.

Hansen, R. S., R. Stoger, et al. (2000). *Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant*. Hum Mol Genet **9**(18):p. 2575-87.

Haqq, C. M., C. Y. King, et al. (1994). *Molecular basis of mammalian sexual determination: activation of Mullerian inhibiting substance gene expression by SRY*. Science **266**(5190):p. 1494-500.

Hayatsu, H., Y. Wataya, et al. (1970). *The addition of sodium bisulfite to uracil and to cytosine*. J Am Chem Soc **92**(3):p. 724-6.

Heard, E. and C. M. Disteche (2006). *Dosage compensation in mammals: fine-tuning the expression of the X chromosome*. Genes Dev **20**(14):p. 1848-67.

Heard, E., C. Rougeulle, et al. (2001). *Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation*. Cell **107**(6):p. 727-38.

Heintzman, N. D., G. C. Hon, et al. (2009). *Histone modifications at human enhancers reflect global cell-type-specific gene expression*. Nature **459**(7243):p. 108-12.

Hellman, A. and A. Chess (2007). *Gene body-specific methylation on the active X chromosome*. Science **315**(5815):p. 1141-3.

Hernandez-Munoz, I., A. H. Lund, et al. (2005). *Stable X chromosome inactivation involves the PRC1 Polycomb complex and requires histone MACROH2A1 and the CULLIN3/SPOP ubiquitin E3 ligase*. Proc Natl Acad Sci U S A **102**(21):p. 7635-40.

Herzing, L. B., J. T. Romer, et al. (1997). *Xist has properties of the X-chromosome inactivation centre*. Nature **386**(6622):p. 272-5.

Hong, Y. K., S. D. Ontiveros, et al. (2000). *A revision of the human XIST gene organization and structural comparison with mouse Xist*. Mamm Genome **11**(3):p. 220-4.

Huang, S., X. Li, et al. (2007). *USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier*. Mol Cell Biol **27**(22):p. 7991-8002.

Inlow, J. K. and L. L. Restifo (2004). *Molecular and comparative genetics of mental retardation*. Genetics **166**(2):p. 835-81.

Jegalian, K. and D. C. Page (1998). *A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated*. Nature **394**(6695):p. 776-80.

Johnston, C. M., F. L. Lovell, et al. (2008). *Large-scale population study of human cell lines indicates that dosage compensation is virtually complete*. PLoS Genet **4**(1):p. e9.

Jonkers, I., T. S. Barakat, et al. (2009). *RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation*. Cell **139**(5):p. 999-1011.

Kalantry, S. and T. Magnuson (2006). *The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation*. PLoS Genet **2**(5):p. e66.

Karachentsev, D., K. Sarma, et al. (2005). *PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis*. Genes Dev **19**(4):p. 431-5.

Kay, G. F., G. D. Penny, et al. (1993). *Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation*. Cell **72**(2):p. 171-82.

Ke, X. and A. Collins (2003). *CpG islands in human X-inactivation*. Ann Hum Genet **67**(Pt 3):p. 242-9.

Keohane, A. M., A. L. Barlow, et al. (1999). *H4 acetylation, XIST RNA and replication timing are coincident and define x;autosome boundaries in two abnormal X chromosomes*. Hum Mol Genet **8**(2):p. 377-83.

Keohane, A. M., P. O'Neill L, et al. (1996). *X-Inactivation and histone H4 acetylation in embryonic stem cells*. Dev Biol **180**(2):p. 618-30.

Kim, J., A. Kollhoff, et al. (2003). *Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3*. Hum Mol Genet **12**(3):p. 233-45.

Kim, T. H., Z. K. Abdullaev, et al. (2007). *Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome*. Cell **128**(6):p. 1231-45.

Kirchgessner, C. U., S. T. Warren, et al. (1995). *X inactivation of the FMR1 fragile X mental retardation gene*. J Med Genet **32**(12):p. 925-9.

Klose, R. J. and A. P. Bird (2006). *Genomic DNA methylation: the mark and its mediators*. Trends Biochem Sci **31**(2):p. 89-97.

Kohlmaier, A., F. Savarese, et al. (2004). *A chromosomal memory triggered by xist regulates histone methylation in x inactivation*. PLoS Biol **2**(7):p. E171.

Komura, J., H. Ikehata, et al. (2007). *Chromatin fine structure of the c-MYC insulator element/DNase I-hypersensitive site I is not preserved during mitosis*. Proc Natl Acad Sci U S A **104**(40):p. 15741-6.

Korenberg, J. R. and M. C. Rykowski (1988). *Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands*. Cell **53**(3):p. 391-400.

Kouzarides, T. (2007). *Chromatin modifications and their function*. Cell **128**(4):p. 693-705.

Lahn, B. T. and D. C. Page (1999). *Four evolutionary strata on the human X chromosome*. Science **286**(5441):p. 964-7.

Lee, J. T. (2002). *Homozygous Tsix mutant mice reveal a sex-ratio distortion and revert to random X-inactivation*. Nat Genet **32**(1):p. 195-200.

Lee, J. T. (2005). *Regulation of X-chromosome counting by Tsix and Xite sequences*. Science **309**(5735):p. 768-71.

Lee, J. T., L. S. Davidow, et al. (1999). *Tsix, a gene antisense to Xist at the X-inactivation centre*. Nat Genet **21**(4):p. 400-4.

Lee, J. T. and R. Jaenisch (1997). *Long-range cis effects of ectopic X-inactivation centres on a mouse autosome*. Nature **386**(6622):p. 275-9.

Lee, J. T. and N. Lu (1999). *Targeted mutagenesis of Tsix leads to nonrandom X inactivation*. Cell **99**(1):p. 47-57.

Lee, J. T., W. M. Strauss, et al. (1996). *A 450 kb transgene displays properties of the mammalian X-inactivation center*. Cell **86**(1):p. 83-94.

Leeb, M. and A. Wutz (2007). *Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells*. J Cell Biol **178**(2):p. 219-29.

Li, N. and L. Carrel (2008). *Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus*. Proc Natl Acad Sci U S A **105**(44):p. 17055-60.

Liberg, D., M. Sigvardsson, et al. (2002). *The EBF/Olf/Collier family of transcription factors: regulators of differentiation in cells originating from all three embryonal germ layers*. Mol Cell Biol **22**(24):p. 8389-97.

Lin, H. and R. Grosschedl (1995). *Failure of B-cell differentiation in mice lacking the transcription factor EBF*. Nature **376**(6537):p. 263-7.

Littlefield, J. W. (1966). *The use of drug-resistant markers to study the hybridization of mouse fibroblasts*. Exp Cell Res **41**(1):p. 190-6.

Lue, Y., P. N. Rao, et al. (2001). *XXY male mice: an experimental model for Klinefelter syndrome*. Endocrinology **142**(4):p. 1461-70.

Luikenhuis, S., A. Wutz, et al. (2001). *Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells*. Mol Cell Biol **21**(24):p. 8512-20.

Luoh, S. W., K. Jegalian, et al. (1995). *CpG islands in human ZFX and ZFY and mouse Zfx genes: sequence similarities and methylation differences*. Genomics **29**(2):p. 353-63.

Lyon, M. F. (1961). *Gene action in the X-chromosome of the mouse (Mus musculus L.)*. Nature **190**:p. 372-3.

Lyon, M. F. (1962). *Sex chromatin and gene action in the mammalian X-chromosome*. Am J Hum Genet **14**:p. 135-48.

Lyon, M. F. (1998). *X-chromosome inactivation: a repeat hypothesis*. Cytogenet Cell Genet **80**(1-4):p. 133-7.

Magdinier, F., T. M. Yusufzai, et al. (2004). *Both CTCF-dependent and -independent insulators are found between the mouse T cell receptor alpha and Dad1 genes*. J Biol Chem **279**(24):p. 25381-9.

Majumder, P. and H. N. Cai (2003). *The functional analysis of insulator interactions in the Drosophila embryo*. Proc Natl Acad Sci U S A **100**(9):p. 5223-8.

Marahrens, Y., B. Panning, et al. (1997). *Xist-deficient mice are defective in dosage compensation but not spermatogenesis*. Genes Dev **11**(2):p. 156-66.

Marks, H., J. C. Chow, et al. (2009). *High-resolution analysis of epigenetic changes associated with X inactivation*. Genome Res **19**(8):p. 1361-73.

Maston, G. A., S. K. Evans, et al. (2006). *Transcriptional regulatory elements in the human genome*. Annu Rev Genomics Hum Genet **7**:p. 29-59.

Matthews, P. M., D. Benjamin, et al. (1995). *Muscle X-inactivation patterns and dystrophin expression in Duchenne muscular dystrophy carriers*. Neuromuscul Disord **5**(3):p. 209-20.

McDaniell, R., B. K. Lee, et al. (2010). *Heritable individual-specific and allele-specific chromatin signatures in humans*. Science **328**(5975):p. 235-9.

McNeil, J. A., K. P. Smith, et al. (2006). *Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]n in the X escape region*. Genome Res **16**(4):p. 477-84.

Mermoud, J. E., C. Costanzi, et al. (1999). *Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation*. J Cell Biol **147**(7):p. 1399-408.

Mermoud, J. E., B. Popova, et al. (2002). *Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation*. Curr Biol **12**(3):p. 247-51.

Mietton, F., A. K. Sengupta, et al. (2009). *Weak but uniform enrichment of the histone variant macroH2A1 along the inactive X chromosome*. Mol Cell Biol **29**(1):p. 150-6.

Migeon, B. R., A. K. Chowdhury, et al. (2001). *Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation*. Am J Hum Genet **69**(5):p. 951-60.

Mikkelsen, T. S., M. J. Wakefield, et al. (2007). *Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences*. Nature **447**(7141):p. 167-77.

Miller, A. P., K. Gustashaw, et al. (1995). *Three genes that escape X chromosome inactivation are clustered within a 6 Mb YAC contig and STS map in Xp11.21-p11.22*. Hum Mol Genet **4**(4):p. 731-9.

Mohandas, T., R. S. Sparkes, et al. (1981). *Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation*. Science **211**(4480):p. 393-6.

Monk, M. and M. I. Harper (1979). *Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos*. Nature **281**(5729):p. 311-3.

Monkhorst, K., B. de Hoon, et al. (2009). *The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties*. PLoS ONE **4**(5):p. e5616.

Monkhorst, K., I. Jonkers, et al. (2008). *X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator*. Cell **132**(3):p. 410-21.

Morey, C., C. Kress, et al. (2009). *Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression*. Genome Res **19**(7):p. 1184-94.

Morishima, A., M. M. Grumbach, et al. (1962). *Asynchronous duplication of human chromosomes and the origin of sex chromatin*. Proc Natl Acad Sci U S A **48**:p. 756-63.

Morleo, M. and B. Franco (2008). *Dosage compensation of the mammalian X chromosome influences the phenotypic variability of X-linked dominant male-lethal disorders*. J Med Genet **45**(7):p. 401-8.

Mueller, J. L., S. K. Mahadevaiah, et al. (2008). *The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression*. Nat Genet **40**(6):p. 794-9.

Murakami, K., T. Ohhira, et al. (2009). *Identification of the chromatin regions coated by non-coding Xist RNA*. Cytogenet Genome Res **125**(1):p. 19-25.

Namekawa, S. H., B. Payer, et al. (2010). *Two-step imprinted X inactivation: repeat versus genic silencing in the mouse*. Mol Cell Biol **30**(13):p. 3187-205.

Navarro, P. and P. Avner (2010). *An embryonic story: analysis of the gene regulative network controlling Xist expression in mouse embryonic stem cells*. Bioessays **32**(7):p. 581-8.

Navarro, P., I. Chambers, et al. (2008). *Molecular coupling of Xist regulation and pluripotency*. Science **321**(5896):p. 1693-5.

Nelson, J. D., O. Denisenko, et al. (2006). *Protocol for the fast chromatin immunoprecipitation (ChIP) method*. Nat Protoc **1**(1):p. 179-85.

Nesterova, T. B., C. M. Johnston, et al. (2003). *Skewing X chromosome choice by modulating sense transcription across the Xist locus*. Genes Dev **17**(17):p. 2177-90.

Nishioka, K., J. C. Rice, et al. (2002). *PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin*. Mol Cell **9**(6):p. 1201-13.

O'Neill, L. P., H. T. Spotswood, et al. (2008). *Differential loss of histone H3 isoforms mono-, di- and tri-methylated at lysine 4 during X-inactivation in female embryonic stem cells*. Biol Chem **389**(4):p. 365-70.

Ogawa, Y. and J. T. Lee (2003). *Xite, X-inactivation intergenic transcription elements that regulate the probability of choice*. Mol Cell **11**(3):p. 731-43.

Ohhata, T., Y. Hoki, et al. (2008). *Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification*. Development **135**(2):p. 227-35.

Ohhata, T., M. Tachibana, et al. (2004). *X-inactivation is stably maintained in mouse embryos deficient for histone methyl transferase G9a*. Genesis **40**(3):p. 151-6.

Ohno, S. (1967). Sex Chromosomes and Sex-Linked Genes. Berlin, New York, Springer-Verlag.

Ohno, S. and T. S. Hauschka (1960). *Allocycly of the X-chromosome in tumors and normal tissues*. Cancer Res **20**:p. 541-5.

Okamoto, I. and E. Heard (2009). *Lessons from comparative analysis of X-chromosome inactivation in mammals*. Chromosome Res **17**(5):p. 659-69.

Oki, M. and R. T. Kamakaka (2005). *Barrier function at HMR*. Mol Cell **19**(5):p. 707-16.

Orstavik, K. H. (2009). *X chromosome inactivation in clinical practice*. Hum Genet:p.

Panning, B., J. Dausman, et al. (1997). *X chromosome inactivation is mediated by Xist RNA stabilization*. Cell **90**(5):p. 907-16.

Pant, V., S. Kurukuti, et al. (2004). *Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance*. Mol Cell Biol **24**(8):p. 3497-504.

Parnell, T. J., M. M. Viering, et al. (2003). *An endogenous suppressor of hairy-wing insulator separates regulatory domains in Drosophila*. Proc Natl Acad Sci U S A **100**(23):p. 13436-41.

Payer, B. and J. T. Lee (2008). *X Chromosome Dosage Compensation: How Mammals Keep the Balance*. Annu Rev Genet:p.

Pegoraro, E., R. N. Schimke, et al. (1994). *Detection of new paternal dystrophin gene mutations in isolated cases of dystrophinopathy in females*. Am J Hum Genet **54**(6):p. 989-1003.

Pegoraro, E., R. N. Schimke, et al. (1995). *Genetic and biochemical normalization in female carriers of Duchenne muscular dystrophy: evidence for failure of dystrophin production in dystrophin-competent myonuclei*. Neurology **45**(4):p. 677-90.

Penny, G. D., G. F. Kay, et al. (1996). *Requirement for Xist in X chromosome inactivation*. Nature **379**(6561):p. 131-7.

Penzkofer, T., T. Dandekar, et al. (2005). *L1Base: from functional annotation to prediction of active LINE-1 elements*. Nucleic Acids Res **33**(Database issue):p. D498-500.

Peters, A. H., J. E. Mermoud, et al. (2002). *Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin*. Nat Genet **30**(1):p. 77-80.

Peterson, C. L. and M. A. Laniel (2004). *Histones and histone modifications*. Curr Biol **14**(14):p. R546-51.

Phillips, J. E. and V. G. Corces (2009). *CTCF: master weaver of the genome*. Cell **137**(7):p. 1194-211.

Plath, K., J. Fang, et al. (2003). *Role of histone H3 lysine 27 methylation in X inactivation*. Science **300**(5616):p. 131-5.

Plath, K., D. Talbot, et al. (2004). *Developmentally regulated alterations in Polycomb repressive complex 1 proteins on the inactive X chromosome*. J Cell Biol **167**(6):p. 1025-35.

Prothero, K. E., J. M. Stahl, et al. (2009). *Dosage compensation and gene expression on the mammalian X chromosome: one plus one does not always equal two*. Chromosome Res **17**(5):p. 637-48.

Raab, J. R. and R. T. Kamakaka (2010). *Insulators and promoters: closer than we think*. Nat Rev Genet **11**(6):p. 439-46.

Ramser, J., M. E. Ahearn, et al. (2008). *Rare missense and synonymous variants in UBE1 are associated with X-linked infantile spinal muscular atrophy*. Am J Hum Genet **82**(1):p. 188-93.

Rao, E., B. Weiss, et al. (1997). *Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome*. Nat Genet **16**(1):p. 54-63.

Rasmussen, T. P., M. A. Mastrangelo, et al. (2000). *Dynamic relocalization of histone MacroH2A1 from centrosomes to inactive X chromosomes during X inactivation*. J Cell Biol **150**(5):p. 1189-98.

Rastan, S. (1983). *Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre*. J Embryol Exp Morphol **78**:p. 1-22.

Rastan, S. and E. J. Robertson (1985). *X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation*. J Embryol Exp Morphol **90**:p. 379-88.

Recillas-Targa, F., M. J. Pikaart, et al. (2002). *Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities*. Proc Natl Acad Sci U S A **99**(10):p. 6883-8.

Ross, J. L., D. Roeltgen, et al. (2000). *The Turner syndrome-associated neurocognitive phenotype maps to distal Xp*. Am J Hum Genet **67**(3):p. 672-81.

Ross, M. T., D. V. Grafham, et al. (2005). *The DNA sequence of the human X chromosome*. Nature **434**(7031):p. 325-37.

Rougeulle, C., J. Chaumeil, et al. (2004). *Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome*. Mol Cell Biol **24**(12):p. 5475-84.

Rougeulle, C., P. Navarro, et al. (2003). *Promoter-restricted H3 Lys 4 di-methylation is an epigenetic mark for monoallelic expression*. Hum Mol Genet **12**(24):p. 3343-8.

Sado, T., M. H. Fenner, et al. (2000). *X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation*. Dev Biol **225**(2):p. 294-303.

Schoeftner, S., A. K. Sengupta, et al. (2006). *Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing*. Embo J **25**(13):p. 3110-22.

Shapiro, L. J., T. Mohandas, et al. (1979). *Non-inactivation of an x-chromosome locus in man*. Science **204**(4398):p. 1224-6.

Shapiro, R., V. DiFate, et al. (1974). *Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction*. J Am Chem Soc **96**(3):p. 906-12.

Sharman, G. B. (1971). *Late DNA replication in the paternally derived X chromosome of female kangaroos*. Nature **230**(5291):p. 231-2.

Sheardown, S. A., S. M. Duthie, et al. (1997). *Stabilization of Xist RNA mediates initiation of X chromosome inactivation*. Cell **91**(1):p. 99-107.

Shibata, S. and J. T. Lee (2004). *Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice*. Curr Biol **14**(19):p. 1747-54.

Shogren-Knaak, M., H. Ishii, et al. (2006). *Histone H4-K16 acetylation controls chromatin structure and protein interactions*. Science **311**(5762):p. 844-7.

Silva, J., W. Mak, et al. (2003). *Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes*. Dev Cell **4**(4):p. 481-95.

Sinclair, A. H., P. Berta, et al. (1990). *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif*. Nature **346**(6281):p. 240-4.

Skaletsky, H., T. Kuroda-Kawaguchi, et al. (2003). *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes*. Nature **423**(6942):p. 825-37.

Soshnev, A. A., X. Li, et al. (2008). *Context differences reveal insulator and activator functions of a Su(Hw) binding region*. PLoS Genet **4**(8):p. e1000159.

Spencer, J. A., A. H. Sinclair, et al. (1991). *Genes on the short arm of the human X chromosome are not shared with the marsupial X*. Genomics **11**(2):p. 339-45.

Stahl, J. M., G. C. Nickel, et al. (2010). *A rapid and quantitative alleleic-expression assay*. Methods in Molecular Biology **in press**:p.

Stavropoulos, N., N. Lu, et al. (2001). *A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice*. Proc Natl Acad Sci U S A **98**(18):p. 10232-7.

Stavropoulos, N., R. K. Rowntree, et al. (2005). *Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation*. Mol Cell Biol **25**(7):p. 2757-69.

Sudbrak, R., G. Wieczorek, et al. (2001). *X chromosome-specific cDNA arrays: identification of genes that escape from X-inactivation and other applications*. Hum Mol Genet **10**(1):p. 77-83.

Sun, B. K., A. M. Deaton, et al. (2006). *A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization*. Mol Cell **21**(5):p. 617-28.

Talebizadeh, Z., S. D. Simon, et al. (2006). *X chromosome gene expression in human tissues: Male and female comparisons*. Genomics **88**(6):p. 675-81.

Tartaglia, N. R., S. Howell, et al. (2010). *A review of trisomy X (47,XXX)*. Orphanet J Rare Dis **5**:p. 8.

Taylor, J. H. (1960). *Asynchronous duplication of chromosomes in cultured cells of Chinese hamster*. J Biophys Biochem Cytol **7**:p. 455-64.

Tsai, C. L., R. K. Rowntree, et al. (2008). *Higher order chromatin structure at the X-inactivation center via looping DNA*. Dev Biol **319**(2):p. 416-25.

Tsuchiya, K. D., J. M. Greally, et al. (2004). *Comparative sequence and x-inactivation analyses of a domain of escape in human xp11.2 and the conserved segment in mouse*. Genome Res **14**(7):p. 1275-84.

Valley, C. M., L. M. Pertz, et al. (2006). *Chromosome-wide, allele-specific analysis of the histone code on the human X chromosome*. Hum Mol Genet **15**(15):p. 2335-47.

Van Esch, H., M. Bauters, et al. (2005). *Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males*. Am J Hum Genet **77**(3):p. 442-53.

Venter, J. C., M. D. Adams, et al. (2001). *The sequence of the human genome*. Science **291**(5507):p. 1304-51.

Veyrunes, F., P. D. Waters, et al. (2008). *Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes*. Genome Res **18**(6):p. 965-73.

Volkel, P. and P. O. Angrand (2007). *The control of histone lysine methylation in epigenetic regulation*. Biochimie **89**(1):p. 1-20.

Wang, Z., H. F. Willard, et al. (2006). *Evidence of influence of genomic DNA sequence on human X chromosome inactivation*. PLoS Comput Biol **2**(9):p. e113.

Waters, P. D., M. C. Wallis, et al. (2007). *Mammalian sex--Origin and evolution of the Y chromosome and SRY*. Semin Cell Dev Biol **18**(3):p. 389-400.

Watson, J. M., J. A. Spencer, et al. (1990). *The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation*. Proc Natl Acad Sci U S A **87**(18):p. 7125-9.

Weber, M., I. Hellmann, et al. (2007). *Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome*. Nat Genet **39**(4):p. 457-66.

Welshons, W. J. and L. B. Russell (1959). *The Y-Chromosome as the Bearer of Male Determining Factors in the Mouse*. Proc Natl Acad Sci U S A **45**(4):p. 560-6.

West, A. G., M. Gaszner, et al. (2002). *Insulators: many functions, many mechanisms*. Genes Dev **16**(3):p. 271-88.

West, A. G., S. Huang, et al. (2004). *Recruitment of Histone Modifications by USF Proteins at a Vertebrate Barrier Element*. Mol Cell **16**(3):p. 453-63.

White, W. M., H. F. Willard, et al. (1998). *The spreading of X inactivation into autosomal material of an x;autosome translocation: evidence for a difference between autosomal and X-chromosomal DNA*. Am J Hum Genet **63**(1):p. 20-8.

Wilson, M. A. and K. D. Makova (2009). *Evolution and survival on eutherian sex chromosomes*. PLoS Genet **5**(7):p. e1000568.

Wu, H., J. Min, et al. (2010). *Structural biology of human H3K9 methyltransferases*. PLoS ONE **5**(1):p. e8570.

Wutz, A. and R. Jaenisch (2000). *A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation*. Mol Cell **5**(4):p. 695-705.

Wutz, A., T. P. Rasmussen, et al. (2002). *Chromosomal silencing and localization are mediated by different domains of Xist RNA*. Nat Genet **30**(2):p. 167-74.

Xie, X., T. S. Mikkelsen, et al. (2007). *Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites*. Proc Natl Acad Sci U S A **104**(17):p. 7145-50.

Xiong, Z., W. Tsark, et al. (1998). *Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension*. Nucleic Acids Res **26**(2):p. 684-6.

Xu, J., P. S. Burgoyne, et al. (2002). *Sex differences in sex chromosome gene expression in mouse brain*. Hum Mol Genet **11**(12):p. 1409-19.

Xu, J., X. Deng, et al. (2008). *Sex-specific expression of the X-linked histone demethylase gene Jarid1c in brain*. PLoS ONE **3**(7):p. e2553.

Xu, J., X. Deng, et al. (2008). *Sex-specific differences in expression of histone demethylases Utx and Uty in mouse brain and neurons*. J Neurosci **28**(17):p. 4521-7.

Xu, J., R. Watkins, et al. (2006). *Sexually dimorphic expression of the X-linked gene Eif2s3x mRNA but not protein in mouse brain*. Gene Expr Patterns **6**(2):p. 146-55.

Xu, N., M. E. Donohoe, et al. (2007). *Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein*. Nat Genet **39**(11):p. 1390-6.

Xu, N., C. L. Tsai, et al. (2006). *Transient homologous chromosome pairing marks the onset of X inactivation*. Science **311**(5764):p. 1149-52.

Yan, H. T., T. Shinka, et al. (2005). *Molecular analysis of TBL1Y, a Y-linked homologue of TBL1X related with X-linked late-onset sensorineural deafness*. J Hum Genet **50**(4):p. 175-81.

Yang, F., T. Babak, et al. (2010). *Global survey of escape from X inactivation by RNA-sequencing in mouse*. Genome Res **20**(5):p. 614-22.

Yasukochi, Y., O. Maruyama, et al. (2010). *X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils*. Proc Natl Acad Sci U S A **107**(8):p. 3704-9.

Yen, Z. C., I. M. Meyer, et al. (2007). *A cross-species comparison of X-chromosome inactivation in Eutheria*. Genomics **90**(4):p. 453-63.

Yoshioka, M., T. Yorifuji, et al. (1998). *Skewed X inactivation in manifesting carriers of Duchenne muscular dystrophy*. Clin Genet **53**(2):p. 102-7.

Young, J. I. and H. Y. Zoghbi (2004). *X-chromosome inactivation patterns are unbalanced and affect the phenotypic outcome in a mouse model of rett syndrome*. Am J Hum Genet **74**(3):p. 511-20.

Yusufzai, T. M., H. Tagami, et al. (2004). *CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species*. Mol Cell **13**(2):p. 291-8.

Zechner, U., M. Wilda, et al. (2001). *A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution?* Trends Genet **17**(12):p. 697-701.

Zeschnigk, M., M. Martin, et al. (2009). *Massive parallel bisulfite sequencing of CG-rich DNA fragments reveals that methylation of many X-chromosomal CpG islands in female blood DNA is incomplete*. Hum Mol Genet **18**(8):p. 1439-48.

Zofall, M. and S. I. Grewal (2006). *Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats*. Mol Cell **22**(5):p. 681-92.

VITA

Katie Eva Prothero

EDUCATION

2003-2010	Pennsylvania State University College of Medicine, Hershey PA Department of Biochemistry and Molecular Biology Ph.D. in Biochemistry and Molecular Biology (Laura Carrel, mentor)
1998-2000	St. Mary's College of Maryland, St. Mary's, MD B.S. in Biology and Chemistry

HONORS

Graduate Alumni Endowment Scholarship, Pennsylvania State University, 2007
Karl H. Beyer, Jr., M.D., Ph.D. Scholarship Pennsylvania State University, 2005
University Graduate Fellowship, Pennsylvania State University, 2003

PUBLICATIONS

Prothero KE, Carrel L. (2010) Identification of a CTCF-independent insulator that delimits X inactivated from expressed transcripts within the *UBA1* gene. *In preparation*

Prothero KE, Stahl JM, Carrel L. (2009) Dosage compensation and gene expression on the mammalian X chromosome: one plus one does not always equal two. Chromosome Res, 17:637-48.

PRESENTATIONS

Prothero KE, Carrel L. Comparative and functional identification of a chromatin insulator that delimits domains on the inactive X chromosome. Medical Genomics Retreat, State College, PA 2010 (Poster)

Prothero KE, Carrel L. An epigenetically demarcated chromatin insulator at a boundary between X chromosome inactivated and expressed transcripts of the *UBA1* gene. American Society of Human Genetics Conference, Philadelphia, PA, 2008 (Poster)

Prothero, KE and Carrel, L. Chromatin Insulators in X Chromosome Inactivation. Graduate Student Research Forum, Penn State College of Medicine, Hershey, PA, 2008 (Oral Presentation)

Prothero, KE and Carrel, L. X Chromosome Inactivated and Expressed Transcripts of the *UBE1* Gene are Delimited by a Chromatin Insulator, Graduate Student Research Forum, Penn State College of Medicine, Hershey, PA, 2007 (Poster)

CONFERENCES

Medical Genomics Retreat, Pennsylvania State University, State College, PA 2010
American Society of Human Genetics, Philadelphia, PA 2008
American Society of Biochemistry and Molecular Biology, New Orleans, LA 2002