XIST AND ITS ROLE IN X-CHROMOSOME INACTIVATION

BY

JENNIFER CHRISTA CHOW

B.Sc Simon Fraser University, 1996

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

THE FACULTY OF GRADUATE STUDIES

Department of Medical Genetics

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 2003

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of MEDICAL GENETICS

The University of British Columbia
Vancouver, Canada

Date AUG 25, 2003

ABSTRACT

Compensation for the two-fold dosage difference in female versus male mammals for X-linked genes involves the formation of an extremely stable heterochromatin structure on one of the two X chromosomes in females. Expression of the untranslated RNA, *XIST*, is required in *cis* for the establishment of the heterochromatic state. It is unclear how this RNA recruits the factors involved in the chromatin alterations, but its unique association with the X chromosome is required for its function. Recent results in mouse have started to elucidate how expression of *Xist* is controlled, including the role of the antisense transcript *Tsix*. To determine the pattern of expression within the *XIST* region through human development I used germ-cell derived, embryonal carcinoma cells as a potential model since some lines do appear to have a limited capacity to recapitulate events in early human development when induced to differentiate in culture. Comparisons between N-Tera2D1 (male embryonal carcinoma cell line) and somatic cells reveal a complex expression pattern in the *XIST* region with the identification of an alternate *XIST* isoform as well as a downstream antisense transcript which appear to be developmentally regulated and specific to N-Tera2D1. Although there were no regions of conservation outside of *XIST* that would point to important regulatory elements, the presence of a developmentally-regulated antisense transcript in humans is suggestive of a possible function. To address this, an *XIST* transformant clone expressing both *XIST* and the overlapping downstream antisense transcript allowed me to analyze the effects of the antisense expression on sense *XIST*. However, unlike mouse *Tsix*, the human antisense does not appear to destabilize or affect sense *XIST* localization. I also examined the factors involved in *XIST* localization and have developed an inducible *XIST* cDNA

construct to determine the domains within the transcript that are essential for localization. Initial deletion constructs have identified two regions within exon 1 that are essential for transcript accumulation and localization. This inducible system will also allow future experiments to address the role of *cis*-DNA elements and transcript levels on the transcript's association with the surrounding chromatin.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BRCA1 = breast cancer 1

BrdU = bromodeoxyuridine

ChIP = chromatin immunoprecipitation

CMV = cytomegalovirus

Dnmt = DNA methyltransferase

Dox = doxycycline

dpy = dumpy

EC = embryonal carcinoma

ES = embryonic stem cells

eed = embryonic ectoderm development

ezh2 = enhancer of zeste homolog 2

FRT = Flp recombinase target

HAT = histone acetyltransferase

HDAC = histone deacetylase

hnRNPC1/C2 = heterogeneous nuclear ribonucleoprotein C (C1/C2)

K = lysine

MCB = macrochromatin body

mH2A = macrohistone H2A

MAR = matrix attachment region

MEF = murine embryonic fibroblast

mix-1 = mitosis and X

mle = maleless

msl = male specific lethal

mof = males absent on the first

MY= million years

ORC = origin recognition complex

PAC = P1 artificial chromosome

PAR = pseudoautosomal region

PNA = peptide nucleic acid

RNAi = RNA interference

rOX = RNA on the X

SDC = sex and dosage compensation

SMC = structural maintenance of chromosomes

SNP = single nucleotide polymorphism

Tet = tetracycline

Xi = inactive X

XIC = X inactivation centre

XIST = Xi specific transcript

YAC = yeast artificial chromosome

CHAPTER 1: INTRODUCTION

Portions of this chapter have been published in the review article:

Chow J.C. and Brown C.J. (2003) Forming facultative heterochromatin: silencing of an X chromosome in mammalian females. Cell. Mol. Life Sci. (in press).

I.1 The evolution of the sex chromosomes and dosage compensation in mammals

Chromosomal mechanisms of sex determination have evolved in many organisms, and often involve a divergent pair of sex chromosomes where the heterogametic sex is either the male, as in mammals and *Drosophila* (XY), or the female, as in birds and snakes (ZW). Based on the graded differences between the Z and W chromosomes in different snake families, Ohno (1967) hypothesized that sex chromosomes originally evolved from a pair of identical autosomes and that the variability in the W chromosome among snake species represented different stages in the gradual breakdown of the chromosome. The mammalian X and Y chromosomes appear to be subject to similar evolutionary processes but are believed to have evolved independently from a different pair of autosomes since they share no homology to the ZW chromosomes of birds and snakes [1].

The acquisition of a sex-determining allele on one chromosome of an originally identical pair has been hypothesized to be the initiating event in sex chromosome differentiation. In mammals, the best candidate for this initiating event is the appearance of the testis-determining gene, *SRY,* on the Y chromosome [2]. With the advent of a sex-determining gene, it became advantageous for other alleles with sex-specific functions to accumulate in the same region. This created a cluster of genes where recombination was repressed such that the entire region could be passed intact to a single sex. In the absence

of recombination, mutations have accumulated on the Y leading to the inactivation of most of the genes on this chromosome [3].

Under these evolutionary pressures, the human Y chromosome has maintained only 60 Mb of sequence with about 50 functional genes compared to the much larger, conserved X chromosome with 165 Mb of sequence and about 1500 genes [4]. These now divergent homologs have, however, retained small regions of homology called pseudoautosomal regions (PARs) at the tips of the long and short arms where they regularly pair and recombine at meiosis. However, most of the genes present on the X have been lost from the Y or are in the process of becoming nonfunctional. A few genes, such as *RBMY* and *SRY* have been modified to perform male-specific functions and have therefore been retained [1].

The evolutionary origins of the eutherian (placental mammal) sex chromosomes can be elucidated through sequence comparisons with the sex chromosomes of other extant mammalian species such as monotremes and marsupials, which diverged from the eutherians about 170 MYA and 130 MYA, respectively. The long arm and pericentromeric regions of the human X appear to be conserved in both monotremes and marsupials and are therefore at least 170 MY old. However, the distal short arm of the human X appears to be a new addition to the eutherian X chromosome since it maps to autosomal regions in marsupials and monotremes [5]. The human X can be further subdivided into four different strata by comparing the relative ages of 19 X-Y gene pairs by nucleotide divergence [6]. On this basis, it has been hypothesized that the evolution of the human sex chromosomes involved at least four rearrangements on the Y

chromosome, possibly inversions, which suppressed recombination and allowed X-Y divergence to proceed.

In contrast to the rapidly degrading Y chromosome, the X chromosome is highly conserved and represents about 5% of the haploid genome in all eutherians [7]. Because the deteriorating Y chromosome leaves the XY male essentially haploid for many of the genes on the X, female mammals have evolved a mechanism to inactivate most of the genes on one of their two X chromosomes to achieve dosage equivalency with males [8]. Ohno therefore proposed that rearrangements involving the X chromosome would be selected against since they would alter the dosage of X-linked genes. The evolution of X-inactivation has been hypothesized to occur on a gene-by-gene basis, triggered by Y-gene decay which leads to increased expression of the X-linked homolog in the males and subsequent inactivation of one copy in females to restore optimal gene dosage [9].

Variability in the X inactivation process among somatic tissues of eutherian and metatherian mammals provides clues about the evolution of X-chromosome inactivation. In general, one of the two X chromosomes in females is chosen to be inactivated and once silenced, this heterochromatic state is heritably maintained through subsequent cell divisions. The initial choice of which chromosome to inactivate is random in human and rodent embryonic tissues; however, in marsupials and in the extraembryonic tissues of rodents and cows, inactivation is imprinted with the paternally-derived X always being inactivated [10-12]. In humans, there does not appear to be a strict imprint on X inactivation in extraembryonic tissues as random or nearly random X inactivation is generally observed [13,14]. The imprinted form of X-chromosome inactivation appears to be less stable as evidenced by genes on the inactivated X in marsupials often

reactivating in culture [15]. This contrasts with the hyperstable nature of inactivation in the somatic tissues of eutherians and may, in part, be due to the absence of methylation at CpG islands [16]. The presence of similar nonrandom inactivation mechanisms in both marsupials and the extraembryonic tissues of some eutherians suggests that imprinted inactivation may be the ancestral mammalian dosage compensation mechanism [10].

Transient inactivation of the single X chromosome is also observed in males during the pachytene stage of spermatogenesis. At this time, the inactivated X pairs with the Y chromosome at the pseudoautosomal regions, forming the sex chromatin body or X-Y body. The function of silencing the single X is unknown. An early suggestion was that male sex chromatin inactivation occurred to silence the X (or Y) from producing products inhibitory to spermatogenesis [17]. Handel has argued that the role may also be to protect against double strand breaks and illegitimate recombination [18]. Another alternative explanation is the prevention of activation of the "synapsis" checkpoint at the end of pachytene by the unpaired X and Y axes [19]. Burgoyne has suggested that the role may be to 'focus' the homology search to the small pseudoautosomal region [20]. Furthermore, the role may be to strengthen sister chromatid cohesion to prevent obligatory distal chiasmata from migrating off the end of the chromosomes [21].

Whatever the function, male sex chromatin inactivation during spermatogenesis shares many features of X-chromosome inactivation in female somatic cells which suggests a common molecular mechanism; however, there are also differences between the processes. Somatic X inactivation in female cells requires *XIST/Xist*, an untranslated RNA that is expressed exclusively from and localizes to the inactive X at interphase. *Xist* expression, and thus the choice of which chromosome to inactivate, is regulated by

another untranslated RNA antisense to *Xist*, called *Tsix* (reviewed in [22]). *Xist* is also expressed during spermatogenesis, but at substantially lower levels [23-25]. However, male mice carrying deletions of *Xist* that prevent somatic X-chromosome inactivation are still fertile [26], and are able to form the XY body in spermatocytes indicating that a functional *Xist* gene is not required for X-chromosome inactivation during spermatogenesis [27,28]. The histone variant, macrohistone H2A1.2, which associates with the somatic inactive X, is also present in the XY body [29,30], as is the BRCA1 protein, which was recently shown to transiently associate with the inactive X in somatic cells [31]. The histone hypoacetylation observed on the somatic inactive X chromosome is not detected on the X/Y body [32], and X-chromosomal DNA from sperm is still competent for DNA transformation, an assay used to demonstrate the inactive X has not been methylated [33]. Therefore, because inactivation of the X chromosome during spermatogenesis has many features distinct from those observed in the somatic tissues it has been suggested that the two processes occur by different mechanisms. However, it has also been proposed that inactivation of the X chromosome during spermatogenesis may be an ancient mechanism important for setting up the chromatin conformation required for imprinted inactivation of the paternal X in female marsupials and the extraembryonic tissues of some eutherians [34].

I.2 Dosage compensation in *Drosophila melanogaster* and *Caenorhabditis elegans*

Species with heteromorphic sex chromosomes have devised diverse mechanisms to compensate for the dosage imbalance between the sexes. In contrast to the chromosome-wide inactivation that occurs in female mammals, males in the fruit fly *Drosophila melanogaster* hypertranscribe genes on their single X approximately two-

fold. This sex-specific hypertranscription is mediated by the male-specific lethal complex which consists of five protein subunits: male-specific lethal proteins (msl-1, msl-2 and msl-3), maleless (mle), and males absent on the first (mof), as well as two untranslated RNA transcripts roX1 and roX2 (RNA on the X) [35]. This entire complex binds to hundreds of sites along the single male X and its binding specificity is dependent upon MSL1, a novel acidic protein, and MSL2 a RING finger protein [36,37]. The complex mediates alterations in chromatin conformation and mediates a two-fold transcriptional enhancement. MOF, a histone acetyltransferase, may have a central role in this conformational change since the male X is highly enriched for an isoform of histone H4 that is acetylated on lysine 16 [38]. It is not clear how the complex recognizes sequences on the X chromosome, but it appears to involve about 35 chromatin entry sites along the X chromosome, two of which are the roX1 and roX2 genes. These sites are hypothesized to act as nucleation centres for the spread of MSL-mediated hypertranscription [39]. The roX RNA transcripts themselves may also have a role in the recruitment or stable association of the complex to the X chromosome since its localization is RNase-sensitive and the histone acetyltransferase component, MOF, has been shown to interact with the roX2 RNA *in vivo* through its chromodomain [40].

Caenorhabditis elegans has come up with yet a different way to deal with dosage differences by partially downregulating transcription from both X chromosomes in the XX hermaphrodite [41]. Mutations in some of the genes involved in dosage compensation (dpy-21, dpy-26, dpy-27, dpy-28, dpy-30) lead to XX-specific lethal or dumpy phenotypes due to the elevated expression levels of X-linked genes. However, DPY-26 and DPY-28 also have a more general role in the segregation of chromosomes

during meiosis in both sexes [42]. Another component, mix-1 (mitosis and X), has an essential role in mitotic chromosome segregation in addition to its role in dosage compensation and is therefore lethal in both sexes [43]. DPY-27 and MIX-1 belong to the SMC (structural maintenance of chromosomes) family of proteins which are required for mitotic chromosome cohesion and condensation in yeast and frogs [41]. Therefore, many of the members of the dosage compensation complex are multifunctional and were probably recruited from more ancestral functions in chromosome structure and condensation to the new role of regulating X chromosome transcription. The targeting of these molecules specifically to the X chromosomes in the hermaphrodite requires SDC-2 and SDC-3 (sex and dosage compensation) which are involved in both hermaphrodite development and dosage compensation [44].

I.3 Mammalian X-chromosome inactivation

The epigenetic silencing of the mammalian X chromosome includes acquisition of many of the features associated with heterochromatin. The timing of the events of inactivation as outlined in Figure 1-1 was determined in female mouse ES cells induced to undergo differentiation [45]. I will discuss the process of inactivation in embryonic lineages or somatic cells, and will note differences from the situation in extraembryonic tissues, or spermatogenesis.

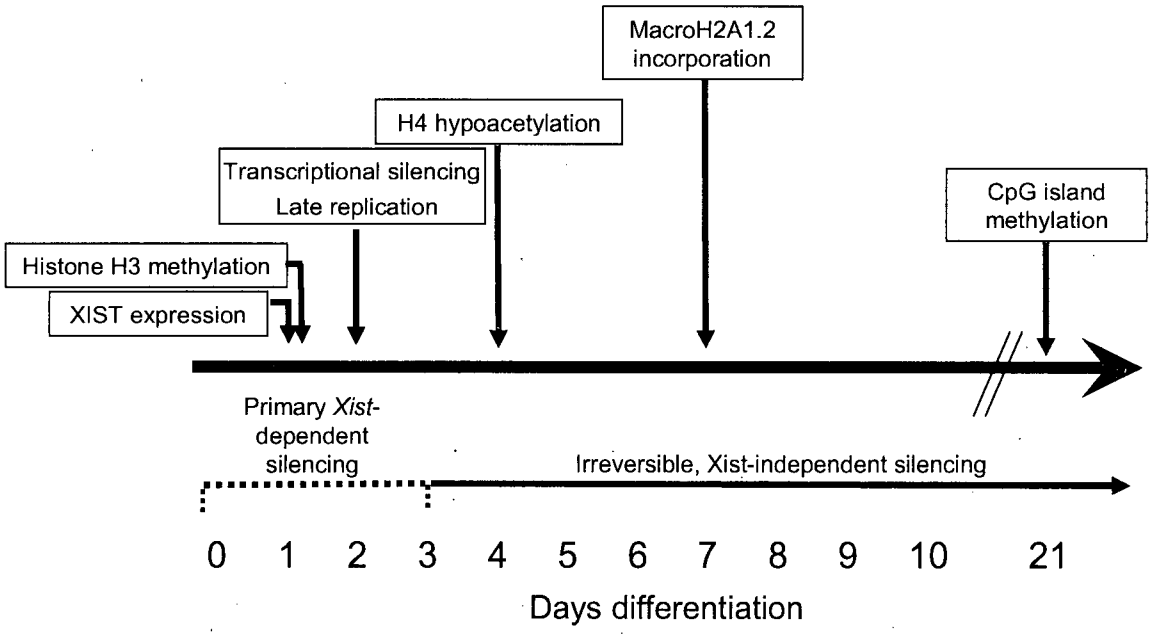


Figure 1-1. Features of inactive heterochromatin are acquired in a sequential fashion in differentiating female mouse ES cells. The approximate timing of the events following induction of differentiation is shown below the arrow [45].

I.3.1 *XIST* Expression

One of the first events in the inactivation cascade is the expression of a transcript exclusively from the inactive X, called the inactive X-specific transcript, or *XIST*. *XIST*

was the first gene to be mapped within the master control region on the X called the X-inactivation centre (XIC), which is required in *cis* for inactivation to occur. *XIST* encodes an untranslated, functional RNA molecule that remains in the nucleus and associates with the inactive X. For human *XIST*, this association is retained only during interphase; however in mouse, *Xist* continues to associate with the X until early metaphase [46-48]. Although the *XIST* transcript remains nuclear, it is still spliced and polyadenylated and is highly heterogeneous with many different isoforms arising from alternate initiation sites, termination sites and splicing variants.

XIST has been identified and sequenced in mouse, voles, cow and humans [49,50]. Comparative analyses between species show that there is a fairly low degree of sequence constraint within the *XIST* primary sequence, with exonic sequences having 66% identity between mouse and human, and 62% identity between mouse and cow [50]. The basic intron-exon structure of the gene, however, has been largely conserved prompting suggestions that structural features of the transcript rather than the primary sequence are important for its function. One of the unique conserved features of the *XIST* transcript is the presence of tandem repeats which make up over one-third of the *XIST* sequence in rodents, humans, and cows [49,50]. Six of these repeats (A-F) have been identified across the *XIST* gene and repeat A, at the 5' end of exon 1 shows the highest conservation in all species examined [46,48,49], see Figure 1-2. The presence and conservation of these repeats are highly suggestive of a functional role, possibly in the binding of specific proteins or in the formation of particular secondary structures.

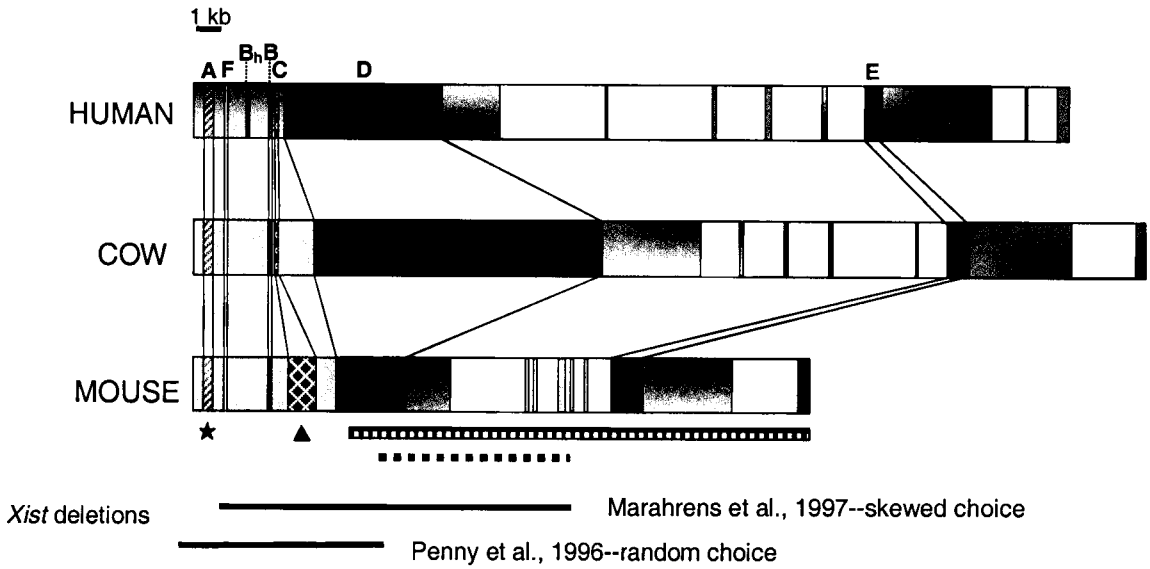


Figure 1-2. Conserved repetitive elements in human, cow and mouse *XIST*. The grey blocks indicate the exonic regions. The repetitive elements are labeled from A-F, with B_h representing a human specific element that may have arisen from an insertion into the B repeat in humans. The conserved A-repeats, marked with the red star, are required for silencing in mouse and also contribute to *XIST* transcript localization [51]. PNA probes directed against the C-repeats (triangle) in differentiating female mouse ES cells also disrupt transcript localization [52]. The rectangle (blue vertical hatching) contains regions(s) required for macroH2A localization [53]. The solid black lines represent *Xist* deletions that disrupt *Xist* function and lead to nonrandom X-chromosome inactivation. Only the deletion spanning exons 1-5 leads to primary nonrandom inactivation implicating the region represented by the dotted line in X-chromosome choice.

Xist expression patterns during the initiation of inactivation have been studied in mouse preimplantation embryos and in ES cells. Prior to inactivation, *Xist* is expressed from both X chromosomes in females and the single X in males as a low-level unstable transcript that can be seen as small punctuate signals by RNA-FISH. Upon differentiation, *Xist* is upregulated on one of the two X chromosomes, forming a large stable signal that coats the X that is subsequently inactivated; whereas on the future active X, *Xist* expression is silenced [54,55]. Because accumulation of stable *Xist* RNA is one of the first events in the inactivation cascade, starting at about day 1 of differentiation, it is thought to recruit factors required for the silencing of the chromosome [56]. *Xist* is also expressed in mouse extraembryonic tissues where X-inactivation is imprinted and the paternally-derived X is preferentially inactivated. *Xist* is expressed exclusively from the paternal X and, as in the cells of the embryo proper, expression precedes inactivation. Therefore, based on these dynamic changes in expression, *Xist* appears to be involved in the initiation of both random and imprinted inactivation [22]. *XIST* expression prior to inactivation is also observed in humans [57,58] and cows [59].

Mouse knockouts have demonstrated that *Xist* is necessary for inactivation, as chromosomes that contain deletions removing the *Xist* promoter or interstitial regions of the gene are unable to be inactivated [26,60]. In addition, *Xist* alone is sufficient for inactivation since an inducible *Xist* cDNA construct introduced into ES cells can induce long-range transcriptional silencing as well as late-replication and histone H4 hypoacetylation of surrounding chromatin [51]. However, the stability and reversibility of *Xist*-induced silencing is dependent on developmentally-regulated factors since its

ability to silence depends on the time at which it is expressed during differentiation (Figure 1-1). In undifferentiated ES cells, the *Xist* transgene can induce silencing, but this "primary" silencing is reversible and dependent on continued *Xist* expression. It is also not associated with other features of inactivation such as histone hypoacetylation and late replication timing. After 72 hours of differentiation, silencing becomes irreversible and *Xist*-independent and other features of the inactive X are acquired. There appears to be a critical window in the first 48 hours of differentiation when *Xist* can initiate silencing, after this time the cell becomes refractive to *Xist* expression. However, there has been a report of *XIST*-mediated silencing in human somatic cells suggesting that this developmental requirement may not be so strict in some cell types [61]. Alternatively, this difference may point to species or experimental differences.

Functional domains have been identified in mouse knockouts using a series of deletions introduced into an *Xist* inducible cDNA construct integrated into the *Hprt* locus on the X chromosome and expressed in ES cells [53]. The conserved A-repeats at the 5' end of *Xist* are required for silencing and also contribute to the localization of the transcripts to the chromosome, but are not absolutely required for the latter function. Transcripts deleted for the A repeats were able to localize to the chromosome, but could not silence the surrounding chromatin, demonstrating that localization of the transcript is not sufficient for silencing. Computational analysis of the A-repeats predicts that the RNA folds into two stem-loops which are repeated 7.5 times in mouse. Specific mutations in these repeats show that the RNA secondary structure, not the specific sequence, is important since mutations that retain the stem loops retain silencing activity. The activity of these repeats is also copy number dependent, requiring at least 5.5 repeat

units for silencing, but position-independent since moving the repeats to the 3' end of the transcript had no effect on function [53]. These conserved stem loops may provide binding sites for specific protein factors that mediate chromatin conformational changes that lead to inactivation. Not only is *Xist* required for silencing, but sequences within *Xist* also appear to be involved in the choice of which chromosome will be inactivated. Although some *Xist* deletions lead to nonrandom inactivation patterns due to secondary selection [60], others appear to result in primary nonrandom inactivation and implicate specific regions within *Xist* in the initial choice of which chromosome will be inactivated [26] (Figure 1-2).

The ability of the *XIST* transcript to associate in *cis* with the chromosome from which it is transcribed is essential for the initiation of silencing. *Xist* deletion constructs created by Wutz et al. [53] have identified regions within *Xist* that are involved in the association of the transcript to the chromosome and found that localization requires multiple domains throughout the *Xist* transcript that work co-operatively and in a functionally redundant manner (Figure 1-2). From these various constructs it was concluded that although localization does not necessarily lead to silencing, silencing does require a localized transcript. A similar result was obtained using Peptide Nucleic Acid (PNA) antisense probes against the conserved C repeats which disrupted the localization of *Xist* and the subsequent Barr body formation in differentiating female ES cells [52].

The formation of functional protein-RNA complexes are likely important for *Xist* action and localization. *Xist* is necessary for the localization of the histone variant macroH2A to the inactive X. MacroH2A will be discussed further below, but association with the inactive X apparently involves the 3' region of *Xist* (Figure 1-2) [53]. A few

additional proteins interacting directly or indirectly with *XIST* have been described, but so far the role of these proteins in X-chromosome inactivation has not been determined. BRCA1, a breast and ovarian tumor suppressor, has recently been shown to co-localize with the inactive X [31]. BRCA1 has been implicated in many cellular functions, and through the analysis of *BRCA1* -\- cancer cells, it appears to also support the localization of *XIST* and the formation and/or stabilization of the silenced state of the inactive X. A general RNA-binding protein hnRNPC1/C2 has been shown to interact with the region of the conserved repeats in the 5' end of *XIST in vitro* [62]. Although this interaction does not appear to be specific for the *XIST* transcript, specialized heteronuclear protein isoforms may be important for *XIST*'s unique transcript processing and localization requirements [56]. In addition, antibodies in the serum of an autoimmune patient were found to recognize an as yet unidentified protein enriched on the inactive X [63]. The *Xist* transcript also appears to be important for the transient association of the Polycomb group proteins, Eed/Ezh2 with the inactivating X chromosome [64]. The Eed/Ezh2 complex has an important role in heritable gene silencing through chromatin modifications and its association with the inactive X suggests that its activity may be recruited to the chromosome by the *Xist* transcript.

The importance of *XIST* expression in the initiation stage of inactivation has clearly been established and its continual synthesis in somatic cells argues that it may also have a role in maintenance. Its role in these later stages, however, may be more redundant because although *XIST* expression appears to contribute to the stability of silencing [65], loss of *XIST* does not lead to reactivation of the chromosome in somatic cells [66-68]. An understanding of the initiation of inactivation will require

understanding how *XIST* expression is regulated, and I will briefly review what is known about the transcriptional regulation of *XIST*, the role of an antisense transcript and chromatin changes in regulation, and *trans*-acting factors potentially involved in *XIST* regulation.

I.3.1.i) *XIST* Regulation

Transcriptional regulation

Because changes in *Xist* steady-state levels have important consequences in the inactivation process, identifying elements that regulate its expression will be key to understanding the mechanism of X-chromosome inactivation. An increase in mouse *Xist* RNA steady state levels is important in the initiation of inactivation and appears to involve a change in transcript stability, which goes from a half life of ~30 min prior to inactivation to ~5 hours after differentiation [54,55]. Although transcriptional run-on experiments have determined that an increase in transcriptional activity does not appear to be involved in *Xist* upregulation at this stage, increases in transcriptional levels do appear to affect *Xist*'s stability and its ability to coat and inactivate surrounding chromosomal material. High levels of *Xist* expression from an inducible promoter are able to initiate inactivation in undifferentiated ES cells which suggests that stabilization may not require developmentally specific stabilization factors [51,53].

To identify developmentally-specific transcription factors that modulate *Xist* transcriptional activity during inactivation, elements within the *Xist* promoter region have been characterized. Mutational analyses in promoter-reporter gene constructs have defined the minimal promoter to be about -81 to +1 in mouse and -93 to +31 in humans relative to the transcriptional start sites [69,70]. The promoter regions in both species

contain binding sites for ubiquitous transcription factors and are constitutively active with no sex-specific activity which suggests that elements responsible for the developmentally-regulated expression patterns of the *XIST* transcript must lie outside this minimal region and act to silence expression on the X chosen to remain active and upregulate expression on the X to be inactivated. Intriguingly, a basepair change in the promoter region of the human *XIST* gene has been found in individuals with non-random inactivation [71,72], although its relevance is unclear since the change is also found in family members who do not have extremely skewed inactivation.

Developmentally-regulated *Xist* expression is dependent on differential methylation at CpG sites in the promoter region and at the 5' end of the first exon of *Xist*. Methylation at these sites correlates with the activity of the gene and precedes the onset of *Xist* expression, which suggests that it has an active role in controlling expression [73]. In addition, targeted disruption of DNA methyltransferase 1 (*Dnmt1*) results in inappropriate expression of *Xist* from the single X in males and from both X's in females upon differentiation. However, although methylation seems to be important for *Xist* regulation after differentiation, the low-level *Xist* expression prior to inactivation is unaffected by the *Dnmt1* mutation [74,75]. Human *XIST* also shows differential promoter methylation [76], and in somatic cells demethylation seems to be sufficient to induce expression from the active X chromosome [77,78].

Tsix* – *antisense regulation?

An antisense transcript, *Tsix*, has been identified in the *Xist* region and has been shown to be an important regulator of *Xist* expression during the initiation stages of inactivation. *Tsix* encodes a heterogeneous transcript and its main initiation site lies 13

kb downstream of *Xist* and transcription spans over 40 kb, overlapping the entire *Xist* gene. *Tsix* is also spliced and, like *Xist*, has many different transcript isoforms with an alternate minor start ~28 kb downstream of *Xist* and several different splicing variants [79,80]. Interestingly, the spliced forms eliminate all overlap with *Xist* except for a 2 kb region at the 5' end of *Xist* which includes the conserved repeats important for silencing.

Tsix expression, like *Xist,* goes through dynamic changes early in development and its expression pattern suggests a role as a negative regulator [79,81,82]. Prior to inactivation both *Xist* and *Tsix* are co-expressed on both X's in females and the single X in males, appearing as pinpoint signals by RNA-FISH. Upon differentiation, *Tsix* expression is retained on the future active X, repressing *Xist* upregulation; whereas on the future inactive X, *Tsix* is silenced and *Xist* accumulation then leads to silencing (Figure 1-3). However, *Xist* regulation may be more complex since the extinction of the *Tsix* signal does not always precede *Xist* upregulation, with some nuclei expressing both an accumulating *Xist* signal as well as a pinpoint *Tsix* signal [83].

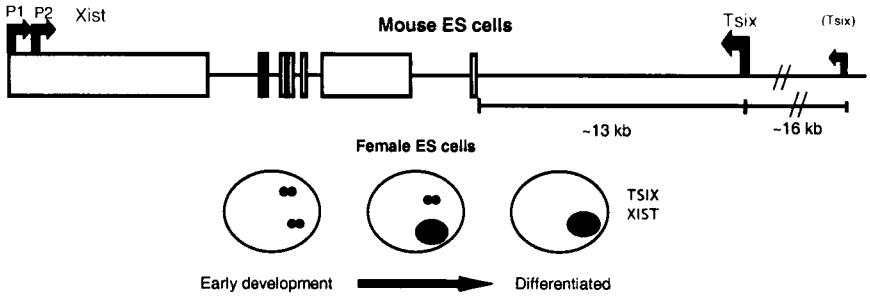


Figure 1-3. RNA-FISH on differentiating female ES cells show that *Xist* and *Tsix* undergo dynamic changes in expression patterns. Prior to differentiation, both X chromosomes co-express low levels of *Xist* (blue) and *Tsix* (red), both appearing as small pinpoint signals. Upon differentiation, the future active X retains *Tsix* expression and represses *Xist* expression whereas on the future inactive X, *Tsix* is silenced

and *Xist* expression is upregulated and coats the chromosome. The active X then silences *Tsix* expression. Both *Xist* and *Tsix* have alternate initiation sites indicated with the blue arrows (P1 and P2) for *Xist* and the red arrows for *Tsix*. The significance of these alternate isoforms is unclear.

Targeted deletions and insertions into the *Tsix* gene have further substantiated its role as a negative regulator of *Xist* and provide clues to the mechanism of its action. Targeted deletions of the *Tsix* main promoter disrupt random choice and lead to upregulation of *Xist* and inactivation on the deleted chromosome [79,82,83]. In addition, ectopic expression of *Tsix* through the targeted insertion of an inducible or constitutively active promoter also leads to suppression of *Xist* upregulation and inactivation [84,85]. Active transcription in the region appears to be important for its function since insertion of transcriptional stop signals also lead to preferential inactivation of the mutated chromosome [84].

Based on *Xist* and *Tsix* expression patterns and the results of *Tsix* knockout and ectopic promoter knock-in experiments, several mechanisms of repression have been proposed [80,81,86]. It is possible that the antisense transcript itself is not important, and *Tsix* may regulate *Xist* through an enhancer competition-type model where activity at a downstream promoter outcompetes *Xist* for a shared enhancer (analogous to early models for the proposed action of the oppositely imprinted genes *H19/IGF2*) [87]. However, insertion of a transcriptional stop signal that truncates the *Tsix* transcript leads to nonrandom inactivation suggesting that simply activating the promoter is not sufficient to suppress *Xist* function and that either the antisense transcript itself or the act of transcription through the *Xist* locus is important [84]. In these transcription-dependent mechanisms, *Tsix* inhibition of *Xist* may or may not require the RNA transcript. A

mature, processed antisense transcript may not be important and the simple act of transcription may function to either modify surrounding chromatin, or inhibit the movement of RNA polymerase in the sense direction. If the transcript is important, *Xist/Tsix* RNA duplex formation could lead to the degradation of the *Xist* transcript or mask important domains on the *Xist* transcript preventing protein interactions and proper function. Interestingly, splicing of the *Tsix* transcript eliminates almost all overlap with *Xist* except in the 5' region of exon 1 where the conserved repeats required for silencing reside [79,80]. Quantitative analyses of *Xist* and *Tsix* transcript levels in undifferentiated mouse ES cells revealed that the *Tsix* RNA is present at much higher levels than *Xist* (>10-100 fold molar excess) supporting transcript-dependent models where sufficient levels of the antisense transcript are required to bind and block the activity of all the *Xist* transcripts [80].

Antisense transcription has also been detected at the human *XIC* (X inactivation centre) and conserved features may provide clues about the mechanism of *Tsix* regulation [88,89] (see Chapter 5). The human *TSIX* transcript only partially overlaps *XIST* and does not extend into the 5' region; therefore, unlike mouse, there is no overlap with the conserved 5' repeats. In addition, the presence of the antisense transcription in humans does not preclude stable, localized *XIST* expression, which may mean that the human homolog has lost its ability to negatively regulate *XIST* expression. However, 30-60% of *Tsix* transcripts in mouse remain unspliced and transcript levels are highest at the 5' end of *Tsix* where the human antisense is also expressed which may be suggestive of a functional role for transcription in this region [80].

Chromatin structure flanking Xist

Changes in chromatin modifications 5' of *Xist* coincide with the initiation of inactivation and thus implicate flanking chromatin in the regulation of *Xist* expression. In particular, the existence of regulatory elements upstream of *Xist* has been suggested by several studies. Johnston et al. [90] originally proposed that stabilization of the *Xist* transcript during the initiation of inactivation may involve a switch from an upstream promoter (Po); however, the presence of several ribosomal protein pseudogenes in the Po region [91], and the *Tsix* antisense transcript which extends upstream of *Xist* may have confounded the original observations of Po expression. Still, the region upstream of *Xist* does appear to contain important regulatory sequences since transgenes deleted for the region, although they express unstable transcripts in undifferentiated ES cells, are unable to express stable *Xist* when differentiated [92]. Several histone modifications also point to important elements in the region. A unique hotspot of histone H3 lysine methylation has been identified 40-100 kb upstream of *Xist* and has been proposed to act as a nucleation site for heterochromatin spread [93]. In addition, a region of histone H4 hyperacetylation specific to undifferentiated female ES cells has been identified up to 120 kb upstream of *Xist*. Upon differentiation, the region becomes hypoacetylated and this change has been hypothesized to be important for *Xist* stabilization and progression of inactivation [94].

***Xist* regulation through *trans*-acting factors**

Trans-acting factors have been hypothesized to be involved in the X-inactivation pathway and the choice of which X to inactivate. Diploid mammalian cells always retain one X chromosome active, which has led to models suggesting that an autosomal factor binds a counting element in the XIC (reviewed in [22]). A 65 kb deletion 3' of *Xist*

abolishes the retention of an active X (even in male cells), and thus may define the location of binding of such a counting element [95]. One way that these hypothetical factors could carry out these functions is to regulate *XIST* expression. Binding sites for the ubiquitous chromatin insulator and transcriptional regulator CTCF have been found in the *Tsix* promoter region and implicate it as a candidate *trans*-acting factor for X chromosome choice. It has been proposed that CTCF could mark the active X by *Xist* suppression through *Tsix* activation or by blocking access to putative enhancers [96].

Through mutagenesis screens two, as yet unidentified, autosomal loci have been implicated in X chromosome choice and may thus also be involved in *Xist* regulation [97]. In a mouse phenotypic screen, females were identified in the progeny of ENU-mutagenized males and screened for changes in the expected X-inactivation ratios. Two females with altered X-inactivation patterns were identified and the mutations were found to segregate in an autosomal dominant manner. Because the alterations in X inactivation ratios are observed early in development, they are thought to represent mutations affecting primary choice rather than secondary cell selection. Different strains of mice show differential tendencies to keep one X active, and this locus has been called the X controlling element (Xce) [98]. Xce maps close to, but downstream, of the murine *Xist* locus [99].

I.3.1.ii) Mechanism of *Xist*-Mediated Silencing

While there is substantial evidence supporting a role for *XIST* in establishing X chromosome silencing, there is growing evidence that an RNA-mediated silencing function may not be unique to *XIST* and X-chromosome inactivation. Naturally occurring antisense transcripts are found both in *cis* and *trans* throughout the human genome. A

recent survey shows that ~2% of mRNA have associated antisense expression [100]. In particular, antisense transcripts seem to be enriched in imprinted domains, which also show epigenetic silencing of one allele, although based upon parental origin. While *Tsix* apparently regulates *Xist* expression, antisense expression is not required for the establishment of inactivation, as a cDNA construct is able to establish silencing [51].

Two untranslated RNA transcripts roX1 and roX2 (RNA on the X) are involved in the sex-specific hypertranscription in male *Drosophila melanogaster,* functioning in conjunction with the male-specific lethal complex consisting of msl-1, msl-2, msl-3, mle, and mof [35]. Recognition of the X chromosome appears to involve about 35 chromatin entry sites along the X chromosome, two of which are the roX1 and roX2 genes, which are hypothesized to act as nucleation centres for the spread of MSL-mediated hypertranscription [39]. MOF, a histone acetyltransferase thought to have an important role in the chromatin conformation changes mediated by the msl complex, may be targeted to the X chromosome by the roX2 RNA since its specific localization to the X chromosome is RNase-sensitive [40].

In addition to their roles in dosage compensation, noncoding RNA transcripts are emerging as crucial components in directing chromatin modifications at other heterochromatic regions. The first link between RNA and chromatin modification was observed in plants where short double stranded RNAs could direct the methylation and silencing of homologous DNA sequences [101]. Recent findings in the fission yeast *Schizosaccharomyces pombe* have implicated the RNA interference (RNAi) pathway and RNA transcripts derived from repetitive sequences at the centromere and the mating type region in the assembly of heterochromatic chromatin [102,103]. A role for an RNA

component in the higher-order structure in pericentromeric heterochromatin also appears to be conserved in mammalian cells since RNase treatment of permeabilized mouse cells results in loss of the histone H3 lysine (K) 9 methylation and HP1 binding normally present at the centromere [104]. The precise role of the RNA transcripts in directing chromatin changes remains unclear; however, models have hypothesized that double-stranded RNA transcripts processed through the RNAi pathway recruit chromatin-modifying proteins like histone methyltransferases or deacetylases through their chromodomains to nucleate heterochromatin formation [102,103]. The chromodomain has been reported to interact directly with RNA [40]. At the *Xist* locus, there has so far been no evidence that the transcripts are processed by the RNAi machinery; however, *Xist* expression is required for the establishment of inactive heterochromatin which involves the recruitment of the Polycomb group proteins Eed and Ezh2. The Eed/Ezh2 complex has recently been shown to localize transiently to the inactive X during both random and imprinted X-chromosome inactivation, and is essential for histone methylation at lysine (K) 9/27 [64,105].

While knockout and transgenic experiments in mice have shown that *Xist* is sufficient to initiate silencing, the mechanism by which *Xist* expression recruits the ensuing features of heterochromatin requires further elucidation. As the initiating event in inactivation, it seems reasonable to expect that *Xist* would be a constant feature of mammalian dosage compensation; however the marsupial homolog has not yet been identified, perhaps due to a low level of sequence conservation. *XIST/Xist* is also expressed during spermatogenesis [23-25], while *Tsix*, if expressed, is expressed at substantially lower relative levels than seen in undifferentiated ES cells [27,28].

However, male mice carrying deletions of *Xist* that prevent somatic X-chromosome inactivation are still fertile [26], and apparently able to undergo meiotic sex-chromosome inactivation [28]. Thus the role of *Xist* in spermatogenesis is unclear.

I.3.2 Replication Timing

That heterochromatin is late replicating has been known for decades [106], and there is now evidence to suggest that this late replication actually contributes to the silencing of DNA, rather than simply being the result of inactive chromatin. Asynchronous replication of the immunoglobulin and T cell receptor genes is established at the time of implantation, and it is the early-replicating allele that tends to be the one to undergo rearrangement and subsequent expression (reviewed in [107]). Furthermore, a recent study demonstrated that exogenous DNA injected into cells early in S phase is more likely to be transcriptionally active than DNA injected in late S phase, supporting the conjecture that DNA synthesized in late S phase is incorporated into a repressive chromatin state [108].

In order to replicate the entire mammalian genome within S phase, there are thousands of origins replicating individual replicons on the order of ~100 kb in size. Replication patterns generally correspond to banding patterns, demonstrating that the timing of replication of several replication origins is coordinately regulated to yield a like-replicating domain, whose boundaries coincide with R/G banding boundaries [107]. The identification of individual metazoan origins has been difficult, and no consensus sequence has been identified, however origins in mammalian chromosomes have been suggested to be enriched in CpG islands [109]. It has also been suggested that epigenetic chromatin structure may contribute to the definition of mammalian origins, an argument

supported by the association of the Origin Recognition Complex (ORC) with HP1 and histone acetyltransferase in humans [110]. While several studies have identified origins on the X chromosome [111,112], only recently has it been shown for origins in the *G6PD* and *HPRT* genes that the same origin is used on both the active and inactive X chromosomes [113].

Evidence for late replication of the X chromosomes was first obtained by observation of the whole chromosome (e.g. [114,115]), and later extended by the study of individual genes (see Table 1-1). The introduction of a combination of bromodeoxy-uridine (BrdU) incorporation and Giemsa staining allowed the more precise definition of replication patterns on a chromosome [116]. While variability is observed between cells of the same type, as well as between cell types, bands Xp22, Xp11, Xq13, and Xq26 are often earlier replicating than the rest of the inactive, late-replicating X chromosome [117-120]. It was suggested that at least some of these early replicating segments on the inactive X chromosome may contain genes escaping inactivation [121]. In some situations the inactive X can be early replicating, as was shown for a mouse thymic lymphoma cell line. Interestingly, the inactive X in this line could be induced to become late replicating by fusion with EC cells, suggesting that replication timing was under the control of *trans*-acting factors [122]. In addition, during mouse development the first evidence of a differential replication pattern for one X (allocyclic replication) is a shift to earlier replication [123].

Studies of X chromosome replication timing for specific regions or loci have used a variety of approaches (see Table 1-1) with the general conclusion that silent genes/regions tend to replicate late. Genes silenced in a tissue-specific manner (e.g. F8,

F9) are generally late replicating, although show somewhat earlier replication on the active X chromosome [124]. A common approach to analyzing replication timing of individual loci has been to use FISH, under the assumption that probes showing two signals (double dot) have replicated, while ones showing a single dot have not yet replicated, so that replication asynchrony is evidenced by the proportion of cells showing single/double signals. Using this technique, the silent *XIST* on the active X chromosome appeared to be late replicating [125,126]; however, an approach based on the isolation of newly replicated DNA that had incorporated BrdU showed the *XIST* locus on the inactive X to be late-replicating [127]. These later authors suggested that the FISH double dot would reflect not only replication but also chromatin conformation allowing separation of the replicated chromatids. The early replication of the silent *XIST* allele on the active X chromosome is further evidenced by a direct quantitative measurement using allele-specific PCR following flow cytometry [128]. Interestingly, the *FMR1* gene is a late-replicating domain that replicates even later when on the inactive X, or silenced by expansion of the 5' CCG repeats in fragile X syndrome. The distal boundary of this domain is 350-600 kb 3' of *FMR1* in proximal q27, supporting the theory of co-ordinate activation of adjacent replicons. Apparently at least 3 replicons can be delayed in their replication after expansion of the FRAXA trinucleotide repeat [129].

Table 1-1. Summary of replication patterns at X-linked loci.

Locus/Region	Result	Technique	Reference
HPRT1 F8	Inactive X late Limited asynchrony	Southern analysis following BrdU incorporation	[124]
HPRT1 FRAXA RPS4X ZFX XIST	Asynchronous Asynchronous Synchronous Synchronous Inactive X early	FISH	[125]
F8 FMR1 XIST	Both X's late Later when mutant Active X late	FISH	[126]
XIST HPRT1 FMR1 F9 SLC16A2 PGK1 IDS G6PD	Inactive X late Inactive X late Inactive X very late Inactive X later Inactive X late Inactive X late Inactive X late Inactive X late	Flow cytometry of BrdU labeled DNA	[127,130,131]
PGK1 XIST	Xa early Xa early	Quantitative allele-specific PCR after flow cytometry.	[128]

Late replication is a consistent feature of silent heterochromatin, and appears to be established early in the silencing of the X chromosome. Replication asynchrony can be disrupted by factors influencing DNA methylation or chromatin modifications; however, a late-replicating X chromosome is observed in marsupials [132], suggesting that it is a conserved event in the inactivation process.

I.3.3 Histone modifications

Over the past few years there has been a rapid increase in information describing the role of histone modifications in controlling gene expression, including the suggestion of a histone regulatory 'code' superimposed upon the genetic code (reviewed in [133]). The histones, which form the protein substructure of chromatin, are small highly basic

proteins and their amino-terminal 'tails' can be modified by acetylation, phosphorylation, methylation, ubiquitination, and ADP-ribosylation (reviewed in [134]). In addition to this plethora of modifications, there are interactions between modifications at different amino acid sites of the different histones (reviewed in [135]). Studies of histone modifications on the inactive X chromosome have focused upon acetylation and methylation. As most analyses of histone modifications are based on reactivity to modification-specific antibodies, caution must be used in interpreting results because individual antibodies may not recognize all levels of modification (for example, mono-, di- and tri-methylation), or may be specific to combinations of modification at adjacent residues (reviewed in [135]).

I.3.3.i) Histone methylation

The inactive X has been reported to be methylated at K9 of histone H3 [93,136] and hypomethylated at K4 of histone H3 [137], similar to modifications seen in pericentromeric heterochromatin. However, recently the inactive X has been shown to display a more distinctive K27 tri-methylation accompanied by Eed/Ezh2 binding as an early, and transient, event in the establishment of inactivation [64,105]. The *C. elegans* homologs of these genes (MES-2 and MES-6) are needed for germline silencing [138].

In somatic cells, histone methylation is most dramatic at the promoter, but can be observed in the body of silenced genes [93] and the differential methylation is also maintained in rodent/human somatic cell hybrids [137]. There are regions of the inactive X that seem to stain positively for antibodies to methylated H3 (K4). These include the Xp pseudoautosomal region, Xp11 and Xq25-26 [137] and may reflect regions with genes that escape X-chromosome inactivation [139].

Many histone methyltransferases have been identified recently, and the SET domain, which is the canonical sequence found in most histone methyltransferases, is found in over 70 genes [134]. The Polycomb group protein Ezh2 contains a SET domain, which presumably accounts for the initial methylation of the inactive X. However, the association of Eed/Ezh2 with the inactive X is transient and appears only during the initiation stages of inactivation, which suggests that other methyltransferases are likely involved in the maintenance of methylated chromatin. The first histone lysine methyltransferase identified was from the SUV39H family that methylates histone H3 at K9. This methylation allows binding of the chromo domain of heterochromatic HP1 proteins, which form complexes to recruit the SUV39H methyltransferases, both of which are enriched at pericentromeric heterochromatin, providing a mechanism for propagation of heterochromatin [140]. However, the histone H3-K9 methylation of the inactive X occurs in the absence of SUV39h [134]. There is also no enrichment of HP1 alpha, beta or gamma on the inactive X in somatic cells [136]. Nevertheless, in the X/Y body seen in spermatogenesis, Suv39h2, a testis-specific isoform of the histone methyltransferase, and M31, the HP1 homologue, are recruited to the silenced chromatin [21].

I.3.3.ii) Histone acetylation

Histone H4 acetylation is generally associated with transcriptionally active chromatin, and immunostaining of chromosomes with anti-acetylated histone H4 mimics an R banding pattern, and is maintained through the cell cycle. The inactive X stains very poorly with such antibodies, although the mouse inactive X chromosome shows a banding pattern, with three bands visible [141]. Similar regions are visible on the human

inactive X after treatment with the histone deacetylase inhibitor, butyrate; and these include the Xp pseudoautosomal region, the proximal short arm (Xp11.2-Xp11.3) and Xq22 [141]. Antibodies to acetylated histone H3 also showed the inactive X to be underacetylated relative to the active X and autosomes. Again, however, bands were observed at the Xp pseudoautosomal region and Xq22 [142]. Hypoacetylation of histone H2A is also seen on the inactive X [142].

Chromatin immunoprecipitation (ChIP) is a procedure involving cross-linking of DNA with chromatin proteins before immunoprecipitation, thereby co-precipitating DNA with antibodies to chromatin-associated proteins. Using ChIP with antibodies to acetylated histone H4, Gilbert et al. demonstrated that H4 acetylation at promoter regions correlated with expression in hamster hybrids retaining the human active or inactive X chromosome. Furthermore, the highest levels of acetylation were found at promoters relative to even a few kb away [143]. In contrast, Morrison and Jeppesen showed underacetylation of a polymorphic marker on the inactive X in a lymphoblast line with non-random inactivation over 20 kb from the promoter [144]. While this could reflect differences acquired in hybrids, the existence of promoter-specific effects is supported by the association of histone acetyltransferases (HATs) and histone deacetylases (HDACs) with the transcriptional machinery. In contrast, cytological observations show that, even in metaphase chromosomes, histone acetylation reflects the R-bands where the majority of constitutively expressed genes are located and shows a global restriction from the inactive X, supporting that differences beyond promoter regions also exist [144]. Thus it seems likely that the strongest difference is found at promoter regions, but that there are

also more globally distributed differences that give rise to the cytologically observed differences in acetylation levels.

While histone acetylation is maintained as a balance between the activity of HATs and HDACS, there are no known histone demethylases, so histone methylation may be a more stable epigenetic mark that is maintained, once established, until replication or replacement of the histone. The balance between various modifications is demonstrated by the negative relationship between methylation of histone H3 K4 (which correlates with transcriptional activity) and methylation at K9 (which correlates with silencing), as well as the fact that K9 of histone H3 can only be acetylated or methylated [135]. Hypoacetylation of histone H4 is also observed for the inactive X chromosome of marsupials, suggesting that it is a primordial feature of mammalian X inactivation [132]. Interestingly, histone hypoacetylation observed on the somatic inactive X chromosome is not detected on the X/Y body [32], while histone methylation is, as discussed above.

I.3.4 Macrohistone H2A and the Macrochromatin Body

In addition to modifications of the core histones, there is also preferential incorporation of a variant histone into the chromatin of the inactive X chromosome. This histone H2A variant is three times the size of conventional H2A, and is therefore called macrohistone H2A1 (mH2A1). The N-terminal third of the gene shows 65% amino acid identity to histone H2A, while the novel C-terminal portion has several putative protein domains, including most notably a putative phosphoesterase domain that is found in many proteins encoded by RNA viruses. This domain is shared by another histone H2A variant, macrohistone H2A2 (mH2A2) that shows ~ 70 % amino acid identity to mH2A1. The non-histone domain of mH2A1 has also been reported to show homology to a gene

overexpressed in aggressive lymphoma, *BAL* [145]. The mH2A1 gene (*H2AFY*) is an intron-containing gene that maps to human chromosome 5q31-32, and encodes two alternative splice variants, mH2A1.1 and mH2A1.2, which differ by the utilization of an alternate pair of sixth exons in the non-histone, leucine-zipper region. MH2A1 is present in both male and female cells; however, in female somatic cells the histone variant is condensed into a 'macrochromatin body' or MCB that colocalizes with the inactive X chromosome [146]. The mH2A2 gene (H2AFY2) maps to human chromosome 10q22.3, and shares a very similar genomic organization to mH2A1, suggesting a relatively recent evolutionary origin by gene duplication. This variant is enriched on the inactive X chromosome in female somatic cells [147], although other histone variants such as H2A.X, H2A.Z or H2A-Bbd (Barr body-deficient) are not enriched, or are excluded from, the inactive X [148].

The N-terminal histone homologous domain is sufficient to direct MCB formation, but the non-histone domain could also direct MCB formation when fused to H2A or H2B domains [149]. The function of mH2A in X-chromosome inactivation is unknown; however, the non-histone region of mH2A1 can repress transcription when recruited to a site by fusion to GAL4 binding sites [150], and the leucine zipper region of the non-histone domain of macroH2A1.2 has been reported to bind the MATH domain of the Spop protein [151]. Furthermore, the non-histone domain has been shown to disrupt transcription factor binding; while the H2A-like domain apparently blocks SWI/SNF nucleosome remodeling [152].

In differentiating ES cells, mH2A1 relocalizes from the centrosome to the inactive X chromosome after initiation and propogation of X inactivation [153,154]. In mouse

embryos, association with the inactive X begins between the 8 and 16 cell stage and therefore MCB formation may be an earlier event in the extraembryonic tissues [155]. Association of mH2A1 with the centrosome is seen in all cell types, but the detection can be influenced by fixation techniques [156]. Centrosomic association is cell-cycle dependent, increasing as the association with the inactive X dissipates in late S phase, perhaps reflecting a degradation pathway possibly shared by several other chromatin proteins [157].

Conditional deletion of *Xist* results in loss of the MCB, despite ongoing maintenance of silencing, and *Xist* transgenes can recruit macroH2A1 localization to autosomal sites [67]. Induction of *Xist* in undifferentiated ES cells did not yield a MCB; however, upon induction of *Xist* expression in differentiated MEF (murine embryonic fibroblast) lines, where inactivation does not occur, MCB formation can be detected [158]. This suggests that *Xist* may recruit macroH2A in a developmentally-regulated manner independent of inactivation. Although a direct interaction has yet to be demonstrated, continued *Xist* expression is required for macroH2A localization and a specific domain at the 3' end of *Xist* has been identified that is required for MCB formation [53]. In metaphase inactive X chromosomes, mH2A1 can be seen to preferentially associate with several regions including Xp22, Xp11, Xq13 (indistinguishable from *XIST* location) and Xq22-24, at the DXZ4 repeat, and overlapping the site of histone H3 lysine-4 methylation [157].

The macroH2A1 gene is conserved beyond mammals – it shows greater than 90% amino acid identity between human and chicken, supporting a role in addition to its involvement in X-chromosome inactivation. Other histone variants show associations

with heterochromatin. For example, a histone H3 variant is associated with centromeric heterochromatin [159], and the H2A.Z variant has recently been shown to be found at transitions between euchromatin and heterochromatin [160]. Macrohistone H2A1.2 is also present in the XY body during spermatogenesis [29,30], as is the histone H2A variant H2AX, which is required for sex body formation and meiotic sex chromosome inactivation [161].

I.3.5 DNA Methylation

DNA methylation is an epigenetic modification that involves the transfer of a methyl group to the 5-position of a cytosine residue in CpG dinucleotides in mammals. DNA cytosine methylation is developmentally regulated and plays an important role in the regulation of gene expression in vertebrate development. A global decrease in methylation during the blastocyst stage is thought to erase gamete-specific methylation patterns and re-establish pluripotency. Methylation then increases during implantation and gastrulation to create stage- and tissue-specific methylation patterns [162]. The majority of CpG sites in the vertebrate genome are methylated and the spontaneous deamination of these methylated cytosines into thymine is thought to be responsible for the underrepresentation of CpG dinucleotides in the mammalian genome. However, certain regions of the genome, called CpG islands, remain largely unmethylated and retain the expected CpG frequency. These clusters of CpGs are often associated with the promoter regions of ubiquitous housekeeping genes and some tissue-specific genes and generally remain unmethylated regardless of their transcription status [163]. However, regulated methylation of CpG islands does occur on the inactive X chromosome, as well

as in imprinted regions, and contributes to the stable silencing of the associated genes [164].

Methylation is catalyzed by DNA methyltransferases (Dnmts) and three active DNA cytosine methyltransferases have been identified in mammals: *Dnmt1, Dnmt3a* and *Dnmt3b*. *Dnmt1* was the first to be cloned and since it has a 5-30 fold preference for hemimethylated DNA compared with unmethylated DNA, it is believed to be a maintenance methylase that ensures the stable transmission of methylation patterns through mitosis [165]. However, there is evidence that *Dnmt1* may also have *de novo* methyltransferase capabilities [166]. *Dnmt3a* and *Dnmt3b* were identified through database screens for the conserved methyltransferase catalytic domain and are required for *de novo* methylation during mammalian development [167,168]. The *Dnmt3* genes have also been shown to interact and functionally co-operate with *Dnmt1* to maintain methylation patterns in human cancer cells and in mouse ES cells [169-171]. Developmentally-specific isoforms of both *Dnmt1* and *Dnmt3* have been characterized and are likely responsible for the establishment and maintenance of the distinct methylation patterns observed at different stages in development [172].

DNA methylation can regulate gene expression by a variety of mechanisms. Methylation at CpG dinucleotides may directly interfere with the binding of transcription factors to their recognition sequences [173,174]. Alternatively, methyl-CpG-binding proteins may directly compete with transcription factors for binding sites at the promoter or may initiate the recruitment of co-repressors to alter the chromatin structure [166]. The methyl-CpG-binding protein *MeCP2* can recognize and bind DNA sequences containing a single methylated CG dinucleotide [175] and through its association with the

transcriptional repressor Sin3A and histone deacetylases *HDAC1* and *HDAC2*, it can alter chromatin structure and silence transcription in a methylation-dependent manner [176]. Similarly, another methyl-CpG-binding protein, *MBD2*, which requires >10 methylated CpGs to bind DNA, forms a complex with the ATP-dependent chromatin remodeling protein, Mi-2, and histone deacetylases *HDAC1* and *HDAC2* [177,178]. Therefore, methylation at CpG sites can target repressor complexes to specific regions which are then deacetylated to form silenced chromatin structures.

DNA-mediated gene transfer experiments provided the first evidence that DNA from the inactive and active X chromosomes was differentially modified and functionally inequivalent [179]. In these experiments, wildtype *HPRT* alleles on an active but not the inactive X were able to transfer *HPRT* activity into recipient *HPRT*-deficient cells, indicating that the DNA on the inactive X is much less efficient for gene transfer. Interestingly, X-chromosomal DNA from sperm is still competent for DNA transformation, suggesting that in this case the inactive X has not been methylated [33]. A role for DNA methylation in at least some of the functional differences between the active and inactive X was supported by early observations that, unlike most CpG islands which remained unmethylated regardless of the transcriptional status of the gene, CpG islands associated with genes on the X chromosome tend to be heavily methylated on the inactive X, but unmethylated on the active X, and genes that escape inactivation are often unmethylated [180]. In addition, treatment of interspecific somatic cell hybrids with the demethylating agent 5-azacytidine resulted in reactivation of the silenced genes on the inactive X, demonstrating a direct role for methylation in the maintenance of inactivation [180].

Targeted disruptions of the DNA methyltransferase loci in mouse have further defined the role of methylation during both the initiation and maintenance stages of X inactivation. Methylation during the initiation of inactivation appears to be important for the regulation of *Xist* expression since hypomethylation at the *Xist* promoter in the *Dnmt1* mutants results in ectopic *Xist* expression and inactivation [74,75]. In addition, methylation by *Dnmt1* appears to be important for maintaining the silenced state in the embryonic lineage since transgenes integrated into the X often reactivate in mutant embryos. However, methylation does not appear to play as crucial a role in imprinted inactivation in the extraembryonic tissues since inactivation appears to be properly maintained in *Dnmt1* mutant embryos [181]. The *Dnmt3* family of methyltransferases also appears to be involved in maintaining X-inactivation. Mutations in the *DNMT3B* gene underlie the ICF immunodeficiency syndrome (immunodeficiency, centromeric decondensation, facial anomalies), which results in the hypomethylation of CpG islands, advanced replication timing and destabilization of the silenced state of X-linked genes normally subject to inactivation [182]. Therefore, the methylase activities of *Dnmt1* and *Dnmt3b* are required for both the initiation and maintenance of the inactive state on the X chromosome.

DNA hypermethylation does not, however, appear to be an essential aspect of X-inactivation in all mammals. Imprinted inactivation in marsupials, as well as inactivation in extraembryonic tissues, occurs without extensive methylation [14,183]. However, inactivation in these cases is less stable and is often incomplete, which suggests that DNA methylation may be a recent acquisition in eutherians associated with more stable, random inactivation [184].

I.3.6 Other Changes Associated with an Inactive X

The presence of the Barr body in female cells was one of the key observations upon which Lyon based her original hypothesis of X-chromosome inactivation [8]. The Barr body stains brightly with DNA stains, which suggests it is more condensed than the active X chromosome, however, more recent cytological examinations suggest that the inactive X chromosome is not dramatically reduced in volume relative to the active X chromosome, rather it shows a differential shape [185]. The Barr body was originally described as being adjacent to the nucleolus [186], and is also often found at the nuclear periphery [187], consistent with its heterochromatic state. Therman suggested that because isodicentric X chromosomes formed 'bipartite' Barr bodies that there was a region in proximal Xp that escapes inactivation [188], which has been supported by recent studies of human genes that escape inactivation [139].

The inactivated X chromosome is also seen to be 'folded' in metaphase spreads at an unusually high frequency [189,190]. This bend is found in several species, and appears to be located at the site of the X inactivation center [191]. Such a bend could result from the telomeres of the inactive X being closer to each other than those of the active X, suggesting a looped structure in interphase [192], although this has not been supported by multiprobe analysis [193]. Finally, the inactive X chromosome is less sensitive to DNase I than the active X chromosome [194,195], consistent with active chromatin being preferentially sensitive to this enzyme [196]. In addition, general DNase sensitivity and the presence of DNase hypersensitive sites at individual genes corresponds to their activity (e.g. [197-199]). However, there does not appear to be a

difference between the active and inactive X chromosomes in their scaffold-associated regions [200].

I.3.7 Gene Silencing

While the majority of X-linked genes are subject to X inactivation and expressed solely from the active X chromosome, there are genes that are expressed from both the active and inactive X chromosome, and thus are said to escape X inactivation. It could be anticipated that genes in common on the X and Y chromosomes would not be silenced by dosage compensation, as silencing would actually cause an imbalance in which females expressed only one copy while males expressed two copies. Indeed, Lyon suggested that genes in the pairing region of the X and Y chromosome would escape inactivation [201], and this has been shown to be the case in both mice (Sts - [202] ; Mid1 - [203]) and humans. As meiotic exchange occurs between the X/Y pairing region, the region does not strictly follow X-linked inheritance patterns, and thus is called a pseudoautosomal region (PAR). In humans the Xp PAR spans approximately 2.6 Mb, and has many genes that escape inactivation [139]. In addition, there is a second PAR on the distal end of the human X chromosome long arm. The Xq PAR is only ~330 kb, and, although recombination is approximately six times higher than in autosomes, there is not an obligate crossover in male meiosis as there is in the Xp PAR. Interestingly, in the Xq PAR the centromeric genes H-SPRY3 and SYBL1 are inactivated on both the X and the Y chromosome, while the more distal IL9R and CXYorf1 genes escape inactivation [204]. Studies on the evolution of the X/Y chromosome pair have shown that different regions were added over evolutionary time (see section above on evolution of the X), and the PAR is the most recent addition, and thus would also be expected to escape XCI as it

has had the least time to acquire the necessary sequences for proper spread of inactivation.

I.3.7.i) Genes escaping X inactivation

Indirect evidence for a gene being subject to inactivation can be derived from phenotypic evaluation, such as whether female carriers of a mutation show non-random inactivation or partial phenotypes, and whether affected females have been associated with translocations disrupting the X-linked gene (reviewed by [205]). As the human X chromosome retains its expression pattern when isolated in rodent cells [206], this has also provided a powerful system to analyze the human active and inactive X chromosome separately. Direct evidence for genes escaping inactivation comes from a demonstration of increased product (RNA or protein) in individuals with multiple X chromosomes, or the demonstration of expression from both alleles. The latter requires an expressed polymorphism and either analysis in single cells or in a population of cells in which the same X chromosome is always inactive. The G6PD A/B isozymes forms heterodimers when produced in a single cell, and this allowed the demonstration of expression from both X chromosomes in oocytes [207]. Carrel and Willard used PCR to analyse single cells to demonstrate expression of the REP-1 gene [208], however, a more common approach has been to use clonal cultures, or cells derived from individuals with X rearrangements or mutations that resulted in non-random inactivation. Analyzable expressed polymorphisms have increased in number and ease of detection as they have evolved from relatively rare protein isoforms to the more common single nucleotide polymorphisms (SNPs) used in recent studies [209,210].

Several studies have used hybrids to examine the inactivation status of genes (e.g. [211]), however the largest profiling using hybrids was by Carrel et al. (1999), who utilized a panel of up to nine inactive X-containing hybrids for the analysis of 224 X-linked transcripts. Of these, they found 177 were subject to XCI, 34 escaped inactivation and 13 were heterogeneous, being expressed in from three to six of the inactive X-containing hybrids. Non-conforming expression in only one or two hybrids was found for almost one-fifth of the genes, and suggests that there may be some instability in the hybrid system. However, for two of the heterogenous class, further studies using expressed polymorphisms have validated that for TIMP1 and REP1 there is expression from the inactive X chromosome in some females, but that in other females the allele on the inactive X is silenced [208,212].

The discovery that such a large proportion of human genes escape inactivation was somewhat surprising. Females who lack an X chromosome manifest Turner syndrome, and over 90 % of 45,X conceptuses are lost. This suggests that the genes escaping inactivation play an important role in normal development. As males also lack an X chromosome, it is presumed that the majority of these important genes have expressed copies on the Y chromosome that provide the necessary dose of gene product for proper development in males. Mice lacking an X chromosome develop normally, which may reflect that there are fewer genes escaping inactivation in mice. Indeed, many of the genes that escape inactivation in humans have been shown to be subject to inactivation in mice [213].

Analysis of the homologous murine genes from a region of the human X chromosome that contains many genes escaping inactivation (Xp11) showed that only

one of the mouse homologs of human genes escaping inactivation also escaped inactivation in mouse (Smcx), suggesting that there is a fundamental difference for the extent of silencing between human and mouse. The study could not determine any obvious difference in promoter structure [214]. The large number of genes escaping inactivation in humans may provide an insight into the process of inactivation.

I.3.7.ii) Sequences involved in the spread of silencing

The genes that escape inactivation are non-randomly distributed across the X chromosome with 31 of 104 sequences analyzed on Xp escaping inactivation, while only three of 120 on Xq escaped inactivation [139]. This discrepancy is at least partially reflective of the more recent evolutionary origin of the human X short arm relative to the long arm [5], however the proximal region of the human X short arm was also X-linked in marsupials and still contains a number of genes that escape inactivation. Furthermore this explanation does not account for the difference in number of genes escaping inactivation between humans and mice. Perhaps the inactivation signal is attenuated by the centromeric heterochromatin of humans, as mice are essentially telocentric.

To address whether escape from inactivation occurs regionally or on a gene-by-gene basis, several studies have focused on examining expression of many genes within a small region and find that genes escaping inactivation do seem to cluster together, although genes that are subject to inactivation and escape inactivation can be in close proximity to each other [215,216]. Further support for 'regionality' comes from the apparent 'banding' observed on the inactive X for features associated with activity such as earlier replication and histone acetylation.

Whether escape from inactivation occurs for a region or individual gene, it is clear that there are differences in X chromatin for inactivation sensitivity, and this differential susceptibility should yield important clues about the signals necessary for silencing. It was proposed that the X chromosome must contain 'way-stations' that would boost the inactivation signal as it spread along the X [179]. Based upon sequence differences amongst chromosomes, Lyon has proposed that such elements could be LINE repetitive elements [217]. In comparing regions with genes that escape inactivation to those that are subject to inactivation, support was garnered for this model [218].

Another approach to studying the spread of inactivation, which is less confounded by the complications of the evolutionary history of the X and Y chromosome, has been to analyze autosomal material translocated onto the X chromosome. In balanced X/autosome translocations there is generally non-random inactivation of the intact X chromosome [219]. This is believed to result from selective lethality of cells with unbalanced expression due to silencing of autosomal material associated with the X carrying the X inactivation center, or failure to silence the X material separated from the XIC, as has been demonstrated in mice [220]. However, in unbalanced cases it is generally the derivative X which is inactivated, and this inactivation can spread into the autosome. Using mouse coat colour markers, Russell noted that spread of inactivation into autosomal material translocated to an X was variable and limited [221]. While the variable spread of late-replication into translocated autosomal material in human has also been recognized for years, and milder phenotypes than might be expected have suggested silencing of autosomal regions, recent studies to correlate gene silencing and other features show a lack of consistency [222].

There have been a limited number of studies that have examined the actual expression of genes in human X;autosome translocations. Interestingly, there were two genes that were informative in two different t(X;6) cases. One of these genes was inactivated in both patients; however the other gene was active in one case and subject to inactivation in the other, suggesting that DNA sequence is not the only factor in determining whether a gene will be inactivated, and that the location of the breakpoints must also contribute [223]. No correspondence was observed between whether silent genes were in G-light or G-dark regions and no absolute correspondence with distance from the centromere was determined, although most distal genes did seem to escape silencing more often [223].

In the limited studies done so far, it appears that histone modifications are better correlates with gene silencing than late replication, however not all regions with active genes showed histone modifications which may reflect the difference between the size of region escaping inactivation and the resolution of the cytological approach to examine histone modifications [223]. Furthermore, all genes examined in late-replicating regions were inactive, and while the presence of a CpG island was not a pre-requisite for inactivation, those that were associated with silenced genes and examined were hypermethylated [223]. Dilution plating of a cell line that had showed variability in the extent of late replication, did not suggest that the patterns were stably inherited, leading to the suggestion that the variability results from lack of maintenance of silencing [224]. Sharp et al. did not detect differences between EBV-transformed lymphoblasts and peripheral blood from a patient, and thus suggested that the current pattern detected was relatively stable [223]. In another study, Hall and coworkers found only partial coverage

of the translocated autosome with *XIST* RNA and suggested that this tenuous relationship results in failure of maintenance of silencing on autosomal material, even when a relatively minor phenotype would suggest that there had been more extensive inactivation than currently detected [225].

I.3.8 Evidence for Complementary and Interactive Roles of Chromatin Changes in X inactivation

Once established, the inactivated state is extremely stable due to the interplay of several different repressive mechanisms. These interactions are shown schematically on Figure 1-4. Expression of *Xist* seems to be necessary and sufficient for the induction of X inactivation, however it is not clear how expression results in the acquisition of heterochromatin and gene silencing. *Xist* expression does recruit macroH2A independently of silencing; however the formation of an MCB has not been associated with other features of X inactivation. *XIST* localization is lost in mouse/human somatic cell hybrids [78,226], which have also been shown to be more susceptible to gene reactivation with demethylating agents [227]. Furthermore, hybrids are reported to lose the peripheral localization of the inactive X [187]. While *XIST* is not essential for maintenance of an inactive X [66-68], loss of *Xist* does result in a slightly elevated reactivation frequency, which is enhanced by inhibition of DNA methylation or histone deacetylation [65].

Replication timing appears to be closely linked with DNA methylation. Changes in replication are induced by treatment with 5-azacytidine, an inhibitor of DNA methylation in a cell-dependent manner [228-231]. 5-aza-deoxycytidine acts late in S phase to inhibit condensation of the late replicating X [232], and also affects the

constitutive heterochromatin of centromeres [233]. Furthermore, ICF cells that are mutated for *DNMT3B*, in addition to being hypomethylated, also display altered replication patterns, with many X-linked loci replicating earlier in S phase. Hypomethylation of satellite sequences is also observed in ICF cell cultures, and some sequences tend to show advanced replication [234]. Inhibitors of HDACs also disrupt replication timing (reviewed in [107]). In somatic cell hybrids, induction of gene expression by 5-azacytidine has been demonstrated to result in early replication of the reactivated gene, however, an advance in replication timing for *XIST* and *F9* could be detected after treatment even when transcription was not induced [131]. The nature of the relationship between DNA methylation and replication timing is not clear, however if CpG rich sequences serve as origins of replication [109], then methylation of these islands might alter origin use or delay the time of firing.

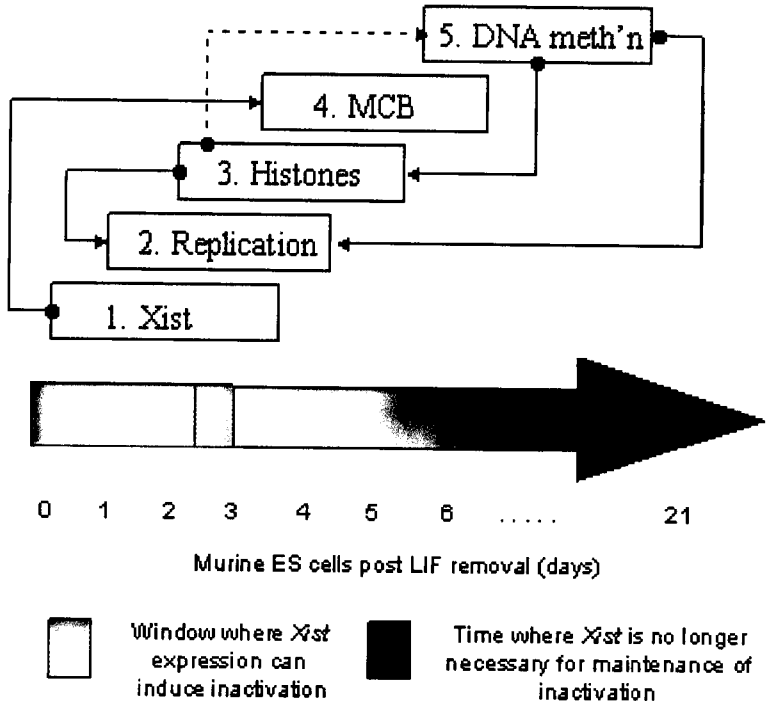


Figure 1-4. Schematic representation of the steps involved in the formation of a heterochromatic inactive X. Interactions between features are shown by connectors. Approximate timing of the events following induction of differentiation in mouse ES cells is shown below the arrow [45]. In mouse, induction of *Xist* first results in reversible inactivation which is later stabilized, while later induction of *Xist* is unable to induce silencing [51].

In *Neurospora crassa* histone methylation is required for DNA methylation, and similarly in *Arabidopsis thaliana* mutations in histone methylation disrupt DNA methylation [235,236]. Such direct interactions have not been described in mammals, so the interaction between histone modifications and DNA methylation is dotted in Figure 1-4. However, many of the methylated DNA-binding proteins recruit HDACs [172], and

chromatin remodeling complexes influence DNA methylation [237,238], so it seems likely that there is at least an indirect influence.

Figure 1-4 demonstrates that rather than being a simple step-wise accumulation of chromatin changes that lead to silencing, X-chromosome inactivation involves a complex interaction amongst the various events. Indeed an inactive X can be retained after loss of *XIST* expression [66,68], changes in replication timing [122], altered histone modifications [239], loss of the MCB [67], or the absence/reduction of DNA methylation (as in marsupials [10], ICF syndrome [234] and extraembryonic tissues [183]). Reactivation of individual genes occurs at higher frequency when a feature is disrupted, such as in somatic cell hybrids that have delocalized *XIST* [227] or marsupials that lack DNA hypermethylation [10]. This reactivation frequency can be elevated synergistically when multiple features are lost [65]. Reactivation of an entire X chromosome is rarely observed – it occurs naturally during oogenesis, and is also observed upon somatic cell fusion of extraembryonic cells [240]. While there is strong evidence that stable expression and localization of *Xist* is the initiating event in X chromosome silencing, it is not yet clear how all the other heterochromatic features are assembled to result in the extremely stable inactivation of the X chromosome. The recent identification of additional players in the process, such as Eed/Ezh2, mH2A, or BRCA1 may start to provide insights into the initiation and maintenance of silencing.

X-chromosome inactivation is clearly essential for the normal development of mammalian females. The choice of chromosome to inactivate is normally random in humans, and thus females tend to be protected from X-linked disease. Skewing of inactivation can occur due to chance, or more commonly due to selective advantages or

disadvantages of certain cell populations, as seen with X chromosome rearrangements or some X-linked mutations. Primary skewing of inactivation due to disruptions in the initial choice of which X chromosome to inactivate is seen in mice, and may be associated with variants in humans, possibly of the *XIST* gene [71,72], but it is not known how common such variants may be. Skewed inactivation has also been reported in individuals who experience recurrent spontaneous abortions [241,242], or are at risk of ovarian or breast cancer [243-245], but no mechanistic association has yet been conclusively demonstrated. In addition to answering questions of clinical importance, the study of X-chromosome inactivation can serve to address general mechanisms of epigenetic silencing. While the *XIST* RNA is unique to mammals, silencing by RNAs may be a more common event than previously thought. The study of epigenetic silencing has gained much attention recently as its role in expression changes associated with cancer [246] and somatic reprogramming [247] has been detailed, but recent reports also suggest roles in diverse areas such as development [248], circadian clocks [249], long-term memory [250], and quantitative traits [251]. In the over 40 years since X-chromosome inactivation was first hypothesized [8] much has been learned, but there is still much more to learn about the silencing of one X chromosome in mammalian females.

Thesis Objectives

Although significant advances have been made in elucidating the sequence of events that occur in the initiation and establishment of inactivation in early mouse development, obtaining suitable models for studying X inactivation in early human

development has been more difficult. In my thesis, I attempt to determine expression patterns within the *XIST* region in somatic versus early development using human embryonal carcinoma cells which are germ-cell derived tumors. Using these cells in conjunction with *XIST* transformant clones, I specifically address the possible role of an antisense transcript within the region in *XIST* regulation and X chromosome inactivation.

In addition, I examine *XIST* transcript localization and the factors involved in its unique association with the X chromosome. By introducing *XIST*-containing PAC transgenes into male somatic cells, I was able to determine the relative contribution of transgene copy number and expression levels on transcript localization. However, to more precisely evaluate the role of specific domains within the *XIST* transcript, *cis*-DNA sequences, and transcript levels on the transcript's ability to associate with the surrounding chromatin, I created an inducible *XIST* cDNA construct that can undergo site-specific integration into a single FRT (Flp recombinase target) site in the recipient cell line. With this system I can control for integration site, copy number and transcript levels which will enable me to more effectively evaluate the role each of these variables on transcript localization.

CHAPTER 2: MATERIALS AND METHODS

II.1 Cell Culture

The N-Tera2D1 embryonal carcinoma cell line was derived from a lung metastasis of a testicular teratocarcinoma [252]. This cell line has been extensively characterized in other studies and have shown that the line is aneuploid but still is able to differentiate into a variety of different somatic cell types in culture [253,254]. N-Tera2D1 (American Type Culture Collection, CRL-1973) was grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and penicillin/streptomycin (Gibco-BRL) at 37°C. When confluent, the cells were passaged by trypsinization and reseeded at a density of about 5 X 10^6 cells per 75 cm^2 flask. BOSC 23, a female human embryonic kidney cell line, and the male somatic cell line HT-1080 were grown similarly. Although full karyotypes for these two cell lines were not determined, they both appear to be aneuploid since RNA-FISH using various probes have indicated the presence of an abnormal number of chromosomes. For *in situ* hybridizations (performed by our collaborators in Dr. Jeannie Lawrence's laboratory), the female diploid lung epithelial line ATCC CCL 75 (WI-38) was used and was grown according to recommendations of the American Type Culture Collection (Rockville, MD). The mouse-human somatic cell hybrids were grown as previously described [212]. The human, female somatic embryonic kidney cell line Flp-In T-Rex 293 was obtained from Invitrogen and grown as recommended by the manufacturer (Cat# R-750-07). The cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and penicillin/streptomycin (Invitrogen) at 37°C. The cells were passaged by trypsinization and split at a ratio of 1:4 when confluent.

II.2 Generation of transformant clones

PAC 92E23 was retrofitted with a neomycin resistance cassette (kindly provided by J. Mejia and A. Monaco) [255] for transfection into mammalian cells. The retrofitted PAC was introduced into HT-1080 cells using Lipfectamine 2000 (Invitrogen) and the transformant clones were maintained in DMEM supplemented with 10% fetal calf serum, penicillin/streptomycin and 400 µg/ml Geneticin (Invitrogen) for selection.

The short 5' and 3' *XIST* fragments were inserted into the multiple cloning site of the pAX72 vector and co-transfected with the puromycin resistance vector pPGKPuro into the recipient cell lines using Lipofectamine 2000 (Invitrogen). Puromycin was added to the media (1.5µg/ml) to select for stable transformant clones.

To make a full-length *XIST* cDNA construct, the *XIST* exon 1 sequence was obtained by PCR amplification using the Expand Long Template PCR system (Roche). The PCR was performed using the supplied Buffer #3, 0.5 mM of each dNTP, 0.5µM of each primer (*XIST* primers AT1: 5' GAACCAACCAAATCACAAAGA 3' and *XIST* exon 1B: ACTGCTGTTGTGATGACTAG 3' which give rise to an 11 317 bp product), 100ng of template genomic DNA and 3.75 U of the supplied enzyme mix in a 50µl reaction. The amplification of the product involved an initial denaturation step of 2 min at 94°C followed by 30 cycles of a 30 second denaturation at 94°C, a 30 second annealing at 54°C and a 4 minute elongation at 68°C. A final 7 min elongation step was then performed at 68°C. The PCR product was then cut with XbaI to obtain a sequence that extends from nucleotides 105-10748 of the *XIST* cDNA sequence (Accession number M97168). The sticky ends were filled with Klenow and the resulting blunted fragment was inserted into the EcoRV site of the pcDNA5/FRT/TO expression vector (Invitrogen,

Figure 6-3a). To obtain the 3' end of *XIST*, *XIST* cDNA clones previously isolated from human heart and fetal brain libraries [256] were used and inserted downstream of the exon 1 fragment, into the NotI and XhoI sites of the pcDNA5/FRT/TO multiple cloning site. The *XIST* spliced variant used contains the 5' end of exon 6 up to the splice donor site at nucleotide 13 809 as well as exons 7 and 8 (Figure 6-3b).

The pcDNA5/FRT/TO inducible expression vector containing the *XIST* cDNA constructs was co-transfected with the pOG44 plasmid expressing Flp recombinase to catalyze a site-specific recombination event between the FRT site integrated into the genome of the host cell line and the expression vector. The cells were then grown in media containing hygromycin (150µg/ml) to select for clones that had undergone the proper site-specific integration event.

II.3 RT-PCR

Total RNA was isolated by acid guanidinium thiocyanate-phenol-chloroform extraction [257] and treated with RNase-free DNase I (10U) for 1h at 37°C to remove any genomic DNA. Each RNA preparation was tested for DNA contamination by performing PCR without the addition of reverse transcriptase (RT). Nuclear and cytoplasmic RNA was isolated after nuclei from lysed cells were separated by centrifugation [258]. Random primed RT reactions were performed using M-MLV reverse transcriptase (200U; Gibco BRL) with 5-12 µg RNA and random hexamers (100 pmol) to prime cDNA synthesis for 2 hours at 42°C followed by 95°C for 5 min. For strand-specific RT-PCR, first strand synthesis was performed using Superscript II reverse transcriptase (200U; Gibco BRL) with 12µg of total RNA and either a sense or antisense primer (25pmol) to amplify antisense or sense transcripts, respectively. The sense

orientation refers to the orientation originally described for *XIST* [46]. The reactions were incubated at 50°C for 1 hour with subsequent heat-inactivation of the enzyme at 80°C for 30 min. Taq polymerase (5U) was then used to amplify the strand-specific or the randomly primed cDNA using paired sense and antisense primers. *XIST* region primers are listed in Table 2-1; additional primer sequences are as follows: *ACTIN* 1, 5'-ATG ATA TCG CCG CGC TCG –3'; *ACTIN* 2, 5'- CGC TCG GTG AGG ATC TTC – 3'; *PGK*r1, 5'- TCG GCT CCC TCG TTG ACC GA –3'; *PGK*r2, 5'- AGC TGG GTT GGC ACA GGC TT –3'; JPX1, 5' – TTG CAC GGA GTC CAA TCA CT – 3'; JPX2, 5' – TCT GCA ACT TCC AAG CTT CG – 3'. For these primers, PCR cycles were 94°C 1min, 54°C 1 min, and 72°C 2 min. Primers amplifying regions 3' of *XIST* required 40 cycles while primers within *XIST* required 30 cycles in female somatic cells and 40 cycles in N-Tera2D1.

Control strand-specific RT-PCR with *in vitro* transcribed 'sense' and 'antisense' RNAs showed some cross-amplification when high levels of transcripts (10^6 molecules) were used as a template for reverse transcription (data not shown), and in fact, reverse transcription in the absence of primer will also yield product at extremely high template concentrations. Such products were detected even at elevated RT temperatures of up to 56°C, and were attributed to non-specific priming as controls without reverse transcriptase were blank. While the extent of this artifact was at least partially primer pair dependent, it precluded reliable detection of strand-specific product if the opposite strand was present in greater than 100,000 fold excess, thus we did not analyze female-derived EC lines since a large population of differentiated cells might result in artefactual detection of an antisense product.

For the nuclear and cytoplasmic RT-PCR in the region 3' of *XIST* in N-Tera2D1, a nested PCR (50 cycles in total) with primer pairs 11 and 11N was done to ensure that a negative result was not simply due to lower RNA concentrations in the cytoplasmic preparations.

II.4 Quantitative Competitive RT-PCR

To synthesize the competitor RNA, a 63 bp AvrII fragment (11629-11692 of *XIST* cDNA sequence accession number M97168) in exon 4 was deleted from an *XIST* cDNA clone in Bluescript and T7 Ampliscribe (Epicentre Technologies) was then used to synthesize transcripts. The concentration of the transcript was determined after DNaseI treatment from the optical density. Several RT reactions were performed each with 5μg of total RNA and known amounts of competitor RNA in 10-fold dilutions. Reactions were random primed as described above, using 200 U of M-MLV reverse transcriptase. Serial 10-fold dilutions of each RT reaction were then used as a template for PCR ensuring that the reactions were in the linear phase of amplification. The PCR primers (*XIST* primer pair 6) spanned the deletion in the competitor so that the competitor PCR product would be smaller than the endogenous. To calculate the approximate number of *XIST* transcripts per cell, we assumed that a typical cell has $\sim 10^{-5}$μg of RNA molecules [258] and the number of competitor RNA molecules present in each reaction was calculated using the formula:

picograms of competitor RNA X [1pmol / (340pg X number of bases)] X 6.02×10^{23}

II.5 Quantitative PCR with SYBR Green Dye

To quantitate *XIST* transcription levels, total RNA (12ug) was reverse transcribed using M-MLV reverse transcriptase and random hexamers. Primers for *XIST* and *ACTIN*

(control) were designed using Primer Express Software version 1.0 (PE Applied Biosystems), with each pair spanning a splice junction to avoid amplification of contaminating genomic DNA. The *XIST* primer pair, q*XIST*1: 5' ACGCTGCATGTGTCCTTAGTAGTC 3' and q*XIST*2: 5' ATTTGGAGCCTCTTATAGCTGTTTG 3', gives rise to a 102 bp product spanning exons 2 and 3. The *ACTIN* primer pair, q*ACTIN*1: 5' TTGCCGACAGGATGCAGAA 3' and q*ACTIN*2: 5' GCCGATCCACACGGAGTACTT 3', gives rise to a 101 bp product and spans exons 5 and 6. The PCR amplification was performed in a GeneAmp 5700 sequence detection system (PE Applied Biosystems). The reactions were set up in 96 well plates in 25µl reaction volumes containing 400nM of each forward and reverse primer and 2X SYBR green PCR master mix (PE Applied Biosystems). The thermal profile used was 95°C for 10 min to activate the DNA polymerase, followed by 40 cycles of 95°C for 15s and 60°C for 1 min. For both *XIST* and *ACTIN*, a dilution series of a plasmid control was run along with the experimental samples to obtain the standard curve. Each reaction was repeated at least 3 times.

To determine the number of transgene integrations in the HT 1080 PAC transformant clones, primers amplifying genomic DNA within an autosomal gene *SDC4* [259] and within the *XIST* region were used. The *SDC4* primer pair, *SDC4F*: 5' CAGGGTCTGGGAGCCAAGT 3' and *SDC4R*: 5' GCACAGTGCTGGACATTGACA 3' gives rise to a 128 bp product. The *XIST* primer pair, q*XIST*gen1 5' TGCTGCATGATGGTGTTAGCT 3' and q*XIST*gen2 5' TTGCGCTGCGACATTGAT 3' gives rise to a 126 bp product. The ratio of *XIST/SDC4* was taken for each clone then normalized to the same ratio in males.

II.6 Transcript Stability and Real-time Quantitative RT-PCR using SYBR Green

For N-Tera2D1, $5X10^7$ cells were grown overnight in a $75cm^2$ flask, then treated with either 4μg/ml Actinomycin D dissolved in DMSO, or DMSO only as a control. For female somatic cells, BOSC 23, derived from the 293 female human embryonic kidney cell line was used. For these experiments, $5X10^6$ cells were plated in a 60mm dish and incubated overnight before treatment. 12μg total RNA was reverse transcribed using random hexamers and the levels of *XIST* and *ACTIN* (control) were determined by real time quantitative PCR using SYBR green dye. Each reaction was repeated at least 3 times. The average variation in the replicates was 18% for *XIST* and 10% for *ACTIN*. The half-life of the transcripts was calculated by taking the ratio of *XIST* and *ACTIN* for a specific treatment time to control for variations in RT and RNA quality. *ACTIN* has a half-life significantly longer (>20h; [260]) than any of the timepoints used in these experiments. Each ratio was then divided by the average ratio of XIST/ACTIN in DMSO control samples and normalized to the zero timepoint.

II.7 Relative quantitation of expression from pAX72 short *XIST* constructs by RT-PCR

For the 5' *XIST* construct, three primers were used in the PCR reaction, XIST1: 5' TGTGGGTTGTTGCACTCTCT 3', XIST2: 5' ACAACGCCTGCCATATTGTC 3', and pAX72#1: 5' GCTTCATGATGTCCCCATAA 3' to give rise to a 695 bp endogenous *XIST* band and a 500 bp transgene band. For the 3' *XIST* construct, the three primers, C8-2: 5' GTCAGGAGGTTCTGTCAAGA 3', C6-1: 5' AGAAGGGGAAGGGGTAACAA 3', and pAX72#1 give rise to a 604 bp endogenous *XIST* band and a 450 bp transgene

band. PCR reactions were performed at 94°C-1 min, 54°C-1 min, and 72°C-2 min for 30 cycles.

II.8 *In Situ* Hybridization and Detection

A monolayer of cells was grown on glass coverslips and permeabilized on ice for 2 min with 0.5% Triton X-100, and 5% vanadyl ribonucleoside complex to preserve RNA (Gibco-BRL) in cytoskeletal (CSK) buffer [261]. Cells were then fixed in 4% paraformaldehyde for 10 min at room temperature, and were stored in 70% ethanol at 4°C.

Hybridization and detection was performed by our collaborators in the Lawrence Lab as described in [262]. The DNA probe, G1A, a ~10 kb genomic plasmid extending from the 4th intron to the 3' end of the XIST gene, was nick-translated using biotin-11-dUTP or digoxigenin-16-dUTP (Boehringer Mannheim, Indianapolis, IN). For RNA hybridization, cells were hybridized under nondenaturing conditions (such that cellular DNA was not accessible) overnight at 37°C in 50% formamide, 2X SSC using a probe concentration of 5 μg/ml. Human Cot-1 was included in the hybridization buffer. Posthybridization washes were performed as follows: 50% formamide, 2X SSC for 30 min at 37°C; 2X SSC for 30 min at 37°C; 1X SSC for 30 min at room temperature with agitation; and 4X SSC for 30 min at room temperature with agitation. Hybridization was detected with either antidigoxigenin antibody (Boehringer Mannheim) coupled with rhodamine or fluorescein at 200 μg/ml in 1% BSA, 4X SSC; or with fluorescein-conjugated avidin (Boehringer Mannheim) at 2.5 mg/ml in 1% BSA, 4X SSC for 1h at 37°C. Postdetection washes were performed as follows: 4X SSC for 20 min at room

temperature with agitation, 4X SSC, 0.1% Triton for 20 min at room temperature with agitation, and then 4X SSC for 20 min at room temperature with agitation.

II.9 Sequence Analysis

The dotplot was generated using PipMaker (http://bio.cse.psu.edu) [263]. To look for matrix-attachment regions (MARs), the MAR-finder program, which searches for sequence motifs known to occur in the vicinity of MARs was used (http://www.futuresoft.org/MAR-Wiz/).

Table 2-1. *XIST* region primers

Primer[a]	Position	Primer	Position
1as	566-585	1s	803-784
2as	784-803	2s	1478-1459
3as	5429-5448	3s	5694-5675
4as	11005-11024	4s	11231-11212
5as	11317-11336	5s	11559-11541
6as	11459-11478	6s	11798-11779
7(a)as	12897-12916	7(a)s	13432-13413
7(b)as	12992-13016	7(b)s	13176-13152
8as	15213-15232	8s	15718-15700
9as	142-161	9s	355-336
10as	1028-1047	10s	1674-1655
11as	2469-2488	11s	3018-2999
11Nas	2494-2513	11Ns	2825-2806
12as	6381-6400	12s	6819-6800
13as	8199-8218	13s	8680-8661
14as	9931-9950	14s	10366-10347
15as	10660-10679	15s	11265-11246

[a]Primers within *XIST* (primers 1-8) are numbered in relation to the *XIST* cDNA sequence (M97168) except for primer pair #9 which is numbered according to *XIST* 'd' (X56196). Primers 3' of *XIST* (primers 10-15) are numbered according to the distance from the end of *XIST* exon 8 (+1 is position 50415 of U80460).

CHAPTER 3: CHARACTERIZATION OF SENSE *XIST* EXPRESSION

Portions of this chapter have been included in two published papers:

Clemson, C.M., Chow, J.C., Brown, C.J. and Lawrence, J.B. (1998) Stabilization and localization of *Xist* RNA are controlled by separate mechanisms and are not sufficient for X inactivation. J. Cell. Biol. 142: 13-23.

Chow, J. C., Hall, L.L., Clemson, C.M. Lawrence, J.B. and Brown, C.J. (2003) Characterization of expression at the human *XIST* locus in somatic, embryonal carcinoma, and transgenic cell lines. Genomics 82: 309-22.

All RNA-FISH analyses presented in this chapter were performed by L.L. Hall and C.M. Clemson.

INTRODUCTION

The process of X-inactivation requires *XIST*, an X-linked gene that is expressed exclusively from the inactive X [256,264,265]. It encodes a 17 kb functional RNA that remains in the nucleus and has a specific association with the inactive X that is important in the initiation and spread of inactivation along the chromosome [46-48]. The transcript is heterogeneous due to alternative splicing, differential promoter usage and polyadenylation [46,48,90,266-268]. The significance of the various isoforms is unknown and may simply reflect a greater selective pressure to maintain the overall structure rather than the primary sequence [49].

In mouse, embryonic stem (ES) cells provide an *in vitro* model of X-inactivation since they can be induced to differentiate and undergo the inactivation process. Prior to inactivation, both X chromosomes in female cells and the single X in male cells express a low-level, unstable (t1/2 ~30 min) form of *Xist* that appears as a small dot of expression

by RNA-FISH. Upon differentiation, *Xist* is upregulated on the future inactive X through the stabilization of the transcripts (t1/2 ~5 hours) that accumulate along the inactive X, coating the chromosome in *cis*. [54,55]. An alternate upstream promoter (P_0) was initially proposed as a mechanism for this stability change [90]; however, this upstream expression has since been attributed to cross-amplification of an upstream pseudogene and to the 3' end of an antisense gene, *Tsix* [268]. However, mechanisms involving different *Xist* isoforms arising from alternate splicing events, as yet unidentified promoters, or alternative polyadenylation may still play a role.

XIST expression has also been examined in human and mouse preimplantation embryos. In mouse, expression has been detected as early as the 2-cell stage [269,270]. This early *Xist* expression is imprinted, being expressed exclusively from the paternally derived X, reflecting the preferential inactivation of the paternal X in extraembryonic tissues [271]. Expression of the maternal *Xist* allele begins shortly before gastrulation when the paternal imprint is erased and random inactivation occurs in the cells of the embryo [272]. In human preimplantation embryos, *XIST* expression is detected in the 1-cell zygote, slightly earlier than in mouse. In addition, transcripts were detected in male embryos as well as females, suggesting that the maternally inherited X in males are able to express *XIST* and are not subject to the same paternal imprint as in mouse [57,58].

Embryonal carcinoma (EC) cells have also been used to study X-inactivation in human development. EC cells are the pluripotent stem cells of germ-cell derived tumors which in some cases are able to recapitulate the processes in early human embryogenesis [273]. One of the most extensively studied embryonal carcinoma lines, TERA2, is derived from a lung metastasis of a testicular teratocarcinoma and can differentiate into

various cell types when exposed to differentiation-inducing agents like retinoic acid and hexamethylene bisacetamide [273]. In mouse, certain EC lines have been induced to undergo X-inactivation upon differentiation [274], and there is evidence for inactivation in human EC cells as well [275].

To determine whether there are changes in *XIST* expression during human development, I compared expression in embryonal carcinoma cells and male and female somatic cells, examining the extent, the half-life, and the quantity of expression in the different cell types. In addition to potential developmental factors affecting transcript stability, *XIST*'s specific association with the inactive X may also influence the transcript's half-life. To determine whether localization to the inactive X affects transcript stability, I compared the half-life of *XIST* in female somatic cells and mouse-human somatic cell hybrids where, possibly due to species differences, the transcripts are no longer able to localize to the inactive X.

RESULTS

Extent of XIST expression in somatic cells and N-Tera2D1

To determine the extent of *XIST* expression in female somatic cells, male somatic cells and the male-derived embryonal carcinoma (EC) line, N-Tera2D1, I performed RT-PCR and subsequent strand-specific RT-PCR with primers specific for regions upstream, downstream and within the body (exons 1-8) of the *XIST* gene (Figure 3-1). Primers upstream of the previously identified initiation site did not give any consistent PCR products in any of the cell lines, demonstrating that in the cell lines tested, there is no evidence for any alternative start sites upstream of the *XIST* gene. However, when primers downstream of exon 8 were used, products were consistently amplified in female

somatic cell lines (4 independent cell lines) in a region 3-10kb downstream of *XIST*. PCR has so far failed to detect any continuity between this downstream sense transcript and *XIST* exon 8 or splicing with other internal *XIST* sites. Male somatic cells do not show expression within or downstream of *XIST* which supports the hypothesis that the 3' transcript observed in female somatic cells is an alternate isoform of the *XIST* gene. Extremely low level transcripts were detected in males in the 3' end of exon 6; however, as these were constrained to a small region and were fainter than any other product observed, they were not further characterized.

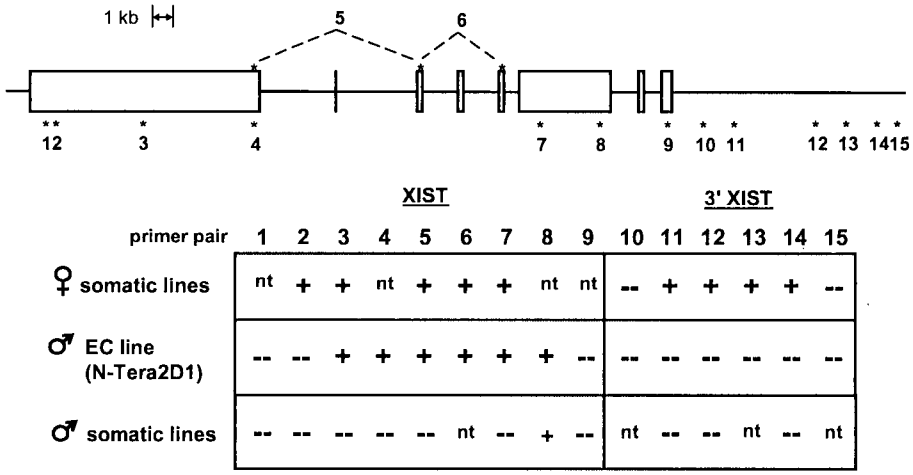


Figure 3-1. Extent of sense *XIST* expression in female somatic cells, male somatic cells and in the male-derived embryonal carcinoma cell line N-Tera2D1. Asterisks mark the positions of primer pairs used for RT-PCR. + = PCR product formation; -- = no expression detected by PCR; nt = not tested.

Sense expression within *XIST* was also detected in the male-derived EC line N-Tera2D1; however, the extent of *XIST* expression is much more limited than in female somatic cells. By strand-specific RT-PCR, sense expression in these EC cells initiates midway through exon 1 and extends into exon 6 of *XIST* (Figure 3-1), and may represent

a developmentally specific isoform, utilizing alternate promoter and polyadenylation sites. A second male-derived embryonal carcinoma line, Hs444, also expresses a similar *XIST* isoform, but this cell line showed a greater tendency to lose expression upon culture (data not shown), precluding further study.

Quantity, Stability and Localization of transcripts

At the time of X-chromosome inactivation, mouse *Xist* undergoes drastic changes in half-life and expression patterns, going from low-level unstable expression from both X chromosomes to a more stable transcript that is expressed from only one X which is then inactivated [54,55]. To determine the levels of the different *XIST* isoforms in human somatic and N-Tera2D1 cells, I utilized competitive RT-PCR. By this method, the number of *XIST* transcripts in female somatic cells was demonstrated to be ~2000 molecules per cell (Figure 3-2), similar to the previously determined quantity for mouse [276]. In the male N-Tera2D1 EC cells, the levels of *XIST*, as measured using sense-specific primers spanning exons 3 to 5, are only about 1/100X that found in female somatic cells (Figure 3-2).

The transcripts 3' of *XIST* exon 8 also appear to be present at lower levels than within the *XIST* gene body in female somatic cells and were undetectable by Northern analysis or RNase protection. Quantitative real-time RT-PCR results show that the levels of the 3' sense transcript are variable, but are generally several hundred-fold lower than levels within the *XIST* (data not shown).

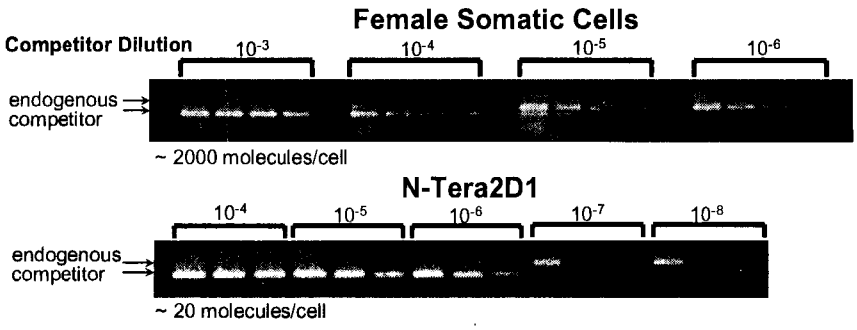


Figure 3-2. *XIST* quantitation in N-Tera2D1 and human female lymphoblasts by quantitative RT-PCR. Each RT was spiked with a known amount of competitor RNA containing a small deletion. The competitor dilutions are shown above the brackets. For each RT, a 10-fold dilution series was performed to ensure that the PCR reactions were in the linear phase of amplification.

In murine cells prior to X inactivation, low levels of *Xist* expression are accompanied by an extremely shortened half-life of the RNA. To determine the half-life of the transcripts in the male N-Tera2D1 and human female somatic cells, we treated cells with Actinomycin D to arrest transcription and then determined the levels of RNA at various timepoints after treatment by real-time quantitative RT-PCR (Figure 3-3). In female somatic cells, *XIST* has a half-life of ~4-5 hours [226] (Figure 3-3a). The 3' sense transcript, however, appears to have a slightly shorter half-life of around 2 hours and may be a further indication that the expression downstream of *XIST* in female somatic cells represents an alternate *XIST* isoform. The sense *XIST* expression detected in N-Tera2D1 had a similar stability to female somatic cells of ~4-6 hours (Figure 3-3b). Therefore, the lower level of *XIST* in these EC cells is not due to instability, but is more likely due to decreased transcription.

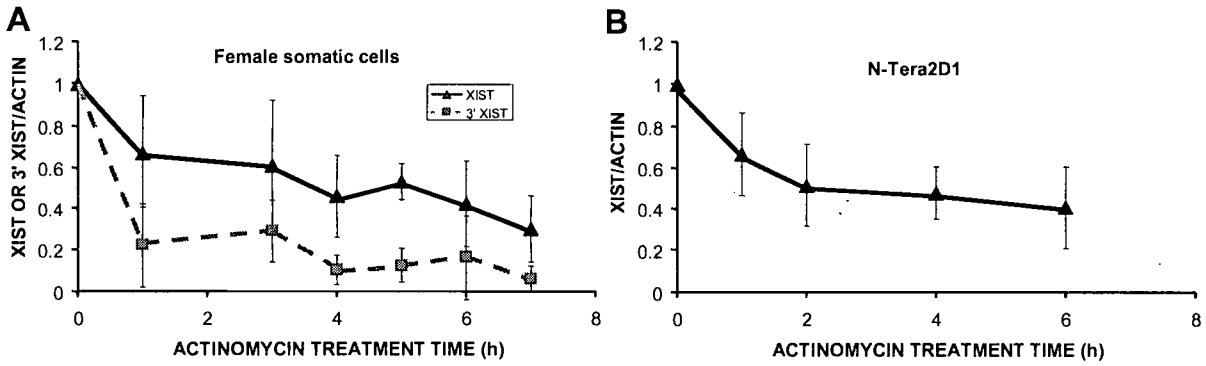


Figure 3-3. Half-life of *XIST* RNA in (A) female somatic cells and (B) N-Tera2D1. The graph shows the levels of RNA determined by real-time quantitative PCR at various timepoints after treatment with the transcriptional inhibitor Actinomycin D (4μg/ml). PCR reactions for each timepoint were performed in triplicate for both *XIST* and *ACTIN*. The point on the graph represents the average of *XIST* values divided by the average of *ACTIN* values. The average variation in the replicates was 18% for *XIST* and 10% for *ACTIN*.

Nuclear and cytoplasmic RNA preparations show that the 3' sense transcript in female somatic cells and the short *XIST* isoform in N-Tera2D1 remain in the nucleus, strengthening the hypothesis that they are a part of the *XIST* locus (Figure 3-4a). RNA-FISH revealed that only about 10% of cells in N-Tera2D1 actually express *XIST*, appearing as a single dot of expression in those cells (Figure 3-4b), with the remainder showing no *XIST* expression. Although N-Tera2D1 cells expressed stable *XIST* RNA, large stable *XIST* signals, as seen in female somatic cells (Figure 3-4c), were never detected in N-Tera2D1. The small *XIST* signal seen in N-Tera2D1 is consistent with the very low expression levels, which could not be entirely accounted for by the low percentage of expressing cells.

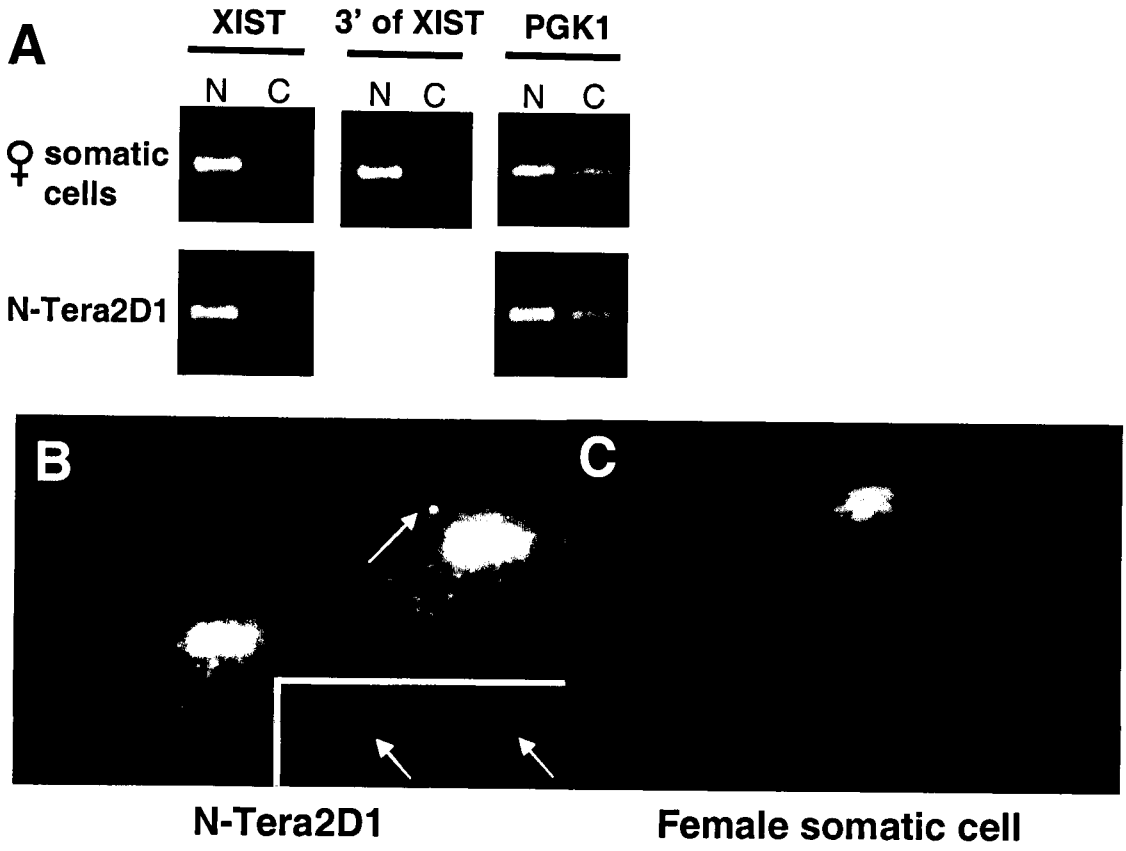


Figure 3-4. Sense *XIST* transcripts in female somatic cells and N-Tera2D1 remain in the nucleus. **(A)** RT-PCR on nuclear (N) and cytoplasmic (C) RNA. The 3' sense and antisense transcripts are found mostly in the nuclear fraction, showing a similar pattern of expression to *XIST* in female somatic cells. *PGK1*, a protein-coding transcript, is shown as a control. **(B,C)** RNA-FISH analysis of *XIST* expression in N-Tera2D1 and female somatic cells. In N-Tera2D1, an overlapping pinpoint *XIST* signal (yellow) can be detected with two different *XIST* probes (green and red) shown in the inset. The blue area is the DAPI-stained nucleus.

To determine if the specific association of the *XIST* transcript with the inactive X chromosome affects its stability, I examined the half-life of the *XIST* in mouse-human

somatic cell hybrids where, most likely due to species differences, human *XIST* is no longer able to localize to the X chromosome. By RNA-FISH, the *XIST* transcripts in these hybrids appear to occupy a more dispersed nuclear territory and are no longer tightly associated with the inactive X [226] (Figure 3-5a). Actinomycin experiments show that the half-life of the delocalized *XIST* transcripts in the hybrids is ~4-6 hours, similar to the half-life of localized *XIST* transcripts in normal female somatic cells (Figure 3-5b). Therefore, in a mouse background, delocalized human *XIST* transcripts are not unstable.

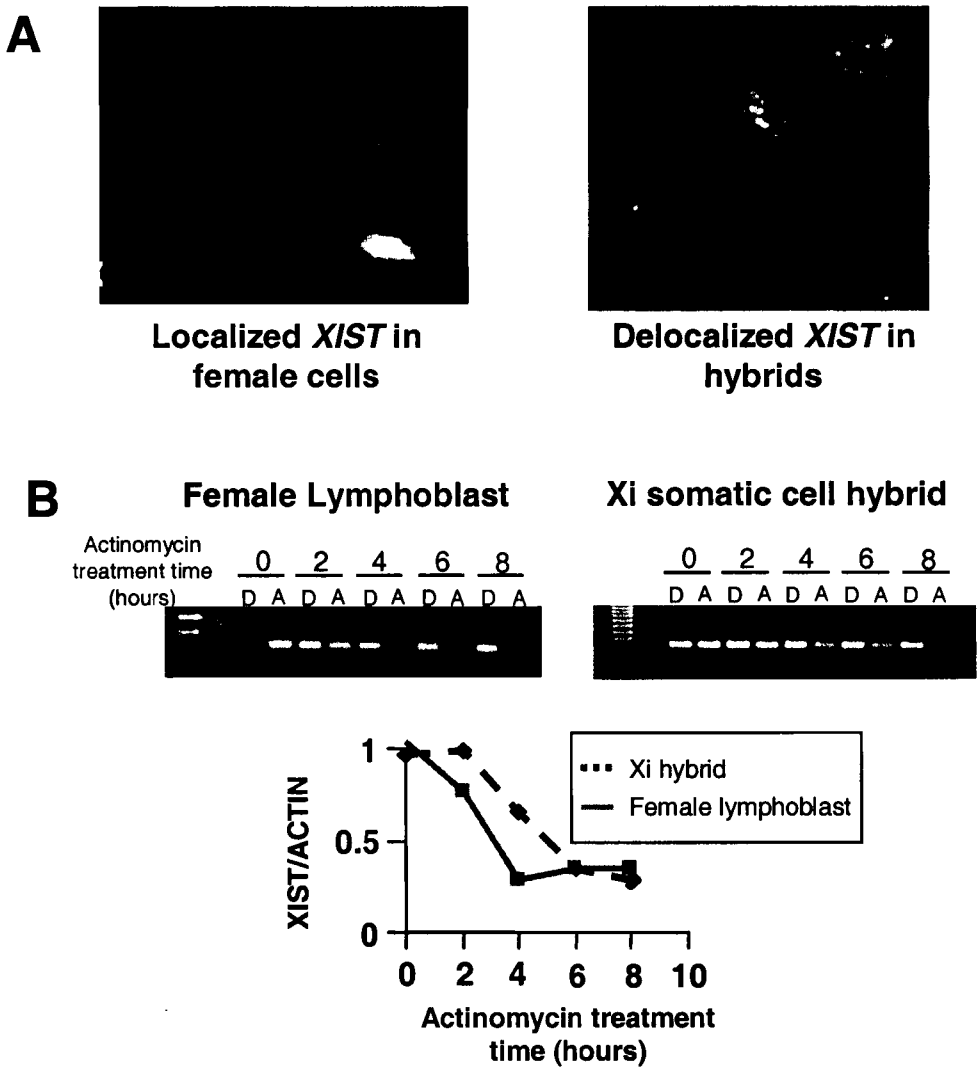


Figure 3-5. Delocalized human *XIST* transcripts in mouse-human somatic cell hybrids are stable. (A) In normal human diploid fibroblasts, (46, XX WI-38 cells), the *XIST* RNA forms a tight, localized signal in the nucleus (red) that co-localizes with the X chromosome (green). In mouse-human somatic cell hybrids that contain a single human inactive X, the *XIST* RNA (red) does not form a tight association with the X chromosome (human whole X chromosome library, green). **(B)** Half-life of the *XIST* transcript in a human inactive X-containing hybrid (t75-2maz-34-1b) and in a normal female lymphoblast cell line after treatment with the transcriptional inhibitor Actinomycin D (4μg/ml). The graph shows the levels of RNA present at various timepoints after treatment and each point represents the ratio of *XIST* levels to *ACTIN* levels. D = untreated DMSO control, A= Actinomycin treated

DISCUSSION

In an attempt to detect changes in *XIST* expression during human development, I have analyzed the extent, levels, localization and stability of sense transcription at the human *XIST* locus in somatic cells and the male-derived embryonal carcinoma cell line N-Tera2D1. In addition to demonstrating that human *XIST* in female somatic cells is present in similar copy number (~2000 transcripts) to that previously reported for mouse *Xist* [276], I now describe two additional sense isoforms at the *XIST* locus. First, sense transcription was detected in the male N-Tera2D1 EC cells. This expression does not extend across the entire *XIST* gene, perhaps reflecting alternate promoter and polyadenylation sites specific to this undifferentiated cell line. No other cell line examined has been shown to express this truncated isoform, including the human embryoid-body-derived cells examined by Migeon *et al.* (2001). RNA-FISH revealed that N-Tera2D1 is actually a mixed population of cells with *XIST* expression, seen as a small dot, in only about 10% of cells. Interestingly, although the sense *XIST* isoform in these cells does not form a large territory of expression, the transcripts are stable, with a half-life of around 5 hours, similar to the stability of *XIST* in female somatic cells, but sharply contrasting to the shortened half-life seen in undifferentiated mouse ES cells [54,55]. Thus, since these cells spontaneously differentiate in culture [254], we conclude that the small focus of expression observed by *in situ* hybridization reflects *XIST* expression in the undifferentiated population.

An additional sense transcript was also found in female somatic cells. Numerous ESTs have been reported at the 3' end of human *XIST* [277] and correspond to exons 7 and 8 [46]. The transcript described in this study, however, is even further 3' of *XIST* and does not match any ESTs in the database. These novel sense transcripts detected

downstream of *XIST* were found only in female somatic cells (four independent cell lines) or in inactive, but not active, X-containing mouse/human somatic cell hybrids (data not shown). These transcripts were localized to the nucleus and not the cytoplasm, consistent with their being related to *XIST* expression. However, the inability of extensive RT-PCR analyses to connect the transcripts indicates that they may represent alternate isoforms. We saw no evidence that this transcript was contiguous with *XIST* RNA using over a dozen different primer combinations and the transcripts were present at too low levels to detect by Northern or RNase protection. Although primers in this region readily amplify genomic DNA, failure to amplify may be due to the inability of the reverse transcriptase to copy certain RNA sequences efficiently. A murine equivalent to the low level 3' sense transcript that we report in human somatic cells has not been described, and we have been unable to correlate presence of this transcript with any cell type. Therefore, the significance of this alternate *XIST* isoform, if any, is unknown.

Because changes in *Xist* transcript stability appear to be important during the initiation stages of inactivation in mouse [54,55], determining the factors that affect *XIST* RNA half-life may be important to understanding how X inactivation is regulated. In mouse, prior to inactivation, *Xist* is expressed as an unstable transcript that appears as a small dot of expression from both the future active and inactive X's in both males and females. After differentiation, the *Xist* on the future inactive X is stabilized and accumulates along the chromosome forming a large domain of expression. Mouse-human somatic cell hybrids provide a unique opportunity to analyze the effect of *XIST*'s specific association with the inactive X chromosome on transcript stability since human *XIST* is unable to localize to the Xi in the mouse background [226]. Actinomycin

experiments on these hybrids show that despite their inability to associate with the X chromosome, delocalized human *XIST* transcripts remain stable which indicates that stability and localization of the transcript can be uncoupled and that stabilization of the transcript does not automatically lead to association with X chromatin [226]. There has been some suggestion that *Xist* localization to chromatin is required for the stability of the transcript in mouse since deletions that disrupt localization also appear to destabilize the transcript [53]. This discrepancy may be due to species differences or cell type differences since the mouse studies were performed in mouse ES cells. However, specific association with chromatin may still be important for human *XIST* since species differences may have prevented degradation of the human *XIST* transcript in the mouse hybrid background.

CHAPTER 4: ECTOPIC *XIST* TRANSCRIPTS IN HUMAN SOMATIC CELLS SHOW VARIABLE EXPRESSION AND LOCALIZATION

The material in this chapter has been published in:

Chow, J.C., Hall, L.L., Lawrence, J.B. and Brown, C.J. (2002) Ectopic XIST transcripts in human somatic cells show variable expression and localization. Cytogenet. Genome Res. 99: 92-8.

All RNA-FISH analyses presented in this chapter were performed by L.L. Hall.

INTRODUCTION

The *XIST* transcript has a very unique localization pattern in interphase nuclei, physically associating with the inactive X [47]. Because its upregulation is one of the earliest events in the inactivation process [56], models have proposed that the *XIST* transcript has a role in the recruitment of chromatin modifying proteins to the X chromosome [93,278]; however, it has not been determined how the *XIST* RNA localizes in *cis* to the chromosome from which it is transcribed. The association likely involves the interaction of the RNA with *cis*-linked DNA sequences and proteins. Transgene deletion studies as well as the use of PNA interference have identified regions of the mouse *Xist* transcript that are important for transcript localization and have determined that its specific association with the X chromosome is essential for the initiation of silencing [52,53]. Several proteins have been found to associate with the inactive X chromosome including the histone variants macroH2A1 and macroH2A2, the Polycomb group proteins Eed and Ezh2 and the BRCA1 protein [31,56,64,105,147]. Human *XIST* transcripts do not localize to the human X chromosome in mouse-human somatic cell hybrids, which

suggests that some factors affecting transcript localization and their interaction with surrounding DNA are species-specific [226]. However, since human *XIST* transgenes in mouse ES cells and developing mouse embryos do appear to be able to localize to surrounding mouse DNA [279,280], it is likely that the species-specific interaction is between specific DNA elements in the surrounding sequence and factors affecting localization.

The importance of *cis*-DNA elements for efficient association of the *XIST* transcript is also evident in X;autosome translocations [222,225,281,282] and *Xist*-containing transgenes [283] where *Xist* localization and silencing of autosomal material is attenuated. Autosomal sequences often acquired only a portion of the features that are normally associated with the inactive X, suggesting that autosomal DNA is unable to support the spread and/or the maintenance of the inactive state. Based on these observations, it has been suggested that "waystation" or "booster" elements on the X chromosome promote the spread of inactivation [179] and LINE-1 elements have been proposed as possible candidates for this function [217]. Sequence analysis of the human X chromosome revealed that, compared to autosomes, the X chromosome is enriched for LINE-1 elements that are distributed unevenly along the X, with regions that escape inactivation having a decreased density of LINE –1 elements compared to inactivated regions [218].

Although progress has been made in defining functional regions of *Xist*, *cis*-acting DNA elements, and interacting proteins, it is still unknown how these factors are involved in the localization of *XIST* and the subsequent chromosomal silencing. Transgenic experiments suggested that an important factor in *Xist* RNA spreading and

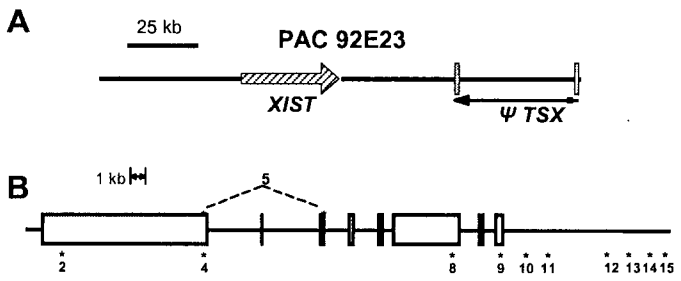
inactivation in *cis* was the number of integrated copies of the transgene. Unlike multicopy arrays, single copy transgenes were unable to induce a mature *Xist* RNA domain and subsequent inactivation of surrounding chromatin [284,285]. The authors proposed that single copy transgenes lacked the required number of elements necessary for *Xist* RNA association and spreading *in cis*; whereas in multicopy arrays, the number of relay elements would be increased allowing *Xist* transcript accumulation and inactivation. However, a single-copy, inducible *Xist* cDNA construct in mouse ES cells is able to form a large *Xist* accumulation, suggesting that it may be transcript levels that are critical for localization [51,56]. In addition, this inducible construct is also able to form a stable *Xist* signal in undifferentiated mouse ES cells, in contrast to the unstable transcripts from the endogenous locus, suggesting that high levels of *Xist* expression may promote the stabilization of *Xist* RNA [51]. Similarly, multiple insertions of a human *XIST* transgene in undifferentiated mouse ES cells form a large, stable *XIST* signal which could be due to an elevated level of transcripts, or due to species-specific factors affecting stability [280,285]. In contrast to mouse where *Xist* transgenes can only initiate inactivation in a discrete window in early development, ectopic human *XIST* transgenes appear to be capable of inducing chromosomal inactivation in human somatic cells [61]. Transcripts from an *XIST*-containing cosmid are able to localize to surrounding autosomal DNA which becomes transcriptionally silenced, hypoacetylated, late replicating and condensed. This suggests either that there is a difference between human and mouse in competence for induction of inactivation in somatic cells, or that the transformed state of the HT-1080 cells provided such competence.

To extend the previous transgene studies and to determine if increased copy number or expression levels are sufficient to promote *XIST* localization, I have introduced an *XIST*-containing PAC into the male fibrosarcoma cell line HT 1080. This PAC contains at least 50 kb of flanking sequences. Because the clones showed variable expression at the *XIST* locus we were able to examine the factors that affect *XIST* localization in human somatic cells. Comparisons of *XIST* transcript levels, the number of transgene integrations and the degree of *XIST* localization in five different transformant clones show that increased expression levels and increased copy number are not sufficient to achieve *XIST* localization.

RESULTS

To study the factors involved in the association of the *XIST* transcript with surrounding chromatin, I introduced an *XIST*-containing PAC into HT 1080, a human male fibrosarcoma cell line. The PAC (PAC 92E23; accession number U80459 and U80460) contains the entire *XIST* gene as well as 50 kb upstream of *XIST* and 85 kb downstream of *XIST* (Figure 4-1A). Five different single cell clones with expression from the transgene were isolated and the extent of expression determined by RT-PCR (Figure 4-1B). Expression continues 3' of exon 8 in three of the clones (L3.4A.7, L3.4B.3 and L3.2A.2). The extent of expression varies between clones, with one clone showing expression 9 kb downstream of exon 8 and the other two extending only 7 kb downstream of *XIST*. Although the transcript was synthesized from the same strand as the *XIST* RNA, I have not detected any continuity between the downstream transcript and expression within the gene body. Primer pairs immediately 3' of exon 8 (primer pair 6) do not give a PCR product and primers at other sites within *XIST* have also failed to detect splicing to this downstream region. Clones L3.4B.2 and L1.10.1 have

transcription that spans the entire *XIST* gene from exon 1 to 8 with no expression detected downstream of exon 8.



primer pair	XIST EXPRESSION 2	4	5	8	9	3' XIST EXPRESSION 10	11	12	13	14	15	PAC copy number
L1.10.1	+	+	+	+	+	--	--	nt	--	--	--	4.8
L1.10.1 subclone	+	+	+	+	+	--	--	nt	--	--	--	3.4
L3.4A.7	+	+	+	+	+	--	+	+	+	--	--	3.5
L3.4B.3	+	+	+	+	+	--	+	+	--	--	--	6.4
L3.4B.2	+	+	+	+	+	--	--	--	--	--	--	4.4
L3.2A.2	+	+	+	+	+	--	+	+	--	--	--	12.1

Figure 4-1. PAC 92E23 and the extent of *XIST* expression from HT-1080 transformant clones. (A) Diagram of PAC 92E23 (accession number U80459 and U80460). The PAC contains the entire *XIST* genomic sequence as well as ~50 kb upstream of the *XIST* initiation site and ~85 kb downstream of *XIST* exon 8. The PAC includes regions that show homology to the mouse testis-specific gene, *Tsx* that is disrupted and nonfunctional in humans [88]. **(B)** The extent of *XIST* expression in HT-1080 transformant clones. Using RT-PCR with primers pairs along the *XIST* gene and downstream regions, the extent of expression was determined in each of the transformant clones. The relative number of integrated PAC transgenes in each of the clones was determined using quantitative PCR. The ratio of *XIST* and an autosomal gene, *SDC4* [259], was taken for each clone, and normalized against a normal male.
+ = PCR product detected; -- = no expression detected by PCR; nt = not tested

RNA-FISH (performed by L. Hall) using an *XIST* genomic probe shows that there is considerable variability in the ability of the ectopic *XIST* transcript to associate with the surrounding chromatin (Figure 4-2). In L1.10.1, cells in early passages contain a highly localized *XIST* signal in the majority of cells (71%) which is associated with DNA condensation at the site of integration forming a well-defined Barr body at the site of ectopic *XIST* expression (Figure 4-2A). Clones L3.4B.2 and L3.4B.3 have highly variable phenotypes with 23% and 12% of cells showing a localized *XIST* phenotype, respectively. However, not all cells with a localized signal are associated with DNA condensation at the site of expression since only about 16% of L3.4B.2 cells have localized *XIST* expression with associated DNA condensation. More than 50% of cells in these two transformant lines have a more dispersed *XIST* signal not associated with any DNA condensation (Figure 4-2B, C). In L3.4A.7 only ~ 19% of the cells express a localized *XIST* signal and less than half of these (7%) were also associated with DNA condensation. The majority of cells in this clone express a small *XIST* signal (63%) (Figure 4-2D). In L3.2A.2, ~16% of cells formed a large, localized *XIST* signal, but only 5% showed DNA condensation. The rest of the cells had either a more dispersed *XIST* signal or a smaller focus of expression that did not seem to show transcript accumulation (Figure 4-2E). In all clones, however, transcripts in many of the cells seem to remain in the area of transcription. We also introduced the PAC into mouse A9 somatic cells to determine whether the localization and concentration of *XIST* transcripts requires human-specific DNA elements. RNA-FISH shows that the ectopic human *XIST* transcripts in mouse somatic cells, as in human somatic cells, are variably localized (Figure 4-2F);

however, the *XIST* transcripts are able to form a concentration of transcripts at the site of integration.

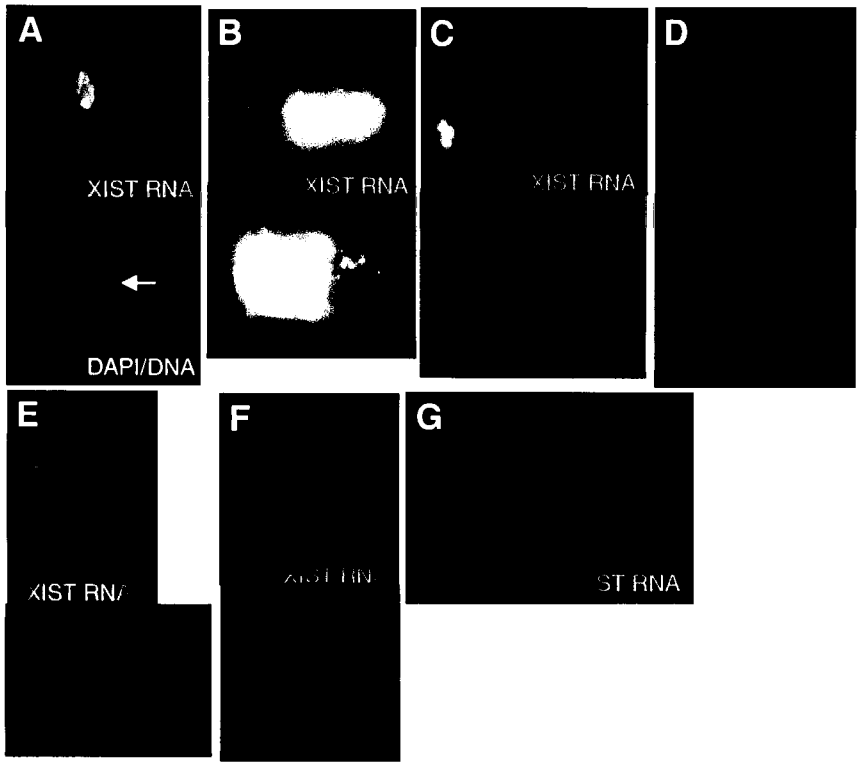


Figure 4-2. *XIST* shows variable localization by RNA-FISH in HT-1080 transformant clones. **(A)** In L1.10.1, the majority of cells (71%) show *XIST* localization (red, upper panel) with associated Barr body formation (DAPI-stain, lower panel). **(B)** L3.4B.2 has a variable *XIST* localization phenotype with 23% of cells with a large localized signal (upper panel). The majority of cells have a more dispersed *XIST* signal as shown in the lower panel. **(C)** L3.4B.3 is similar to L3.4B.2 with both localized (12%, upper panel) and delocalized *XIST* signals (lower panel). **(D)** In L3.4A.7 the majority of cells express a small *XIST* signal (lower panel) with only 19% showing a large localized *XIST* signal (upper panel). **(E)** In L3.2A.2, 16% of cells formed a large, localized *XIST* signal (upper panel), and over 70% have a dispersed signal (lower panel). **(F)** Mouse A9 somatic cells transformed with the human *XIST* PAC. Human *XIST* is variably localized, but transcripts remain concentrated in the area of transcription. **(G)** Later passages of L1.10.1 show a more dispersed *XIST* localization phenotype compared to earlier passages.

To determine whether differences in the number of integrated transgenes contribute to the variability in localization, we determined the relative transgene copy number in each of the HT-1080-derived clones by quantitative PCR (Figure 4-1B). The integrated copy numbers range from 12 copies in L3.2A.2 to 3-6 copies in the other four clones. As transcript levels may be another factor important for proper localization of the *XIST* RNA, we determined the levels of *XIST* expression by quantitative RT-PCR using primers spanning exons 2 to 3 (Figure 4-3). Clone L3.2A.2, with 12 copies of the transgene, has the highest level of expression with about two-fold more *XIST* RNA than a normal female lymphoblast line. L1.10.1, the clone with the highest percentage of nuclei with a localized *XIST* signal (71%), has similar *XIST* levels to a normal female. L3.4A.7 and L3.4B.2 both have expression levels that are about one half the levels found in normal females, while L3.4B.3 has the lowest transcript levels, being about one third normal female levels.

A recent report has indicated that ectopic human *XIST* transgenes can induce inactivation of surrounding chromosomal material in somatic cells [61]. Therefore, depending on the gene content of the surrounding autosomal material, there may be selection against *XIST* expression or localization. To determine if selection contributes to the variability in *XIST* expression and localization, we analyzed the *XIST* signal by RNA-FISH and quantitative PCR in early passages of L1.10.1 and compared them to cells in later passages. The majority (71%) of cells in early passages of L1.10.1 had a large, tightly localized *XIST* signal as described previously. In later passages, however, although the *XIST* RNA still seems to form a focus of transcript accumulation, the signal is not as compact, possibly due to selection against cells that inactivate surrounding genes

(Figure 4-2G). In addition, cells from later passages have lower *XIST* levels which are about one-half the transcript levels detected in earlier passages (Figure 4-3).

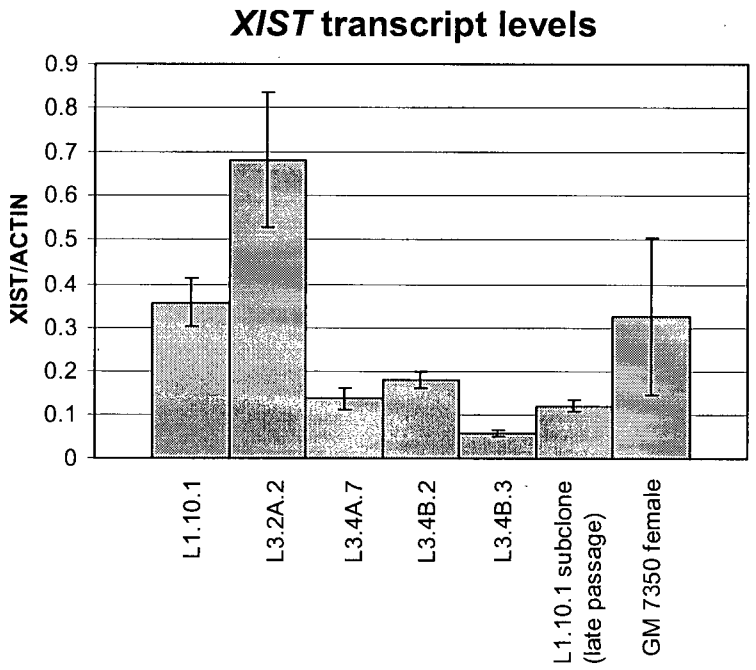


Figure 4-3. Quantitation of *XIST* expression in HT-1080 transformant clones. *XIST* transcript levels were determined by quantitative real-time PCR using SYBR green dye. PCR reactions were done in triplicate and normalized to *ACTIN*.

DISCUSSION

The *XIST* expression in the PAC transformant clones is generally similar to the expression observed from the endogenous *XIST* locus of female somatic cells. In three of the five isolated clones, *XIST* expression extends 3' of exon 8 which we have also observed in female somatic cells (see Chapter 3). There does appear to be some variability in the extent of the downstream transcripts since, in female somatic cells, expression extends ~10kb downstream of exon 8 whereas in the transformant clones, expression extends only to 7-9 kb. Because the PAC was not linearized before transfection, this variability could be due to differences in integration. We have not detected any continuity between the downstream transcript and expression within the *XIST* gene body, which is also true in female cells. Primer pairs immediately downstream of exon 8 do not give a PCR product (primer pair 10) and primers at other sites within *XIST* have also failed to detect splicing to this downstream region. Despite the lack of continuity with *XIST*, the downstream transcript is probably *XIST*-related because expression is female-specific and the transcripts are nuclear. However, based on the clones analyzed in this study, this downstream variant transcript does not appear to be required for localization, since expression in L1.10.1, which has the highest degree of localization, does not extend downstream of exon 8.

Several factors including the presence of "booster" elements, levels of the *XIST* transcript and the copy number of integrated transgenes have been proposed to affect the accumulation and spread of *XIST* along the chromosome. L3.2A.2, the clone with the highest number of integrated copies (12) has a similar *XIST* localization phenotype to clone L3.4B.3, which has only 6 integrated copies. L1.10.1, the clone with the highest percentage of nuclei with a large, localized *XIST* signal, has 4 copies of the PAC..

Therefore, it does not appear that increased copy number with necessarily lead to a more localized signal.

Localization was also not apparently dependent upon transcript levels. Clone L3.2A.2, with 12 copies of the transgene, has the highest level of expression with about two-fold more *XIST* RNA than a normal female lymphoblast line. However, despite the elevated transcript levels, only 16% of cells were able to form a large localized focus of *XIST* expression. L1.10.1, the clone with the highest percentage of cells with a localized *XIST* signal (71%), has similar *XIST* levels to a normal female. L3.4A.7 and L3.4B.2 both have expression levels that are about one-half the levels found in normal females, and 19% and 23% of cells have a localized *XIST* signal, respectively. L3.4B.3 has the lowest transcript levels, being about one-third normal female levels, and 12% of cells are still able to form a large, localized *XIST* focus. Therefore, greater levels of *XIST* expression do not always lead to increased transcript accumulation, although it remains possible that intraclonal expression differences lead to the variability in localization observed within a clone.

In a previous study by Hall et al. (2002) one of their three transformant lines showed good localization, and in this study, only one of five PAC clones showed a high percentage of localizing *XIST* signals. Thus the ability to form a well-localized domain of *XIST* may be unique to only a limited subset of clones. However, all clones contained a low percentage of cells with localized *XIST* transcripts and signs of associated DNA condensation; therefore, is seems that many sites are able to support some limited *XIST* concentration. The variable transcript localization suggests that autosomal DNA may not be able to maintain its interaction with *XIST* as efficiently as X chromosome DNA which

is supported by observations in some X;autosome translocation cases where the autosomal material is often not associated with *XIST* RNA despite being transcriptionally silenced [225]. We have not been able to identify the site of integration in these random integrants which would be required to determine whether the ectopic transcripts are in fact associating with *cis*-linked DNA sequences and whether there is a bias in integration sites that allow *XIST* expression and localization.

Although the *XIST* expression was analyzed by RNA-FISH within three passages of clone isolation, the decreased number of nuclei with a localized *XIST* signal could be due to selection against *XIST* localization and inactivation of surrounding genes. Transgene experiments in mouse ES cells suggest that initiation of X-inactivation can only occur within a discrete window early in mouse development; however, expression of an ectopic human *XIST* transgene in human somatic HT-1080 cells can initiate inactivation of surrounding autosomal material [61]. Early passages of clone L1.10.1 show the highest percentage of cells with a localized *XIST* signal with 71% of cells showing a tight *XIST* focus along with Barr body formation. Later passages, however, have a more dispersed *XIST* signal and a decreased incidence of Barr body formation. In addition, later passages of L1.10.1 tend to show a two-fold decrease in the levels of *XIST*, and silencing of the *XIST* transgene may also be a mechanism to prevent inactivation of important genes.

To determine whether *cis*-acting elements important for *XIST* accumulation are species-specific, we introduced the human *XIST*-containing PAC into mouse somatic A9 cells. Like the HT-1080 transformant clones, the A9 transformant clone has variable *XIST* expression and localization, but despite the mouse background, the human *XIST*

transcripts still concentrate at the site of integration. Human *XIST* transcripts appear to be able to localize more efficiently in this A9 transformant line than in mouse-human somatic cell hybrids containing a human inactive X. This supports the conjecture that the inability of the human *XIST* transcript to localize to the human X in mouse-human somatic cell hybrids may be due to species differences in the *cis*-acting elements required for transcript accumulation. Therefore, mouse *trans*-acting factors may be unable to localize the human transcript to surrounding human X DNA whereas in the A9 transformant line, the mouse factors are better able to localize human *XIST* to the surrounding mouse DNA.

In conclusion, the extent of expression and localization of *XIST* transcripts from ectopic transgenes is not dependent solely upon expression level or copy number. It is likely that many interacting factors including the number of transgene integrations, the level of expression, and the site of integration influence these phenotypes. To further study the specific effects of *XIST* transcription, DNA elements and proteins on X-inactivation, I have developed an inducible human *XIST* construct that can be integrated in single copy using site-specific integration to control for these various factors and minimize the effects of selection and silencing during cell culture (see Chapter 6).

CHAPTER 5: ANTISENSE EXPRESSION AT THE HUMAN *XIST* LOCUS

The material in this chapter has been published in:

Chow, J. C., Hall, L.L., Clemson, C.M. Lawrence, J.B. and Brown, C.J. (2003) Characterization of expression at the human *XIST* locus in somatic, embryonal carcinoma, and transgenic cell lines. Genomics 82: 309-22.

RNA-FISH results presented in this chapter were performed by L.L. Hall and C.M. Clemson.

INTRODUCTION

X inactivation initiates early in embryogenesis and is associated with an increase in stable, localized *Xist* transcripts in the mouse. Prior to inactivation, low levels of unstable *Xist* are seen from both mouse X chromosomes [54,55]. Expression of human *XIST* in very early male and female embryos [57,58] suggests that the situation may be similar in humans. The low level early *Xist* expression in mouse is accompanied by expression of an RNA antisense to *Xist*, *Tsix* [79,81]. Analyses of expression patterns of *Xist* and *Tsix* in differentiating ES and mouse embryos, as well as targeted insertions and deletions of the *Tsix* gene body and control regions suggest a regulatory role for *Tsix* in early *Xist* expression (reviewed in [286]). Prior to differentiation, *Xist* and *Tsix* are expressed at low levels from both X chromosomes in XX cells and from the single X in XY cells. Upon differentiation *Tsix* is silenced, and, coincidentally, *Xist* is upregulated on the future inactive X, while on the future active X both genes are silenced [82,83]. Expression of an antisense transcript at the human *XIST* locus has been described in murine ES cells containing a YAC including human *XIST*, as well as in human embryoid-body-derived cells [88]. Although expression of the antisense appears to be transient and

is not present in somatic cells of adult females, it may not be under the same strict developmental regulation as mouse *Tsix* since some expression does persist in human fetal somatic cells after the establishment of X inactivation, surprisingly coincident with the presence of the inactive X [89]. To address the potential role of antisense transcripts in the regulation of the *XIST* RNA in humans, I examined the transcriptional patterns at the human *XIST* locus in various cell types. There are many similarities, but also key differences, between X inactivation in mouse and humans so that the comparison of expression patterns provides insights into the potential role of transcripts from the human *XIST* locus in the important developmental process of X inactivation.

It has been suggested that differences between human and mouse *Tsix* expression could be attributed to differences between imprinting of X inactivation in the two organisms [89]. In mouse, preferential inactivation of the paternal X is seen in the extraembryonic tissues [11], although the somatic tissues of the embryo proper undergo random inactivation. In humans, there does not appear to be a strict imprint on X inactivation in extraembryonic tissues, as random or nearly random X inactivation is generally observed (reviewed in [13]). Non-coding RNAs, including antisense transcripts, have been associated with several imprinted loci [287,288], and in murine extraembryonic tissues, *Tsix* expression is imprinted, being expressed exclusively from the maternal X [289]. However, targeted deletions of the *Tsix* promoter suggest that *Tsix* is involved in the regulation of *Xist* expression during both imprinted and random X inactivation [79,82,83,289]. There are additional differences between human and mouse X inactivation, including the apparently greater number of human genes expressed from

the inactive X chromosome [213] and the ongoing association of murine but not human *Xist* RNA with the inactive X chromosome at metaphase [47,282].

Transgene experiments in mice have attempted to determine the extent of sequences required for complete X inactivation centre function, including sequences required for the establishment and spread of the inactivation signal. Using a cDNA *Xist* construct, under the control of an inducible promoter, Wutz *et al.* (2000) demonstrated that *Xist* expression alone is sufficient to initiate silencing. However, multiple copies of genomic YAC *Xist* transgenes were also found to be required for the initiation of silencing, suggesting that the endogenous *Xist* promoter requires additional sequences for the establishment of a particular chromatin domain for inactivation [285]. Sequence elements 5' of the *Xist* gene have been shown to be important for its function. For example, changes in chromatin modifications extending 5' of *Xist* coincide with the initiation of inactivation and point to important regulatory elements upstream of the gene [93,94]. Furthermore, transgenes deleted for the region upstream of *Xist*, although they express unstable transcripts in undifferentiated ES cells, are unable to express stable *Xist* when differentiated [92]. Sequences immediately upstream of *Xist/XIST* have been shown to be conserved between humans and mouse [69], and in general mouse and human *XIST* share similar gene structures. However, complete sequences of the X inactivation centre region have recently been analyzed in mouse, human and cow, and little conservation is observed in regions further outside of *XIST* [50]. There are small regions of homology 3' of *Xist* representing rearranged exons of the testis-specific gene, *Tsx*, which is present in mouse but has become an inactive pseudogene in humans; however, despite the importance of the CpG island associated with the *Tsix* transcript in

mouse [82,83], the sequences in this region do not appear to have been conserved in humans. Interestingly, the nearest gene upstream of *XIST, JPX/ENOX,* is conserved in humans, cows and mice and encodes an untranslated RNA that partially escapes inactivation in mouse [50,290]. We demonstrate that this gene is fully expressed from the human inactive X in human/mouse somatic cell hybrids, delineating the 5' end of the Xi-specific transcriptional domain in humans.

To address whether there is an equivalent to *Tsix* in human embryonic cells, I examined the extent of both antisense transcription at the human *XIST* locus in embryonal carcinoma (EC) cells. By strand-specific RT-PCR, low-level antisense transcripts were detected in a subset of N-Tera2D1 cells. I further assessed antisense transcription in clones of a human male cell line transformed with a *XIST*-containing PAC recently demonstrated to show variable *XIST* expression (Chow *et al.*, in press; see Chapter 4). The presence of stable, localized *XIST* in a clone with ongoing antisense expression suggests that the human antisense does not serve to destabilize *XIST* expression. Although there are significant differences in the extent and stability of transcripts compared to the murine homologues, conservation of an antisense transcript at the human *XIST* locus suggests that it may be an important regulatory feature not only associated with imprinted transcripts.

RESULTS

Extent of antisense expression in N-Tera2D1 and an XIST transgenic clone

I determined the extent and orientation of antisense expression in the *XIST* region in the male-derived EC line, N-Tera2D1 and in male and female somatic cell lines by strand-specific RT-PCR. Antisense expression was detected 3' of *XIST* in N-Tera2D1,

but not in somatic cells (male or female). This antisense expression could be detected in a region from 3 to 10 kb downstream of *XIST* and also in the 3' end of *XIST*, overlapping exons 6 to 8 (Figure 5-1, see also Figure 5-5). It is not clear whether this antisense expression represents a single RNA transcript or multiple smaller ones as long-range reverse transcription has failed to amplify a single continuous RNA species, and no product immediately 3' of exon 8 (primer pair 10) has been detected (Figure 5-1). The transcript(s) were present at low levels and generally required more than twice the amount of RNA in the reverse-transcription reaction combined with 40 cycle PCR to detect bands. While these findings were reproducible, expression was often lost upon continued culturing of the cells.

I also characterized expression from an *XIST*-containing PAC introduced into male somatic HT-1080 cells (PAC 92E23 contains at least 50 kb of flanking genomic DNA; accession numbers U80459 and U80460). Six different transformant clones with some expression from the transgene have been identified (Chow et al., in press; see Chapter 4). Of these, four showed no antisense expression, one showed only a very low level of limited antisense expression with no sense expression (data not shown) and one clone (L1.10.1) showed both sense and antisense expression. By strand-specific RT-PCR, this antisense transcript is similar to the antisense expression in N-Tera2D1, initiating 10kb downstream of *XIST*. However, expression extends a bit farther into *XIST* to the end of exon 1 (Figure 5-1a), and is continuous without the gap 3' of exon 8 that is present in N-Tera2D1 (primer pair 10).

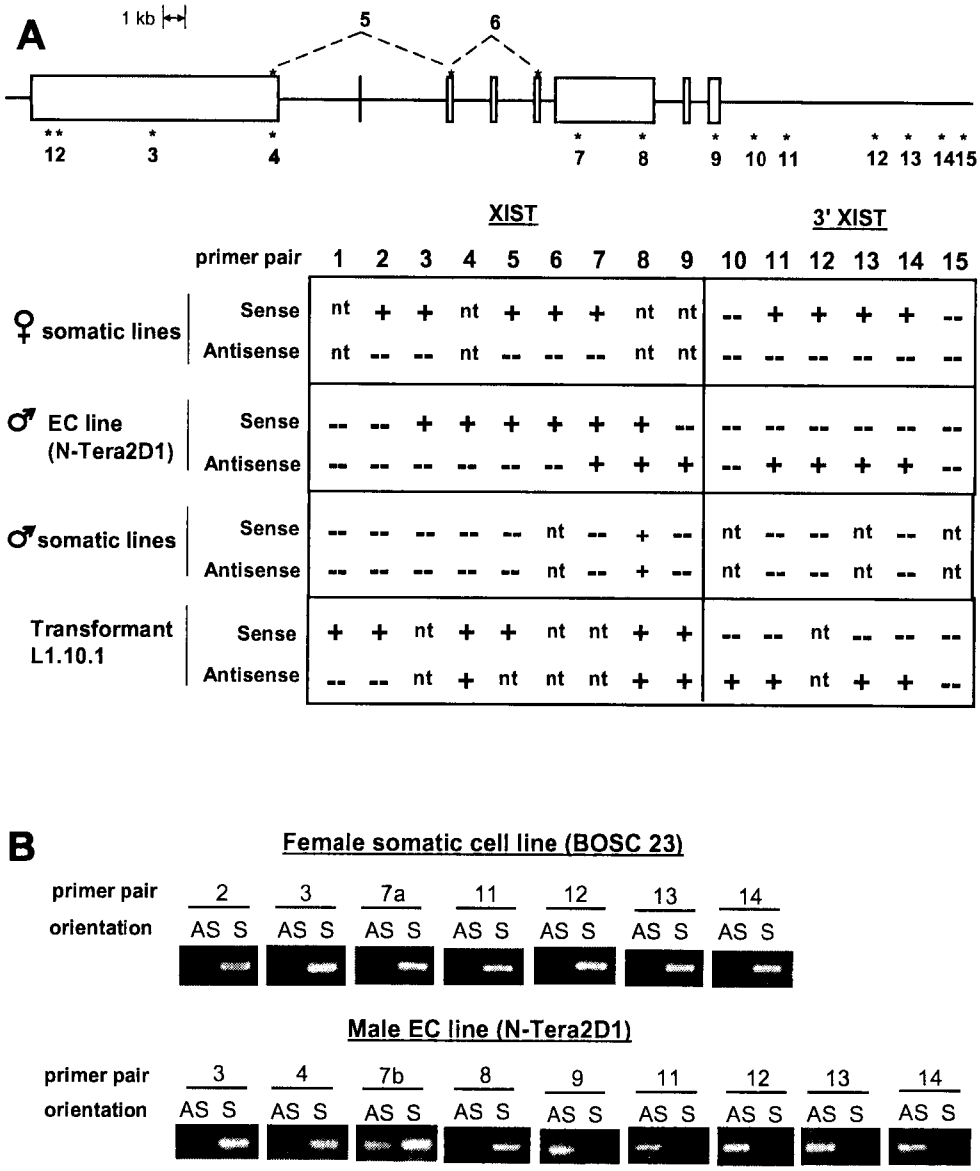


Figure 5-1. Extent and orientation of expression within *XIST* and 3' of *XIST* in somatic cells, N-Tera2D1 EC cells and from an ectopic transgene. (A) Strand-specific RT-PCR to determine the orientation of transcripts. The asterisks mark the location of primer pairs used for RT-PCR. To detect a transcript in the sense orientation, an antisense primer was used to prime cDNA synthesis while for antisense transcripts, the sense primer was used to prime cDNA synthesis. + = PCR product formation; -- = no PCR product; nt = not tested. Very faint bands were sometimes observed with high cycle PCR for male

cDNA using primer 8. **(B)** Representative strand-specific RT-PCR in female somatic cells and in N-Tera2D1. AS (antisense) and S (sense) indicate the orientation of the transcript amplified by RT-PCR.

Quantity, Stability and Localization of Antisense Transcripts

The antisense transcript(s) in N-Tera2D1 are found at very low levels. By real-time quantitative PCR, and consistent with the number of PCR cycles required for detection in regular RT-PCR, it appears that the antisense transcript is expressed at similar or just slightly lower levels than the sense transcript in N-Tera2D1 which is about 1/100X the levels of *XIST* in female somatic cells (see Chapter 3).

In murine cells prior to X inactivation, low levels of *Xist* expression are accompanied by an extremely shortened half-life of the RNA thought to be due to the expression of the overlapping antisense transcript, *Tsix*. To determine if the sense transcripts in N-Tera2D1 and the L1.10.1 transgenic clone are also destabilized by the presence of an antisense transcript, we treated cells with Actinomycin D to arrest transcription and then determined the levels of RNA at various timepoints after treatment by real-time quantitative RT-PCR (Figure 5-2). In female somatic cells, the half-life of *XIST* is ~4-5 hours, as previously described [226] (see Chapter 3). In both N-Tera2D1 and L1.10.1, despite the presence of an antisense transcript, the *XIST* transcript stability is similar to that in female somatic cells. Half-life experiments also suggest that the 3' antisense transcript in N-Tera2D1 is not particularly unstable with a half-life of 1-4 hours (data not shown).

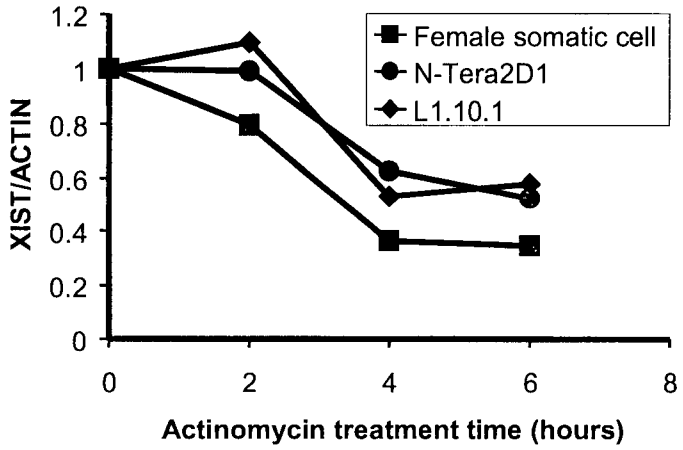


Figure 5-2. Sense *XIST* expression is not destabilized by the presence of antisense transcription in N-Tera2D1 and the PAC transformant clone L1.10.1. Half-life of *XIST* RNA in female somatic cells (squares), N-Tera2D1 (circles), and L1.10.1, an HT 1080 clone transformed with an *XIST*-containing PAC (diamonds). The graph shows the levels of RNA present at various timepoints after treatment with the transcriptional inhibitor Actinomycin D (4μg/ml). PCR reactions for each timepoint were performed in triplicate for both *XIST* and *ACTIN*. Each point on the graph represents the average of *XIST* values divided by the average of *ACTIN* values. The average variation in the replicates was 18% for *XIST* and 10% for *ACTIN*.

RT-PCR of RNA isolated from nuclear and cytoplasmic fractions shows that the 3' antisense transcript in male-derived N-Tera2D1 cells remains in the nucleus (Figure 5-3a), which suggests that it is a part of the *XIST/TSIX* locus. Unfortunately, the low transcript levels precluded further analysis of expression by FISH; however, we were able to further characterize the antisense expression in the L1.10.1 PAC transformant clone. In L1.10.1, RNA-FISH has shown that the ectopic *XIST* transcripts can form a

localized signal similar to the *XIST* signal in female somatic cells at the site of integration in about 70% of nuclei (Figure 5-3b and [61]). Probes specific to the antisense transcript show that the 3' transcript is present in all nuclei with sense *XIST* expression and both co-localize and are therefore expressed from the same transgene integration site (Figure 5-3b). The antisense transcript signal is weaker than the *XIST* signal, consistent with real-time quantitative PCR results that show the antisense to be present at about 1/10X the levels of the sense transcript (data not shown). Furthermore, although the antisense signal localizes to the same territory as *XIST*, the antisense territory is consistently smaller than the *XIST* territory (Figure 5-3b). Previous results have shown that the transformant clone L1.10.1 has about 5 integrated copies of the *XIST*-containing PAC (see Chapter 4), but expresses *XIST* at levels that are very similar to those from the endogenous locus in female somatic cells (Chow et al., in press; see Chapter 4).

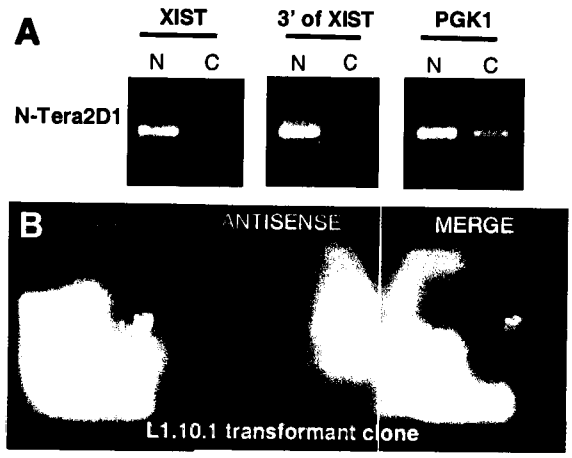


Figure 5-3. Antisense transcripts 3' of *XIST* are found predominantly in the nucleus and co-localize with the sense *XIST* transcript. **(A)** RT-PCR on nuclear (N) and cytoplasmic (C) RNA from N-Tera2D1. *PGK1*, a protein-coding transcript, is shown as a control. **(B)** RNA-FISH analysis of *XIST* (green) and antisense expression downstream of *XIST* (red) from the transgene PAC 92E23 in the HT-1080 transformant clone L1.10.1. The transgene-derived sense *XIST* transcript forms a localized signal similar to female somatic cells. An antisense specific probe shows that the antisense transcript co-localizes with *XIST*.

Extent of the XIST region

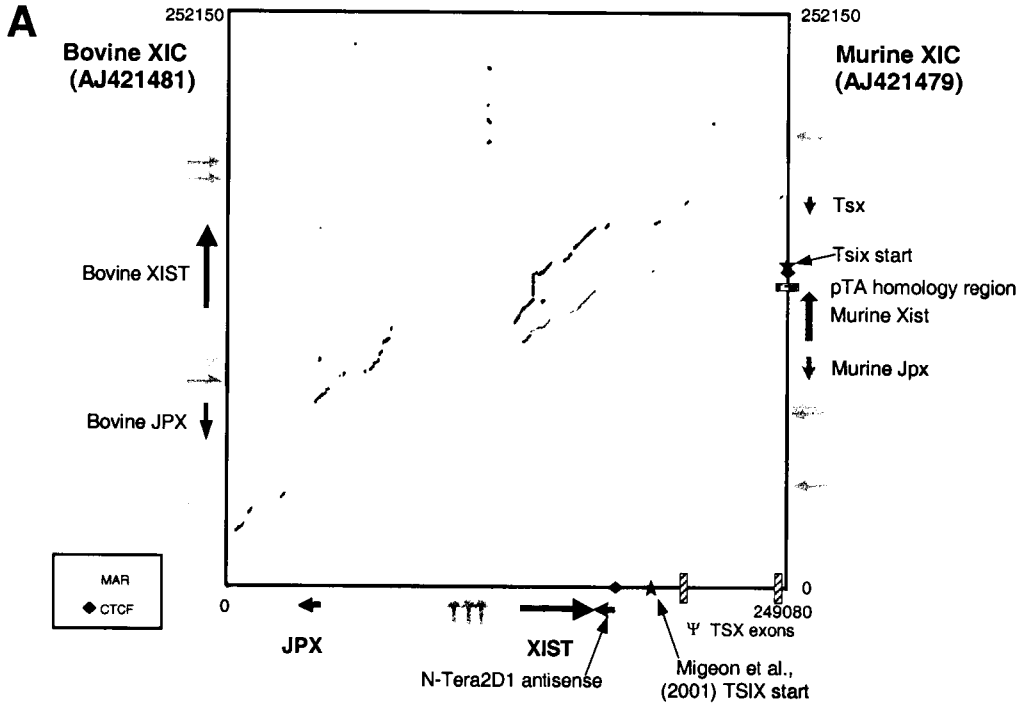
In order to identify conserved sequences that might reflect important functional elements, we compared the sequence of human, mouse and cow *XIST* regions (Figure 5-4a). Extensive similarity is seen for exons 1 to 8 with mouse and cow sequence [50]. The antisense transcription described in this study initiates ~10 kb downstream of exon 8, overlapping a small region of homology with mouse that is, however, not shared with

cow. Additional small regions of homology between human and mouse correspond to exons of the murine testis-specific gene *Tsx*, which is present as a rearranged pseudogene in humans [88] and absent from the region in cows [50]. Primers in these regions reveal no expression in the human cell lines tested (data not shown). In mouse, a region 3' of *Xist* shows homology to the pre-T cell receptor α upstream enhancer and may have arisen from a duplication of the *pTα* region to the *Xist* locus [291], and this region is absent in humans and cows. In human *XIST*, a 100 bp stretch at the beginning of exon 7 shows 100% homology to an orphan nuclear hormone receptor gene (accession number, Z30425), and this region shows limited homology to both the cow and mouse sequences. The importance of these rearrangements for the function of *XIST* and X inactivation, if any, is unknown.

To determine the extent of inactive-X specific expression I analyzed transcription of *Jpx/Enox*, which has recently been identified 10 kb 5' to the murine *Xist* on the antisense strand [50,290]. This sequence is conserved in all 3 species, and the human homolog, *JPX/ENOX*, initiates ~90 kb upstream of *XIST*. Based on ESTs, the human *JPX/ENOX* transcript, like the murine homolog appears to be alternatively spliced (*e.g.* AV714079, AK056172). Human-specific primers from human exon 1 and exon 2 (which is homologous to mouse exon 1) were used to amplify cDNA from human/mouse somatic cell hybrids retaining a human active or inactive X chromosome (Figure 5-4b). Expression was detected from both the active and inactive X chromosomes, and therefore the gene appears to escape inactivation; an expression pattern distinct from *XIST* sense or antisense transcription. A comparison of PCR product band intensities in a reverse transcription reaction dilution series shows that *JPX/ENOX* levels are similar in both Xa

and Xi hybrids, suggesting that the gene likely completely escapes inactivation (data not shown).

Regions between *JPX* and *XIST* in human, mouse and cow show significant scores on MAR-finder, a program designed to identify DNA regions involved in the attachment to the nuclear matrix (http://www.futuresoft.org/MAR-Wiz/) (Figure 5-4a). In addition, a putative MAR is detected near the human antisense start site; however, mouse sequences do not have a similarly located MAR at the *Tsix* initiation site, with the first significant MAR score appearing after *Tsx*. A recent report has demonstrated that the insulator and transcription factor CTCF binds to the mouse *Tsix* region and that the human sequence also contains ~10 CTCF motifs that appear to be clustered near the start site of the antisense transcript seen in this study [96]; however, no substantial enrichment for CTCF sites is seen in the cow *XIST*-containing region.


A
252150
Bovine XIC
(AJ421481)
252150
Murine XIC
(AJ421479)
Bovine XIST
Bovine JPX
Tsx
Tsix start
pTA homology region
Murine Xist
Murine Jpx
MAR
CTCF
0
JPX
XIST
N-Tera2D1 antisense
0
249080
Ψ TSX exons
Migeon et al.,
(2001) TSIX start
Human XIC (AL353804+U80460)

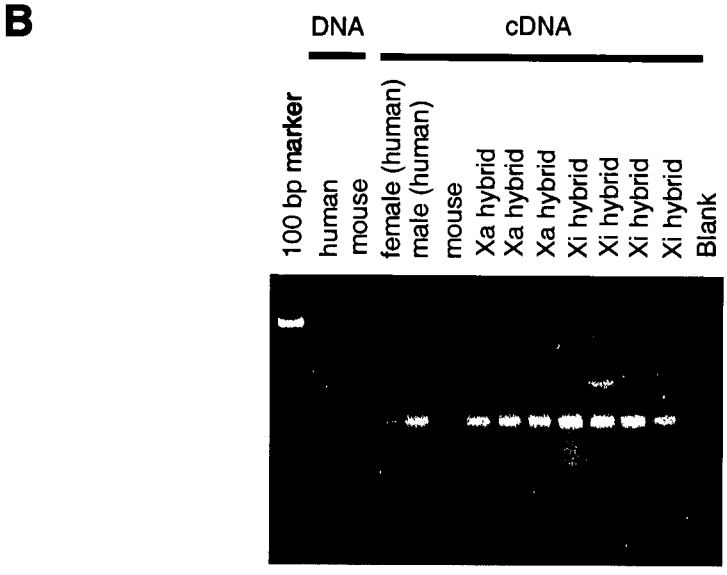

B
DNA
cDNA
100 bp marker
human
mouse
female (human)
male (human)
mouse
Xa hybrid
Xa hybrid
Xa hybrid
Xi hybrid
Xi hybrid
Xi hybrid
Xi hybrid
Blank

Figure 5-4. Sequence conservation at the X inactivation centre region in human, mouse, and cow. **(A)** PIPmaker dotplot comparison of human (AL353804+U80460), mouse (AJ421479) and bovine (AJ421481) X inactivation centre regions. The grey markers indicate regions of homology between mouse and human, and the black markers represent homologies between cow and human sequences. The diagonal lines represent the gap-free segments in alignments generated by BlastZ that have at least 50% identity between the two species. The vertical grey arrows indicate regions that showed significant scores (average strength of at least 0.6) using the MAR-finder program (http://www.futuresoft.org/MAR-Wiz/). **(B)** *JPX* expression from the inactive X chromosome. Primers flanking the 118 bp intron 1 of JPX were used to amplify human and mouse DNA or cDNA, as well as cDNA from mouse/human somatic cell hybrids containing a human active X (Xa: t60-12, AHA-11aB1, A23-1aCl5) or inactive X (Xi: t86-B1maz1b-3a, t11-4Aaz5, t75-2maz34-4a, LT23-1E2Buv5Cl26-7A2)

DISCUSSION

Antisense Transcription at the XIST Locus

I have analyzed the extent, orientation, stability, and quantity of transcription at the human *XIST* locus (summarized in Figure 5-5). As in undifferentiated mouse ES cells, the human EC line N-Tera2D1 expresses a low-level of antisense transcription that can be detected overlapping, and extending 3' to *XIST*. This antisense expression appears to be unique to undifferentiated cell types since we do not detect it in any somatic lines (lymphoblast and fibroblast cell lines). The antisense transcript reported by Migeon et al. (2001) in human embryoid-body-derived cells as well as from a YAC containing the human *XIST* locus introduced into mouse ES cells appears to be longer than reported here for N-Tera2D1, initiating ~17kb further downstream of *XIST* than the transcripts we have detected. In both studies, however, the antisense transcript shows limited overlap with *XIST*, ending before exon 1. The main transcriptional start for mouse *Tsix* lies 12 kb

downstream of *Xist* and spans 40 kb, completely overlapping the *Xist* locus; however, about 30-60% of *Tsix* transcripts are spliced, eliminating *Xist/Tsix* homology except at the 5' end of exon 1 [80]. Recently, Migeon has described persistent antisense expression in additional fetal cell types [89] coincident with *XIST* expression from the Xi. We have not examined fetal somatic cells, however, the male N-Tera2D1 cells never show a large focus of *XIST* expression, and thus the antisense reported here is not associated with an inactive X, but rather appears to be a unique developmentally-regulated transcript. Because of the difficulty in detecting the low-level antisense transcripts in the N-Tera2D1 EC line we were unable to demonstrate that the sense and antisense transcripts are concurrently expressed, which led us to analyze expression in a series of *XIST*-containing transgenic clones.

The antisense transcript in L1.10.1 is similar to that in N-Tera2D1, however, it has slightly more overlap with sense *XIST*, and appears to be continuous (see summary Figure 5-5). Because the antisense transcript can be expressed from a transgene introduced into somatic cells, it is likely that it is normally silenced in development by chromatin modifications. Intriguingly, this is a clone that has previously been shown to have localized *XIST* expression [61], suggesting that antisense expression is not incompatible with *XIST* localization. However, the transgene is present in multiple copies and although *XIST* and the antisense transcripts co-localize by RNA-FISH, we cannot rule out that they are expressed from different transgenes at the same integration site.

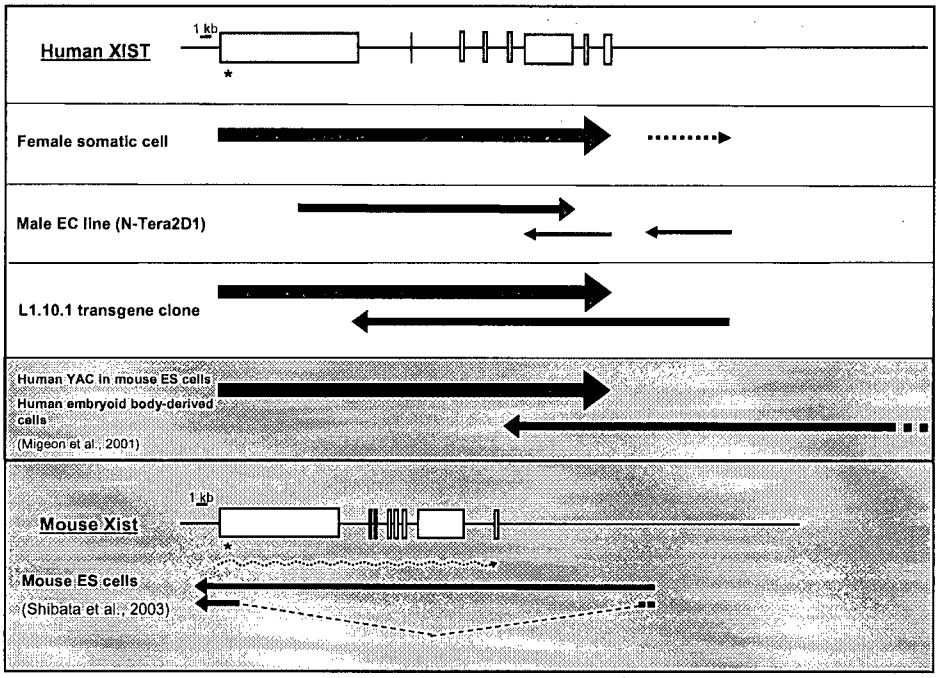


Figure 5-5. Summary of expression at the *XIST* locus. The thickness of the arrows reflects the relative abundance of the transcripts. In female somatic cells, sense *XIST* transcripts are present at about 2000 copies/cell. Variable downstream sense transcription is also detected (dotted line) and is present at about 1/100X the expression levels within the *XIST* gene body. In the transformant clone L1.10.1, *XIST* transcripts from the PAC transgene are expressed at levels similar to female somatic cells. An overlapping antisense transcript, present at about 1/10X the levels of the sense transcript, initiates downstream of *XIST* and extends into exon 1. In N-Tera2D1, a male embryonal carcinoma cell line, a truncated *XIST* isoform is expressed at 1/100X the levels of *XIST* in female somatic cells. Since only 10% of cells appear to be expressing this truncated transcript, the *XIST* transcript levels in these cells must be 1/10X that in female somatic cells. An overlapping antisense transcript is also expressed in N-Tera2D1 and initiates 10 kb downstream of *XIST* and extends into exon 6. A similar antisense transcript has been reported by Migeon et al. (2001), but it is longer initiating ~22-27 kb downstream of *XIST* exon 8. The antisense expression from the human *XIST* locus differs from mouse *Tsix* detected in undifferentiated ES cells which is expressed at 10-100 fold molar excess over the sense and completely overlaps the *Xist* transcript which is unstable (wavy dotted line) [54,55,80]. The asterisk marks the conserved 5' repeats important for silencing [53].

Role of Antisense Transcription at the XIST Locus

Is the antisense transcript described here reflective of a human homolog of the mouse *Tsix* gene? The human and mouse antisense transcripts both initiate downstream of *XIST/Xist*, and are normally expressed only during early development; however, the EC and transgenic cells are clearly non-physiological systems to study early development. Although EC cells are germ-cell derived lines, determining the stage in development that they mimic is problematic and a recent study has shown that *Xist* levels can vary substantially between different cells of the early embryo [292]. The N-Tera2D1 line tends to spontaneously differentiate in culture [254] and often loses *XIST* expression, precluding a more thorough *in situ* analysis of expression in these cells. Furthermore, the region downstream of *XIST* contains many repetitive elements and transcription from retroelements may be less rigidly controlled than in the normal epiblast cells of a human embryo, such that the antisense transcription detected may reflect promiscuous transcription. However, the presence of an antisense transcript at the human *XIST* locus in several cell types both in this study and as described by Migeon et al. [88,89], in addition to the similarity in the transcription patterns between the mouse and human transcripts, argues in favor of the antisense expression at the human locus reflecting biologically-relevant transcription. Thus comparison between expression at the human and mouse *Xist/Tsix* locus may provide insights into its role in mammalian X-chromosome inactivation.

The small dot of *XIST* expression in N-Tera2D1 appears very similar to the *Xist* signal seen in undifferentiated mouse ES cells and mouse preimplantation embryos, and suggests an inability of the transcripts at these early stages to accumulate on the X chromosome. In mouse, this low-level expression has been attributed to the short half-

life of the transcript (t1/2 ~ 30min). However, in N-Tera2D1, the sense *XIST* transcript is apparently not destabilised, therefore, the RNA's inability to accumulate along the chromosome likely reflects a lower rate of transcription. Antisense expression in the L1.10.1 transformant clone also does not appear to affect sense *XIST* stability or localization. Furthermore, we do not detect substantial instability of the antisense transcript, as might be expected if sense/antisense duplexes were being formed and degraded.

Differential transcript stability may be attributable to the presence of the antisense in mouse cells even though the antisense in humans is not destabilizing as there are significant differences between human and mouse *Xist/Tsix* transcription. First, destabilization may require an excess of antisense transcripts, as quantitative analyses of *Xist* and *Tsix* transcript abundance in undifferentiated mouse ES cells revealed that the *Tsix* RNA is present at 10-100 fold higher levels than the *Xist* RNA [80], while the antisense transcripts in the N-Tera2D1 EC line and the L1.10.1 PAC transgenic cell line are present at lower levels than the sense transcript. Second, the human antisense does not completely overlap the *XIST* transcript. Notably, it does not extend into the 5' conserved A repeats that are important for silencing [53], and where both the spliced and unspliced versions of mouse *Tsix* overlap with *Xist* (see Figure 5-5).

Extent of the conserved Xist/Tsix domain

While our findings support the conservation of a developmentally-regulated antisense transcript initiating 3' to *XIST/Xist*, there is no substantial sequence conservation in this region between human, mice and cows. Preferential paternal X-chromosome inactivation, as seen in mice, but not humans, has recently been reported in

bovine extraembryonic tissues [12]. Thus, the lack of sequence homology, including the concentration of CTCF sites, between species is not simply due to the lack of imprinted *XIST* expression in human extraembryonic tissues. *Tsx*, a testis-specific gene in mice, is present only as rearranged exons in humans, and thus may demarcate an evolutionary breakpoint between mouse and humans [88], and is also likely to reflect the end of a domain of inactive-X specific expression.

Two recent reports have identified another apparently non-translated RNA transcript, named *Jpx*, or *Enox* that is only 10 kb upstream of the beginning of mouse *Xist* [50,290]. It seemed somewhat surprising that a second non-coding gene would be found adjacent to *XIST*, so we examined its expression from the human X. *Jpx/Enox* is predominantly expressed from the mouse active X chromosome, with low level transcription being observed from the inactive X chromosome [290]. However, in humans, we have shown that *JPX/ENOX*, which initiates ~90 kb upstream of *XIST*, fully escapes X chromosome inactivation, being expressed approximately equally from both the active and inactive X chromosome. Thus the extent of silencing on the active X chromosome does not extend to human or mouse *JPX/Jpx,* and there is no evidence that expression of this non-coding RNA is related to *XIST* expression.

The substantial differences in transcription in the *XIST* region may be reflective of differences in the inactivation process between mouse and humans. It has been suggested that as human *TSIX* is not required for imprinting that it may no longer act as an antagonist of *XIST* expression [89]. Our finding that the stability is not disrupted would be consistent with such a hypothesis; however, we observe developmental restriction of the antisense transcript, which is also consistent with a conserved function, but perhaps

operating through interactions other than destabilization of the sense transcript. The human antisense expression may simply be an evolutionary remnant with little of the function of the mouse counterpart; however retention of transcription in this region despite a general loss of sequence conservation seems to suggest some function for this transcription. While the spliced *Tsix* transcript shows little relation to the human antisense, a considerable amount of the mouse *Tsix* transcript does remain unspliced, and levels are highest at the 5' end of *Tsix,* which may indicate a functional role for transcripts in this region where the human antisense is also expressed [80].

CHAPTER 6: IDENTIFICATION OF *XIST* LOCALIZATION DOMAINS

INTRODUCTION

The *XIST* transcript's unique association with the inactive X chromosome involves several factors and requires not only specific sequences and structural features of the *XIST* RNA, but also *cis*-acting DNA elements and likely *trans*-acting protein factors. In addition, transcript levels and stability appear to be important in the RNA's ability to accumulate along the X chromosome. Because of these multiple variables, randomly integrated *XIST* transgenes exhibit widely variable phenotypes with respect to *XIST* localization (see Chapter 4).

Functional domains in mouse *Xist* have been identified through a series of inducible *Xist* deletion constructs integrated into the *Hprt* locus on the X chromosome in mouse [53]. The conserved A- repeats at the 5' end of exon 1 are required for transcriptional repression and mutational analysis of this repetitive element has demonstrated that the silencing activity of this motif is independent of its position in the RNA since silencing still occurs when the element is moved to the 3' end of the RNA. However, its silencing function does require the formation of a particular stem loop structure and this structure has been hypothesized to be the motif recognized by putative binding factors [53]. Localization of the *Xist* transcript appears to be accomplished by multiple, functionally redundant domains that do not appear to have any common motifs; therefore, association with surrounding chromatin may involve many low-affinity binding sites that act in a co-operative fashion [53]. In addition, the transcript's ability to form a localized structure may be influenced by other factors like the chromatin structure of the surrounding region as well as the levels of the transcript.

Because transcript levels may play a role in *XIST* function, the promoter used in *XIST* transgenes is an important consideration. Although utilizing the endogenous *XIST* promoter in a PAC or a YAC transgene may confer a more accurate reflection of the endogenous regulation, as described in Chapter 4, these constructs, even with considerable flanking DNA in the construct, were still susceptible to position effects and were often silenced over time. In addition, the number of transgene integrations is not easily controlled which leads to variability in transcript levels depending on the copy number. Selection is also a potential problem if expression of a localized, ectopic *XIST* transcript provides a selective disadvantage, which may be the case even in somatic cells [61]. Therefore, an inducible promoter capable of high-level expression would be an ideal system to study *XIST* function.

In my initial attempts to identify important domains within the *XIST* sequence, I created 2 different constructs containing fragments of *XIST* with expression driven by a constitutive CMV promoter. Because of the multiple variables described above and in Chapter 4, these constructs were subject to position effects and low expression levels which precluded reliable conclusions regarding their ability to localize to surrounding chromatin. Therefore, to more effectively examine the role of specific sequence elements within the *XIST* transcript on localization, I created an *XIST* cDNA construct under the control of a tetracycline inducible promoter. The inducible expression vector, pcDNA5/FRT/TO (Invitrogen), contains a human cytomegalovirus (CMV) promoter into which two tandem copies of the Tetracycline (Tet) operator sequence have been inserted. These *E. coli*-derived regulatory elements allow tetracycline-regulated expression in cell lines expressing the Tetracycline repressor protein [293]. Although the levels of

induction vary depending on the cell line, studies have reported 50-100 fold increases in expression levels upon tetracycline induction [294-296]. Therefore, the inducible construct will not only be able produce high-levels of expression to overcome potential position effects, but will also allow me to vary and examine the effects of different transcript levels as well as minimize the effects of selection. The expression vector also contains a Flp-recombinase target site (FRT) derived from *Saccharomyces cerevisiae.* This 34 bp sequence serves as the binding site for for Flp recombinase which mediates a recombination event between 2 FRT sites and will therefore allow site-specific integration into a cell line containing a single FRT site integrated into its genome [297,298]. Each transgenic clone will therefore have a single copy of the *XIST* transgene in a specified site in the genome, thereby controlling both copy number and integration site.

In addition to a full-length *XIST* cDNA construct, I have created 3 deletion constructs. In the first (delXB), I deleted a 796 bp region which removes the conserved A repeats at the 5' end of exon 1 required for silencing in mouse that also contribute to transcript localization [53]; the second (delPflMI) removes a 3.8 kb fragment in the central region of exon 1 that contains the C repeats homologous to the a mouse region implicated in *Xist* localization by peptide nucleic acid (PNA) interference mapping [52]; the third construct removes the 3' end of *XIST* and contains only exon 1 (del 3' *XIST*). These constructs along with the full length *XIST* cDNA have been introduced into a female somatic cell line and assayed for transcript localization by RNA-FISH. The initial results will be presented in this chapter.

RESULTS

Expression of short XIST fragments in somatic cells

Two different fragments of *XIST* were inserted into the pAX72 vector which contains a constitutive CMV promoter to drive expression (Figure 6-1). The first construct contains a 1196 bp insert from the 5' end of *XIST* exon 1 which includes the conserved A repeats important for silencing and localization in mouse. The second construct contains a 2235 bp insert that extends from the end of exon 1 to exon 6. These constructs were co-transfected with a puromycin resistance vector into female somatic BOSC 23 cells (human embryonic kidney cell line) as well as the mouse somatic cell line tsA1S9. After transfection, cells were grown in puromycin and single cell clones were isolated.

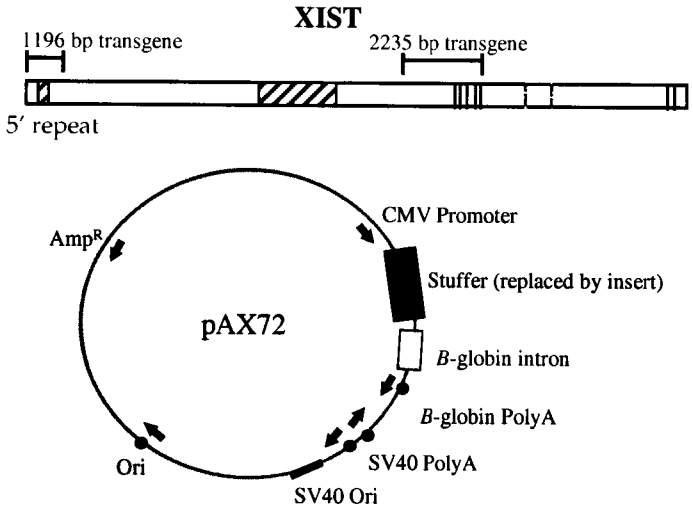


Figure 6-1. *XIST* 5' and 3' fragment constructs in the pAX72 expression vector. An 1196 bp fragment from the 5' end of *XIST* and a 2.2 kb region from the end of exon 1 to exon 6 were inserted into the pAX72 vector which contains a constitutively active CMV promoter to drive expression. These two constructs were co-transfected with a puromycin resistance vector into BOSC 23 female somatic cells and the mouse somatic cell line tsA1s9.

For the 5' *XIST* construct, only ~10-20% of the single cell clones showed expression of the transgene by RT-PCR (Table 6-1). The level of expression in these positive clones varied widely, presumably due to position effects, and although expression was detectable by RT-PCR in the early passages after picking the clones, many of them were silenced over time. Relative levels of endogenous *XIST* compared to the transgene expression levels were obtained by performing a RT-PCR on the transformant clones using 3 primers in one reaction tube: one common to both endogenous *XIST* and the transgene and the other two specific to either the transgene or endogenous *XIST*. This primer combination gave distinct endogenous and transgene bands whose intensity was compared (Figure 6-2a). One clone, HB16, which expressed the highest levels of the transgene still only appeared as a small dot by RNA-FISH (Figure 6-2b). Actinomycin experiments suggest that the ectopic transcripts were not unstable with a half-life similar to endogenous *XIST* (Figure 6-2c); however, with the background of large endogenous *XIST* signals and the relatively low expression levels of the transgene, it was difficult to detect the presence of delocalized transcripts by RNA-FISH. A similar variable expression pattern was detected when the construct was introduced into the mouse somatic cell line tsA1S9; however, none of these transformant clones had high enough expression levels that could be observed by RNA-FISH (data not shown). The 3' *XIST* construct also had low transfection efficiencies and expression levels from the transgene and none of the clones gave signals by RNA-FISH. RT-PCR on nuclear and cytoplasmic RNA fractions of each of the expressing clones show that the ectopic *XIST* transcripts seem to remain in the nucleus (Figure 6-2d); however, without a

detectable signal by RNA-FISH, it is unclear whether the transcripts remain associated with the surrounding chromatin.

The pAX72 vector also contains a rabbit beta-globin intron meant to enhance the stability of the expressed transcripts. Because *XIST* is a nuclear transcript and the presence of splicing signals derived from a protein-coding transcript may alter its normal localization pattern, I deleted this region in the constructs described so far in order to be able to determine the importance of specific *XIST* sequences. However, to analyze the effect an ectopic splicing signal has on *XIST* transcript localization, I made one construct with the 5' *XIST* repeats that retained the rabbit beta-globin intron. When transfected into the BOSC 23 cells, as with the other constructs, the transfection efficiency was fairly low and the transgene was often silenced over time. However, several clones did retain expression of the transgene which was detectable by PCR (Figure 6-2a). Interestingly, unlike the other constructs without the intron, these ectopic transcripts appear to be present in the cytoplasm. RT-PCR on nuclear and cytoplasmic RNA preparations reveal that the intron-containing 5' *XIST* transcripts, although predominantly nuclear, also show some amplification in the cytoplasmic fraction in a pattern similar to PGK, a protein-coding gene whose RNA transcripts normally get exported out of the nucleus (Figure 6-2d). Therefore, the presence of ectopic splicing signals appears to be sufficient to alter *XIST*'s retention in the nucleus. Unfortunately, RNA-FISH could not detect the ectopic transcripts, probably due to the relatively low levels of expression, but also potentially because the transcripts, now dispersed in the cytoplasm, are much more difficult to detect.

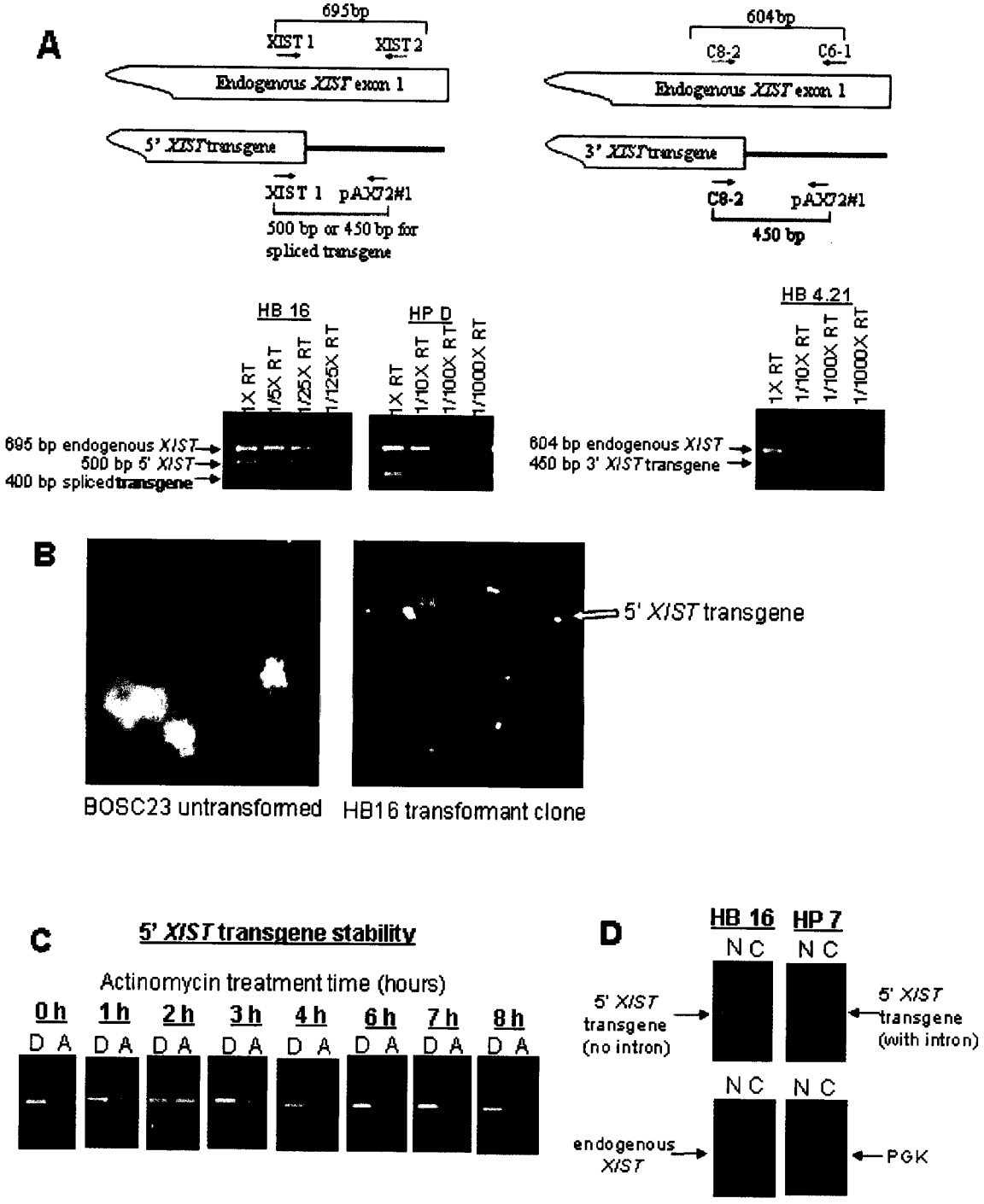


Figure 6-2. *XIST* 5' and 3' fragments are expressed at low levels in the transformant clones and are silenced over time. **(A)** Relative quantitation of *XIST* transgene expression. Three different primers were used in one PCR reaction, one common primer and the other two specific to either the transgene or the

endogenous *XIST* sequence, which resulted in a transgene-specific and an endogenous *XIST* band whose intensities could be compared to obtain relative transcript levels. A dilution series of the RT reactions were used as template to ensure that the PCR reactions were not saturated. Transformant clone HB 16 is expressing the 5' *XIST* transgene fragment without the rabbit beta-globin intron in the vector while the construct in clone HP D contains the intron which is spliced out as expected and results in a 450 bp spliced PCR product. **(B)** RNA-FISH with an *XIST* probe (G1A, green) and X paint (red) on BOSC 23 female somatic cells transformed with 5' transgene (clone HB16). The cell line contains 4 X chromosomes, two of which are inactivated and coated with *XIST*. (C) Stability of the 5' *XIST* transgene fragment expressed in the BOSC 23 transformant clone HB 8. The cells were treated with Actinomycin to arrest transcription and RNA was then harvested at various timepoints after treatment for RT-PCR analysis. A= actinomycin treated; D= DMSO control **(D)** RT-PCR on nuclear and cytoplasmic RNA isolated from the transformant clones containing the 5' *XIST* construct with and without the rabbin beta-globin intron.

Table 6-1. Summary of clones transformed with small *XIST* fragments in the pAX72 expression vector

Cell type	*XIST* fragment introduced	Rabbit beta-globin intron	Number of clones picked	Clones with stable transgene integration	Clones with transgene expression	Clones Analysed
BOSC 23	5' XIST	—	25	6	3	HB C ** HB 8 ** HB 16
t86 (X_i hybrid)	5' XIST	—	28	7	2	HB N HB O
tsA1S9 (mouse)	5' XIST	—	37	16	11	tsHB 3, tsHB 7, tsHB 8, tsHB 9, tsHB 10, tsHB 11, tsHB D, tsHB E, tsHB K, tsHB L, tsHB M
BOSC 23	5' XIST	+	96	24	10	HP 50; HP 54; HP 36 ; HP D; HP 7; HP 38; HP A (3 silenced over time)
BOSC 23	3' XIST	—	117	30	12	HB 3.35; HB 3.37; HB 3.38; HB 3.40; HB 3.41; HB 3.42; HB 3.44; HB 3.45; HB 4.15; HB 4.21 (2 silenced over time)

** **expression silenced over time**

Expression of a full length XIST cDNA construct in somatic cells

To make the full length *XIST* cDNA, I PCR amplified exon 1 using the Expand long PCR template system (Roche) using female genomic DNA as a template. The PCR product was then cut with XbaI to obtain a sequence that extends from nucleotides 105-10748 of the *XIST* cDNA sequence (Accession number M97168). The sticky ends were filled with Klenow and the resulting blunted fragment was inserted into the EcoRV site of the pcDNA5/FRT/TO expression vector (Invitrogen, Figure 6-3a). To obtain the 3' end of *XIST*, *XIST* cDNA clones previously isolated from human heart and fetal brain libraries [256] were used and inserted downstream of the exon 1 fragment, into the NotI and XhoI sites of the pcDNA5/FRT/TO multiple cloning site. The *XIST* spliced variant used contains the 5' end of exon 6 up to the splice donor site at nucleotide 13 809 as well as exons 7 and 8 (Figure 6-3b).

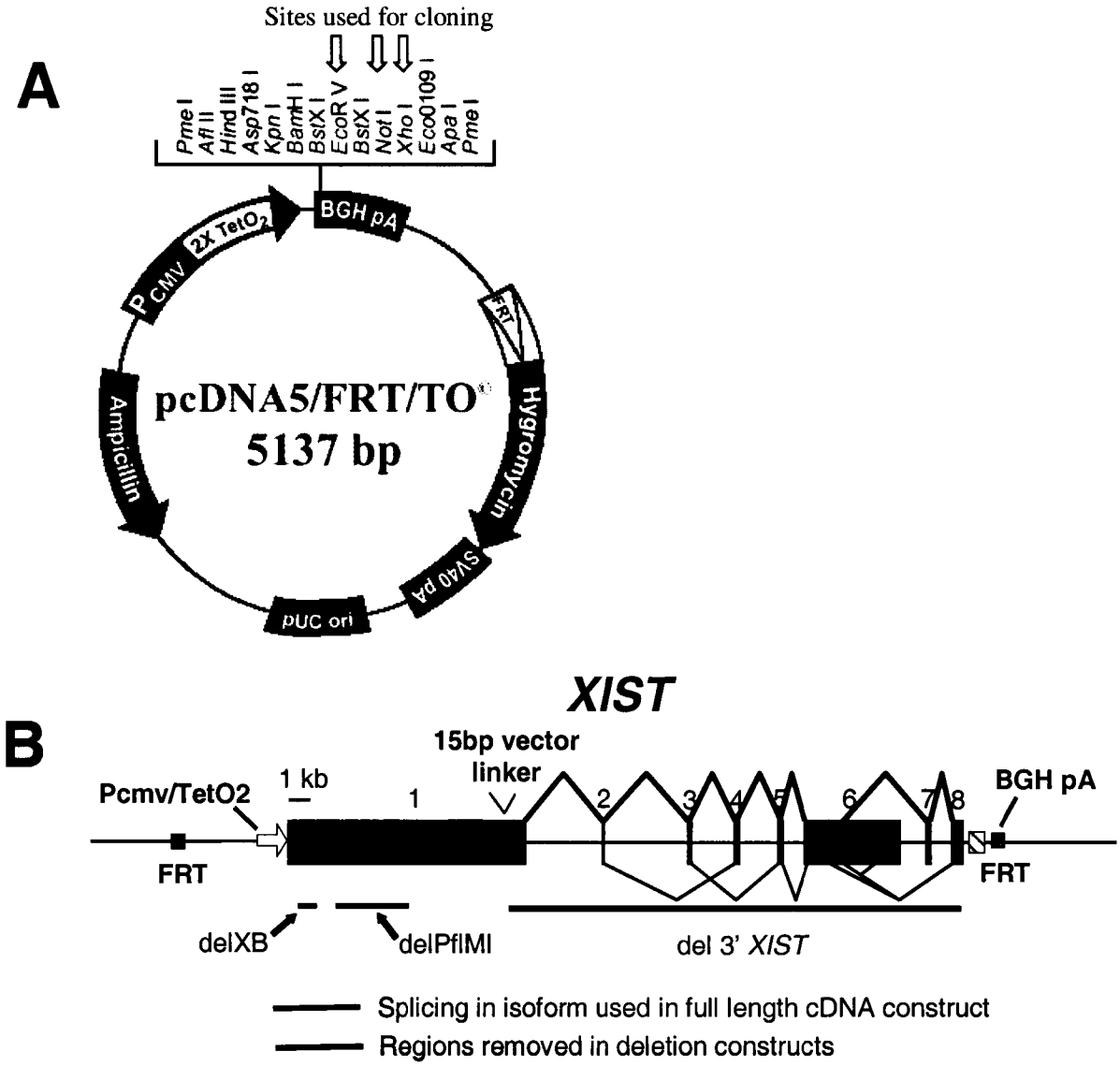


Figure 6-3. Inducible *XIST* cDNA construct. (A) pcDNA5/FRT/TO tetracycline inducible vector (adapted from Invitrogen) (B) The *XIST* transcript is alternatively spliced and the isoform used in the full length cDNA construct is indicated with the red lines. The regions removed in deletion constructs are shown in green. The numbers above the blue boxes represent the *XIST* exons. A 15 bp region of the vector between the EcoRV and the NotI cloning sites is retained in the construct (vector linker region).

The *XIST* cDNA construct was introduced into the Flp-In T-Rex 293 host cell line (Invitrogen) which is a female somatic cell line that contains a single FRT site integrated into its genome and expresses the tetracycline repressor protein important for tetracycline inducible expression. The *XIST* construct was co-transfected with a Flp-recombinase expression plasmid (pOG44, Invitrogen) which mediates a homologous recombination event between the FRT site integrated into the genome and the site within the expression vector. Proper integration events were then selected for using hygromycin and single cell clones isolated. Expression of the transgene after induction was first analyzed by quantitative real time PCR using primers that amplify a transgene-specific product. Doxycycline treatment led to a 10-20-fold induction and resulted in transgene expression levels about 2-3X endogenous *XIST* levels (Figure 6-4a). When the transformant clones were analyzed by RNA-FISH using an *XIST* probe and a vector-specific probe (presumably detecting portions of the vector that are being transcribed along with the *XIST* insert), the ectopic *XIST* transcripts were able to form a large localized signal (Figure 6-4b).

Expression of XIST deletion constructs in female somatic cells

Three different deletion constructs were made and transfected into the Flp-In T-Rex 293 cells (Figure 6-3b). The first deletion (delXB) removed the conserved 5' A repeats in exon one and was created by cutting the full-length construct with XcmI (nucleotide 421) and BlpI (nucleotide 1217), filling in the sticky ends with Klenow and religating the blunt ends. The second deletion removed a 3.8 kb central portion of exon 1 (2181-5996) using PflMI restriction sites (delPflMI). The final deletion construct contains only exon 1 and lacks the 3.6 kb portion spanning exons 2-8 (del 3' *XIST*).

When analyzed by RNA-FISH, both the delXB and the delPflMI deletion transcripts were unable to form a localized signal and appeared only as a small spot of expression. In contrast, transcripts from the *XIST* exon 1 construct (del 3'*XIST*) appeared to be present at higher levels than the other two deletions and were able to form a localized signal (Figure 6-4e). Therefore, sequences within exon 1 appear to be sufficient for proper transcript localization and do not require the 3' end of the *XIST* transcript. However, deleting the conserved 5' end of exon 1 (delXB) and the central portion of exon 1 (delPflMI) appear to disrupt the transcripts' ability to accumulate and form a localized signal.

To determine whether the small RNA-FISH signals in the delXB and delPflMI deletions were due to lowered transcript levels, I determined the transgene expression levels using real-time quantitative PCR (Figure 6-4a). Upon doxycycline induction, the delPflMI clones expressed the mutant transgene at levels similar to endogenous *XIST*. In contrast, the delXB clones had much lower levels of transgene expression, with very little increase in transcript levels with induction. Therefore, deletion of the conserved 5' repeats seems to affect the steady state levels of the ectopic transcript while the PflMI deletion, although it does not affect transcript levels, precludes transcript accumulation and localization to the surrounding chromatin.

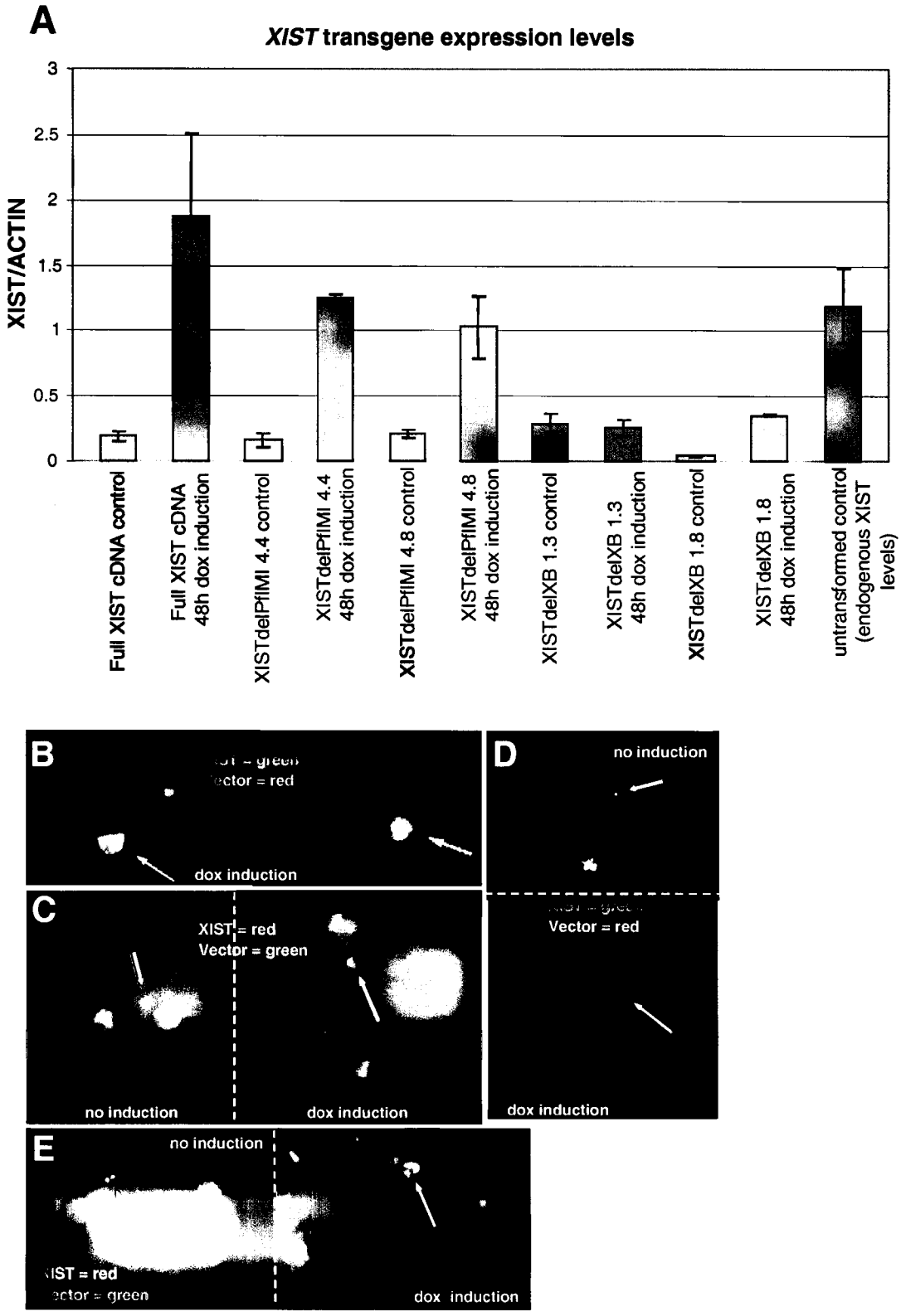

A
XIST transgene expression levels
XIST/ACTIN
3
2.5
2
1.5
1
0.5
0
Full XIST cDNA control
Full XIST cDNA 48h dox induction
XISTdelPflMI 4.4 control
XISTdelPflMI 4.4 48h dox induction
XISTdelPflMI 4.8 control
XISTdelPflMI 4.8 48h dox induction
XISTdelXB 1.3 control
XISTdelXB 1.3 48h dox induction
XISTdelXB 1.8 control
XISTdelXB 1.8 48h dox induction
untransformed control (endogenous XIST levels)
B
ector = red
dox induction
C
XIST = red
Vector = green
no induction
dox induction
D
no induction
Vector = red
dox induction
E
no induction
IST = red
ector = green
dox induction

Figure 6-4. Expression of an ectopic *XIST* transgene in doxycycline-induced transformant clones. **(A)** Quantitation of ectopic *XIST* transgene levels by real-time quantitative PCR. Results from two different clones for each deletion construct are shown. **(B)** RNA-FISH using an *XIST* probe (G1A, green) and a vector probe (red) shows that the ectopic full length *XIST* transcript forms a large, localized *XIST* signal similar to endogenous *XIST*(clone L3.7.10.6, doxycycline induced). **(C-E)** RNA-FISH on transformant clones containing the XISTdelXB **(C)**, XISTdelPflMI **(D)**, and the 3' *XIST* **(E)** deletion constructs

DISCUSSION

The association of the *XIST* transcript to surrounding chromatin is facilitated by a variety of factors including the sequence as well as the chromatin conformation of the surrounding regions. In addition, sequence elements within the *XIST* RNA transcript as well as levels of the RNA also appear to be important in the formation of the particular structure required for transcript accumulation [22]. Because of the multiple variables that affect localization, my initial attempts at defining the sequence elements required for localization using short *XIST* fragments in the pAX72 expression vector were largely unsuccessful. The relatively small transgenes were highly susceptible to position effects and were often silenced over time. In addition, because the transgenes are randomly integrated into the genome in different copy numbers, it is not possible to conclude whether the absence of a localized transcript is due to missing sequence elements within the *XIST* transgene, the surrounding DNA, or inadequate levels of expression.

To address these problems, I created an *XIST* cDNA construct under the control of a tetracycline/doxycycline inducible promoter. The expression vector also contains an FRT site which allows site-specific integration of a single copy of the transgene into a specified site in the genome of the recipient somatic cell line. Therefore, with this

expression system, transgene expression levels, integration site and copy number can be controlled. In my initial experiments, I have introduced a full length *XIST* cDNA construct, as well as three deletion constructs into a female somatic 293 cell line which contains a single FRT site integrated into its genome (see summary Table 6-2). The full-length *XIST* cDNA construct is expressed at similar levels to endogenous *XIST* and forms a large, localized signal. Therefore, a single integrated copy of the *XIST* transgene is sufficient to form a localized signal in human somatic cells. A similar result was observed in mouse embryonic stem cells where a single integrated copy of an inducible *Xist* cDNA construct is sufficient to form a localized signal capable of silencing surrounding chromatin [53]. This contrasts with the requirement for multiple integrated copies of a YAC transgene for *Xist* localization and silencing [284,285] and suggests that the high levels of expression from the inducible constructs may be an important factor for transcript localization and silencing.

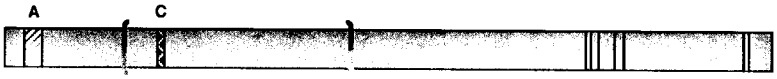
Deletion of the 3' end of *XIST* does not seem to affect transcript accumulation since the *XIST* exon 1 construct is also expressed at high levels and forms a localized signal. In contrast, the removal of the 5' conserved repeats (delXB) leads to lower steady-state levels of the ectopic transcript with almost no or very little increase in transcript levels with doxycycline induction. In addition, the transcripts are unable to form large, localized signal and appear as a small spot of expression by RNA-FISH. This is different from the results in mouse where a similar construct removing the conserved A repeats is unable to initiate silencing, but is still able to form a localized signal by RNA-FISH. Although further deletions in mouse found the A repeats to work co-operatively with other spatially separated domains to contribute to transcript localization, they appear

to play a more integral role in humans, since their removal abolishes transcript accumulation. Currently, Actinomycin experiments are being planned to determine if the removal of the A repeats affects the stability of the ectopic transcript.

Removal of the central portion of *XIST* exon 1 also leads to an ectopic transcript that can no longer form a large, accumulated signal by RNA-FISH. However, unlike the deletions that removed the 5' conserved repeats, the delPflMI deletion transcripts had a higher steady-state level, similar to endogenous *XIST*. Therefore, deletion of this region does not appear to affect stability (since transcription rate should be the same), but may affect transcript localization. This result agrees with PNA-interference studies which implicated the homologous region in mouse for *XIST* transcript localization [52].

For future studies, additional deletion constructs to further define important sequence elements within the *XIST* transcript will be examined. In addition, the effects of varying transcript levels on localization will be addressed by altering doxycycline induction levels. Several recipient somatic cell lines are being constructed with different FRT integration sites which will allow comparisons among different integration sites to determine the features of the surrounding DNA that facilitate *XIST* transcript spread and accumulation. There is also evidence that other features of inactivation can be recruited by ectopic *XIST* expression in human somatic cells ([61], see Chapter 4), so it will also be interesting to determine whether a single integrated transgene is able to alter surrounding chromatin and initiate features of inactivation.

Table 6-2. Summary of expression from inducible *XIST* cDNA deletion constructs

XIST cDNA construct	Transgene RNA-FISH signal	Transgene expression levels (relative to female somatic cells)
 Full length XIST cDNA	Localized signal	1.58
 delXB	Pinpoint low level expression	0.26
 delPflMI	Pinpoint low level expression	0.96
 del3' XIST	Localized signal	Not determined

*A and C denote conserved repeat elements described in Figure 1-2. The red brackets indicate the regions deleted in each construct.

CHAPTER 7: DISCUSSION

Complexity of expression patterns at the XIST locus

Analysis of expression in somatic, embryonal carcinoma and *XIST* transgene cell lines has revealed a complex expression pattern in the human *XIST* region. Unique sense and antisense transcripts identified in N-Tera2D1, the embryonal carcinoma cell line, have not been detected in somatic cells and thus appear to be developmentally-specific. Like the mouse antisense, *Tsix*, the antisense transcript in N-Tera2D1 initiates downstream of *XIST* and has some overlap with the *XIST* gene. However, unlike mouse *Tsix*, the human antisense does not extend to the 5' region of *XIST*, and does not appear to be associated with an unstable *XIST* transcript. In addition the ratio of antisense to sense transcripts in N-Tera2D1 is much lower than mouse where *Tsix* is expressed at 10-100X higher levels than *Xist* [80]. Given these differences and the nonphysiological nature of the embryonal carcinoma cell line, it is unclear whether the antisense transcription observed in humans is biologically relevant.

Sequence comparisons between mouse, human and cow reveal very little conservation outside the *XIST* gene. In addition, the regions surrounding the *XIST* locus are highly repetitive and have undergone several rearrangements through evolution. Therefore, the human antisense transcript may have lost some of the functions that are observed in mouse; however, the retention of expression in a developmentally restricted manner is suggestive of a possible function. In fact, expression within the mouse Xic also shows a complex pattern of expression. The *Tsix* antisense transcript itself is subject to alternative splicing and has multiple initiation sites. The role of this heterogeneity is as yet unclear [79,80]. In addition, extensive, low level intergenic transcripts have also been detected between *Tsix* and the upstream *Tsx* gene [299], and although the transcripts

themselves do not seem to be important, targeted deletions of the intergenic transcriptional start sites results in the downregulation of *Tsix* in *cis* and skewed X inactivation ratios. Therefore, low-level, heterogeneous transcription appears to be common to the *XIST* region in both mouse and humans and further supports a possible role for the antisense transcription observed in humans.

Factors affecting XIST localization

The association of the *XIST* RNA to the X chromosome is an important event in the inactivation pathway, required for the recruitment of complexes involved in chromatin modifications like the Polycomb group proteins Eed and Ezh2 as well as the histone variants macroH2A1 and macroH2A2 [56,64,105]. The nature of this association; however, is unclear and appears to involve many interacting factors, including specific sequences and likely structural elements within the RNA, *cis*-acting DNA elements as well as *trans*-acting protein factors. In addition, increased transcript levels appear to contribute to the formation of a stable, localized *XIST* signal [51,56]. Ectopic *XIST* transgenes have provided some clues about the role and contribution of these variables and suggest that increased transgene integrations lead to a more efficient formation of a localized *Xist* domain that is capable of initiating inactivation of surrounding chromatin [284,285]. This increased competency could possibly be due to a more permissive chromatin structure provided by the multicopy array, an increase in the number of *cis*-DNA elements required for *XIST* association, or an elevation in the transcript levels. However, based on the results presented in Chapter 4, increased copy number and transgene expression levels alone will not lead to a more localized signal.

To further examine the factors that affect *XIST* localization, I have created an inducible cDNA construct that will integrate single-copy into a specific site in the host cell line. My initial experiments focus on the identification of important domains within the *XIST* transcript that are required for its association with surrounding chromatin. A single copy of the full length *XIST* cDNA construct integrated into a female somatic cell line is able to form a large, localized signal by RNA-FISH upon induction. This is similar to the mouse inducible *XIST* cDNA which also forms a large *XIST* domain when present as a single integrated transgene [51]. Also similar to mouse, it appears that sequences within exon 1 are sufficient for localization since transcripts without sequences at the 3' end of the gene are still able to form a large domain. The conserved A repeats at the 5' end of exon 1 appear to be essential for *XIST* upregulation and localization since constructs expressing deletions of this region appear only as small spots by RNA-FISH. Removal of the region likely affects transcript stability since quantitation by real-time PCR revealed that levels of the RNA did not increase with doxycycline induction. These repeats may play a more prominent role in human *XIST* stability and localization than in mouse since in the latter; although removal of the conserved repeats disrupt silencing, transcripts still form a localized signal [51]. The central region of exon 1 also appears to play a role in *XIST* accumulation since transcripts lacking a 3.8 kb region containing the C repeats can only form a small signal by RNA-FISH. However, transcript levels do increase upon doxycycline induction, so unlike the A repeat deletion, these transcripts remain stable, but are unable to accumulate to form a localized signal.

Ectopic *XIST* transcripts have been shown to recruit features of inactivation in human somatic cells [61]; therefore, for future studies it will be interesting to determine

whether the transformant clones described in this study are also able to initiate features of inactivation and induce changes in chromatin structure. In addition, with an inducible promoter it is possible to vary transcript levels by modulating the amount of doxycycline which will allow us to determine the importance of transcript levels on localization. Finally, to determine the nature of the *cis*-acting DNA elements that facilitate transcript localization, we are in the process of creating several different recipient host cell lines with different FRT integration sites whose ability to localize *XIST* can then be compared to identify features in the DNA that are more able to support *XIST* accumulation.

Although initiation of inactivation in human cells does not appear to require a strict developmental window, it is likely that inactivation initiated outside of a normal developmental context does not completely recapitulate the series of events that occur during normal embryogenesis and probably results in a less stable inactive state. Therefore, it will be important to determine the effects of ectopic *XIST* expression in a system that more closely resembles early human development. Embryonic stem cells, which can be induced to undergo differentiation and recapitulate the events of X-chromosome inactivation, have been provided an invaluable model system for mouse early development. Although embryonic stem cells have been isolated from humans, evidence suggests that the available lines are not at the correct developmental stage and have already undergone inactivation. A more feasible way to mimic the conditions in early development may be to introduce the inducible human *XIST* transgene into mouse embryonic stem cells. Although human and mouse *XIST* sequences show considerable sequence divergence, human *XIST* YAC transgenes are recognized as an X-inactivation centre in mouse embryonic stem cells and can initiate inactivation [280]. In addition to

the problems that may arise due to species differences, the *XIST* cDNA construct also does not include sequences outside of *XIST*, including potential regulatory elements like *Tsix* antisense transcript downstream of *Xist* in mouse. Therefore, elements of count and choice may not be accurately represented with this system; however, given the difficulties in finding a suitable model for early human development, experiments using the *XIST* cDNA in mouse ES cells will likely still be informative.

CHAPTER 8: REFERENCES

1 Graves, J. (2002) The rise and fall of *SRY*. Trends Genet. **18:** 259-264

2 Foster, J. W. and Graves, J. A. M. (1994) An *SRY*-related sequence on the marsupial X chromosome: Implications for the evolution of the mammalian testis-determining gene. Proc. Natl. Acad. Sci. USA **91:** 1927-1931

3 Charlesworth, B. (1991) The evolution of sex chromosomes. Science **251:** 1030-1033

4 Consortium, I. H. G. S. (2001) Initial sequencing and analysis of the human genome. Nature **409:** 860-921

5 Graves, J., Disteche, C. and Toder, R. (1998) Gene dosage in the evolution and function of mammalian sex chromosomes. Cytogenet. Cell Genet. **80:** 94-103

6 Lahn, B. and Page, D. (1999) Four evolutionary strata on the human X chromosome. Science **286:** 964-967

7 Ohno, S. *Sex Chromosomes and Sex Linked Genes*, (Springer Verlag, Berlin, 1967).

8 Lyon, M. F. (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature **190:** 372-373

9 Jegalian, K. and Page, D. (1998) A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. Nature **394:** 776-780

10 Cooper, D. W., Johnston, P. G., Watson, J. M. and Graves, J. A. M. (1993) X-inactivation in marsupials and monotremes. Sem. Dev. Biol. **4:** 117-128

11 Takagi, N. and Sasaki, M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature **256:** 640-642

12 Xue, F., Tian, X. C., Du, F., Kubota, C., Taneja, M., Dinnyes, A., Dai, Y., Levine, H., Pereira, L. V. and Yang, X. (2002) Aberrant patterns of X chromosome inactivation in bovine clones. Nat. Genet. **31:** 216-220

13 Brown, C. J. and Robinson, W. P. (2000) The causes and consequences of random and non-random X chromosome inactivation in humans. Clin. Genet. **58:** 353-363

14 Penaherrera, M. S., Ma, S., Ho Yuen, B., Brown, C. J. and Robinson, W. P. (2003) X-chromosome inactivation (XCI) patterns in placental tissues of a paternally derived bal t(X;20) case. Am. J. Med. Genet. **118A:** 29-34

15 Migeon, B. R., Jan de Beur, S. and Axelman, J. (1989) Frequent derepression of G6PD and HPRT on the marsupial inactive X chromosome associated with cell proliferation in vitro. Exp. Cell Res. **182:** 597-609

16 Kaslow, D. C. and Migeon, B. R. (1987) DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. Proc. Natl. Acad. Sci. USA **84:** 6210-6214

17 Lifschytz, E. and Lindsley, D. L. (1972) The role of X-chromosome inactivation during spermatogenesis. Proc. Natl. Acad. Sci. USA **69:** 182-186

18 Handel, M. A., Park, C. and Kot, M. (1994) Genetic control of sex-chromosome inactivation during male meiosis. Cytogenet. Cell Genet. **66:** 83-88

19 Jablonka, E. and Lamb, M. J. (1988) Meiotic pairing constraints and the activity of the sex chromosomes. J. Theor. Biol. **133:** 23-36

20 Turner, J. M. A., Mahadevaiah, S. K., Benavente, R., Offenberg, H. H., Heyting, C. and Burgoyne, P. S. (2000) Analysis of male meiotic "sex body" proteins during XY female meiosis provides new insights into their function. Chromosoma **109:** 426-432

21 Turner, J. M. A., Burgoyne, P. S. and Singh, P. B. (2001) M31 and macroH2A1.2 colocalise at the pseudoautosomal region during mouse meiosis. J. Cell Sci. **114:** 3367-3375

22 Plath, K., Mlynarczyk-Evans, S., Nusinow, D. and Panning, B. (2002) XIST RNA and the mechanism of X chromosome inactivation. Ann. Rev. Genet. **36:** 233-278

23 Salido, E. C., Yen, P. H., Mohandas, T. K. and Shapiro, L. J. (1992) Expression of the X-inactivation-associated gene *XIST* during spermatogenesis. Nat. Genet. **2:** 196-199

24 McCarrey, J. R. and Dilworth, D. D. (1992) Expression of *Xist* in mouse germ cells correlates with X-chromosome inactivation. Nat. Genet. **2:** 200-203

25 Richler, C., Soreq, H. and Wahrman, J. (1992) X inactivation in mammalian testis is correlated with inactive X-specific transcription. Nat. Genet. **2:** 192-195

26 Marahrens, Y., Panning, B., Dausman, J., Strauss, W. and Jaenisch, R. (1997) *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. **11:** 156-166

27 McCarrey, J. R., Watson, C., Atencio, J., Ostermeier, G. C., Marahrens, Y., Jaenisch, J. and Krawetz, S. A. (2002) X-chromosome inactivation during

spermatogenesis is regulated by an *Xist/Tsix*-independent mechanism in the mouse. Genesis **34:** 257-266

28 Turner, J. M. A., Mahadevaiah, S. K., Elliott, D. J., Garchon, H., Pehrson, J. R., Jaenisch, R. and Burgoyne, P. S. (2002) Meiotic sex chromosome inactivation in male mice with targeted disruptions of *Xist*. J. Cell Sci. **115:** 4097-4105

29 Hoyer-Fender, S., Costanzi, C. and Pehrson, J. (2000) Histone macroH2A1.2 is conentrated in the XY-body by the early pachytene stage of spermatogenesis. Exp. Cell Res. **258:** 254-260

30 Richler, C., Dhara, S. K. and Wahrman, J. (2000) Histone macroH2A1.2 is concentrated in the XY compartment of mammalian male meiotic nuclei. Cytogenet. Cell Genet. **89:** 118-120

31 Ganesan, S., Silver, D. P., Greenberg, R. A., Avni, D., Drapkin, R., Miron, A., Mok, S. C., Randrianarison, V., Brodie, S., Salstrom, J., Rasmussen, T. P., Klimke, A., Marrese, C., Marahrens, Y., Deng, C., Feunteun, J. and Livingston, D. M. (2002) BRCA1 supports XIST RNA concentration on the inactive X chromosome. Cell **111:** 393-405

32 Armstrong, S. J., Hulten, M. A., Keohane, A. M. and Turner, B. M. (1997) Different strategies of X-inactivation in germinal and somatic cells: histone H4 underacetylation does not mark the inactive X chromosome in mouse male germline. Exp. Cell Res. **230:** 399-402

33 Venolia, L. and Gartler, S. M. (1983) Transformation of Hprt gene with sperm DNA. Somat. Cell Mol. Genet. **9:** 617-627

34 McCarrey, J. R. X-chromosome inactivation during spermatogenesis: the original dosage compensation mechanism in mammals? in *Gene families: studies of DNA, RNA, enzymes and proteins* (eds. Xue, G. *et al.*) 59-72 (World Scientific, River Edge, NJ, 2001).

35 Meller, V. (2000) Dosage compensation: making 1X equal 2X. Trends Cell Biol. **10:** 54-59

36 Copps, K., Richman, R., Lyman, L., Chang, K., Rampersad-Ammons, J. and Kuroda, M. (1998) Complex formation by the Drosophila MSL proteins: Role of the MSL2 RING finger in protein complex assembly. EMBO **17:** 5409-5417

37 Scott, M., Pan, L., Cleland, S., Knox, A. and Heinrich, J. (2000) MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in *Drosophila*. EMBO **19:** 144-155

38 Marin, I., Siegal, M. and Baker, B. (2000) The evolution of dosage-compensation mechanisms. Bioessays **22:** 1106-1114

39 Kelley, R., Meller, V., Gordadze, P., Roman, G., Davis, R. and Kuroda, M. (1999) Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. Cell **98:** 513-522

40 Akhtar, A., Zink, D. and Becker, P. (2000) Chromodomains are protein-RNA interaction modules. Nature **407:** 405-409

41 Meyer, B. (2000) Sex in the worm: counting and compensating X-chromosome dose. Trends Genet. **16:** 247-253

42 Lieb, J., Capowski, E., Meneely, P. and Meyer, B. (1996) DPY-26, a link between dosage compensation and meiotic chromosome segregation in the nematode. Science **274:** 1732-173

43 Lieb, J. D., Albrecht, M. R., Chuang, P. and Meyer, B. J. (1998) MIX1: an essential component of the C. elegans mitotic machinery executes X chromosome dosage compensation. Cell **92:** 265-277

44 Dawes, H., Berlin, D., Lapidus, D., Nusbaum, C., Davis, T. and Meyer, B. (1999) Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate. Science **284:** 1800-1804

45 Keohane, A. M., Lavender, J. S., O'Neill, L. P. and Turner, B. M. (1998) Histone acetylation and X inactivation. Dev. Genet. **22:** 65-73

46 Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J. and Willard, H. F. (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:** 527-542

47 Clemson, C. M., McNeil, J. A., Willard, H. F. and Lawrence, J. B. (1996) *XIST* RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell. Biol. **132:** 259-275

48 Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. and Rastan, S. (1992) The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell **71:** 527-542

49 Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M. E., Rogozin, I. B., Kolesnikov, N. N., Brockdorff, N. and Zakian, S. M. (2001) Characterization of the genomic *Xist* locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res. **11:** 833-849

50 Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P. and Duret, L. (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human and bovine. Genome Res. **12:** 894-908

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

52 Beletskii, A., Hong, Y. K., Pehrson, J., Egholm, M. and Strauss, W. M. (2001) PNA interference mapping demonstrates functional domains in the noncoding RNA *Xist*. Proc. Natl. Acad. Sci. USA **98:** 9215-9220

53 Wutz, A., Rasmussen, T. P. and Jaenisch, R. (2002) Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. Nat. Genet. **30:** 167-174

54 Sheardown, S. A., Duthie, S. M., Johnston, C. M., Newall, A. E. T., Formstone, E. J., Arkell, R. M., Nesterova, T. B., Alghisi, G., Rastan, S. and Brockdorff, N. (1997) Stabilization of *Xist* RNA mediates initiation of X chromosome inactivation. Cell **91:** 99-107

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

56 Brockdorff, N. (2002) X-chromosome inactivation: closing in on proteins that bind *Xist* RNA. Trends Genet. **18:** 352-358

57 Daniels, R., Zuccotti, M., Kinis, T., Serhal, P. and Monk, M. (1997) *XIST* expression in human oocytes and preimplantation embryos. Am. J. Hum. Genet. **61:** 33-39

58 Ray, P. F., Winston, R. M. L. and Handyside, A. H. (1997) *XIST* expression from the maternal X chromosome in human male preimplantation embryos at the blastocyst stage. Hum. Mol. Genet. **6:** 1323-1327

59 de la Fuente, R., Hahnel, A., Basrur, P. K. and King, W. A. (1999) X inactive-specific transcript (*Xist*) expression and X chromosome inactivation in the preattachment bovine embryo. Biol. Reprod. **60:** 769-775

60 Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. and Brockdorff, N. (1996) Requirement for *Xist* in X chromosome inactivation. Nature **379:** 131-137

61 Hall, L. L., Byron, M., Sakai, K., Carrel, L., Willard, H. F. and Lawrence, J. B. (2002) An ectopic human *XIST* gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. USA **99:** 8677-8682

62 Brown, C. J. and Baldry, S. E. L. (1996) Evidence that heteronuclear proteins interact with the *XIST* RNA *in vitro*. Somatic Cell Mol. Genet. **22:** 403-417

63 Brooks, W. H., Satoh, M., Hong, B., Reeves, W. H. and Yang, T. P. (2002) Autoantibodies from an SLE patient immunostain the Barr body. Cytogenet. Genome Res. **97:** 28-31

64 Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B. and Zhang, Y. (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science **300:** 131-135

65 Csankovszki, G., Nagy, A. and Jaenisch, R. (2001) Synergism of *Xist* RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J. Cell Biol. **153:** 773-783

66 Rack, K. A., Chelly, J., Gibbons, R. J., Rider, S., Benjamin, D., Lafreniere, R. G., Oscier, D., Hendriks, R. W., Craig, I. W., Willard, H. F., Monaco, A. P. and Buckle, V. J. (1994) Absence of the *XIST* gene from late-replicating isodicentric X chromosomes in leukemia. Hum. Mol. Genet. **3:** 1053-1059

67 Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. and Jaenisch, R. (1999) Conditional deletion of *Xist* disrupts histone macroH2A localization but not maintenance of X inactivation. Nat. Genet. **22:** 323-324

68 Brown, C. and HF, W. (1994) The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature **368:** 154-156

69 Hendrich, B., Plenge, R. and Willard, H. (1997) Identification and characterization of the human *XIST* gene promoter: implications for models of X chromosome inactivation. Nucleic Acids Res. **25:** 2661-2671

70 Pillet, N., Bonny, C. and Schorderet, D. (1995) Characterization of the promoter region of the mouse *Xist* gene. Proc. Natl. Acad. Sci. USA **92:** 12515-12519

71 Tomkins, D. J., McDonald, H. L., Farrell, S. A. and Brown, C. J. (2002) Lack of expression of *XIST* from a small ring X chromosome containing the XIST locus in

a girl with short stature, facial dysmorphism and developmental delay. Eur. J. Hum. Genet. **10:** 44-51

72 Plenge, R. M., Hendrich, B. D., Schwartz, C., Arena, J. F., Naumova, A., Sapienza, C., Winter, R. M. and Willard, H. F. (1997) A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nat. Genet. **17:** 353-356

73 Norris, D., Patel, D., Kay, G., Penny, G., Brockdorff, N., Sheardown, S. and Rastan, S. (1994) Evidence that random and imprinted *Xist* expression is controlled by preemptive methylation. Cell **77:** 41-51

74 Panning, B. and Jaenisch, R. (1996) DNA hypomethylation can activate *Xist* expression and silence X-linked genes. Genes Dev. **10:** 1991-2002

75 Beard, C., Li, E. and Jaenisch, R. (1995) Loss of methylation activates *Xist* in somatic but not in embryonic cells. Genes Dev. **9:** 2325-2334

76 Hendrich, B. D., Brown, C. J. and Willard, H. F. (1993) Evolutionary conservation of possible functional domains of the human and murine *XIST* genes. Hum. Mol. Genet. **2:** 663-672

77 Tinker, A. V. and Brown, C. J. (1998) Induction of *XIST* expression from the human active X chromosome in mouse/human somatic cell hybrids by DNA demethylation. Nucl. Acids Res. **26:** 2935-2940

78 Hansen, R. S., Canfield, T. K., Stanek, A. M., Keitges, E. A. and Gartler, S. M. (1998) Reactivation of *XIST* in normal fibroblasts and a somatic cell hybrid: abnormal localization of *XIST* RNA in hybrid cells. Proc. Natl. Acad. Sci. USA **95:** 5133-5138

79 Sado, T., Wang, Z., Sasaki, H. and Li, E. (2001) Regulation of imprinted X-chromosome inactivation in mice by *Tsix*. Development **128:** 1275-1286

80 Shibata, S. and Lee, J. (2003) Characterization and quantitation of differential *Tsix* transcripts: implications for *Tsix* function. Hum. Mol. Genet. **12:** 125-136

81 Lee, J., Davidow, L. and Warshawsky, D. (1999) *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. Nat. Genet. **21:** 400-404

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

83 Debrand, E., Chureau, D., Arnaud, P., Avner, P. and Heard, E. (1999) Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression. Mol. Cell. Biol. **19:** 8513-8525

84 Luikenhuis, S., Wutz, A. and Jaenisch, R. (2001) Antisense transcription through the *Xist* locus mediates *Tsix* function in embryonic stem cells. Mol. Cell. Biol. **21:** 8512-8520

85 Stavropoulos, N., Lu, N. and Lee, J. T. (2001) A function role for *Tsix* transcription in blocking *Xist* RNA accumulation but not in X chromosome choice. Proc. Natl Acad. Sci. USA **98:** 10232-10237

86 Boumil, R. M. and Lee, J. T. (2001) Forty years of decoding the silence in X-chromosome inactivation. Hum. Mol. Genet. **10:** 2225-2232

87 Bartolomei, M. S., Webber, A. L., Brunkow, M. E. and Tilghman, S. M. (1993) Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev **7:** 1663-73

88 Migeon, B. R., Chowdhury, A. K., Dunston, J. A. and McIntosh, I. (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:** 951-960

89 Migeon, B. R., Lee, C. H., Chowdhury, A. K. and Carpenter, H. (2002) Species differences in *TSIX/Tsix* reveal the roles of these genes in X-chromosome inactivation. Am. J. Hum. Genet. **71:** 286-293

90 Johnston, C. M., Nesterova, T. B., Formstone, E. J., Newall, A. E., Duthie, S. M., Sheardown, S. A. and Brockdorff, N. (1998) Developmentally regulated *Xist* promoter switch mediates initiation of X inactivation. Cell **94:** 809-817

91 Romer, J. and Ashworth, A. (2000) The upstream region of the mouse *Xist* gene contains two ribosomal protein pseudogenes. Mamm. Genome **11:** 461-463

92 Lee, J. (1999) Genetic analysis of the mouse X inactivation center defines an 80-kb multifunction domain. Proc. Natl. Acad. Sci. USA **96:** 3836-3841

93 Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, D. and Spector, D. L. (2001) Methylation of histone H3 at lys-9 is an early mark on the X chromosome during X inactivation. Cell **107:** 727-738

94 O'Neill, L., Keohane, A., Lavender, J., McCabe, V., Heard, E., Avner, P., Brockdorff, N. and Turner, B. (1999) A developmental switch in H4 acetylation upstream of *Xist* plays a role in X chromosome inactivation. EMBO **18:** 2897-2907

95 Clerc, P. and Avner, P. (1998) Role of the region 3' to *Xist* exon 6 in the counting process of X-chromosome inactivation. Nat. Genet. **19:** 249-253

96 Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S. and Lee, J. T. (2002) CTCF, a candidate *trans*-acting factor for X-inactivation choice. Science **295:** 345-347

97 Percec, I., Plenge, R., Nadeau, J., Bartolomei, M. and Willard, H. (2002) Autosomal dominant mutations affecting X inactivation choice in the mouse. Science **296:** 1136-1139

98 Johnston, P. G. and Cattanach, B. M. (1981) Controlling elements in the mouse IV. Evidence of non-random X-inactivation. Genet. Res. **37:** 151-160

99 Clerc, P. and Avner, P. (2003) Multiple elements within the Xic regulate random X inactivation in mice. Semin Cell Dev Biol **14:** 85-92

100 Lehner, B., Williams, G., Capmbell, R. D. and Sanderson, C. M. (2002) Antisense transcripts in the human genome. Trends in Genet. **18:** 63-65

101 Matzke, M., Matzke, A. and Kooter, J. (2001) RNA: Guiding gene silencing. Science **293:** 1080-1083

102 Volpe, T., Kidner, C., Hall, I., Teng, G., Grewal, S. and Martienssen, R. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. **297:** 1833-1837

103 Hall, I., Shankaranarayana, G., Noma, K., Ayoub, N., Cohen, A. and Grewal, S. (2002) Establishment and maintenance of a heterochromatin domain. Science **297:** 2232-2237

104 Maison, C., Bailly, D., Peters, A., Quivy, J., Roche, D., Taddei, A., Lachner, M., Jenuwein, T. and Almouzni, G. (2002) Higher-order structure in pericentric

heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat. Genet. **30:** 329-334

105 Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T., Webster, Z., Peters, A., Jenuwein, T., Otte, A. and Brockdorff, N. (2003) Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell **4:** 481-495

106 Lima-de-Faria and Jaworksa, H. (1968) Late DNA synthesis in heterochromatin. Nature **217:** 138-142

107 Gilbert, D. M. (2002) Replication timing and transcriptional control: beyond cause and effect. **14:** 377-383

108 Zhang, J., Xu, F., Hashimshony, T., Keshet, I. and Cedar, H. (2002) Establishment of transcriptional competence in early and late S phase. Nature **420:** 198-202

109 Delgado, S., Gomez, M., Bird, A. and Antequera, F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO **17:** 2426-2435

110 Mechali, M. (2001) DNA replication origins: from sequence specificity to epigenetics. Nat. Rev. Genet. **2:** 640-645

111 Verbovaia, L. V. and Razin, S. V. (1997) Mapping of replication origins and termination sites in the Duchenne muscular dystrophy gene. Genomics **45:** 24-30

112 Rivella, S., Palermo, B., Pelizon, C., Sala, C., Arrigo, G. and Toniolo, D. (1999) Selection and mapping of replication origins from a 500-kb region of the human X chromosome and their relationship to gene expression. Genomics **62:** 11-20

113 Cohen, S., Brylawski, B., Cordeiro-Stone, M. and Kaufman, D. (2003) Same origins of DNA replication function on the active and inactive human X chromosomes. J. Cell. Biochem. **88:** 923-931

114 Taylor, J. H. (1960) Asynchronous duplication of chromosomes in cultured cells of chinese hamster. Biophysic. Biochem. Cytol. **7:** 455-463

115 Morishima, A., Grumbach, M. M. and Taylor, J. H. (1962) Asynchronous duplication of human chromosomes and the origin of sex chromatin. Proc. Natl. Acad. Sci. USA **48:** 756-763

116 Latt, S. (1974) Microfluorometric analysis of DNA replication in human X chromosomes. Exp. Cell Res. **86:** 412-415

117 Willard, H. F. and Latt, S. A. (1976) Analysis of deoxyribonucleic acid replication in human X chromosomes by fluorescence microscopy. Am. J. Hum. Genet. **28:** 213-227

118 Willard, H. F. (1977) Tissue-specific heterogeneity in DNA replication patterns of human X chromosomes. Chromosoma **61:** 61-73

119 Babu, A., Chemitiganti, S. and Verma, R. S. (1986) Heterochromatinization of human X-chromosomes: classification of replication profile. Clin. Genet. **30:** 108-111

120 Schwemmle, S., Mehnert, K. and Vogel, W. (1989) How does inactivation change timing of replication in the human X chromosome? Hum. Genet. **83:** 26-32

121 Schempp, W. and Meer, B. (1983) Cytologic evidence for three human X-chromosomal segments escaping inactivation. Hum. Genet. **63:** 171-174

122 Yoshida, I., Kashio, N. and Takagi, N. (1993) Cell fusion-induced quick change in replication time of the inactive mouse X chromosome: an implication for the maintenance mechanism of late replication. EMBO **12:** 4397-4405

123 Takagi, N. (1974) Differentiation of X chromosomes in early female mouse embryos. Exp. Cell Res. **86:** 127-135

124 Schmidt, M. and Migeon, B. (1990) Asynchronous replication of homologous loci on human active and inactive X chromosomes. Proc. Natl. Acad. Sci. **87:** 3685-3689

125 Boggs, B. A. and Chinault, A. C. (1994) Analysis of replication timing properties of human X-chromosomal loci by fluorescence in situ hybridization. Proc. Natl. Acad. Sci. **91:** 6083-6087

126 Torchia, B., Call, L. and Migeon, B. (1994) DNA replication analysis of FMR1, XIST, and factor 8C loci by FISH shows nontranscribed X-linked genes replicate late. Am. J. Hum. Genet. **55:** 96-104

127 Hansen, R. S., Canfield, T. K. and Gartler, S. M. (1995) Reverse replication timing for the *XIST* gene in human fibroblasts. Hum. Mol. Genet. **4:** 813-820

128 Xiong, Z., Tsark, W., Singer-Sam, J. and Riggs, A. D. (1998) Differential replication timing of X-linked genes measured by a novel method using single-nucleotide primer extension. Nucl. Acids Res. **26:** 684-686

129 Hansen, R. S., Canfield, T. K., Field, A. D., Laird, C. D. and Gartler, S. M. (1997) A variable domain of delayed replication in *FRAXA* fragile X chromosomes: X inactivaton-like spread of late replication. Proc. Natl. Acad. Sci., USA **94:** 4587-4592

130 Hansen, R., Canfield, T., Lamb, M., Gartler, S. and Laird, C. (1993) Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell **73:** 1403-1409

131 Hansen, R. S., Canfield, T. K., Fjeld, A. D. and Gartler, S. M. (1996) Role of late replication timing in the silencing of X-linked genes. Hum. Mol. Genet. **5:** 1345-1353

132 Wakefield, M. J., Keohane, A. M., Turner, B. M. and Graves, J. A. M. (1997) Histone underacetylation is an ancient component of mammalian X chromosome inactivation. Proc. Natl. Acad. Sci. USA **94:** 9665-9668

133 Jenuwein, T. and Allis, C. D. (2001) Translating the histone code. Science **293:** 1074-1080

134 Lachner, M. and Jenuwein, T. (2002) The many faces of histone lysine methylation. Curr. Opin. Cell Biol. **14:** 286-298

135 Turner, B. M. (2002) Cellular memory and the histone code. Cell **111:** 285-91

136 Peters, A. H., Mermoud, J. E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N. and Jenuwein, T. (2002) Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat. Genet. **30:** 77-80

137 Boggs, B. A., Cheung, P., Heard, E., Spector, D. L., Chinault, A. C. and Allis, C. D. (2002) Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat. Genet. **30:** 73-6

138 Pirrotta, V. (2002) Silence in the germ. Cell **110:** 661-4

139 Carrel, L., Cottle, A. A., Goglin, K. C. and Willard, H. F. (1999) A first-generation X-inactivation profile of the human X chromosome. Proc. Natl. Acad. Sci., USA **96:** 14440-14444

140 Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. and Grewal, S. I. (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science **292:** 110-3

141 Jeppeson, P. and Turner, B. (1993) The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell **74:** 281-189

142 Belyaev, N. D., Keohane, A. M. and Turner, B. M. (1996) Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells. Hum. Genet. **97:** 573-578

143 Gilbert, S. L. and Sharp, P. A. (1999) Promoter-specific hypoacetylation of X-inactivated genes. Proc. Natl. Acad. Sci., USA **96:** 13825-13830

144 Morrison, H. and Jeppesen, P. (2002) Allele-specific underacetylation of histone H4 downstream from promoters is associated with X-inactivation in human cells. Chromosome Res. **10:** 579-95

145 Aguiar, R. C., Yakushijin, Y., Kharbanda, S., Salgia, R., Fletcher, J. A. and Shipp, M. A. (2000) BAL is a novel risk-related gene in diffuse large B-cell lymphomas that enhances cellular migration. Blood **96:** 4328-34

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

147 Chadwick, B. P. and Willard, H. F. (2001) Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Hum. Mol. Genet. **10:** 1101-1113

148 Chadwick, B. P. and Willard, H. F. (2001) A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. J. Cell Biol. **152:** 375-384

149 Chadwick, B. P., Valley, C. M. and Willard, H. F. (2001) Histone variant macroH2A contains two distinct macrochromatin domains capable of directing macroH2A to the inactive X chromosome. **29:** 2699-2705

150 Perche, P.-Y., Vourc'h, C., Konecny, L., Souchier, C., Robert-Nicoud, M., Dimitrov, S. and Khochbin, S. (2000) Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density. Curr. Biol. **10:** 1531-1534

151 Takahashi, I., Kameoka, Y. and Hashimoto, K. (2002) MacroH2A1.2 binds the nuclear protein Spop. Biochim. Biophys. Acta **1591:** 63-68

152 Angelov, D., Molla, A., Perche, P., Hans, F., Cote, J., Khochbin, S., Bouvet, P. and Dimitrov, S. (2003) The histone variant macroH2A interferes with transcription factor binding and SWI-SNF nucleosome remodeling. Mol. Cell. **11:** 1033-1041

153 Mermoud, J. E., Costanzi, C., Pehrson, J. R. and Brockdorff, N. (1999) Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J. Cell Biol. **147:** 1399-1408

154 Rasmusson, T. P., Matstrangelo, M.-A., Eden, A., Pehrson, J. and Jaenisch, R. (2000) Dynamic relocalization of histone macroH2A1 from centrosomes to inactive X chromosomes during X inactivation. J. Cell Biol. **150:** 1189-1198

155 Costanzi, C., Stein, P., Worrad, D. M., Schultz, R. M. and Pehrson, J. R. (2000) Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryoes. Development **127:** 2283-2289

156 Mermoud, J. E., Tassin, A.-M., Pehrson, J. and Brockdorff, N. (2001) Centrosomal association of histone macroH2A1.2 in embryonic stem cells and somatic cells. Exp. Cell Res. **268:** 245-251

157 Chadwick, B. P. and Willard, H. F. (2002) Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome. J. Cell Biol. **157:** 1113-1123

158 Rasmussen, T. P., Wutz, A., Pehrson, J. R. and Jaenisch, R. (2001) Expression of *Xist* RNA is sufficient to initiate macrochromatin body formation. Chromosoma **110:** 411-420

159 Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H. and Margolis, R. L. (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. USA **88:** 3734-3738

160 Meneghini, M., Wu, M. and Madhani, H. (2003) Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell **112:** 725-736

161 Fernandez-Capetillo, O., Mahadevaiah, S., Celeste, A., Romanienko, P., Camerini-Otero, R., Bonner, W., Manova, K., Burgoyne, P. and Nussenzweig, A.

(2003) H2AX is required for chromatin remodelling and inactivation of sex chromosomes in male mouse meiosis. Dev. Cell. **4:** 497-508

162 Goto, T. and Monk, M. (1998) Regulation of X-chromosome inactivation in development in mice and humans. Microbiol. Mol. Biol. Rev. **62:** 362-378

163 Larsen, F., Gundersen, G., Lopez, R. and Prydz, H. (1992) CpG islands as gene markers in the human genome. Genomics **13:** 1095-1107

164 Bird, A. (2002) DNA methylation patterns and epigenetic memory. Genes Dev. **16:** 6-21

165 Yoder, J., Soman, N., Verdine, G. and Bestor, T. (1997) DNA methyltransferases in mouse tissues and cells. Studies with a mechanism-based probe. J. Mol. Biol. **270:** 385-395

166 Attwood, J., Yung, R. and Richardson, B. (2002) DNA methylation and the regulation of gene transcription. Cell. Mol. Life Sci. **59:** 241-257

167 Okano, M., Xie, S. and Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat. Genet. **19:** 219-220

168 Okano, M., Bell, D., Haber, D. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99:** 247-257

169 Kim, G., Ni, J., Kelesoglu, N., Roberts, R. and Pradhan, S. (2002) Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO **21:** 4183-4195

170 Liang, G., Chan, M., Tomigahara, Y., Tsai, Y., Gonzales, F., Li, E., Laird, P. and Jones, P. (2002) Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol. Cell. Biol. **22:** 480-491

171 Rhee, I., Bachman, K., Park, B., Jair, K., Chiu Yen, R., Schuebel, K., Cui, H., Feinberg, A., Lengauer, C., Kinzler, K., Baylin, S. and Vogelstein, B. (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature **416:** 552-556

172 Meehan, R. (2003) DNA methylation in animal development. Semin. Cell. Dev. Biol. **14:** 53-65

173 Holmgren, C., Kanduri, C., Dell, G., Ward, A., Mukhopadhya, R., Kanduri, M., Lobanenkov, V. and Ohlsson, R. (2001) CpG methylation regulates the Igf2/H19 insulator. Curr. Biol. **11:** 1128-1130

174 Comb, M. and Goodman, H. (1990) CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. Nucleic Acids res. **18:** 3975-3982

175 Lewis, J., Meehan, R., Henzel, W., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell **69:** 905-914

176 Nan, X., Ng, H., Johnson, C., Laherty, C., Turner, B., Eisenman, R. and Bird, A. (1998) Transcriptional repression by the methyl-cpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393:** 386-389

177 Wade, P., Gegonne, A., Jones, P., Ballestar, E., Aubry, F. and Wolffe, A. (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat. Genet. **23:** 62-66

178 Zhang, Y., Ng, H., Erdjument-Bromage, H., Tempst, P., Bird, A. and Reinberg, D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. **13:** 1924-1935

179 Gartler, S. M. and Riggs, A. D. (1983) Mammalian X-chromosome inactivation. Ann. Rev. Genet. **17:** 155-190

180 Grant, S. and Chapman, V. (1988) Mechanisms of X-chromosome regulation. Annu. Rev. Genet. **22:** 199-233

181 Sado, T., Fenner, M., Tan, S., Tam, P., Shioda, T. and Li, E. (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev. Biol. **225:** 294-303

182 Hansen, R., Stoger, R., Wijmenga, C., Stanek, A., Canfield, T., Luo, P., Matarazzo, M., D'Esposito, M., Feil, R., Gimelli, G., Weemaes, C., Laird, C. and Gartler, S. (2000) Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum. Mol. Genet. **9:** 2575-2587

183 Kratzer, P. G., Chapman, V. M., Lambert, H., Evans, R. E. and Liskay, R. M. (1983) Differences in the DNA of the inactive X chromosomes of fetal and extraembryonic tissues of mice. Cell **33:** 37-42

184 Graves, J. (1996) Mammals that break the rules: genetics of marsupials and monotremes. Annu. Rev. Genet. **30:** 233-260

185 Eils, R., Dietzel, S., Bertin, E., Schrock, E., Speicher, M. R., Ried, T., Robert-Nicoud, M., Cremer, C. and Cremer, T. (1996) Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. J. Cell Biol. **135:** 1427-1440

186 Barr, M. L. and Bertram, E. G. (1949) A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature **163:** 676-677

187 Dyer, K. A., Canfield, T. K. and Gartler, S. M. (1989) Molecular cytological differentiation of active from inactive X domains in interphase: implications for X chromosome inactivation. Cytogenet. Cell Genet. **50:** 116-120

188 Therman, E., Sarto, G. E., Disteche, C. and Denniston, C. (1976) A possible active segment on the inactive human X chromosome. Chromosoma **59:** 137-145

189 Flejter, W. L., Van Dyke, D. L. and Weiss, L. (1984) Bends in human mitotic metaphase chromosomes, including a bend marking the X-inactivation center. Am. J. Hum. Genet. **36:** 218-226

190 Van Dyke, D. L., Worsham, M. J. and Weiss, L. (1987) The human inactivated X chromosome folds in early metaphase, prometaphase, and prophase. Hum. Genet. **77:** 57-59

191 Flejter, W. L., Van Dyke, D. L. and Weiss, L. (1986) Location of the X inactivation center in primates and other mammals. Hum. Genet. **74:** 63-66

192 Walker, C. L., Cargile, C. B., Floy, K. M., Delannoy, M. and Migeon, B. R. (1991) The Barr body is a looped X chromosome formed by telomere association. Proc. Natl. Acad. Sci. USA **88:** 6191-6195

193 Dietzel, S., Eils, R., Satzler, K., Bornfleth, H., Jauch, A., Cremer, C. and Cremer, T. (1998) Evidence against a looped structure of the inactive human X-chromosome territory. Exp. Cell Res. **240:** 187-196

194 Kerem, B.-S., Goten, R., Richler, C., Marcus, M. and Cedar, H. (1983) In situ nick-translation distinguishes between active and inactive X chromosomes. Nature **304:** 88-90

195 Dyer, K. A., Riely, D. and Gartler, S. M. (1985) Analysis of inactive X chromosome structure by in situ nick translation. Chromosoma **92:** 209-213

196 Elgin, S. C. R. (1981) DNase I-hypersensitive sites of chromatin. Cell **27:** 413-415

197 Riley, D. E., Canfield, T. K. and Gartler, S. M. (1984) Chromatin structure of active and inactive human X chromosomes. Nucl. Acids Res. **12:** 1829-1845

198 Yang, T. P. and Caskey, T. (1987) Nuclease sensitivity of the mouse HPRT gene promoter region: differential sensitivity on the active and inactive X chromosomes. Mol. Cell. Biol. **7:** 2994-2998

199 Lin, D. and Chinault, A. C. (1988) Comparative study of DNase I sensitivity at the X-linked human HPRT locus. Somat. Cell Mol. Genet. **14:** 261-272

200 Beggs, A. H. and Migeon, B. R. (1989) Chromatin loop structure of the human X chromosome: relevance to X inactivation and CpG clusters. Mol. Cell. Biol. **9:** 2322-2331

201 Lyon, M. F. (1962) Sex chromatin and gene action in the mammalian X-chromosome. Am. J. Hum. Genet. **14:** 135-145

202 Salido, E. C., Li, X. M., Yen, P. H., Martin, N., Mohandas, T. K. and Shapiro, L. J. (1996) Cloning and expression of the mouse pseudoautosomal steroid sulphatase gene (Sts). Nat. Genet. **13:** 83-6

203 Dal Zotto, L., Quaderi, N. A., Elliott, R., Lingerfelter, P. A., Carrel, L., Valsecchi, V., Montini, E., Yen, C.-H., Chapman, V., Kalcheva, I., Arrigo, G., Zuffardi, O., Thomas, S., Willard, H. F., Ballabio, A., Disteche, C. M. and Rugarli, E. I. (1998) The mouse Mid1 gene: implications for the pathogenesis of Opitz syndrome and the evolution of the mammalian pseudoautosomal region. Hum. Mol. Genet. **7:** 489-499

204 Ciccodicola, A., D'Esposito, M., Esposito, T., Gianfrancesco, F., Migliaccio, C., Miano, M. G., Matarazzo, M. R., Vacca, M., Franze, A., Cuccurese, M., Cocchia, M., Curci, A., Terracciano, A., Torino, A., Cocchia, S., Mercanante, G., Pannone, E., Archidiacono, N., Rocchi, M., Schlessinger, D. and D'Urso, M. (2000) Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudoautosomal region. Hum. Mol. Genet. **9:** 395-401

205 Brown, C. J. and Willard, H. F. (1993) Molecular and genetic studies of human X chromosome inactivation. Adv. Dev. Biol. **2:** 37-72

206 Migeon, B. R., Sprenkle, J. A. and Do, T. T. Studies of human-mouse cell hybrids with respect to X-chromosome inactivation. in *Genetic mosaics and chimeras in mammals*, Vol. 12 (ed. Russell, L.B.) 329-337 (Plenum Press, New York, 1978).

207 Gartler, S. M., Andina, R. and Gant, N. (1975) Ontogeny of X-chromosome inactivation in the female germ line. Experimental Cell Research **9:** 444-457

208 Carrel, L. and Willard, H. F. (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. USA **96:** 7364-7369

209 Kutsche, R. and Brown, C. J. (2000) Determination of X-chromosome inactivation status using X-linked expressed polymorphisms identified by database searching. Genomics **65:** 9-15

210 Vasques, L. R. and Pereira, L. V. (2001) Allele-specific X-linked gene activity in normal human cells assayed by expressed single nucleotide polymorphisms (cSNPs). DNA Res. **8:** 173-7

211 Esposito, T., Gianfrancesco, F., Ciccodicola, A., D'Esposito, M., Nagaraja, R., Mazzarella, R., D'Urso, M. and Forabosco, A. (1997) Escape from X inactivation of two new genes associated with DXS6974E and DXS7020E. Genomics **43:** 183-190

212 Anderson, C. L. and Brown, C. J. (1999) Polymorphic X-chromosome inactivation of the human TIMP1 gene. Am. J. Hum. Genet. **65:** 699-708

213 Disteche, C. M. (1999) Escapees on the X chromosome. **96:** 14180-14182

214 Tsuchiya, K. D. and Willard, H. F. (2000) Chromosomal domains and escape from X inactivation: comparative X inactivation analysis in mouse and human. Mamm. Genome **11:** 849-854

215 Miller, A. P. M., Gustashaw, K., Wolff, D. J., Rider, S. H., Monaco, A. P., Eble, B., Schlessinger, D., Gorski, J. L., van Ommen, G.-J., Weissenbach, J. and Willard, H. F. (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:** 731-739

216 Carrel, L., Clemson, C. M., Dunn, J. M., Miller, A. P., Hunt, P. A., Lawrence, J. B. and Willard, H. F. (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:** 391-402

217 Lyon, M. F. (1998) X-chromosome inactivation: a repeat hypothesis. Cytogenet. Cell Genet. **80:** 133-137

218 Bailey, J. A., Carrel, L., Chakravarti, A. and Eichler, E. E. (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. Proc. Natl. Acad. Sci. USA **97:** 6634-6639

219 Schmidt, M. and Du Sart, D. (1992) Functional disomies of the X chromosome influence the cell selection and hence the inactivation pattern in females with balanced X;autosome translocations: a review of 122 cases. Am. J. Med. Genet. **42:** 161-169

220 Disteche, C. M., Eicher, E. M. and Latt, S. A. (1979) Late replication in an X-autosome translocation in the mouse: Correlation with genetic inactivation and evidence for selective effects during embryogenesis. Proc. Natl. Acad. Sci. U.S.A. **76:** 5234-5238

221 Russell, L. B. (1963) Mammalian X-chromosome action: inactivation limited in spread and in region of origin. Science **140:** 976-978

222 Sharp, A., Robinson, D. O. and Jacobs, P. (2001) Absence of correlation between late-replication and spreading of X inactivation in an X;autosome translocation. Hum. Genet. **109:** 295-302

223 Sharp, A. J., Spotswood, H. T., Robinson, D. O., Turner, B. M. and Jacobs, P. A. (2002) Molecular and cytogenetic analysis of the spreading of X inactivation in X;autosome translocations. Hum. Mol. Genet. **11:** 3145-3156

224 Schanz, S. and Steinback, P. (1989) Investigation of the "variable spreading" of X inactivation into a translocated autosome. Hum. Genet. **82:** 244-248

225 Hall, L. L., Clemson, C. M., Byron, M., Wydner, K. and Lawrence, J. B. (2002) Unbalanced X;Autosome translocations provide evidence for sequence specificity. Hum. Mol. Genet. **in press:**

226 Clemson, C. M., Chow, J. C., Brown, C. J. and Lawrence, J. B. (1998) Stabilization and localization of *Xist* RNA are controlled by separate mechanisms and are not sufficient for X inactivation. J. Cell. Biol. **142:** 13-23

227 Gartler, S. M. and Goldman, M. A. (1994) Reactivation of inactive X-linked genes. Dev. Genet. **15:** 504-514

228 Hors-Cayla, M. C., Couturier, J., Homman, N., Heuerts, S. and Frézal, J. (1987) Inactive X chromosome gene reactivation induced by 5-azacytidine, is not correlated with the modification of the pattern of replication. Ann. Genet. **30:** 153-157

229 Schmidt, M., Wolf, S. F. and Migeon, B. R. (1985) Evidence for a relationship between DNA methylation and DNA replication from studies of the 5-azacytidine-reactivated allocyclic X chromosome. Exp. Cell Res. **158:** 301-310

230 Gregory, P. and Wang, N. (1986) Alterations in the X chromosome replication pattern induced by 5-azacytidine in a human tumor line. Cancer Genet. Cytogenet. **20:** 263-267

231 Shafer, D. A. and Priest, J. H. (1984) Reversal of DNA methylation with 5-azacytidine alters chromosome replication patterns in human lymphocyte and fibroblast cultures. Am. J. Hum. Genet. **36:** 534-545

232 Haaf, T., Ott, G. and Schmid, M. (1988) Inhibition of condensation in the late-replicating X chromosome induced by 5-azadeoxycytidine in human lymphocyte cultures. Hum. Genet. **79:** 18-23

233 Kokalj-Vokac, N., Almeida, A., Viegas-Peguignot, E., Jeanpierre, M., Malfoy, B. and Dutrillaux, B. (1993) Specific induction of uncoiling and recombination by azacytidine in classical satellite-containing constitutive heterochromatin. Cytogenet. Cell Genet. **63:** 11-15

234 Hassan, K. M. A., Norwood, T., Gimelli, G., Gartler, S. M. and Hansen, R. S. (2001) Satellite 2 methylation patterns in normal and ICF syndrome cells and association of hypomethylation with advanced replication. Hum. Genet. **109:** 45-462

235 Tamaru, H. and Selker, E. U. (2001) A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature **414:** 277-83

236 Jackson, J. P., Lindroth, A. M., Cao, X. and Jacobsen, S. E. (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416:** 556-60

237 Dennis, K., Fan, T., Geiman, T., Yan, Q. and Muegge, K. (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev. **15:** 2940-4

238 Santoro, R., Li, J. and Grummt, I. (2002) The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. Nat. Genet. **32:** 393-6

239 Leal, C. A., Ayala-Madrigal, M. L., Figuera, L. E. and Medina, C. (1998) Histone H4 acetylation analyses in patients with polysomy X: implications for the mechanisms of X inactivation. Hum. Genet. **103:** 29-33

240 Luo, S., Torchia, B. S. and Migeon, B. R. (1995) XIST expression is repressed when X inactivation is reversed in human placental cells: a model for study of XIST regulation. Somat. Cell Mol. Genet. **21:** 51-60

241 Sangha, K. K., Stephenson, M. D., Brown, C. J. and Robinson, W. P. (1999) Extremely skewed X-chromosome inactivation is increased in women with recurrent spontaneous abortion. Am. J. Hum. Genet. **65:** 913-917

242 Pegoraro, E., Whitaker, J., Mowery-Rushton, P., Surti, U., Lanasa, M. and Hoffman, E. (1998) Familial skewed X inactivation: a molecular trait associated with high spontaneous-abortion rate maps to Xq28. Am. J. Hum. Genet. **61:** 160-170

243 Kristiansen, M., Langerod, A., Knudsen, G. P., Weber, B. L., Borresen-Dale, A.-L. and Orstavik, K. H. (2002) High frequency of skewed X inactivation in young breast cancer patients. J. Med. Genet. **39:** 30-33

244 Buller, R. E., Sood, A. K., Lallas, T., Buekers, T. and Skilling, J. S. (1999) Association between nonrandom X-chromosome inactivation and BRCA1 mutation in germline DNA of patients with ovarian cancer. J. Natl. Cancer Inst. **91:** 339-346

245 Cheng, P. C., Gosewehr, J., Kim, T. M., Velicescu, M., Wan, M., Zheng, J., Felix, J. C., Cofer, K. F., Luo, P., Biela, B., Godorov, G. and Dubeau, L. (1996) Potential role of the inactivated X chromosome in ovarian epithelial tumor development. J. Nat. Cancer Inst. **88:** 510-518

246 Plass, C. (2002) Cancer epigenomics. Hum. Mol. Genet. **11:** 2479-88

247 Dean, W., Santos, F. and Reik, W. (2003) Epigenetic reprogramming in early mammalian development and following somatic nuclear transfer. Semin. Cell Dev. Biol. **14:** 93-100

248 Hove, J. R., Koster, R. W., Forouhar, A. S., Acevedo-Bolton, G., Fraser, S. E. and Gharib, M. (2003) Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis. Nature **421:** 172-7

249 Etchegaray, J. P., Lee, C., Wade, P. A. and Reppert, S. M. (2003) Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. Nature **421:** 177-82

250 Guan, Z., Giustetto, M., Lomvardas, S., Kim, J. H., Miniaci, M. C., Schwartz, J. H., Thanos, D. and Kandel, E. R. (2002) Integration of long-term-memory-related

synaptic plasticity involves bidirectional regulation of gene expression and chromatin structure. Cell **111:** 483-93

251 Rutherford, S. L. and Henikoff, S. (2003) Quantitative epigenetics. Nat. Genet. **33:** 6-8

252 Fogh, J. and Trempe, G. New human tumor cell lines. in *Human tumor cells in vitro* (ed. Fogh, J.) 115-159 (Plenum Press, New York, 1975).

253 Andrews, P. W. (1984) Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. Dev. Biol. **103:** 285-293

254 Andrews, P. W., Damjanov, I., Simon, D., Banting, G. S., Carlin, C., Dracopoli, N. C. and Fogh, J. (1984) Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Lab. Invest. **50:** 147-162

255 Mejia, J. E. and Monaco, A. P. (1997) Retrofitting vectors for *Escherichia coli*-based artificial chromosomes (PACs and BACs) with markers for transfection studies. Genome Res. **7:** 179-186

256 Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R. and Willard, H. F. (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature **349:** 38-44

257 Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156-159

258 Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual. 2nd edn (Cold Spring Harbor Laboratory Press, Plainview, New York, 1989).

259 De Preter, K., Speleman, F., Combaret, V., Lunec, J., Laureys, G., Eussen, B. H. J., Francotte, N., Board, J., Pearson, A. D. J., De Paepe, A., Van Roy, N. and Vandesompele, J. (2002) Quantification of *MYCN*, *DDX1*, and *NAG* gene copy number in neuroblastoma using real-time quantitative PCR assay. Mod. Pathol. **15:** 159-166

260 Del Pozzo, G. and Guardiola, J. (1996) The regulation mechanism of *HLA* class II gene expression at the level of mRNA stability. Immunogenetics **44:** 453-458

261 Fey, E. G., Krochmalnic, G. and Penman, S. (1986) The non-chromatin substructures of the nucleus: the ribonucleoprotein RNP-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. J. Cell Biol. **102:** 1654-1665

262 Tam, R., Johnson, C., Shopland, L., McNeil, J. and Lawrence, J. B. *Applications of RNA FISH for visualizing gene expression and nuclear architecture*, (Oxford University Press, Oxford, 2002).

263 Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R. and Miller, W. (2000) PipMaker--A web server for aligning two genomic DNA sequences. Genome Res. **10:** 577-586

264 Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H. F., Avner, P. and

Ballabio, A. (1991) Characterization of a murine gene expressed from the inactive X chromosome. Nature **251:** 325-329

265 Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., G.D., P., Patel, D. and Rastan, S. (1991) Conservation of position and exclusive expression of mouse *Xist* from the inactive X chromosome. Nature **351:** 329-351

266 Memili, E., Hong, Y., Kim, D., Ontiveros, S. D. and Strauss, W. M. (2001) Murine *Xist* RNA isoforms are different at their 3' ends: a role for differential polyadenylation. Gene **266:** 131-137

267 Hong, Y., Ontiveros, S. D., Chen, C. and Strauss, W. M. (1999) A new structure for the murine *Xist* gene and its relationship to chromosome choice/counting during X-chromosome inactivation. Proc. Natl Acad. Sci. USA **96:** 6829-6834

268 Warshawsky, D., Stavropoulos, N. and Lee, J. T. (1999) Further examination of the Xist promoter-switch hypothesis in X inactivation: Evidence against the existence and function of a Po promoter. Proc. Natl Acad. Sci. USA **96:** 14424-14429

269 Hartshorn, C., Rice, J. E. and Wangh, L. J. (2002) Developmentally-regulated changes of *Xist* RNA levels in single preimplantation mouse embryos, as revealed by quantitative real-time PCR. Mol. Reprod. Dev. **61:** 425-436

270 Zuccotti, M., Boiani, M., Ponce, R., Guizzardi, S., Scandroglio, R., Garagna, S. and Redi, C. A. (2002) Mouse *Xist* expression begins at zygotic genome activation and is timed by a zygotic clock. Mol. Reprod. Dev. **61:** 14-20

271 Kay, G. F., Penny, G. D., Patel, D., Ashworth, A., Brockdorff, N. and Rastan, S. (1993) Expression of *Xist* during mouse development suggests a role in the initiation of X chromosome inactivation. Cell **72:** 171-182

272 Kay, G. F., Barton, S. C., Surani, M. A. and Rastan, S. (1994) Imprinting and X chromosome counting mechanisms determine *Xist* expression in early mouse development. Cell **77:** 639-650

273 Andrews, P. W. (1998) Teratocarcinomas and human embryology: Pluripotent human EC cell lines. APMIS **106:** 158-168

274 Takagi, N. and Martin, G. R. (1984) Studies of the temporal relationship between the cytogenetic and biochemical manifestations of X-chromosome inactivation during the differentiation of LT-1 teratocarcinoma stem cells. Dev. Biol. **103:** 425-433

275 Looijenga, L. H. J., Gillis, A. J. M., van Gurp, R. J. H. L. M., Verkerk, A. J. M. H. and Oosterhuis, J. W. (1997) X inactivation in human testicular tumors: *XIST* expression and androgen receptor methylation status. Am. J. Pathol. **151:** 581-590

276 Buzin, C. H., Mann, J. R. and Singer-Sam, J. (1994) Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. Development **120:** 3429-3536

277 Hong, Y., Ontiveros, S. D. and Strauss, W. M. (2000) A revision of the human *XIST* gene organization and structural comparison with mouse *Xist*. Mamm. Genome **11:** 220-224

278 Mermoud, J. E., Popova, B., Peters, A. H., Jenuwein, T. and Brockdorff, N. (2002) Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation. Curr. Biol. **12:** 247-251

279 Heard, E., Mongelard, F., Arnaud, D., Chureau, C., Vourc'h, C. and Avner, P. (1999) Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA **96:** 6841-6846

280 Migeon, B. R., Kazi, E., Haisley-Royster, C., Hu, J., Reeves, R., Call, L., Lawler, A., Moore, C. S., Morrison, H. and Jeppesen, P. (1999) Human X inactivation center induces random X chromosome inactivation in male transgenic mice. Genomics **59:** 113-121

281 Keohane, A. M., Barlow, A. L., Waters, J., Bourn, D. and Turner, B. M. (1999) H4 acetylation, XIST RNA and replication timing are coincident and define X;autosome boundaries in two abnormal X chromosomes. Hum. Mol. Genet. **8:** 337-383

282 Duthie, S. M., Nesterova, T. B., Formstone, E. J., Keohane, A. M., Turner, B. M., Zakian, S. M. and Brockdorff, N. (1999) *Xist* RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in *cis*. **8:** 195-204

283 Lee, J. T. and Jaenisch, R. (1997) Long-range *cis* effects of ectopic X-inactivation centres on a mouse autosome. Nature **386:** 275-279

284 Migeon, B. R., Winter, H., Kazi, E., Chowdhury, A. K., Hughes, A., Haisley-Royster, C., Morrison, H. and Jeppesen, P. (2001) Low-copy-number human

transgene is recognized as an X inactivation center in mouse ES cells, but fails to induce cis-inactivation in chimeric mice. Genomics **71:** 156-162

285 Heard, E., Mongelard, F., Arnaud, D. and Avner, P. (1999) *Xist* yeast artifical chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. Mol. Cell. Biol. **19:** 3156-3166

286 Avner, P. and Heard, E. (2001) X-chromosome inactivation: counting, choice and initiation. Nat. Rev. Genet. **2:** 59-67

287 Nakabayashi, K., Bentley, L., Hitchins, M. P., Mitsuya, K., Meguro, M., Minagawa, S., Bamforth, J. S., Stanier, P., Preece, M., Weksberg, R., Oshimura, M., Moore, G. E. and Scherer, S. W. (2002) Identification and characterization of an imprinted antisense RNA (*MESTIT1*) in the human *MEST* locus on chromosome 7q32. Hum. Mol. Genet. **11:** 1743-1756

288 Sleutels, F., Barlow, D. P. and Lyle, R. (2000) The uniqueness of the imprinting mechanism. Curr. Opin. Genet. Dev. **10:** 229-233

289 Lee, J. T. (2000) Disruption of imprinted X inactivation by parent-of-origin effects at *Tsix*. Cell **103:** 17-27

290 Johnston, C. M., Newall, A. E. T., Brockdorff, N. and Nesterova, T. B. (2002) *Enox*, a novel gene that maps 10kb upstream of *Xist* and partially escapes X inactivation. Genomics **80:** 236-244

291 Reizis, B., Lee, J. T. and Leder, P. (2000) Homologous genomic fragments in the mouse pre-T cell receptor alpha and *Xist* loci. Genomics **63:** 149-152

292 Hartshorn, C., Rice, J. and Wangh, L. (2003) Differential pattern of *Xist* RNA accumulation in single blastomeres isolated from 8-cell stage mouse embryos following laser zona drilling. Mol. Reprod. Dev. **64:** 41-51

293 Yao, F., Svensjo, T., Winkler, T., Lu, M., Eriksson, C. and Eriksson, E. (1998) Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. Hum. Gene Ther. **9:** 1939-1950

294 Audibert, A., Weil, D. and Dautry, F. (2002) In vivo kinetics of mRNA splicing and transport in mammalian cells. Mol. Cell Biol. **22:** 6706-6718

295 Tate, C., Haase, J., Baker, C., Boorsma, M., Magnani, F., Vallis, Y. and Williams, D. (2003) Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. Biochim. Biophys. Acta **1610:** 141-153

296 Feng, Y., Wente, S. and Majerus, P. (2001) Overexpression of the inositol phophatase SopB in human 293 cells stimulates cellular chloride influx and inhibits nuclear mRNA export. Proc. Natl. Acad. Sci. USA **98:** 875-879

297 Sauer, B. (1994) Site-specific recombination: developments and applications. Curr. Opin. Biotechnol. **5:** 521-527

298 Craig, N. (1988) The mechanism of conservative site-specific recombination. Ann. Rev. Genet. **22:** 77-105

299 Ogawa, Y. and Lee, J. (2003) *Xite*, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol Cell **11:** 731-743