*XIST*ential wanderings: The role of XIST RNA in X-chromosome inactivation

Steven C. Spusta and Michael A. Goldman*

Department of Biology, San Francisco State University, San Francisco, CA 94132-1722, USA

RESEARCH during the past decade has led to the discovery of an interesting class of non-coding RNAs. Unlike rRNA, tRNA, cRNA and snRNAs, these novel non-coding RNAs are transcribed by RNA polymerase II as are protein-encoding genes, but lack significant ORFs. Examples include the *H19* gene, which is imprinted in mammals[1], *rox*1, which is involved in the dosage compensation pathway in *Drosophila*[2], and the *XIST* locus, which is involved in mammalian X chromosome inactivation[3-5].

X chromosome inactivation

Dosage compensation is a mechanism developed in organisms with heteromorphic sex chromosomes to keep the number of expressed genes from these chromosomes constant between males and females. Experimentally, dosage compensation has been studied in a variety of organisms such as *Drosophila, C. elegans* and mammals. In *Drosophila,* males harbour a single X chromosome in which twice as much transcription occurs relative to each transcriptionally active X chromosome in the female. For the nematode *C. elegans*, dosage compensation is accomplished by females down-regulating both X chromosomes. In contrast to the mechanisms of hyper- and hypotranscription, eutherian mammals transcriptionally silence most genes on one of the female X chromosomes, thereby maintaining dosage equivalence between males and females, a process known as X chromosome inactivation (XCI)[6]. Much of what has been learned recently about the XCI process is due to the discovery and characterization of the *XIST* gene[3-5,7], found in humans, mice and most other eutherian mammals examined to date. This gene is necessary and sufficient in the developmental initiation and establishment of XCI.

XCI as a dosage compensation mechanism in mammals was proposed by Lyon. The basic tenets of the Lyon hypothesis are (1) X chromosome inactivation occurs, (2) that a single active X chromosome exists in all cells, (3) the choice of which chromosome is inactivated is random, (4) the inactivated X chromosome is somatically inherited, and (5) the process occurs early in embryogenesis.

The inactive X chromosome has a number of unique characteristics that distinguish it from the active X chromosome within the same cell. The inactive X chromosome at interphase is a morphologically distinct structure known as the Barr body, which is highly condensed and heterochromatic. The Barr body is generally found at the periphery of the cell nucleus, and can be followed through metaphase by a characteristic bend, suggesting that this heterochromatic state is retained throughout the cell cycle. Active genes replicate earlier than their inactive counterparts[8]. Studies of the inactive X chromosome by 5-bromodeoxyuridine incorporation and fluorescent *in-situ* hybridization (FISH) show that genes on the inactive X chromosome replicate later than their counterparts on the active X chromosome[9-11].

Chromatin structure differs between the active and inactive X chromosomes in several respects. Regulatory regions of genes on the active X chromosome are more sensitive to nuclease digestion than the corresponding genes on the inactive X[12-14]. A global difference in nuclease sensitivity between the two chromosomes has also been demonstrated cytologically[15,16]. Histone biochemistry differs markedly between the active and inactive X chromosomes. Experiments using anti-acetylated histone H4 antibodies and FISH analysis showed that the histone H4 of the inactive X chromosome is hypoacetylated relative to its active X counterpart, correlating with its lack of transcriptional activity[17,18]. It has also been noted that a new histone subtype, mH2A1, is enriched within the inactive X chromosome of female mammals, suggesting a possible function of this protein in XCI[19].

Generally genes that are transcriptionally active are found to be hypomethylated and inactive genes are found to be hypermethylated in key regulatory regions. Genes on the active X are generally hypomethylated, but homologous genes on the inactive X chromosome are hypermethylated[6,20,21]. Agents that inhibit DNA methylation may lead to reactivation of previously silent genes on the inactive X chromosome[20]. DNA methylation is therefore thought to play a strong role in the maintenance of XCI.

Any proposed mechanistic explanation for XCI must account for several key features of the process. (1) The

*For correspondence. (e-mail: goldman@sfsu.edu)

number of X chromosomes per cell is counted relative to the number of autosomes present. (2) One X per diploid set of chromosomes remains active, any additional X chromosomes being inactivated. (3) XCI apparently initiates at one point on each X-chromosome undergoing inactivation. (4) XCI spreads *in cis* along the X-chromosome(s). (5) The inactive state of the X is stable and somatically inherited throughout subsequent mitotic divisions.

The X inactivation center (XIC)

Several lines of evidence suggest that there is a nucleation point from which an inactivation signal originates and spreads in *cis* along the X chromosome. This nucleation center has been called the X Inactivation Center or XIC. Two genetic loci identified in humans and mice, the X inactivation center (*XIC/Xic*)[22,23] and the X controlling element (*Xce*)[24], are often used interchangeably. We believe that they are part of a complex genetic locus which we refer to as XIC.

The first genetic evidence suggesting the existence of an *XIC/Xic* came from X-autosome translocations studied in human and mouse. Russell[22] first observed that in such translocations only one of the derived chromosomes underwent inactivation. Studies suggested that a crucial region of the X chromosome was required for X inactivation. Later observations using Cattanach's translocation (between mouse chromosomes seven and the X) showed that the inactivation seemed to skip certain regions, leading to the belief in more than one *Xic*. It was later determined that these regions undergo inactivation but are later reactivated, favouring the hypothesis of a single *Xic*[25]. By analysing break points within somatic cell hybrids, the mouse *Xic* has localized to distal band D[26,27], and the human *XIC* to within one megabase of band Xq13.2 (ref. 7).

A second line of evidence for a single site for nucleation of XCI concerns a locus responsible for skewing or non-random inactivation of chromosomes in mice. This locus is called the X controlling element or *Xce*[24]. To date, four *Xce* alleles have been described in mice; Xce^a, Xce^b, Xce^c, and Xce^d [28–30]. The *Xce* maps to within the *XIC*, but is separable from *Xist* itself, lying to the 3′ end of *Xist*[31]. In Xce^a/Xce^b heterozygotes, the Xce^a-bearing X chromosome was more likely to become inactivated, and in Xce^b/Xce^c heterozygotes, the Xce^b was more likely to become inactivated. Xce^a/Xce^c heterozygotes show the greatest skewing of inactivation, although the Xce^d allele seems to be the strongest.

The strength of the *Xce* allele is inversely proportional to the degree of *Xist* expression from the inactive X, in that mice containing the Xce^d allele had lower levels of *Xist* expression than those with the Xce^b allele[4]. The possible relationships between *Xist*, the *Xic*, and the *Xce* will be explored further below.

X-inactive-specific transcript (XIST)

The molecular nature of the XIC remained a mystery until 1991, when Brown announced the discovery of the *XIST* gene in humans[5]. The *XIST* cDNA was originally identified from a screen for cDNAs of the human steroid sulphatase gene, and studies began when it was apparent that it mapped to Xq13, making it the only known gene to map to the human *XIC*. The mouse gene was discovered and mapped to the *Xic* soon after[3,4]. RNA analysis in human somatic cells and somatic cell hybrids showed that *XIST* was expressed only in female cells or cell hybrids containing an inactive X, and not male cells or active X hybrids, giving *XIST* its name (X-inactive specific transcripts). Accepted nomenclature is *XIST* in humans, *Xist* in mice, and *XIST/Xist* when referring to both loci. For simplicity in this review, we use *XIST* instead of *XIST/Xist*.

Characterization of the human *XIST* gene by northern blot analysis identified it as a very large, heterogeneous, polyadenylated transcript, 17 kb in size. RT-PCR revealed that the *XIST* RNA was alternatively spliced. The mouse *Xist* gene is 15 kb long and shares most of the same features as the human *XIST* gene.

One of the most notable features of *XIST* is its lack of protein coding capacity. By using GRAIL ORF analysis, both human and mouse sequences showed a deficiency of good Kozak initiation sequences, and frequent initiation codons upstream from putative ORFs. Analysis of some potential mouse ORFs produced a potential protein of 25 to 57 amino acids, with amino acid similarities lower than nucleic acid similarities when compared to human[32–34].

The mouse and human genes are composed of six and eight exons, respectively, with the size of two of the human exons unusually large (11 kb and 4.5 kb). Overall sequence identity between the two cDNA sequences is approximately 76%. This sequence similarity is interrupted multiple times by blocks of sequence found within human *XIST* and not in mouse, suggesting a lack of evolutionary constraint between many regions of *XIST*. The strongest sequence similarities between the mouse and human *XIST* is between five repetitive sequences found within exons 1 and 6. Two out of five of these direct tandem repeats are highly conserved between human and mouse. The most conserved sequence is at the 5′ end of exon 1, with nine repeats in human and eight in mouse. This repeat homology is the strongest evolutionarily conserved sequence between human and mouse *XIST* and may function as a protein binding site on the RNA or as a control sequence by differential methylation relating to the *Xce*[32–34].

Evidence to support an RNA–protein interaction comes from Brown[35]. They found upon UV crosslinking, *XIST* bound to two nuclear proteins that immunoprecipitate with antibodies to hnRNPC1/C2. These proteins act

as possible RNA chaperones, and may help with splicing or localization of *XIST* transcript. Further work using chromatin immunoprecipitation, or affinity chromatography using *XIST* RNA, should allow the isolation and characterization of nuclear proteins involved specifically with the *XIST*–chromatin complex.

The replication timing of *XIST* also correlates with its exclusive expression from the inactive X chromosome. It was found by FISH analysis that the *XIST* allele on the inactive X chromosome replicates before its other allele on the active X, in contrast to the majority of late replicating genes on the inactive X[9,11,36–38].

Localization of XIST RNA

Another remarkable feature of *XIST* is the subcellular localization of its RNA. Nuclear and cytoplasmic fractions were examined for the presence of XIST RNA. XIST was shown, using RT-PCR, to localize to the nuclear fraction. To further localize XIST transcript, FISH analysis was performed. It was found that XIST not only localizes to the nucleus, but that XIST co-localized to the Barr body, coating the inactive X chromosome. Analysis of male cells showed no XIST hybridization. In aneuploid cells such as 47,XXX and 49,XXXXX lines, two and four XIST RNA hybridization signals, respectively, were detected, corresponding to the number of inactive X chromosomes present[32,33]. Further FISH studies of XIST localization showed that there are two variants of XIST RNA, a minor type associated with transcriptional processing, and a major type that is a fully-spliced transcript associating with the inactive X. Digestion of chromosomal DNA and chromatin showed that XIST RNA remained bound within the nucleus, presumably bound to insoluble protein components of the nuclear matrix[39]. These studies strongly suggest that XIST RNA may be a structural element of the nuclear architecture of the inactive X chromosome.

Two recent studies show a nice colocalization of XIST RNA and chromosomal regions undergoing X inactivation. In human cells carrying either an X;6 translocation, or an X;6 insertion, Keohane[40] determined that the characteristics of XCI (XIST localization, H4 hypoacetylation, and late replication timing) were excluded from the autosomal component of these chromosomes. Duthie *et al.*[41] also used FISH analysis to study Xist RNA propagation and localization. Using the unique morphological characteristics of metaphase X chromosomes in voles, they discovered that Xist did not bind along the entire X chromosome, but exhibited a specific banded pattern of localization. The zones from which Xist was excluded were constitutive heterochromatin. Xist banding was also seen in mouse and rat cell lines. It remains to be seen if the same localization pattern is present in human cells, and whether this pattern is consistent, since previous FISH analysis showed XIST in a more granular and punctate form, spread across the whole chromosome[39,42,43]. Duthie went on to examine Xist localization in mouse X:A rearrangements, and found similar results as Keohane, with little or no spreading of Xist into autosomal regions.

XIST expression in development

Since X chromosome inactivation is a developmental process, researchers began to study *XIST* expression and its regulation in a variety of tissues, such as the germline, embryonic stem (ES) cells, preimplantation embryos, and extraembryonic tissues. In female somatic cells, *XIST* is only expressed from the inactive X chromosome[3–5]. In the germline, however, *XIST* is expressed during both male and female gametogenesis[44]. In males X inactivation only occurs in the testis during spermatogenesis, in which the single X chromosome is inactivated. The amount of XIST RNA in spermatogenesis seems lower than that in female somatic cells[44]. XIST RNA is also found in mouse newborn and adult testes as well as human testes with normal spermatogenesis[45–47]. XIST transcripts have been detected in the urogenital ridges of mouse midgestation embryos, suggesting that X inactivation may transiently occur in somatic cells during gonadal development[48]. The pattern of *XIST* expression interestingly overlaps that of the testis determining gene, *Sry*.

The counting mechanism that seems to involve *XIST* also seems to function in males, as seen in studies of testicular cancers. Cancers derived from testicular germ cells have been found to gain X chromosomes, and these chromosomes are subject to *XIST* expression and inactivation[49]. XIST RNA has also been seen to localize in male meiosis in the same fashion as in female somatic cells[50]. In mouse oocytes, *Xist* expression is only detected 12.5 to 13.5 days postcoitum (dpc), but not 13.5 to 18.5 dpc, which is coincident with the time the oocyte enters meiosis, and the X chromosome is reactivated[44]. Female X chromosomes generally remain active throughout ovulation and fertilization until inactivation occurs during the preimplantation stage. Thus *Xist* expression seems to be transient in gametogenesis, but maintains similar characteristics between males and females. In mouse preimplantaion embryos, no *Xist* expression is seen until the four to eight cell stage, just before the onset of X inactivation[45]. Unstable *Xist* transcript is then expressed from both X chromosomes at low levels until one is chosen for inactivation. The X being inactivated then alters *Xist* expression by stabilizing its transcript which then accumulates *in cis*. The active X continues to produce unstable transcript for 24–48 h, then begins to silence expression[42,51,52]. In humans, XIST RNA has been detected in preimplantation

embryos from the five to ten cell stage onwards, with XIST RNA produced from both female and male embryos, indicating that *XIST* is expressed from the maternal X chromosome in males[53,54]. This questions the influence of imprinting in human *XIST* regulation.

Regulation of the *XIST* gene

Many of the studies relating to *XIST* expression and development have focused on the regulation of *XIST* by DNA methylation of its promoter, and its possible relation to genomic imprinting. In humans a minimal *XIST* promoter has been identified with at least two *cis*-acting regulatory elements[55]. *In vitro,* these elements bind to common transcription factors such as SP1, YY1, and TBP. Studies with this promoter suggest that the regulatory elements strongly influence *XIST* promoter activity in transient transfection assays. These studies suggested that the *XIST* promoter in humans acts constitutively, and binds common transcription factors. This agrees with studies on the mouse *Xist* promoter. The mouse *Xist* minimal promoter also illustrated constitutive activity in reporter assays, and binding with common transcription factors such as SP1 (refs 56, 57). *In vivo* DNA-protein footprinting experiments examining both active and inactive *Xist* promoters found that the chromatin structure of the promoter was consistent with the activity of the allele. The results show an absence of footprints on the silent *Xist* allele on the active X, but on the inactive X, footprints were observed at the sequences for a CCAAT box, two SP1 sites and a TATA box[58]. Since the active and inactive somatic *Xist* alleles are hypomethylated and hypermethylated respectively, researchers sought possible methylated-DNA binding proteins that might influence transcriptional activity. One group identified a single 100 kDa protein[59], and another group identified two other proteins named Met1 (75 kDa) and Met2 (<120 kDa)[60]. These proteins preferentially bound to methylated *Xist* promoter sequences, and repressed promoter activity in reporter assays. These results illustrate the influence of methylation upon *Xist* promoter activity, suggesting how *Xist* might be regulated during gametogenesis by methylation induced imprinting, and the resulting silencing of one *Xist* allele in XCI.

Genomic imprinting is a phenomenon in which one allele of maternal or paternal origin is preferentially silenced[61,62]. This often correlates with the differential methylation of promoters or other regulatory regions of the maternal or paternal genes in gametes. An example of imprinting is seen in the preference of the paternal X chromosome to undergo XCI in extra-embryonic tissues such as trophectoderm, and to express *XIST*[45]. The promoter and the 5′ end of the first exon of *Xist*, in mouse gametes, were found to be differentially methylated; sperm are hypomethylated and eggs are hypermethylated[63,64]. This pattern seemed to persist after the genome wide demethylation that occurs during the preimplantation stage, suggesting that this imprint allows for the preferential paternal *Xist* expression as seen in the trophectoderm[63–65]. This pre-emptive imprint is seen in both XY and XX ES cells prior to the onset of *Xist* expression[66]. It is possible that this methylation preemptively silences the *Xist* gene, and may mark that chromosome as the active one, and the unmethylated allele the inactive. Later studies using a *Xist* promoter-luciferase construct injected into one-celled mouse embryos further illustrate that DNA methylation regulates *Xist* expression[67]. *In vitro* methylation of the construct before injection was seen to repress luciferase activity. Studies of six transgenic lines showed *Xist*–luciferase expression only in the testes, correlating with previously reported endogenous *Xist* activity in males. Paternal transmission of the construct in preimplantation embryos showed expression at the morula stage, regardless of transmission from the mother or father. In the testes, it was found that this construct would express even if methylated. These results show that gametic methylation patterns could influence *Xist* expression, but other factors are involved. These factors may include proteins found in the embryo, but not in the testes, binding to the methylated *Xist* promoter and directing transcriptional silencing. This parental imprint has been shown to be erased in some female ES cells, leading to random inactivation and *Xist* expression[68].

Although methylation is differential between gametes and remains so before the onset of random X inactivation, other studies in ES cells suggest that it may be mosaic in nature, and that the imprint can persist after proliferation *in vitro*. Clones derived from single female ES cells show the absence of the allelic-specific methylation pattern, suggesting that ES *Xist* alleles are mosaically methylated, rather than differentially methylated[68]. Furthermore, recent studies using bisulphite genomic sequencing have shown that although there is differential methylation of gametes in certain areas in the 5′ region of *Xist*, it is not maintained during preimplantation development[69]. We cannot rule out the possibility that the actual region responsible for imprinting is outside the region studied. It is clear that methylation is involved in *XIST* expression, but the above conflicting data suggest that more detailed analysis is needed.

To study further links between methylation and allelic *Xist* expression, Jaenisch and coworkers studied *Xist* expression in ES cells and embryos in DNA methyltransferase (*Dnmt*) deficient mice[70]. *Dnmt* deficiency resulted in *Xist* expression in male mutant mouse embryos and in differentiated ES cells. In undifferentiated ES cells, *Xist* activity seemed independent of methylation status, suggesting that methylation may only be essential in differentiating ES cells and in later de-

velopment. Further studies in differentiated *Dnmt* mutant ES cells and embryos showed *Xist* expression to be localized with the single active X in males, and with both X chromosomes in females. Genes on the inactive X chromosome, coated with *Xist* RNA, were silenced. In differentiating *Dnmt* mutant ES cells, active *Xist* transcription correlated with an increase in apoptosis, suggesting that improper *Xist*-mediated transcriptional silencing can lead to cell death and lethality in the *Dnmt* mice[71]. The results also suggest that *Xist* can mediate X inactivation in the absence of appropriate methylation of X-linked genes, and that methylation of *Xist* is required for the repression of *Xist* on the active X chromosome, to maintain its transcriptionally active state.

Studies in somatic cells show that *Xist* expression can become reactivated or altered by a number of processes such as 5-azacytosine (5AC) induced demethylation. 5AC prevents methylation at cytosine bases, and in the process leads to the reactivation of previously hypermethylated, silenced genes[20]. In two separate reports Hansen *et al.*[72] and Tinker and Brown[73] reported the reactivation of *XIST* expression in somatic cell hybrids containing a single human active X chromosome , or in XY human fibroblasts. Stable expression of *XIST* was found after several rounds of demethylation in some somatic cell hybrid lines, but the resulting *XIST* expression was insufficient to silence expressed genes from the same chromosome[73]. FISH studies, showed that the XIST RNA localization was also abnormal, in that it was much more diffuse around its parent chromosomes. This suggests that factors other than transcription, perhaps species-specific, are necessary to localize the XIST RNA properly in these somatic cell hybrids[72]. These data suggest that methylation does play a role in the regulation of *XIST* transcription in somatic cells, and that stabilization and proper localization of XIST transcript may be required for *XIST* mediated gene repression.

Similar results were found in reactivation experiments involving fusions between murine embryonal carcinoma (EC) cells and female lymphocytes[74,75]. These EC cells are divided into reactivation-competent and reactivation-incompetent classes[74]. Cell fusions with reactivation-competent EC cells showed partial methylation at the 5′ end of *Xist*, and in reactivation-incompetent EC cell fusions containing a single X chromosome, *Xist* showed full methylation. The results from the partial methylation of *Xist* and the ability to reactivate X chromosomes introduced by fusion into these cells suggest methylation-dependent regulation of *Xist* expression similar to that found in embryonic cells prior to random XCI. The same group later published results of another cell fusion in which a somatic cell hybrid containing a single human inactive X chromosome was fused into two EC cell lines. Analysis of twenty clones showed reactivation of the donor inactive X chromosome by replication timing and expression of previously inactive X linked genes. The resulting fusions had continued *XIST* expression with *XIST* methylation ranging from fully methylated to fully unmethylated[75]. The demethylation of inactive-X-bearing somatic cell hybrids by 5AC was shown to reactivate silenced genes, but suprisingly, *XIST* expression continued, even though the *XIST* promoter remained unmethylated. Furthermore, *XIST* replication timing, early before 5AC treatment, replicated even earlier[9]. These results illustrate that altering methylation can reactivate X-linked genes, but expression of *XIST*, and its silencing ability in somatic cell hybrids, is dependent upon other controlling factors in addition to the processes of methylation and demethylation. Such factors are most likely species and developmental-stage specific.

Improper or skewed expression of *XIST* by deletions and/or mutations has been analysed in humans in a clinical setting. A number of studies have examined the phenotypes associated with ring chromosomes and the changes in *XIST* expression[76–78]. Ring chromosomes are believed to be formed by breakage in Xp and Xq followed by fusion of the proximal termini. Generally, in conceptuses that survive X chromosome abnormalities, the phenotype is relatively normal, with one normal active X chromosome and a second abnormal X chromosome that is inactivated. However, females that are mosaic with small ring chromosomes [45,X/46,Xr(X)] are much more severely affected[76]. Analysis has shown that the ring X chromosomes are active, with the *XIST* locus either silent or absent. These results show that a lack of XIST alters X inactivation in these patients[78]. In another study, Plenge *et al.*[79] found *XIST* promoter mutations in two families exhibiting skewed X inactivation, with random 50:50 X inactivation replaced with 95:5 X inactivation. The preferentially inactivated X had a cytosine to guanine mutation in the *XIST* minimal promoter. In transfection assays, this promoter was two to five times less active than the normal allele. These data suggest that the *XIST* promoter is crucial in the choosing of one of the two X chromosomes for inactivation. The fact that a reduction in transcription level of the mutated *XIST* actually increases the likelihood of inactivation suggests that the amount of XIST RNA is not limiting.

An antisense *Xist* transcript, cleverly named *Tsix*, is transcribed on the opposite strand from a promoter downstream of *Xist*, in a region implicated in the counting and selection mechanism[80]. As is the case of *Xist*, both alleles are transcribed initially. Transcription of *Tsix* from the future inactive X is silenced at the onset of XCI, while *Tsix* on the future active X transcribes only until the adjacent *Xist* gene is silenced. It is possible that *Tsix* transcription plays a key role in silencing *Xist* or in preventing the formation of a stable *Xist* RNA.

XIST transgenes and knockout experiments

The clearest evidence that *XIST* plays a pivotal role in the initiation of XCI comes from experiments with *XIST* transgenes and *XIST* knockout experiments in cells and in animals. These experiments establish the *XIST* locus as necessary and sufficient for initiation of XCI, although it is not entirely clear whether or not continued presence and transcription of the locus is required for the maintenance of the inactive state.

To date, there have been two major studies using *Xist* knockouts. Penny *et al.*[81] created a targeted deletion of *Xist* in XX ES cell lines heterozygous for the *Xce* by removing the 36 bp of minimal promoter and 7 kb of the first exon using homologous recombination. A single line bearing the deletion was isolated. Both chromosomes were active prior to ES differentiation, but upon differentiation this line was found to inactivate one of the two X chromosomes, showing the counting function of the *XIC* was intact, and not within the deleted *XIST* region. These ES cells and chimaeric embryos created with these cells undergo complete non-random inactivation with 65% having the non-targeted *Xist* allele inactivated, the remainder retaining two active X chromosomes. These results show that the intact *Xist* gene is required *in cis* for inactivation to occur, and that the counting and choice functions of the *XIC* are not affected by the 7 kb deletion. A second deletion experiment by Marahrens *et al.*[82], replaced 15 kb of *Xist* with the neomycin resistance gene, deleting exons 1 through 5, leaving the promoter and part of the 5′ region of exon 1, in an XY ES line. FISH analysis showed that the integrated *Xist-neo* deletion construct was expressed. Using these cells, male and female chimeras were made and the progeny examined. Males who inherited the deleted *Xist* were normal and healthy. Spermatogenesis was not impaired and *Xist*-deficient males could produce offspring, although all were male. This finding contrasts with previous work implicating *Xist* involvement in male X inactivation[44,46,47]. Further examination showed female embryonic lethality in embryos heterozygous for normal and deleted *Xist* alleles. Chimeras were created and female mice were generated with the mutant *Xist* allele on a maternal X chromosome. Female offspring inheriting the maternal chromosome were viable; with only the paternal X chromosome bearing the wild type *Xist* inactivated in every cell. It was concluded that the original lethality in female offspring with paternal mutant *Xist* exerted its effect in extraembryonic tissues. This lethality is believed to be due to the failed imprinted *Xist*- mediated X inactivation of the paternal X chromosome. Lack of inactivation of the imprinted male X chromosome in the trophoblast would lead to genetically unbalanced cells, inappropriate gene expression, and defects in trophoblast function. This interpretation is substantiated by the fact that XO mice inheriting the mutant *Xist* allele are viable, with only one chromosome active in the trophectoderm[82]. These results again point to the need for proper *Xist* expression in XCI.

Numerous *Xist* transgene experiments have also supported a role for X*ist* in XCI. Lee *et al.*[83] introduced a 450 kb *XIST* YAC into male ES cells. Three ES lines were analysed, with the *XIST* YAC integrating in multiple copies onto different autosomes. When these transgenic XY ES lines were induced to differentiate into embryoid bodies, ectopic *Xist* expression on autosomes, as well as expression of the endogenous *XIST,* was seen. Xist transcript levels were proportional to transgene number. Since either the ectopic or the endogenous *Xist* can be chosen for expression, the elements needed for counting and/or choosing must be present in the transgene. FISH analysis showed that *Xist* transgene RNA associates with the autosome into which the transgene is inserted, and can silence a *lacZ* marker gene *in cis*.

Lee and Jaenisch[84] studied the 450 kb YAC inserted onto murine chromosome 12. They found that ectopic Xist RNA completely coats chromosome 12, and that genes on the autosome were silenced over a 50 centimorgan region. The chromosome replicated later in S phase than the normal chromosome 12, and exhibited H4 hypoacetylation. This report demonstrated that long range *cis* inactivation can occur on autosomal DNA, and that the *Xist* transgene could alter chromatin structure.

Herzing *et al.*[85] created another XY ES transgenic line using an approximately 35-kb YAC containing *Xist*. They discovered that this smaller YAC, lacking most of the flanking sequences used in previous studies, was sufficient to induce *cis* inactivation on an autosome, and that the RNA coated that autosome. In some cells, endogenous *Xist* expression was seen, suggesting that the YAC contains a counting element within, or in close proximity to *Xist*[85].

Animals with *Xist* transgenes have also been reported. Matsuura *et al.*[86] used a 350 kb YAC integrated onto autosomes, and found that *Xist* was not expressed in the two lines derived. However, a YAC construct with a 110 kb deletion, integrated into a heterochromatic region of the long arm of the Y chromosome, did express ectopic *Xist*, with expression levels similar to that in female mice. The autosomal inserts were hypermethylated in the 5′ region of *Xist*, whereas the Y-chromosome-integrated *Xist* transgene was hypomethylated, consistent with the expression status of *Xist*. These results contrast with the results of Heard *et al.*[87], who created trangenic mice with a 460 kb YAC autosomally integrated, but with no *Xist* expression in either males or females. It is possible that the discrepancies between these reports (between ES cells and various trangenic lines), could be due to factors such as copy number, YAC rearrangement, site of integration, or

the methods of introduction of YACs into ES cells and oocytes. Recently, Heard *et al.*[88] have shown that YAC *Xist* transgenes function when present in multicopy arrays (two to seven copies) but not when present as single copy inserts. Multiple copies are required both for inactivation and for counting of X chromosomes. Since the inserts do not differ in primary sequence, it is possible that some sort of repeating pattern of the same sequence may be required for correct inactivation of the chromosome.

Mechanisms of *XIST* action

A picture of *XIST* function and mechanism of action is beginning to emerge. The initiation of XCI begins with the counting of the number of X chromosomes present, and then the choice of which X chromosome to inactivate. Experiments with transgenes have shown that elements within the region surrounding *Xist* (35 kb) have the ability to be counted as an *XIC*[85], and that deletions of *Xist* from exon 1 to exon 5 exhibited inactivation of wild type X chromosomes in every cell[82]. In another deletion experiment in which a 65 kb region 3′ to *Xist* exon 6 was deleted, undifferentiated ES cells had reduced *Xist* expression, and in differentiating ES cells X inactivation was never initiated from the normal XIST allele[89]. In XO ES cells, the 3′ *XIST* deletion was able to initiate X inactivation, even in the absence of another X chromosome. These observations suggest that the counting mechanism resides 3′ to the *XIST* gene itself. Further *Xist* RNA itself is not involved in the counting mechanism[89]. This is consistent with the finding that in a region 15 kb 3′ to *Xist*, methylation levels differ between *Xce* alleles[90]. The counting region 3′ to *XIST* may be important to the imprinting (paternal X inactivation in extraembryonic tissues) and to the skewing of random X inactivation seen with various *Xce* alleles.

Prior to gastrulation, biallelic transcription of *Xist* begins, but the transcripts are apparently unstable. Subsequently, the choice of which chromosome to inactivate is made, leading to inactivation of all but a single X chromosome. Inactivation occurs coincident with the accumulation of Xist RNA around the chromosome, and is attributed to an increased stability of Xist RNA from the X chromosome(s) to be inactivated[42,51]. *XIST* is also transcribed from the active X for a short period, following which the locus is silenced and methylated[42,51]. An antisense transcript, Tsix, may be involved in regulation of *Xist* transcription[80,91].

The switch from unstable to stable Xist RNA is mediated by use of alternative promoters[52]. The unstable transcript apparently initiates at P_0, about 6.5 kb upstream from the minimal promoter region. Stable transcripts, on the other hand, initiate either at P_1 or P_2. Imprinted *Xist* expression from the paternal X chromosome in the trophectoderm is from the P_1/P_2 promoters. It is thought that the isoforms of *Xist* RNA produced from the P_1/P_2 promoters contribute to differential stability of Xist RNA, possibly due to ribonuclease sites in the 5′ end of the P_0 transcript.

Although *XIST* is essential for the initiation and propagation of *cis* inactivation, it is probably not essential for the maintenance of XCI, and it is unclear why XIST transcripts persist in somatic cells. In deletions of *XIST* in somatic cell hybrids, in human leukemia, and in radiation induced X-autosome translocations in human cells, the silence of the X is maintained despite the absence of *XIST* and the *XIC*[92–94]. Further, the X-chromosome segment detached from the X in somatic translocations apparently retains a hypoacetylation pattern characteristic of the inactive X, suggesting that continued *Xist* transcription or contact is not necessary for the maintenance of the inactive state once it has been established[92–94].

XIST transcription alone cannot, in fact, maintain a stable inactive state in somatic cell hybrids treated with 5-azacytidine, even though stable XIST transcript is produced[9,72,73]. However, the interpretation of these experiments is complicated by the fact that human XIST may not be properly localized in murine cells, interrupting the normal function of XIST RNA.

Based on the data we have reviewed here, we propose the following model for the initiation and propagation of X inactivation. Our model is consistent with the waystation model of Riggs *et al.*[95]. Both *XIST* alleles produce transcripts during early development, but the transcripts are unstable. The initial choice of an X chromosome for inactivation is probably dependent on interactions at the 3′ end of the *XIST* locus, which is also involved in counting. The locus may be imprinted in some tissues, ensuring paternal-preferential X inactivation. Allelic differences in the 3′ region may render one X chromosome more likely than others to undergo inactivation. However, if the 3′ regions on both X chromosomes are equivalent, random inactivation is expected. If a promoter mutation affects transcription of *XIST*, this might also result in skewed X inactivation. The immediate result of the selection of a particular X for inactivation is the stabilization of the XIST transcript (or the production of a stable transcript from an alternative promoter) on that X chromosome. Once an X chromosome has accumulated the stable XIST RNA, XIST begins to interact with proteins that may be bound to sites along the X chromosome, creating conformational changes allowing for local heterochromatization. In regions that escape inactivation, these high affinity sites may be absent or inaccessible, or protected by boundary elements, eventually looping out into open chromatin domains[96]. The process results in the condensation of the inactive X into a Barr body within its own

heterochromatic nuclear compartment. The sites interacting with XIST and protein could be X-specific sequences, or sequences enriched on the X chromosome, such as Lyon's LINES[97] or the LCR-like X-linked inactivator regions (XLIRs) we have proposed[98]. A similar sequence may appear in some autosomal domains, accounting for the occasional spreading of the inactive state into autosomal regions in X-autosome translocations.

Progress in our understanding of XCI has been swift since the identification of *XIST* in 1991. Yet the molecular mechanisms of *XIST* function are still not fully understood. Some of the specific questions we need to answer are the following. (1) How is *XIST* transcription, and the use of alternative promoters, regulated? (2) How is an X chromosome chosen for inactivation, whether imprinted or random? (3) What keeps XIST RNA localized around the chromosome from which it was transcribed? (4) With what protein or DNA sequences does XIST RNA interact directly? (5) What are the specific sequences on the X chromosome (the way-stations or XLIRs) responsible for interacting with XIST and transmitting the X inactivation signal to the local chromatin?

1. Brannan, C. I., Dees, E. C., Ingram, R. S. and Tilghman, S. M., *Mol. Cell. Biol.*, 1990, **10**, 28–36.
2. Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. and Davis, R. L., *Cell*, 1997, **88**, 445–457.
3. Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D. and Lawrence, C., *Nature*, 1991, **351**, 325–329.
4. Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D. and Rastan, S., *Nature*, 1991, **351**, 329–331.
5. Brown, C. J., Ballabio, A., Rupert, J. L., Lafreniere, R. G., Grompe, M., Tonlorenzi, R. and Willard, H. F., *Nature*, 1991, **349**, 38–44.
6. Gartler, S. M., Dyer, K. A. and Goldman, M. A., *Mol. Genet. Med.*, 1992, **2**, 121–160.
7. Brown, C. J., Lafreniere, R. G., Powers, V. E., Sebastio, G., Ballabio, A., Pettigrew, A. L., Ledbetter, D. H., Levy, E., Craig, I. W. and Willard, H. F., *Nature*, 1991, **349**, 82–84.
8. Goldman, M. A., Holmquist, G. P., Gray, M. C., Caston, L. A. and Nag, A., *Science*, 1984, **224**, 686–692.
9. Hansen, R. S., Canfield, T. K., Fjeld, A. D. and Gartler, S. M., *Hum. Mol. Genet.*, 1996, **5**, 1345–1353.
10. Takagi, N. and Oshimura, M., *Exp. Cell Res.*, 1973, **78**, 127–135.
11. Boggs, B. A. and Chinault, A. C., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 6083–6087.
12. Riley, D. E., Goldman, M. A. and Gartler, S. M., *Somatic Cell Mol. Genet.*, 1986, **12**, 73–80.
13. Yang, T. P. and Caskey, C. T., *Mol. Cell Biol.*, 1987, **7**, 2994–2998.
14. Lin, D. and Chinault, A. C., *Somatic Cell Mol. Genet.*, 1988, **14**, 261–272.
15. Kerem, B., Goitein, R., Richler, C., Marcus, M. and Cedar, H., *Nature*, 1983, **304**, 88–90.
16. Kerem, B.-S., Kottusch-Geiseler, V., Kalscheuer, V., Goitein, R., Sperling, K. and Marcus, M., *Chromosoma*, 1988, **96**, 227–230.
17. Jeppesen, P. and Turner, B. M., *Cell*, 1993, **74**, 281–289.
18. Keohane, A. M., Lavender, J. S., O'Neill, L. P. and Turner, B. M., *Dev. Genet.*, 1998, **22**, 65–73.
19. Costanzi, C. and Pehrson, J. R., *Nature*, 1998, **393**, 599–601.
20. Gartler, S. M., Goldman, M. A., *Dev. Genet.*, 1994, **15**, 504–514.
21. Migeon, B. R., *Trends Genet.*, 1994, **10**, 230–235.
22. Russell, L. B., *Science*, 1963, **140**, 876–978.
23. Rastan, S. and Brown, S. D., *Genet. Res.*, 1990, **56**, 99–106.
24. Cattanach, B. M., Pollard, C. E. and Perez, J. N., *Genet. Res.*, 1969, **14**, 223–235.
25. Cattanach, B. M., *Genet. Res.*, 1974, **23**, 291–306.
26. Rastan, S., *J. Embryol. Exp. Morphol.*, 1983, **78**, 1–22.
27. Rastan, S. and Robertson, E. J., *J. Embryol. Exp. Morphol.*, 1985, **90**, 379–388.
28. Cattanach, B. M. and Williams, C. E., *Genet. Res.*, 1972, **19**, 229–240.
29. Johnston, P. G. and Cattanach, B. M., *Genet. Res.*, 1981, **37**, 151–160.
30. Cattanach, B. M. and Rasberry, C., *Mouse Genome*, 1991, **89**, 565–566.
31. Simmler, M. C., Cattanach, B. M., Rasberry, C., Rougeulle, C. and Avner, P., *Mamm. Genome*, 1993, **4**, 523–530.
32. Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. and Rastan, S., *Cell*, 1992, **71**, 515–526.
33. Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, R. G., Xing, Y., Lawrence, J. and Willard, H. F., *Cell*, 1992, **71**, 527–542.
34. Hendrich, B. D., Brown, C. J. and Willard, H. F., *Hum. Mol. Genet.*, 1993, **2**, 663–672.
35. Brown, C. J. and Baldry, S. E., *Somatic Cell Mol. Genet.*, 1996, **22**, 403–417.
36. Hansen, R. S., Canfield, T. K. and Gartler, S. M., *Hum. Mol. Genet.*, 1995, **4**, 813–820.
37. Torchia, B. S., Call, L. M. and Migeon, B. R., *Am. J. Hum. Genet.*, 1994, **55**, 96–104.
38. Torchia, B. S. and Migeon, B. R., *Somat. Cell Mol. Genet.*, 1995, **21**, 327–333.
39. Clemson, C. M., McNeil, J. A., Willard, H. F. and Lawrence, J. B., *J. Cell Biol.*, 1996, **132**, 259–275.
40. Keohane, A. M., Barlow, A. L., Waters, J., Bourn, D. and Turner, B. M., *Hum. Mol. Genet.*, 1999, **8**, 377–383.
41. Duthie, S. M., Nesterova, T. B., Formstone, E. J., Keohane, A. M., Turner, B. M., Zakian, S. M. and Brockdorff, N., *Hum. Mol. Genet.*, 1999, **8**, 195–204.
42. Panning, B., Dausman, J. and Jaenisch, R., *Cell*, 1997, **90**, 907–916.
43. Clemson, C. M., Chow, J. C., Brown, C. J. and Lawrence, J. B., *J. Cell Biol.*, 1998, **142**, 13–23.
44. McCarrey, J. R. and Dilworth, D. D., *Nature Genet.*, 1992, **2**, 200–203.
45. Kay, G. F., Penny, G. D., Patel, D., Ashworth, A., Brockdorff, N. and Rastan, S., *Cell*, 1993, **72**, 171–182.
46. Richler, C., Soreq, H. and Wahrman, J., *Nature Genet.*, 1992, **2**, 192–195.
47. Salido, E. C., Yen, P. H., Mohandas, T. K. and Shapiro, L. J., *Nature Genet.*, 1992, **2**, 196–199.
48. Dolci, S., Geremia, R., Albanesi, C. and Rossi, P., *Biochem. Biophys. Res. Commun.*, 1994, **205**, 334–340.
49. Looijenga, L. H., Gillis, A. J., van Gurp, R. J., Verkerk, A. J. and Oosterhuis, J. W., *Am. J. Pathol.*, 1997, **151**, 581–590.
50. Ayoub, N., Richler, C. and Wahrman, J., *Chromosoma*, 1997, **106**, 1–10.
51. 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., *Cell*, 1997, **91**, 99–107.

52. Johnston, C. M., Nesterova, T. B., Formstone, E. J., Newall, A. E. T., Duthie, S. M., Sheardown, S. A. and Brockdorff, N., *Cell*, 1998, **94**, 809–817.
53. Daniels, R., Zuccotti, M., Kinis, T., Serhal, P. and Monk, M., *Am. J. Hum. Genet.*, 1997, **61**, 33–39.
54. Ray, P. F., Winston, R. M. and Handyside, A. H., *Hum. Mol. Genet.*, 1997, **6**, 1323–1327.
55. Hendrich, B., Plenge, R. and Willard, H., *Nucleic Acids Res.*, 1997, **25**, 2661–2671.
56. Pillet, N., Bonny, C. and Schorderet, D. F., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 12515–12519.
57. Sheardown, S. A., Newall, A. E. T., Norris, D. P., Rastan, S. and Brockdorff, N., *Gene*, 1997, **203**, 159–168.
58. Komura, J., Sheardown, S. A., Brockdorff, N., Singer-Sam, J. and Riggs, A. D., *J. Biol. Chem.*, 1997, **272**, 10975–10980.
59. Huntriss, J., Lorenzi, R., Purewal, A. and Monk, M., *Biochem. Biophys. Res. Commun.*, 1997, **235**, 730–738.
60. Allaman-Pillet, N., Djemai, A., Bonny, C. and Schorderet, D. F., *Gene Exp.*, 1998, **7**, 61–73.
61. Monk, M., *Nature*, 1987, **328**, 203–204.
62. Monk, M., *Genes Dev.*, 1988, **2**, 921–925.
63. Ariel, M., Robinson, E., McCarrey, J. R. and Cedar, H., *Nat. Genet.*, 1995, **9**, 312–315.
64. Zuccotti, M. and Monk, M., *Nat. Genet.*, 1995, **9**, 316–320.
65. Monk, M., Boubelik, M. and Lehnert, S., *Development*, 1987, **99**, 371–382.
66. Norris, D. P., Patel, D., Kay, G. F., Penny, G. D., Brockdorff, N., Sheardown, S. A. and Rastan, S., *Cell*, 1994, **77**, 41–51.
67. Goto, T., Christians, E. and Monk, M., *Mol. Reprod. Dev.*, 1998, **49**, 356–367.
68. Sado, T., Tada, T. and Takagi, N., *Dev. Dyn.*, 1996, **205**, 421–434.
69. McDonald, L. E., Paterson, C. A. and Kay, G. F., *Genomics*, 1998, **54**, 379–386.
70. Beard, C., Li, E. and Jaenisch, R., *Genes Dev.*, 1995, **9**, 2325–2334.
71. Panning, B. and Jaenisch, R., *Genes Dev.*, 1996, **10**, 1991–2002.
72. Hansen, R. S., Canfield, T. K., Stanek, A. M., Keitges, E. A. and Gartler, S. M., *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 5133–5138.
73. Tinker, A. V. and Brown, C. J., *Nucleic Acids Res.*, 1998, **26**, 2935–2940.
74. Mise, N., Sado, T., Tada, M., Takada, S. and Takagi, N., *Exp. Cell Res.*, 1996, **223**, 193–202.
75. Yoshida, I., Nishita, Y., Mohandas, T. and Takagi, N., *Exp. Cell Res.*, 1997, **230**, 208–219.
76. Dennis, N. R., Collins, A. L., Crolla, J. A., Cockwell, A. E., Fisher, A. M. and Jacobs, P. A., *J. Med. Genet.*, 1993, **30**, 482–486.
77. Migeon, B. R., Luo, S., Stasiowski, B. A., Jani, M., Axelman, J., Van Dyke, D. L., Weiss, L., Jacobs, P. A., Yang-Feng, T. L. and Wiley, J. E., *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 12025–12029.
78. Migeon, B. R., Luo, S., Jani, M. and Jeppesen, P., *Am. J. Hum. Genet.*, 1994, **55**, 497–504.
79. Plenge, R. M., Hendrich, B. D., Schwartz, C., Arena, J. F., Naumova, A., Sapienza, C., Winter, R. M. and Willard, H. F., *Nat. Genet.*, 1997, **17**, 353–356.
80. Lee, J. T., Davidow, L. S. and Warshawsky, D., *Nat. Genet.*, 1999, **21**, 400–404.
81. Penny, G. D., Kay, G. F., Sheadown, S. A., Rastan, S. and Brockdorff, N., *Nature*, 1996, **379**, 131–137.
82. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. and Jaenisch, R., *Genes Dev.*, 1997, **11**, 156–166.
83. Lee, J. T., Strauss, W. M., Dausman, J. A. and Jaenisch, R., *Cell*, 1996, **86**, 83–94.
84. Lee, J. T. and Jaenisch, R., *Nature*, 1997, **386**, 275–279.
85. Herzing, L. B. K., Romer, J. T., Horn, J. M. and Ashworth, A., *Nature*, 1997, **386**, 272–275.
86. Matsuura, S., Episkopou, V., Hamvas, R. and Brown, S. D., *Hum. Mol. Genet.*, 1996, **5**, 451–459.
87. Heard, E., Kress, C., Mongelard, F., Courtier, B., Rougeulle, C., Ashworth, A., Vourch, C., Babinet, C. and Avner, P., *Hum. Mol. Genet.*, 1996, **5**, 441–450.
88. Heard, E., Mongelard, F., Arnaud, D. and Avner, P., *Mol. Cell Biol.*, 1999, **19**, 3156–3166.
89. Clerc, P. and Avner, P., *Nature Genet.*, 1998, **19**, 249–253.
90. Courtier, B., Heard, E. and Avner, P., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 3531–3535.
91. Heard, E., Lovell-Badge, R. and Avner, P., *Nature Genet.*, 1999, **21**, 343–344.
92. Brown, C. J. and Willard, H. F., *Nature*, 1994, **368**, 154–156.
93. Rack, K. A., Chelly, J., Gibbons, R. J., Rider, S., Benjamin, D., Lafreniére, R. G., Oscier, D., Hendriks, R. W., Craig, I. W. and Willard, H. F., *Hum. Mol. Genet.*, 1994, **3**, 1053–1059.
94. Surralles, J. and Natarajan, A. T., *Cytogenet. Cell Genet.*, 1998, **82**, 58–66.
95. Riggs, A. D., Singer-Sam, J. and Keith, D. H., *Prog. Clin. Biol. Res.*, 1985, **198**, 211–222.
96. Goldman, M. A., *BioEssays*, 1988, **9**, 50–55.
97. Lyon, M. F., *Cytogenet. Cell Genet.*, 1998, **80**, 133–137.
98. Goldman, M. A., Reeves, P. S., Wirth, C. M., Zupko, W. J., Wong, M. A., Edelhoff, S. and Disteche, C. M., *Chromosome Res.*, 1998, **6**, 397–404.

ACKNOWLEDGEMENTS. This work was supported in part by an Academic Research Enhancement Award (GM52645) and Minority Biomedical Research Support Grants (GM52588) from the National Institutes of Health. We would like to thank Michael A. Wong for helpful comments in the preparation of this manuscript.