A Boundary Element Between Tsix and Xist Binds the Chromatin Insulator Ctcf and Contributes to Initiation of X-Chromosome Inactivation

Rebecca J. Spencer, Brian C. del Rosario, Stefan F. Pinter, Derek Lessing, Ruslan I. Sadreyev, and Jeannie T. Lee
Howard Hughes Medical Institute, Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114

ABSTRACT In mammals, X-chromosome inactivation (XCI) equalizes X-linked gene expression between XY males and XX females and is controlled by a specialized region known as the X-inactivation center (Xic). The Xic harbors two chromatin interaction domains, one centered around the noncoding Xist gene and the other around the antisense Tsix counterpart. Previous work demonstrated the existence of a chromatin transitional zone between the two domains. Here, we investigate the region and discover a conserved element, RS14, that presents a strong binding site for Ctcf protein. RS14 possesses an insulatory function suggestive of a boundary element and is crucial for cell differentiation and growth. Knocking out RS14 results in compromised Xist induction and aberrant XCI in female cells. These data demonstrate that a junction element between Tsix and Xist contributes to the initiation of XCI.

X-chromosome inactivation (XCI) equalizes sex chromosome-linked gene expression between male (XY) and female (XX) mammals (Lyon 1961). XCI is routinely studied in mouse embryonic stem (ES) cells, an ex vivo model that recapitulates the random form of XCI during cell differentiation (reviewed in Avner and Heard 2001; Lucchesi et al. 2005; Wutz and Gribnau 2007; Payer and Lee 2008). In ES cells and the early epiblast, all X chromosomes are active. As differentiation proceeds, each cell makes an autonomous decision regarding whether to inactivate one X. This decision is governed by a counting mechanism that determines X-chromosome number and inactivates all but one X in a diploid cell. At the same time, a choice mechanism randomly selects one of the Xs for inactivation.

Control of XCI is mediated by the X-inactivation center (Xic) (Brown et al. 1991b), an X-linked region that produces a number of noncoding RNAs (Figure 1A) (Simmler et al. 1996; Chureau et al. 2002). The X-inactive specific transcript (Xist) is responsible for initiating silencing along the X as the RNA accumulates and recruits silencing factors to the X in cis (Borsani et al. 1991; Brockdorff et al. 1991; Brown et al. 1991a; Clemson et al. 1996; Penny et al. 1996; Wutz et al. 2002). A major regulator of Xist is Tsix, which encodes another noncoding RNA, overlaps the Xist gene, and is transcribed from the opposite strand of DNA as Xist (Lee et al. 1999). Tsix expression suppresses Xist upregulation in cis and thereby designates the active X chromosome (Xa) (Lee and Lu 1999; Sado et al. 2001). Two other noncoding genes have been identified at the Xic. Xite lies upstream of Tsix and positively regulates the antisense RNA (Ogawa and Lee 2003; Stavropoulos et al. 2005). Jpx/Enox resides upstream of Xist (Chureau et al. 2002; Johnston et al. 2002), makes looping contacts with Xist (Tsai et al. 2008), and has recently been shown to activate Xist expression (Tian et al. 2010).

Genetic and physical interplay among Xite, Tsix, and Xist is central to the control of XCI. Its initiation is marked by homologous pairing of the Xs through the 5’ ends of Xite and Tsix (Bacher et al. 2006; Xu et al. 2006, 2007) in an act that is necessary for counting and choice. Following pairing, Tsix expression is extinguished on the future inactive X chromosome (Xi) with the consequent activation of Xist and global silencing of the X in cis. On the future Xa, Tsix expression persists to block Xist induction and to ensure the continued active state of the X in cis. Several studies have shown that
the coupling of Tsix and Xist chromatin states controls whether Xist will be transactivated. On the Xa, Tsix transcription through the Xist promoter results in recruitment of DNA methylation and repressive chromatin modifications to the Xist promoter. By contrast, loss of Tsix transcription on the Xi allows activating chromatin modifications to occur at the Xist promoter for transcriptional induction of Xist (Navarro et al. 2005; Sado et al. 2005; Sun et al. 2006).

A recent study has revealed a series of complex three-dimensional interactions underlying the genetic interactions between the noncoding genes (Tsai et al. 2008). It is proposed that the Xic is partitioned into two active chromatin hubs (ACHs) (Figure 1A)—permissive higher-order structures within which looping interactions between genes facilitate developmentally specific chromatin interactions (Tolhuis et al. 2002; Kosak and Groudine 2004; Splinter et al. 2006). ACH-1 is thought to encompass Xist and Jpx, delineating a domain in which physical interaction between the Xist promoter and a 5’ region of Jpx occurs and defines a state that is “poised” for Xist transactivation on the future Xi. ACH-2 encompasses Xite and the 5’ half of Tsix, defining a domain in which the Xite enhancer and Tsix promoter can physically interact and maintain Tsix expression during cell differentiation to block Xist expression on the future Xa.

It was also shown that ACH-1 and ACH-2 are separated by a chromatin transitional zone located at the 3’ terminus of Xist (Figure 1A) (Tsai et al. 2008), which was proposed to buffer opposing chromatin forces between the Tsix- and the Xist-centered domains. How this region does so is not known, but the 3’ end of Xist has a number of noteworthy features, including a strong DNase I hypersensitive site (Tsai et al. 2008). Previously thought to terminate in exon 7 (formerly exon 6), Xist is now known to extend 3 kb to include a small eighth exon (Brockdorff et al. 1992; Hong et al. 1999). Truncating Xist after exon 7 does not affect the ability of Xist to initiate XCI in the context of a 65-kb multi-gene deletion (Figure 1A) (Clerc and Avner 1998), suggesting that the 3’ end is not required for Xist’s chromosome-wide silencing properties. Other studies have proposed that this region of Xist may harbor a counting element (Herzing et al. 1997; Clerc and Avner 1998; Morey et al. 2004). Thus, several potentially interesting functions have been ascribed to the chromatin transitional zone. What is the nature of the chromatin transitional zone? Established examples of “boundaries” include a trRNA gene in yeast (Oki and Kamakaka 2005), Ctf-binding domains in vertebrates (Lewis and Murrell 2004; Murrell et al. 2004; Splinter et al. 2004; Kurukuti et al. 2006; Ling et al. 2006; Wallace and Felsenfeld 2007;
Lunyak 2008; Wan and Bartolomei 2008; Phillips and Coreces 2009), and a SINE B2 repeat in mammals (Lunyak et al. 2007). Here, we seek to gain molecular and functional understanding of the transitional zone at the Xic and discover a conserved element that binds Ctcf (Ohlsson et al. 2001; West et al. 2002) and serves a boundary function.

Materials and Methods

Cell lines and culture

EL16 female mouse ES lines, derived as described (Lee and Lu 1999), and the 40 XY male ES line J1 (Li et al. 1993), were used for targeting and transgenic experiments. EL16 contains X chromosomes from two different strain derivations, 129 and Mus castaneus. As is the case with all female ES cell lines, EL16 contains a mixture of diploid and tetraploid cells. ES cell maintenance and differentiation via embryoid body formation upon leukemia inhibitory factor (LIF) withdrawal were performed as described (Lee and Lu 1999). For differentiation with retinoic acid, 13 nM all-trans-retinoic acid (Sigma) was added to the LIF-free media.

Identification of conserved sequence and Ctcf sites

Multipipmaker (Schwartz et al. 2000) was used to align sequences corresponding to the following: Mus musculus chrX: 100,565,208–100,688,561; Homo sapiens chrX: 72,871,740–73,007,631 [University of California at Santa Cruz (UCSC) genome browser database coordinates] (Karolchik et al. 2003); and GenBank accession nos. 13445265 (Microtus arvalis), 13445266 (Microtus rossiaemeridionalis), 13445263 (Microtus transascpius), 1575005 (Equus caballus), 1575009 (Oryctolagus cuniculus), and 21425595 (Bos taurus). To identify potential Ctcf-binding sites, the Blat function (Kent 2002) was used with the UCSC genome browser database (Karolchik et al. 2003) to find sequences from various species corresponding to the 2.3-kb Clal–BamHI fragment (bp 126,980–129,252 of GenBank accession no. AJ421479) from the 3′ end of M. musculus Xist, and sequences from Rattus norvegicus (chrX: 914,45,420–91,447,696; UCSC genome browser database coordinates for this and subsequent species), Oryctolagus cuniculus (scaffold, 214795; 36610:38552), Equus caballus (chrUn: 222,426,533–222,428,947), Canus familiaris (chrX: 60,374,176–60,367,292), Pan troglodytes (chrX: 73,151,334–73,153,817), Felis catus (scaffold 203485:24–1902), and Macaca mulatta (chrX: 72,942,833–72,945,325). These sequences, along with sequence from H. sapiens Xic (bp 3974–6554 of GenBank accession no. AL353804), were searched with the Ctcf-binding site matrix (Kim et al. 2007) using the program rVista (Loots et al. 2002) (parameters: 0.95 Core Matrix Similarity, 0.75 Matrix Similarity) to locate potential binding sites within RS14.

Electrophoretic mobility shift assay

Complementary single-stranded oligonucleotides were ordered from Integrated DNA Technologies (RS14d) or from the Massachusetts General Hospital DNA Core Facility. The complementary oligonucleotides were as follows: H19, 5′-CAGTTTGCGCAGGGTTATAGCGGAGGTGCCTGGT GGCGCAGAAAATCGATTGCGCAAAACCTAAAGAG-3′ (MS1) (Hark et al. 2000); H19-reverse, 5′-CTTITTAGGTTTGGGCAATCGATTGCGCCACTCGCCGGGTA AACCCCAACAGT-3′; RS14c, 5′-TGTATACCCGTGAGGC CAGCAGGGTTGTGCGATTCACAGCGGTA TATAACA-3′; RS14c-mut, 5′-TGTATACCCGTGAGGC CAAAATGTGCGATCCACAGCGGTA TATAACA-3′; RS14c-mut-reverse, 5′-TTCCTTGGGATAGCCGACATTGTGCGCTGCTGCTGCAA CGAAGGGGATCGCCTG-3′; and RS14d-reverse, 5′-CAGGCT GGACCCTTTCCTTCACTCGGCCTCCTGTCCCACCCTCTGG ACCCCCACAACTG-3′. To identify potential Ctcf-binding sites, the Blat function (Kent 2002) was used with the UCSC genome browser database (Karolchik et al. 2003) to find sequences from various species corresponding to the 2.3-kb Clal–BamHI fragment (bp 126,980–129,252 of GenBank accession no. AJ421479) from the 3′ end of M. musculus Xist, and sequences from Rattus norvegicus (chrX: 914,45,420–91,447,696; UCSC genome browser database coordinates for this and subsequent species), Oryctolagus cuniculus (scaffold, 214795; 36610:38552), Equus caballus (chrUn: 222,426,533–222,428,947), Canus familiaris (chrX: 60,374,176–60,367,292), Pan troglodytes (chrX: 73,151,334–73,153,817), Felis catus (scaffold 203485:24–1902), and Macaca mulatta (chrX: 72,942,833–72,945,325). These sequences, along with sequence from H. sapiens Xic (bp 3974–6554 of GenBank accession no. AL353804), were searched with the Ctcf-binding site matrix (Kim et al. 2007) using the program rVista (Loots et al. 2002) (parameters: 0.95 Core Matrix Similarity, 0.75 Matrix Similarity) to locate potential binding sites within RS14.

Electrophoretic mobility shift assay

Complementary single-stranded oligonucleotides were ordered from Integrated DNA Technologies (RS14d) or from the Massachusetts General Hospital DNA Core Facility. The complementary oligonucleotides were as follows: H19, 5′-CAGTTTGCGCAGGGTTATAGCGGAGGTGCCTGGT GGCGCAGAAAATCGATTGCGCAAAACCTAAAGAG-3′ (MS1) (Hark et al. 2000); H19-reverse, 5′-CTTITTAGGTTTGGGCAATCGATTGCGCCACTCGCCGGGTA AACCCCAACAGT-3′; RS14c, 5′-TGTATACCCGTGAGGC CAGCAGGGTTGTGCGATTCACAGCGGTA TATAACA-3′; RS14c-mut, 5′-TGTATACCCGTGAGGC CAAAATGTGCGATCCACAGCGGTA TATAACA-3′; RS14c-mut-reverse, 5′-TTCCTTGGGATAGCCGACATTGTGCGCTGCTGCTGCAA CGAAGGGGATCGCCTG-3′; and RS14d-reverse, 5′-CAGGCT GGACCCTTTCCTTCACTCGGCCTCCTGTCCCACCCTCTGG ACCCCCACAACTG-3′. The enhancer-blocking assay was carried out essentially as described (Lee et al. 2002) on each of day 0 and day 4 male ES cell lines, and the results were averaged for three independent biological replicates. α-Ctcf antibodies (sc-15914) and IgG polyclonal serum (sc-2028) were obtained from Santa Cruz Biotechnology. Quantitative PCR was performed using an iCycler iQ real-time PCR detection system (Bio-Rad) by amplifying with primer pairs specific to RS14c (p142/p139, 87-bp product), RS14d (p140/p141, 102-bp product), control site A (p134/135, 87-bp product, which is 1.2 kb from RS14c), and control site B (p136/137, 86-bp product, which is 0.8 kb from RS14d (1.0 kb from RS14c)). Primer sequences were the following: p134, ACTCATCATGTTGATATGTTGAGTTGCC; p135, CTCCAAGAAGGCCCATGACCCGAGC; p136, GAAATCCTTGT TAGAACAGGCGGAA; p137, CCCACAGTTTGGCTCAGGATTG AATAC; p139, TAGGTTTGGTGTATACCCGTTAGGC; p142, CCTTAATCCATAGATACAGAGGCCTAGG; p141, AATGGCTGTG CTTTGACCCCTTGC; and p142, GGCTCACAACAGATGGCGCT GTCATACCTG.

Enhancer-blocking assay

The enhancer-blocking assay was carried out essentially as described (Chung et al. 1993; Bell et al. 1999). To produce the RS14 forward and reverse plasmids, a 2.3-kb Clal–
*Bam*HI fragment [bp 126,980–129,252 of GenBank no. AJ421479 (Chureau et al. 2002)] at the 3’ end of Xist was ligated, using *AscI* linkers, into the *Ascl* site of the no-insulator plasmid in the forward and reverse directions. Constructs were linearized by *SalI* digestion, phenol–chloroform extracted, and ethanol precipitated. Control constructs with β-globin insulators or λDNA are as described (Chung et al. 1993). A transfection efficiency control vector, pTK-hygroycin (Chao et al. 2002), was linearized with *BglII* and prepared in a similar way. K562 cells were grown in Iscove’s Modified Dulbecco’s Medium with L-glutamine, 10% fetal bovine serum, 1% penicillin–streptomycin (Gibco 15140-122), and 1.5 g/liter sodium bicarbonate. K562 cells (10^7) were electroporated with 3 pmol of each construct and 23 μg of pTK-hygroycin and were recovered for 2 days in K562 media. Electroporated cells were split into two pools and were plated in soft agar composed of K562 media with 0.42% agar, 1 μg/ml fungizone, 10 μg/ml amphotericin, and either 1 mg/ml of neomycin (G418) or 200 μg/ml ganciclovir. G418 and hygromycin-resistant colonies were counted 3–4 weeks after plating, and sample pairs with <50 hygromycin colonies were discarded. The ratios of G418 to hygromycin-resistant colonies were normalized to the value obtained with the no-insulator construct. *P*-values were calculated using the unpaired two-tailed Student’s *t*-test in pairwise comparisons.

**Targeted mutagenesis**

The PGKneobpAlox2DTA vector (Soriano 1997) was used to construct the targeting vector for RS14. Fragments corresponding to bases 119,800–126,978 and 129,252–1,322,274 of GenBank no. AJ421479 (Chureau et al. 2002) were ligated into the *SacI* site and *SalI–HindIII* sites, respectively, to create a targeting vector with a 7.2-kb homologous arm, followed by a PGK-driven neomycin resistance expression cassette flanked by *loxP* sites, a 3-kb homologous arm, and a diphtheria toxin (PGK-DTApBA; negative selection) expression cassette in a pBluescript II KS+ backbone. *PvuI* was used to linearize the construct, and 40 μg was electroporated into ~10^7 J1 and EL16 cells. Colonies were selected with G418, and homologous recombination events were screened by Southern blot using Spel and probe 1, a probe external to the targeting construct (bp 118,989–119,739 of GenBank no. AJ421479). Single insertion of the vector and correct targeting of the 129 allele was confirmed with Southern blot of a *BglI* digest using probe 2, which is located internal to the targeting construct (bp 129,826–130,568 of GenBank no. AJ421479). Targeted clones were transfected via electroporation with pMC-CRE to delete Neo via Cre-mediated recombination, and deletion was confirmed by Southern blot using BpiI and probe 1. As is the case with all female ES cell lines, the targeted female clones have a heterogeneous mixture of diploid and tetraploid cells.

**Fluorescent in situ hybridization and antibody stains**

RNA/DNA fluorescent *in situ* hybridization (FISH) was carried out as previously described (Lee and Lu 1999; Zhang et al. 2007). Wild-type male (J1) (Li et al. 1993) and female (EL16) (Lee and Lu 1999) ES cells and RS14 mutant male (B7Δ2.3) and female (A9Δ2.3) ES cells generated in this study were grown and treated under identical conditions. Embryoid body differentiation was induced by LIF withdrawal, and cells were trypsinized and harvested at days 0, 4, and 10. Cells were cytospun and fixed onto superfrost plus slides (Fisherbrand). Detection of Xist RNA was carried out using a pSx9 probe labeled by a nick translation kit (Roche) using Cy3-dUTP (Amersham). Images of the cells were taken with 0.5-μm z-sections using Velocity software (Improvision) and a Nikon Eclipse 90i microscope. DNA FISH with FITC-labeled X-paint (Cambio) was subsequently used to identify X chromosomes. Signals corresponding to overlapping Xist RNA coating of an X chromosome were counted using the extended focus view of Velocity. A minimum of 271 cells was characterized at each time point. For co-labeling of the Xic (FITC) and Neo cassette (Cy3), DNA FISH was performed with nick-translated probes generated from pSx9 and a 2.3-kb *BstI–XcmI* fragment from the RS14 deletion plasmid. Prior to DNA FISH slide denaturation, samples were treated with both RNaseA and RNaseH. Images of the cells were taken with 0.2-μm z-sections.

Immunofluorescence was performed as previously described (Zhang et al. 2007). Undifferentiated ES cells and day 8 cells from embryoid bodies (EBs) were fixed and incubated with rabbit anti-Nanog antibody (Novus) at 1:750 dilution and visualized with goat anti-rabbit-alexa555 (Invitrogen) at 1:1000 dilution. For all FISH and antibody imaging, nuclei were stained with DAPI in Vectashield mounting medium.

**Quantitative RT-PCR**

RNA was isolated from embryoid bodies derived from wild-type male (J1) and female (EL16.7) cells and RS14 mutant male (B7Δ2.3) and female (A9Δ2.3) cells using Trizol (Invitrogen). First-strand synthesis was performed with SuperScript III RT (Invitrogen) using gene-specific primers to Xist (BD127, 5’-TAAACAGCCAGTCTGATGTAACGG-3’), Tsix (NS19) (Stavropoulos et al. 2001), and tubulin α1A (YJ2, 5’-CTTGCCAGCTCCTGTCTC-3’). Real-time reactions were performed in a CFX96 Real-Time PCR System using iQ SYBR Green Supermix (Bio-Rad). Xist was amplified with primers NS66 (Stavropoulos et al. 2001) and BD127. Tsix was amplified with NS18 and NS19 (Stavropoulos et al. 2001). Tubulin α1A was amplified with primers YJ1 (5’-CTGCCAGCTCCTGTCTGTT-3’) and YJ2. Expression of Xist and Tsix was calculated relative to tubulin α1A normalized to primer pair efficiency.

**Allele-specific RT-PCR**

Allele-specific RT-PCR was performed as described (Stavropoulos et al. 2001). Briefly, total RNA was isolated from retinoic acid differentiated cells, and 2 μg was reverse transcribed with random hexamer primer. Primers NS33 (5’-
CAGAGTAGCGAGGACTTGAAGAG-3' and NS66 (5'-GCTG GTCGTCCTATCTTGTGGG-3') and 32 cycles of PCR were used to amplify polymorphic Mus musculus castaneus and 129 fragments which were digested with ScrfI. These fragments were fractionated on an agarose gel, blotted, and probed with the 32P-end-labeled probe NS67 (5'-CCA GAGTCTGATGTAACGGAGG-3'). Relative allelic amounts were quantitated by phosphorimage using the software ImageQuant. P-values were calculated using the unpaired two-tailed Student's t-test in pairwise comparison.

**Quantitation of cell death**

Day 0 cells were trypsinized and 6 x 10⁶ cells were differentiated with 100 nM retinoic acid. On day 1, cells were plated on gelatinized plates (0.2%) with 200 nM retinoic acid, and for all subsequent days were differentiated with...
13 nM of all trans-retinoic acid. At each time point, the level of cell death was measured in triplicate using the Multitox-Fluor Multiplex Cytotoxicity Assay (Promega).

**Analysis of sequence motif occurrence in Ctcf ChIP-seq data sets**

We used multiple sequence alignments of Rs14c and Rs14d in mammals to generate sequence motifs with GLAM2 (Bailey et al. 2009) and performed FIMO (Grant et al. 2011) searches against the sequence data sets corresponding to Ctcf peaks defined by genome-wide ChIP-seq experiments in mouse ES cells (Chen et al. 2008; Nitzsche et al. 2011). A FIMO q-value of 0.05 was used as a significance cutoff for the produced sequence matches.

**Results**

**Computational analysis identifies a conserved element, RS14**

To identity candidate elements for boundary function, we carried out Pipmaker analysis (Schwartz et al. 2000; Flint et al. 2001) across the Xic to search for conserved elements among eight eutherian mammalian species, including *O. caniculus* (rabbit), *B. taurus* (cow), *E. caballus* (horse), *M. transascipicus* (vole), *M. rossiaeermidionalis* (vole), *M. arvalis* (vole), *M. musculus* (mouse), and *H. sapiens* (human). We detected a region at the 3' end of Xist that exhibited a relatively high level of cross-species conservation (Figure 1B). This region, which we call “RS14,” covers a 2.3-kb region spanning Xist intron 7 and exon 8. One motif within it, designated RS14d, is relatively long (81 bp) and exceptionally well conserved (Figure 1C). On average, the Xist locus has an average sequence identity of 66% between mouse and human, approximately the same conservation level as 5' and 3' untranslated regions of protein-coding genes (Chureau et al. 2002) and the introns of beta-globin (Hendrich et al. 1993). By contrast, the 81-bp RS14d has 86% identity between mouse and human sequences and 80% identity between mouse and bovine sequences. RS14d coincides with a strong DNase I hypersensitive site (Tsai et al. 2008).

**RS14 binds Ctcf protein and displays enhancer-blocking activity**

Given the presence of a boundary-like activity at the 3' end of Xist, we asked if RS14 might bind the only known mammalian chromatin insulator, Ctcf, especially in view of the fact that Ctcf is implicated in formation of an ACH at the *H19/Igf2* imprinted locus (Murrell et al. 2004; Kurukuti et al. 2006; Ling et al. 2006) and at the beta-globin locus (Splinter et al. 2004). A search with the 20-bp Ctcf-binding site matrix (Kim et al. 2007) revealed two potential elements within RS14 that are conserved among multiple eutherian mammals. One is located in RS14d and shows conservation among nine eutherian species (Figure 2, A and B). We also found a cluster of closely spaced potential Ctcf motifs in rodents downstream of RS14d in a 110-bp domain that we designate RS14c, the best conserved of which is shown in Figure 2B.

To determine if the sites could bind Ctcf in vitro, we performed electrophoretic mobility shift assays (EMSA) using purified recombinant Ctcf protein and labeled 46- and 80-bp oligonucleotides corresponding to RS14c and RS14d, respectively. Although RS14d contains a conserved Ctcf motif, the DNA did not shift Ctcf protein under our conditions (data not shown). A more careful examination of the RS14d motif indicated that the highly conserved C-nucleotide in the sixth position of the Ctcf consensus is a T nucleotide instead in RS14d (Figure 2B, arrow). This difference may explain the inability to bind Ctcf protein. By contrast, murine RS14c carries the C nucleotide and consistently shifted Ctcf protein (Figure 2C). The protein–DNA complex was eliminated when four conserved G nucleotides within the core motif were mutated (RS14c-mut) (Figure 2B, asterisks). Furthermore, the protein–DNA complex was competed away by addition of unlabeled self-DNA or H19, which contains a
known Ctcf-binding site (Bell and Felsenfeld 2000; Hark et al. 2000; Kanduri et al. 2002), but was still present with the addition of unlabeled RS14c-mut (Figure 2C). Combined, these data indicate that Ctcf specifically binds RS14c in vitro.

To determine if RS14 is occupied by Ctcf in vivo, we carried out ChIPs using α-Ctcf antibodies in male and female ES cells before (day 0) and during XCI (day 4) and compared the results to control IgG ChIP (Figure 2D). Consistent with EMSA, we observed significant enrichment of Ctcf at RS14c over background (IgG) in both male and female ES cells and at both day 0 and day 4, indicating that Ctcf strongly occupies RS14c in vivo. At RS14d, the occupancy was much lower than at RS14c, although the enrichment was statistically significant in day 4 male and female ES cells when compared to the results of IgG ChIP. Analysis of two published ChIP-seq data sets for ES cells showed that RS14c but not RS14d was detected above the significance cutoffs in both data sets (Chen et al. 2008; Nitzsche et al. 2011).

We then performed a reciprocal bioinformatics analysis and asked whether the RS14c motif appeared more frequently in known Ctcf-binding sites, as defined by the published ChIP-seq data sets for ES cells. We generated sequence motifs based on alignments of RS14d and RS14c sequences in various species and used them as queries to search the published binding sites with the FIMO method (Grant et al. 2011). In total, the RS14c motif matched 438 sequences in the Chen et al. (2008) data set and 10,008 sequences in the Nitzsche et al. (2011) data set with a significance cutoff of q-value < 0.05 (Figure 2E; Supporting Information, Table S1; Table S2). By contrast, the RS14d motif did not identify any significant matches in either data set [the best q-values were 0.345 and 0.277 for the Chen et al. (2008) and Nitzsche et al. (2011) data sets, respectively].

Taken together, the in vivo, in vitro, and in silico analyses argue strongly that RS14c is a bona fide binding site for Ctcf. Regarding RS14d, we consider one of three possibilities: (1) that RS14d is also directly bound by Ctcf in vivo, albeit at a much reduced level that could be identified by ChIP-qPCR (Figure 2D) but escaped detection by ChIP-seq (Figure 2E); (2) that RS14d indirectly interacts with Ctcf via other motifs and proteins; or (3) that RS14d’s proximity to RS14c resulted in its co-immunoprecipitation with the RS14c–Ctcf complex. Given EMSA results and the base difference at the conserved “C” in RS14d, one of the latter two explanations may be more likely. Likewise, the control flanking sites A and B, which do not have any obvious Ctcf motifs, also showed very slight enrichment for

**Figure 4** Targeted deletion of RS14. (A) Targeting scheme for the 2.3-kb RS14 deletion. The expanded region depicts the targeted area of the Xic. Blue rectangles, RS14. Black rectangles, Xist exons. WT, wild type. Spe, Bgl II, B, Btg I. B9, BtgI restriction site found only on 129 allele. HI, BamHI. C, Cla I. M, Mnl I. Black triangles, LoxP sites. Neo, neomycin-resistance gene. DTA, diphtheria toxin A. Δ2.3, RS14 knockout allele. ΔNeo, neomycin-resistance reporter excised with transient Cre expression. (B–D) Southern blot analysis of new RS14 alleles. Knockout alleles are indicated by “Δ2.3” or “ΔNeo” labels of the expected bands hybridizing to the relevant probes. (B) Wild-type male ES (J1), wild-type female ES (EL16), knockout female A9Δ2.3 and B5Δ2.3, and knockout male B7Δ2.3 and D7Δ2.3. SpeI-digested genomic DNA was hybridized with external probe 1 from A. (C) Same lines as in B. BtgI-digested genomic DNA was hybridized with external probe 2 from A. (D) ΔNeo lines produced from Cre-mediated excision of the NeoR cassette. Bgl II-digested genomic DNA was hybridized with probe 1 from A.
Ctcf binding in day 4 female ES cells, perhaps due to similar reasons. We conclude that Ctcf strongly occupies RS14c in vivo.

We next investigated if RS14 might possess enhancer blocking activity (Figure 3A). In this assay, a test sequence is inserted between the β-globin locus control region and a neomycin-resistance reporter (NeoR) to determine the degree to which the test element can decrease the number of NeoR colonies in a K562 background [a flanking insulator (INS) protects against position effects] (Chung et al. 1993; Bell et al. 1999). Whereas a test sequence containing λDNA or no insulator yielded plentiful colonies, two copies of the established β-globin insulator led to a more than fourfold reduction in colony number, as expected (Figure 3B). A 2.3-kb test fragment containing RS14 likewise resulted in a significant reduction of NeoR colonies when compared to λ or no insulator controls \( (P < 0.005) \), consistent with RS14 acting as a chromatin insulator. RS14 could block enhancer–promoter interactions in both forward and reverse orientations, as is the case with insulators in other contexts (Ohlsson et al. 2001; West et al. 2002). Taken together with the EMSA and ChIP experiments discussed above, these data demonstrate that RS14 binds Ctcf and exhibits classic enhancer blocking activity.

**Targeted deletion of RS14**

To investigate the function of RS14 during XCI in vivo, we generated a deletion of a 2.3-kb sequence corresponding to RS14 (Figure 4A). One concern that arises when creating Xist deletions in this context is the need to preserve the Ctcf binding in day 4 female ES cells, perhaps due to similar reasons. We conclude that Ctcf strongly occupies RS14c in vivo.

We next investigated if RS14 might possess enhancer blocking activity (Figure 3A). In this assay, a test sequence is inserted between the β-globin locus control region and a neomycin-resistance reporter (NeoR) to determine the degree to which the test element can decrease the number of NeoR colonies in a K562 background [a flanking insulator (INS) protects against position effects] (Chung et al. 1993; Bell et al. 1999). Whereas a test sequence containing λDNA or no insulator yielded plentiful colonies, two copies of the established β-globin insulator led to a more than fourfold reduction in colony number, as expected (Figure 3B). A 2.3-kb test fragment containing RS14 likewise resulted in a significant reduction of NeoR colonies when compared to λ or no insulator controls \( (P < 0.005) \), consistent with RS14 acting as a chromatin insulator. RS14 could block enhancer–promoter interactions in both forward and reverse orientations, as is the case with insulators in other contexts (Ohlsson et al. 2001; West et al. 2002). Taken together with the EMSA and ChIP experiments discussed above, these data demonstrate that RS14 binds Ctcf and exhibits classic enhancer blocking activity.

**Targeted deletion of RS14**

To investigate the function of RS14 during XCI in vivo, we generated a deletion of a 2.3-kb sequence corresponding to RS14 (Figure 4A). One concern that arises when creating Xist deletions in this context is the need to preserve the
silencing function of Xist. Two previous studies indicated that the 2.3-kb deletion would not affect Xist’s silencing function. In one study, Xist cDNA transgenes lacking the 3’ end of Xist and RS14 retained the RNA’s ability to bind the X, spread in cis, and establish XCI (Wutz et al. 2002). In a second study, a 65-kb deletion of the Xic (Δ65; see Figure 1A), which also removed the 2.3-kb sequence in addition to Tsix and Xite, resulted in a fully functional Xist RNA that

Figure 6 Aberrant XCI in female ES cells lacking RS14. (A) EBs derived from wild-type and mutant female cell lines. The mutant EB phenotype is similar to that of the transgenic Xite11 line (Lee 2005) (B) Anti-Nanog staining of wild-type and ΔRS14 (A9Δ2.3 clone shown) ES cells, either undifferentiated (day 0) or after 8 days of differentiation (day 8). (C) Cell death analysis of mutant female cell lines relative to the parental female EL16 line. The graph shows averages and standard errors of three independent differentiation experiments. Asterisks indicate samples that show a statistically significant difference (P = <0.05) when compared to wild-type ES cells at the same time point. (D) Sequential RNA/DNA FISH in wild type and ΔRS14 (A9Δ2.3 clone). Xist RNA FISH to visualize Xist RNA (red) is followed by cell denaturation and DNA FISH to detect the X chromosome (green). The ΔRS14 female lines contain a mixture of diploid and tetraploid cells. Arrows, low-level Xist expression. Asterisks, upregulated Xist clouds. A rare example of upregulated Xist in day 8 mutant female cells is shown at bottom right. (E) qRT-PCR of Tsix and Xist RNA levels in wild-type female and one representative ΔRS14 mutant on day 0 and day 8 of differentiation. RNA levels were normalized to tubulin α1A levels and the day 0 levels were set to 1.0.
The X chromosome of M. castaneus (cas) Xs confirmed that the deletion was targeted to the 129 chromosome in each case.

**Female-specific defects associated with ∆RS14**

To characterize the effect of ∆RS14, we examined the behavior of the mutant ES cells during differentiation between day 0 and day 10, a time window during which XCI proceeds to completion. Mutant male ES cells grew normally in the undifferentiated state (data not shown) and also differentiated normally into EBs upon removal of LIF (Figure 5A), with no statistically significant increases in cell death evident in any ∆RS14 male line (Figure 5B). To determine the state of Xist expression, we carried out sequential RNA/DNA FISH in which we first performed RNA FISH to visualize Xist RNA clusters followed by cell denaturation and X-chromosome painting (DNA FISH) to locate the X chromosome (Figure 5C). No ectopic Xist clusters were detectable in the ∆RS14 male cells at any time before or during cell differentiation (0%, n = 279–356 for each sample and time point). Quantitative RT-PCR (qRT-PCR) of Tsix and Xist levels revealed no differences between wild-type and mutant male ES cells before and after differentiation (Figure 5D). Therefore, the ∆RS14 allele does not recapitulate the Xist phenotype and the hypothesized counting defect in ∆65-kb XY and XO cells (Clerc and Avner 1998; Morey et al. 2004). We conclude that RS14 is dispensable in male ES cells and does not obviously serve as a counting element.

By contrast, heterozygous female knockout cells exhibited a number of anomalies. First, although they grew normally in the undifferentiated state, the EBs grew poorly when placed under differentiation conditions (Figure 6A). While the parental female ES line produced large EBs with abundant outgrowth and a variety of differentiated cell types at day 8, the mutant female lines yielded very sparse outgrowth after the EBs were adhered onto gelatinized plates on day 4. Very few mutant cells migrated outward from the embryoid body cluster, and overall colony growth resembled that of undifferentiated ES cells. This phenotype was reminiscent of those seen in cell lines defective for XCI choice or pairing (Stavropoulos et al. 2001; Lee 2005; Xu et al. 2006) (Figure 6A, Xite transgenic line shown as a representative line), and the outgrowth phenotype did not appear to be caused by a general differentiation defect, since both wild-type and mutant cells lost expression of the pluripotency marker NANOG after 8 days (Figure 6B).

heterozygous knockout female lines (A9Δ2.3, B5Δ2.3) and their Neo− derivatives (A9D7ΔNeo, B5D3ΔNeo) showed relatively increased cell death between day 4 and day 8, compared to that observed between day 2 and day 3 (Figure 6C). However, we note that the apparent increase in cell death was very modest and not uniformly statistically significant, suggesting that cells may arrest in growth rather than die in large numbers. These results suggested that ∆RS14 female cells do not differentiate well and that they either arrest in growth or die during differentiation.

To determine whether the defect in growth and viability stemmed from XCI anomalies, we examined Xist RNA expression in the ES cells before and during cell differentiation. In the undifferentiated state (day 0), mutant female ES cells exhibited low-level Xist expression, similar to that typically observed in wild-type cells (Figure 6D). When placed in differentiation conditions, few mutant colonies outgrew differentiated cells. We then examined Xist expression by serial RNA/DNA FISH and observed that, whereas 29.5% (n = 332) of female cells displayed Xist clouds by day 4, only 6.2% (n = 276) of A9Δ2.3 female cells did (Figure 6D; asterisks point to Xist clouds). Similarly on day 10, 83.0% (n = 271) of wild-type cells exhibited a Xist cloud, whereas only 5.4% (n = 480) of mutant cells did. The majority of mutant female cells retained low-level Xist expression similar to that seen in the undifferentiated state (Figure 6D, arrows). qRT-PCR confirmed compromised XCI, as Tsix levels remained relatively high and Xist levels were aberrantly reduced on day 8 in ∆RS14 cells (Figure 6E). Taken together, these data showed that ∆RS14 precludes Xist upregulation in female cells and that this deficiency may be responsible for their poor differentiation and cell death or growth arrest.

These results indicated that some mutant cells could upregulate Xist (5–6%) but the vast majority could not. Because RS14 was deleted from only one allele, we suspected that this difference may stem from cis-effects of ∆RS14. If so, we would expect to observe allelic skewing of Xist expression. To test this idea, we performed allele-specific RT-PCR to examine the origin of Xist expression. The wild-type parental female cell line is a hybrid ES line that carries X chromosomes of Mus musculus musculus origin (X129, strain 129) and M. m. castaneus origin (Xcas) (Lee and Lu 1999). Because the X chromosomes carry different Xce modifier alleles (Cattanach and Isaacs 1967), XCI in the parental ES line is ordinarily biased toward inactivating X129 by a 70:30 ratio (Figure 7, A and B). Analysis of allelic choice in the ∆RS14 lines revealed a significant deviation of Xist ratios from the normal pattern. Indeed, quantitation of Xist allelic ratios in A9Δ2.3, B5Δ2.3, and their Neo−derivatives indicated that the bias was the reverse of what was expected in wild-type cells: Xcas was preferentially chosen for inactivation (Figure 7, A and B). Allele-specific RNA/DNA FISH analysis using a Neo probe to discern mutant A9Δ2.3(Neo−) from wild-type cells showed that Xist was
upregulated exclusively from Xcas in 100% of cells, thereby confirming skewed allelic ratios (Figure 7C). These data indicated that linkage to the ΔRS14 mutation compromised the Xist allele in cis. Thus, we conclude that RS14 is required in cis for the proper activation of Xist and that its deficiency precludes cell differentiation and XCI in cis.

Figure 7 Allele-specific analysis of Xist expression in mutant female cells. (A) Allele-specific RT-PCR analysis of Xist expression (ScrFI polymorphism) on days 0 and 8 of differentiation. Bands represent Xist RNA from the 129 and the castaneus allele (cas), as indicated. One representative experiment is shown. (B) Quantitation of allelic ratios of Xist RNA on day 8 of differentiation. The open circles on the left of each set represent quantitation of three to four independent allele-specific RT-PCR experiments. The means and standard errors are shown to the right by the black circles and bars, respectively. *P*, significance of the difference in pairwise comparisons between wild-type EB and the indicated mutant EB, as calculated by a two-tailed unpaired Student’s t-test. (C) Serial RNA/DNA FISH analysis of wild-type and mutant (A9Δ2.3, Neo+) female cells. Xist RNA is performed first and is followed by DNA FISH using Xic and Neo probes to distinguish between wild-type and mutant alleles. Four types of hybridization patterns are shown for mutant cells. Note that Xist RNA is always upregulated from the wild-type chromosome (Xist RNA and Neo probe signals are not coincident).
Discussion

In this study, we identify a conserved genetic element, RS14, at the junction of Tsix and Xist. RS14 binds Ctf protein and demonstrates enhancer-blocking activity, correlating with the chromatin transition zone previously shown to separate chromatin hubs centered at Tsix and Xist (Tsai et al. 2008). We propose that RS14 is necessary to achieve proper transcriptional induction of the Xist allele in cis. It is clear that the RS14 sequence is not required for Xist RNAs silencing function, as a 65-kb deletion including RS14 (shown in Figure 1) does not affect silencing in cis (Clerc and Avner 1998; Morey et al. 2004). Therefore, the disruption of Xist function in cis cannot be attributed to defects in the silencing step per se. Our results suggest that RS14 regulates a step upstream of silencing—probably not the counting step, since ΔRS14 male cells properly block XCI. Given that deleting RS14 leads to a measurable effect on allelic choice, we believe that ΔRS14 affects the ability of the female cell to choose the active Xist allele.

Effects on allelic choice may have two causes (primary vs. secondary). In primary nonrandom XCI, the allelicskewing would result from a primary mutation in the choice apparatus, which would preclude the selection of the mutated X (X129 in this case) as Xi. In this situation, Xist upregulation would occur on Xcas, and differentiating ES cells would survive and differentiate into normal-looking EBs. This is clearly not what we observed. Examples of a primary mechanism include deletions of the Tsix (Lee and Lu 1999) and Xist (Marahrens et al. 1998) promoters. In secondary nonrandom XCI, allelic skewing of Xist expression would result from selection against cells that have chosen an unfavorable allele (Stavropoulos et al. 2001). In the case of ΔRS14 cells, those selecting X129 would either die or arrest in growth and only those that select Xcas would continue to divide. Because of slightly elevated cell death and the failure of EBs to outgrow cells, we believe that ΔRS14 female cells exhibit secondary nonrandom XCI (Figure 8). In this scenario, RS14 +/− cells continue to choose either X129 or Xcas for inactivation at the expected 70:30 ratio, but choosing X129 (XΔRS14) would lead to growth retardation, arrest, or death due to compromised Xist induction and failure of XCI in cis. Only those cells selecting wild-type Xcas as Xi would be viable and continue to divide and differentiate.

Why does ΔRS14 disrupt Xist upregulation in cis? We propose that the association of Ctf with RS14 provides insulation between chromatin hubs centered at Tsix and Xist. Since Tsix dominates over and silences Xist in ES cells, deleting RS14 might upset the relative transcriptional balance in favor of Tsix to persist during cell differentiation and decrease the probability with which the linked Xist allele would be turned on, which is consistent with our analysis (Figure 6). In female cells, loss of enhancer-blocking activity within RS14 would then result in persistent Tsix expression and preclude Xist activation. In male cells, there would be no obvious consequence, as Xist is normally not activated.

We speculate that RS14 may organize chromatin architecture at the Xic. Apart from ΔRS14, Ctf also binds to enhancer elements within Xite, to the S′ end of Tsix around the DXPas34 repeat, and to the Xist promoter (Chao et al. 2002; Pugacheva et al. 2005; Xu et al. 2007). Interestingly, all of these Ctf-binding sites occur at the bases of known looping interactions (Tsai et al. 2008). It is known that Ctf can homodimerize to form large multimers that knit together many chromatin loops (Yusufzai et al. 2004). Therefore, dimerization between Ctf proteins bound at Xist, Tsix, Xite, and RS14 could give rise to the looping structures and discrete chromatin hubs at the Xic. Ctf is also involved in formation of interchromosomal bridges that underlie pairing interactions observed at the onset of XCI (Xu et al. 2007). Because the switch from trans to cis regulation is a key
RS14 Is Required for Xist Activation


Loots, G. G., I. Ovcharenko, L. Pachter, I. Dubchak, and E. M. Rubin, 2002 rVista for comparative sequence-based discovery of func-

Acknowledgments

We are grateful to all lab members for discussion and support. R.I.S. is supported by the Medical Scientist Training Program at Harvard Medical School. S.F.P. is supported by a Deutscher Forschungsgemeinschaft research fellowship. This work is funded by a grant from the National Institutes of Health (ROI-GM58839). J.T.L. is an Investigator of the Howard Hughes Medical Institute.

Literature Cited


Ohlsson, R., R. Renkawitz, and V. V. Lobanenkov, 2001 CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet. 7: 520–527.


Communicating editor: J. A. Birchler

454 R. J. Spencer et al.
A Boundary Element Between *Tsix* and *Xist* Binds the Chromatin Insulator Ctcf and Contributes to Initiation of X-Chromosome Inactivation

Rebecca J. Spencer, Brian C. del Rosario, Stefan F. Pinter, Derek Lessing, Ruslan I. Sadreyev, and Jeannie T. Lee
Tables S1 and S2

Supporting Data

Tables S1: Significant hits to the RS14c motif in the ChIP-seq dataset of Chen et al., 2008

Table S2: Significant hits to the RS14c motif in the ChIP-seq dataset of Nitzche et al., 2011

Available for download as text files at http://www.genetics.org/content/early/2011/08/10/genetics.111.132662/suppl/DC1.