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[1]

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

doi: 10.1534/genetics.111.132662
Manuscript received December 30, 2010; accepted for publication July 15, 2011
Supporting information is available online at http://www.genetics.org/content/suppl/2011/08/12/genetics.111.132662.DC1.
[1]Corresponding author: Massachusetts General Hospital/Harvard Medical School, 50 Blossom St., Boston, MA 0214. E-mail: lee@molbio.mgh.harvard.edu

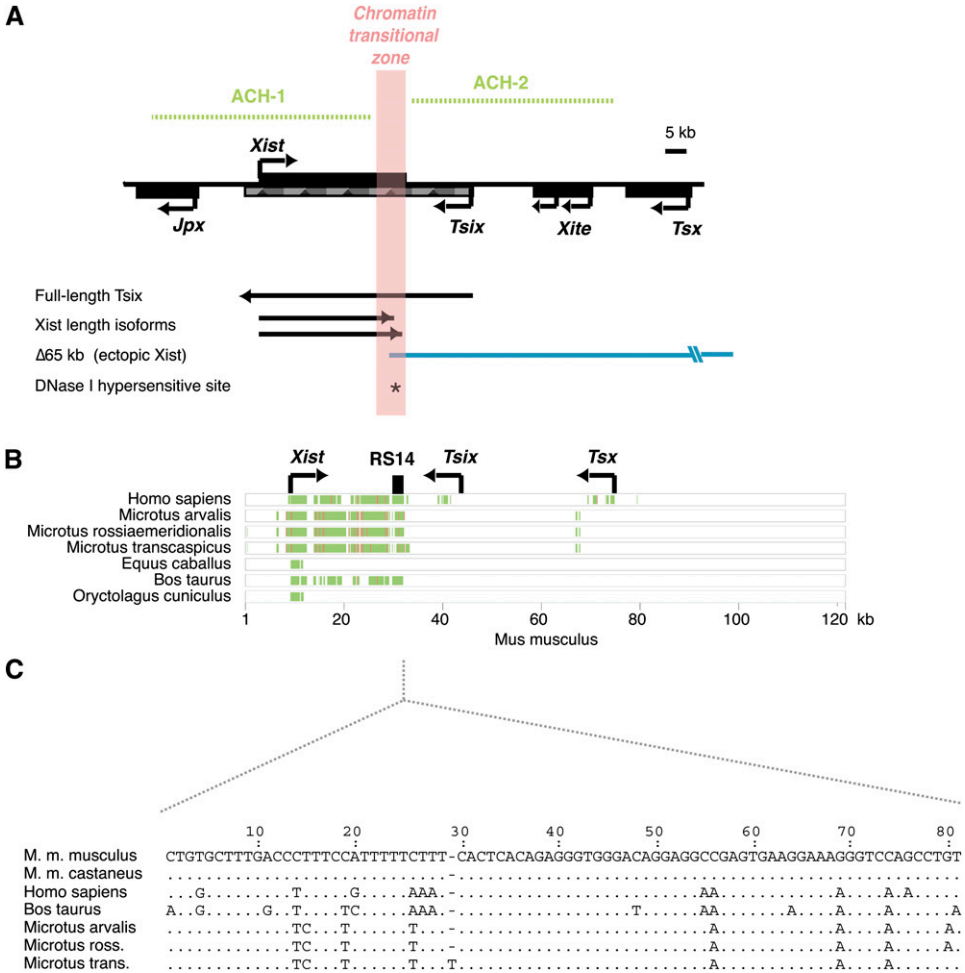


Figure 1 *RS14*, a conserved element at the *Xist–Tsix* junction. (A) Map of the *Xic* region showing its noncoding genes and other features of interest. ACH, active chromatin hub. (B) PipMaker figure showing areas of general cross-species conservation within the *Xic*. Green, general regions of conservation. Red, regions of high conservation: at least 70% nucleotide identity >100 bp, without gaps. *RS14* is an ~2-kb element that lies at the junction between *Xist* and *Tsix* (black box). (C) Alignment of *RS14d*, an exceptionally well-conserved motif within *RS14*. *M. m. musculus* sequence corresponds to chrX: bp 100,656,551–100,656,631 (UCSC Genome Browser).

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 tRNA gene in yeast (Oki and Kamakaka 2005), Ctcf-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 Corces 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 transcaspicus*), 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 *Cla*I–*Bam*HI 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: 222,426,453–222,428,947), *Canus familiaris* (chrX: 60,374,176–60,376,292), *Pan troglodytes* (chrX: 73,151,334–73,153,817), *Felis catus* (scaffold_203485:24-1902), and *Macaca mulatta* (chrX: 72,942,832–72,945,325). 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′-CAGTTGTGGGGTTTATACGCGGGAGTTGCCGCGTGGTGGCAGCAAAATCGATTGCGCCAAACCTAAAGAG-3′ (MS1) (Hark *et al.* 2000); H19-reverse, 5′-CTCTTTAGGTTTGGCGCAATCGATTTTGCTGCCACCACGCGGCAACTCCCGCGTATAAACCCCACAACTG-3′; RS14c, 5′-TGTTATACCCGTGTAGGCCAGCAGAGGGTGTCGGATCCCAAGGAAA-3′; RS14c-reverse, 5′-TTTCCTTGGGATCCGACACCCTCTGCTGGCCTACACGGGTATAACA-3′; RS14c-mut, 5′-TGTTATACCCGTGTAGGCCAGCAAAAAATGTCGGATCCCAAGGAAA-3′; RS14c-mut-reverse, 5′-TTTCCTTGGGATCCGACATTTTTTGCTGGCCTACACGGGTATAACA-3′; RS14d, 5′-CTGTGCTTTGACCCTTTCCATTTTTCTTTCACTCACAGAGGGTGGGACAGGAGGCCGAGTGAAGGAAAGGGTCCAGCCTG-3′; and RS14d-reverse, 5-CAGGCTGGACCCTTTCCTTCACTCGGCCTCCTGTCCCACCCTCTGTGAGTGAAAGAAAAATGGAAAGGGTCAAAGCACAG. The complementary single-stranded oligonucleotides were suspended together in duplex buffer from Integrated DNA Technologies (RS14d) or PCR buffer, heated to 90° for 2 min, and cooled to room temperature to form double-stranded oligonucleotides. Double-stranded oligonucleotides were labeled using phosphonucleotide kinase and γ^{32}P-ATP. For gel-shift analyses, full-length purified recombinant Ctcf protein (Novus Biologicals catalog no. H00010664-P01, amino acids 1–728) and 0.2 pmol ^{32}P-labeled oligonucleotides were incubated at room temperature for 20 min in 20 mM HEPES (pH 7.5), 50 mM KCl, 5 mM $MgCl_2$, 3.3 μM $ZnSO_4$, 1 mM dithiothreitol, 0.3 mg/ml BSA, 0.5 μg poly(dI:dC), 5% glycerol, and 0.5% triton X-100. Cold competitors were added at 1000× molar excess. Shifts were resolved in 5% acrylamide/0.5× TBE gels at 4° which were fixed in 50:10:40 methanol:acetic acid:water and dried.

Chromatin immunoprecipitations

Chromatin immunoprecipitation (ChIP) analysis was carried out as described (Lee *et al.* 2006) on each of day 0 and day 4 male and female ES cells, 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, ACTCATCAGTGGATACGTTTGAGTGCC C; p135, CTCCAAGAAGCCCATGAAACCTTACC; p136, GCAAATCCTTGTAGAACAGGCGAACC; p137, CCCACAGTTTGCCTCAGAGGTTGAATAC; p139, TAGGTTTGGGTGTTATACCCGTGTAGGC; p14, CTTAGATTCCAGATAGACAGGCTGGACC; p141, AATGCCTGTGCTTTGACCCTTTCC; and p142, GGCTACACACAAGATGGCGTCTGTAACTTG.

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 *Cla*I–

*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 *AscI* site of the no-insulator plasmid in the forward and reverse directions. Constructs were linearized by *Sal*I 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–hygromycin (Chao *et al.* 2002), was linearized with *Bgl*I and prepared in the same 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–hygromycin 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 ketoconazole, and either 1 mg/ml of neomycin (G418) or 200 μg/ml hygromycin. 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 *Sac*II site and *Sal*I–*Hin*dIII 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-DTAbpA; negative selection) expression cassette in a pBluescript II KS+ backbone. *Pvu*I 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 *Spe*I 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 *Btg*I 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 *Blp*I 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 the 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 Volocity 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 Volocity. 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 *Bts*I–*Xcm*I 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.

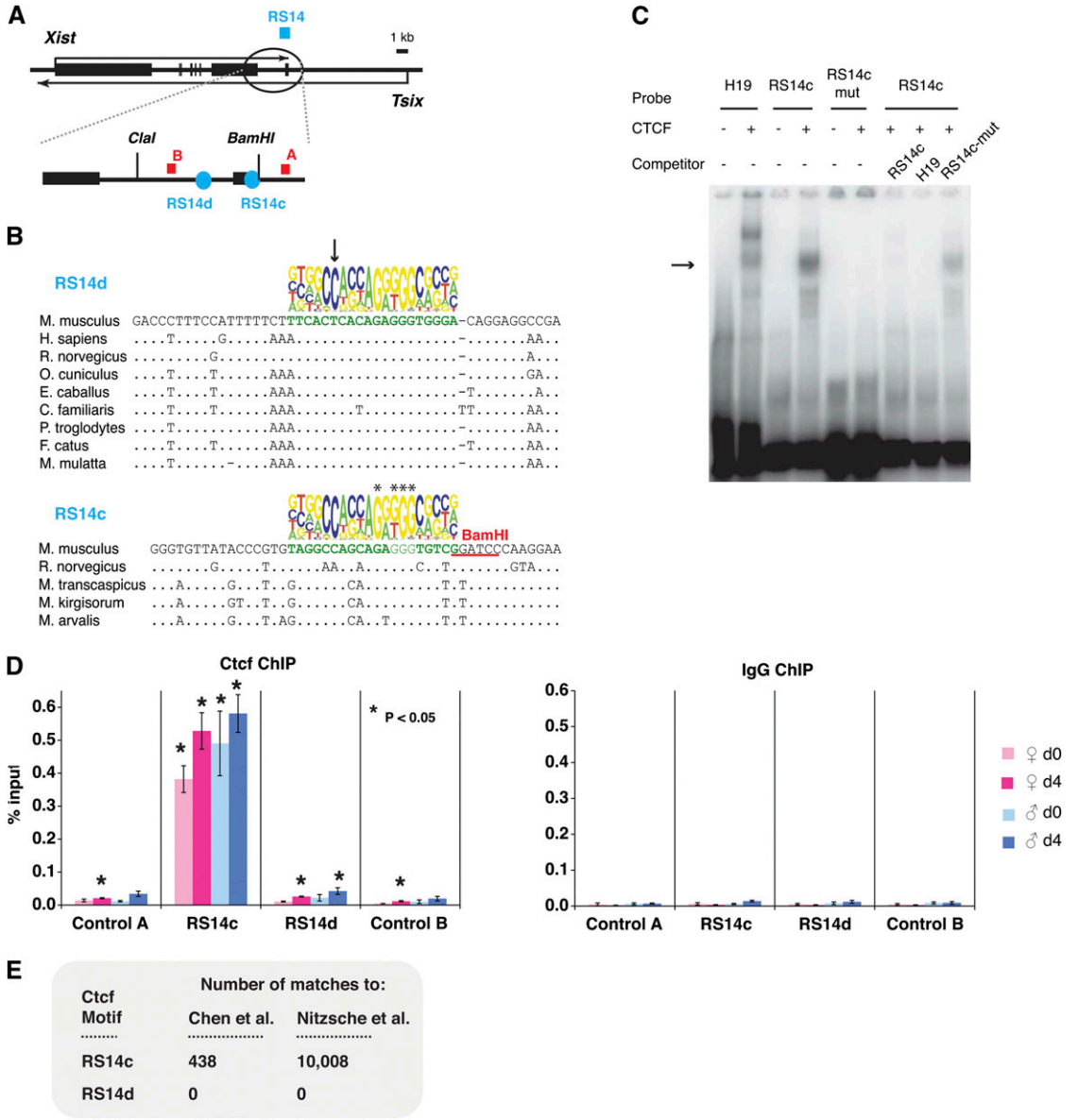
Immunofluorescence was performed as previously described (Zhang *et al.* 2007). Undifferentiated ES cells and day 8 cells from embryoid bodies (EBs) were fixed and incubated with rabbit anti-Nanog antibody (Novus) at 1:750 dilution and visualized with goat anti-rabbit-alexa555 (Invitrogen) at 1:1000 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′-TAAACAGCCAGAGTCTGATGTAACGG-3′), Tsix (NS19) (Stavropoulos *et al.* 2001), and tubulin α1A (YJ2, 5′-CTTGCCAGCTCCTGTCTCAC-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′-CTCGCCTCCGCCA TCCACCC-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′-



E

Ctcf Motif	Number of matches to: Chen et al.	Nitzsche et al.
RS14c	438	10,008
RS14d	0	0

Figure 2 *RS14* binds Ctcf *in vitro* and *in vivo*. (A) The RS14 region contains two putative Ctcf-binding sites within a *Cla*I–*Bam*HI restriction fragment that harbors *RS14d* and *RS14c* (locations shown by blue circles). Control sites for Ctcf ChIP are designated "A" and "B." (B) Multispecies alignment of the consensus Ctcf motif (Kim *et al.* 2007) at *RS14d* and *RS14c*. The *RS14d M. musculus* sequence corresponds to chrX: bp 100,656,561–100,656,608 (UCSC Genome Browser); *RS14c M. musculus* sequence corresponds to chrX: bp 100,655,589–100,655,636. Control sites A and B lack obvious Ctcf motifs and are 1.2 kb from *RS14c* and 0.8 kb from *RS14d*, respectively. The *Bam*HI site next to *RS14c* is underlined in red. Asterisks mark mutated bases in RS14c-mut. Arrow marks the highly conserved C nucleotide of the Ctcf-binding consensus. (C) EMSA shows that RS14c binds Ctcf protein and depends upon four core G nucleotides for binding. Radioactively labeled probe RS14c was shifted with purified Ctcf protein with or without unlabeled competitor DNA as indicated. Arrow, specific protein–DNA complex. (D) Ctcf and control IgG ChIPs at RS14 and flanking positions in day 0 and day 4 male and female ES cells. Statistical significance ($P < 0.05$) is determined by a paired two-tailed Student's *t*-test comparing Ctcf ChIP to the corresponding IgG ChIP samples. Statistically significant localizations are marked by an asterisk.

CAGAGTAGCGAGGACTTGAAGAG-3′) and NS66 (5′-GCTG GTTCGTCTATCTTGTGGG-3′) and 32 cycles of PCRwere used to amplify polymorphic *Mus musculus castaneus* and 129 fragments which were digested with *Scrf*I. These fragments were fractionated on an agarose gel, blotted, and probed with the ^{32}P-end-labeled probe NS67 (5′-CCA GAGTCTGATGTAACGGAGG-3′). Relative allelic amounts were quantitated by phorphorimage 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 6×10^6 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

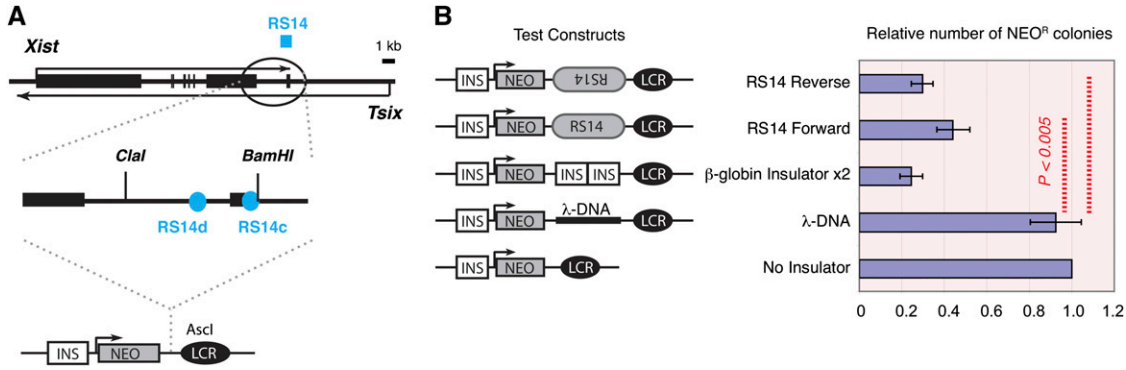


Figure 3 *RS14* contains enhancer-blocking activity. (A) The schematic shows the RS14 test region and the vector construct (below) used in the enhancer-blocking assay. INS, β-globin insulator. NEO, neomycin resistance gene. LCR, locus control region (enhancer). (B) (Left) The enhancer-blocking assay using the constructs. (Right) Graph showing the relative number of neomycin-resistant (NEO^R) colonies, normalized to hygromycin-resistant colonies (the transfection efficiency control) and presented relative to the number of colonies from the "no insulator" construct (standardized to 1.0). The average and standard errors are shown for 10–11 experiments. A 2.3-kb *RS14* fragment produced a statistically significant reduction in Neo^R colony number in either orientation when compared to λ or no insulator controls ($P < 0.005$), as determined by unpaired two-tailed Student's *t*-tests in pairwise comparisons.

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. cuniculus* (rabbit), *B. taurus* (cow), *E. caballus* (horse), *M. transcaspicus* (vole), *M. rossiaemeridionalis* (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 β-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 β-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

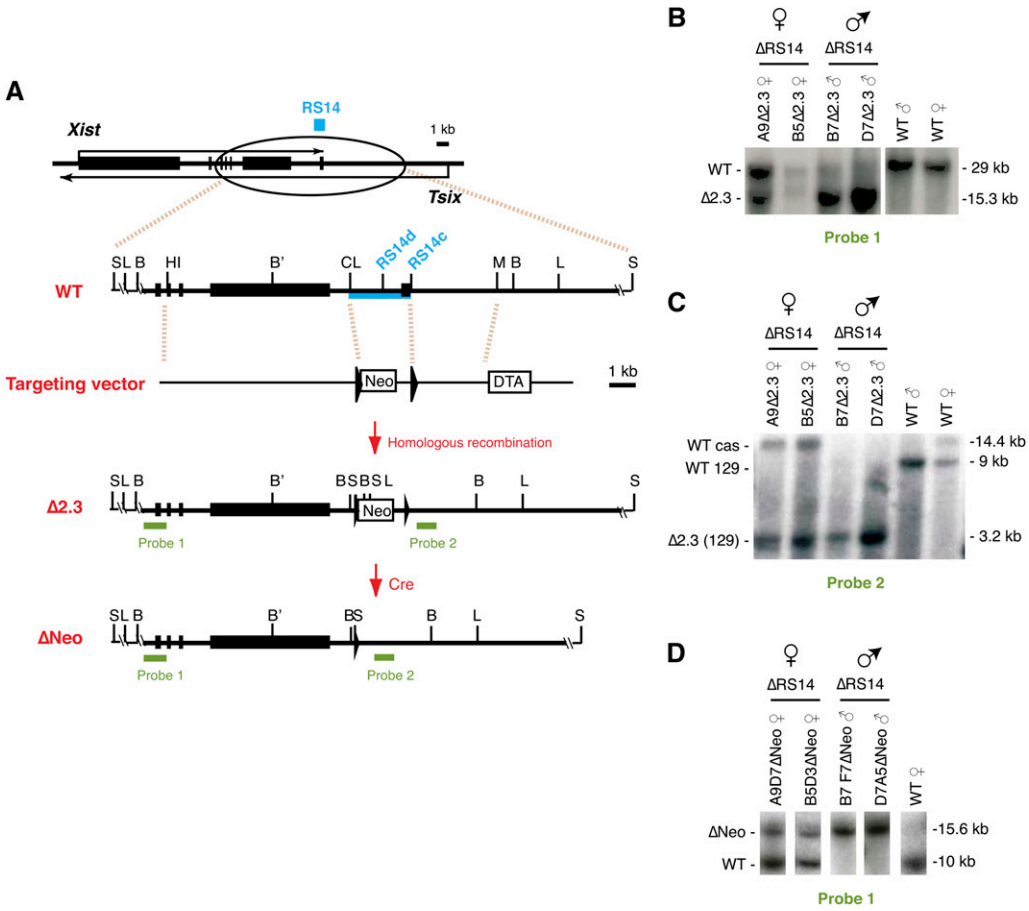


Figure 4 Targeted deletion of *RS14*. (A) Targeting scheme for the 2.3-kb *RS14* deletion. The expanded region depicts the targeted area of the *Xic*. Blue rectangles, *RS14*. Black rectangles, *Xist* exons. WT, wild type. S, *Spe*I. L, *Blp*I. B, *Btg*I. B′, *Btg*I restriction site found only on 129 allele. HI, *Bam*HI. 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. *Spe*I-digested genomic DNA was hybridized with external probe 1 from A. (C) Same lines as in B. *Btg*I-digested genomic DNA was hybridized with external probe 2 from A. (D) ΔNeo lines produced from Cre-mediated excision of the Neo^R cassette. *Blp*I-digested genomic DNA was hybridized with probe 1 from 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

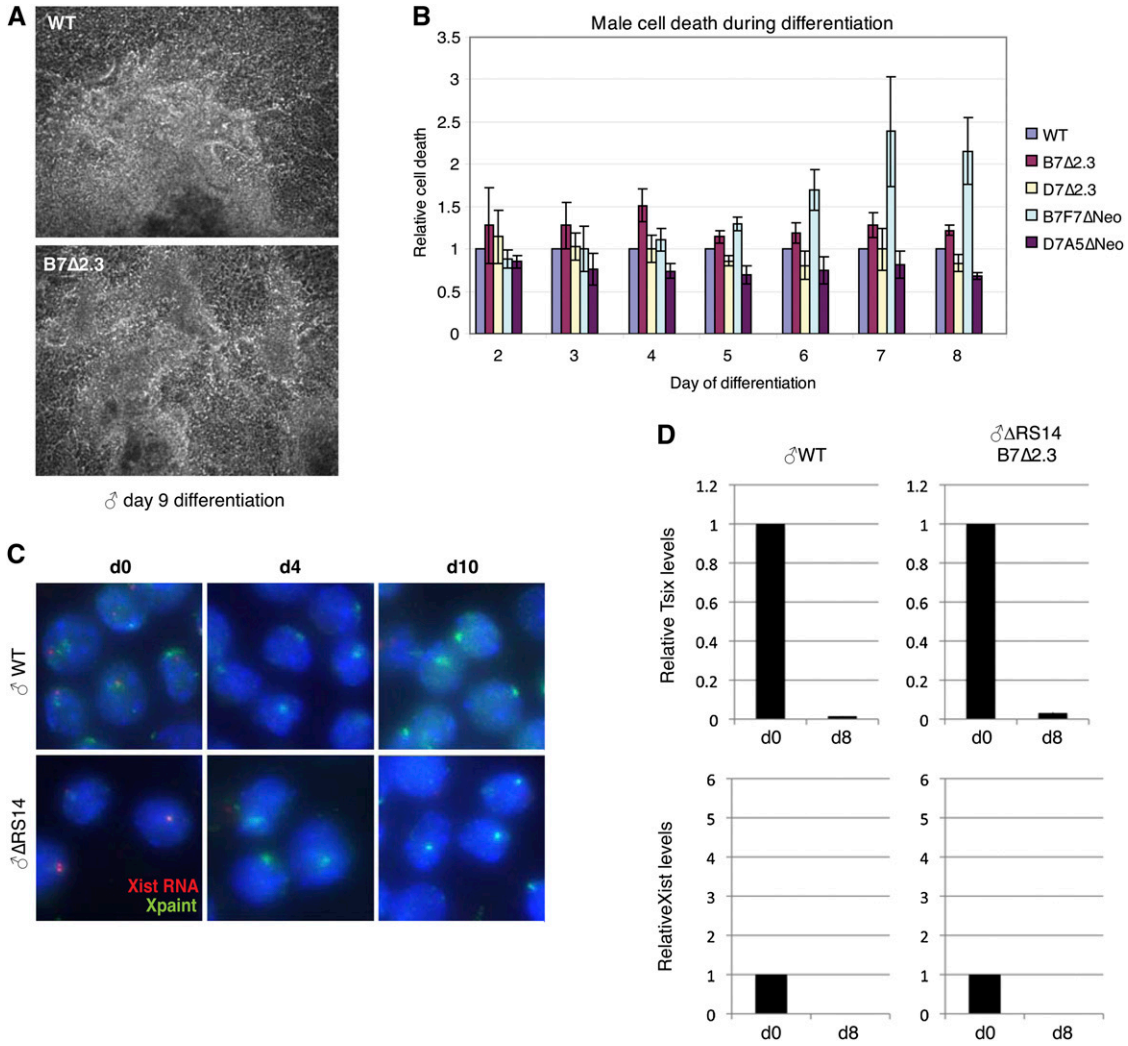


Figure 5 Δ*RS14* knockout male cells are normal. (A) EBs derived from knockout male ES cells (representative line shown: B7Δ2.3) are normal at all time points of cell differentiation. Day 9 images are shown. (B) Cell death relative to knockout *vs.* wild-type male ES cells. The graph shows averages and standard errors of three independent differentiation experiments. (C) Sequential RNA/DNA FISH in wild type and Δ*RS14* (B7Δ2.3 clone shown) on day 0, day 4, and day 10 of differentiation. Xist RNA FISH to visualize Xist RNA (red) is followed by cell denaturation and DNA FISH to detect the X chromosome (green). Sample size (*n*) is 279–356 for each sample and time point. Zero percent showed Xist clouds in both wild type and mutant. (D) qRT-PCR of Tsix and Xist RNA levels in wild-type male and one representative Δ*RS14* 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.

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 (Neo^R) to determine the degree to which the test element can decrease the number of Neo^R 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 either 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 Neo^R 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

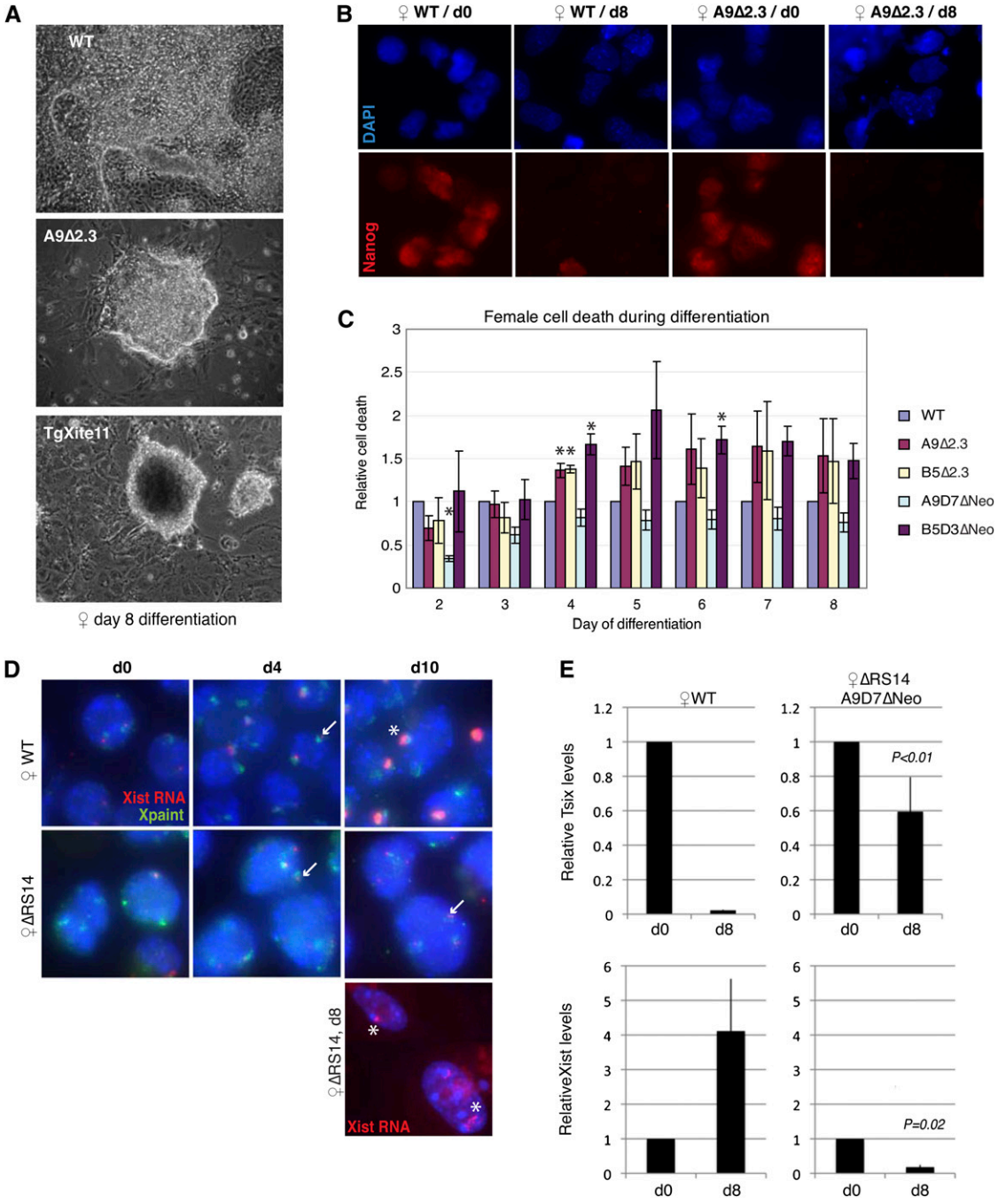


Figure 6 Aberrant XCI in female ES cells lacking *RS14*. (A) EBs derived from wild-type and mutant female cell lines. The mutant EB phenotype is similar to that of the transgenic Xite11 line (Lee 2005) (B) Anti-Nanog staining of wild-type and Δ*RS14* (A9Δ2.3 clone shown) ES cells, either undifferentiated (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 Δ*RS14* (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 Δ*RS14* 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 Δ*RS14* 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.

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

constitutively silenced the X in *cis* (Clerc and Avner 1998; Morey *et al.* 2004). Here, we matched the 5′ border of Δ2.3 to Δ65 to preserve Xist RNA function. Among the ~2500 XY and 1700 XX clones picked, two correctly targeted clones (Neo^+) of each sex were identified (XY, B7Δ2.3 and D7Δ2.3; XX, A9Δ2.3 and B5Δ2.3) and verified in Southern blot analysis using 5′ and 3′ external probes (Figure 4, B and C). Two Neo^- clones were subsequently derived for each (Figure 4D). A *Btg*I polymorphism (B′) present between the *M. musculus* (129) and *M. castaneus* (cas) Xs confirmed that the deletion was targeted to the 129 chromosome in each case.

Female-specific defects associated with ΔRS14

To characterize the effect of Δ*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 Δ*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 Δ*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 Δ*RS14* allele does not recapitulate the *Xist* phenotype and the hypothesized counting defect in Δ65-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). Both 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 Δ*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 ΔRS14 cells (Figure 6E). Taken together, these data showed that Δ*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 Δ*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 (X^{129}, strain 129) and *M. m. castaneus* origin (X^{cas}) (Lee and Lu 1999). Because the X chromosomes carry different *Xce* modifier alleles (Cattanach and Isaacson 1967), XCI in the parental ES line is ordinarily biased toward inactivating X^{129} by a 70:30 ratio (Figure 7, A and B). Analysis of allelic choice in the Δ*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: X^{cas} 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

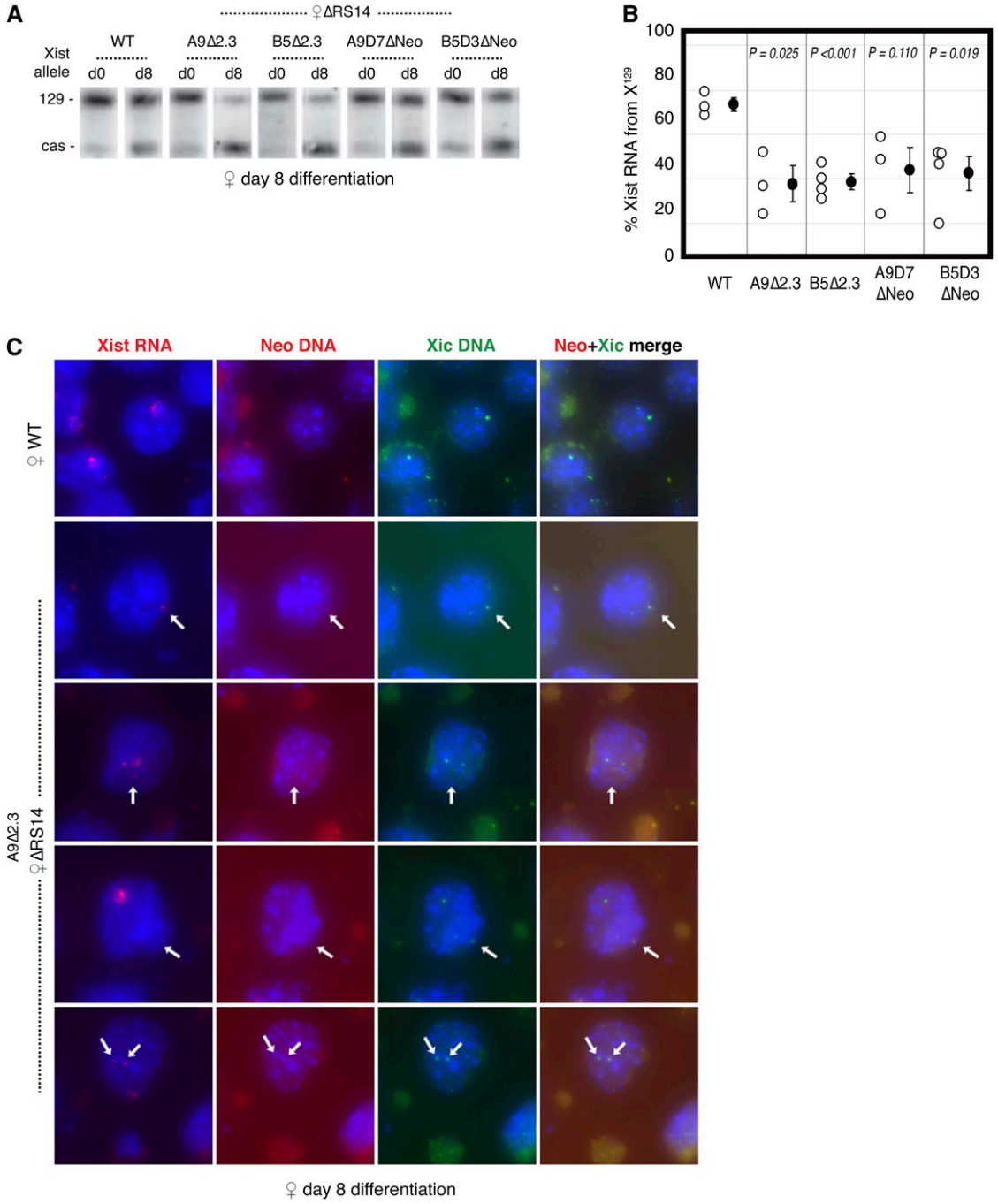


Figure 7 Allele-specific analysis of Xist expression in mutant female cells. (A) Allele-specific RT-PCR analysis of *Xist* expression (ScrFI polymorphism) on days 0 and 8 of differentiation. Bands represent Xist RNA from the 129 and the *castaneus* allele (cas), as indicated. One representative experiment 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).

upregulated exclusively from X^{cas} in 100% of cells, thereby confirming skewed allelic ratios (Figure 7C). These data indicated that linkage to the Δ*RS14* 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*.

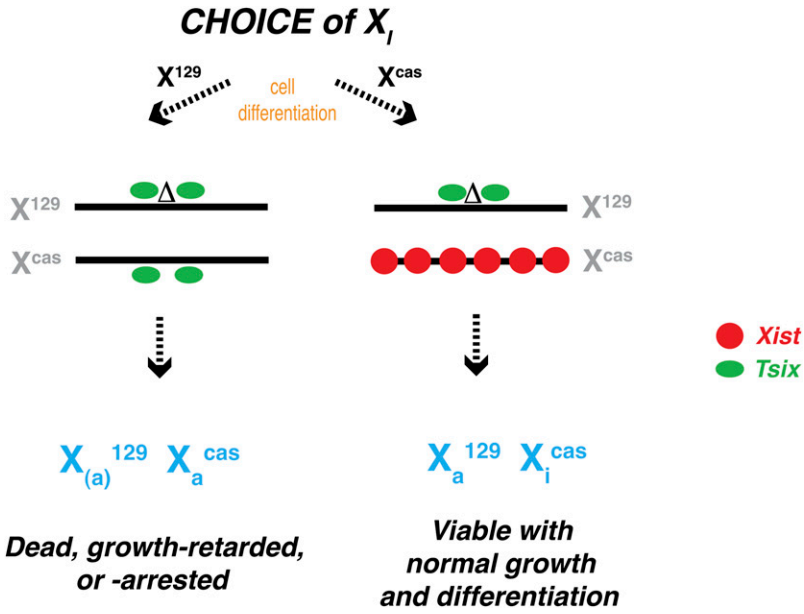


Figure 8 Model of Δ*RS14*'s effect on *Xist* allelic choice. The data are most consistent with skewed XCI being caused by a secondary mechanism caused by selection against cells that have made an unfavorable allelic choice, rather than a primary effect on the choice apparatus itself. Δ*RS14* female cells selecting X^{129} would either die or arrest in growth, and only those that select X^{cas} would continue to divide. This model supposes that *RS14*+/− cells retain the ability to choose either X^{129} or X^{cas} for inactivation (at the expected 70:30 ratio), but choosing X^{129} ($X^{\Delta RS14}$) 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 X^{cas} as Xi would be viable and continue to divide and differentiate.

Discussion

In this study, we identify a conserved genetic element, *RS14*, at the junction of *Tsix* and *Xist*. *RS14* binds Ctcf 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 RNA's 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 Δ*RS14* 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 allelic skewing would result from a primary mutation in the choice apparatus, which would preclude the selection of the mutated X (X^{129} in this case) as Xi. In this situation, Xist upregulation would occur on X^{cas}, 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 Δ*RS14* cells, those selecting X^{129} would either die or arrest in growth and only those that select X^{cas} would continue to divide. Because of slightly elevated cell death and the failure of EBs to outgrow cells, we believe that Δ*RS14* female cells exhibit secondary nonrandom XCI (Figure 8). In this scenario, *RS14* +/− cells continue to choose either X^{129} or X^{cas} for inactivation at the expected 70:30 ratio, but choosing X^{129}—the X carrying Δ*RS14*–would result in growth retardation, arrest, or death due to compromised *Xist* induction and failure of XCI *in cis*. Thus, only those cells selecting wild-type X^{cas} as Xi would be viable and continue to divide and differentiate.

Why does Δ*RS14* disrupt *Xist* upregulation in *cis*? We propose that the association of Ctcf 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 Δ*RS14*, Ctcf also binds to enhancer elements within *Xite*, to the 5′ 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 Ctcf-binding sites occur at the bases of known looping interactions (Tsai *et al.* 2008). It is known that Ctcf can homodimerize to form large multimers that knit together many chromatin loops (Yusufzai *et al.* 2004). Therefore, dimerization between Ctcf proteins bound at *Xist*, *Tsix*, *Xite*, and *RS14* could give rise to the looping structures and discrete chromatin hubs at the *Xic*. Ctcf 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

feature at the initiation of XCI, future research will focus on how *RS14* affects interactions in *cis* and in *trans* at the *Xic*.

Acknowledgments

We are grateful to all lab members for discussion and support. R.J.S. is supported by the Medical Scientist Training Program at Harvard Medical School. S.F.P. is supported by a Deutche Forschungsgemeinschaft research fellowship. This work is funded by a grant from the National Institutes of Health (RO1-GM58839). J.T.L. is an Investigator of the Howard Hughes Medical Institute.

Literature Cited

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

Bacher, C. P., M. Guggiari, B. Brors, S. Augui, P. Clerc *et al.*, 2006 Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat. Cell Biol. 8: 293–299.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant *et al.*, 2009 MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37: W202–W208.

Bell, A. C., and G. Felsenfeld, 2000 Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 405: 482–485.

Bell, A. C., A. G. West, and G. Felsenfeld, 1999 The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98: 387–396.

Borsani, G., R. Tonlorenzi, M. C. Simmler, L. Dandolo, D. Arnaud *et al.*, 1991 Characterization of a murine gene expressed from the inactive X chromosome. Nature 351: 325–329.

Brockdorff, N., A. Ashworth, G. F. Kay, P. Cooper, S. Smith *et al.*, 1991 Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature 351: 329–331.

Brockdorff, N., A. Ashworth, G. F. Kay, V. M. McCabe, D. P. Norris *et al.*, 1992 The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71: 515–526.

Brown, C. J., A. Ballabio, J. L. Rupert, R. G. Lafreniere, M. Grompe *et al.*, 1991a A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349: 38–44.

Brown, C. J., R. G. Lafreniere, V. E. Powers, G. Sebastio, A. Ballabio *et al.*, 1991b Localization of the X inactivation centre on the human X chromosome in Xq13. Nature 349: 82–84.

Cattanach, B. M., and J. H. Isaacson, 1967 Controlling elements in the mouse X chromosome. Genetics 57: 331–346.

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

Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss *et al.*, 2008 Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133: 1106–1117.

Chung, J. H., M. Whiteley, and G. Felsenfeld, 1993 A 5′ element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell 74: 505–514.

Chureau, C., M. Prissette, A. Bourdet, V. Barbe, L. Cattolico *et al.*, 2002 Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res. 12: 894–908.

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

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

Flint, J., C. Tufarelli, J. Peden, K. Clark, R. J. Daniels *et al.*, 2001 Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum. Mol. Genet. 10: 371–382.

Grant, C. E., T. L. Bailey, and W. S. Noble, 2011 FIMO: scanning for occurrences of a given motif. Bioinformatics 27: 1017–1018.

Hark, A. T., C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse *et al.*, 2000 CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405: 486–489.

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

Herzing, L. B., J. T. Romer, J. M. Horn, and A. Ashworth, 1997 Xist has properties of the X-chromosome inactivation centre. Nature 386: 272–275.

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

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

Kanduri, C., G. Fitzpatrick, R. Mukhopadhyay, M. Kanduri, V. Lobanenkov *et al.*, 2002 A differentially methylated imprinting control region within the Kcnq1 locus harbors a methylation-sensitive chromatin insulator. J. Biol. Chem. 277: 18106–18110.

Karolchik, D., R. Baertsch, M. Diekhans, T. S. Furey, A. Hinrichs *et al.*, 2003 The UCSC Genome Browser Database. Nucleic Acids Res. 31: 51–54.

Kent, W. J., 2002 BLAT—the BLAST-like alignment tool. Genome Res. 12: 656–664.

Kim, T. H., Z. K. Abdullaev, A. D. Smith, K. A. Ching, D. I. Loukinov *et al.*, 2007 Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128: 1231–1245.

Kosak, S. T., and M. Groudine, 2004 Gene order and dynamic domains. Science 306: 644–647.

Kurukuti, S., V. K. Tiwari, G. Tavoosidana, E. Pugacheva, A. Murrell *et al.*, 2006 CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc. Natl. Acad. Sci. USA 103: 10684–10689.

Lee, J. T., 2005 Regulation of X-chromosome counting by Tsix and Xite sequences. Science 309: 768–771.

Lee, J. T., and N. Lu, 1999 Targeted mutagenesis of Tsix leads to nonrandom X inactivation. Cell 99: 47–57.

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

Lee, T. I., S. E. Johnstone, and R. A. Young, 2006 Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. 1: 729–748.

Lewis, A., and A. Murrell, 2004 Genomic imprinting: CTCF protects the boundaries. Curr. Biol. 14: R284–R286.

Li, E., C. Beard, and R. Jaenisch, 1993 Role for DNA methylation in genomic imprinting. Nature 366: 362–365.

Ling, J. Q., T. Li, J. F. Hu, T. H. Vu, H. L. Chen *et al.*, 2006 CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science 312: 269–272.

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

tional transcription factor binding sites. Genome Res. 12: 832–839.
Lucchesi, J. C., W. G. Kelly, and B. Panning, 2005 Chromatin remodeling in dosage compensation. Annu. Rev. Genet. 39: 615–651.
Lunyak, V. V., 2008 Boundaries. Boundaries...Boundaries??? Curr. Opin. Cell Biol. 20: 281–287.
Lunyak, V. V., G. G. Prefontaine, E. Nunez, T. Cramer, B. G. Ju *et al.*, 2007 Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science 317: 248–251.
Lyon, M. F., 1961 Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190: 372–373.
Marahrens, Y., J. Loring, and R. Jaenisch, 1998 Role of the Xist gene in X chromosome choosing. Cell 92: 657–664.
Morey, C., P. Navarro, E. Debrand, P. Avner, C. Rougeulle *et al.*, 2004 The region 3′ to Xist mediates X chromosome counting and H3 Lys-4 dimethylation within the Xist gene. EMBO J. 23: 594–604.
Murrell, A., S. Heeson, and W. Reik, 2004 Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat. Genet. 36: 889–893.
Navarro, P., S. Pichard, C. Ciaudo, P. Avner, and C. Rougeulle, 2005 Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. Genes Dev. 19: 1474–1484.
Nitzsche, A., M. Paszkowski-Rogacz, F. Matarese, E. M. Janssen-Megens, N. C. Hubner *et al.*, 2011 RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. PLoS ONE 6: e19470.
Ogawa, Y., and J. T. Lee, 2003 Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. Mol. Cell 11: 731–743.
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.
Oki, M., and R. T. Kamakaka, 2005 Barrier function at HMR. Mol. Cell 19: 707–716.
Payer, B., and J. T. Lee, 2008 X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. 42: 733–772.
Penny, G. D., G. F. Kay, S. A. Sheardown, S. Rastan, and N. Brockdorff, 1996 Requirement for Xist in X chromosome inactivation. Nature 379: 131–137.
Phillips, J. E., and V. G. Corces, 2009 CTCF: master weaver of the genome. Cell 137: 1194–1211.
Pugacheva, E. M., V. K. Tiwari, Z. Abdullaev, A. A. Vostrov, P. T. Flanagan *et al.*, 2005 Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. Hum. Mol. Genet. 14: 953–965.
Sado, T., Z. Wang, H. Sasaki, and E. Li, 2001 Regulation of imprinted X-chromosome inactivation in mice by Tsix. Development 128: 1275–1286.
Sado, T., Y. Hoki, and H. Sasaki, 2005 Tsix silences Xist through modification of chromatin structure. Dev. Cell 9: 159–165.
Schwartz, S., Z. Zhang, K. A. Frazer, A. Smit, C. Riemer *et al.*, 2000 PipMaker—a web server for aligning two genomic DNA sequences. Genome Res. 10: 577–586.
Simmler, M. C., D. B. Cunningham, P. Clerc, T. Vermat, B. Caudron *et al.*, 1996 A 94 kb genomic sequence 3′ to the murine Xist gene reveals an AT rich region containing a new testis specific gene Tsx. Hum. Mol. Genet. 5: 1713–1726.
Soriano, P., 1997 The PDGF alpha receptor is required for neural crest cell development and for normal patterning of the somites. Development 124: 2691–2700.
Splinter, E., F. Grosveld, and W. de Laat, 2004 3C technology: analyzing the spatial organization of genomic loci in vivo. Methods Enzymol. 375: 493–507.
Splinter, E., H. Heath, J. Kooren, R. J. Palstra, P. Klous *et al.*, 2006 CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. 20: 2349–2354.
Stavropoulos, N., N. Lu, and J. T. Lee, 2001 A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. Proc. Natl. Acad. Sci. USA 98: 10232–10237.
Stavropoulos, N., R. K. Rowntree, and J. T. Lee, 2005 Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. Mol. Cell. Biol. 25: 2757–2769.
Sun, B. K., A. M. Deaton, and J. T. Lee, 2006 A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol. Cell 21: 617–628.
Tian, D., S. Sun, and J. T. Lee, 2010 The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell 143: 390–403.
Tolhuis, B., R. J. Palstra, E. Splinter, F. Grosveld, and W. de Laat, 2002 Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol. Cell 10: 1453–1465.
Tsai, C. L., R. K. Rowntree, D. E. Cohen, and J. T. Lee, 2008 Higher order chromatin structure at the X-inactivation center via looping DNA. Dev. Biol. 319: 416–425.
Wallace, J. A., and G. Felsenfeld, 2007 We gather together: insulators and genome organization. Curr. Opin. Genet. Dev. 17: 400–407.
Wan, L. B., and M. S. Bartolomei, 2008 Regulation of imprinting in clusters: noncoding RNAs *vs.* insulators. Adv. Genet. 61: 207–223.
West, A. G., M. Gaszner, and G. Felsenfeld, 2002 Insulators: many functions, many mechanisms. Genes Dev. 16: 271–288.
Wutz, A., and J. Gribnau, 2007 X inactivation Xplained. Curr. Opin. Genet. Dev. 17: 387–393.
Wutz, A., T. P. Rasmussen, and R. Jaenisch, 2002 Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. 30: 167–174.
Xu, N., C. L. Tsai, and J. T. Lee, 2006 Transient homologous chromosome pairing marks the onset of X inactivation. Science 311: 1149–1152.
Xu, N., M. E. Donohoe, S. S. Silva, and J. T. Lee, 2007 Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. Nat. Genet. 39: 1390–1396.
Yusufzai, T. M., H. Tagami, Y. Nakatani, and G. Felsenfeld, 2004 CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell 13: 291–298.
Zhang, L. F., K. D. Huynh, and J. T. Lee, 2007 Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell 129: 693–706.

Communicating editor: J. A. Birchler

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2011/08/12/genetics.111.132662.DC1

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

DOI: 10.1534/genetics.111.132662

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.