A novel *cis* regulatory element regulates human *XIST* in CTCF-dependent manner

Rini Shah[1,2], Ashwin Kelkar[1] & Sanjeev Galande[1]*

1. Centre of Excellence in Epigenetics, Department of Biology, Indian Institute of Science Education and Research, Dr. Homi Bhabha Road, Pune, 411008, India.

2. Present Address**:** Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA 02115

***To whom correspondence should be addressed**

Address for Correspondence:

Sanjeev Galande
Centre of Excellence in Epigenetics
Indian Institute of Science Education and Research
Dr. Homi Bhabha Road, Pune 411008, India
Tel: 91-20-25908060, Fax: 91-20-20251566
email: sanjeev@iiserpune.ac.in

Abstract

The long non-coding RNA XIST is the master regulator for the process of X chromosome inactivation in mammalian females. Here we report the existence of a hitherto uncharacterized *cis* regulatory element within the first exon of human *XIST*, which by associating with the promoter region through chromatin looping defines the transcriptional status of *XIST*. This interaction is brought about by CTCF, which in turn assists towards the maintenance of YY1 binding at the promoter and governs *XIST* transcription. Strikingly, the *cis* element is competitively bound by pluripotency factors and CTCF, wherein the enrichment of the former disrupts its interaction with the-promoter, leading to downregulation of *XIST*. Based on sequence similarity of this element across species we propose that this mechanism is conserved between the human and mouse systems. Collectively, our study uncovers

multiple transcription and epigenetic factors influencing *XIST* expression during the initiation and maintenance phases of inactivation.

Introduction

X chromosome inactivation is a gene dosage compensatory phenomenon in the mammals. XIST, the Xi-specific transcript, is the long non-coding RNA that co-ordinates the process of X chromosome inactivation (XCI) in the eutherian mammalian females (Brockdorff et al., 1991; Brown et al., 1991; Borsani et al., 1991). The phenomenon of XCI corrects for the X-linked gene dosage disparity between the males (XY) and females (XX) of mammalian species (Lyon, 1961). The first and the foremost event for the initiation of XCI is the mono-allelic and sustained upregulation of *Xist* which occurs in the epiblast cells during the implantation of the mouse embryos. At this stage, one of the two X chromosomes is randomly chosen for silencing (Kay et al., 1993; Williams et al., 2011). Xist/XIST lncRNA physically coats the chosen Xi *in cis* initiating epigenetic reprogramming characterized by the sequential removal of active chromatin marks, RNA Pol II exclusion and the establishment of repressive modifications (Li et al., 2012; Marks et al., 2009; Plath, 2003; Wakefield et al., 1997), macroH2A recruitment (Mermoud et al., 1999) and CpG methylation of gene promoters (Cotton et al., 2015; Sharp et al., 2011). Consequently, the entire X chromosome undergoes heterochromatinization, rendering it stably inactive in a mitotically heritable manner (Chow et al., 2007; Clemson et al., 1996; Duthie et al., 1999; Hall et al., 2002a; Heard et al., 1999; Herzing et al., 1997; Lee et al., 1996; Tang et al., 2010; Wutz and Jaenisch, 2000). Since Xist/XIST lncRNA is the central player for establishing the process of chromosome-wide transcriptional silencing, it is incredibly essential to regulate its levels and function to ensure proper initiation of XCI as well as maintenance.

Several studies have demonstrated that the temporal activation of *Xist* during a specific window of development is brought about by the concerted action of a network of activators and repressors either encoded from the X-inactivation centre (Xic) locus or regulating Xic (reviewed in Brockdorff and Turner, 2015). In addition to *Xist*, the Xic locus codes for a number of

lncRNAs that participate in regulating *Xist*. One of the most critical *cis*-regulators of *Xist* is its antisense lncRNA–Tsix, which unlike Xist, is exclusively produced from the future active X (Xa) and negatively regulates *Xist* by modulating the chromatin architecture of its promoter (Lee and Lu, 1999; Navarro et al., 2005; Nesterova et al., 2003; Ohhata et al., 2008; Sado et al., 2005). More recent studies demonstrate that Xic is partitioned into two distinct TADs – (i) Xist TAD (~550 Kb) containing Xist and its positive regulators such as Ftx and Jpx, (ii) Tsix TAD (~300 Kb) harbouring negative regulators of Xist such as Tsix and Xite. Both these TADs show opposite transcriptional behaviour on the chosen Xi, with the expression of genes on Xist TAD increasing and that on Tsix TAD decreasing during differentiation of mESCs (Nora *et al*., 2012, van Bemmel and Galupa *et al*., 2019). While this manuscript was under preparation, a report highlighting the role of promoter of another long non-conding RNA, *Linx* serving as a cis-regulatory silencer of Xist was published. Remarkably, *Linx* is located on Tsix TAD and serves as a silencer of Xist independent of its transcript or transcription or Tsix. However, when it is placed in Xist TAD, it serves as an enhancer of Xist (Galupa *et al*., 2020). Besides the X-chromosomal cis-acting modulators, the autosomally encoded trans-acting factors such as the core pluripotency factors – Oct4, Sox2, Nanog, Klf4 and c-Myc link the two developmental events of differentiation and XCI (Schulz et al., 2014), by regulating *Xist* (Navarro et al., 2008; Nesterova et al., 2011), *Tsix* (Donohoe et al., 2009; Navarro et al., 2010) and *Rnf12* (Navarro et al., 2011), Linx (Galupa *et al*., 2020). Thus, it can be convincingly stated that Xist expression is robustly regulated by a multitude of *cis* and *trans* factors acting either synergistically or independently to ensure accurate execution of the developmentally important process of XCI.

For the past two decades the mouse has been the preferred model system to study the molecular pathways leading to the initiation and establishment of XCI. Our understanding of the mechanism(s) regulating *XIST* and the molecular dynamics of XCI in other eutherian mammals is rather limited. Deciphering the XCI pathways in multiple systems is important to address the question of conservation and evolution of the process of XCI.

Although mouse and human *Xist*/*XIST* were discovered almost simultaneously (Brockdorff et al., 1991; Brown et al., 1991; Borsani et al., 1991), our understanding of human *XIST* regulation has remained poor as compared to its mouse counterpart. It has been demonstrated that human and mouse *XIST*/*Xist* are functionally conserved since ectopic insertion of human *XIST* in murine and human cells induces XCI (Hall et al., 2002b, 2002a; Heard et al., 1999; Migeon et al., 1999). However, there is only 49% conservation at the sequence level, with the maximum homology observed in the first exon which harbours repeat elements (A-F) (Brockdorff et al., 1992; Hendrich et al., 1993; Nesterova et al., 2001). Most notably, *Tsix*, the negative regulator of *Xist* in the mouse is a pseudogene in the humans and the potential regulatory sequences of XIST also do not show any conservation between humans and mouse (Migeon et al., 2001, 2002, Chang and Brown, 2010). Hence, apart from a plethora of other factors regulating *Xist*, the critical mechanism governed by *Tsix* does not seem to be conserved. It is known that Xist/XIST is not only induced in both the male as well as female blastocyst from the maternal as well as paternal X chromosomes, it also coats the X chromosomes, leading to partial silencing of X-linked genes in mouse but not in humans (Okamoto *et al*., 2011, Sousa *et al*., 2018). This discrepancy can be attributed to the human specific lncRNA, XACT, which specifically coats Xa and co-accumulates with XIST in human preimplantation as well huESCs, possibly tempering with XIST silencing function (Vallot *et al*., 2013 and 2015). This suggests that although the process of XCI is conserved across eutherians and is dependent on XIST RNA, this process could be manifested in diverse ways in different species.

To date three reports have addressed the regulation of human *XIST*. Hendrich *et al*.,1997 compared the putative promoter sequences from human, horse, rabbit and mouse and discovered that the first 100 base pairs (bp) upstream of the Transcription Start Site (TSS) exhibit maximum conservation and hence assigned it as the promoter of human *XIST* (Hendrich et al., 1997). Through a series of *in vitro* biochemical assays, the authors identified three transcription factors – SP1, YY1 and TBP as

potential regulators of *XIST*. The question whether any of these factors can bind *XIST* promoter in cells and regulate its expression was addressed by two independent studies where they uncovered the role of YY1 as the key transcription factor regulating *XIST* (Chapman et al., 2014; Makhlouf et al., 2014). Here, we demonstrate for the first time that besides YY1, *XIST* is also regulated by the pluripotency factors – OCT4, SOX2 and NANOG as well as CTCF. More specifically, *XIST* is repressed by the pluripotency factors and activated by YY1 and CTCF. We report the existence of a novel evolutionarily conserved *cis* regulatory element (cRE) at the *XIST* locus located in the first exon that seems to be bound by the aforementioned factors in a sex-specific and X inactivation-dependent manner. Further, this element could presumably act as a crucial determinant of transcriptional outcome from the *XIST* promoter during the initiation as well as maintenance phases of XCI.

Results

Cell line models provide the context for initiation and maintenance phases of XCI

XCI can be broadly categorised into two stages, (i) Initiation - when differentiating embryonic stem cells perform XCI for the first time and (ii) Maintenance - where any cell carrying an inactive X ensures continued expression of XIST RNA from the inactive X upon subsequent cell divisions. In the case of human XCI, YY1 is the only known regulatory factor that influences XIST expression from the inactive X. However, the question of alternate regulatory molecules affecting XIST expression is still unanswered. This is especially of importance in lieu of the fact that the antisense RNA *TSIX* is truncated/non-functional in humans (Migeon et al., 2001, 2002).

To study the regulation of *XIST* promoter concerning factors other than YY1 in the contexts of initiation and maintenance phases of XCI, we employed a human male embryonic carcinoma cell line – NTERA-2 clone D1 (NT2/D1), and differentiated cells such as HEK 293T respectively. NT2/D1 cells express pluripotency factors, low amounts of XIST as assessed by RNA-FISH (Chow et al., 2003) and respond to retinoic acid (RA) mediated

differentiation cues to give rise to neuronal progenitors (Hirka et al., 1991). Therefore, this cell line provides a good system to probe for the dynamic pattern of *XIST* promoter activity during the initial expression of XIST RNA. HEK 293T cell line is a female line and exhibits consistent expression of XIST denoting the maintenance phase of XCI (Xiao et al., 2007).

Firstly, we determined the expression of XIST in NT2/D1, female cell line - HEK 293T and a male cell line - DLD1 by semi-quantitative RT-PCR and observed that XIST is synthesized only in HEK 293T cells **(Fig. 1A)**. To understand the regulation of *XIST* during the initiation phase, we performed RA-mediated differentiation of NT2/D1 cells and show that the levels of pluripotency factors – OCT4 (POU5F1), SOX2, NANOG decline, and the expression of neuronal progenitor specific marker – PAX6 is upregulated during differentiation, thus corroborating the previous report (Hirka et al., 1991) **(Fig. 1B,C)**. Next, we monitored the expression of *XIST* by performing quantitative real-time PCR and RNA FISH and observed that it is upregulated progressively upon differentiation of NT2/D1 cells **(Fig. 1D, E, F)**. RNA FISH analysis indicated that over 20% NT2/D1 cells exhibit XIST RNA FISH signal upon differentiation of NT2/D1 for 5 days **(Fig. 1E, F)**. Based on these results, we believe that NT2/D1 cells can serve as a good model to address the question of human *XIST* regulation, especially, since human ES cells haven't served as an ideal model for the purpose so far.

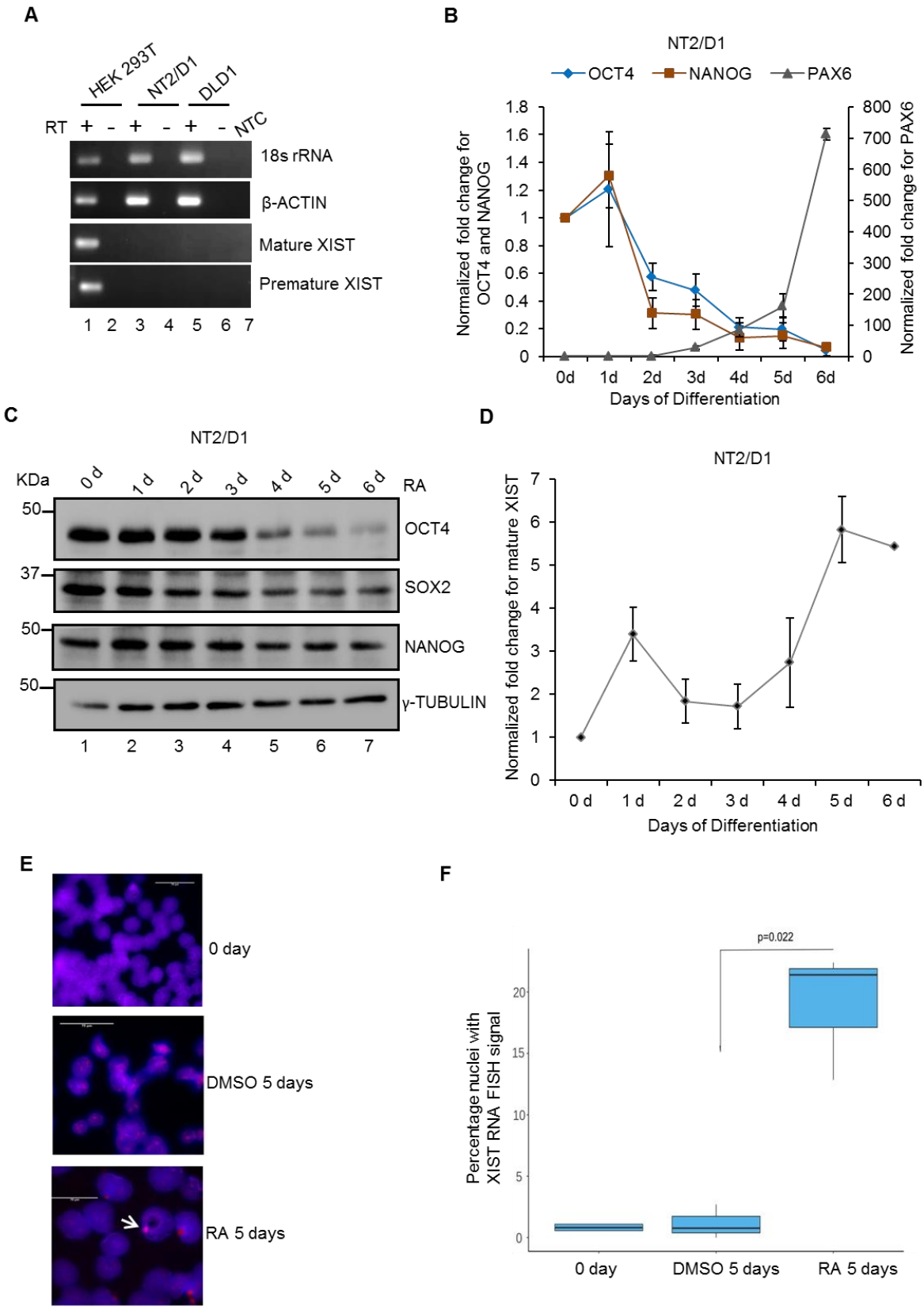


Figure 1. NT2/D1 and HEK 293T cells provide the contexts for initiation and maintenance phases of XCI. (A) Semi-quantitative RT-PCR for mature and premature XIST using cDNA prepared from HEK 293T (female), NT2/D1 (male) and DLD1 (male) cells. 18s rRNA and β-ACTIN serve as controls. (B)

qRT-PCR depicting a decrease in the levels of OCT4 and NANOG and increase in PAX6 levels upon RA-mediated differentiation of NT2/D1 cells for 6 days. X-axis represents the differentiation time-point and Y-axis represents the fold change normalized to 18s rRNA. Each point on the graph represents values from 5 independent experiments and error bar represents ±S.E.M. (C) Immunoblotting showing a decrease in OCT4, SOX2 and NANOG levels upon RA-mediated differentiation of NT2/D1 cells for 6 days. γ-TUBULIN serves as an equal loading control. (D) qRT-PCR depicting an increase in the levels of XIST upon RA-mediated differentiation of NT2/D1 cells for 6 days. X-axis represents the differentiation time-point and Y-axis represents the fold change normalized to 18s rRNA. Each point on the graph represents values from 3 independent experiments and error bar represents ±S.E.M. (E) RNA FISH for mature *XIST* for undifferentiated (0 day, 5 day DMSO) and differentiating (5 day, RA) NT2/D1 cells. Arrowheads indicate the FISH signal. Scale bar = 75 microns. (F) Quantification for the RNA FISH signals in 0 day, 5 day DMSO and 5 day RA treated NT2/D1 cells. N=3, 200 nuclei were counted for each replicate and statistical significance was ascertained by Student's T-test.

Transcription from *XIST* promoter is governed by the promoter as well as exon1 of *XIST*

Identification of the *XIST* promoter element would be the first step to understand the regulation of *XIST*. The pioneering study attempting to characterize the promoter of human *XIST* restricted to +50 to -50 bp from the TSS, since it was found to be conserved across four mammalian species (Hendrich et al., 1997). We chose to test the larger promoter region to uncover the unique potential regulatory elements for human *XIST*. Hence, we cloned the genomic region ranging from +50 bp to -4408 bp upstream of *XIST* TSS into a promoter-less reporter vector – pGL3 Basic **(Supp Fig. 1)**. These DNA constructs were then transfected into NT2/D1 and HEK 293T cells and assayed for the presence of promoter by measuring the induced firefly luciferase activity. Similar to the first study on characterizing *XIST* promoter elements (Hendrich et al., 1997), it was observed that the +50 bp to -51 bp region was sufficient to drive the transcription of the luciferase gene **(Fig. 2A,B)**. Also, the promoter activities for all other fragments with increasing distance from the TSS (except for the fragment +50 bp to -260 bp, bar 3 in **Fig. 2A,B**) remained constant when compared to the vector control. As a control, we also cloned one of the fragments, +50 bp to -1050 bp in the anti-sense (AS) orientation. Since *XIST* promoter is unidirectional, the

fragment cloned in the AS direction fails to transcribe the reporter gene and hence does not exhibit any measurable reporter activity **(**AS**, Fig. 2(A, B))**.

The tested promoter fragments exhibited a similar trend in both cell lines tested. This implies that the assessment of transcription activity is independent of the cellular context wherein all the necessary transcription factors are present in both cell types. This opens up the question as to why does *XIST* fail to express in the undifferentiated NT2/D1 cells. To address the effect of differentiation status of NT2/D1 cells on the *XIST* promoter fragments, we analysed the same in differentiating NT2/D1 cells. Since the increase in the expression of XIST correlates with the differentiation of NT2/D1 cells, we expected a similar increase in the luciferase activity upon differentiation. But surprisingly we observed a decline in luciferase activity instead **(Fig. 2C)**. Upon measuring the expression of previously identified transcriptional activators of *XIST* - SP1 and YY1, we found a decrease in their levels upon differentiation **(Fig. 2D)** reaffirming the fact that they could positively regulate *XIST*. The disparity between the two results **(Fig. 1D and Fig. 2C)** could be explained by the incorporation of additional regulatory factors into the model of *XIST* transcriptional activation.

In the mouse system, pluripotency factors negatively regulate *Xist* and therefore in the NT2/D1 cells, which mimic undifferentiated embryonic stem cells, these were the logical candidates for the role. Towards this, we scanned the *XIST* locus for the binding sites of pluripotency factors by re-analysing the available ChIP-sequencing datasets for OCT4 and NANOG in the H9 human embryonic stem cells (hESC) and OCT4 in NT2/D1 cells **(Supp Fig. 2A)**. The rationale for the speculation was the negative correlation between XIST and the stem cell marker levels during differentiation **(Fig. 1B,C,D)** as well as previous reports demonstrating the negative regulation of *Xist* by them (Navarro et al., 2008). ChIP-seq analysis revealed that OCT4 and NANOG are enriched in the 1st exon, ~+4.5 Kb from the TSS in both the cell types independent of sex **(Supp Fig. 2A)**. Therefore, we validated their binding at this particular location by ChIP and observed significant enrichment of OCT4 and SOX2 in undifferentiated (0 day) as compared to 5 day differentiated NT2/D1 cells **(Fig. 2E)**. This indicates that

the pluripotency factors binding might serve to repress *XIST* in undifferentiated state.

The next prominent question is to identify the factors positively regulating *XIST* in this scenario. SP1 and YY1 have been proposed to be the transcriptional activators of *XIST* (Hendrich et al., 1997), and YY1 has been demonstrated to bind to a site +1.5 Kb from the TSS (Chapman et al., 2014; Makhlouf et al., 2014). Therefore, we determined their binding at +1.5 Kb from the TSS and observed a significant enrichment of YY1 in 5 day differentiated NT2/D1 cells **(Fig. 2F)**. Also, both SP1 and YY1 are enriched on the putative promoter of *XIST* of the inactive X chromosome since their occupancies can be observed only in the female cells (HEK 293T) and not in the male cells (DLD1) **(Supp Fig. 2B)**. We also determined the occupancies of the pluripotency factors at the +1.5 Kb site and did not observe any significant enrichment. The *NANOG* promoter region comprising ChIP-seq peaks for OCT4 and SOX2 served as a positive control for the ChIP assay **(Supp Fig. 9C,D)**. These results suggest that the enrichment of the pluripotency factors, 4.5 Kb downstream of the TSS negatively correlates with XIST expression. And the occupancy of YY1, the known transcriptional activator of *XIST* at +1.5 Kb correlates with the induction of *XIST* in NT2/D1 differentiation model. Therefore, it seems likely that similar to the mouse system, the pluripotency factors act as the potential repressors of human *XIST*, the hypothesis tested further in the current study.

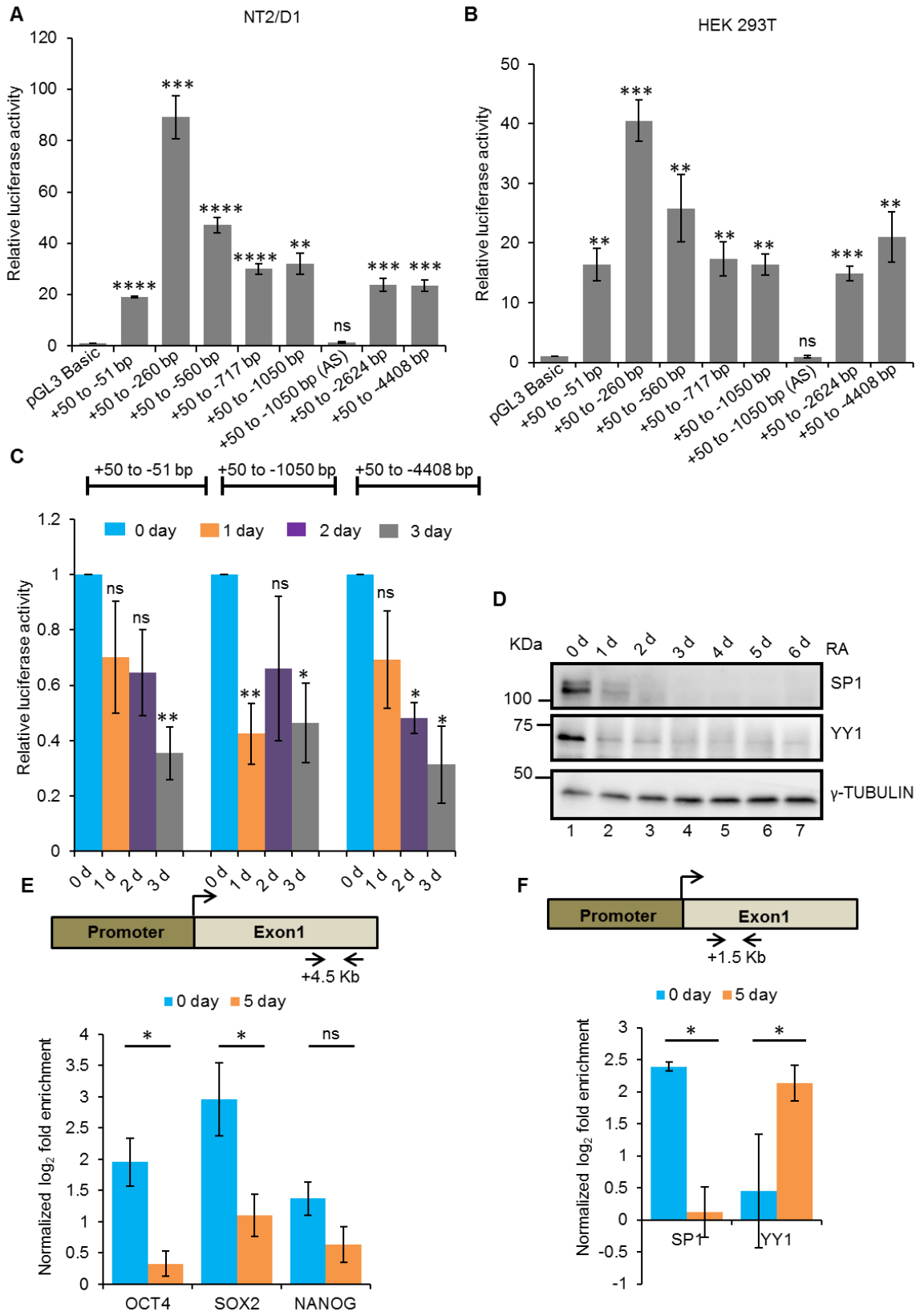


Figure 2. Induction of XIST during differentiation of NT2/D1 cells is governed by the promoter as well as exon1 of *XIST*. (A,B) Luciferase reporter activities for *XIST* promoter constructs transfected into NT2/D1 cells (A) or HEK 293T

(B) compared to pGL3 Basic. Firefly luciferase activities represented here are normalized to Renilla luciferase activity, which serves as an internal control. X-axis indicates DNA constructs transfected and Y-axis represents the normalized fold change in firefly luciferase activity. Each bar represents values from three independent experiments. Error bar represent ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, ***P value < 0.001, **P value < 0.01, *P value < 0.05, ns = non-significant). (C) Luciferase reporter activities of +50 bp to -51 bp, +50 bp to -1050 bp and +50 bp to -4408 bp promoter constructs decrease upon RA-mediated differentiation of NT2/D1 for 3 days. Firefly luciferase activities represented here are normalized to Renilla luciferase activity, which serves as an internal control. X-axis indicates differentiation time-points and Y-axis represents the normalized fold change in the firefly luciferase activity. Each bar represents values from three independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (**P value < 0.01, *P value < 0.05, ns = non-significant). (D) SP1 and YY1 proteins decrease upon RA-mediated differentiation of NT2/D1 for 6 days. γ-TUBULIN serves as a loading control. (E) ChIP-qPCR analysis showing a decrease in OCT4, SOX2 and NANOG enrichment on the *XIST* at +4.5 Kb (as shown in the schematic above) in undifferentiated (0 day) versus 5 day differentiated NT2/D1 cells. (F) ChIP-qPCR analysis demonstrating a change in the occupancies of SP1 and YY1 on *XIST* promoter-proximal region (+1.5 Kb) (as shown in the schematic above) in undifferentiated (0 day) versus 5 day differentiated NT2/D1 cells. (E,F) X-axis represents the immunoprecipitated factor and Y-axis represents the normalized $\log_2$ fold enrichment over IgG. Each bar represents values from 2 independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over undifferentiated cells (0 day) as per Student's T-test (*P value < 0.05, ns = non-significant).

Pluripotency factors negatively regulate *XIST*

To ascribe the roles of repressors to the pluripotency factors, we ectopically overexpressed FLAG-tagged OCT4 or SOX2 or NANOG **(Fig. 3A)** in a female cell line, HEK 293T, which expresses XIST and does not express any of these factors. A significant decrease in the levels of both mature and premature XIST was observed implying that the pluripotency factors negatively regulate *XIST* **(Fig. 3B)**. It is noteworthy that levels of YY1 (a well-established activator of *XIST*) remained unperturbed under these experimental conditions **(Supp Fig. 3A)**. To lend further support to this hypothesis, we tested the repressive role of pluripotency factors by employing NT2/D1 differentiation model. Towards this, NT2/D1 cells were differentiated for 4 days, at which point the levels of pluripotency factors

would begin to decline **(Fig. 1B,C)**, and XIST expression would be induced **(Fig. 1D)**, followed by the over-expression pluripotency factors. The experimental scheme is depicted in **Fig. 3C**. We observed a significant decline in XIST expression under these experimental conditions as well, thereby strengthening the proposed repressor role for OCT4, SOX2 and NANOG **(Fig. 3D,E)**. In addition to this, we also performed a knockdown experiment in which OCT4, SOX2 or NANOG were targeted by specific siRNA in undifferentiated NT2/D1 cells and observed a significant increase in XIST levels upon single knockdown of OCT4 or in combination with SOX2 or NANOG **(Supp Fig. 3B, D)**. Increase in the expression of neuronal marker – PAX6 is expected upon depletion of pluripotency factors and hence serves as a control **(Supp Fig. 3C)**. The immunoblot analysis performed to confirm the knockdown efficiencies demonstrate that depletion of OCT4 also leads to a significant decrease in the levels of SOX2 and NANOG **(Supp Fig. 3B)**. While knockdown of SOX2 leads to measurable reduction in OCT4 levels as well, depletion of NANOG does not cause any appreciable change in the expression of the other two pluripotency factors. Therefore, we observed a significant upregulation in XIST expression only when the levels of all the three factors were severely reduced, which is under the condition where OCT4 is knocked down.

Furthermore, we performed ChIP using FLAG antibody in both the cell lines (HEK 293T and NT2/D1) to determine if the overexpressed proteins (FLAG-OCT4/SOX2/NANOG) are capable of binding the +4.5 Kb site. We observed a significant enrichment of all the three factors on the exon 1, +4.5 Kb from the TSS **(Fig. 3F,G)** providing a clear evidence that the pluripotency factors negatively regulate transcription from *XIST* promoter by directly binding to this site. These results led us to conclude that OCT4, SOX2 and NANOG can act as the repressors of *XIST* in the embryonic stem cells, wherein it is necessary to prevent premature induction of XIST. Considering the effect of OCT4, SOX2 and NANOG on XIST expression in HEK 293T cells, we postulated that occupancy of +4.5 Kb site by the pluripotency factors could have a more potent effect on the expression of XIST in

conjunction with the binding of the known activator – YY1 on the promoter-proximal region.

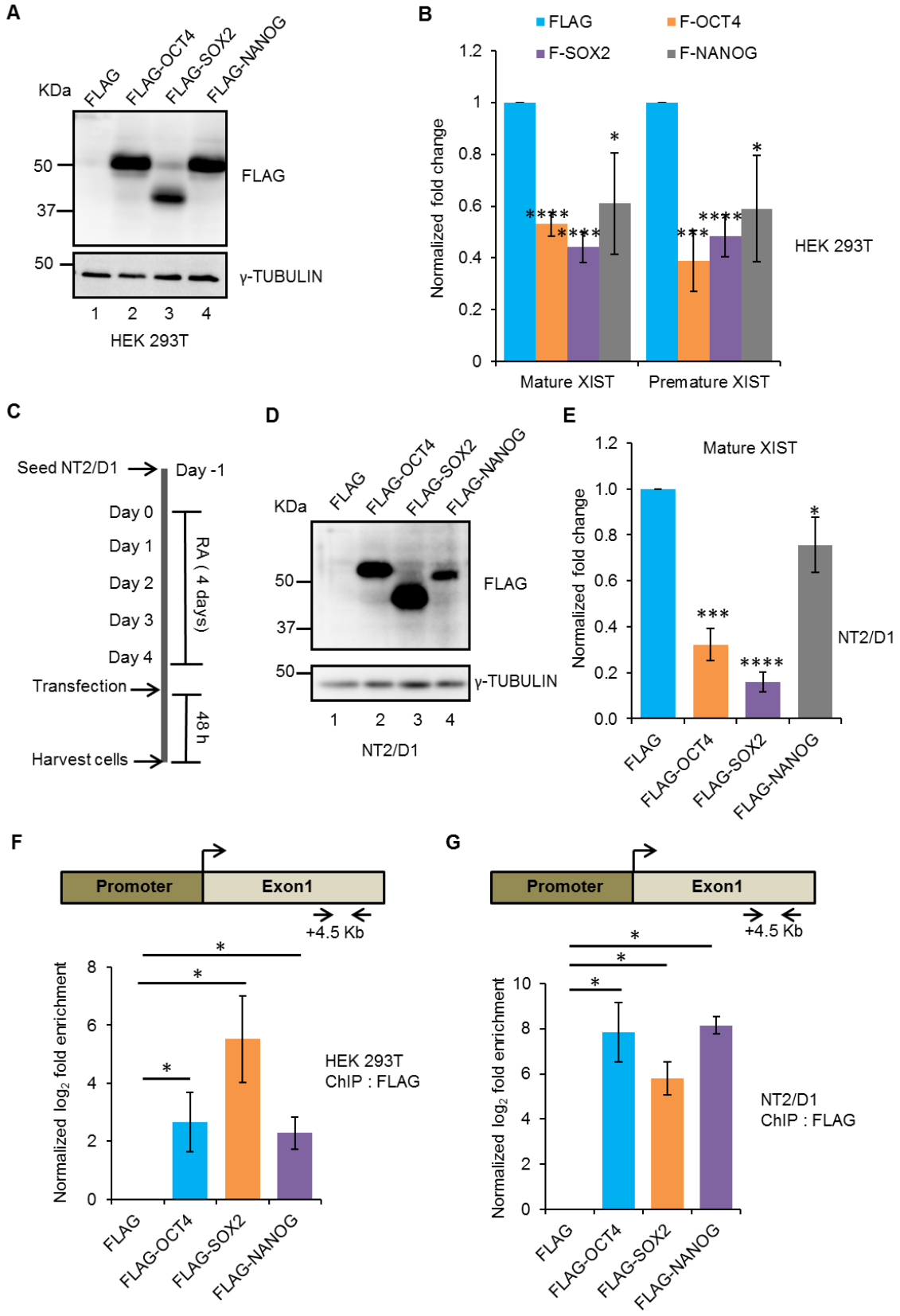


Figure 3. Pluripotency factors repress *XIST* by binding to exon 1 (+4.5 Kb) site. (A) Immunoblotting for FLAG to confirm the over-expression of OCT4, SOX2, NANOG in HEK 293T cells. γ-TUBULIN serves as an equal loading control. (B) qRT-PCR showing a significant reduction in mature and premature XIST upon over-expression of pluripotency factors in HEK 293T cells. X-axis represents the mature or premature XIST and Y-axis represents the fold change normalized to 18s rRNA. Each point on the graph represents values from 5 independent experiments and error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, ***P value < 0.001, *P value < 0.05). (C) Experimental scheme to overexpress OCT4, SOX2, NANOG in NT2/D1 cells differentiated for 4 days. (D) Immunoblotting for FLAG to confirm the over-expression of OCT4, SOX2, NANOG in NT2/D1 cells differentiated for 4 days. γ-TUBULIN serves as an equal loading control. (E) qRT-PCR showing a significant reduction in mature XIST upon over-expression of pluripotency factors in NT2/D1 cells differentiated for 4 days. X-axis represents transfected DNA and Y-axis represents the fold change normalized to 18s rRNA. Each point on the graph represents values from 3 independent experiments and error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, ***P value < 0.001, *P value < 0.05). (F,G) ChIP showing significantly higher occupancies of OCT4, SOX2, NANOG on the exon 1 (+4.5 Kb) site upon their over-expression in HEK 293T and NT2/D1 cells differentiated for 4 days. X-axis represents the transfected DNA and Y-axis represents the fold change normalized to FLAG control. Each point on the graph represents values from 3 (HEK 293T) and 2 (NT2/D1) independent experiments and error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (*P value < 0.05).

The pluripotency factor bound regulatory site (+4.5 Kb) on *XIST* exon 1 is a potential *cis* regulatory element (cRE)

The results obtained thus far provided compelling evidence to further characterize the regulatory region of the human *XIST* exon1 (+4.5 Kb from the TSS). It seemed likely that this element harbours regulatory potential whose function is modulated by the binding of pluripotency factors. To assess the functional significance of this site, we performed ChIP for active (H3K4me1, H3K27ac) and inactive (H3K27me3) histone modifications in female (HEK 293T) as well as male (DLD1) cells and observed a significant enrichment of active chromatin marks only in the female cells indicating that we are indeed scoring for the *XIST* locus from the inactive X (Xi) **(Fig. 4A)**. Additionally, there is no significant difference between H3K27me3 enrichment on this site in the male and female cells suggesting it to be enriched on the *XIST* locus of active X (Xa) **(Fig. 4A)**.

Another feature of the newly identified cRE is the significant enrichment of chromatin organizer protein CTCF in female cells (MCF7, HeLa, HEK 293 and female skin epithelium) versus male (DLD1 and male skin epithelium) cells as evident from the analysis of available ChIP-seq datasets **(Fig. 4B)**. The inclusion of primary cell data lends confidence in the findings and excludes the possibility of observing CTCF occupancy merely as an artefact of cultured cell lines. It is notable that the peaks for CTCF and OCT4/NANOG are overlapping (**Fig. 4B**), which could essentially point towards the functionality of cRE. However, a previous report ruled out the role of CTCF in regulating *XIST* since they observed no appreciable difference in its enrichment on the part of *XIST* locus tested between male and female cells (Makhlouf et al., 2014). Nonetheless, this report has not examined the site identified in our study and hence, we carefully examined its role in governing *XIST* transcription.

Firstly, we perturbed levels of CTCF as well as that of SP1 or YY1 in HEK 293T cells and observed a decrease in both mature and premature XIST levels upon depletion of YY1 or CTCF **(Fig. 4C,D, Supp Fig. 4A,B)**. The effect of CTCF depletion on SP1 and YY1 levels was virtually unnoticeable **(Fig. 4C)**. Next, we over-expressed SP1, YY1 or CTCF in undifferentiated NT2/D1 cells which expresses XIST at very low levels and observed a significant increase in XIST levels upon over-expression of YY1 or CTCF **(Supp Fig. 5 A, B)**. We also determined that the over-expression of SP1 or YY1 or CTCF in NT2/D1 cause no appreciable decrease in the levels of pluripotency factors, thereby ruling out the indirect effect of over-expressions **(Supp Fig. 5C)**. These results strongly suggest that in addition to YY1, CTCF is an essential regulator of *XIST*. Furthermore, we validated *in silico* analysis finding of CTCF binding by performing ChIP for CTCF in HEK 293T, DLD1 cells **(Fig. 4E (i))** and differentiating NT2/D1 cells **(Fig. 4E (ii))** and observed a differential enrichment on the cRE only in the female cells or 5 day differentiated NT2/D1 cells, thereby positively correlating with XIST expression **(Fig. 4D)**. To confirm whether CTCF is indeed bound to the cRE only on the inactive X, we performed sequential ChIP in HEK 293T cells using H3K27ac (mark enriched at +4.5 Kb site only in female cells as per

Fig. 4A) or H3K27me3 (mark enriched at +4.5 Kb site in both female and male cells as per Fig. 4A) antibodies for the first ChIP followed by second ChIP using anti-CTCF. We observed a remarkably higher pulldown of CTCF when the first ChIP was performed using H3K27ac antibody **(Fig. 4E (iii)**, **blue bar)** compared to H3K27me3 antibody **(Fig. 4E (iii), brown bar)**.

Additionally, there is no difference in the CTCF enrichment between female (HEK 293T) and male (DLD1) cells for the previously reported sites (+1.5 Kb and -3.0 Kb from the TSS) consistent with the published report and the *in silico* analyses (Makhlouf et al., 2014)**(Fig. 4B, Supp Fig. 4C,D)**. However, we did observe a statistically significant difference between H3K27ac enrichment on both these sites **(Supp Fig. 4C,D)**. Henceforth, we denote +1.5 Kb and -3.0 Kb sites as the promoter-proximal and promoter-distal regions respectively. We performed ChIP-qPCR for a control region in HEK 293T and DLD1 cells (~1.3 Kb from the unique CTCT binding site on the *XIST* locus) that does not show any CTCF peaks and found that neither YY1 nor CTCF binds this particular site. ChIP analysis using antibodies against H3, H3K27ac and H3K27me3 does exhibit enrichment, thereby serving as a positive control ascertaining the quality of the chromatin **(Supp Fig. 9A, B)**.

Altogether, the results obtained suggest the existence of a cRE on the exon1 of *XIST* whose function is modulated by the binding of CTCF or the pluripotency factors.

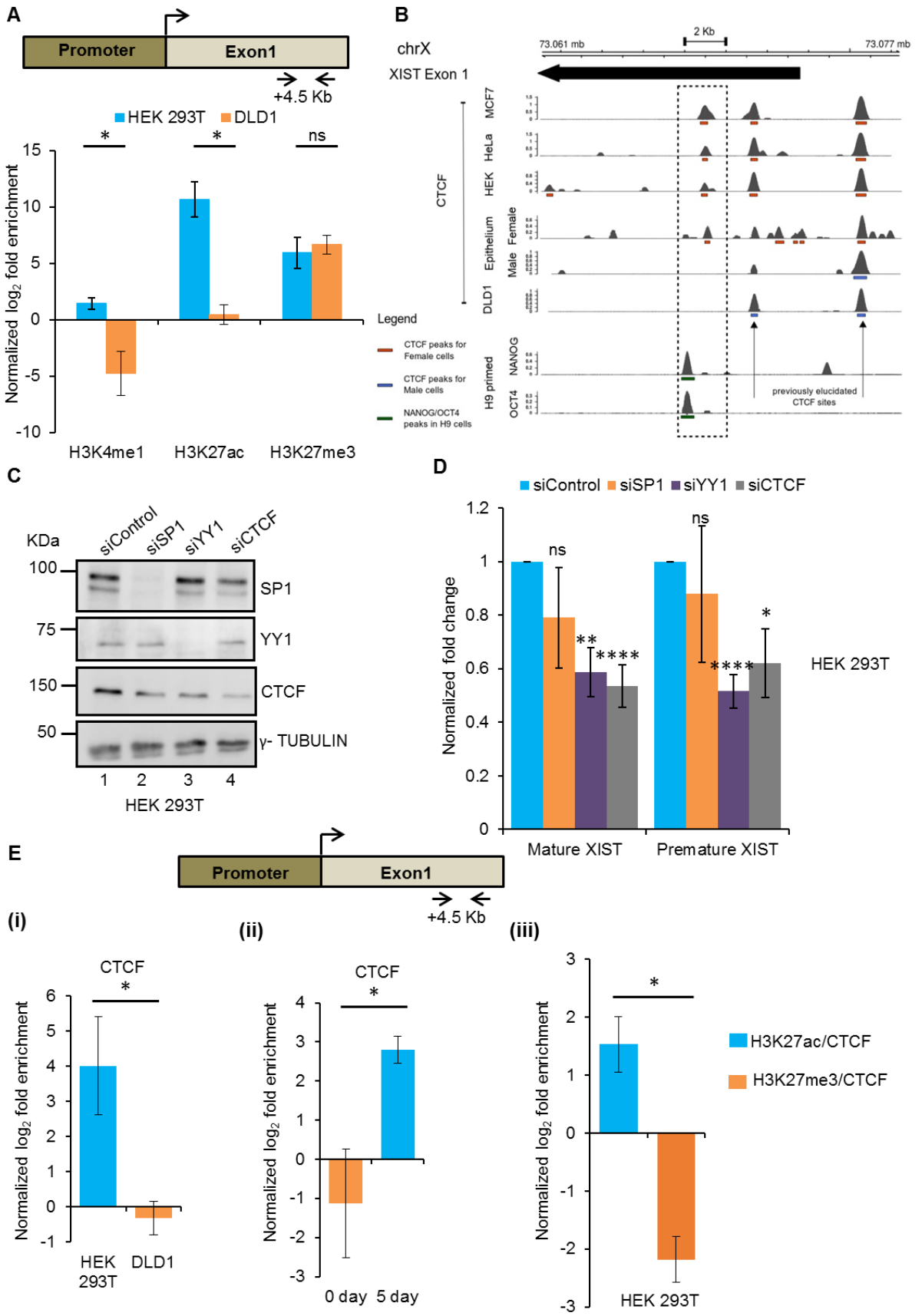


Figure 4. Pluripotency factor binding element is the potential cRE. (A) ChIP-qPCR analysis showing enrichment of active histone marks H3K4me1,

H3K27ac and repressive histone mark H3K27me3 for *XIST* cRE (+4.5 Kb, as shown in the schematic above) on exon1 in HEK 293T (female, blue bar) and DLD1 (male, yellow bar) cells. X-axis represents the antibodies used for ChIP and Y-axis represents the normalized log_2fold enrichment over IgG. Each bar represents values from 3 independent experiments. Error bar represent ±S.E.M. Asterisks represent the significance over DLD1 ChIP as per Student's T-test (*P value < 0.05, ns = non-significant). (B) ChIP-seq peaks for OCT4 and NANOG for the *XIST* locus in human embryonic stem cell line H9 and OCT4 ChIP-seq peak for NT2/D1 cells. CTCF ChIP-seq for male (DLD1, skin epithelium) and female cell (MCF7, HeLa, HEK 293, skin epithelium) lines. The arrowhead indicates the direction of *XIST* transcription and tail denoted the TSS. All peaks were confirmed to have a p value < 0.05 as reported by MACS2 callpeak function. (C) Immunoblotting to confirm siRNA mediated knockdown of SP1, YY1, CTCF in HEK 293T cells. γ-TUBULIN serves as an equal loading control. (D) qRT-PCR demonstrating reduction in XIST (both mature and premature) levels upon knockdown of YY1 or CTCF in HEK 293T cells. X-axis represents the mature or premature XIST, and Y-axis represents the fold change normalized to 18s rRNA. Each point on the graph represents values from 4 independent experiments and error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, **P value < 0.01, *P value < 0.05, ns=non-significant). (E) ChIP-qPCR analysis showing enrichment of CTCF on *XIST* cRE (+4.5 Kb) on exon1 in (i) HEK 293T (female, blue bar) and DLD1 (male, brown bar) cells (ii) 0 day (brown bar) and 5 day RA treated (blue bar) NT2/D1 cells and (iii) Sequential ChIP in HEK 293T cells. X-axis represents the cell lines or the differentiation time-points, and Y-axis represents the normalized log_2fold enrichment over IgG. Each bar represents values from 3 (for (i) and (ii)) or 2 (for (iii)) independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance as per Student's T-test (*P value < 0.05).

Binding of CTCF to the cRE is required for the activation of *XIST*

To address the relevance of CTCF binding on the cRE, we sought to determine its occupancy upon over-expression of pluripotency factors or depletion of CTCF in HEK 293T cells. As demonstrated earlier, XIST levels decrease under both these conditions **(Fig. 3B, 4D)**. Also, there is a higher enrichment of pluripotency factors (especially SOX2) on the cRE **(Fig. 3F)**, indicating that the downregulation of XIST may be a consequence of their direct binding to this element. By performing ChIP for CTCF upon over-expression of pluripotency factors in HEK 293T cells, we observed that CTCF bound to cRE is displaced by SOX2 **(Fig. 5C)**. Interestingly, over-expression of YY1 or CTCF coupled with depletion of OCT4 in NT2/D1 cells led to higher XIST upregulation **(Fig. 5A,B)** when compared with the over-

expressions **(Supp Fig. 5B)** or the depletion of pluripotency factors alone **(Supp Fig. 3D)**. This provides a clue that it is the balance of activators (YY1, CTCF) and repressors (OCT4, SOX2, NANOG) and, their binding on the same genomic element governs the transcriptional activation of *XIST*. It is noteworthy that the siRNA-mediated knockdown of OCT4 depletes not only OCT4 but also SOX2 **(Fig. 5A)**.

Further confirmation that CTCF binding to the cRE directly regulates XIST transcription came from the observation that the enrichment of CTCF decreases and repressive mark H3K27me3 increases upon depletion of CTCF in HEK 293T cells **(Fig. 5D)**. We also show that the CTCF occupancies as well as enrichment of H3K27ac on the CTCF peaks common between males and females are reduced upon knockdown of CTCF, accompanied by the concomitant increase in H3K27me3 **(Supp Fig. 6A,B)**. Taken together, these findings underscore the significance of CTCF in shaping the epigenetic landscape of *XIST* locus and hence, mediating the transcription of *XIST*.

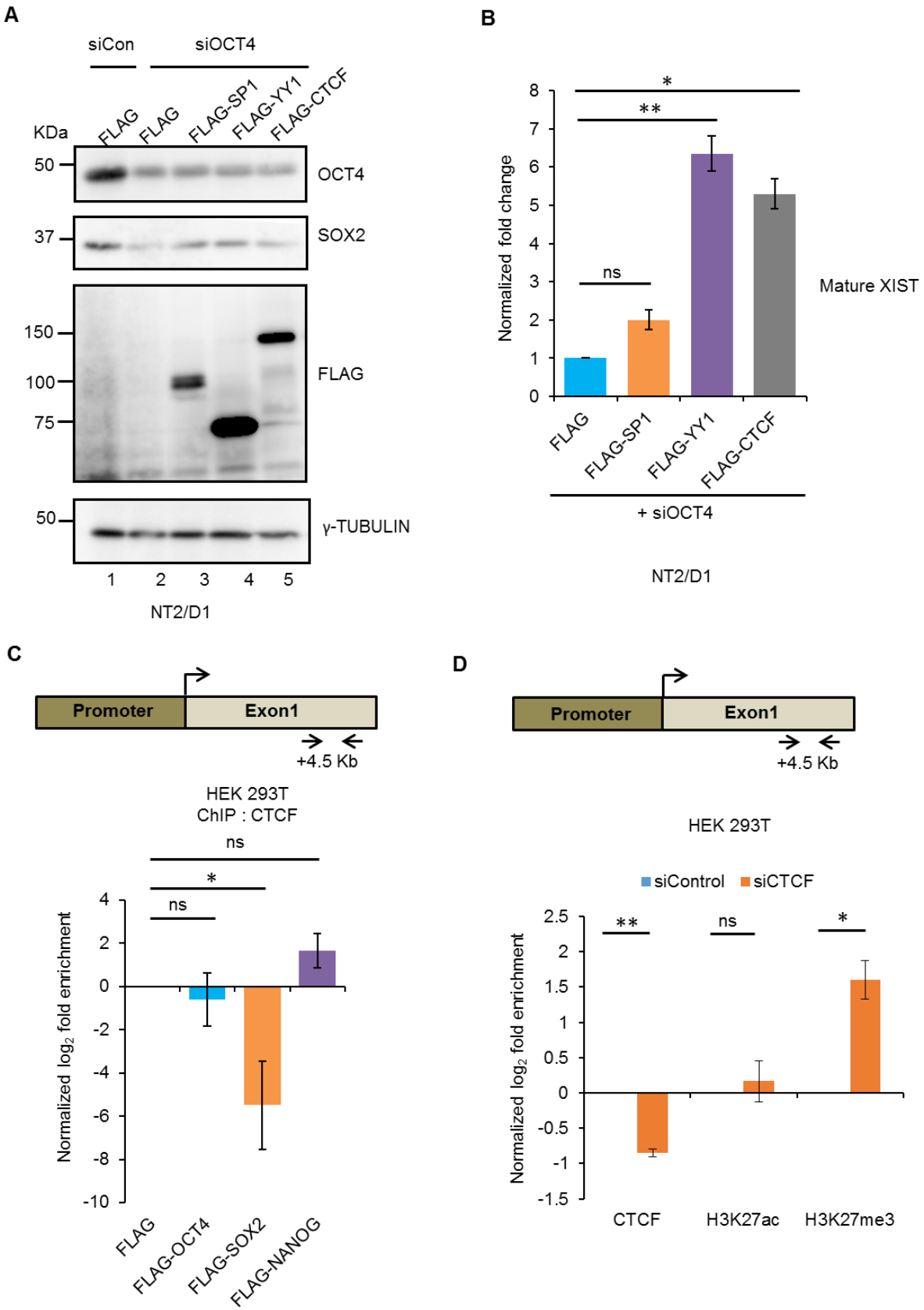

A
siCon
siOCT4
KDa
FLAG
FLAG
FLAG-SP1
FLAG-YY1
FLAG-CTCF
50
OCT4
37
SOX2
150
100
75
FLAG
50
γ-TUBULIN
1 2 3 4 5
NT2/D1
B
Normalized fold change
ns
**
*
Mature XIST
FLAG
FLAG-SP1
FLAG-YY1
FLAG-CTCF
+ siOCT4
NT2/D1
C
Promoter
Exon1
+4.5 Kb
HEK 293T
ChIP : CTCF
ns
*
ns
Normalized log2 fold enrichment
FLAG
FLAG-OCT4
FLAG-SOX2
FLAG-NANOG
D
Promoter
Exon1
+4.5 Kb
HEK 293T
siControl
siCTCF
**
ns
*
Normalized log2 fold enrichment
CTCF
H3K27ac
H3K27me3

Figure 5. CTCF activates *XIST* by binding to cRE. (A) Immunoblotting for OCT4 and SOX2 upon the knockdown of OCT4 and for FLAG to confirm the over-expression of SP1, YY1, CTCF in NT2D1 cells. γ -TUBULIN serves as an equal loading control. (B) qRT-PCR for mature XIST upon knockdown of OCT4 and over-expression of SP1, YY1, CTCF in NT2D1 cells. X-axis represents DNA and siRNA transfected and Y-axis represent the fold change normalized to 18s rRNA. Each bar represents values from 2 independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (**P value < 0.01, *P value < 0.05, ns = non-significant). (C) ChIP-qPCR analysis showing a reduction of CTCF occupancy on the cRE (+4.5 Kb) on exon1 upon over-expression of OCT4, SOX2, NANOG in HEK 293T cells. X-axis represents the transfected DNA, and Y-axis represents the normalized log_2fold enrichment over FLAG vector control. Each bar represents values from 3 independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance for CTCF ChIP over vector control as per Student's T-test (*P value < 0.05, ns=non-significant). (D) ChIP-qPCR analysis demonstrating a significant decrease in CTCF enrichment and an increase in H3K27me3 enrichment upon siRNA mediated knockdown of CTCF in HEK 293T cells (brown bar). X-axis represents the antibody used for ChIP and Y-axis represents the normalized fold enrichment over control siRNA. Each bar represents values from 2 independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance for ChIP for siRNA control as per Student's T-test (**P value < 0.01,*P value < 0.05, ns=non-significant).

CTCF binding to the cRE assists in stable YY1 binding at the *XIST* promoter-proximal site

While addressing the question whether YY1 and CTCF operate via the same regulatory pathway, we monitored XIST levels upon double knockdown of YY1 and CTCF in HEK 293T cells and observed that there is no cumulative effect of dual knockdown on XIST expression and the decrease is the same as that obtained upon single knockdown of YY1 or CTCF **(Supp Fig. 4A,B)**. This led us to ask whether CTCF occupancy on the cRE influences YY1 binding on the promoter-proximal region. Towards this, we performed ChIP for YY1 upon depletion of CTCF in HEK 293T cells and indeed observed a significant decrease in YY1 occupancy on its previously identified cognate site **(Fig. 6A)** (Makhlouf et al., 2014, Chapman et al., 2014). We also observed a decline in H3K27ac and an increase in H3K27me3 enrichment on the YY1 binding site **(Supp Fig. 6A)**. Furthermore, the decrease in XIST levels observed upon knockdown of CTCF in HEK 293T cells is restored by over-expression of YY1 **(Fig. 6B,C)**.

Thus, it can be concluded that CTCF governs the transcription from the *XIST* promoter by assisting in the maintenance of YY1 binding at the promoter-proximal site.

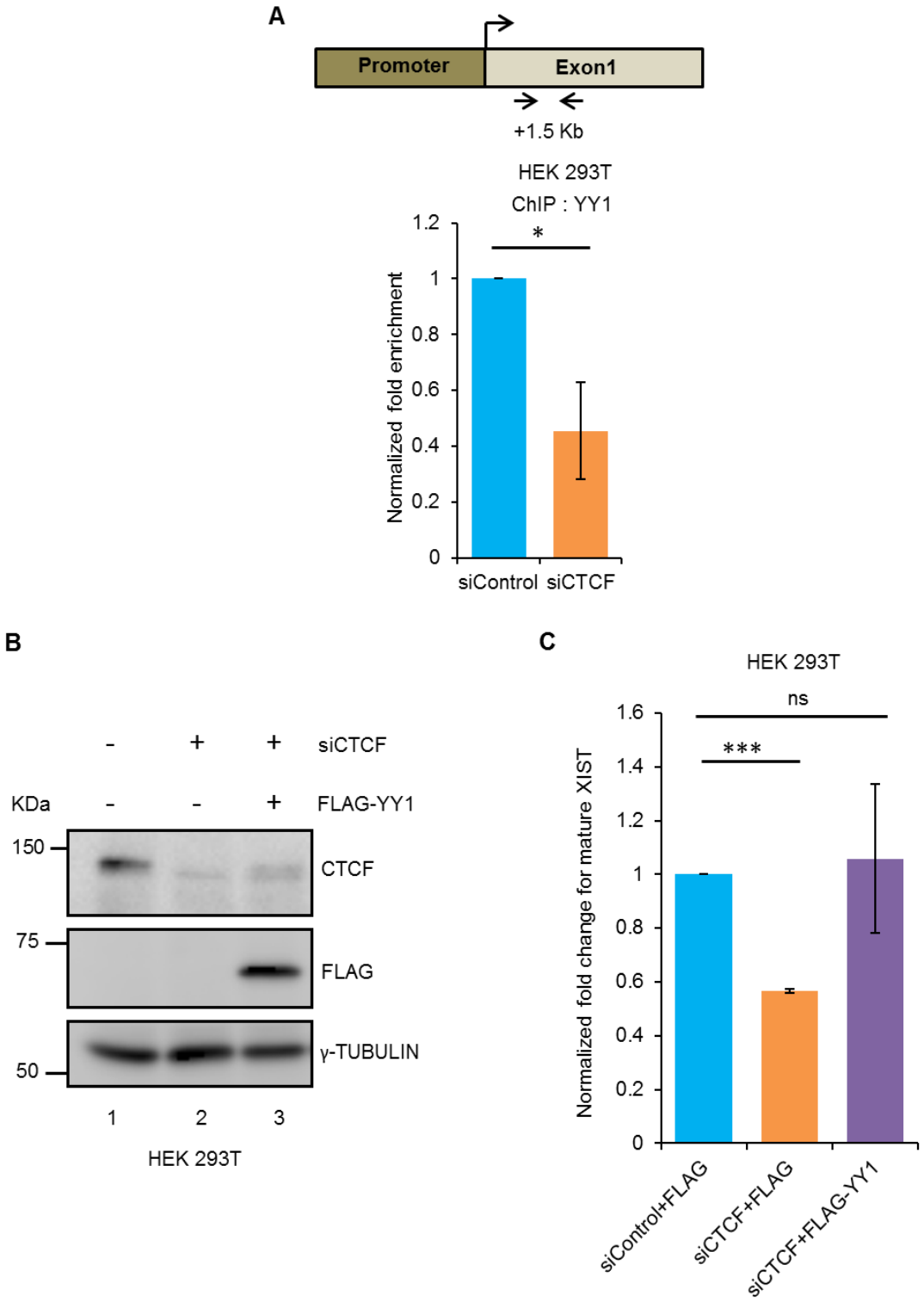


Figure 6. CTCF assists in the recruitment or maintenance of YY1 binding to the promoter-proximal region of *XIST*. (A) ChIP-qPCR analysis showing a reduction of YY1 occupancy on the promoter-proximal site (+1.5 Kb) upon siRNA mediated knockdown of CTCF in HEK 293T cells (brown bar). X-axis represents the antibody used for ChIP, and Y-axis represents the normalized fold enrichment over control siRNA (blue bar). Each bar represents values from 3 independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over YY1 ChIP for vector control as per Student's T-test (*P value < 0.05). X-axis represents the transfected siRNA, and Y-axis represents the normalized fold enrichment over control siRNA. (B) Immunoblotting to confirm knockdown of CTCF and over-expression of FLAG-tagged YY1 in HEK 293T cells. γ-TUBULIN serves as an equal loading control. (C) qRT-PCR for mature XIST upon depletion of CTCF and/or over-expression of FLAG-YY1 in HEK 293T cells. X-axis represents the DNA and siRNA transfected and Y-axis represents the normalized fold enrichment over 18s rRNA. Each bar represents values from 2 independent experiments. Error bar represents ±S.E.M. Asterisks represent the significance over control (1st bar) as per Student's T-test (***P value < 0.001, ns=non-significant).

CTCF mediates the interaction between the cRE and the promoter of *XIST* by looping

The finding that the cRE bound by CTCF is necessary for the promoting the binding of YY1 near the promoter prompted us to explore the underlying mechanism. It has long been known that a cRE such as an enhancer element functions by physically contacting the corresponding gene promoter, thereby activating its transcription. Such *cis* interaction is manifested by the proteins bound to these regulatory regions, and CTCF is a well-studied organizer in this context.

The chromatin interaction map of the *XIST* locus in presence and absence of CTCF is presented in **Figure 7A**. The HiC Analysis of the entire XIC locus was carried out using the HiCPro pipeline using data from the control HEK-293T cells and upon knockdown of CTCF. The contact matrix was plotted for the XIC using the script HiC-Plotter. The inset of the location of the *XIST* gene was plotted separately using the same plotting tool. Differences between local contact frequencies were determined using the 'compare' function of HiCplotter. The HiC analysis revealed no global differences in the chromatin organization of the *XIST* locus in HEK293T cells. Nevertheless, upon CTCF knockdown, a very small change was observed in

the region in the vicinity of the cRE (Fig. 7A, marked by asterix). This is indicative of CTCF having a very specific and localized effect on the cRE that seems to govern *XIST* regulation in *cis*. The presence or absence of CTCF at the cRE could dictate its engagement with the promoter of *XIST* without affecting the organization of the locus at large.

To test this possibility further we carried out chromosome conformation capture (3C) assay (Dekker et al., 2002) using libraries prepared from HEK 293T, CTCF depleted HEK 293T cells and SOX2 overexpressing HEK 293T cells. Strategy for the assay marking CTCF peaks, restriction enzyme site and the primers is depicted schematically in **Figure 7B**. We scored for three specific interactions between the regions near the three CTCF binding sites (i) cRE (+4.5 Kb) and promoter-proximal regions (+1.5 Kb) '**a**' (ii) cRE (+4.5 Kb) and promoter-distal sites (-3.0 Kb) '**b**' and (iii) promoter-proximal (+1.5 Kb) and promoter distal regions (-3.0 Kb) '**c**'. Using the specific primer pairs we monitored the possible chromatin looping events by performing quantitative real time PCR. We observed that all the tested interactions occur at relatively higher frequencies in control HEK 293T cells compared to the conditions in which either CTCF was depleted or SOX2 was overexpressed **(Fig. 7C, Supp Fig. 7A, B, C)**.

Thus, it can be stated that the *XIST* locus from Xi adopts a conformation wherein (i) cRE bound by CTCF (+4.5 Kb) interacts with the promoter-proximal (+1.5 Kb) as well as the distal CTCF (-3.0 Kb) sites, and (ii) promoter-proximal (+1.5 Kb) and promoter-distal (-3.0 Kb) CTCF sites also interact with each other **(Fig. 7C)**. Based on these results, it can be speculated that such a conformation allows for the establishment of appropriate active histone marks (H3K27ac) uniformly **(Fig. 4A, Supp Fig. 4C,D)**, thereby ensuring persistent expression of XIST during the maintenance phase of XCI.

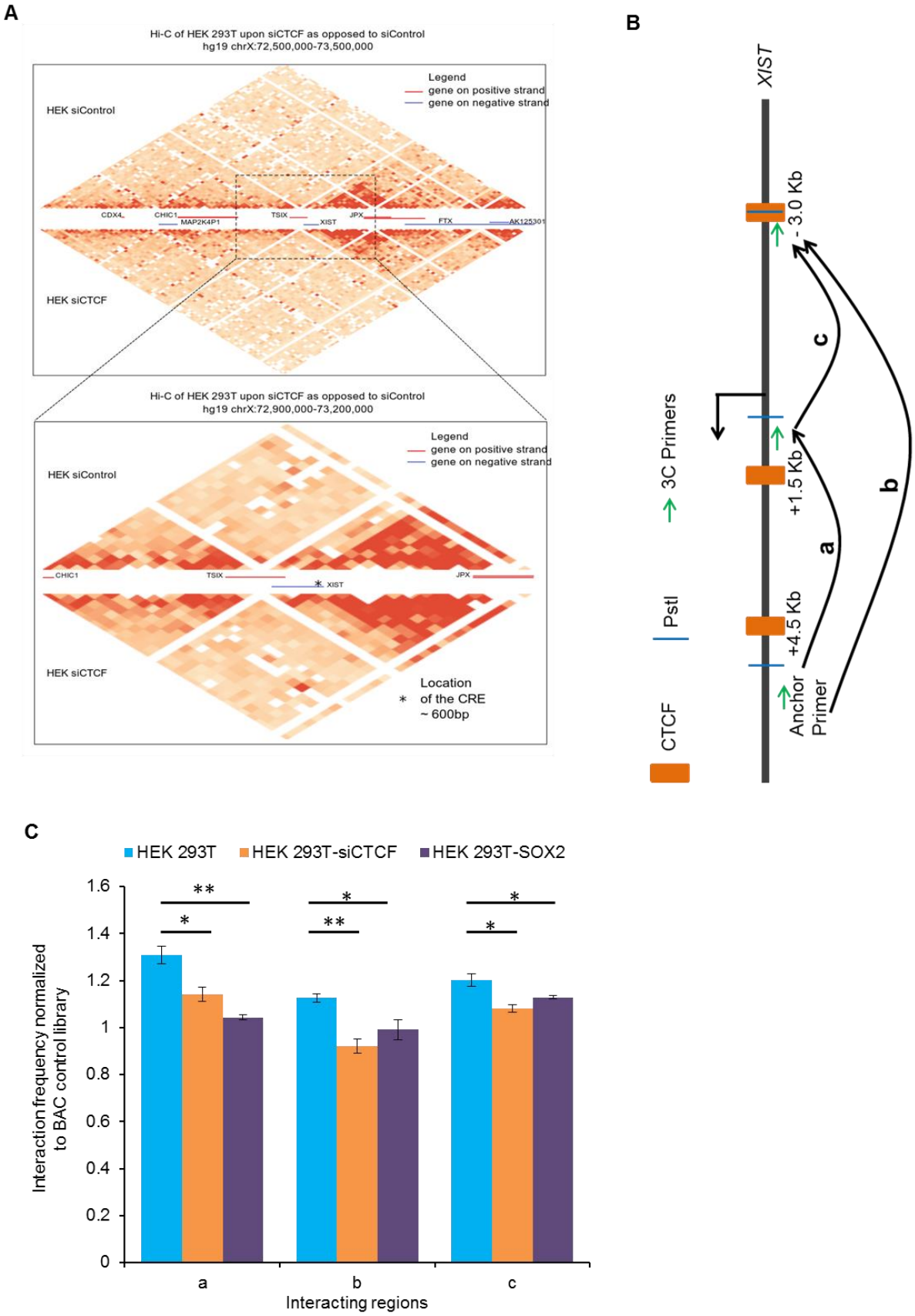

A
Hi-C of HEK 293T upon siCTCF as opposed to siControl
hg19 chrX:72,500,000-73,500,000
Legend
gene on positive strand
gene on negative strand
HEK siControl
HEK siCTCF
CDX4
CHIC1
MAP2K4P1
TSIX
XIST
JPX
FTX
AK125301
Hi-C of HEK 293T upon siCTCF as opposed to siControl
hg19 chrX:72,900,000-73,200,000
Location
of the CRE
~ 600bp
B
XIST
- 3.0 Kb
+1.5 Kb
+4.5 Kb
3C Primers
PstI
CTCF
Anchor
Primer
a
b
c
C
HEK 293T
HEK 293T-siCTCF
HEK 293T-SOX2
Interaction frequency normalized
to BAC control library
Interacting regions

Figure 7. CTCF mediates interaction between the cRE and the promoter-proximal and distal regions of *XIST* by looping. (A) HiC analysis for the XIC locus was performed in sicontrol (top triangle) and siCTCF (bottom triangle) transfected HEK293 cells using the datasets described in Materials and Methods. The HiC interaction matrix for the *XIST* locus on HEK293T cells upon siCTCF denotes no large scale changes in chromatin organization but for the location of the newly identified cRE as indicated by asterix. This is indicative of a very specific role of CTCF occupying the cRE and affecting XIST expression. (B) Strategy for 3C assay depicting position of CTCF binding sites (+4.5 Kb, +1.5 Kb and -3.0 Kb), PstI restriction enzyme sites and location of primers. Curved arrows denote the putative interacting regions, where a = chrX:73066941-73071641, b = chrX:73066941-73075470 and c = chrX:73071641-73075470. (C) Relative interaction frequencies between the cRE and promoter-proximal (a), cRE and promoter-distal (b) and promoter proximal to promoter distal (c) in control (blue bar), CTCF depleted (brown bar) and SOX2 overexpressed (violet bar) HEK 293T cells. Each bar indicates values from three biological replicate. Asterisks represent the significance over HEK 293T control (1st bar) as per Student's T-test (**P value < 0.01,*P value < 0.05).

The identified *cis*-regulatory element dictates *XIST* transcriptional status

To provide a clearer evidence for the functionality of cRE, we deleted the cRE in NT2/D1 cells using the CRISPR-Cas9 methodology. The strategy employed is depicted in **Fig. 8A**. Since, NT2/D1 cells require paracrine signalling to maintain their pluripotent state, it was not possible to isolate single cell clones so as to get a pure population harboring the cRE deletion. Therefore, we worked with a heterogeneous population containing cRE deleted cells **(Fig. 8B)** and performed qRT-PCR **(Fig. 8C)** as well RNA FISH for XIST **(Fig. 8D)**. Strikingly, although the mutant cells are less compared to the WT cells **(as per Fig. 8B)**, we observed a remarkable increase in XIST expression in undifferentiated heterogeneous population of NT2/D1 cells harboring the cRE deletion **(Fig. 8C,D)**. Such increase could potentially be the consequence of removal of the element that is repressed by the pluripotency factors in NT2/D1 cells. As a result, the pluripotency factors are unable to maintain XIST repressed in the fraction of cells with the cRE deletion, leading to the dramatic upregulation of *XIST*.

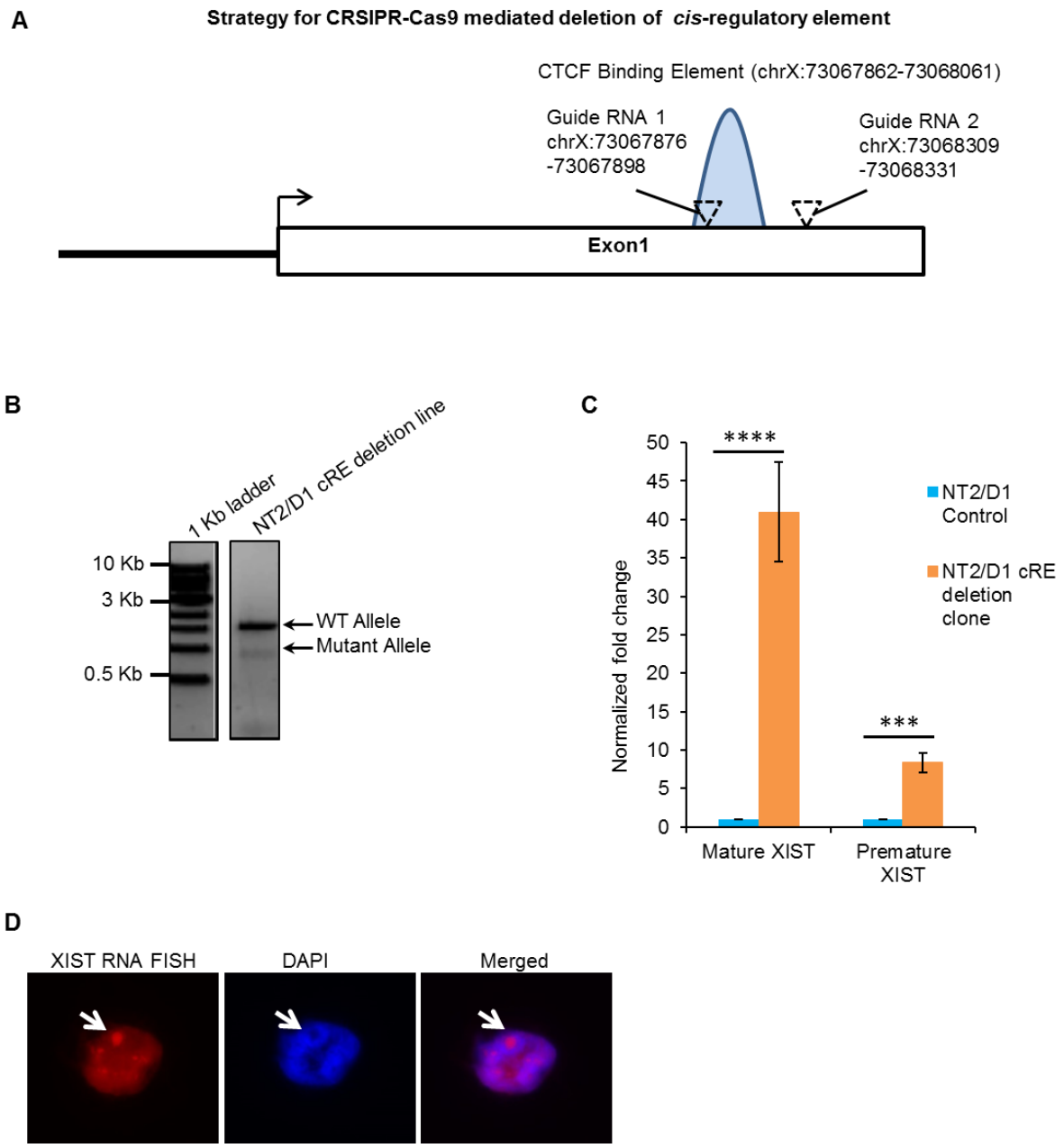


Figure 8. Deletion of cRE in NT2/D1 cells leads to induction of XIST. (A) Schematic depicting the strategy for CRISPR-Cas9 mediated deletion of cRE (+4.5 Kb) in NT2/D1 cells. (B) PCR based screening of deletion for NT2/D1. As expected the wild type allele shows a 1.4 Kb band whereas the mutant allele shows 0.9 Kb band. (C) qRT-PCR for mature and premature XIST in the heterogeneous population of NT2/D1 cells harbouring deletion of cRE. X-axis represents the mature and premature XIST transcripts and Y-axis represent the fold change normalized to 18s rRNA. Each bar represents values from 4 independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, ***P value < 0.001). (D) RNA FISH for XIST in undifferentiated heterogeneous population of NT2/D1 cells harboring the cRE deletion.

The corresponding conserved sequence in mouse exhibits histone modification profile similar to human

Next, we sought to determine whether the cRE discovered in this study is specific just to the primates or conserved between human and mouse. To consider the possibility of conservation of this CTCF site and CTCF binding that is related to XIST expression, we analyzed the sequence for the *XIST* promoter and 5 Kb flanking region for the presence of conserved sequences. In light of the poor sequence conservation of the *XIST* locus (Brockdorff et al., 1992; Hendrich et al., 1993; Nesterova et al., 2001), any conservation of a small stretch of sequence could be indicative of conserved regulatory element. To address this possibility, ECR browser (Ovcharenko et al., 2004) was used to identify such sequence(s) across multiple mammalian genomes. We observed that this cRE was conserved in Macaque, Cow and Dog genomes but no overlapping ECR was found in Rodent genomes **(Fig. 9A)**. We also performed a dot-plot analysis comparing the specific sequences between mouse and human using the online tool DotMatcher. The sequences that contained the peak of interest on the human genome and the ECR that is closest to this location were used. This encompasses a region of ~ 300 bp with the peak of interest at the 5' region of the sequences chosen. This analysis also confirmed the presence of an ECR **(Fig. 9B)**. Next we analyzed enrichment of histone marks and CTCF occupancy at the ECR in mouse, we observed a significantly higher enrichment for H3K4me1, H3K27ac, H3K4me3 as well as CTCF and YY1 specifically in female MEFs (as identified by PCR, **Supp Fig. 8 A,B**). Notably, the corresponding region in the mouse genome has been proposed as an alternative promoter in the mouse system (Johnston et al., 1998) and is reportedly bound by CTCF (Navarro et al., 2006).

The occurrence of enrichment of histone marks and CTCF, YY1 that corresponds to a potential regulatory element are present in two species of mammals and thus provide evidence to support the role of this region as a cRE even in the mouse system. A previous report (Navarro et al., 2006) demonstrates the enrichment of CTCF and H3K4me3 on a region of mouse *Xist* locus that corroborates the ECR analysis carried out in our study. Thus

we hypothesize the presence of a cRE that is involved in the regulation of both human and mouse Xist.

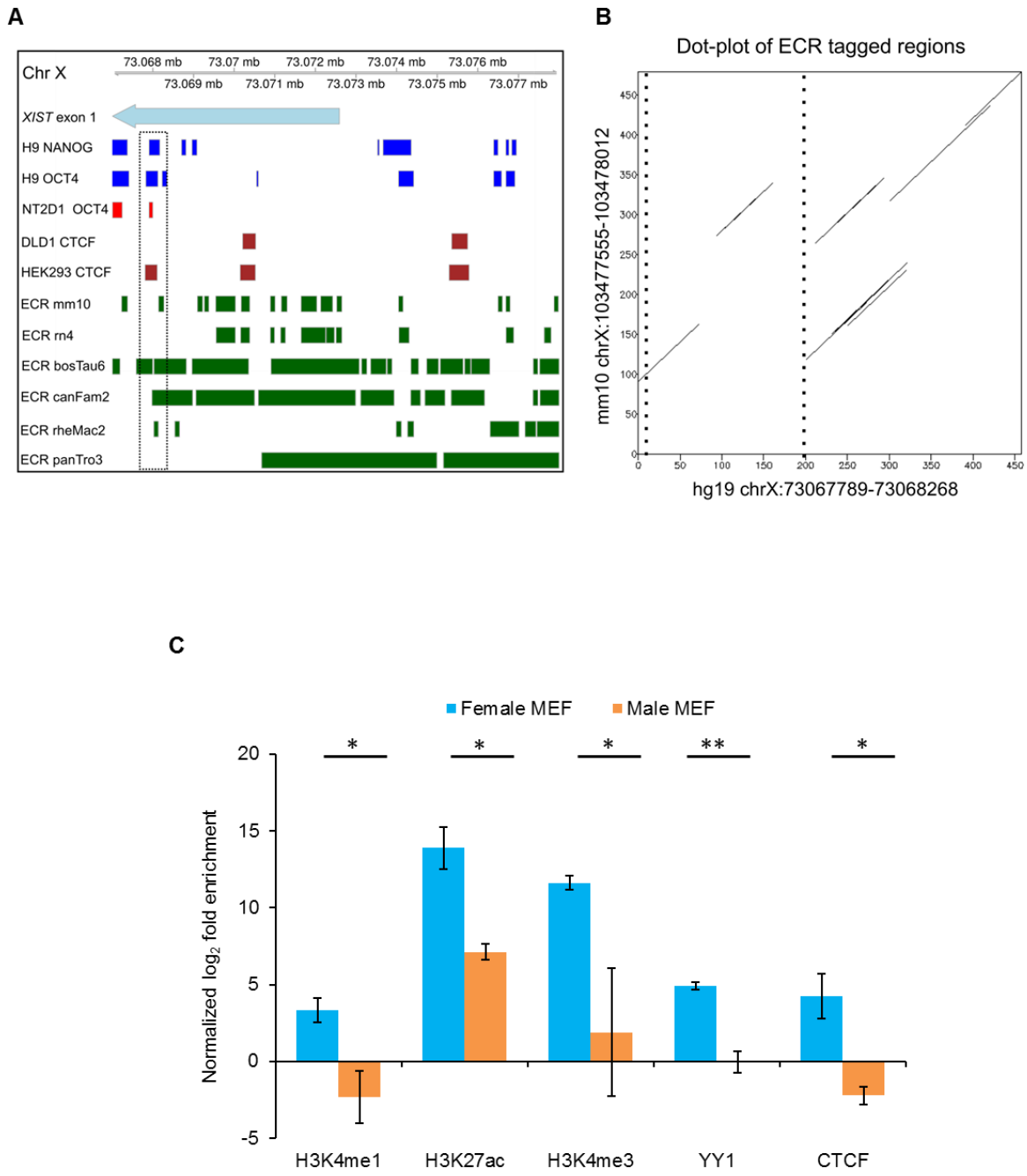


Figure 9. The newly identified cRE is evolutionarily conserved between humans and mouse. (A) Sequence similarity between various mammalian species of the *XIST* promoter and flanking regions analyzed using the ECR browser. The figure shows significant conservation at or near the CTCF site in *Bos taurus* (cow), *Canis familiaris* (Dog) and *Macaca mulatta* (Rhesus macaque). The presence of an ECR that is abutting the cRE in *Mus musculus* is of note. This region has been probed for the binding of histones and factors described in Fig. 8b. (B) Dot-plot showing the conservation of the cRE between human and mouse. The DNA sequence conservation was

plotted as described in 'Materials and Methods'.(C) ChIP-qPCR for H3K4me1, H3K27ac, H3K4me3, YY1 and CTCF for male and female MEFs. X-axis represents the antibodies used for ChIP and Y-axis represents the normalized log_2fold enrichment over IgG. Each bar represents values from 2 independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over male MEF ChIP as per Student's T-test (**P value < 0.01, *P value < 0.05).

Discussion

XIST is the critical regulator of the phenomenon of XCI, a mechanism to correct for the gene dosage imbalance between males and females of the placental mammals (Brockdorff et al., 1991; Brown et al., 1991; Borsani, et al., 1991; Lyon, 1961). Being central to the process of XCI, it is essential to modulate XIST levels temporally to ensure the faithful execution of the process and hence the developmental programs. Despite significant conservation of X-linked genes, chromosomal synteny as well as the process of XCI guided by XIST lncRNA between several eutherian mammals, XCI is initiated in diverse ways in different species. One of the possible explanations could be evolutionary alterations in the developmental programs across species necessitating robustness in the regulation of XCI to accommodate these changes (Okamoto et al., 2011). Extensive studies using mouse system have highlighted the multiple ways in which the transcription of *Xist* is regulated during the initiation and the maintenance phases of XCI. However, only YY1 has been discovered to be the common factor regulating *XIST* in mouse and human (Chapman et al., 2014; Makhlouf et al., 2014). Here, we have uncovered a complex regulatory pathway involving the pluripotency factors, YY1 and CTCF that shapes the transcriptional outcome from the *XIST* promoter **(Fig. 10)**.

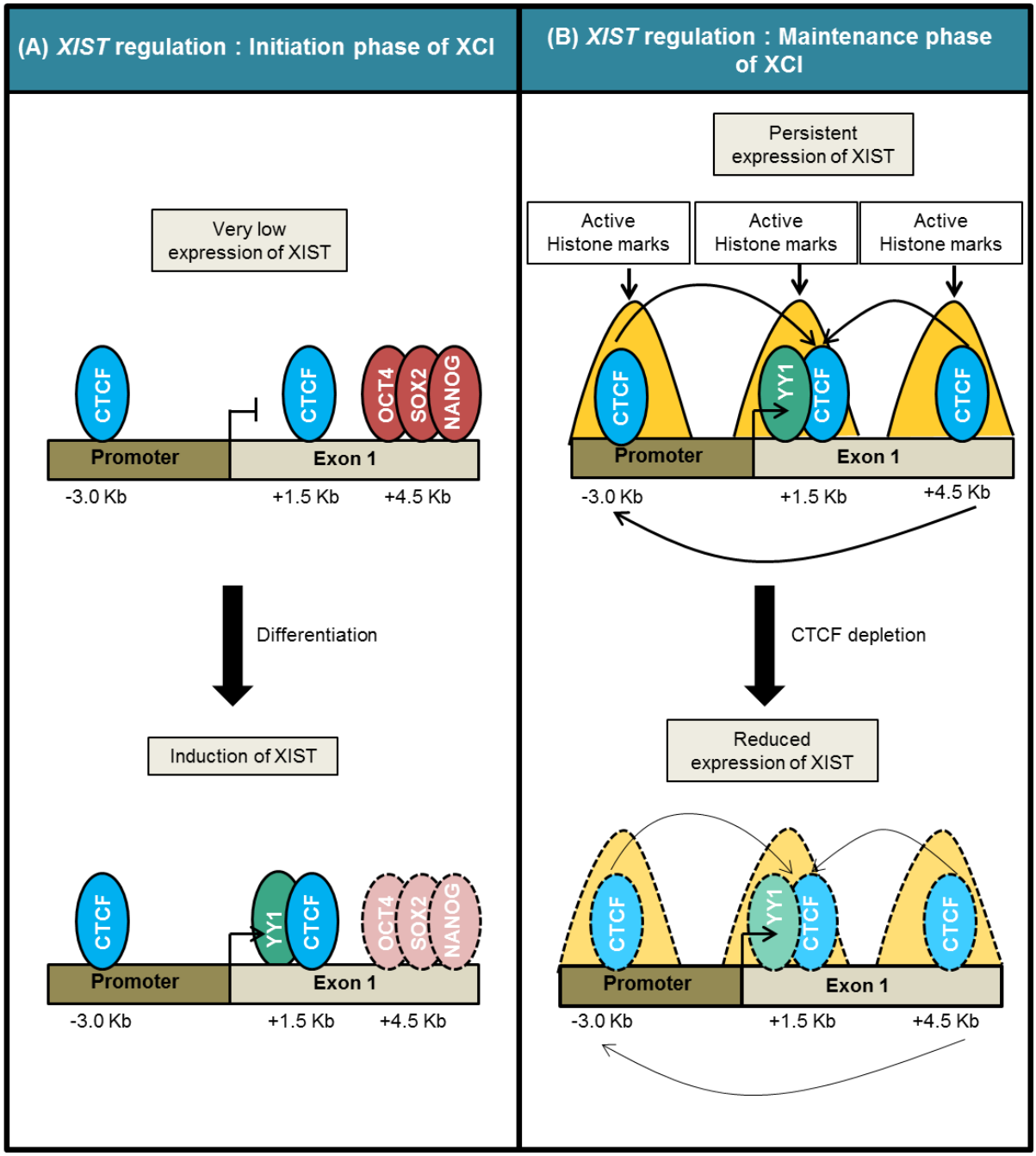


Figure 10. A model illustrating the role of CTCF bound cRE in dictating the transcription from *XIST* promoter. (A) In undifferentiated ES cells, the cRE of *XIST* remains bound by the pluripotency factors (OCT4, SOX2, NANOG) Upon differentiation, enrichment of pluripotency factors on the cRE decrease which is now occupied by CTCF. Subsequently, YY1 also occupies the promoter leading to the upregulation of XIST. (B) In differentiated cells, CTCF bound to the 3 sites mediates interaction between (i) cRE and the promoter-proximal regions (ii) cRE and promoter-distal regions as well as (iii) promoter-proximal –and distal CTCF sites. As a result, there is a consistent H3K27ac mark on all the three sites leading to persistent expression of XIST in the maintenance phase of XCI. However, upon CTCF depletion, the relative frequencies of all the three interactions are decreased. This disrupts

the loop architecture possibly causing a reduction in H3K27ac and YY1 enrichment on the indicated regions and a concomitant increase in H3K27me3 at all the three sites. This effectively leads to decrease in the expression of XIST. The thickness of the arrows signifies the frequency of interactions. Broken lines in association with the lighter colour shade indicate loss of enrichment.

Unlike mouse ES cells, the human counterpart has not served as a tractable model to address the aspects of XCI and *XIST* regulation. Multiple groups have reported that XIST expression, as well as XCI status shows considerable variation between human ES lines over passages (Bruck et al., 2013; Silva et al., 2008). However, some of the recent studies describe a method of culturing human ES cell lines, exhibiting consistent XIST expression as well as XCI upon differentiation (Lengner et al., 2010; Patel et al., 2017). The difficulty of human ES cells as a model notwithstanding, we have worked with NT2/D1 and HEK 293T as models of initiation and maintenance phases of XCI respectively. Moreover, we have reanalysed the data relevant to our study from multiple cell types including a human ES line and primary tissues of the female and male origin to discover physiologically relevant Xi-specific factors contributing to *XIST* regulation.

To address *XIST* regulation in the context of initiation phase of XCI, we have extensively worked with NT2/D1 cells, human male embryonic carcinoma cell line. NT2/D1 cell line provides the context for the initiation phase of XCI since it harbors properties similar to the embryonic stem cells and has been extensively used to understand the human stem cell biology (Andrews et al., 1980; Lee and Andrews, 1986; Skotheim et al., 2005). More importantly, it also expresses low levels of XIST (Chow et al., 2003) analogous to that seen for human and mouse ES cells as well blastocyst (Kay et al., 1993; Ray et al., 1997; Sheardown et al., 1997; van den Berg et al., 2009; Okamoto et al., 2011; Sousa *et al*., 2018). We show for the first time that XIST expression can be induced upon RA-mediated differentiation of these cells and hence, it can serve as a model to determine the induction of *XIST* **(Fig. 1D-F)**. To further characterize this behaviour, the identified promoter elements were subjected to the luciferase reporter assay upon RA-

mediated differentiation, which however, provided contrasting results **(Fig. 2C)** It was observed that while XIST expressed in induced upon differentiation of NT2/D1 cells **(Fig. 1D-F)**, the promoter activities showed a downward trend **(Fig. 2C)**. The latter result correlated with the levels of YY1 decreasing **(Fig. 2C,D)**, thereby corroborating the previously reported role of YY1 in positively regulating the *XIST* promoter (Hendrich et al., 1997). These counterintuitive observations suggested that the induction of *XIST* could be governed by previously unidentified regulatory region(s). The fact that there is a negative correlation between *XIST* induction and loss of OCT4, SOX2 and NANOG during differentiation, prompted us to speculate their role in governing XIST transcription. Moreover, there have been studies reporting pluripotency factors to be important positive and negative regulators of mouse *Xist* and *Tsix* respectively (Donohoe et al., 2009; Navarro et al., 2008, 2010). Upon reanalysing published ChIP-sequencing datasets **(Supp Fig. 2A)** and the experimental validation, we observed consistent enrichment of the pluripotency factors at +4.5 Kb region in differentiation-dependent manner **(Fig. 2E).** Furthermore, we demonstrate that the pluripotency factors can bind the +4.5 Kb (as measured from TSS) site even in the cells which have either low (4 day differentiated NT2/D1) **(Fig. 3G)** or consistent expression of XIST (HEK 293T) **(Fig. 3F)** and downregulate XIST transcriptionally **(Fig. 3B,E)**. Interestingly, although levels of YY1 decline during differentiation, its occupancy on the previously reported promoter-proximal site (+1.5 Kb) increases, causing a significant increment in XIST expression by day 5 of differentiation **(Fig. 1D, Fig. 2F)** (Makhlouf *et al*., 2014, Chapman *et al*., 2014). We think that a regulation mediated by the pluripotency network could be an essential way to control upregulation of XIST temporally. These findings provide compelling evidence suggestive of this gene body regulatory region to be a potential Xi-specific *cis* regulatory element. Indeed, a higher enrichment of activation associated histone marks – H3K4me1, H3K27ac **(Fig. 4A)**, as well as the known chromatin organizer protein CTCF **(Fig. 4E)** on the region of interest specifically in the female and not in the male cells strongly suggests it to be the case.

Role of CTCF in *XIST* regulation was ruled out based on the observation that it does not exhibit a differential enrichment between male and female fibroblasts or XIST^{+} and XIST^{-} human ES cells (Makhlouf et al., 2014). CTCF occupancies on the sites identified in this study are identical irrespective of male and female cells, a finding also corroborated by our experiments **(Supp Fig. 4C, D)**. However, the results obtained in our study unequivocally demonstrate it to be specifically enriched on the cRE only on Xi or XIST-induced cells **(Fig. 4E (i-iii))**. Besides this, it has been demonstrated to be enriched on the TSS of *XIST*/*Xist* specifically in the female cells or embryos (Pugacheva et al., 2005) and we also observe it to be the case in the ChIP experiment performed in the male and female cells (data not shown). Moreover, it is possibly required for the higher enrichment of H3K27ac not just on the cRE **(Fig. 5D)** but even on the promoter-proximal (+1.5 Kb) as well as distal (-3.0 Kb) CTCF binding sites **(Supp Fig. 6A, B)**. This is consistent with the evidence of CTCF playing the causal role in erasing H3K27me3 mark and making the chromatin more accessible (Weth et al., 2014) and mediating the chromatin interactions between these sites as verified by 3C-based assay (Fig. 7C). Consequently, YY1 binding to the promoter-proximal region is dependent on the CTCF binding to the cRE **(Fig. 6A)**. Furthermore, all the 3 CTCF sites (+4.5 Kb – cRE, +1.5 Kb- promoter proximal and -3.0 Kb – promoter distal sites) contact each other, thereby converging at the promoter and most likely increasing the transcription from the TSS facilitating persistent expression of XIST during the maintenance phase of XCI (Fig. 7).

Additionally, by performing knockdown and over-expression experiments, we show that XIST is positively regulated by CTCF in addition to YY1 **(Fig. 4D, Supp Fig. 5B)**. Interestingly, XIST levels increase dramatically when the over-expression of YY1 or CTCF is coupled with the depletion of OCT4 and SOX2 in NT2/D1 cells suggesting that the balance of activators (YY1 and CTCF) and repressors (pluripotency factors) determines *XIST* transcription suggestive of competition to bind cRE. **(Fig. 5A,B)**. Indeed we show that CTCF bound to cRE is displaced upon overexpression of SOX2 in HEK 293T cells as revealed by ChIP **(Fig. 5C)**. Most notably, we report a dramatic

upregulation in XIST expression as determined by qRT-PCR and RNA FISH upon CRISPR-Cas9 mediated deletion of cRE in undifferentiated NT2/D1 cells (**Fig. 8**). Therefore, it can be proposed that at the onset of XCI, the cRE is kept repressed by remaining bound to the pluripotency factors. However, in the differentiated cells which do not express pluripotency factors, the cRE on the Xi is now bound by CTCF and enriched for H3K27ac, whereas it is enriched for H3K27me3 on the active X **(Fig. 10 legend)**. Consequently, XIST continues to be synthesized from the Xi and is involved in maintaining silencing. Whether the inconsistency in XIST expression observed in the various female human ES lines can plausibly be a consequence of the variability in the cRE function as dependent on occupancies of pluripotent factors or CTCF would be a promising line of future investigation.

Interestingly, the cRE seems to be conserved between human and mouse **(Fig. 9A)**. This is particularly important given that *Tsix*, a critical negative regulator of *Xist* in mouse shows poor conservation across species. A recent study identified another potential and evolutionarily conserved modulator of *Xist*, called *Linx*. The report highlights the contextual role played by *Linx* as a silencer of Xist when in its original location in *Tsix* TAD and as an enhancer when artificially placed in *Xist* TAD (Galupa *et al*., 2020)

The conserved cRE identified in mouse is actually designated as the P2 promoter that has been demonstrated to bind YY1 and CTCF (Johnston et al., 1998; Makhlouf et al., 2014; Navarro et al., 2006; Sun et al., 2013). As opposed to our observation of female-specific enrichment of CTCF in mouse (Xi) **(Fig. 9C)**, a previous study found it to be enriched on Xa and hence, acting as a repressor (Sun et al., 2013). Interestingly, the authors in this study speculate a looping between the P2 and the P1 promoters based on their observation of CTCF binding to the P1 promoter. The epigenetic characterization of the mouse *Xist* P2 promoter in our study shows that it is enriched with activate histone marks such as H3K4me1, H3K27ac and H3K4me3 **(Fig. 9C)**.

In conclusion, we identify a novel and evolutionarily conserved cRE for *XIST*/*Xist* that is controlled by developmentally regulated factors, OCT4,

SOX2 and NANOG as the repressors and the global chromatin organizer CTCF as an activator. Since all the identified regulators are expressed to a similar extent in male and female cells, it is intriguing that they are able to regulate the *XIST* expression in Xi-specific manner. This is suggestive of additional factors/ mechanisms contributing towards such specificity, which deserves further investigation. Determining if this feature is conserved in other eutherians as well will provide useful clues regarding the conserved mechanisms governing *XIST* transcription and provide key insights into evolutionary conservation of the phenomenon of X inactivation.

Materials and methods

Cell Culture

Human embryonic carcinoma cell line NT2/D1 (NTERA2-clone D1) was obtained as a kind gift from Dr. Peter Andrews, University of Sheffield, UK. They were grown in Dulbecco's Modified Eagle's Medium with sodium pyruvate, high glucose (DMEM, Sigma-Aldrich, St. Louis, Missouri, USA) supplemented with 10% foetal bovine serum (Invitrogen, Carlsbad, California, USA), 2 mM L-glutamine (Invitrogen, Carlsbad, California, USA) and penicillin-streptomycin (Invitrogen, Carlsbad, California, USA) and maintained at 37°C under 5% CO2 atmosphere. NT2/D1 cells were passaged upon reaching 70% confluency by gentle scraping and trypsin was not used for passaging. Human embryonic kidney cells (HEK 293T) were grown in DMEM (Invitrogen, Carlsbad, California, USA) without sodium pyruvate, high glucose, supplemented with 10% foetal bovine serum (Invitrogen, Carlsbad, California, USA) and penicillin-streptomycin (Invitrogen, Carlsbad, California, USA) and maintained at 37°C under 5% CO2 atmosphere. HEK 293T cells were maintained passaged upon reaching 70-80% confluency using 0.05% trypsin (Invitrogen, Carlsbad, California, USA).

RA-mediated differentiation of NT2/D1 cells

All-trans-retinoic acid (RA) (procured from Sigma-Aldrich, St. Louis, Missouri, USA) used for inducing differentiation of NT2/D1 cells was reconstituted at a concentration of 5 mg/ml in DMSO (Sigma-Aldrich, St. Louis, Missouri, USA)

and stored at -80°C. For differentiation experiments, NT2/D1 cells were harvested using 0.05% trypsin, resuspended in fresh medium and seeded at a density of 0.15 X 10^6 cells in 6-well plate or 1 X 10^6 cells in 100 mm tissue culture dish (Corning, New York, USA). Cells were allowed to grow for 24 hours following which few cells were harvested for RNA and protein extractions or crosslinked for ChIP as 0 day control. RA was added to the remaining wells/plates at a concentration of 13.7 μM for the remaining 6 days. Each day cells were either replenished with fresh medium and RA or harvested as day 1/2/3/4/5/6 samples.

Isolation of MEFs

C56BL/6 mice were crossed and E13.5 day embryos dissected after sacrificing the pregnant females. Individual embryos were dissected out from the uterine horns and the head and the foetal liver was completely removed. The remaining tissue was minced with sterile scissors in the presence of 0.25% Trypsin EDTA. The minced tissue was plated in 100 mm dishes in DMEM supplemented with 10% FBS and Non-Essential Amino Acids (NEAA 100x Invitrogen). After 48 hours the fibroblasts were expanded into 100 mm dishes and used for further experiments.

Molecular Cloning

To characterize *XIST* promoter, genomic regions ranging from +50 bp to -4408 bp upstream of *XIST* TSS were cloned into a promoter- less reporter vector – pGL3 Basic procured from Promega (Wisconsin, Madison, USA). *XIST* genomic regions were PCR amplified from the genomic DNA extracted from NT2/D1cells using specific primer pairs (Supp Table 1) followed by restriction enzyme digestion of the vector and insert and ligation using T4 DNA ligase (New England Biolabs, MA, USA). The positive clones were confirmed by sequencing.

Transfection of DNA and siRNA

NT2/D1 or HEK 293T cells were transfected with the equimolar of DNA constructs using Lipofectamine 2000 and siRNA using Lipofectamine

RNAiMax (Invitrogen, Carlsbad, California, USA) as per manufacturer's guidelines.

Luciferase Reporter Assay

Luciferase reporter assays were performed using the Dual luciferase assay kit from Promega (Madison, Wisconsin, USA). Cells were transfected with Firefly luciferase and Renilla luciferase DNA constructs. After harvesting, the cells were lysed using 1X Passive lysis buffer (Promega, Madison, Wisconsin, USA) as per manufacturer's instructions. The lysates and substrates were mixed in the optical bottom 96-well plate (ThermoFisher Scientific, Waltham, Massachusetts, United States) according to the guidelines provided by Promega. The reporter activities were measured using luminometry mode on the Varioskan machine (ThermoFisher Scientific, Waltham, Massachusetts, United States). In all the assays, Renilla luciferase activity measurement served as an internal control. The fold change was calculated with respect to either the vector control or 0 day control as and when mentioned.

RNA extraction and cDNA synthesis

Total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, California, USA). 1 to 2 µg of RNA was used for DNase treatment (Promega, Madison, Wisconsin, USA) as per the protocol provided by the manufacturer. This was followed by cDNA synthesis using reverse transcriptase kits from Applied Biosystems (High capacity cDNA synthesis kit) (Foster City, California, USA). Additionally, minus RT control was set up to verify the efficiency of DNase treatment. The synthesized cDNA was used to set up quantitative real-time PCR (qRT-PCR).

Quantitative real-time PCR

Quantitative real time PCR was performed using SYBR green chemistry (Roche) with specific set of primer pairs (Supp Table 1) on the Applied Biosystems ViiA7 thermal cycler. Changes in threshold cycles were calculated by subtracting the Ct values of the gene of interest from that of housekeeping control (for qRT-PCR) [Ct(target genes) – Ct(18s rRNA)]. ΔCt

values of specific target genes from the experimental samples were then subtracted from their respective control samples to generate ΔΔCt values. The fold changes were calculated using the formula : 2^(-(ΔΔCt value)). For quantification after ChIP, DNA recovered post ChIP and %Input are used as templates for the PCR with specific set of primers. For quantification of enrichment, the efficiency of chromatin immunoprecipitation of particular genomic locus was calculated as follows :

% (ChIP/ Total input)= 2^[(Ct(x%input) – log(x%)/log2) - Ct(ChIP)]x 100%

Relative occupancy was calculated as a ratio of specific signal over background: Occupancy= % input (specific loci) / % input (background loci)

Protein extraction and immunoblotting

Cell pellets were resuspended in RIPA buffer (10 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0), 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCl) containing 1X protease inhibitors (procured from Roche, Basel, Switzerland) and lysed by repeated freeze-thaw cycles. The lysates were centrifuged at 14000 rpm, 4°C, 30 minutes to eliminate the cellular debris. The supernatant was collected in the fresh microfuge tube. The concentrations of protein were determined by performing BCA assay (purchased from ThermoFisher Scientific, Waltham, MA, USA). Equal amounts of protein lysates were boiled in 1X Laemmli buffer (0.5 M Tris-HCl pH 6.8, 28% glycerol, 9% SDS, 5% 2-mercaptoethanol, 0.01% bromophenol blue) for 10-15 minutes and subjected to electrophoresis on a polyacrylamide gel. The separated proteins were transferred onto PVDF membrane (Millipore, Billerica, Massachusetts, USA) using phosphate based transfer buffer (10 mM sodium phosphate monobasic, 10 mM sodium phosphate dibasic) at 4°C, 600 mA, 2 hours. After the completion of transfer, membranes were blocked in 5% skimmed milk, incubated overnight at 4°C with the primary antibodies prepared in 5% BSA. The membranes were washed thrice with the buffer containing 20 mM Tris buffer pH 7.4, 500 mM NaCl and 0.1% tween 20 (TST) the next day and incubated with the appropriate secondary antibodies conjugated with horseradish peroxidase for an hour at room temperature. Following this, the membranes were again washed thrice with TST buffer. The blots were developed using Immobilon

Western Chemiluminescent HRP Substrate (Millipore, Billerica, MA, USA) and detected using ImageQuant LAS 4000 (GE Healthcare, Piscataway, NJ, USA) according to the manufacturer's instructions.

Chromatin immunoprecipitation

ChIP was performed as per X-ChIP protocol. Briefly, 10 µg of sonicated chromatin (average length 150-400 bp) was incubated with 1 µg of specific antibody and incubated on an end-to-end rocker at 4°C overnight. Each ChIP was performed at least 2 to 3 times as indicated in the figure legends and quantitative real time PCR was set up. Primers used for the ChIP-PCR analysis are listed in the Supplementary Table1. Gel images for end-point PCR for all the ChIP experiments can be found in Supplementary figures 12 to 14.

Sequential ChIP assay

The method followed for the sequential ChIP assay was same as that described above. ChIP was started with 200 µg of chromatin and the 1st elution was done using the buffer containing 0.1 M $NaHCO_3$, 1% SDS and 10 mM DTT. The eluate was subjected to second ChIP with either Rabbit IgG or CTCF antibodies.

Chromosome conformation capture (3C) assay

The 3C assay was performed essentially as described in Naumova *et al*. and Hagège et al. (Hagege et al., 2007; Naumova et al., 2012). Briefly, 10^6 cells were crosslinked with 1% formaldehyde for 10 minutes, quenched using 125 mM glycine for 5 minutes, followed by washing with ice-cold PBS. Nuclei were isolated by incubating the cell pellet in the hypotonic buffer (10 mM Tris-Cl pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630) on ice for 15 minutes followed by 20 strokes of Dounce homogenizer (loose pestle). Nuclear pellet was washed with 1X CutSmart Buffer (NEB) twice and resuspended in 500 µl of 1.2X CutSmart Buffer. 0.3% SDS was added to remove uncrosslinked proteins and incubated at 37°C, 900 rpm, 1 hour. 2% Triton X-100 was added to quench the SDS and incubated at 37°C, 900 rpm, 1 hour. This was followed by addition of 1500 U of PstI-HF (NEB) and incubated at 37°C, 900

rpm, 16 hours. PstI-HF was inactivated by adding 1.6% SDS and incubating at 80°C for 30 minutes. This was followed by addition of the following ligation cocktail:

10X ligation buffer (500 mM Tris-Cl pH 8.0, 100 mM $MgCl_2$, 100 mM Dithiothreitol) – 800 µl

10% Triton X-100 – 800 µl

10 mg/ml BSA – 80 µl

100 mM ATP – 80 µl

H_2O – to 8000 µl

The mixture was incubated at 37°C, 8 rpm, 1 hour.

5000 U of T4 DNA Ligase (NEB) was then added to the ligation mix and incubated further for 6 to 8 hours at 16°C. This was followed by further incubation at 22°C for 1 hour. Ligated DNA was crosslinked by adding 300 µg of Proteinase K followed by incubation at 65°C, overnight. This was followed by RNaseA treatment (300 µg) at 37°C for 1 hour. 3C library was purified by Phenol:Chloroform:Isoamyl alcohol extraction and precipitated using Sodium acetate and absolute ethanol. The volume of the ligated DNA was increased to 2 times with TE to avoid precipitation of DTT. The purified 3C library was quantified and 300 ng (1), 150 ng (1/2), 75 ng (1/4), 37.5 ng (1/8) of DNA was used as a template for quantitative real-time PCR. The interaction frequencies for each primer pairs and experimental conditions were calculated by normalizing to the control library generated using BAC harboring XIST locus of interest (RP13-183A17). The BAC was procured from BACPAC Resources, Children's Hospital Oakland Research Institute (CHORI), Oakland, USA.

RNA FISH

The RNA FISH protocol followed was as described by Chaumeil et al. 2005. Briefly, the probe for RNA FISH was prepared with 2 µg of BAC RP183A17 using Nick translation kit (10976776001) purchased from Roche (Basel, Switzerland). Salmon sperm DNA and cot1 DNA were added to the probe to mask the repetitive sequence. Prior to hybridizations, the cells on the coverslips were washed with RNase-free PBS followed by permeabilization using freshly made Cystoskeletal buffer [10 mM PIPES (pH 7.0), 100 mM

NaCl, 300 mM Sucrose, 3 mM $MgCl_2$] containing 0.5% TritonX-100 for 4 to 5 minutes on ice. Nuclei are fixed using 4% paraformaldehyde for 10 minutes at room temperature. The coverslips were washed twice with 70% ethanol and then subjected to dehydration by sequentially incubating the coverslips in 80%, 95% and 100% ethanol for 3 minutes each, followed by air drying. The nuclei were then subjected to hybridization step in the dark and humidified chamber at 37°C for 24 hours. After incubation, the coverslips are washed thrice with 50% formamide + 2X SSC buffer (pH 7.2) for 5 minutes each at 42°C, followed by washing thrice with 2X SSC buffer for 5 minutes each at 42°C. Counterstained with DAPI and washed twice with 2X SSC buffer. The coverslips were then mounted on a clean slide for microscopic observation.

Antibodies, siRNAs and other reagents

SP1 (5931S) antibody for immunoblotting were procured from Cell Signaling Technology (Danvers, USA), YY1 (ab12132) antibodies for western blotting was purchased from Abcam (Cambridge, UK), SOX2 (AF2018) and NANOG (AF1997) antibodies for immunoblotting and ChIP were purchased from R & D (Menomonie, USA), OCT4 antibody (sc-9081) for western blotting and ChIP and goat-HRP secondary antibody were purchased from Santacruz Biotechnologies (Dallas, Texas, USA), CTCF antibody (07-729) procured from Millipore/Upstate (Billerica, Massachusetts, USA) was used for ChIP and CTCF antibody (sc-21298) purchased from Santacruz Biotechnologies (Dallas, Texas, USA) was used for immunoblotting, Normal rabbit IgG (12-370) and Normal mouse IgG (12-371) for ChIP were purchased from Millipore (Billerica, Massachusetts, USA), β-ACTIN (VMA00048) and γ-TUBULIN primary antibodies, mouse-HRP and rabbit-HRP secondary antibodies were purchased from Biorad Laboratories (Hercules, California, USA). Histone modification antibodies used for the ChIP were as follows - H3K4me (Millipore, 07-436), H3K27ac (Abcam, ab4729), H3K27me3 (Abcam, ab6002), H3K4me3 (Millipore, 17-614) and H3K9me3 (Millipore, 07-523).

Sequence analysis for ChIP-Seq

SRR/DRR/ERR IDs of the datasets used are provided.

Datasets used tabulated by cell line, the files can be downloaded from the link (https://www.ebi.ac.uk/ena) by providing the ids given below as search terms.

H9 cells: OCT4 (SRR5642847), Nanog (SRR5642845); HEK 293 cells: CTCF (DRR014670); DLD1 cells: CTCF (DRR014660); Female breast epithelium cells (ENCODE): CTCF (SRR6213076, Input SRR6213541); Male breast epithelium cells (ENCODE): CTCF (SRR6213724, Input SRR6214303); NT2/D1: OCT4 (SRR1640282) (This dataset is generated on the platform SOLiD3.0).

All Illumina datasets were aligned with bowtie2 (Langmead, et al. 2012) to the hg19 reference assembly. Following which macs2 (Zhang, et al. 2008) was used to identify significant peaks relative to control where available. The p value cut-off was set at 0.05, macs2 callpeak uses the *t* test to determine significance of the enrichment at a certain location over control. The resulting .narrowPeak files were used to plot the figures regarding ChIP Seq peaks using the GViz package in Rstudio. SOLiD 3.0 dataset for the NT2/D1 OCT4 ChIP-Seq was aligned to the hg19 reference genome using bowtie1 following which the downstream processing was as indicated above.

Hi-C Analysis

The Hi-C Analysis was carried out using the HiCPro (https://github.com/nservant/HiC-Pro) pipeline using the following datasets using the HEK-293T cells for Hi-C upon knockdown of CTCF. siControl (SRR710071, SRR710072, SRR710073, SRR710074, SRR710075, SRR710076, SRR710077) and siCTCF (SRR710078, SRR710079, SRR710080, SRR710081, SRR710082, SRR710083). Briefly, each fast file for the forward and reverse reads was independently aligned to the genome using bowtie2. The generated alignment files were then used as an input for calculating a contact matrix at resolutions of 5 Kb, 8 Kb and 10 Kb, The data depicted in figure 7A is at the 5 Kb resolution. The contact matrix was plotted

for the XIC using the script HiC-Plotter (https://github.com/kcakdemir/HiCPlotter). The inset of the location of the XIST gene was plotted separately using the same plotting tool. Differences between local contact frequencies were determined using the 'compare' function of HiCplotter. There were no differences observed in the entire region tested as is described in detail in the results section corresponding to Figure 7A.

Dot plot for evolutionarily conserved region (ECR) mapping

The online tool DotMatcher (http://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher) was used to construct the dot-plot. The sequences that contained the peak of interest on the human genome and the ECR that is closest to this location were used. This encompasses a region of ~ 300 bp with the peak of interest at the 5' region of the sequences chosen. There are no ECRs upstream to the region chosen and this sequence was not included in this analysis. The regions used were human - hg19 chrX:73067789-73068268, mouse - mm10 chrX:103477555-103478012. The window size for the sliding window was kept at 35 and the threshold for matching was 25.

Competing interests

The authors declare no conflict of interest.

Acknowledgements

Authors wish to thank Dr. Peter Andrews (University of Sheffield) for kindly gifting NT2/D1 cells, Drs. Kundan Sengupta and Ajay Labade (IISER Pune) for their help with the RNA FISH experiment, Mukul Rawat and Sneha Tripathi (IISER Pune) for helping with the imaging and analysis, Dr. Deepa Subramanyam (NCCS Pune) for MEFs. Work was supported by the grant from the Centre of Excellence in Epigenetics program of the Department of Biotechnology (BT/COE/34/SP17426/2016), Government of India to SG. RS was supported by the fellowship from the Council of Scientific and Industrial Research, India. AK was supported by the Wellcome Trust-DBT India Alliance Early Career Fellowship.

Author Contributions

RS and SG conceived the project and designed experiments. RS performed all experiments, interpreted data and wrote the manuscript. AK performed bioinformatics analysis and interpretations and also helped with acquisition of the RNA FISH images and manuscript writing. SG interpreted data, supervised the project and wrote the manuscript. All authors read and approved the final manuscript.

References

Andrews, P.W., Bronson, D.L., Benham, F., Strickland, S., and Knowles, B.B. (1980). A comparative study of eight cell lines derived from human testicular teratocarcinoma. Int. J. Cancer *26*, 269–280. DOI:https://doi.org/10.1002/ijc.2910260304

Borsani, G., Tonlorenzi, R. M., Simmler, C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H.F., Avner, P., and Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *354*, 56–58. DOI:https://doi.org/10.1038/351325a0

Brockdorff, N., and Turner, B.M. (2015). Dosage Compensation in Mammals. Cold Spring Harb. Perspect. Biol. *6*, 1–25. DOI:https://doi.org/10.1101/cshperspect.a019406

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*, 329–331. DOI:https://doi.org/10.1038/351329a0

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515–526. DOI:https://doi.org/10.1016/0092-8674(92)90519-I

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44. DOI:https://doi.org/10.1038/349038a0

Bruck, T., Yanuka, O., and Benvenisty, N. (2013). Human pluripotent stem cells with distinct X inactivation status show molecular and cellular differences controlled by the X-Linked ELK-1 gene. Cell Rep. *4*, 262–270.

DOI:https://doi.org/10.1016/j.celrep.2013.06.026

Chang, S.C., and Brown, C.J. (2010). Identification of regulatory elements flanking human XIST reveals species differences. BMC Molecular Biology *11*, 1-10. DOI: https://doi.org/10.1186/1471-2199-11-20

Chapman, A.G., Cotton, A.M., Kelsey, A.D., and Brown, C.J. (2014). Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. BMC Genet. *15*, 89. DOI:https://doi.org/10.1186/s12863-014-0089-4

Chow, J.C., Hall, L.L., Clemson, C.M., Lawrence, J.B., and Brown, C.J. (2003). Characterization of expression at the human XIST locus in somatic, embryonal carcinoma, and transgenic cell lines. Genomics *82*, 309–322. DOI:https://doi.org/10.1016/S0888-7543(03)00170-8

Chow, J.C., Hall, L.L., Baldry, S.E.L., Thorogood, N.P., Lawrence, J.B., and Brown, C.J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. Proc. Natl. Acad. Sci. U. S. A. *104*, 10104–10109. DOI:https://doi.org/10.1073/pnas.0610946104

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: Evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275. DOI:https://doi.org/10.1083/jcb.132.3.259

Cotton, A.M., Price, E.M., Jones, M.J., Balaton, B.P., Kobor, M.S., and Brown, C.J. (2015). Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum. Mol. Genet. *24*, 1528–1539. DOI:https://doi.org/10.1093/hmg/ddu564

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311. DOI:https://doi.org/10.1126/science.1067799

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature *460*, 128–132. DOI:https://doi.org/10.1038/nature08098

Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M., and Brockdorff, N. (1999). Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. Hum. Mol.Genet.*8*, 195–204. DOI:https://doi.org/10.1093/hmg/8.2.195

Galupa, R., Nora, E.P., Worsley-Hunt, R., Picard, C. Gard, C., van Bemmel, J.G., Servant, N., Zhan, Y., El Marjou, F., Johanneau, C., Diabangouaya, P., Le Saux, A., Lameiras, S., da Fonseca, J.P., Loos, F., Gribnau, J., Baulande, S., Ohler, U., Giorgetti, L., and Heard, E. (2020). A Conserved noncoding locus regulates random monoallelic Xist expression across a topological boundary. Molecular Cell 77, 1-16. DOI: https://doi.org/10.1016/j.molcel.2019.10.030

Gontan, C., Achame, E.M., Demmers, J., Barakat, T.S., Rentmeester, E., van IJcken, W., Grootegoed, J.A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature *485*, 386–390. DOI:https://doi.org/10.1038/nature11070

Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., De Laat, W., and Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3c-qpcr). Nat. Protoc. *2*, 1722–1733. DOI:https://doi.org/10.1038/nprot.2007.243

Hall, L.L., Clemson, C.M., Byron, M., Wydner, K., and Lawrence, J.B. (2002a). Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum. Mol. Genet. *11*, 3157–3165. DOI:https://doi.org/10.1093/hmg/11.25.3157

Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., and Lawrence, J.B. (2002b). An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. U. S. A. *99*, 8677–8682.
DOI:https://doi.org/10.1073/pnas.132468999

Heard, E., Mongelard, F., Arnaud, D., Chureau, C., Vourc ’h, C., and Avner, P. (1999). Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. Proc Natl Acad Sci U S A *96*, 6841–6846.
DOI:https://doi.org/10.1073/pnas.96.12.6841

Hendrich, B.D., Brown, C.J., and Willard, H.F. (1993). Evolutionary conservation of possible functional domains of the human and murine XIST genes. Hum. Mol. Genet. *2*, 663–672.
DOI:https://doi.org/10.1093/hmg/2.6.663

Hendrich, B.D., Plenge, R.M., and Willard, H.F. (1997). Identification and characterization of the human XIST gene promoter: Implications for models of X chromosome inactivation. Nucleic Acids Res. *25*, 2661–2671.
DOI: https://doi.org/10.1093/nar/25.13.2661

Herzing, L.B., Romer, J.T., Horn, J.M., and Ashworth, A. (1997). Xist has properties of the X-chromosome inactivation centre. Nature *386*, 272–275. DOI:https://doi.org/10.1038/386272a0

Hirka, G., Prakash, K., Kawashima, H., Plotkin, S.A., Andrews, P.W., and Gonczol1, E. (1991). Differentiation of Human Embryonal Carcinoma Cells Induces Human Immunodeficiency Virus Permissiveness Which Is Stimulated by Human Cytomegalovirus Coinfection. J. Virol. *65*, 2732–2735.

Johnston, C.M., Nesterova, T.B., Formstone, E.J., Newall, A.E.T., Duthie, S.M., Sheardown, S.A., and Brockdorff, N. (1998). Developmentally regulated Xist promoter switch mediates initiation of X inactivation. Cell *94*, 809–817. DOI:https://doi.org/10.1016/S0092-8674(00)81739-0

Kay, G.F., Penny, G.D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. (1993). Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. Cell *72*, 171–182.
DOI:https://doi.org/10.1016/0092-8674(93)90658-D

Langmead, B. and Salzberg,S.L.(2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357-359.
DOI: https://doi.org/10.1038/nmeth.1923

Lee, J.T., and Lu, N. (1999). Targeted Mutagenesis of Tsix Leads to Nonrandom X Inactivation. Cell *99*, 47–57.
DOI:https://doi.org/10.1016/S0092-8674(00)80061-6

Lee, V.M., and Andrews, P.W. (1986). Differentiation of NTERA-2 clonal human embryonal carcinoma cells into neurons involves the induction of all three neurofilament proteins. J. Neurosci. *6*, 514–521.
DOI:https://doi.org/10.1523/JNEUROSCI.06-02-00514.1986

Lee, J.T., Strauss, W.M., Dausman, J.A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell *86*, 83–94.
DOI:https://doi.org/10.1016/S0092-8674(00)80079-3

Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J., et al. (2010). Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell *141*, 872–883.
DOI:https://doi.org/10.1016/j.cell.2010.04.010

Li, Y., Tan, T., Zong, L., He, D., Tao, W., and Liang, Q. (2012). Study of methylation of histone H3 lysine 9 and H3 lysine 27 during X chromosome inactivation in three types of cells. Chromosom. Res. 20, 769–778.
DOI:https://doi.org/10.1007/s10577-012-9311-2

Lyon, M.F. (1961). Gene Action in the X-chromosome of the Mouse (Mus musculus L.). Nature *4773*, 372–373.
DOI: https://doi.org/10.1038/190372a0

Makhlouf, M., Ouimette, J.-F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014). A prominent and conserved role for YY1 in Xist transcriptional activation. Nat. Commun. *5*, 1–12.
DOI:https://doi.org/10.1038/ncomms5878

Marks, H., Chow, J.C., Denissov, S., Françoijs, K.-J., Brockdorff, N., Heard, E., and Stunnenberg, H.G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res. *19*, 1361–1373.
DOI:https://doi.org/10.1101/gr.092643.109

Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. J. Cell Biol. *147*, 1399–1408.
DOI:https://doi.org/10.1083/jcb.147.7.1399

Migeon, B.R., Kazi, E., Haisley-Royster, C., Hu, J., Reeves, R., Call, L., Lawler, A., Moore, C.S., Morrison, H., and Jeppesen, P. (1999). Human X inactivation center induces random X chromosome inactivation in male transgenic mice. Genomics *59*, 113–121.
DOI:https://doi.org/10.1006/geno.1999.5861

Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and Mcintosh, I. (2001). Identification of TSIX , Encoding an RNA Antisense to Human XIST , Reveals Differences from its Murine Counterpart : Implications for X Inactivation. Am. J. Hum. Genet. *69*, 951–960.
DOI:https://doi.org/10.1086/324022

Migeon, B.R., Lee, C.H., Chowdhury, A.K., and Carpenter, H. (2002). Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. Am. J. Hum. Genet. *71*, 286–293.
DOI:https://doi.org/10.1086/341605

Naumova, N., Smith, E.M., Zhan, Y., and Dekker, J. (2012). Analysis of long-range chromatin interactions using Chromosome Conformation Capture. NIH Public Access. Methods *58*, 1–27.
DOI:https://doi.org/10.1016/j.ymeth.2012.07.022

Navarro, P., Pichard, S., Ciaudo, C., Avner, P., and Rougeulle, C. (2005). Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: Implications for X-chromosome inactivation. Genes Dev. *19*, 1474–1484.
DOI:https://doi.org/10.1101/gad.341105

Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Dev. *20*, 2787–2792.
DOI:https://doi.org/10.1101/gad.389006

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693–1695.
DOI:https://doi.org/10.1126/science.1160952

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.
DOI:https://doi.org/10.1038/nature09496

Navarro, P., Moffat, M., Mullin, N.P., and Chambers, I. (2011). The X-inactivation trans-activator Rnf12 is negatively regulated by pluripotency factors in embryonic stem cells. Hum. Genet. *130*, 255–264.
DOI:https://doi.org/10.1007/s00439-011-0998-5

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., and Zakian, S.M. (2001). Characterization of the Genomic Xist Locus in Rodents Reveals Conservation of Overall Gene Structure and Tandem Repeats but Rapid Evolution of Unique Sequence. Genome Res. *11*, 833–849.

DOI:https://doi.org/10.1101/gr.174901

Nesterova, T.B., Johnston, C.M., Appanah, R., Newall, A.E.T., Godwin, J., Alexiou, M., and Brockdorff, N. (2003). Skewing X chromosome choice by modulating sense transcription across the Xist locus. Genes Dev. *17*, 2177–2190.
DOI:https://doi.org/ 10.1101/gad.271203

Nesterova, T.B., Senner, C.E., Schneider, J., Alcayna-Stevens, T., Tattermusch, A., Hemberger, M., and Brockdorff, N. (2011). Pluripotency factor binding and Tsix expression act synergistically to repress Xist in undifferentiated embryonic stem cells. Epigenetics & Chromatin *4*, 17.
DOI:https://doi.org/10.1186/1756-8935-4-17

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. and Heard, E. (2102). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.
DOI: https://doi:10.1038/nature11049

Ohhata, T., Hoki, Y., Sasaki, H., and Sado, T. (2008). Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. Development *135*, 227–235.
DOI:https://doi.org/ 10.1242/dev.008490

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.
DOI: https://doi.org/10.1038/nature09872

Ovcharenko, I., Nobrega, M.A., Loots, G.G., and Stubbs, L. (2004). ECR Browser: A tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res. *32*, 280–286.
DOI:https://doi.org/10.1093/nar/gkh355

Patel, S., Bonora, G., Sahakyan, A., Kim, R., Chronis, C., Langerman, J., Fitz-Gibbon, S., Rubbi, L., Skelton, R.J.P., Ardehali, R., et al. (2017). Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. Cell Rep. *18*, 54–67.
DOI:https://doi.org/10.1016/j.celrep.2016.11.054

Plath, K. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science *300*, 131–135.
DOI:https://doi.org/10.1126/science.1084274

Pugacheva, E.M., Tiwari, V.K., Abdullaev, Z., Vostrov, A.A., Flanagan, P.T., Quitschke, W.W., Loukinov, D.I., Ohlsson, R., and Lobanenkov, V. V. (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.
DOI:https://doi.org/10.1093/hmg/ddi089

Ray, P.F., Winston, R.M.L., and Handyside, A.H. (1997). XIST expression from the

maternal X chromosome in human male preimplantation embryos at the blastocyst stage. Hum. Mol. Genet. *6*, 1323–1327.
DOI:https://doi.org/10.1093/hmg/6.8.1323

Sado, T., Hoki, Y., and Sasaki, H. (2005). Tsix silences Xist through modification of chromatin structure. Dev. Cell *9*, 159–165.
DOI:https://doi.org/10.1016/j.devcel.2005.05.015

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. Cell Stem Cell *14*, 203–216.
DOI:https://doi.org/ 10.1016/j.stem.2013.11.022

Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of human active and inactive X chromosomes. Genome Res. *21*, 1592–1600.
DOI:https://doi.org/10.1101/gr.112680.110

Sheardown, S.A., Duthie, S.M., Johnston, C.M., Newall, A.E.T., Formstone, E.J., Arkell, R.M., Nesterova, T.B., Alghisi, G.C., Rastan, S., and Brockdorff, N. (1997). Stabilization of Xist RNA mediates initiation of X chromosome inactivation. Cell *91*, 99–107.
DOI:https://doi.org/10.1016/S0092-8674(01)80012-X

Silva, S.S.S., Rowntree, R.K.K., Mekhoubad, S., and Lee, J.T.T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. Proc Natl Acad Sci U S A *105*, 4820–4825.
DOI:https://doi.org10.1073/pnas.0712136105

Skotheim, R.I., Lind, G.E., Monni, O., Nesland, J.M., Abeler, V.M., Fosså, S.D., Duale, N., Brunborg, G., Kallioniemi, O., Andrews, P.W., et al. (2005). Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal development. Cancer Res. *65*, 5588–5598.
DOI:https://doi.org/10.1158/0008-5472.CAN-05-0153

Sousa, E.J., Stuart, H.T., Bates, L.E., Ghorbani, M., Nichols, J., Dietmann,S., and Silva, C.R.J. (2018). Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. Cell Stem Cell *22*, 919-928.
DOI: https://doi.org/10.1016/j.stem.2018.05.001

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA activates Xist by Evicting CTCF. Cell *153*, 1537–1551.
DOI:https://doi.org/10.1016/j.cell.2013.05.028

Tang, Y.A., Huntley, D., Montana, G., Cerase, A., Nesterova, T.B., and Brockdorff, N. (2010). Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation. Epigenetics Chromatin *3*, 1–11.
DOI:https://doi.org/10.1186/1756-8935-3-10

van den Berg, I.M., Laven, J.S.E., Stevens, M., Jonkers, I., Galjaard, R.J., Gribnau, J., and Hikke van Doorninck, J. (2009). X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos. Am. J. Hum. Genet. *84*, 771–779.

DOI:https://doi.org/10.1016/j.ajhg.2009.05.003

Vallot, C., Huret, C., Lesecque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A., Duret, L., and Rougeulle, C. (2013). *XACT*, a long noncoding transcript coating the active X chromosome in human pluripotent cells. Nature Genetics *45*, 239-241. DOI:https://doi:10.1038/ng.2530

Vallot,C., Patrat, C., Collier, A.J., Huret,C., Casanova, M., Liyakat Ali,T.M., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J., and Rougeulle, C. (2017). XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. Cell Stem Cell *20*, 102-111. DOI:http://dx.doi.org/10.1016/j.stem.2016.10.014

van Bemmel, J.G., Galupa, R., Gard, C., Servant, N., Picard, C., Davies, J., Szempruch, A.J., Zhan, Y., Żylicz, J.J., Nora, E.P., Lameiras, S., de Wit, E., Gentien, D., Baulande, S., Giorgetti, L., Guttman, M., Hughes, J.R., Higgs, D.R., Gribnau, J. and Heard,E. (2019). The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. Nature Genetics *51*, 1024-1034.
DOI:https://doi.org/10.1038/s41588-019-0412-0

Wakefield, M.J., Keohane, A.M., Turner, B.M., and Graves, J. a (1997). Histone underacetylation is an ancient component of mammalian X chromosome inactivation. Proc. Natl. Acad. Sci. U. S. A. *94*, 9665–9668.
DOI:https://doi.org/10.1073/pnas.94.18.9665

Weth, O., Paprotka, C., Katharina, G., Schulte, A., Baierl, M., Leers, J., Galjart, N., and Renkawitz, R. (2014). CTCF induces histone variant incorporation , erases the H3K27me3 histone mark and opens chromatin. *42*, 11941–11951.
DOI:https://doi.org/10.1093/nar/gku937

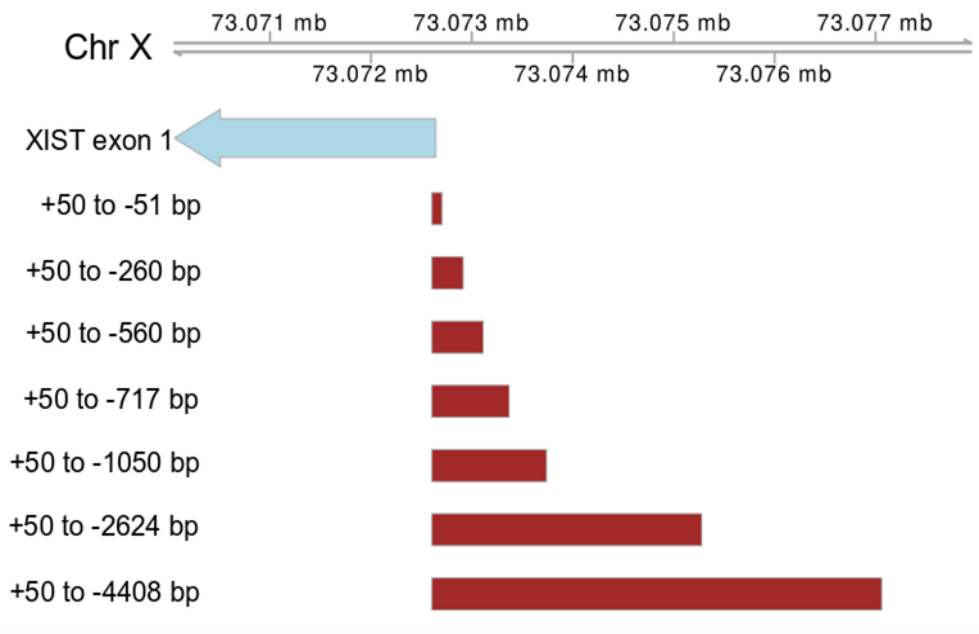
Williams, L.H., Kalantry, S., Starmer, J., and Magnuson, T. (2011). Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass. Development *138*, 2049–2057.
DOI:https://doi.org/10.1242/dev.061176

Wutz, A., and Jaenisch, R. (2000). A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation. Mol. Cell *5*, 695–705.
DOI:https://doi.org/10.1016/S1097-2765(00)80248-8

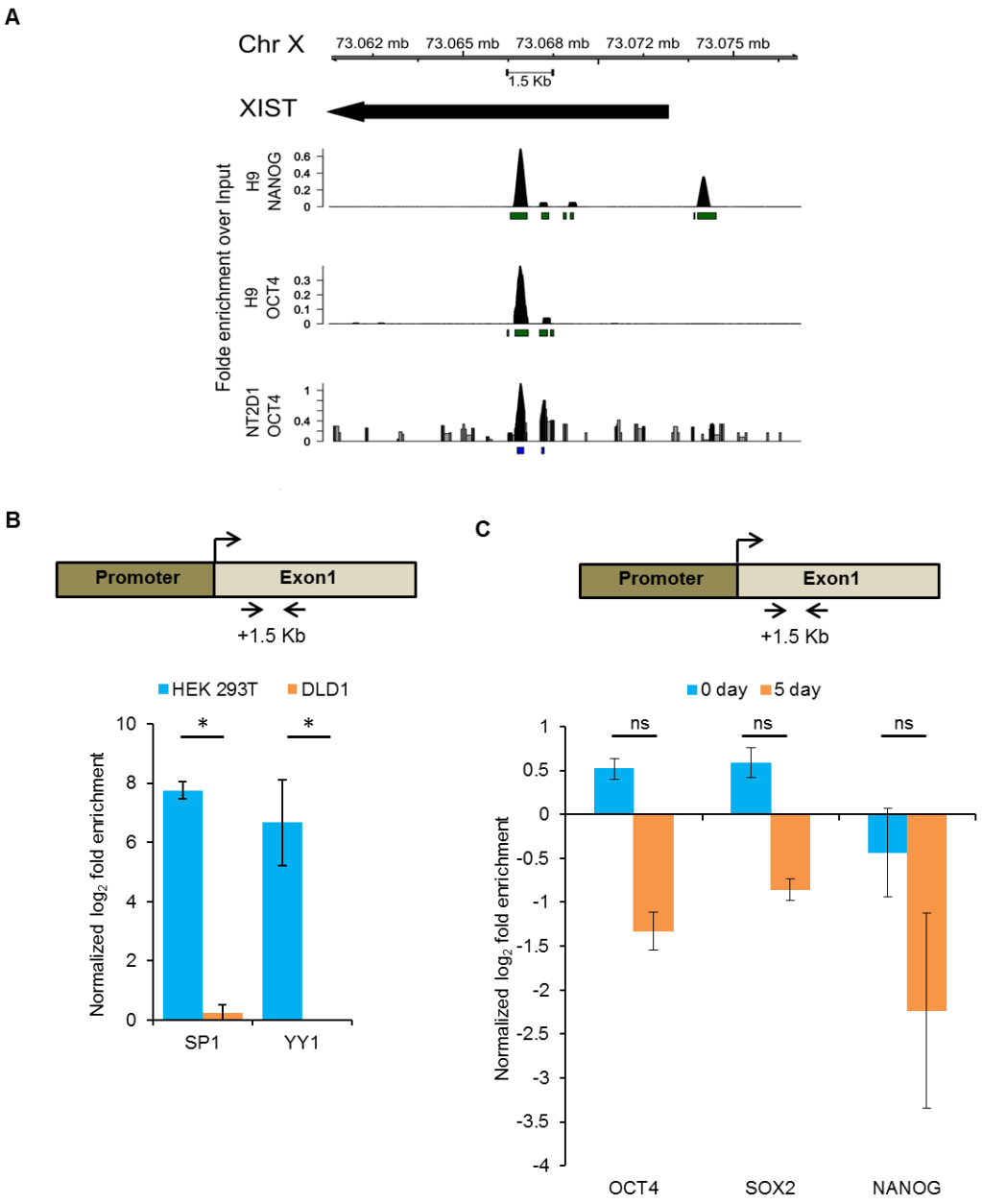
Xiao, C., Sharp, J.A., Kawahara, M., Davalos, A.R., Difilippantonio, M.J., Hu, Y., Li, W., Cao, L., Buetow, K., Ried, T., et al. (2007). The XIST Noncoding RNA Functions Independently of BRCA1 in X Inactivation. Cell *128*, 977–989.
DOI:https://doi.org/10.1016/j.cell.2007.01.034

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum,C., Myers, R.M., Brown, M., Li, W. and Liu, S.X. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biology *9*, R137.
DOI:https://10.1186/gb-2008-9-9-r137

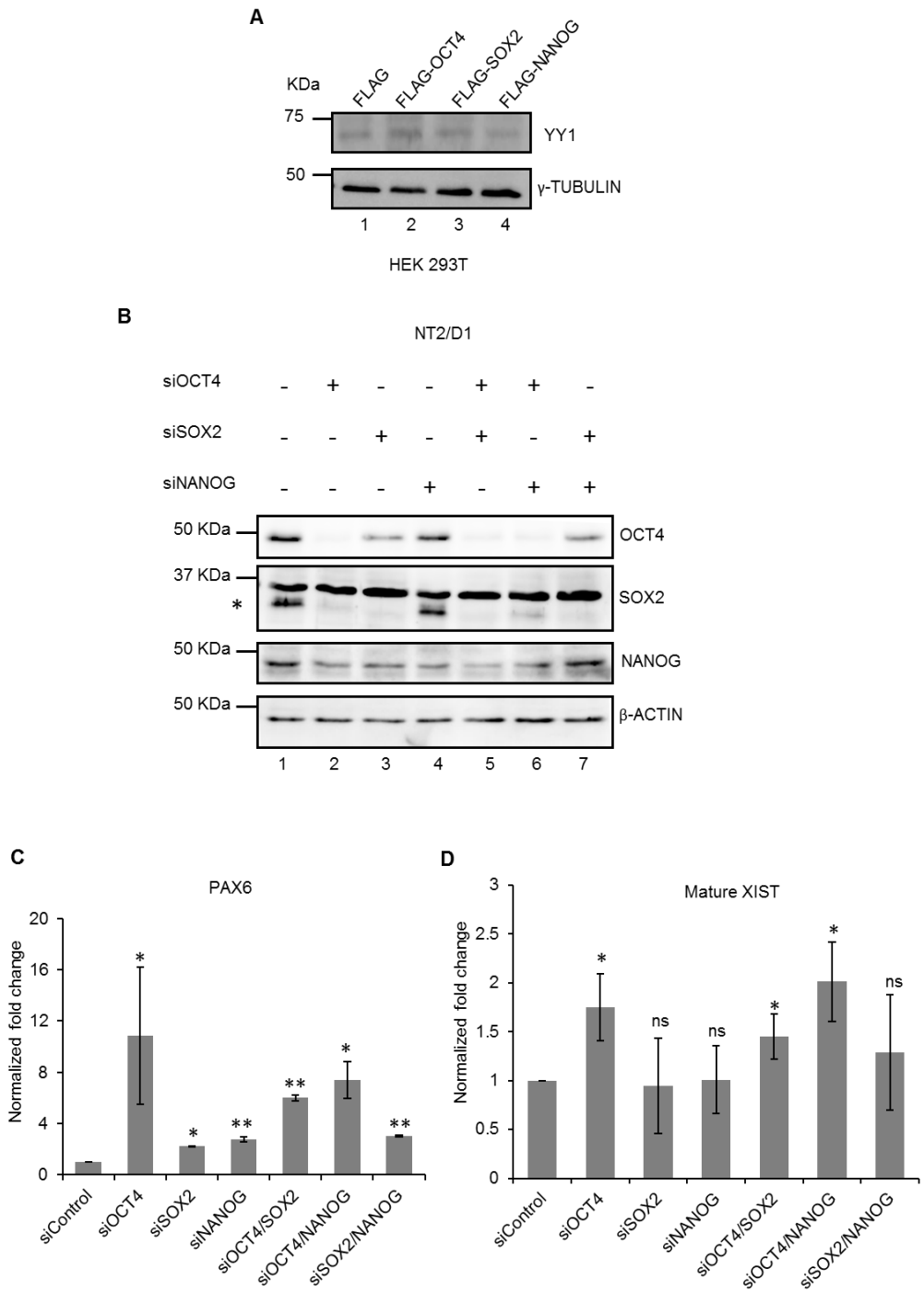
Supplementary Information



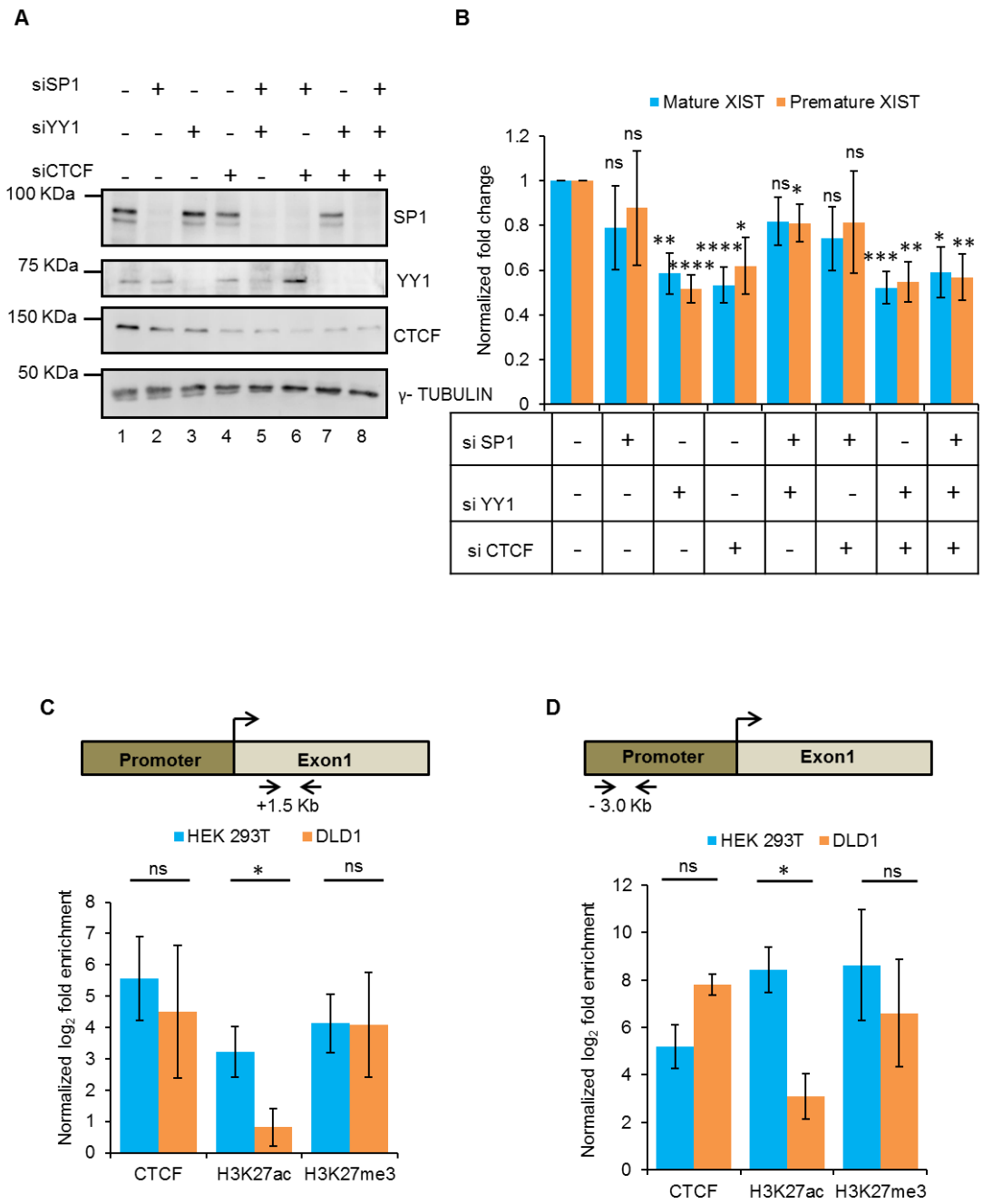
Supplementary figure 1. Promoter constructs for *XIST* cloned in the luciferase reporter vector pGL3 Basic. The brown boxes denote the fragments of the region upstream to XIST cloned into pGL3 Basic Luciferase reporter for testing potential promoter activity. The blue arrow indicates the first exon of *XIST* with the arrowhead denoting the direction of transcription and the tail indicating the TSS. The numbers flanking each brown box denote the relative location of the fragment with respect to the Transcription Start Site (TSS) of XIST. **(Related to Fig. 2A, B)**



Supplementary figure 2. Induction of XIST during differentiation of NT2D1 cells is governed by promoter as well as exon1 of XIST. (A) ChIP-seq peaks for OCT4 and NANOG for the *XIST* locus in human embryonic stem cell line H9 and OCT4 ChIP-seq peak for NT2/D1 cells. The arrowhead indicates direction of *XIST* transcription. All peaks were confirmed to have a p value < 0.05 as reported by MACS2 callpeak function. (B) ChIP-qPCR analysis for SP1 and YY1 on *XIST* promoter proximal region (~ +1.5 Kb) (as shown in the schematic above) in HEK 293T (female, blue bar) and DLD1 (male, brown bar) cells. (C) ChIP-qPCR analysis for OCT4, SOX2, NANOG on *XIST* promoter proximal region (~ +1.5 Kb) (as shown in the schematic above) in 0 day (blue bar) and 5 day (brown bar) differentiated NT2/D1 cells. X-axis represents the immunoprecipitated factor and Y-axis represents the normalized log_2 fold enrichment over IgG. Each bar represents values from 2 independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (*P value < 0.05). **(Related to Fig. 2E, F)**.

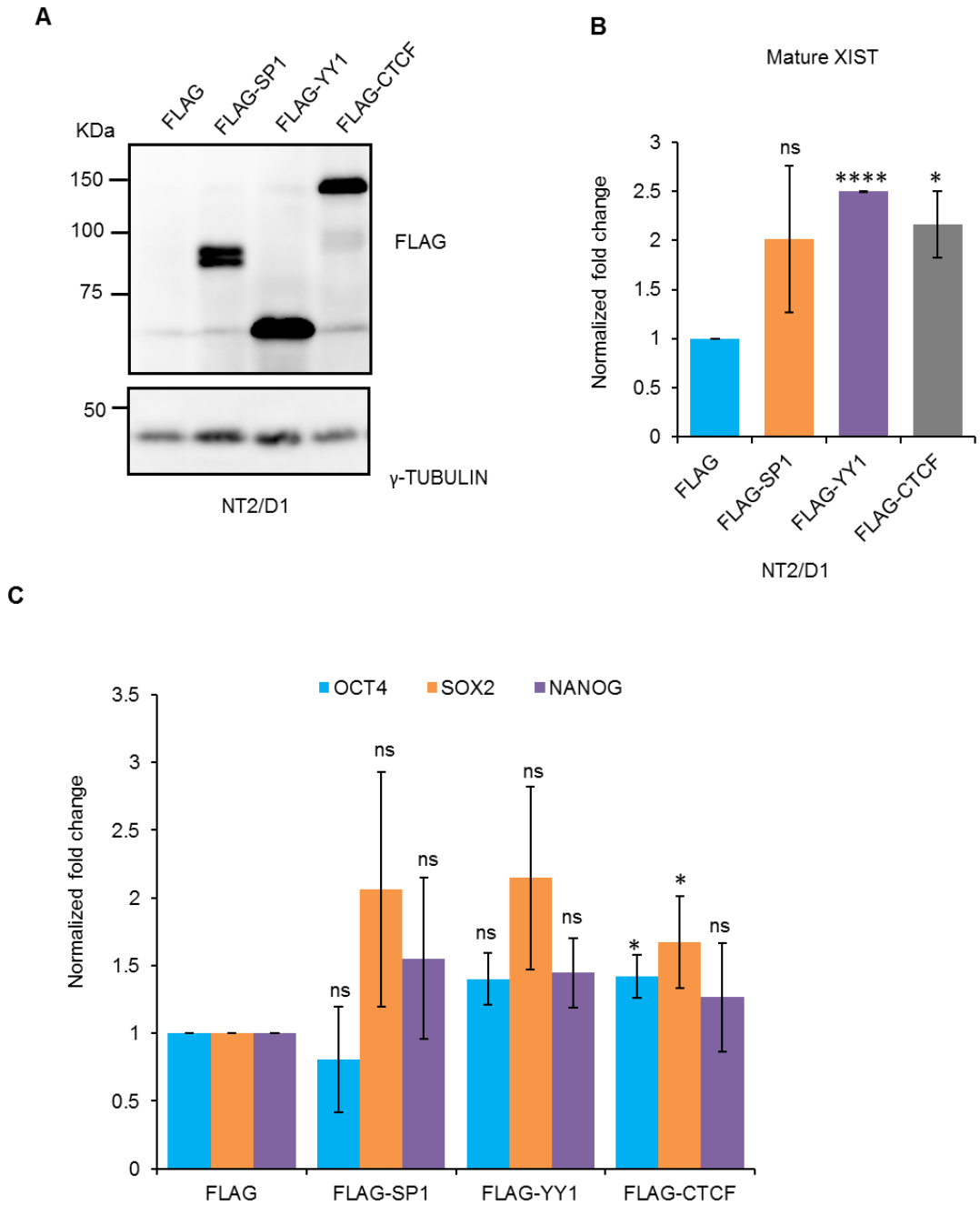


Supplementary figure 3. Pluripotency factors repress *XIST*. (A) Over-expression of pluripotency factors does not affect the levels of YY1 in HEK 293T cells. Immunoblotting for YY1 indicate that its expression does not change upon over-expression of pluripotency factors in HEK 293T cells. (Related to Fig. 3A,B). (B) Immunoblotting to determine the knockdown efficiencies of OCT4, SOX2 and NANOG in NT2/D1 cells. (C,D) qRT-PCR for PAX6 (C) upon siRNA mediated knockdown of OCT4, SOX2, NANOG in NT2/D1 cells. X-axis represents siRNA transfected and Y-axis represents the fold change normalized to 18s rRNA. Each bar represents values from 3 independent experiments. Error bar represents the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (**P value < 0.01, *P value < 0.05, ns=non-significant) **(Related to Fig. 3)**.

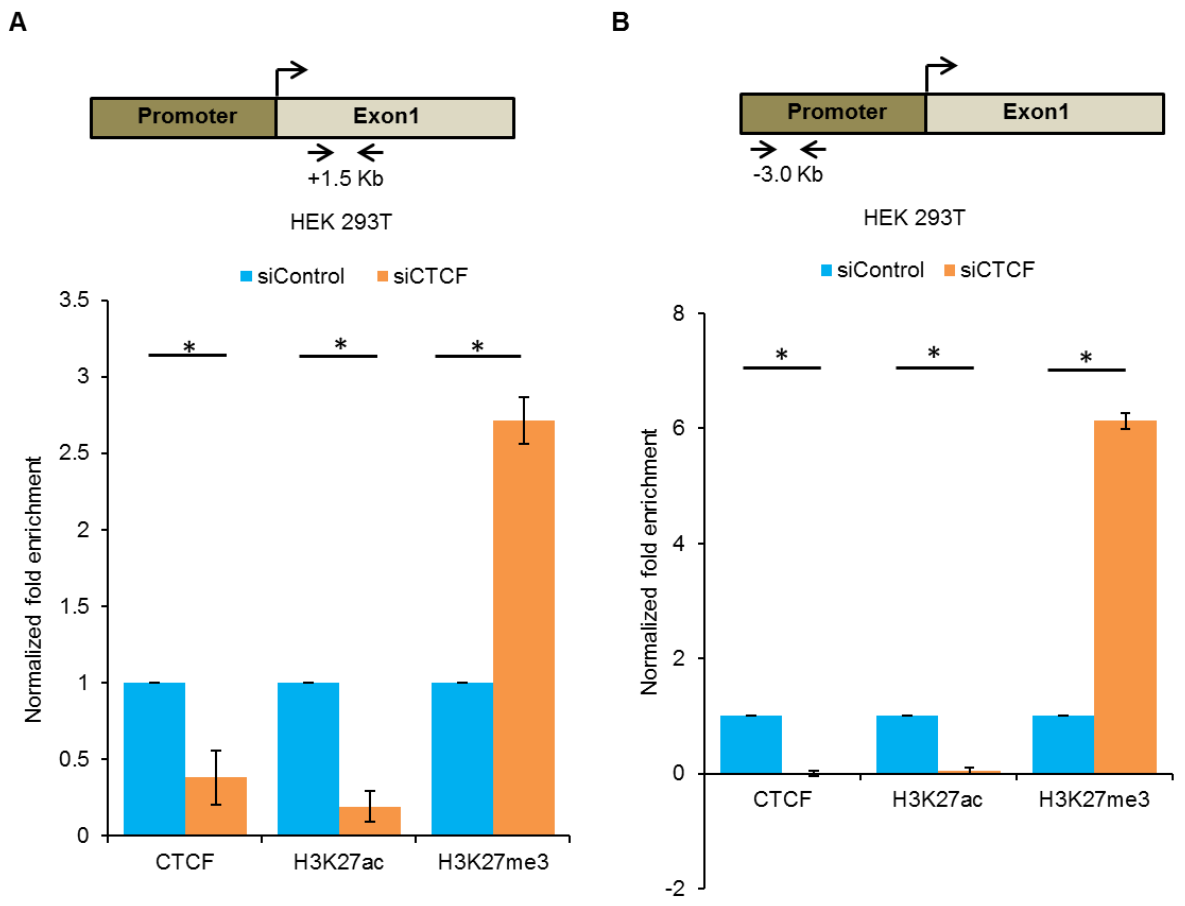


Supplementary figure 4. YY1 and CTCF positively regulate XIST transcription. (A) Immunoblotting to confirm knockdown of SP1, YY1, CTCF in HEK 293T cells. γ-TUBULIN serves as an equal loading control. Western blot for SP1, YY1, CTCF for single knockdown and γ-TUBULIN is the same as in figure 4C. (B) qRT-PCR for mature and premature XIST upon siRNA mediated knockdown of SP1, YY1 and CTCF in HEK 293T cells. X-axis represents siRNA transfected and Y-axis represents the fold change normalized to 18s rRNA. Each bar represents values from 4 independent experiments. Error bar represents the ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, ***P value < 0.001, **P value < 0.01, *P value < 0.05, ns=non-significant). qRT-PCR graph for siControl, siSP1, siYY1 and siCTCF is the same as in the main figure 4D.

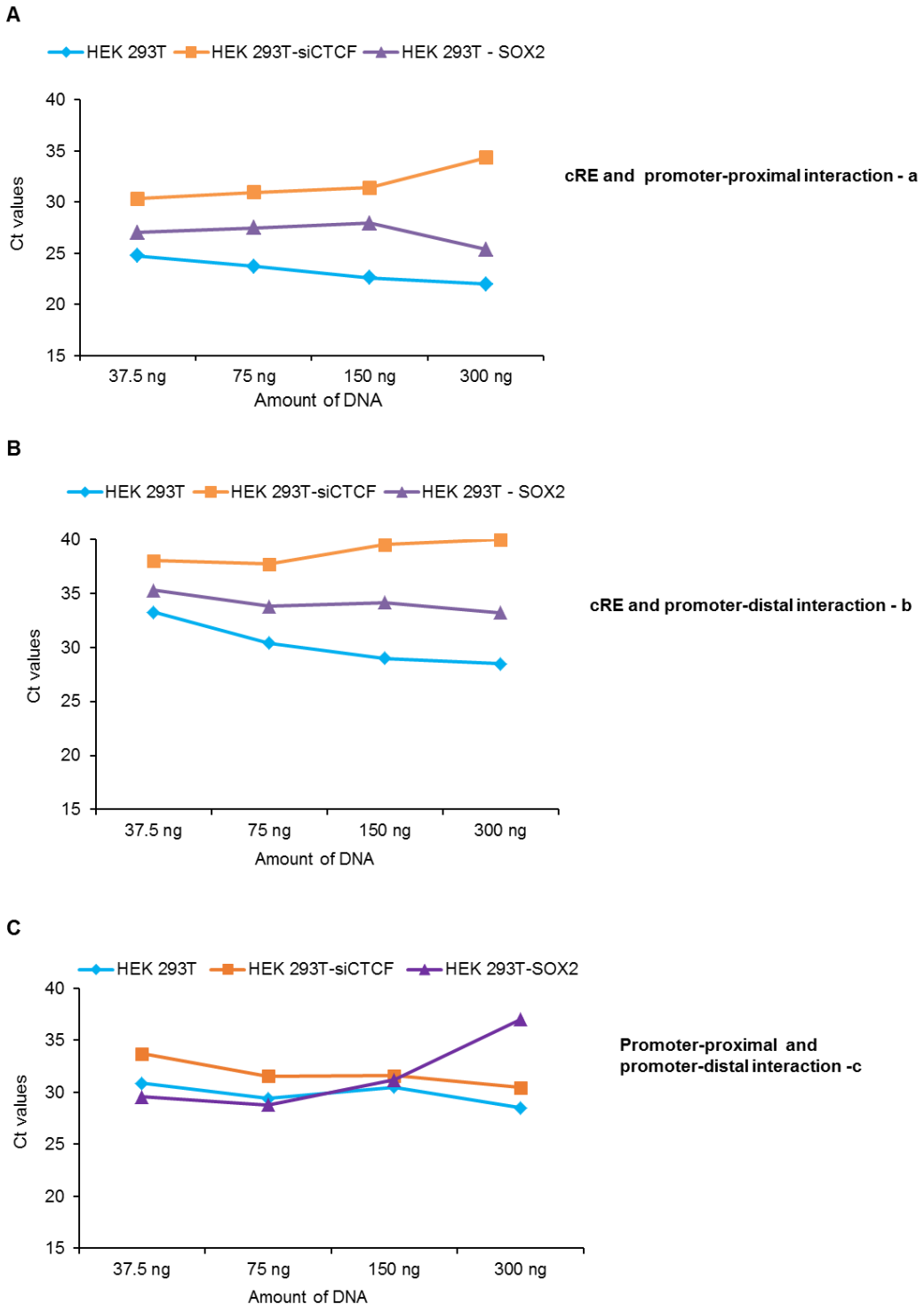
(C,D) ChIP-qPCR analysis showing no difference between CTCF and H3K27me3 and a significant difference in H3K27ac enrichment between HEK 293T (female, blue bar) and DLD1 (male, yellow bar) cells for promoter proximal site (+1.5 Kb) (C) and for promoter distal site (-3.0 Kb) (D) (as shown in the schematic above). X-axis represents the antibodies used for ChIP and Y-axis represents the normalized $\log_2$ fold enrichment over IgG. Each bar represents values from 2 independent experiments. Error bars represent the ±S.E.M. Asterisks represent the significance over DLD1 ChIP as per Student's T-test (ns = non-significant) **(Related to Fig. 4)**.



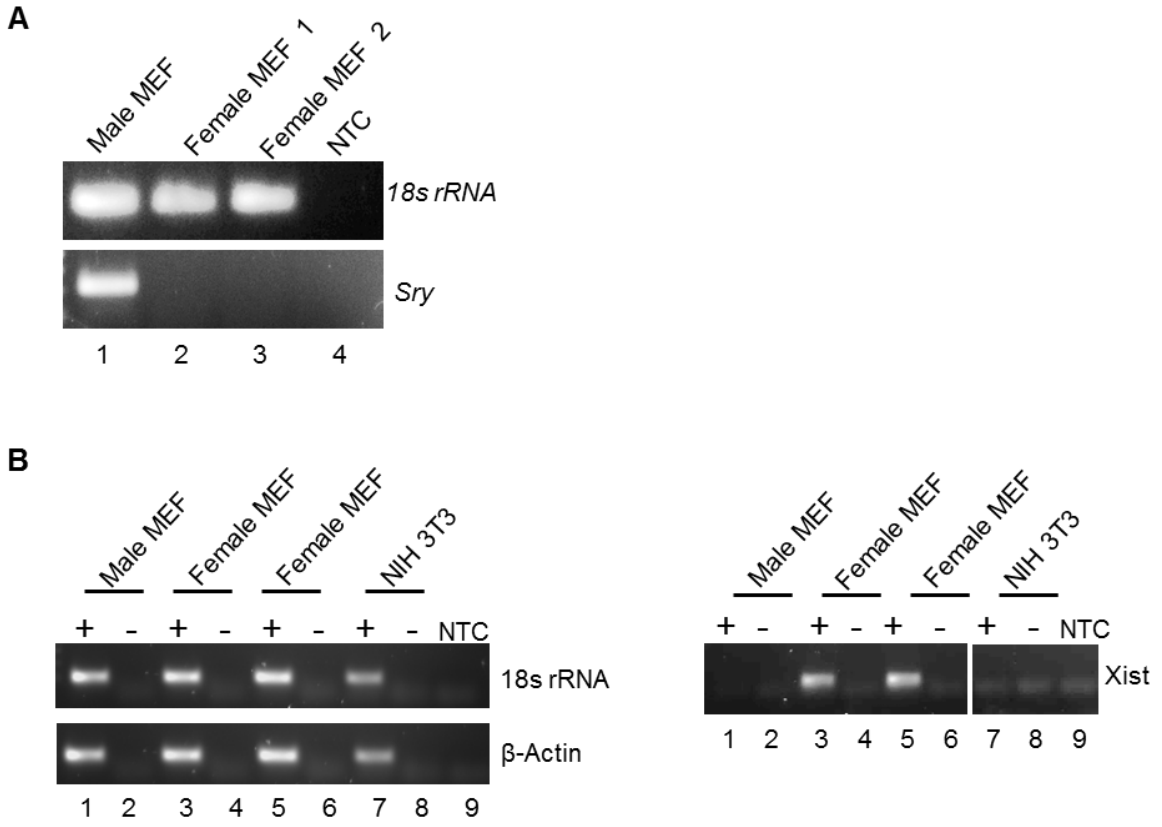
Supplementary figure 5. YY1 and CTCF positively regulate XIST transcription. (A) Immunoblotting to confirm over-expression of SP1, YY1, CTCF in NT2/D1 cells. γ-TUBULIN serves as an equal loading control. (B) qRT-PCR demonstrating a significant increase in XIST expression upon over-expression of YY1 or CTCF in NT2/D1 cells. (C) qRT-PCR for OCT4, SOX2 and NANOG upon over-expression of SP1, YY1, CTCF in NT2D1 cells. X-axis represents the transfected DNA and Y-axis represents the fold change normalized to 18s rRNA. Each point on the graph represents values from 3 independent experiments and error bar represents ±S.E.M. Asterisks represent the significance over vector control as per Student's T-test (****P value < 0.0001, *P value < 0.05, ns=non-significant) **(Related to Figure 4)**.



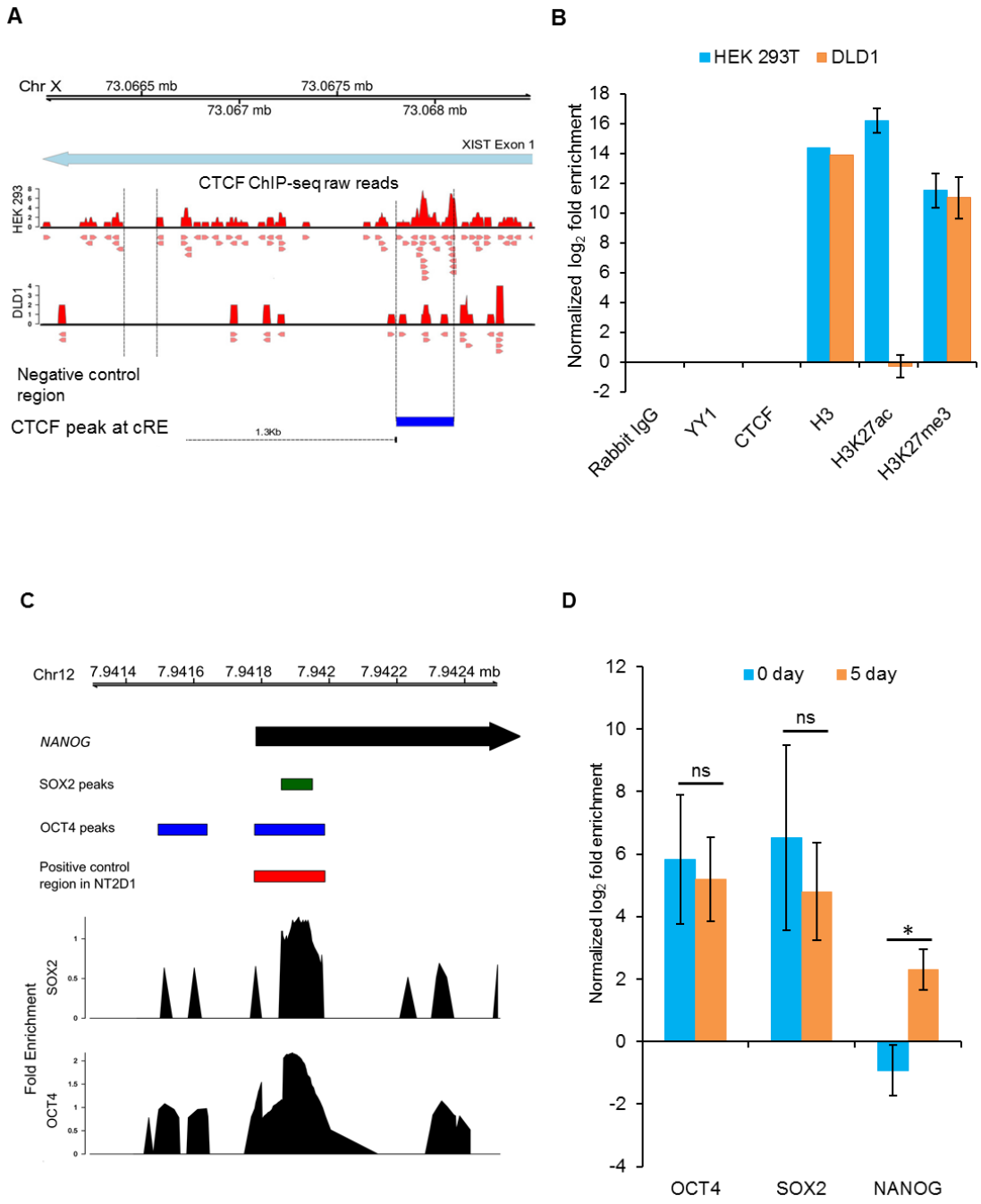
Supplementary figure 6. CTCF shapes the epigenetic landscape of *XIST* locus. (A, B) ChIP-qPCR analysis demonstrating a significant decrease in CTCF occupancy and H3K27ac enrichment and a significant increase in H3K27me3 enrichment on the promoter proximal (+1.5 Kb) (A) and promoter distal (-3.0 Kb) (B) sites upon knockdown of CTCF in HEK 293T cells. (*P value < 0.05) **(Related to Fig. 5)**.



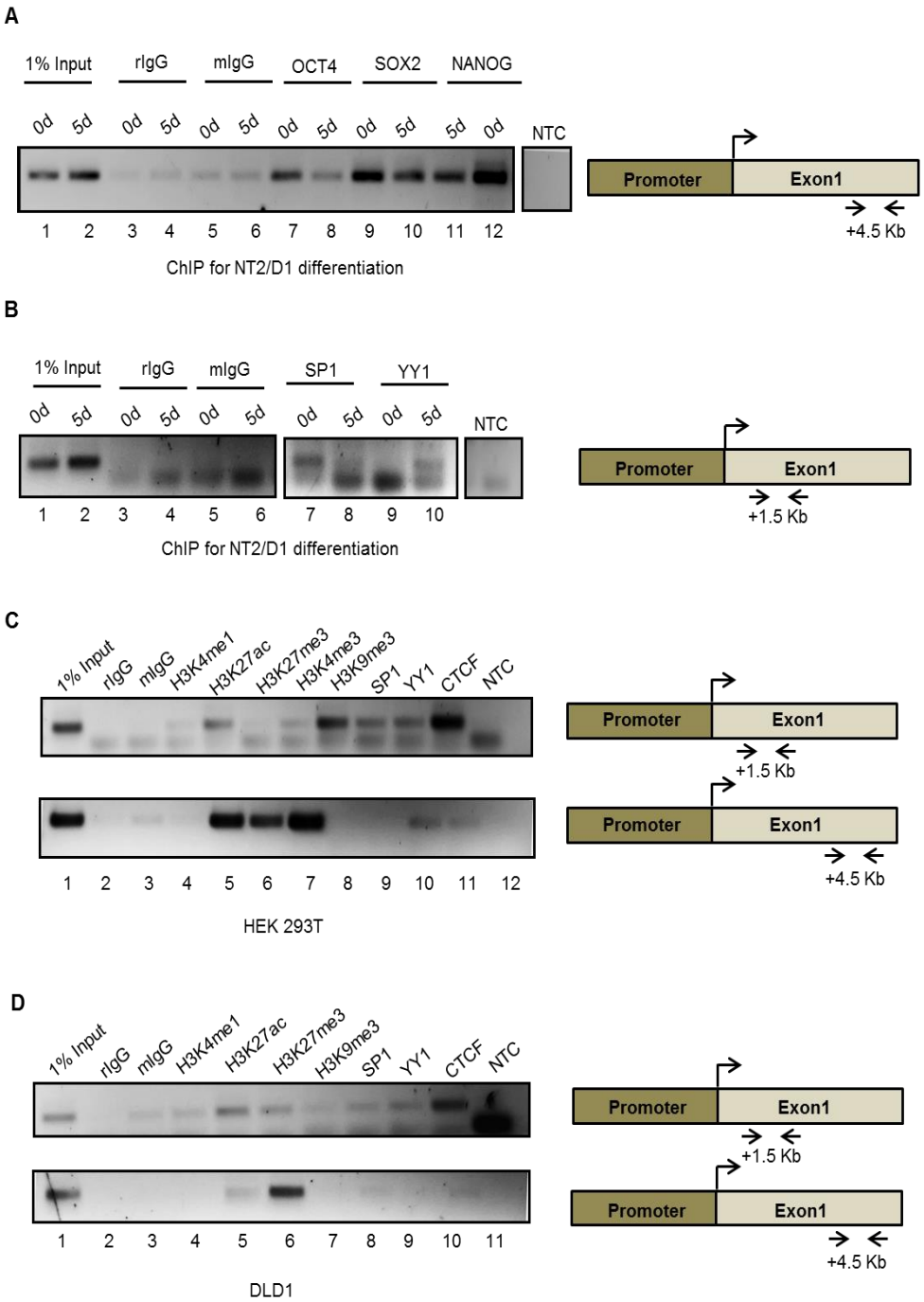
Supplementary figure 7. Plot for 3C library DNA concentration and Ct values for qRT-PCR for cRE to promoter-proximal interaction (+4.5 Kb to +1.5 Kb) (A), cRE to promoter-distal interaction (+4.5 Kb to -3.0 Kb) and promoter-proximal and promoter-distal interaction (+1.5 Kb to -3.0 Kb). Ct values for HEK 293T cells are lower compared to the depletion of CTCF or overexpression of CTCF in HEK 293T cells for any given amount of 3C library DNA. This is indicative of lower interaction frequencies in these two experimental conditions **(Related to figure 7)**.



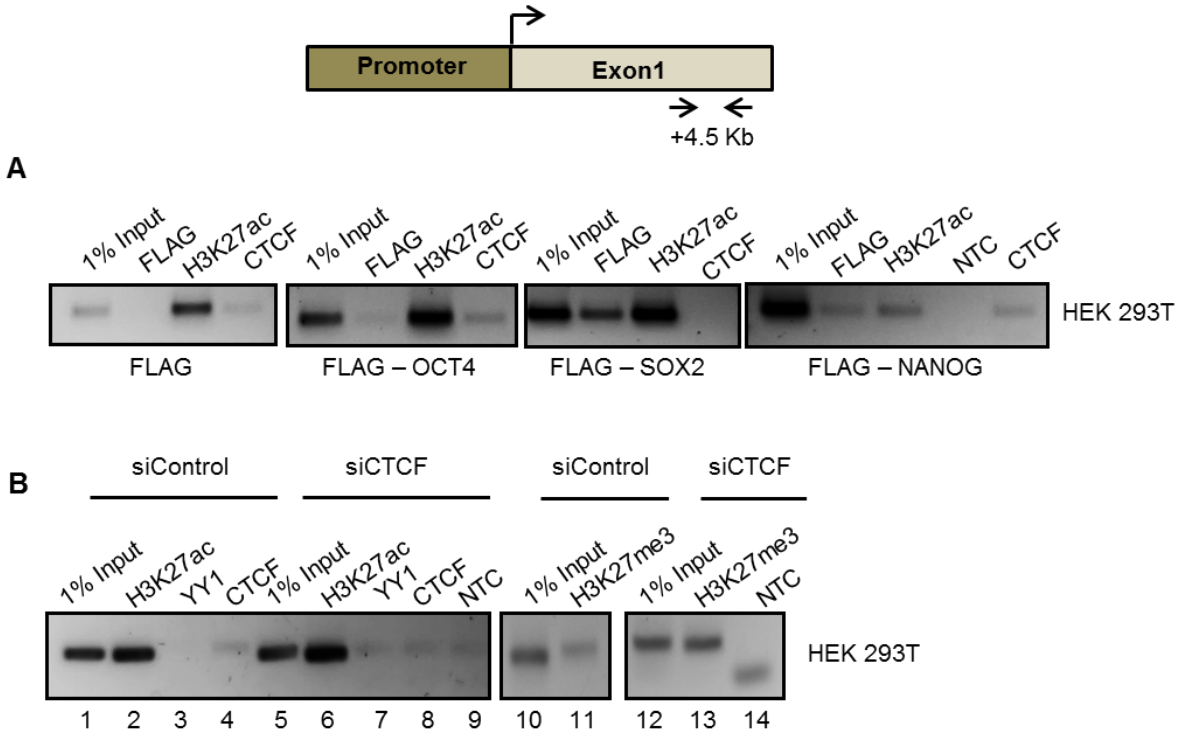
Supplementary figure 8. Identification of male and female MEFs. (A) PCR for *SRY* and *18s rRNA* using genomic DNA isolated from 3 MEF lines. (B) RT-PCR for Xist using cDNA prepared from the RNA isolated from 3 MEF lines and NIH 3T3 cells **(Related to Fig. 9)**.



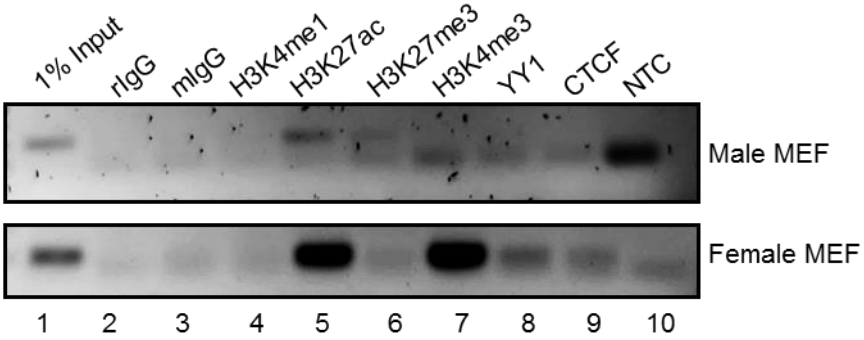
Supplementary figure 9. Controls for ChIP-qPCR analysis. (A) The locations of the negative control for ChIP shown overlayed with the reads from the Chip Seq raw data. The negative control is a region which is devoid of reads and does not contain any repeat regions. The distance between the negative control and the peak of interest is indicated at 1.3 kb. (B) ChIP-qPCR analysis for the control region showing enrichment for histone H3 but no enrichment for YY1 or CTCF. (C) A positive control for the ChIP of OCT4/POU5F1 in NT2D1. The positive control is selected based on the location of two existing peaks in NT2D1 cells. (D) ChIP-qPCR analysis for the control region showing enrichment of pluripotency factors on NANOG promoter.



Supplementary figure 10. ChIP-PCR for monitoring enrichment of the pluripotency factors, SP1, YY1, CTCF and histone modifications. (A) ChIP-PCR for OCT4, SOX2 and NANOG for 0 day and 5 day RA treated NT2/D1 cells using primers specific for cRE site (+4.5 Kb) as depicted in the schematic. (B) ChIP-PCR for SP1 and YY1 for 0 day and 5 day RA treated NT2/D1 cells using primers specific for promoter proximal region (+1.5 Kb) as depicted in the schematic above. Rabbit IgG and mouse IgG serve as negative controls. NTC stands for no template control. **(Related to Fig. 2E, F)**. (C,D) ChIP-PCR for active histone marks – H3K4me, H3K27ac, H3K4me3, repressive histone marks – H3K9me3, H3K27me3, SP1, YY1 and CTCF using primers specific for cRE (+4.5 Kb) and promoter proximal region (+1.5 Kb) for female (HEK 293T) (C) and male (DLD1) cells (D). Rabbit IgG and mouse IgG serve as negative controls. NTC stands for no template control **(Related to Fig. 4A,E)**.



Supplementary figure 11. ChIP-PCR for upon over-expression of pluripotency factors or knockdown of CTCF in HEK 293T cells. (A) ChIP-PCR using FLAG, H3K27ac and CTCF antibodies upon over-expression of FLAG-tagged OCT4, SOX2 or NANOG in HEK 293T cells. (Related to Fig. 3F) (B) ChIP-PCR for H3K27ac, H3K27me3, YY1 and CTCF upon siRNA mediated knockdown of CTCF in HEK 293T cells. The primers specific for the cRE region (+4.5 Kb) are used for both the PCRs. NTC stands for no template control **(Related to Fig. 5C,D)**.



Supplementary figure 12. ChIP-PCR for histone marks in male and female MEF lines. ChIP-PCR for active histone marks – H3K4me, H3K27ac and H3K4me3, repressive histone mark H3K27me3 and YY1, CTCF for male and female MEFs. Rabbit IgG and mouse IgG serve as negative controls. NTC represents no template control **(Related to Fig. 9C)**.

Supplementary Table 1. List of oligonucleotides

Description	Sequence (5' to 3')
Primers designed for cloning	
+50 to -51 bp, Forward	CTGCAGCAGCGAATTGCAG
+50 to -51 bp, Reverse	CAAAGATGTCCGGCTTTC
+50 to -260 bp, Forward	CTGCAGCAGCGAATTGCAG
+50 to -260 bp, Reverse	GCAGTTTATGGAGGATTTTAGC
+50 to -560 bp, Forward	CTGCAGCAGCGAATTGCAG
+50 to -560 bp, Reverse	GGAATGGGAAGTCCCTTGAAG
+50 to -717 bp, Forward	CTGCAGCAGCGAATTGCAG
+50 to -717 bp, Reverse	GCCATTCTATGAAATGTCTTTC
+50 to -1050 bp, Forward	CTGCAGCAGCGAATTGCAG
+50 to -1050 bp, Reverse	GCCAGTGGGAGGGTAATGTA
+50 to -1050 bp, Antisense, Forward	CTGCAGCAGCGAATTGCAG
+50 to -1050 bp, Antisense, Reverse	GCCAGTGGGAGGGTAATGTA
+50 to -2624 bp, Forward	CTGCAGCAGCGAATTGCAG
+50 to -2624 bp, Reverse	TGGAGCCAAGCAGTAGTGAA
+50 to -4408 bp, Forward	CTGCAGCAGCGAATTGCAG
+50 to -4408 bp, Reverse	TTCCTCCCTCTCCCTAGTGTTT
Human SP1, Forward	ATGAGCGACCAAGATCACTCC
Human SP1, Reverse	TCAGAAGCCATTGCCACTG
Human YY1, Forward	ATGGCCTCGGGCGACACCCTC
Human YY1, Reverse	TCACTGGTTGTTTTTGGCC
Human CTCF, Forward	ATGGAAGGTGATGCAGTCG
Human CTCF, Reverse	TCACCGGTCCATCATGCTGAG
Human OCT4, Forward	ATGGCGGGACACCTGGCTTC
Human OCT4, Reverse	TCAGTTTGAATGCATGGGAG
Human SOX2, Forward	ATGTACAACATGATGGAGAC
Human SOX2, Reverse	TCACATGTGTGAGAGGGGCAG
Human NANOG, Forward	ATGAGTGTGGATCCAGCTTG
Human NANOG, Reverse	TCACACGTCTTCAGGTTGC
Primers designed for qRT-PCR	
Human GAPDH, Forward	CTGCACCACCAACTGCTTAG
Human GAPDH, Reverse	GTCTTCTGGGTGGCAGTGAT
Human 18s rRNA, Forward	CGCCGCTAGAGGTGAAATTCT
Human 18s rRNA, Reverse	CGAACCTCCGACTTTCGTTCT
Human mature XIST, Forward	GATGTCAAAAGATCGGCCCA
Human mature XIST, Reverse	CAATGGGATCAGCCAGAGACT
Human Premature XIST, Forward	TGCTTTAGCATCAAAGCCCT
Human Premature XIST, Reverse	GCCTTAGATTCCCAGTTCCA
Human OCT4, Forward	AGCAAAACCCGGAGGAGT
Human OCT4, Reverse	CCACATCGGCCTGTGTATATC
Human SOX2, Forward	TGCTGCCTCTTTAAGACTAGGA C
Human SOX2, Reverse	CCTGGGGCTCAAACTTCTCT
Human NANOG, Forward	CCTGAACCTCAGCTACAAACAG
Human NANOG, Reverse	GCTATTCTTCGGCCAGTTGT

Human PAX6, Forward	GGCACACACACATTAACACACT T
Human PAX6, Reverse	GGTGTGTGAGAGCAATTCTCAG
Mouse Xist, Forward	CAAGAGCGGGGAACTACTTA
Mouse Xist, Reverse	TCACAGTTGGTGGGAAGATG
Primers designed for ChIP-PCR	
Human XIST locus promoter proximal (+1.5 Kb) (hg19, chrX:73070859+73071010), Forward	TGCTAATTCACCCAGGTCTTC
Human XIST locus promoter proximal (+1.5 Kb) (hg19, chrX:73070859+73071010), Reverse	GAGAAAAGGTGGGATGGACA
Human XIST locus cRE (+4.5 Kb) (hg19, chrX:73,067,862-73,068,061) , Forward	AGGGAAGTGAGTGGGGTCTT
Human XIST locus cRE+E33 (+4.5 Kb) (hg19,chrX:73,067,862-73,068,061) , Reverse	TTACAGCAGGGGGTACTTGG
Human XIST locus promoter distal (-3.0 Kb) (hg19, chrX:73,075,478-73,075,627), Forward	ATCTGTACATCCTCTCCTCTGC
Human XIST locus promoter distal (-3.0 Kb) (hg19, chrX:73,075,478-73,075,627), Reverse	GATTCTGGTCCTGCCCTGTA
Mouse XIST locus ECR (mm10, chrX:100,676,896-100,676,997), Forward	GTGCCCATTTCTGACGAGTT
Mouse XIST locus ECR (mm10, chrX:100,676,896-100,676,997), Reverse	ATGGTAGGGTGTGGTTGGTC
Primers designed for 3C-PCR	
XIST locus 3C primer_Anchor primer (+4.5Kb) (hg19,chrX:73066109+73076691)	GTACTTCCAGCTGGGATGTAAA TACAGTGG
XIST locus 3C primer_Test primer (+1.5Kb) (hg19,chrX:73066109+73076691)	GCAATGCACATGACTTCCTCTG CCTGACCTG
XIST locus 3C primer_Test primer (-3.0Kb) (hg19,chrX:73066109+73076691)	GATTCAGATTATTCCAGTTCTAG GATTGTC
sgRNA sequence for deleting cRE	
Human XIST exon 1 cRE deletion sgRNA 1 (hg19, chrX:73068309-73068331) _sense	CACCGAGGGTTAGTGCACAAGA TCGTGG
Human XIST exon 1 cRE deletion sgRNA 1 (hg19, chrX:73068309-73068331) _antisense	AAACCCACGATCTTGTGCACTA ACCCTC
Primers designed for screening cRE deletion by CRISPR-Cas9	
cRE deletion flanking primer (hg19, chrX:73,067,693-73,068,381)_Forward	ATTCATAGTCCCAGGGAAAGGT
cRE deletion flanking primer (hg19, chrX:73,067,693-73,068,381)_Reverse	TGATCACCCAAGGACTCTTCTT