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
Department of Biology
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*. Collectively, our study uncovers the combinatorial effect of multiple epigenetic and transcription 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 (1–3). The phenomenon of XCI corrects for the X-linked gene dosage disparity between the males (XY) and females (XX) of mammalian species (4). 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 (5, 6). 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 (7–10), macroH2A recruitment (11) CpG methylation of gene promoters (12, 13). Consequently, the entire X chromosome undergoes heterochromatinization, rendering it stably inactive in a mitotically heritable manner (14–22). Since Xist/XIST lncRNA is the central player for establishing the process of chromosome-wide transcriptional silencing, it is 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 (23). 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 (24–28). 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 (29, 30). A recent report highlighted the role of promoter of another long non-conding RNA, *Linx* as a *cis*-regulatory silencer of *Xist*. *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 (31). 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 (32), by regulating *Xist* (33, 34),*Tsix* (35, 36) and *Rnf12* (37), Linx (31). 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 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 (1–3), 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 (17, 18, 38, 39). 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) (40–42). 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 (43–45). 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 (46, 47). 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 (48, 49). 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.

Three studies have addressed the regulation of human *XIST*. Hendrich *et al*. 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* (50). 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* (51, 52). 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 *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 undergo 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 (43, 44).

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 (57) and respond to retinoic acid (RA) mediated differentiation cues to give rise to neuronal progenitors (58). 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 (59).

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 **(Figure 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 (58) **(Figure 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 **(Figure 1D, E, F)**. RNA FISH analysis indicated that over 20% NT2/D1 cells exhibit robust XIST RNA FISH signal upon differentiation of NT2/D1 for 5 days **(Figure 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 have not served as an ideal model for the purpose so far.

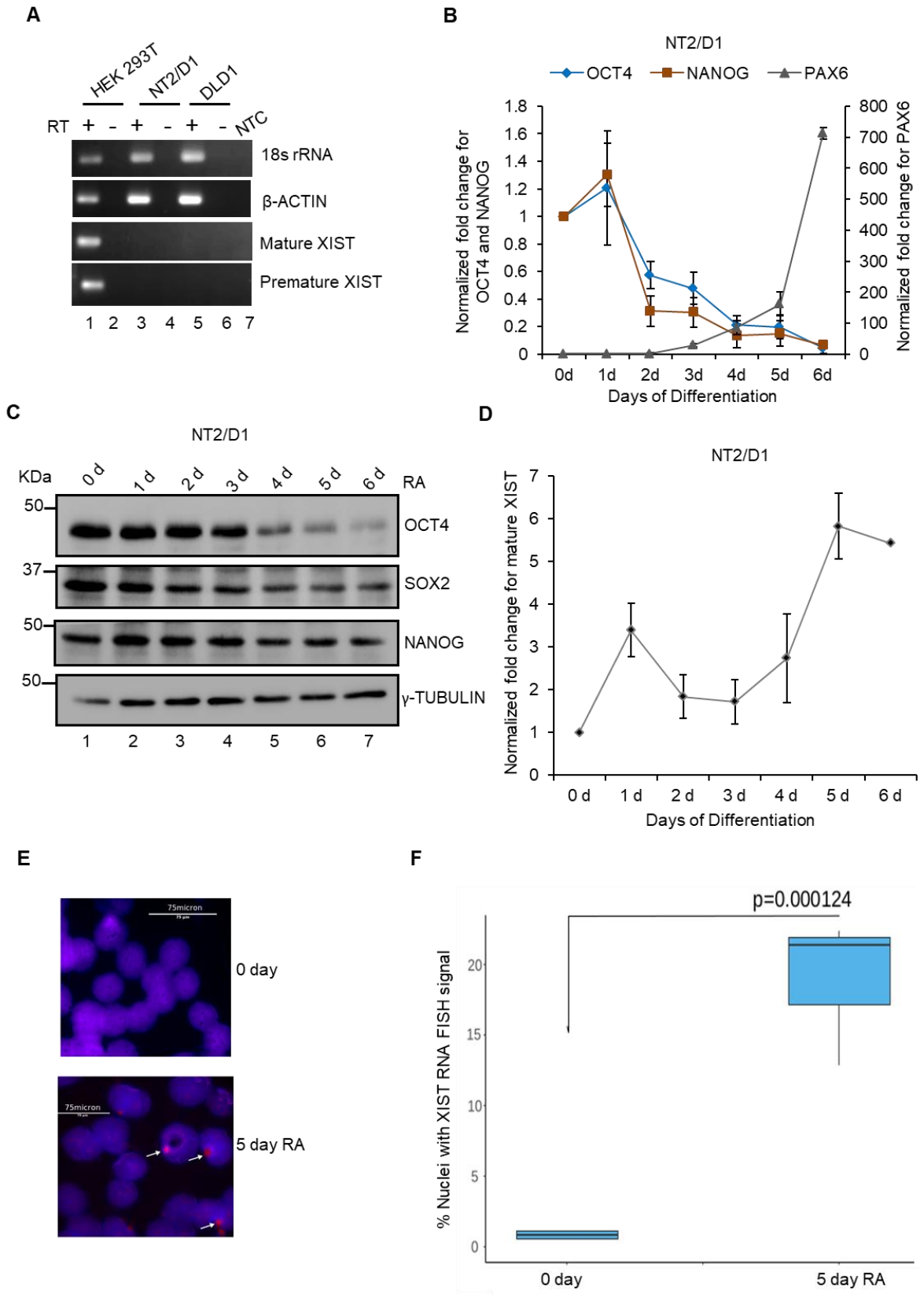


Figure 1. NT2/D1 and HEK 293T cells provide the contexts for initiation and maintenance phases of XCI. (A) Semi-quantitative RT-PCR for mature and premature XIST using cDNA prepared from HEK 293T (female), NT2/D1 (male) and

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) 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 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 exon 1 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 (50). 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 **(Figure S1)**. 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 (50), it was observed that the +50 bp to -51 bp region was sufficient to drive the transcription of the luciferase gene **(Figure 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 **Figure 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**, Figure 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 raises the question as to why *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 analyzed 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 a decline in luciferase activity was observed instead **(Figure 2C)**. The expression of previously identified transcriptional activators of *XIST* - SP1 and YY1 also decreased upon differentiation **(Figure 2D)** reaffirming that they could positively regulate *XIST*. The disparity between the two results **(Figure 1D and Figure 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 **(Figure S2A)**. The rationale for the speculation was the negative correlation between XIST and the stem cell marker levels during differentiation **(Figure 1B,C,D)** as well as previous reports demonstrating the negative regulation of *Xist* by them (33). ChIP-sequencing 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 **(Figure S2A)**. 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 **(Figure 2E)**. This indicates that binding of the pluripotency factors might serve to repress *XIST* in an undifferentiated state.

The next prominent question is to identify the factors positively regulating *XIST* in this scenario. YY1 binds to a site +1.5 Kb from the TSS and serves as a transcriptional activator of XIST (51, 52). Therefore, we determined its binding at

+1.5 Kb from the TSS and observed a significant enrichment of YY1 in 5 day differentiated NT2/D1 cells **(Figure 2F)**. Additionally, YY1 is enriched on the putative promoter of *XIST* of the inactive X chromosome since its occupancy can be observed only in the female cells (HEK 293T) and not in the male cells (DLD1) **(Figure S2B)**. The *NANOG* promoter region comprising ChIP-seq peaks for OCT4 and SOX2 served as a positive control for the ChIP assay **(Figure S2C,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*.

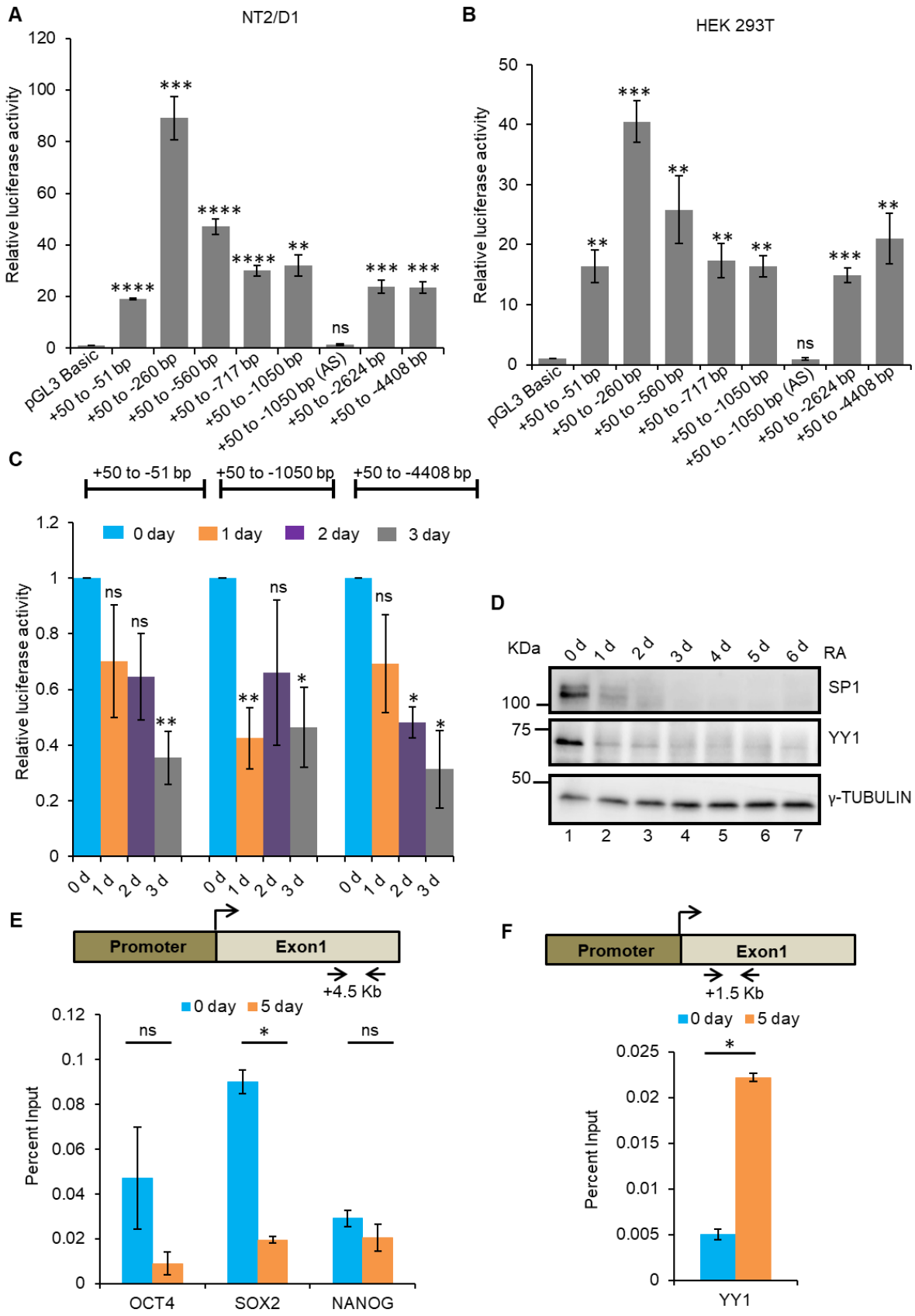


Figure 2. Induction of XIST during differentiation of NT2/D1 cells is governed by the promoter as well as exon1 of *XIST*. (A, B) Luciferase reporter activities for *XIST* promoter constructs transfected into NT2/D1 cells (A) or HEK 293T (B)

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 enrichment calculated as percent input. 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 **(Figure 3A)** in a female cell line, HEK 293T, which expresses XIST and does not express the pluripotency factors. A significant decrease in the levels of both mature and premature XIST was observed implying that the pluripotency factors negatively regulate *XIST* **(Figure 3B)**. By performing ChIP using FLAG antibody, we show that two of the overexpressed proteins (FLAG-OCT4/SOX2) are significantly enriched at the +4.5 Kb region on the exon 1 of *XIST*, providing a clear evidence that the pluripotency factors negatively regulate transcription of *XIST* by directly binding to this site **(Figure 3C)**. Levels of YY1 (a well-established activator of *XIST*) remained unperturbed under these experimental conditions **(Figure S3A)**. In addition to this, we also performed a siRNA mediated knockdown of OCT4, SOX2 or NANOG 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 **(Figure 3E)**. Increase in the expression of the neuronal marker – PAX6 is expected upon knockdown of pluripotency factors and hence serves as a control **(Figure S3B)**. Immunoblot analysis demonstrated that knockdown of OCT4 also leads to a significant decrease in the levels of SOX2 and NANOG **(Figure 3D)**. While knockdown of SOX2 leads to measurable reduction in OCT4 levels as well, knockdown of NANOG does not cause any appreciable change in the expression of the other two pluripotency factors. Thus, a significant upregulation in XIST expression was observed only when the levels of all the three factors were highly reduced, which is under the condition where OCT4 is knocked down. To lend further support to this hypothesis, we tested the repressive role of pluripotency factors by employing the NT2/D1 differentiation model. Towards this, NT2/D1 cells were differentiated in presence of retinoic acid (RA) for 4 days, at which point the levels of pluripotency factors would begin to decline **(Figure 1B,C)**, and XIST expression would be induced **(Figure 1D)**, followed by the over-expression of the pluripotency factors. The experimental scheme is depicted in **Figure 3F**. 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 **(Figure 3G,H)**. 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 the +4.5 Kb site by the pluripotency factors could exert a more potent effect on the expression of XIST in conjunction with the binding of the known activator – YY1 on the promoter-proximal region.

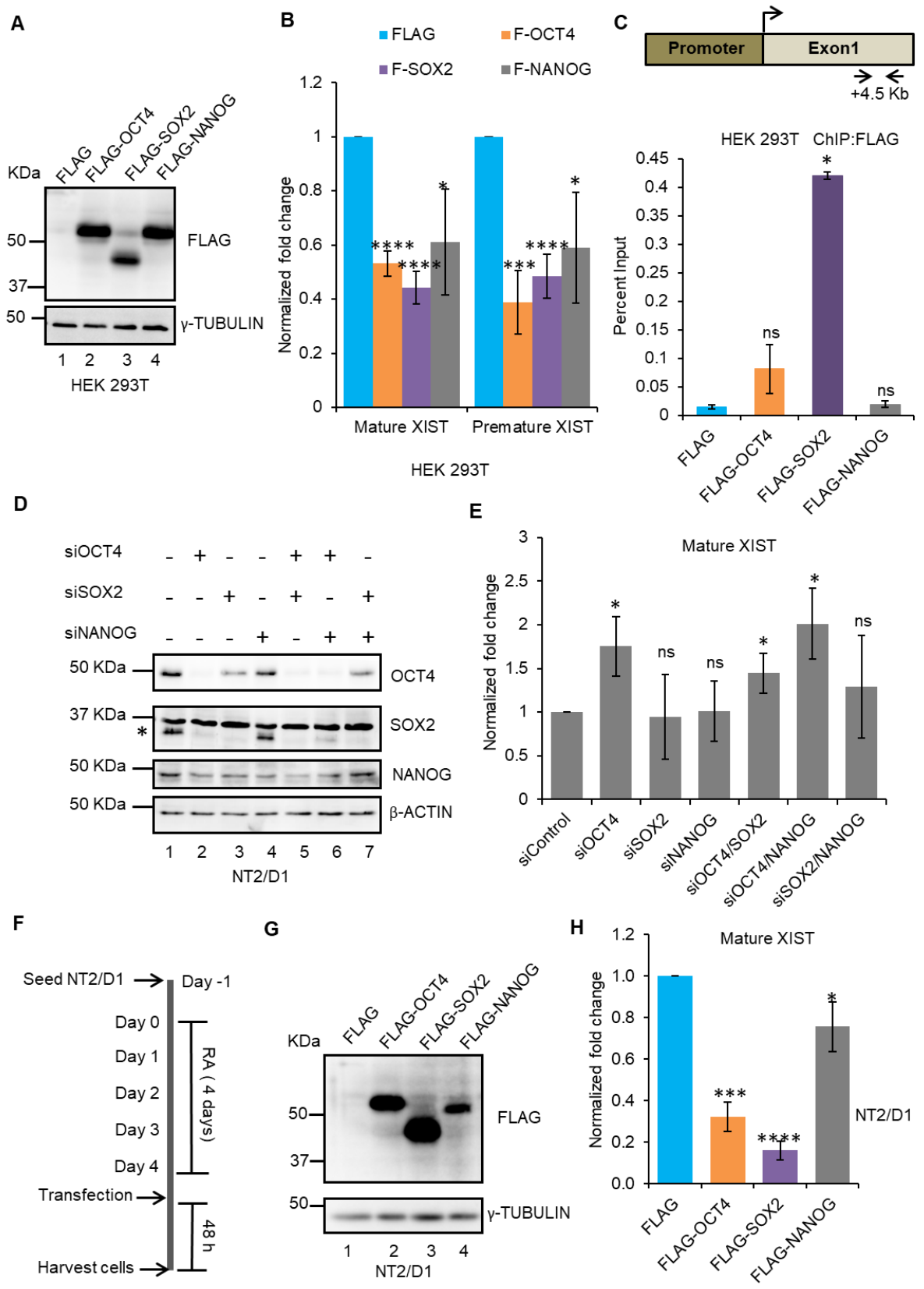


Figure 3. Pluripotency factors repress *XIST* by binding to exon 1 (+4.5 Kb) site. (A) Immunoblotting for FLAG to confirm the over-expression of OCT4, SOX2, NANOG in HEK 293T cells. γ-TUBULIN serves as an equal loading control. (B) qRT-PCR showing a significant reduction in mature and premature XIST 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) ChIP-qPCR showing occupancies of OCT4, SOX2, NANOG on the exon 1 (+4.5 Kb) site upon their over-expression in HEK 293T cells. X-axis represents the transfected DNA and Y-axis represents the enrichment calculated as percent input. Each point on the graph represents values from 2 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). (D) Immunoblotting to determine the knockdown efficiencies of OCT4, SOX2 and NANOG in NT2/D1 cells. β-ACTIN serves as an equal loading control (E) qRT-PCR for mature XIST 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) (F) Experimental scheme to overexpress OCT4, SOX2, NANOG in NT2/D1 cells differentiated for 4 days. (G) 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. (H) 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).

The pluripotency factor bound regulatory site 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* exon 1 (+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 (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) **(Figure 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) **(Figure 4A)**. Another feature of the newly identified cRE is the significant enrichment of the chromatin organizer protein CTCF specifically 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-sequencing datasets **(Figure 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 linesA 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 investigated 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 knockdown of YY1 or CTCF **(Figure 4C,D, S4A,B Fig)**. The effect of CTCF knockdown on SP1 and YY1 levels was virtually unnoticeable **(Figure 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 **(Figure S5A,B)**. We also determined that the over-expression of SP1 or YY1 or CTCF in NT2/D1 caused no appreciable decrease in the levels of pluripotency factors, thereby ruling out the indirect effect of over-expressions **(Figure S5C)**. These results strongly suggest that in addition to YY1, CTCF is an essential regulator of *XIST*. Furthermore, we validated the CTCF ChIP-sequencing analysis finding by performing ChIP for CTCF in HEK 293T, DLD1 cells **(Figure 4E (i,ii))** and observed a enrichment on the cRE in the female cells by an order of magnitude which positively correlates with XIST expression **(Figure 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 higher enrichment of CTCF when the first ChIP was performed using H3K27ac antibody **(Figure 4E (iii)**, **blue bar)** compared to H3K27me3 antibody **(Figure 4E (iii), brown bar)**. Further confirmation for the direct role of CTCF in regulating XIST transcription is evident by the decreased enrichment of CTCF **(Figure 4E (iv))** and increased H3K27me3 levels at the cRE observed upon knockdown of CTCF in HEK 293T cells **(Figure 4E (v))**.

Notably, CTCF occupancies as well as enrichment of H3K27ac on the previously reported CTCF peaks common between males and females as per Makhlouf *et al* 2014 are reduced upon knockdown of CTCF, accompanied by the concomitant increase in H3K27me3 **(Figure S6A,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*. Henceforth, we denote +1.5 Kb and -3.0 Kb sites as the promoter-proximal and promoter-distal regions respectively. We also performed ChIP-qPCR for a control region in HEK 293T and DLD1 cells (~1.3 Kb from the unique CTCF 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 **(Figure S7A,B)**. Altogether, these results indicate the existence of a cRE on the exon 1 of *XIST* whose function is modulated by the binding of CTCF or the pluripotency factors.

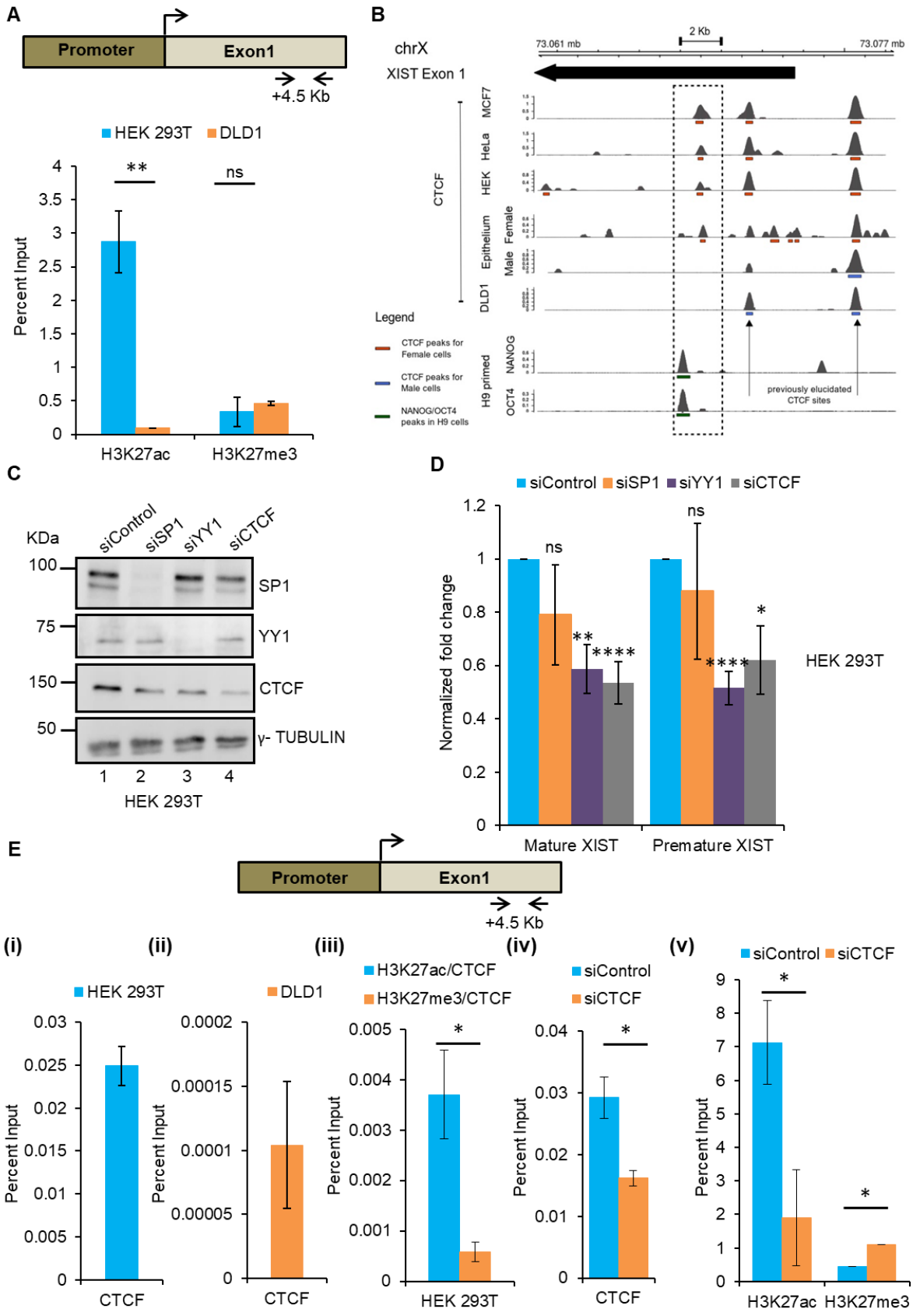


Figure 4. Pluripotency factor binding element is the potential cRE. (A) ChIP-qPCR analysis showing enrichment of active histone mark H3K27ac and repressive histone mark H3K27me3 for *XIST* cRE (+4.5 Kb, as shown in the schematic above)

on exon1 in HEK 293T (female, blue bar) and DLD1 (male, brown bar) cells. X-axis represents the antibodies used for ChIP and Y-axis represents the enrichment calculated as percent input. 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 (ii) DLD1 (male, brown bar) cells (iii) Sequential ChIP in HEK 293T cells (iv-v) ChIP-qPCR analysis for CTCF (iv) and H3K27ac, H3K27me3 (v) upon siRNA mediated knockdown of CTCF in HEK 293T cells (brown bars). Each bar represents values from 3 (for (i)) or 2 (for (ii-v)) independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance as per Student's T-test (*P value < 0.05).

Balance of activators (YY1, CTCF) and repressors (pluripotency factors) govern the transcription of XIST

To address the interplay between the transcriptional activators and repressors of XIST, we overexpressed SP1 or YY1 or CTCF coupled with the knockdown of OCT4 in NT2/D1 cells. Interestingly, we observed a significantly higher upregulation of XIST when the overexpression of YY1 or CTCF was coupled with the knockdown of OCT4 **(Figure 5A,B)** when compared with only the over-expressions **(Figure S5B)** or the knockdown of pluripotency factors alone **(Figure 3E)**. This suggests that the balance of activators (YY1, CTCF) and repressors (OCT4, SOX2, NANOG) govern the transcriptional activation of *XIST*. It is noteworthy that the siRNA-mediated knockdown of OCT4 depletes not only OCT4 but also SOX2 **(Figure 5A)**. To address the relevance of CTCF binding on the cRE, we sought to determine its occupancy upon over-expression of pluripotency factors since XIST levels decrease upon overexpression of OCT4, SOX2 or NANOG **(Figure 3B,H)**. Further, a higher enrichment of pluripotency factors (especially SOX2) is observed at the cRE **(Figure 3C)**, 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 **(Figure 5C)**. These results provide substantial evidence suggestive of a competition between SOX2 and CTCF for binding to cRE, leading to contrasting transcriptional outputs (repression and activation respectively) from the *XIST* promoter.

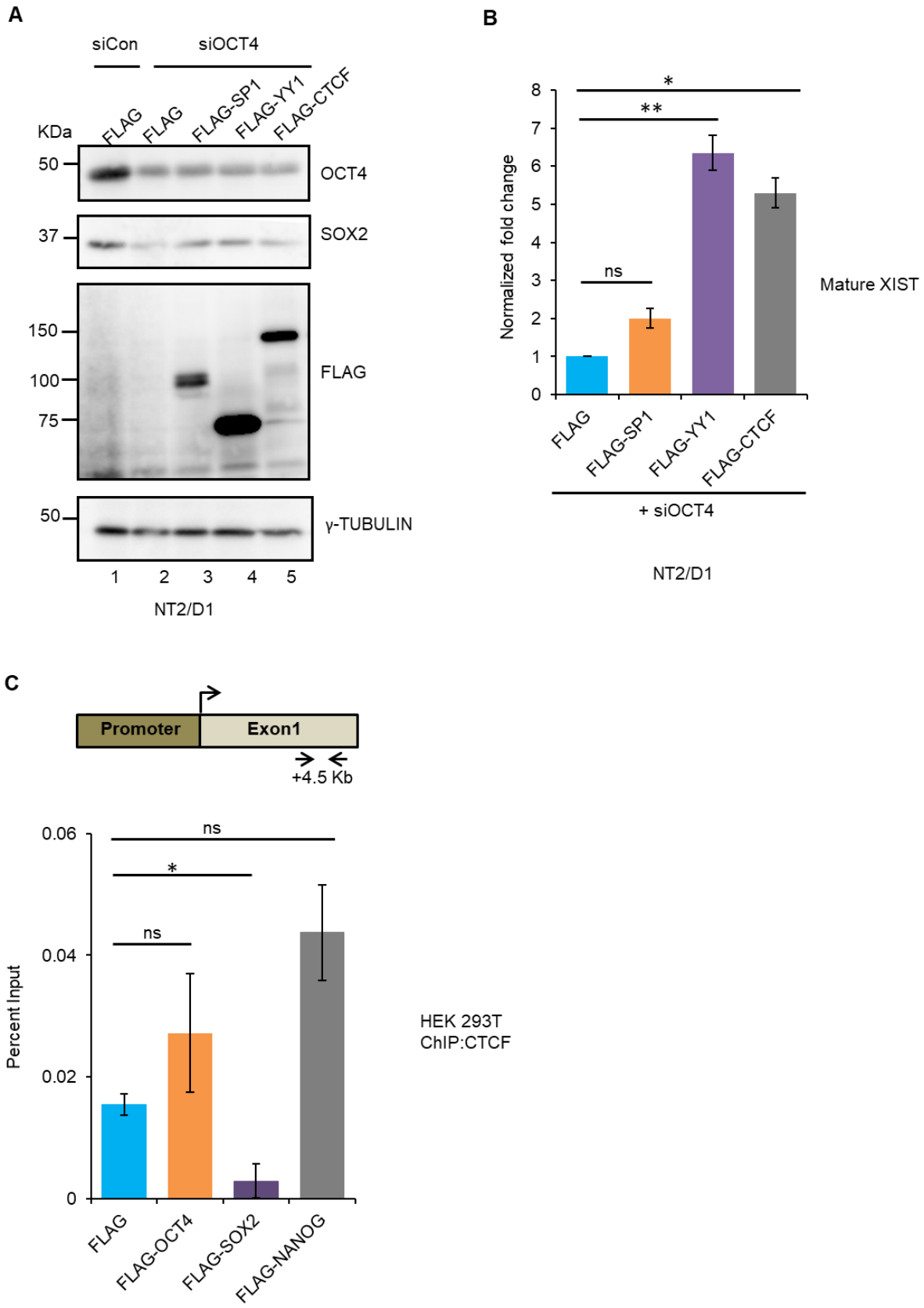


Figure 5. CTCF activates *XIST* by binding to cRE. (A) Immunoblotting for OCT4 and SOX2 upon the knockdown of OCT4 and for FLAG to confirm the over-expression of SP1, YY1, CTCF in NT2D1 cells. γ -TUBULIN serves as a loading

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 enrichment calculated as percent input. Each bar represents values from 2 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).

CTCF binding to the cRE assists in stabilizing YY1 binding at the *XIST* promoter-proximal site through chromatin looping

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 **(Figure S4A,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 knockdown of CTCF in HEK 293T cells and indeed observed a significant decrease in YY1 occupancy on its previously identified cognate site **(Figure 6A)** (51, 52). We also observed a decline in H3K27ac and an increase in H3K27me3 enrichment on the YY1 binding site **(Figure S6A)**. Furthermore, the decrease in XIST levels observed upon knockdown of CTCF in HEK 293T cells is restored by over-expression of YY1 **(Figure 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. It is long 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. To determine if the newly identified cRE can interact with the promoter of XIST, a HiC-based interaction 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.

Such 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 (**Figure S8A**, 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 using libraries prepared from HEK 293T, HEK 293T cells upon CTCF knockdown and SOX2 overexpressing HEK 293T cells. Strategy for the assay marking CTCF peaks, restriction enzyme site and the primers are depicted schematically in **Figure 6D**. 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 PCRs. We observed that all the tested interactions occur at modestly higher frequencies in control HEK 293T cells compared to the conditions in which either CTCF was knocked down or SOX2 was overexpressed **(Figure 6E, Figure S8A-C)**. It is important to note that the knockdown of CTCF or the overexpression of SOX2 does not completely abrogate CTCF binding from the cRE (Figure 4E(iv), 5C, S10A-B). This explains the modest changes in the observed 3C interaction frequencies. Nevertheless, these results indicate 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 (-3.0 Kb) CTCF binding sites, and (ii) promoter-proximal (+1.5 Kb) and promoter-distal (-3.0 Kb) CTCF sites also interact with each other **(Figure 6E)**. Based on these results, it can be speculated that such a conformation presumably allows for the establishment of appropriate active histone marks (H3K27ac) uniformly **(Figure 4A E(v), Figure S6A,B)**, thereby ensuring persistent expression of XIST during the maintenance phase of XCI.

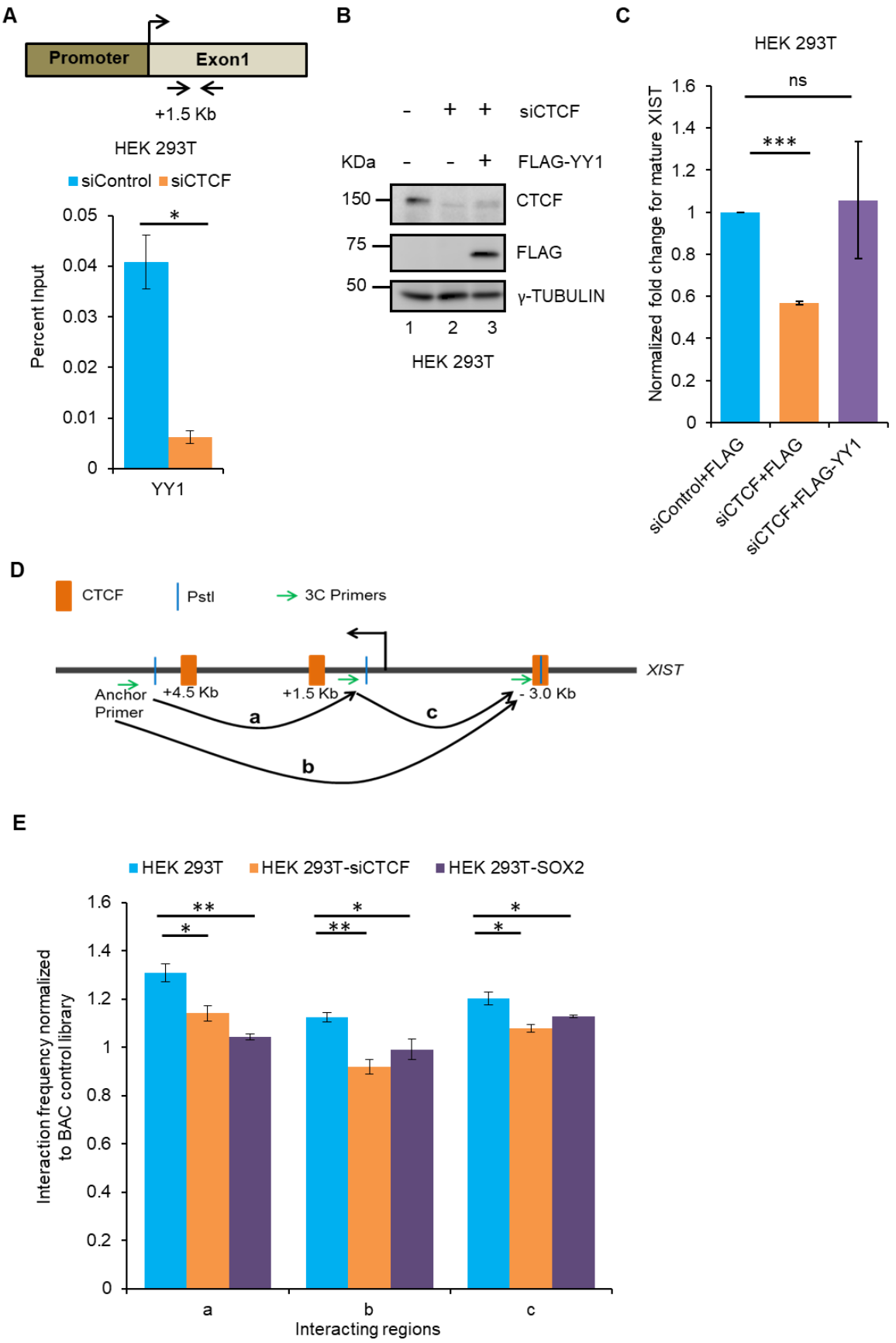

A
Promoter
Exon1
+1.5 Kb
HEK 293T
siControl
siCTCF
Percent Input
0.05
0.04
0.03
0.02
0.01
0
*
YY1
B
siCTCF
KDa
FLAG-YY1
150
75
50
CTCF
FLAG
γ-TUBULIN
1 2 3
HEK 293T
C
HEK 293T
ns
***
Normalized fold change for mature XIST
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
siControl+FLAG
siCTCF+FLAG
siCTCF+FLAG-YY1
D
CTCF
PstI
3C Primers
XIST
Anchor Primer
+4.5 Kb
+1.5 Kb
- 3.0 Kb
a
b
c
E
HEK 293T
HEK 293T-siCTCF
HEK 293T-SOX2
Interaction frequency normalized to BAC control library
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
a
b
c
Interacting regions

Figure 6. CTCF assists in the recruitment or maintenance of YY1 binding to the promoter-proximal region of *XIST* through chromatin looping. (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 knockdown 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). (D) 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. (E) 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).

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 (1–4). 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 (46). Extensive studies using the 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 (51, 52). Here, we have uncovered a complex regulatory network involving the pluripotency factors, YY1 and CTCF that shapes the transcriptional outcome from the *XIST* promoter **(Figure 7)**.

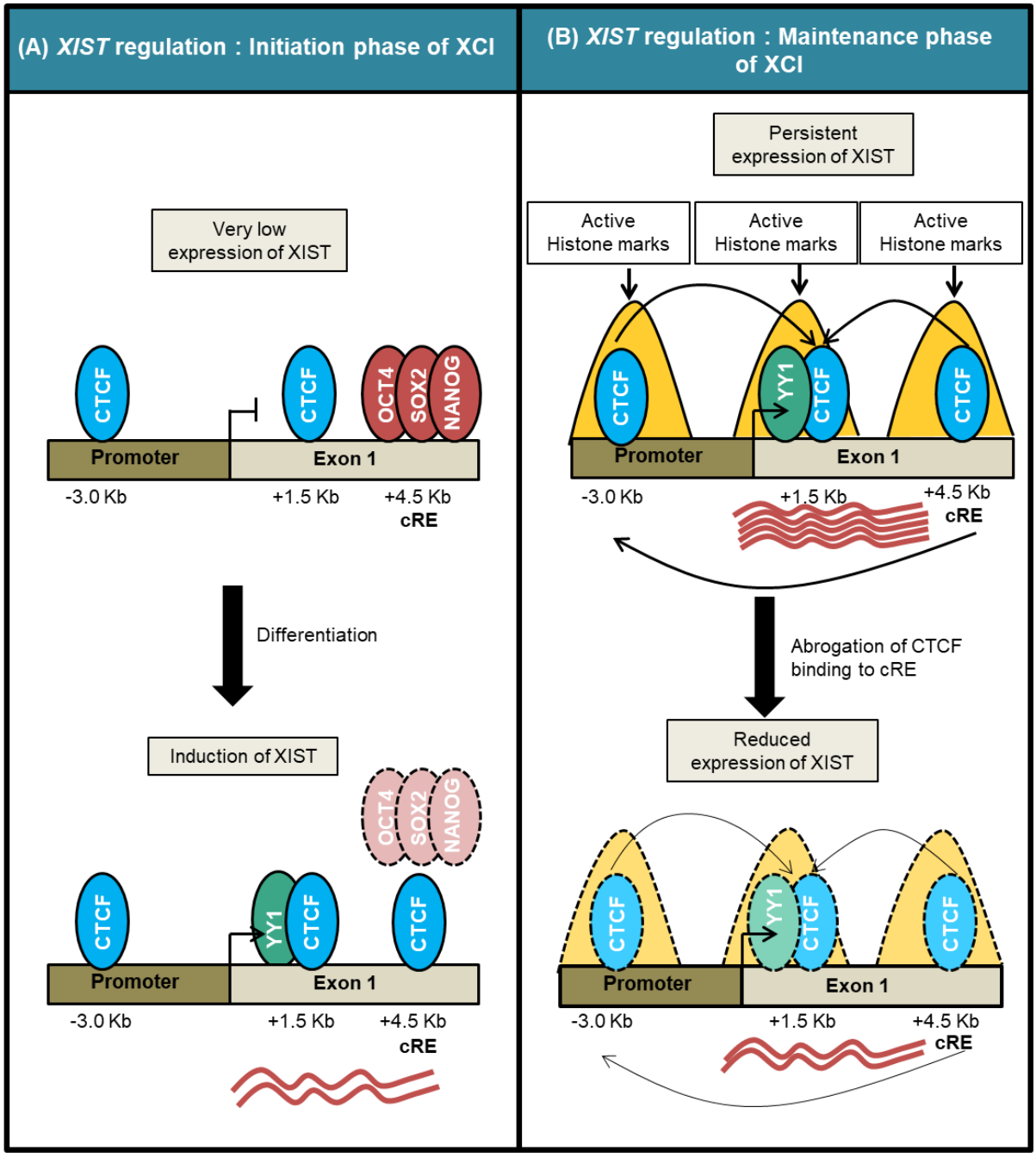


Figure 7. A model illustrating the role of CTCF bound cRE in dictating the transcription from *XIST* promoter. (A) In undifferentiated ES cells, the cRE of *XIST* remains bound by the pluripotency factors (OCT4, SOX2, NANOG). Upon differentiation, enrichment of pluripotency factors on the cRE decrease which is now presumably occupied by CTCF. Subsequently, YY1 also occupies the promoter leading to the upregulation of XIST marking the initiation phase of XCI. (B) In the maintenance phase of XCI , 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 knockdown, the relative frequencies of all the three interactions decreases. This disrupts the loop architecture 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 (63, 64). However, a few of the recent studies describe a method of culturing human ES cell lines, exhibiting consistent XIST expression as well as XCI upon differentiation (65, 66). 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 harbours properties similar to the embryonic stem cells and has been extensively used to understand the human stem cell biology (67–69). More importantly, it also expresses low levels of XIST (57) analogous to that seen for human and mouse ES cells as well blastocyst (5, 46, 47, 70–72). 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*. 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. It was observed that while XIST expressed in induced upon differentiation of NT2/D1 cells, the promoter activities showed a downward trend. The latter result

correlated with the levels of YY1 decreasing, thereby corroborating the previously reported role of YY1 in positively regulating the *XIST* promoter (50). 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 (33, 35, 36). Upon reanalysing published ChIP-sequencing datasets and the experimental validation, we observed consistent enrichment of the pluripotency factors at +4.5 Kb region in differentiation-dependent manner**.** We also demonstrate that the overexpressed pluripotency factors can bind the +4.5 Kb (as measured from TSS) site in the cells with consistent expression of XIST (HEK 293T) and downregulate XIST transcriptionally. 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 (51, 52). We believe that 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 – H3K27ac, as well as the known chromatin organizer protein CTCF 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 (51). However, the results obtained in our study unequivocally demonstrate it to be specifically enriched on the cRE only on Xi or XIST-induced cells. Moreover, it is possibly required for the higher enrichment of H3K27ac not just on the cRE but even on the promoter-proximal (+1.5 Kb) as well as distal (-3.0 Kb) CTCF binding sites. This is consistent with the evidence of CTCF playing the causal role in erasing H3K27me3 mark and making the chromatin more accessible (73) and mediating the chromatin interactions between these sites as verified by 3C-based assay. Consequently, YY1 binding to the promoter-proximal

region is dependent on the CTCF binding to the cRE. 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.

Additionally, by performing knockdown and over-expression experiments, we show that XIST is positively regulated by CTCF in addition to YY1. Interestingly, XIST levels increase dramatically when the over-expression of YY1 or CTCF is coupled with the knockdown 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. Indeed we show that CTCF bound to cRE is displaced upon overexpression of SOX2 in HEK 293T cells as revealed by ChIP. 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 **(Figure 7 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.

In conclusion, we have identified a novel 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.

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 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 Table 1**. Gel images for end-point PCR for all the ChIP experiments can be found in **Figures S9-S10**).

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. (53, 54). 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)

Sequencing 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 (55) to the hg19 reference assembly. Following which macs2 (56) 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 S8A 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 S8A.

DATA AVAILABILITY

Files for the datasets used for analysis can be downloaded from the link (https://www.ebi.ac.uk/ena)

ACCESSION NUMBERS

H9 cells: OCT4 ChIP-seq (SRR5642847), Nanog ChIP-seq (SRR5642845)
HEK 293 cells: CTCF ChIP-seq (DRR014670)
DLD1 cells: CTCF ChIP-seq (DRR014660)
Female breast epithelium cells (ENCODE): CTCF ChIP-seq (SRR6213076, Input SRR6213541)
Male breast epithelium cells (ENCODE): CTCF ChIP-seq (SRR6213724, Input SRR6214303)
NT2/D1: OCT4 ChIP-seq (SRR1640282)
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)

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.

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. SG interpreted data, supervised the project and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

Work was supported by the Centre of Excellence in Epigenetics program (Phase II) of the Department of Biotechnology (BT/COE/34/SP17426/2016), Government of India, and the JC Bose Fellowship (JCB/2019/000013) from the Science and Engineering Research Board, 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.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. 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.

2. 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.

3. 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.

4. Lyon,M.F. (1961) Gene Action in the X-chromosome of the Mouse (Mus musculus L.). *Nature*, **4773**, 372–373.

5. 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.

6. 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.

7. 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.

8. 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**, 1361–73.

9. Plath,K. (2003) Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science.*, **300**, 131–135.

10. 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**, 9665–8.

11. 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.

12. 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**, 1528–39.

13. 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.

14. 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**, 10104–9.

15. 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.

16. 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.

17. Hall,L.L., Byron,M., Sakai,K., Carrel,L., Willard,H.F. and Lawrence,J.B. (2002) An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 8677–82.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. Brockdorff,N. and Turner,B.M. (2015) Dosage Compensation in Mammals. *Cold Spring Harb. Perspect. Biol.*, **6**, 1–25.

24. Lee,J.T. and Lu,N. (1999) Targeted Mutagenesis of Tsix Leads to Nonrandom X Inactivation. *Cell*, **99**, 47–57.

25. 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.

26. 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.

27. 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**, 227–35.

28. Sado,T., Hoki,Y. and Sasaki,H. (2005) Tsix silences Xist through modification of chromatin structure. *Dev. Cell*, **9**, 159–165.

29. 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., *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.

30. Bemmel,J.G. Van, Galupa,R., Gard,C., Servant,N., Picard,C., Davies,J., Szempruch,A.J., Zhan,Y., Żylicz,J.J., Nora,E.P., *et al.* (2019) The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. *Nat. Genet.*, **51**, 1024–1034.

31. Galupa,R., Nora, Pierre Elphege Worsley-hunt,R., Ohler,U., Giorgetti,L. and Heard,E. (2020) A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary. *Mol. Cell*, **77**, 1–16.

32. 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.

33. 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**, 1693–5.

34. 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.

35. 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.

36. 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**, 457–60.

37. 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.

38. Hall,L.L., Clemson,C.M., Byron,M., Wydner,K. and Lawrence,J.B. (2002) Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. *Hum. Mol. Genet.*, **11**, 3157–65.

39. 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.

40. 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.

41. Hendrich,B.D., Browp,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.

42. 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.

43. 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.

44. 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.

45. Chang,S.C. and Brown,C.J. (2010) Identification of regulatory elements flanking human XIST reveals species differences. *BMC Mol. Biol.*, **11**, 20.

46. 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**, 370–4.

47. Sousa,E.J., Stuart,H.T., Bates,L.E., Ghorbani,M., Nichols,J., Dietmann,S., Silva,C.R., Sousa,E.J., Stuart,H.T., Bates,L.E., *et al.* (2018) Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. *Cell Stem Cell*, **22**, 919–928.

48. 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. *Nat. Publ. Gr.*, **45**, 2012–2014.

49. Vallot,C., Ouimette,J.-F., Makhlouf,M., Feraud,O., Pontis,J., Come,J., Martinat,C., Bennaceur-Griscelli,A., Lalande,M. and Rougeulle,C. (2015) Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. *Cell Stem Cell*, **16**, 533–546.

50. 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.

51. 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.

52. 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**, 1–11.

53. 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.

54. Naumova,N., Smith,E.M., Zhan,Y. and Dekker,J. (2012) NIH Public Access. *Methods*, **58**, 1-27.

55. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.

56. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nussbaum,C., Myers,R.M., Brown,M., Li,W., *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**.

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

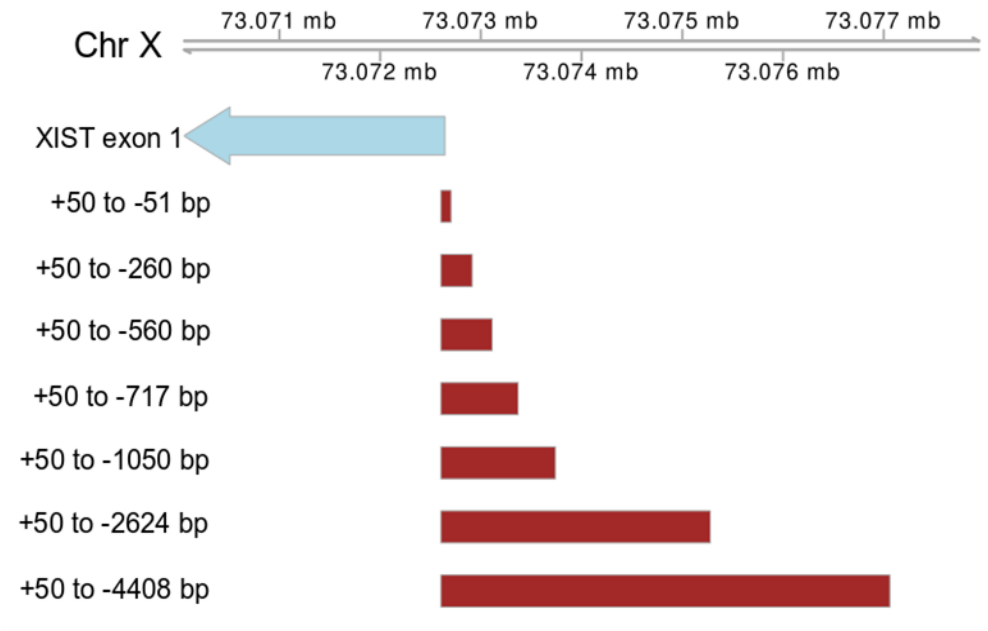
72. van den Berg,I.M., Laven,J.S.E., Stevens,M., Jonkers,I., Galjaard,R.J., Gribnau,J. and Hikke van Doorninck,J. (2009) X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos. *Am. J. Hum. Genet.*, **84**, 771–779.

73. 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.

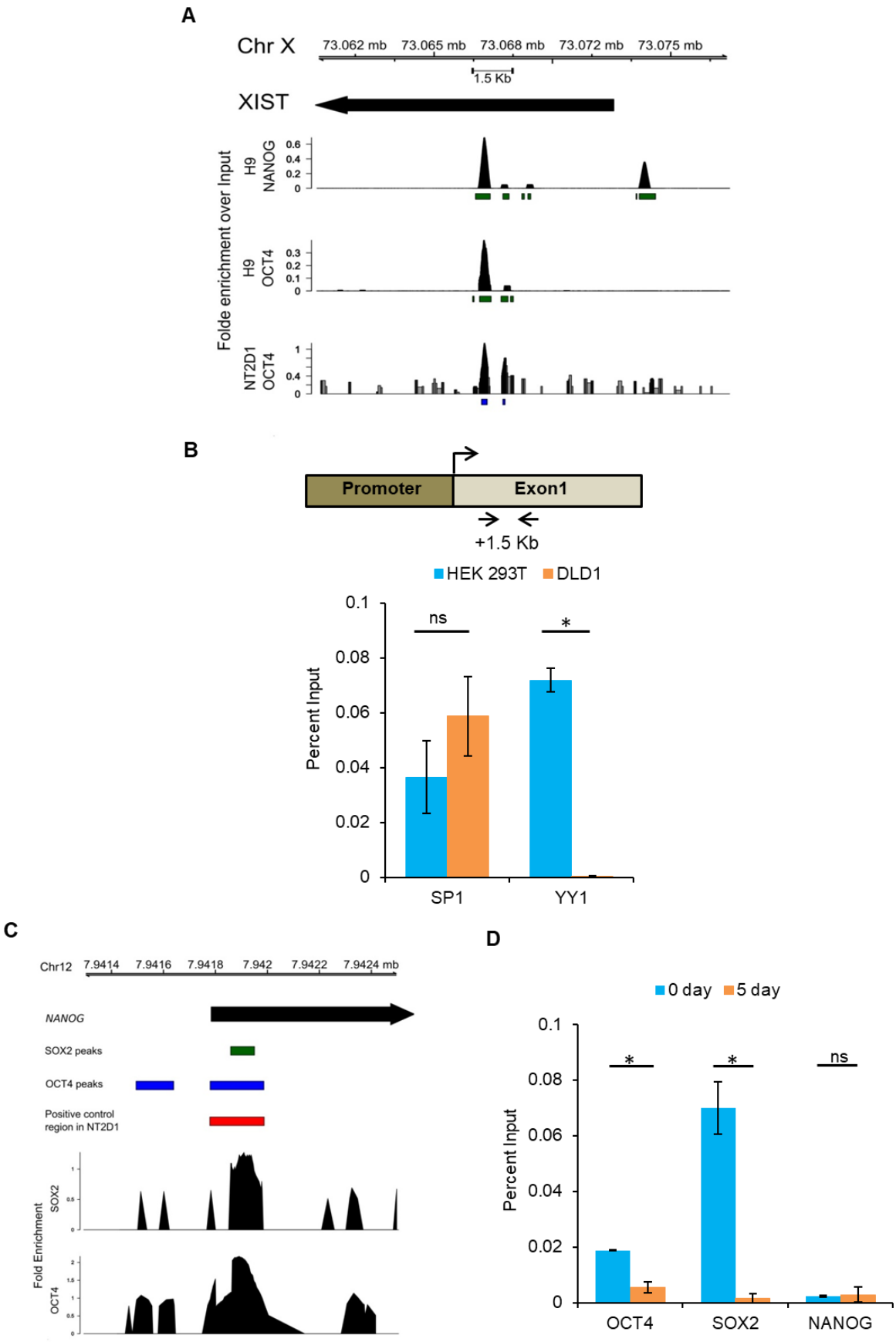
Supporting Information

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

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

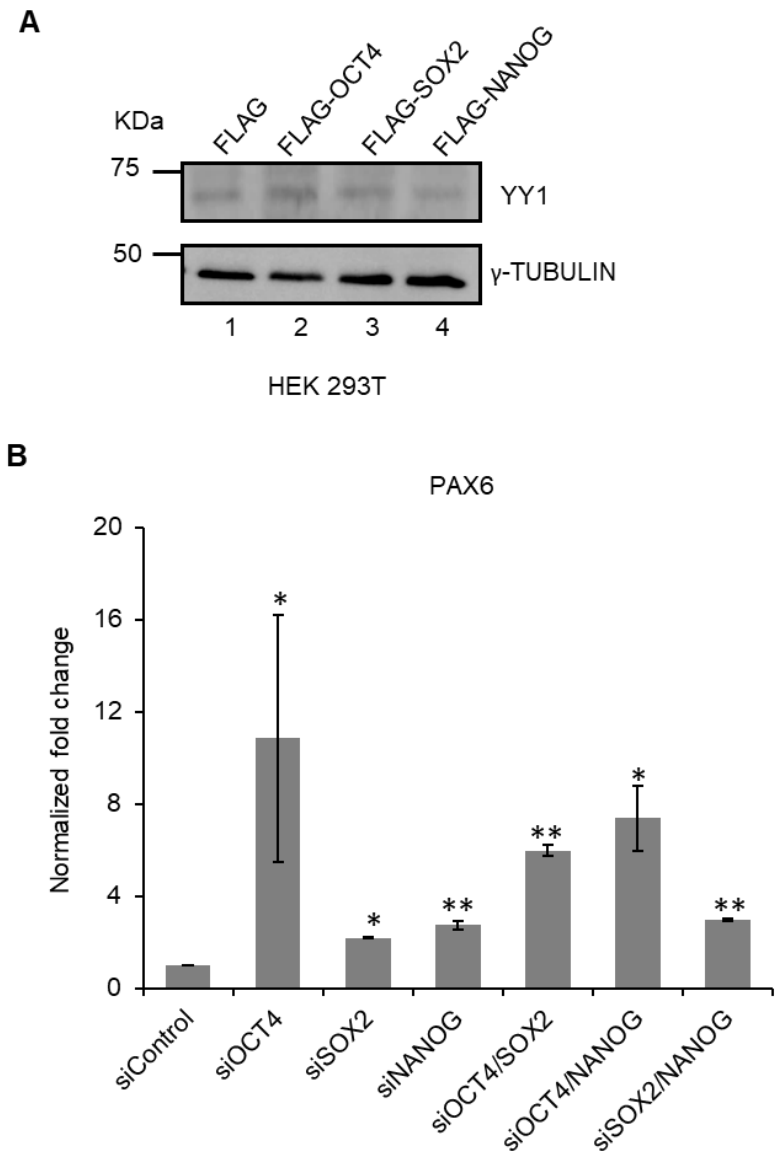


Supplementary figure 1. Promoter constructs for *XIST* cloned in the luciferase reporter vector pGL3 Basic. The brown boxes denote the fragments of the region upstream to XIST cloned into pGL3 Basic Luciferase reporter for testing potential promoter activity. The blue arrow indicates the first exon of *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).**

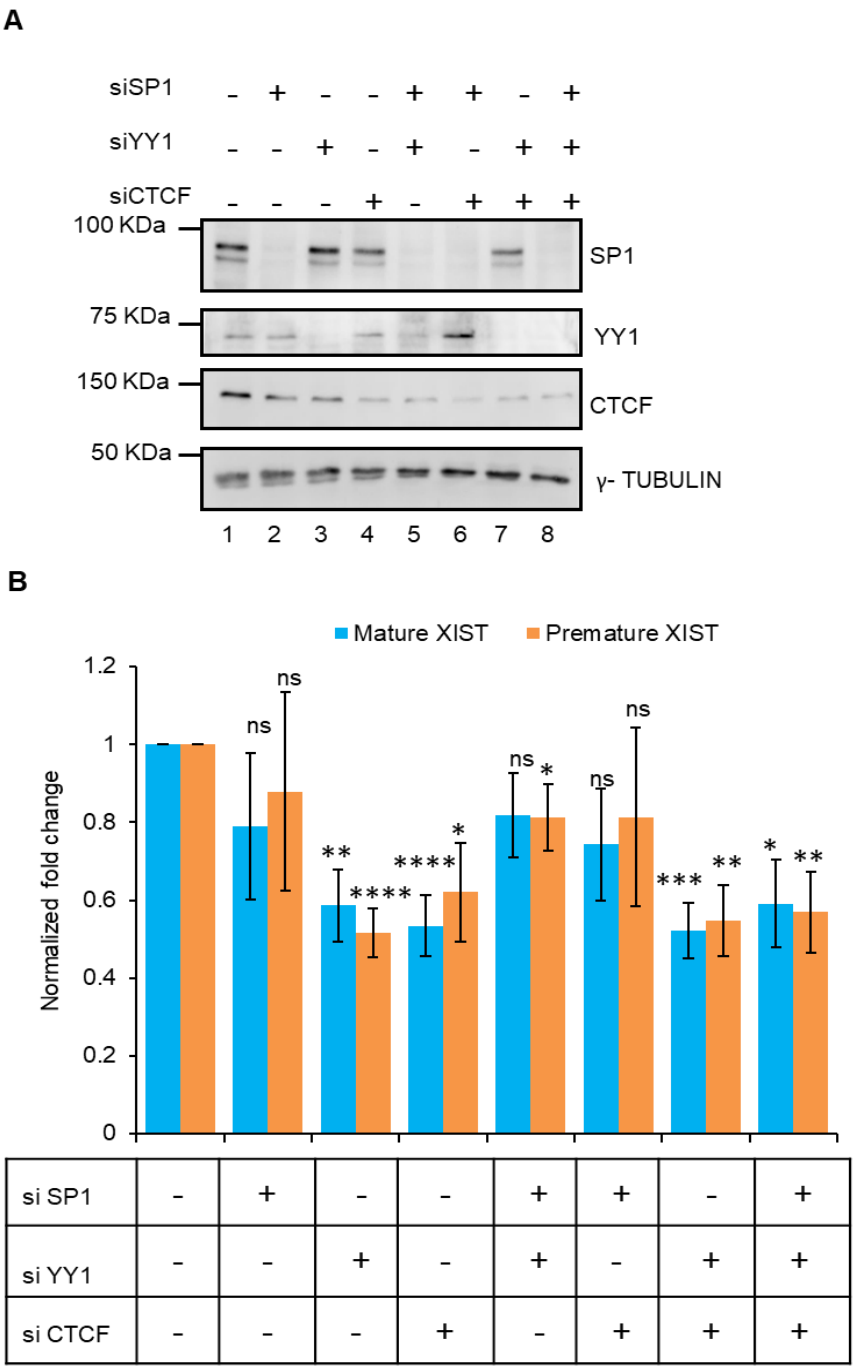


Supplementary figure 2. Induction of XIST during differentiation of NT2D1 cells is governed by promoter as well as exon1 of XIST. (A) ChIP-seq peaks for

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) 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. Each bar represents values from 2 independent experiments. Error bar represent the ±S.E.M. Asterisks represent the significance over DLD1 as per Student's T-test (*P value < 0.05, ns=non-significant) **(Related to Fig. 2E, F)**.

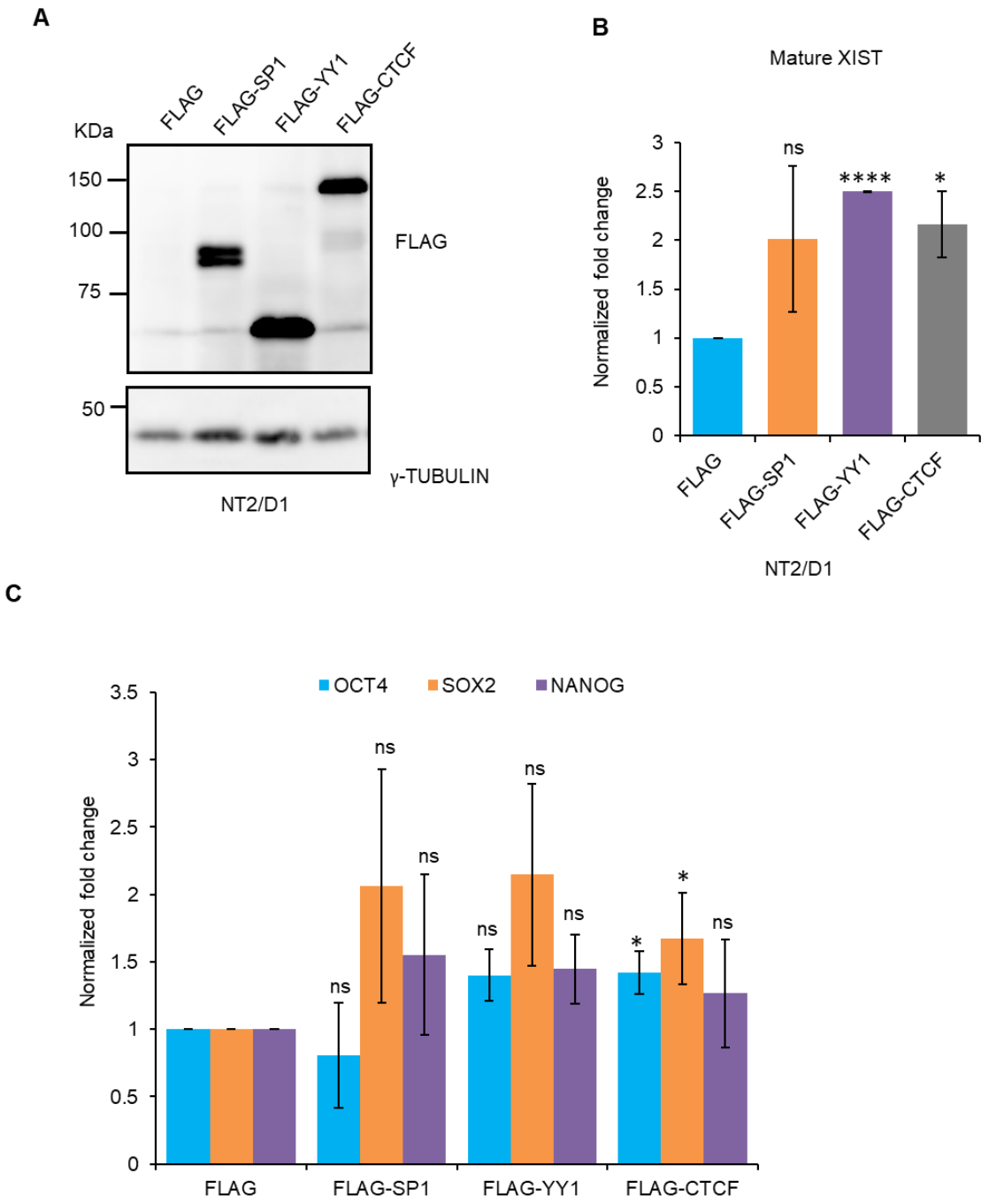


Supplementary figure 3. Pluripotency factors repress *XIST*. (A) Over-expression of pluripotency factors does not affect the levels of YY1 in HEK 293T cells. Immunoblotting for YY1 indicate that its expression does not change upon over-expression of pluripotency factors in HEK 293T cells. (Related to Fig. 3A,B). (B) qRT-PCR for PAX6 (upon siRNA mediated knockdown of OCT4, SOX2, NANOG in NT2/D1 cells (Related to Fig. 3D, E). 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)**.



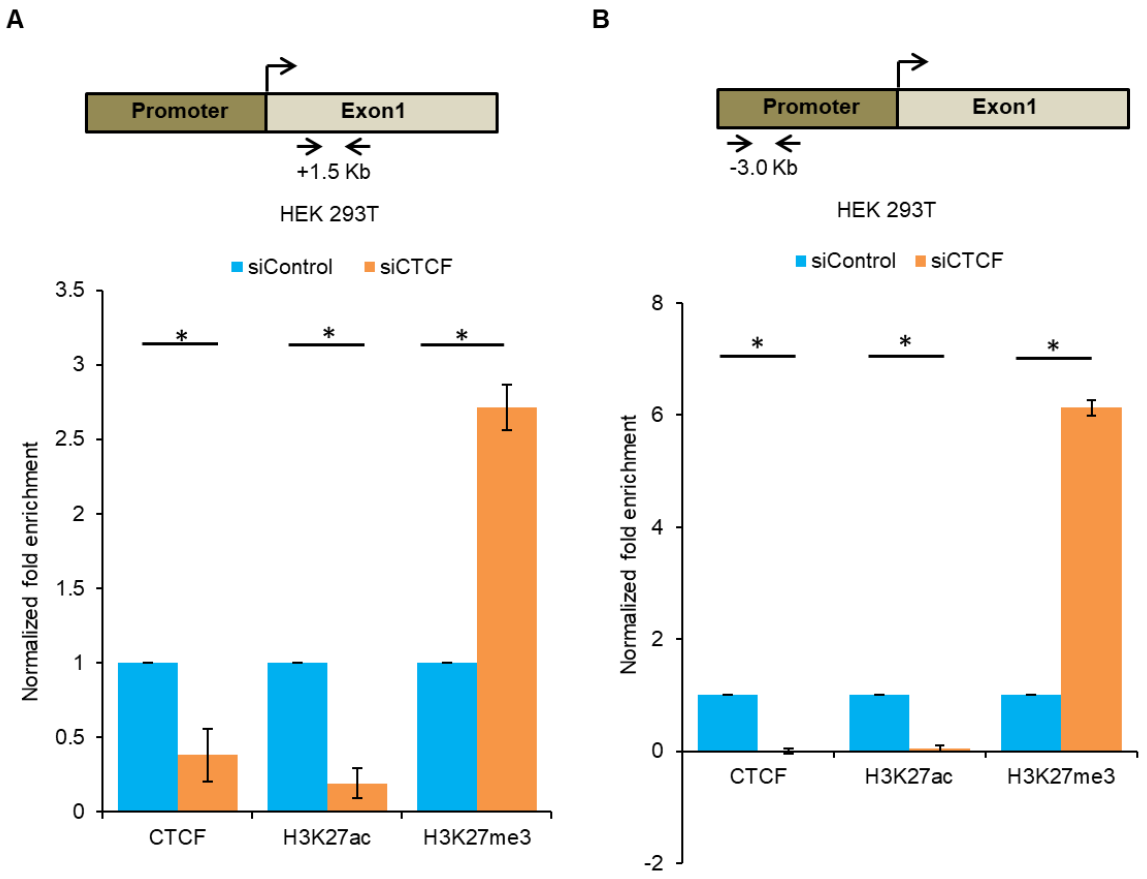
Supplementary figure 4. YY1 and CTCF positively regulate XIST transcription. (A) Immunoblotting to confirm knockdown of SP1, YY1, CTCF in HEK 293T cells. γ-TUBULIN serves as an equal loading control. Western blot for SP1, YY1, CTCF for single knockdown and γ-TUBULIN is the same as in figure 4C. (B) qRT-PCR 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. **(Related to Fig. 4)**.

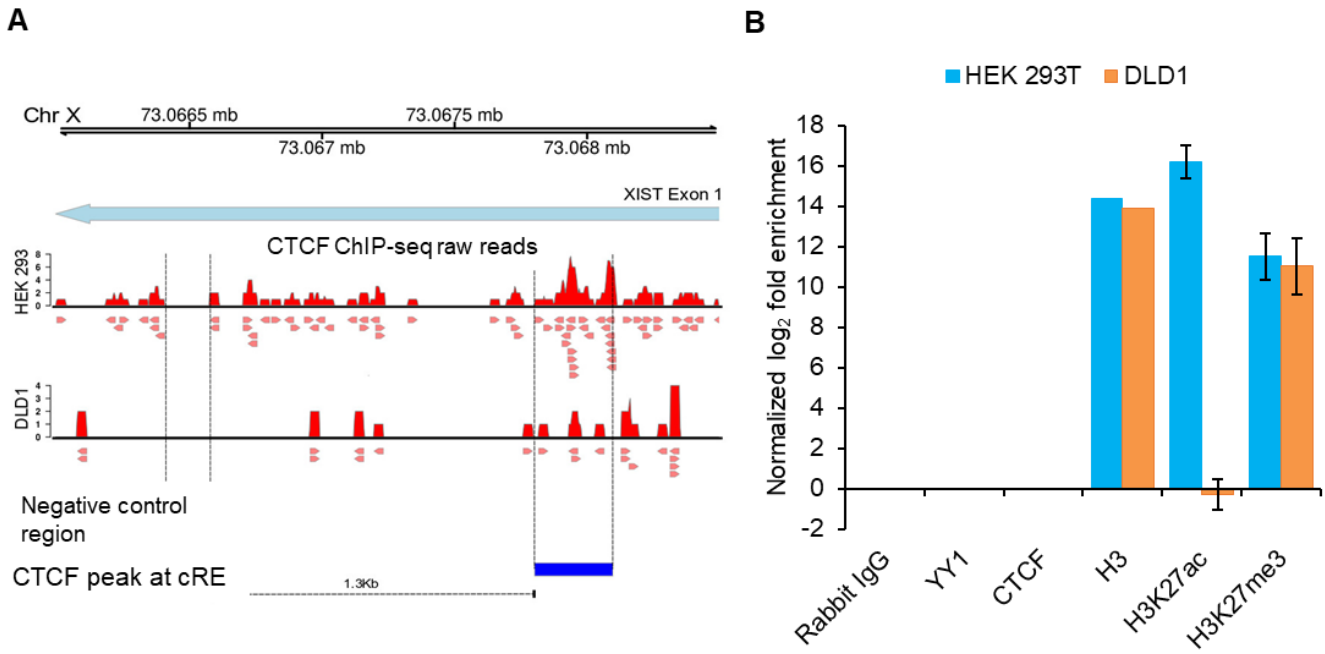


Supplementary figure 5. YY1 and CTCF positively regulate XIST transcription. (A) Immunoblotting to confirm over-expression of SP1, YY1, CTCF in NT2/D1 cells. γ-TUBULIN serves as an equal loading control. (B)

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)**.

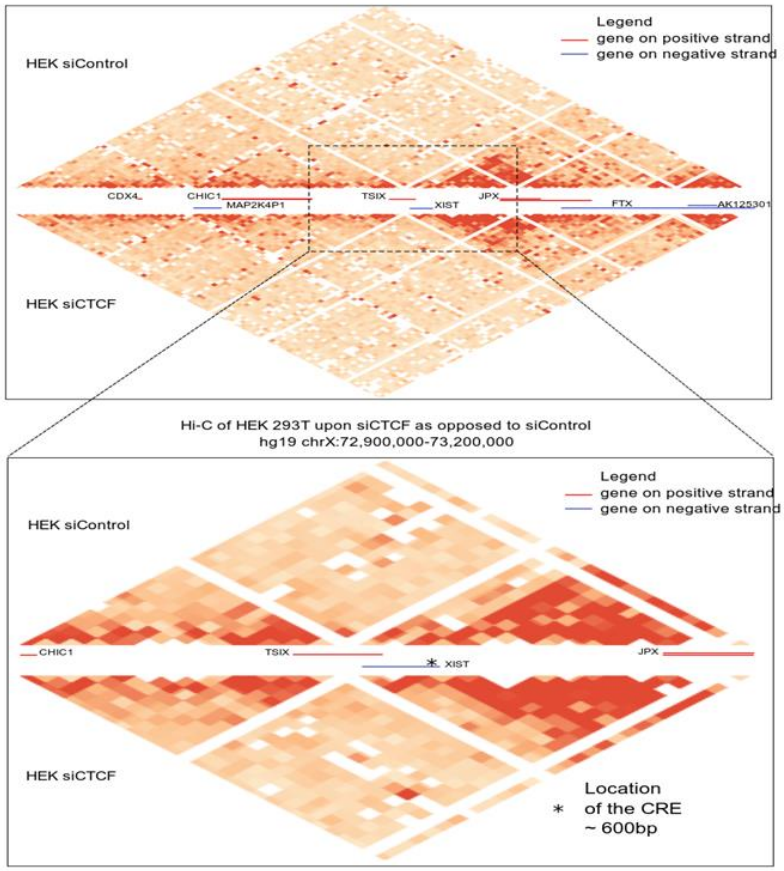


Supplementary figure 6. CTCF shapes the epigenetic landscape of *XIST* locus. (A, B) ChIP-qPCR analysis demonstrating a significant decrease in CTCF occupancy and H3K27ac enrichment and a significant increase in H3K27me3 enrichment on the promoter proximal (+1.5 Kb) (A) and promoter distal (-3.0 Kb) (B) sites upon knockdown of CTCF in HEK 293T cells. (*P value < 0.05) **(Related to Fig. 4E(v))**.

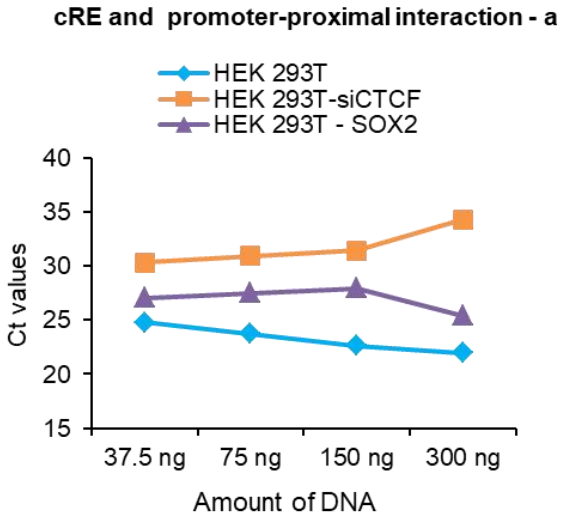


Supplementary figure 7. Controls for ChIP-qPCR analysis. (A) The locations of the negative control for ChIP shown overlayed with the reads from the ChIP-seq raw data. The negative control is a region which is devoid of reads and does not contain any repeat regions. The distance between the negative control 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 (**Related to Fig. 4E**).

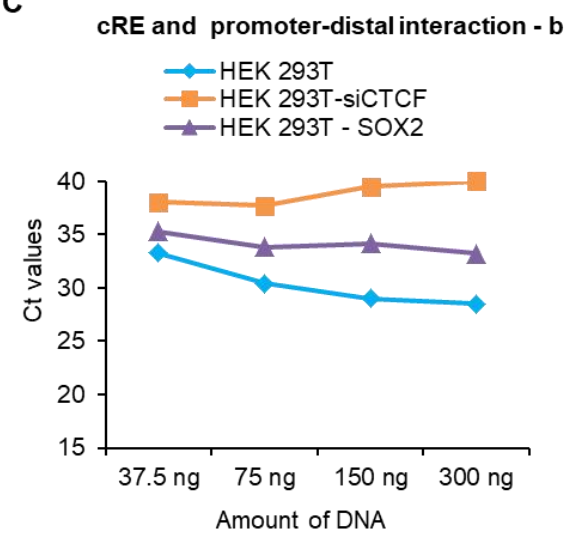
A



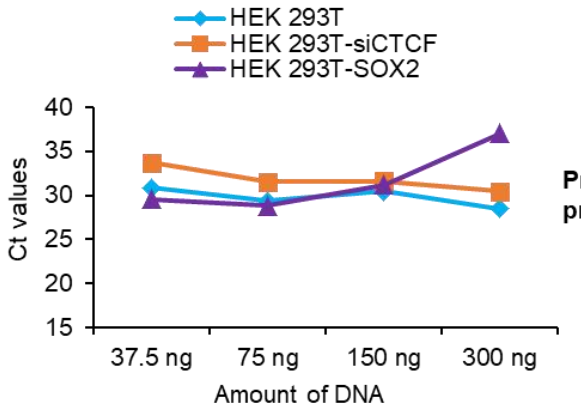
B



C

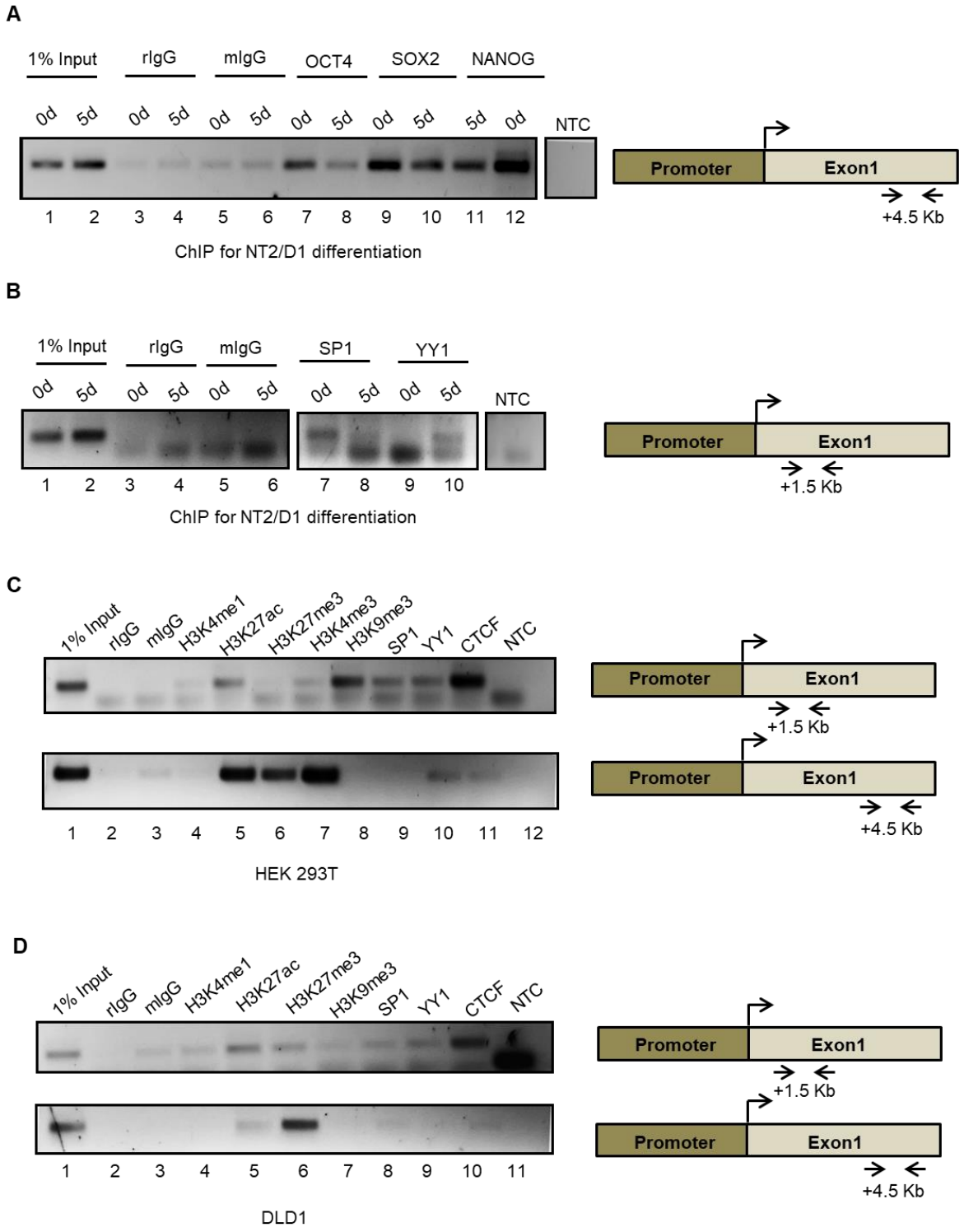


D



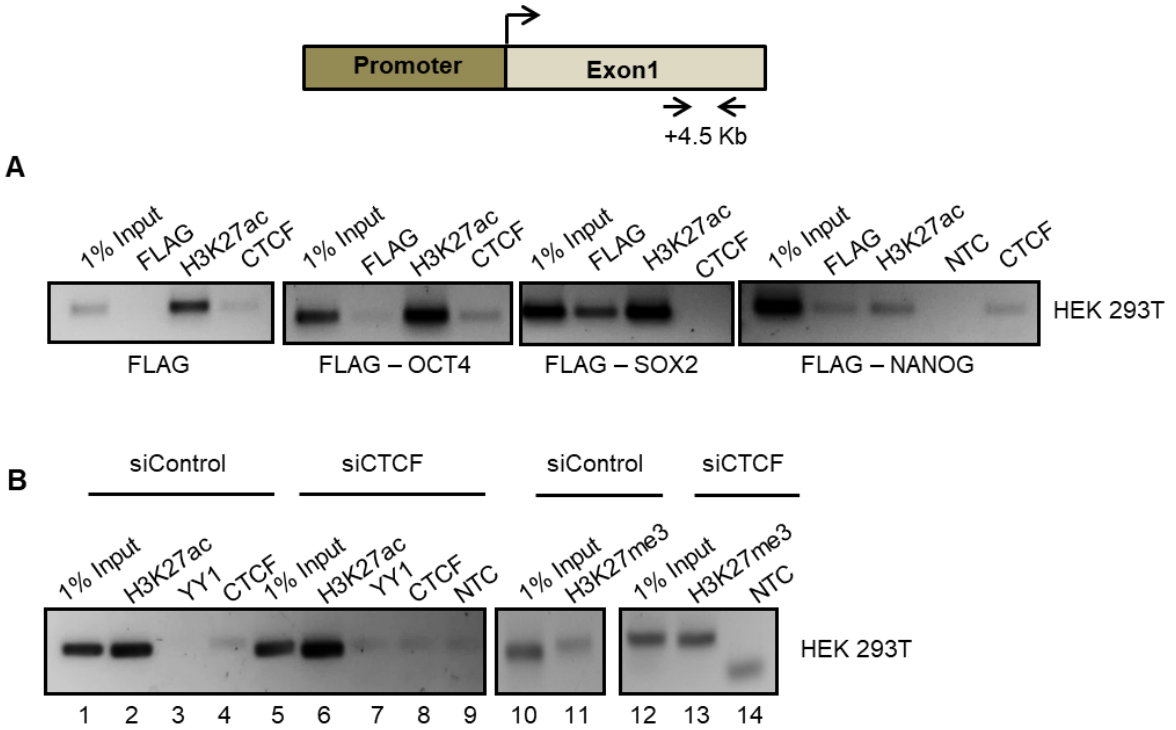
Promoter-proximal and promoter-distal interaction -c

Supplementary figure 8. Chromatin interation between cRE and the promoter-proximal and promoter-distal CTCF binding sites is modulated by CTCF and SOX2. (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-D) 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) (B), cRE to promoter-distal interaction (+4.5 Kb to -3.0 Kb) (C) and promoter-proximal and promoter-distal interaction (+1.5 Kb to -3.0 Kb) (D). 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 Fig. 6)**.



Supplementary figure 9. 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 – H3K4me1, 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(i,ii))**.



Supplementary figure 10. 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. 3C)** (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. 4E(v))**.

Supplementary Table 1. List of oligonucleotide primers used.

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	TGCTGCCTCTTTAAGACTAGGAC
Human SOX2, Reverse	CCTGGGGCTCAAACTTCTCT
Human NANOG, Forward	CCTGAACCTCAGCTACAAACAG

Human NANOG, Reverse	GCTATTCTTCGGCCAGTTGT
Human PAX6, Forward	GGCACACACACATTAACACACTT
Human PAX6, Reverse	GGTGTGTGAGAGCAATTCTCAG
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
Primers designed for 3C-PCR	
XIST locus 3C primer_Anchor primer (+4.5Kb) (hg19,chrX:73066941 - 73066970)	GTACTTCCAGCTGGGATGTAAATACAGTGG
XIST locus 3C primer_Test primer (+1.5Kb) (hg19,chrX:73071546 - 73071576)	GCAATGCACATGACTTCCTCTGCCTGACCTG
XIST locus 3C primer_Test primer (-3.0Kb) (hg19,chrX:73075439 - 73075468)	GATTCAGATTATTCCAGTTCTAGGATTGTC