

Systematic Allelic Analysis Defines the Interplay of Key Pathways in X Chromosome Inactivation

Tatyana B Nesterova^{1,6}, Guifeng Wei^{1,6}, Heather Coker^{1,6}, Greta Pintacuda^{1,2,6}, Joseph S Bowness¹, Tianyi Zhang¹, Mafalda Almeida¹, Bianca Bloechl¹, Benoit Moindrot^{1,3}, Emma J Carter¹, Ines Alvarez Rodrigo^{1,4}, Qi Pan¹, Ying Bi⁵, Chun-Xiao Song⁵, and Neil Brockdorff^{1*}.

1. Developmental Epigenetics, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom

2. present address; Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Avenue, Cambridge MA 02138, USA

3. present address; I2BC Paris-Sud University, Gif-Sur-Yvette, France

4. present address; Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, United Kingdom

5. Ludwig Institute for Cancer Research, Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK

6. These authors contributed equally

*Correspondence to neil.brockdorff@bioch.ox.ac.uk

Abstract

Xist RNA, the master regulator of X chromosome inactivation, acts *in cis* to induce chromosome silencing through the stepwise recruitment of factors that modify underlying chromatin structure. Whilst considerable progress has been made towards defining key silencing factors and the elements to which they bind, their relative contribution to silencing different genes, and their relationship with one another is poorly understood. Here we describe a systematic analysis of Xist-mediated allelic silencing in ES cell-based models. We show that Spen, recruited through the Xist A-repeat, plays a central role, being critical for silencing of all except a subset of weakly expressed genes. Polycomb, recruited through the Xist B/C-repeat, also plays a key role, favouring silencing of genes with pre-existing H3K27me3 chromatin. LBR and the Rbm15/ m6A-methyltransferase complex, previously proposed to have a central role, make at most a minor contribution to gene silencing. We integrate our findings in a comprehensive model for Xist-mediated chromosome silencing.

Introduction

X chromosome inactivation is a developmentally regulated process that evolved in mammals to equalise the levels of X-linked gene expression in XX females relative to XY males¹. X inactivation is established by the X inactive specific transcript (Xist), a 17kb non-coding RNA, which accumulates *in cis* across the entire inactive chromosome elect²⁻⁵. Xist RNA recruits several factors that collectively modify chromatin/chromosome structure to silence transcription⁶. Xist transgenes can recapitulate Xist function in the context of an autosome^{5,7}, albeit with somewhat reduced efficiency⁸⁻¹⁰.

Xist-mediated chromosome silencing occurs in a stepwise manner with specific chromatin/chromosome modifications occurring concurrent with the onset of Xist RNA expression, and others at later timepoints¹¹. Several silencing factors have been identified, including in recent studies, RNA binding proteins (RBPs) that interact directly with specific elements within Xist RNA. The RBP Spen has been identified as an important factor for establishment of Xist-mediated silencing¹²⁻¹⁵, functioning by recruiting the NCoR-HDAC3 complex to induce chromosome-wide histone deacetylation¹⁵. Spen binds to the Xist A-repeat, a short tandem repeat region at the 5' end of the transcript^{12,16}, that is a key element required for Xist-mediated silencing¹⁷. Two other factors whose recruitment is linked to the A-repeat are Wtap, a regulatory subunit of the N6-Methyladenosine (m6A) methyltransferase complex^{13,14}, and Rbm15, an RBP related to Spen (Patil et al., 2016; Moindrot et al., 2015). Wtap and Rbm15 have moreover been linked with one another, with the latter directly recruiting the m6A-methyltransferase complex via binding to Xist A-repeat (Patil et al., 2016). Accordingly, several studies have identified sites of m6A deposition in Xist RNA¹⁸⁻²¹. Both Rbm15 and the m6A-methyltransferase (m6A-MTase) complex have been implicated in Xist-mediated silencing^{13,14,21}.

A third recently identified silencing factor is the Lamin B receptor (LBR), mutation of which has been reported to strongly abrogate Xist-mediated silencing^{15,22}. LBR binds to several sites across Xist RNA, as determined by CLIP-seq, with the most prominent site, referred to as LBS, being located around 0.2 kb downstream of the A-repeat. Deletion of LBS results in abrogated silencing which is complemented by tethering LBR to the mutant RNA²². However, *Lbr* mutations in mouse do not show a female-specific phenotype^{23,24}, suggesting that the role of the protein in X inactivation requires further investigation.

In addition to the aforementioned factors, Xist RNA recruits the Polycomb repressive complexes PRC1 and PRC2 that catalyse the histone modifications H2AK119ub and H3K27me3 respectively²⁵⁻²⁷. Polycomb recruitment is initiated by the RBP hnRNPK which binds to a different Xist RNA region, the B/C-repeat (herein PID region),¹⁰. hnRNPK interacts with the Pcgf3/5-PRC1 complex which initiates a Polycomb cascade resulting in recruitment of other PRC1 complexes and PRC2²⁸. Polycomb is important for chromosome-wide silencing by Xist RNA²⁸, although the relative contribution of PRC1 and PRC2 has not been determined.

Although key silencing factors and their binding sites have been identified, their relationship with one another and the relative contribution they make to repressing individual genes or gene subsets is not well understood. A key limitation is that each factor/pathway has been analysed in different systems, for example XY mouse embryonic stem cells (mESCs), XX mESCs, trophoctoderm tissue of XX embryos or XX somatic cells, and/or at different timepoints. Additionally, several different approaches have been used to assay silencing, including nascent RNA FISH/single molecule FISH of selected genes, chromosome-wide analysis, either with or without an interspecific background to determine allelic ratios, and indirect assays, for example escape from cell death following induction of Xist RNA

expression on the single X chromosome in XY mESCs. With this limitation in mind we have carried out systematic chromosome-wide analyses to define the contribution of all of the major silencing pathways at defined timepoints in two complementary models, interspecific mESCs carrying an inducible Xist transgene located either on an autosome (chromosome 3), or on the X chromosome in an XX mESCs. Our results demonstrate that establishment of Xist-mediated silencing is largely attributable to synergistic actions of Spen/A-repeat and Polycomb/B-repeat.

Results

The dynamics of Xist-mediated silencing in Chr3 and ChrX mESC models

To define the contribution of different pathways in Xist-mediated silencing we analysed two complementary models, a previously described XY interspecific M.m. domesticus (129S) x M.m. castaneus (Cast) mESC line with a multi-copy tetracycline inducible Xist transgene randomly integrated on chromosome 3^{10,29}, referred to herein as iXist-Chr3, and a novel 129S x Cast XX ES cell line, (iXist-ChrX), with a single endogenous *Xist* allele driven by a tetracycline inducible promoter (Fig. 1a). Thus, an expression construct encoding the TetOn transactivator was introduced into the TIGRE locus³⁰, enabling efficient and synchronous induction of Xist RNA (129S allele) in ~95% of cells (Supplementary Fig. 1a). Subsequent allelic sequencing analysis (see below) revealed that a recombination event between 129S and Cast X chromosomes occurred in the process of modifying the Xist promoter such that informative SNPs are present only for the 103Mb proximal to *Xist* (Fig. 1a).

Xist-mediated gene silencing was assayed using ChrRNA-seq, a sensitive method that utilises abundant intronic in addition to exonic SNPs to enable accurate determination of allelic expression of the majority of genes³¹. Thus, we analysed iXist-Chr3 mESCs 1 day and 3 days after Xist RNA induction, the latter in differentiating conditions. Subsequent timepoints were not analysed as selection against autosomal monosomy progressively depletes the proportion of cells with detectable Xist RNA expression. The location of the Xist transgene (125.8097Mb-125.8145 Mb, mm10, Cast allele) was determined using chimeric Xist-Chr3 paired-end sequence tags from ChIP-seq analysis (see below). Silencing levels were determined for genes which had at least 10 SNP informative reads (n~500). As shown in Fig. 1b, c, at 1 day we observed silencing of most genes chromosome-wide, albeit to a

reduced degree at the distal end of the chromosome. Silencing was further reinforced at 3 days (Fig. 1b, c).

iXist-ChrX mESCs were analysed after 1 day of Xist RNA induction, and then at a later timepoint, 6 days, under differentiating conditions. Silencing levels were again determined for genes which had at least 10 SNP informative reads ($n \sim 250$). At 1 day we observed significant silencing of most genes using both allelic analysis for *Xist* proximal genes (Fig. 1d, e), and using non-allelic analysis for the entire ChrX (Supplementary Fig. 1b, c). Overall silencing efficiency was greater than that observed with iXist-Chr3 (Fig. 1f), consistent with previous reports that Xist RNA functions more efficiently on ChrX⁸⁻¹⁰. Silencing in iXist-ChrX cells was more robust after 6 days of Xist RNA induction (Fig. 1d, e), although interestingly, for many genes silencing is incomplete relative to XX somatic cells (Supplementary Fig. 1d).

We went on to explore if specific gene/chromatin characteristics (2D/3D distance relative to *Xist*, mESC gene expression level, and chromatin features of genes centred on promoters), correlate with silencing efficiency in iXist-Chr3 and iXist-ChrX cells 1 day after Xist RNA induction. In the case of 3D distance from *Xist*, recent work has shown that Xist RNA spreading on the X chromosome is initially determined by topological proximity to the *Xist* locus³². We were able to demonstrate a similar correlation for iXist-Chr3 cells using, as a surrogate marker, Polycomb-mediated histone modifications, H2AK119ub and H3K27me3, determined using allelic calibrated ChIP-seq (Supplementary Fig. 1e). Using a published cohesin HiChIP dataset³³, we found a close correlation ($r=0.76-0.77$) between Xist-dependent gain of Polycomb-mediated histone modifications and topological proximity to the *Xist* transgene (Supplementary Fig. 1e, f). To classify gene promoter chromatin states, we employed ChromHMM³⁴ to integrate multiple ENCODE ChIP-seq datasets defining

chromatin states in mESC. Genes were further classified according to the chromatin state surrounding the TSS region (see STAR methods).

To rank the contribution of the aforementioned parameters to silencing efficiency after 1 day of Xist RNA induction, we applied a machine learning approach to train a classifier to discriminate high or low silencing efficiency based on combinatory features for both iXist-Chr3 and iXist-ChrX lines (see STAR Methods), and then scored the importance of each feature (Fig. 1g, h). The most informative feature for Xist-mediated chromosome silencing is 2D/3D proximity to *Xist*. The correlation for 3D proximity in relation to silencing efficiency is shown for iXist-Chr3 ($r=0.41$) and iXist-ChrX ($r=0.41$) (Supplementary Fig. 1f, g). mESC gene expression levels also ranked as being of relatively high importance (Fig. 1g, h). Further analysis demonstrated that this reflects more efficient silencing of weakly expressed genes (Fig. 1i). Accordingly, promoter proximal chromatin states that correlated with silencing efficiency both in iXist-Chr3 and iXist-ChrX were state 7 (pre-active, negative correlation) and state 5 (pre-H3K27me3, positive correlation). Further analysis shows that this reflects more efficient silencing of pre-H3K27me3 associated genes (typically weakly expressed), and less efficient silencing of pre-active genes (Fig. 1j, k, Supplementary Fig. 1h, i).

A key role for *Spn* and the Xist A-repeat in chromosome-wide gene silencing

We generated a null allele of *Spn* by CRISPR/Cas9 mutagenesis essentially as described previously¹², initially in iXist-Chr3 mESCs (see STAR methods). iXist-Chr3 *Spn* null cells were viable and morphologically indistinguishable from unmodified controls. Robust Xist RNA domains were formed in the majority of cells following transgene induction (Supplementary Fig. 2a), although levels of Xist RNA were reduced relative to wild-type (WT) iXistChr3 mESCs (Supplementary Fig. 2b). We went on to analyse allelic silencing in

iXist-Chr3 Δ *Spn* mESCs 1 day after Xist RNA induction. Highly reproducible results, obtained using four independent *Spn* null mESC lines (STAR methods), were thus averaged. As shown in Fig. 2a, b, silencing of genes across chromosome 3 was strongly abrogated in the iXist-Chr3 *Spn* null cells, in most cases to the extent that expression levels were indistinguishable from uninduced control cells. To quantify our findings we defined thresholds to assign silencing degree as high, low and weak/none (STAR methods). On this basis a subset of chromosome 3 genes were silenced (24.8%; 127/512). In most cases the degree of silencing was classified as low (119/127). Further analysis revealed that genes subject to silencing in *Spn* null cells are generally weakly expressed in mESCs (Fig. 2c, left panel).

We also deleted *Spn* in iXist-ChrX mESCs. Viability and ES cell morphology were again unaffected, as was the frequency of cells inducing Xist RNA expression 1 day after addition of doxycycline (Supplementary Fig. 2a). The levels of induced Xist RNA were relatively low, similar to *Spn* null iXist-Chr3 cells (Supplementary Fig. 2b), and additionally, Xist RNA domains were in some instances relatively weak and diffuse compared with WT iXist-ChrX cells (Supplementary Fig. 2a). Allelic chromatin RNA-seq performed 1 day after Xist RNA induction (3 independent *Spn* null cell lines), demonstrated a dramatic reduction in Xist-mediated silencing, similar to *Spn* null iXist-Chr3 cells (Fig. 2d, e). Also, as noted for *Spn* null iXist-Chr3 cells, a subset of genes (40/261) exhibited a low degree of silencing. In most cases these genes are weakly expressed in mESCs (Fig. 2c, right panel).

We went on to compare the effect of *Spn* deletion in iXist-ChrX cells with deletion of the Xist A-repeat, the critical silencing element linked to *Spn* recruitment^{12,14}. A precise deletion of the A-repeat, (Xist Δ A), was engineered in iXist-ChrX cells using CRISPR/Cas9 mediated mutagenesis (Supplementary Fig. 2c). This deletion is somewhat shorter than

previously published A-repeat deletions, which in most cases additionally remove 185nt of sequence 3' to the A-repeat element. This point is significant in relation to other experiments described below. Xist RNA-FISH analysis demonstrated formation of Xist domains at the same frequency as in WT iXist-ChrX cells, although similar to *Spem* null cells, levels of Xist RNA were significantly reduced (Supplementary Fig. 2b). Allelic ChrRNA-seq in Xist Δ A mESCs after 1 day of Xist RNA induction demonstrated strongly abrogated silencing for most genes, similar to the *Spem* null iXist-ChrX cells (Fig. 2e, f). Induction of Xist RNA for 6 days in differentiating conditions gave similar results (Supplementary Fig. 2d, e). In both cases a subset of genes (12.3%; 32/261 at day 1) undergo some degree of silencing, and as noted for *Spem* null iXist-Chr3 and iXist-ChrX cell lines, these correlate with genes that are expressed weakly in mESCs (Fig. 2c, right panel). Most genes silenced in *Spem* null iXist-ChrX cells are also silenced in Xist Δ A mESCs ($p=0.012$; hypergeometric test, Supplementary Fig. 2f). Similarly, the subset of genes silenced in A-repeat mutant iXist-ChrX cells after 1 day or 6 day of Xist RNA induction also correlate (Supplementary Fig. 2g). Taken together, our findings demonstrate that Xist A-repeat and the A-repeat bound factor *Spem* play a major role in silencing at all except a small subset of weakly expressed genes.

Rbm15/m6A-MTase complex and LBR play a minor role in Xist-mediated silencing

Rbm15, a homologue of *Spem*, has been implicated in Xist-mediated silencing, and has been proposed to function by recruiting the m6A-MTase complex^{13,14,21}. To determine the contribution of Rbm15/m6A-MTase complex in Xist-mediated silencing we generated knockouts of the *Rbm15*, *Mettl3* and *Wtap* genes in iXist-Chr3 cells, using CRISPR/Cas9-mediated mutagenesis (Supplementary Fig. 3a). None of the mutations had detectable effects on Xist RNA domains (Supplementary Fig. 3b), or Xist RNA levels (Supplementary Fig. 2b). As above, we assessed Xist-mediated silencing using allelic ChrRNA-seq after 1 day of Xist RNA induction. Surprisingly, none of the three mutations abrogated gene silencing by Xist RNA, and in the case of *Rbm15* null mESCs, silencing efficiency was if

anything enhanced (Fig. 3a, b, Supplementary Fig. 3c). To investigate how the mutations affect m6A genome-wide we performed m6A quantitation using MS/MS and m6A-seq (according to the method of ³⁵, in the *Mettl3* and *Wtap* knockout lines. Notably, we found that whilst overall m6A levels are reduced (Supplementary Fig. 3d), residual m6A is detectable at some peaks, including the major peak in Xist RNA, which maps immediately downstream of the A-repeat (Supplementary Fig. 3e). This observation likely reflects non-essentiality in the case of the *Wtap* knockout. For the *Mettl3* knockout, residual m6A may indicate partial redundancy with the Mettl14 subunit, or alternatively, that the allele we analysed is hypomorphic due to the presence of additional AUG sites (see Supplementary Fig. 3a). Despite repeated efforts we were unable to generate ES cells entirely lacking a functional Rbm15/m6A-MTase complex (knockout of *Rbm15* together with its homologue *Rbm15b*, deletion of the entire *Mettl3* locus, and knockout of the *KIAA1429/virilizer* gene, encoding another core subunit of the complex), suggesting that residual activity of the complex may be essential for ES cell viability. This is likely a consequence of the key role m6A plays in several steps of mRNA biogenesis, and is in keeping with a previous report ²¹.

We went on to derive and validate *Rbm15* and *Wtap* null iXist-ChrX cell lines (Supplementary Fig. 3a, b, Supplementary Fig. 2b), and performed allelic ChrRNA-seq 1 day after Xist RNA induction. In contrast to iXist-Chr3 cells we observed a minor but reproducible reduction in silencing across the chromosome, both for *Rbm15* and *Wtap* null cell lines (Fig. 3c, d, Supplementary Fig. 3c). We did not observe differential effects based on classification of ChrX genes by position, expression level, pre-existing chromatin state, or silencing efficiency.

Given the different effects on Xist-mediated silencing reported above and in prior studies ^{13,21}, together with the fact that it was not possible to assess complete loss of function of the

Rbm15/m6A-MTase complex, we went on to determine the effect of deleting specific m6A sites in the 5' region of Xist RNA in the iXist-ChrX cell line. Thus, we generated and validated two cell lines Xist Δ m6A/3A and Xist Δ m6A/11G, with overlapping short deletions of this region (Supplementary Fig. 2b, c, Supplementary Fig. 3b). We then assayed Xist-mediated silencing 1 day after induction using allelic ChrRNA-seq. The results are summarised in Fig. 3e-g. The most proximal deletion (Xist Δ m6A/3A), resulted in a small deficit in Xist-mediated silencing across the length of the chromosome, consistent with the effects seen in *Rbm15* and *Wtap* null iXist-ChrX cells. In contrast, the distal deletion (Xist Δ m6A/11G) had no detectable effect on chromosome-wide silencing. The latter observation may indicate a hierarchical relationship of different m6A target sites, or alternatively that the proximal deletion is closer to the A-repeat and disrupts normal function of this element. Taken together our results indicate that the Rbm15/m6A-MTase complex plays only a minor role in Xist-mediated chromosome silencing (see also discussion).

To investigate the contribution of LBR to Xist-mediated silencing we generated a null allele of the *Lbr* gene by deleting the entire locus using CRISPR/Cas9-mediated mutagenesis, initially in iXist-Chr3 mESCs (Supplementary Fig. 4a). The deletion had no effect on Xist transgene expression (Supplementary Fig. 2b, Supplementary Fig. 4b). Strikingly, allelic ChrRNA-seq 1 day after Xist RNA induction revealed no detectable effects on Xist-mediated silencing (Fig. 4a, b). To test if LBR may be important at a later stage, we repeated the ChrRNA-seq experiment 3 days after Xist RNA induction in differentiating conditions, but again saw no effect on chromosome-wide silencing (Supplementary Fig. 4c, d).

We went on to investigate LBR function in Xist-mediated silencing in the context of ChrX. Thus, we generated and validated *Lbr* null iXist-ChrX cell lines (Supplementary Fig. 4a, b). Assaying Xist-mediated silencing by ChrRNA-seq revealed a small reduction in silencing

across the X chromosome (Fig. 4c, d). No differential effects were observed in relation to gene location, expression level, or silencing efficiency, although there was a minor preference towards silencing of genes with a pre-existing H3K27me3 chromatin environment (see below).

LBR has been proposed to directly bind to Xist RNA via the LBS element located between the A- and B-repeats²². We therefore generated and validated an LBS deletion in iXist-ChrX cells (Supplementary Fig. 2c), and performed allelic ChrRNA-seq after 1 day of Xist induction (Fig. 4d, e). Xist-mediated silencing was more strongly affected than in *Lbr* null iXist-ChrX cells, although it was nevertheless a relatively weak effect. The correlation of gene silencing in Xist Δ LBS and *Lbr* null iXist-ChrX is significant ($r=0.72$), but not exact (Supplementary Fig. 4e). The latter observations indicate that recruitment of LBR may not be the sole function of the LBS region, and LBS, encompassing the entire F-repeat, may also recruit another factor(s). No differential effects were observed for gene silencing in Xist Δ LBS in relation to gene location, expression level, pre-existing chromatin environment or silencing efficiency.

The role of PRC1 and PRC2 in establishment of Xist-mediated silencing

The Polycomb complexes PRC1 and PRC2 which catalyse the histone modifications H2AK119ub and H3K27me3 respectively, have been linked to Xist-mediated silencing in several independent studies. Recently we demonstrated that a specific PRC1 complex, Pcgf3/5-PRC1, recruited by hnRNPK bound to the Xist B-repeat region, initiates recruitment of both PRC1 and PRC2 complexes^{10,28}. We therefore set out to make *Pcgf3/5* double-knockout (DKO) in iXist-Chr3 cells. In prior studies it was necessary to use a conditional knockout strategy to generate *Pcgf3/5* DKO cells (Almeida et al, 2017). We therefore developed a *Pcgf3* allele tagged with an Auxin inducible degron³⁶, on a *Pcgf5* null

background (Fig. 5a, Supplementary Fig. 5a-d). The GFP-degron-tagged Pcgf3 protein was degraded rapidly, within 2 hours of addition of auxin (Supplementary Fig. 5c). Acute Pcgf3/5 loss of function led to a significant reduction in global levels of H2AK119ub (Supplementary Fig. 5e). Using calibrated allelic ChIP-seq analysis (see STAR methods), we found that there was no gain of H2AK119ub or H3K27me3 on the Cast Chr3 after induction of Xist RNA for 1 day (Supplementary Fig. 5f), consistent with our previous analysis demonstrating that Pcgf3/5–PRC1 activity is upstream of PRC2/H3K27me3 in Xist-mediated silencing. We went on to perform allelic ChrRNA-seq 1 day after Xist RNA induction and analysed the effect on Xist-mediated silencing. Consistent with our prior analysis, silencing was significantly reduced across the length of the transgene-bearing chromosome 3 (Fig. 5b, c).

Because the *Pcgf3/5* DKO abolishes Xist-dependent H3K27me3 and H2AK119ub, it is not possible to determine the relative contribution of the two histone modifications to gene silencing. To address this we made use of a previously described iXist-Chr3 cell line in which the gene encoding Suz12, a core subunit of PRC2 required for catalytic activity, is deleted, and as a consequence, H3K27me3 is completely abolished (²⁹ and see Supplementary Fig. 5e, f). Allelic ChrRNA-seq, performed 1 day after Xist induction, revealed a very small reduction in Xist-mediated silencing (Fig. 5d, e). Given prior evidence suggesting that PRC2 is important for X inactivation *in vivo* ^{25,37}, we also analysed *Suz12* null cells 3 days after Xist RNA induction (under differentiating conditions). As illustrated in Fig. 5f, g, Supplementary Fig. 5g, *Suz12* knockout does compromise silencing after 3 days, although not to the extent seen in *Pcgf3/5* loss of function cells (compare Fig. 5c and 5g). Further analysis demonstrated that genes affected by PRC2 loss after 3 days of Xist RNA induction (63/538) are generally expressed at a relatively high level in mESCs (Supplementary Fig. 5h-j). Consistently, we found that active chromatin regions (determined by ChromHMM analysis), show a greater gain of PRC1-mediated H2AK119ub in WT compared to *Suz12* null mESCs (Supplementary Fig. 5k), suggesting that PRC2-mediated

H3K27me3 plays a role in enhancing Xist-dependent gain of H2AK119ub to silence highly expressed genes. Together these experiments suggest that PRC1/H2AK119ub is the critical determinant of Polycomb function in the context of Xist-mediated silencing after 1 day of Xist RNA induction, but that PRC2/H3K27me3 plays a supporting role in maintaining/reinforcing gene repression at later timepoints.

Next we sought to assess the contribution of Polycomb in the context of Xist-mediated silencing on the X chromosome. Thus, we generated and validated a deletion encompassing the B/C-repeat in iXist-ChrX ES cells ([Supplementary Fig. 2c](#), [Supplementary Fig. 6a](#)). This deletion, which extends into the downstream C-repeat, is equivalent to the Xist transgene deletion Δ PID that we previously showed abolishes both PRC1 and PRC2 recruitment by Xist RNA¹⁰. Consistently we observed complete loss of Xist-dependent H2AK119ub and H3K27me3 after 1 day of Xist RNA induction in Xist Δ PID mESCs, as determined by calibrated allelic ChIP-seq ([Supplementary Fig. 6b](#)). Allelic ChrRNA-seq performed 1 day after Xist RNA induction demonstrated abrogation of Xist-mediated silencing across the chromosome ([Fig. 6a, b](#)). Interestingly, silencing was abrogated to a greater degree after 6 days of Xist RNA induction in differentiating conditions ([Fig. 6c-e](#)).

We went on to investigate if absence of Polycomb affects Xist-mediated silencing of all genes to an equivalent extent. Genes with pre-existing H3K27me3 compared with genes with a pre-existing active chromatin signature (classified by ChromHMM analysis), showed a greater reliance on Polycomb, both in Xist-dependent silencing in iXist-Chr3 and iXist-ChrX cells ([Fig. 6f](#)). To further investigate this observation we examined the acquisition of Polycomb-mediated histone modifications, H2AK119ub and H3K27me3 at the different groups based on chromatin signature. We observed that gain of Polycomb histone modifications was significantly higher at genes and regions classified as having pre-existing

H3K27me3 chromatin in mESCs (Fig. 6g-j, Supplementary Fig. 6c-f). This, however, was not evident in the ChrX Xist Δ PID line. We conclude that enhanced accumulation of Polycomb modifications at pre-existing sites may contribute to preferential silencing of nearby genes, but that other variables also play a role in the observed differential effects.

Integrated analysis of Xist-mediated silencing

A comparison of the contribution made by different silencing pathways in iXist-Chr3 and iXist-ChrX lines after 1 day of Xist RNA induction is shown in Supplementary Fig. 7a,b. Based on this analysis we conclude that in both the X chromosome and autosomal models, Spen/A-repeat is the principal silencing pathway, with Polycomb (predominantly PRC1)/PID also playing an important role. Rbm15/m6A-MTase complex and Lbr play a relatively minor role, detected only in the X chromosome model. These latter conclusions are supported by analysis of Xist RNA elements linked to binding/activity of the factors.

Our conclusions are further illustrated by principal component analysis (PCA) of individual biological replicates for iXist-Chr3 and iXist-ChrX after 1 day of Xist RNA induction (Fig. 7a, b). Thus, all uninduced controls, which are tightly clustered together, are clearly separated from induced datasets, in accord with the occurrence of Xist-mediated silencing. Notably, Xist-induced datasets for *Spn* knockout and the A-repeat deletion are close to or overlapping uninduced controls, indicating a dramatic silencing deficiency. Knockouts of genes encoding subunits of the Rbm15/m6A-MTase complex, LBR and associated *Xist* deletions (after Xist RNA induction), on the other hand, cluster closer to WT induced, with somewhat greater separation evident for iXist-ChrX compared to iXist-Chr3 cells. *Pcgf3/5* and Δ PID datasets for iXist-Chr3 and iXist-ChrX respectively, are further separated with a distinct direction relative to all the other induced WT and mutant datasets. This observation implies a unique role for the Polycomb system in the Xist-mediated silencing cascade.

To further depict relative silencing deficiencies we performed hierarchical clustering of allelic silencing (induced - uninduced), for datasets in iXist-ChrX cells (Fig. 7c). The results emphasise the dramatic and equivalent abrogation of silencing in Spen null and A-repeat deletion lines. Of the other pathways, Polycomb, predominately PRC1-linked H2AK119ub (inferred from deletion of Xist PID region), is the most important factor, with Rbm15/m6A-MTase and LBR playing a relatively minor role (Fig. 7c, d). LBR/LBS and Rbm15/m6A-MTase broadly impact silencing of all genes to an equivalent degree, but Polycomb has a more pronounced effect towards specific subsets of genes, notably those that are normally silenced most efficiently and associated with pre-existing H3K27me3 chromatin (Fig. 7c, d). This tendency was in fact also observed for the *Lbr* knockout cell lines, but to a much weaker degree (Fig. 7d). In most cases the different silencing pathways do not differentially affect silencing in relation to 2D/3D proximity to *Xist* (Fig. 7e, Supplementary Fig. 7c). However, A-repeat and/or Spen-independent silencing occurs preferentially in close proximity to *Xist* (Fig. 7e, Supplementary Fig. 7c). This may indicate a defect in Xist RNA spreading/propagation, although it needs to be considered that there are only a small number of genes in the group showing partial silencing.

Discussion

In this study we provide insights into the interplay and relative importance of different pathways required for the establishment of chromosome silencing by Xist RNA. Through the use of two mESC model systems we were able to gauge the contribution of the key silencing pathways in the context of Xist RNA action on the X chromosome and on an autosome. By using a sensitive approach to assay silencing chromosome-wide, we were able to determine how individual factors act on different subsets of genes.

Our results show that recruitment of the RBP Spen by the Xist A-repeat is the principal pathway for establishment of Xist-mediated silencing, affecting most genes to an equivalent degree. There are however some genes which undergo a degree of silencing in the absence of Spen/A-repeat. Consistent with this finding, a recent study, which utilised allelic RNA-seq, also reported residual Xist-mediated silencing in response to constitutively driven expression of A-repeat-deleted Xist RNA in mouse extraembryonic tissues³⁸. The degree of silencing observed in this study was higher than we find in the ES cell models, possibly reflecting differences in the level of Xist RNA and/or developmental stage/cell-type. In common with our findings, weakly expressed genes were more resistant to loss of the A-repeat.

The fact that there is only very limited silencing following Spen/A-repeat loss of function could suggest that other pathways, Polycomb, and to a lesser degree, Rbm15/m6A-MTase and LBR, function downstream of Spen. However, a potentially confounding factor is that Spen/A-repeat deletion reduces Xist RNA levels. The basis for this effect is not clear but may relate to previous findings indicating that the spread of Xist RNA from early nucleation sites is abrogated when the A-repeat is deleted³². Thus, the ability of alternative silencing pathways to compensate for Spen/A-repeat deletion may be compromised by incomplete Xist RNA spreading. A further caveat is that we only assayed silencing in Spen/A-repeat null

mESCs up to 6 days after Xist RNA induction in differentiating mESCs, and it is possible that gene silencing increases at later timepoints. In support of this possibility we found that the rate of silencing in WT iXist-ChrX cells varies considerably for different genes, and even after 6 days of Xist RNA induction, gene silencing is in many cases incomplete.

A surprising conclusion from our study is that the Rbm15/m6A-MTase complex and LBR, previously proposed to be key factors for Xist-mediated silencing^{21,22}, play only a minor role. In the case of the Rbm15/m6A-MTase complex we observed a small reduction in silencing efficiency specifically in the iXist-ChrX model, a finding that contrasts with prior studies^{13,21}. Whilst redundancy of Rbm15/m6A-MTase complex subunits and essentiality of the m6A system must be taken into account, this conclusion is supported by our analysis of deletions around the major m6A target region in Xist RNA. Possible reasons for the discrepancy in our findings compared to prior studies include the use of different silencing assays (selected loci versus chromosome-wide), and acute (RNAi-mediated knockdown) versus chronic (mESC lines with CRISPR mediated deletions) loss of function.

Similar to the Rbm15/m6A-MTase complex, deletion of *Lbr* had only a minor effect on Xist-mediated silencing, again specifically in the iXist-ChrX model. Our findings provide an explanation for why *Lbr* knockout mouse models show no apparent female-specific phenotypes^{23,24}. Deletion of the proposed LBR binding region, LBS, in Xist RNA led to a more prominent silencing deficiency. We conclude from this that LBS is important for other functions in addition to LBR binding. Discrepancies with prior studies^{15,22} may again be linked to use of different silencing and loss of function assays, as well as allelic vs non-allelic analysis.

The reason for the difference in silencing effects seen for Rbm15/m6A-MTase and LBR loss of function in iXist-ChrX compared to iXist-Chr3 systems is not known. One possibility is that higher levels of Xist RNA in iXist-Chr3 cells, due to the presence of a multicopy Xist transgene, mask the relatively small effect of Rbm15/m6A-MTase and LBR deficiency. Alternatively, the Rbm15/m6A-MTase and LBR pathways may only be relevant in an X chromosome context, for example contributing to the higher silencing efficiency that has been observed to occur on the X chromosome relative to autosomes ⁹.

Our findings substantiate the principal role of the PID region and PRC1 in hierarchical recruitment of Polycomb complexes by Xist RNA, and the proposal that Xist-dependent recruitment of PRC2 is attributable solely to prior H2AK119ub deposition by PRC1 ^{10,28}. Our analysis further demonstrates that PRC1-mediated H2AK119ub is the principal determinant of Polycomb silencing mediated by Xist RNA. H3K27me₃, catalysed by PRC2 contributed to a much lesser extent, although this effect did increase at a later timepoint. The basis for the progressive defect in silencing is not clear but may reflect changes in levels of H3K27me₃ reader proteins in ESC differentiation reported in previous studies ^{39,40}.

In this work we have analysed pathways that function in establishment of silencing at the onset of Xist expression. However, previous studies using XX mESCs have shown that there are some factors, for example histone macroH2A ⁴¹, and the chromosomal protein Smchd1 ⁴², that are recruited to Xi in an Xist-dependent manner, but only after prolonged Xist expression/cell differentiation. Notably, Smchd1 is important for maintaining silencing of a significant proportion of Xi genes ⁴³⁻⁴⁵. Thus, in future work it will be important to define the relationship between early- and late- acting pathways in establishment versus maintenance of Xist-mediated silencing.

Acknowledgements

We thank Amanda Williams for assistance with NGS sequencing and Andrew Bassett for CRISPR/Cas9 sgRNA design for Wtap and Rbm15 mutagenesis; Neil Blackledge for providing the Tir targeting construct; Martin Houliard for providing TIGRE gRNA and HR construct; Michal Gdula for preliminary analysis of LBR datasets. This work was funded by grants to NB from the Wellcome Trust (103768) and the European Research Council (340081). C.-X. Song is supported by the Ludwig Institute for Cancer Research, Cancer Research UK (C63763/A26394), NIHR Oxford Biomedical Research Centre, and Conrad N. Hilton Foundation. Y. Bi is supported by China Scholarship Council.

Author Contributions

TB, GW, HC, and GP contributed to conceiving the study, experimental work, data analysis and manuscript preparation. JSB, TZ, MA, BB, BM, EJC, IAR, QP, and YB contributed to experimental work. CS contributed to analysis of m6A mass spectrometry data. NB contributed to conceiving the study and writing the manuscript.

Competing Interests

The authors declare no competing interests.

References

- 1 Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372-373 (1961).
- 2 Brockdorff, N. *et al.* 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 (1992).
- 3 Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137, doi:10.1038/379131a0 (1996).
- 4 Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-542 (1992).
- 5 Lee, J. T. & Jaenisch, R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386**, 275-279, doi:10.1038/386275a0 (1997).
- 6 Gendrel, A. V. & Heard, E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu Rev Cell Dev Biol* **30**, 561-580, doi:10.1146/annurev-cellbio-101512-122415 (2014).
- 7 Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell* **5**, 695-705 (2000).
- 8 Tang, Y. A. *et al.* Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation. *Epigenetics Chromatin* **3**, 10, doi:10.1186/1756-8935-3-10 (2010).
- 9 Loda, A. *et al.* Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. *Nature communications* **8**, 690, doi:10.1038/s41467-017-00528-1 (2017).
- 10 Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Molecular cell* **68**, 955-969 e910, doi:10.1016/j.molcel.2017.11.013 (2017).
- 11 Chow, J. & Heard, E. X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol* **21**, 359-366, doi:10.1016/j.ceb.2009.04.012 (2009).

- 12 Monfort, A. *et al.* Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell Rep* **12**, 554-561, doi:10.1016/j.celrep.2015.06.067 (2015).
- 13 Moindrot, B. *et al.* A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Rep* **12**, 562-572, doi:10.1016/j.celrep.2015.06.053 (2015).
- 14 Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416, doi:10.1016/j.cell.2015.03.025 (2015).
- 15 McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236, doi:10.1038/nature14443 (2015).
- 16 Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267-1279, doi:10.1016/j.cell.2016.04.028 (2016).
- 17 Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature genetics* **30**, 167-174, doi:10.1038/ng820 (2002).
- 18 Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201-206, doi:10.1038/nature11112 (2012).
- 19 Ke, S. *et al.* A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes & development* **29**, 2037-2053, doi:10.1101/gad.269415.115 (2015).
- 20 Linder, B. *et al.* Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature methods* **12**, 767-772, doi:10.1038/nmeth.3453 (2015).
- 21 Patil, D. P. *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369-373, doi:10.1038/nature19342 (2016).
- 22 Chen, C. K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**, 468-472, doi:10.1126/science.aae0047 (2016).
- 23 Cohen, T. V. *et al.* The lamin B receptor under transcriptional control of C/EBPepsilon is required for morphological but not functional maturation of neutrophils. *Hum Mol Genet* **17**, 2921-2933, doi:10.1093/hmg/ddn191 (2008).

- 24 Shultz, L. D. *et al.* Mutations at the mouse ichthyosis locus are within the lamin B receptor gene: a single gene model for human Pelger-Huet anomaly. *Hum Mol Genet* **12**, 61-69 (2003).
- 25 Silva, J. *et al.* Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Developmental cell* **4**, 481-495 (2003).
- 26 de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental cell* **7**, 663-676, doi:10.1016/j.devcel.2004.10.005 (2004).
- 27 Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-135, doi:10.1126/science.1084274 (2003).
- 28 Almeida, M. *et al.* PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081-1084, doi:10.1126/science.aal2512 (2017).
- 29 Cooper, S. *et al.* Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nature communications* **7**, 13661, doi:10.1038/ncomms13661 (2016).
- 30 Zeng, H. *et al.* An inducible and reversible mouse genetic rescue system. *PLoS genetics* **4**, e1000069, doi:10.1371/journal.pgen.1000069 (2008).
- 31 Gdula, M. R. *et al.* The non-canonical SMC protein SmcHD1 antagonises TAD formation on the inactive X chromosome. *bioRxiv*, doi:10.1101/342147 (2018).
- 32 Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973, doi:10.1126/science.1237973 (2013).
- 33 Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature methods* **13**, 919-922, doi:10.1038/nmeth.3999 (2016).
- 34 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
- 35 Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq

- based on immunocapturing and massively parallel sequencing. *Nat Protoc* **8**, 176-189, doi:10.1038/nprot.2012.148 (2013).
- 36 Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods* **6**, 917-922, doi:10.1038/nmeth.1401 (2009).
- 37 Wang, J. *et al.* Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nature genetics* **28**, 371-375, doi:10.1038/ng574 (2001).
- 38 Sakata, Y. *et al.* Defects in dosage compensation impact global gene regulation in the mouse trophoblast. *Development* **144**, 2784-2797, doi:10.1242/dev.149138 (2017).
- 39 Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**, 47-62, doi:10.1016/j.stem.2011.12.006 (2012).
- 40 O'Loughlen, A. *et al.* MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **10**, 33-46, doi:10.1016/j.stem.2011.12.004 (2012).
- 41 Mermoud, J. E., Costanzi, C., Pehrson, J. R. & Brockdorff, N. Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J Cell Biol* **147**, 1399-1408 (1999).
- 42 Gendrel, A. V. *et al.* Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Developmental cell* **23**, 265-279, doi:10.1016/j.devcel.2012.06.011 (2012).
- 43 Mould, A. W. *et al.* Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* **6**, 19, doi:10.1186/1756-8935-6-19 (2013).
- 44 Blewitt, M. E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature genetics* **40**, 663-669, doi:10.1038/ng.142 (2008).

- 45 Gendrel, A. V. *et al.* Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol Cell Biol* **33**, 3150-3165, doi:10.1128/MCB.00145-13 (2013).

Figure Legends

Fig. 1 | mESC models for analysis of Xist-mediated silencing. **a**, Schematic summarising engineering of iXist-Chr3 and iXist-ChrX mESCs. **b**, Allelic silencing across Chr3 in uninduced (No Dox), 1 day and 3 day induced (Dox) iXist-Chr3 cells. Mean value of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on Chr3 ideogram indicates location of Xist transgene. **c**, Violin plot summarising distribution of ratio values in **(b)**. p-values were calculated by one-sided Wilcoxon rank sum test. **d**, Allelic silencing on ChrX in uninduced (NoDox), 1 day and 6 days induced (Dox) iXist-ChrX cells. Mean value of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on ChrX ideogram indicates location of *Xist* locus. **e**, Violin plot summarising data in **(d)**. p-values were calculated by one-sided Wilcoxon rank sum test. **f**, Comparison of gene silencing in iXist-Chr3 and iXist-ChrX cell lines. **g, h**, Importance of parameters for predicting silencing efficiency in iXist-Chr3 (G) and iXist-ChrX (H) cells using machine learning. AUC (area under the curve) is 0.70 and 0.71 for iXist-Chr3 and iXist-ChrX, respectively. **i**, Allelic silencing for quantiles E1-E3 based on mESC gene expression level in iXist-Chr3 and iXist-ChrX mESCs 1 day after Xist RNA induction, where E1 is lower expressed gene group and E3 is the highest. **j, k**, Scatterplots and boxplots illustrating more efficient silencing of genes that have pre-existing mESC H3K27me3 (pre-H3K27me3) chromatin/low expression level in iXist-Chr3 cells (J) and iXist-ChrX cells (K). p-values calculated using two-sided Wilcoxon rank sum test.

Fig. 2 | Effect on Xist-mediated silencing after deletion of Spen/A-repeat. **a**, Allelic silencing across Chr3 after 1 day of Xist RNA induction in WT and *Spn* null (KO) iXist-Chr3 cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on Chr3 ideogram indicates location of Xist transgene. **b**, Violin plot summarising data in **(a)**, p-values were calculated using one-sided Wilcoxon rank sum test. **c**, Distribution of genes by silencing degree in relation to mESC expression levels (3 equal sized quantiles, E1<E2<E3) in WT and *Spn* null (KO) iXist-Chr3 cells (left panel) and WT, *Xist* Δ A and *Spn* null iXist-ChrX cells (right panel). **d**, Allelic silencing across ChrX after 1 day of Xist RNA induction in WT and *Spn* null (KO) iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on ChrX ideogram indicates location of the *Xist* locus. **e**, Violin plot summarising data in **(d)** and **(f)**. p-values were calculated using one-sided Wilcoxon rank sum test. **f**, Allelic silencing across ChrX after 1 day of Xist RNA induction in WT and *Xist* Δ A iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on ChrX ideogram indicates location of the *Xist* locus.

Fig. 3 | The role of Rbm15/m6A-MTase complex in Xist-mediated silencing. **a**, Allelic silencing across Chr3 after 1 day of Xist RNA induction in WT and *Wtap* null (KO) iXist-Chr3 cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on Chr3 ideogram indicates location of the Xist transgene. **b**, Violin plot summarising data for *Wtap*

null in **(a)**, and additionally, data for *Mettl3* mutant and *Rbm15* null shown in **Supplementary Fig. 3c**. p-values were calculated using one-sided Wilcoxon rank sum test. **c**, Allelic silencing across ChrX after 1 day of Xist RNA induction in WT and *Rbm15* null (KO) iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on ChrX ideogram indicates location of the *Xist* locus. **d**, Violin plot summarising data in **(c)**. p-values were calculated using one-sided Wilcoxon rank sum test. **e, f**, Allelic silencing across ChrX after 1 day of Xist RNA induction in WT and cell lines with deletions within the major m6A peak region, *Xist*Δm6A/3A (E) and *Xist*Δm6A/11G (F) in iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. **g**, Violin plot summarising data in **(e)**, **(f)**. p-values were calculated using one-sided Wilcoxon rank sum test.

Fig. 4 | The role of LBR in Xist-mediated silencing. **a**, Allelic silencing across Chr3 after 1 day of Xist RNA induction in WT and *Lbr* null (KO) iXist-Chr3 cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on Chr3 ideogram indicates location of the *Xist* transgene. **b**, Violin plot summarising data in **(a)**. p-values were calculated using one-sided Wilcoxon rank sum test. **c**, Allelic silencing across ChrX after 1 day of Xist RNA induction in WT and *Lbr* null (KO) iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on ChrX ideogram indicates location of the *Xist* locus. **d**, Violin plot summarising data in **(c)** and **(e)**. p-values were calculated using one-sided Wilcoxon rank sum test. **e**, Allelic silencing across ChrX after 1 day of Xist RNA induction in WT and *Xist*ΔLBS iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods.

Fig. 5 | Divergent roles for PRC1 and PRC2 in Xist-mediated silencing. **a**, Immunofluorescence illustrating Auxin-induced degradation of Pcgf3-GFP-AID (green) and resulting loss of Xi-H2AK119ub domains (red) as seen in iXist-Chr3 *Pcgf5* null mESCs following Xist RNA induction for 1 day. DNA is counterstained with DAPI (blue). Scale bar is 10μm. **b**, Allelic silencing across Chr3 after 1 day of Xist RNA induction in WT and *Pcgf3/5* DKO iXist-Chr3 cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on Chr3 ideogram indicates location of the *Xist* transgene. **c**, Violin plot summarising data in **(b)**. p-values were calculated using one-sided Wilcoxon rank sum test. **d**, Allelic silencing across Chr3 after 1 day of Xist RNA induction in WT and *Suz12* null (KO) iXist-Chr3 cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. **e**, Violin plot summarising data in **(d)**. p-values were calculated using one-sided Wilcoxon rank sum test. **f**, Allelic silencing across Chr3 after 3 days Xist RNA induction in WT and *Suz12* null (KO) iXist-Chr3 cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. **g**, Violin plot summarising data in **(f)**. p-values were calculated using one-sided Wilcoxon rank sum test.

Fig. 6 | The role of Polycomb complexes in Xist-mediated silencing. **a**, Allelic silencing across ChrX after 1 day of Xist RNA induction in WT and Xist Δ PID iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. Green line on ChrX ideogram indicates location of the *Xist* locus. **b**, Violin plot summarising data in **(a)**. p-values were calculated using one-sided Wilcoxon rank sum test. **c**, Allelic silencing across ChrX after 6 days of Xist RNA induction in WT and Xist Δ PID iXist-ChrX cells. Mean value of difference of allelic ratio for each gene with an informative SNP was calculated from biological replicates as detailed in STAR methods. **d**, Violin plot summarising data in **(c)**. p-values were calculated using one-sided Wilcoxon rank sum test. **e**, Comparison of chromosome silencing in WT and Xist Δ PID after 1 day and 6 days of Xist RNA induction (left panel). Boxplot representation of data (right panel). **f**, Comparison of silencing for genes with pre-existing active (pre-active) or H3K27me3 (pre-H3K27me3) chromatin signatures over promoters for iXist-Chr3 in Pcgf3/5 DKO (top), and for iXist-ChrX in Xist Δ PID (bottom) after 1 day of Xist RNA induction. p-values were calculated using two-sided Wilcoxon rank sum test. **g, h**, Metaprofile showing gain of H3K27me3 (**g**) or H2AK119ub (**h**) over genes with pre-existing active or H3K27me3 chromatin in WT, Pcgf3/5 DKO or Suz12 KO iXist-Chr3 ES cells 1 day after Xist RNA induction. **i, j**, Metaprofile showing gain of H3K27me3 (**i**) or H2AK119ub (**j**) over genes with pre-existing active and H3K27me3 chromatin following 1 day of Xist RNA induction in WT and Xist Δ PID mESCs.

Fig. 7 | Integrated analysis of key pathways in Xist-mediated silencing. **a, b**, PCA analysis of biological replicates of ChrRNA-seq analysis for all uninduced and Xist induced experiments in iXist-Chr3 (A) and iXist-ChrX (B) for WT and mutants. **c**, Cluster analysis and heatmap for silencing efficiency of ChrX genes in selected mutant mESCs, as indicated after 1 day of Xist RNA induction. Gene order is ranked by silencing efficiency in WT iXist-ChrX cells. Chromatin signature defined by ChromHMM is shown below. **d**, Silencing levels in different mutants in iXist-ChrX cells illustrated as one-sided Wilcoxon rank sum significance. Bar charts illustrate separate significance scores for genes classified according to pre-existing active (pre-active) and H3K27me3 (pre-H3K27me3) chromatin signatures, corrected for sample size. Error bars indicate significance range for 180 permutations. **e**, Correlation analysis for topological distance and silencing efficiency for WT and selected mutant mESCs in iXist-ChrX model. 95% of the population is encircled. Spearman correlation and p-values are indicated below.

Figure 1

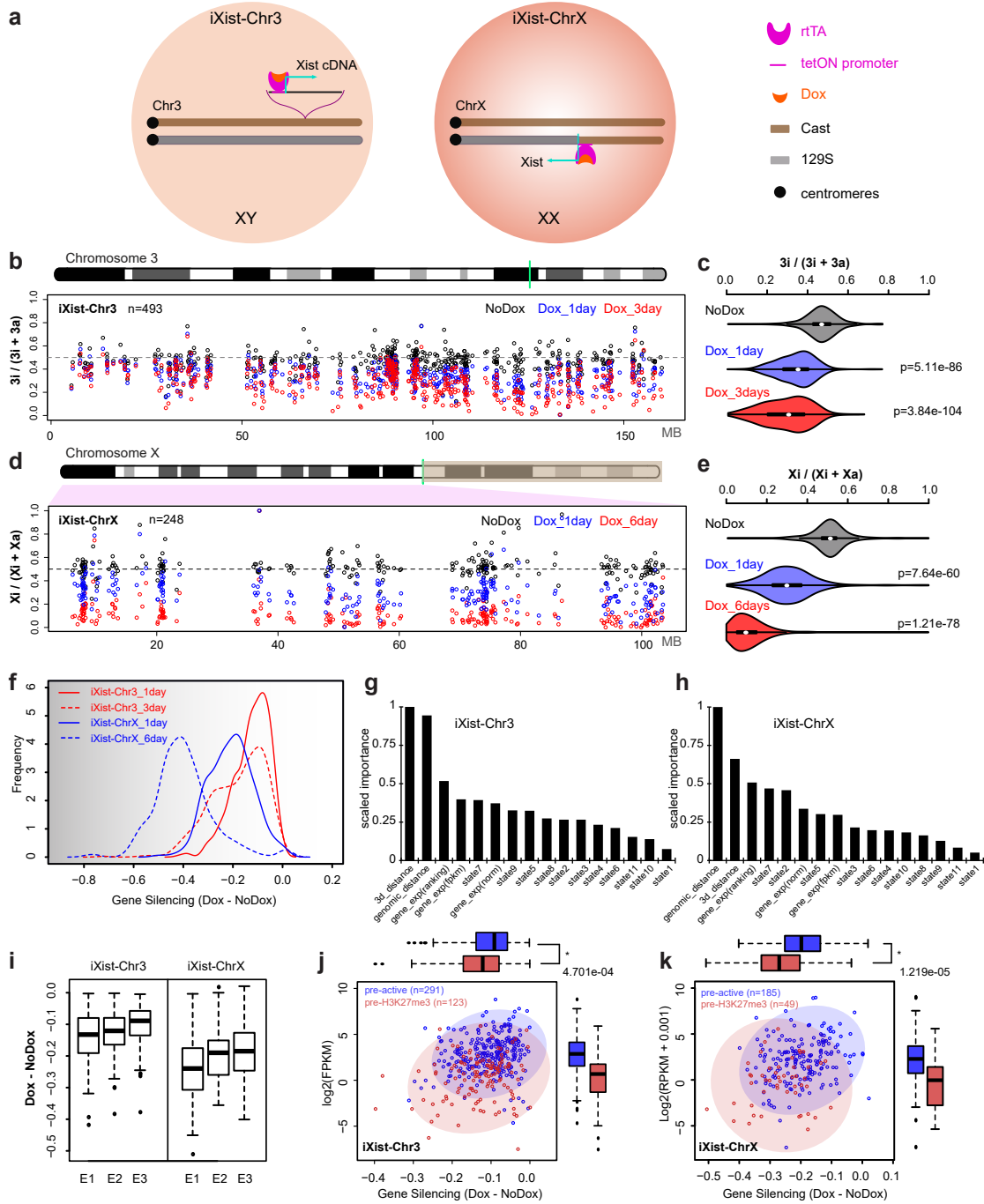


Figure 2

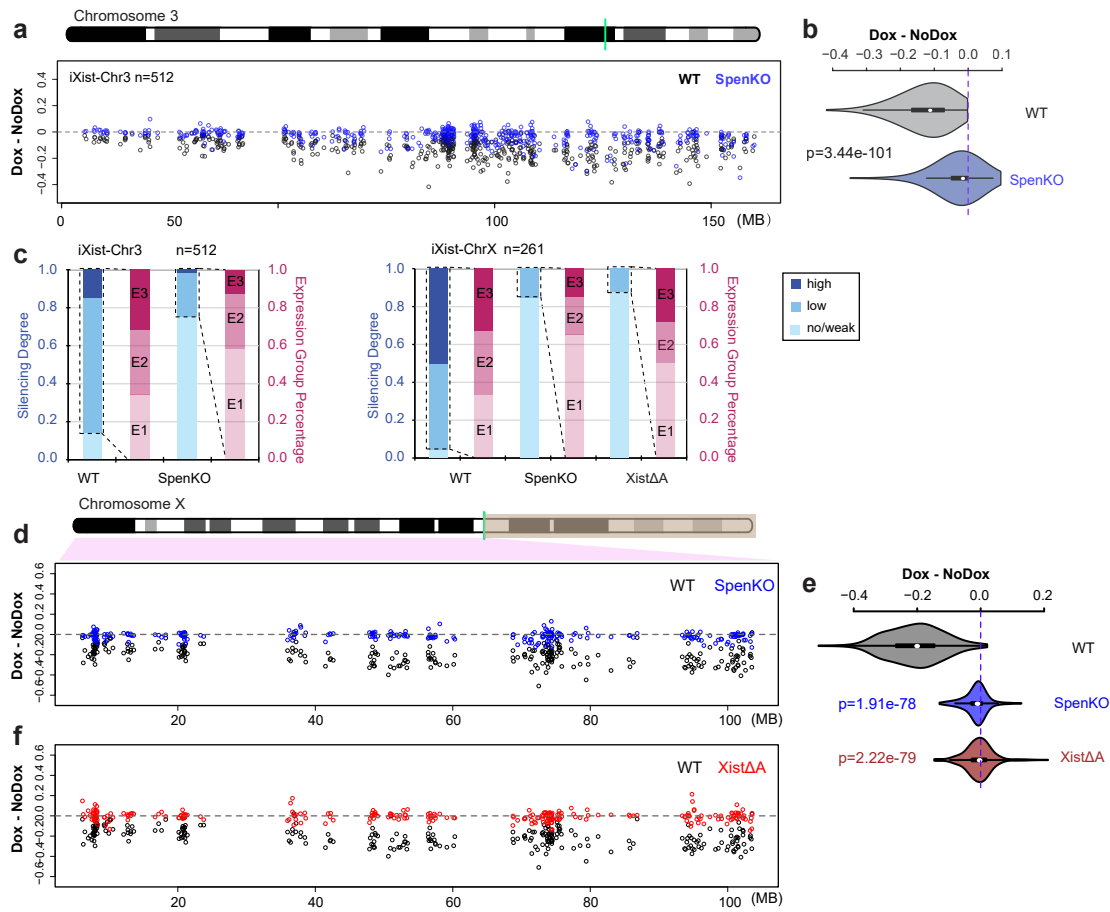


Figure 3

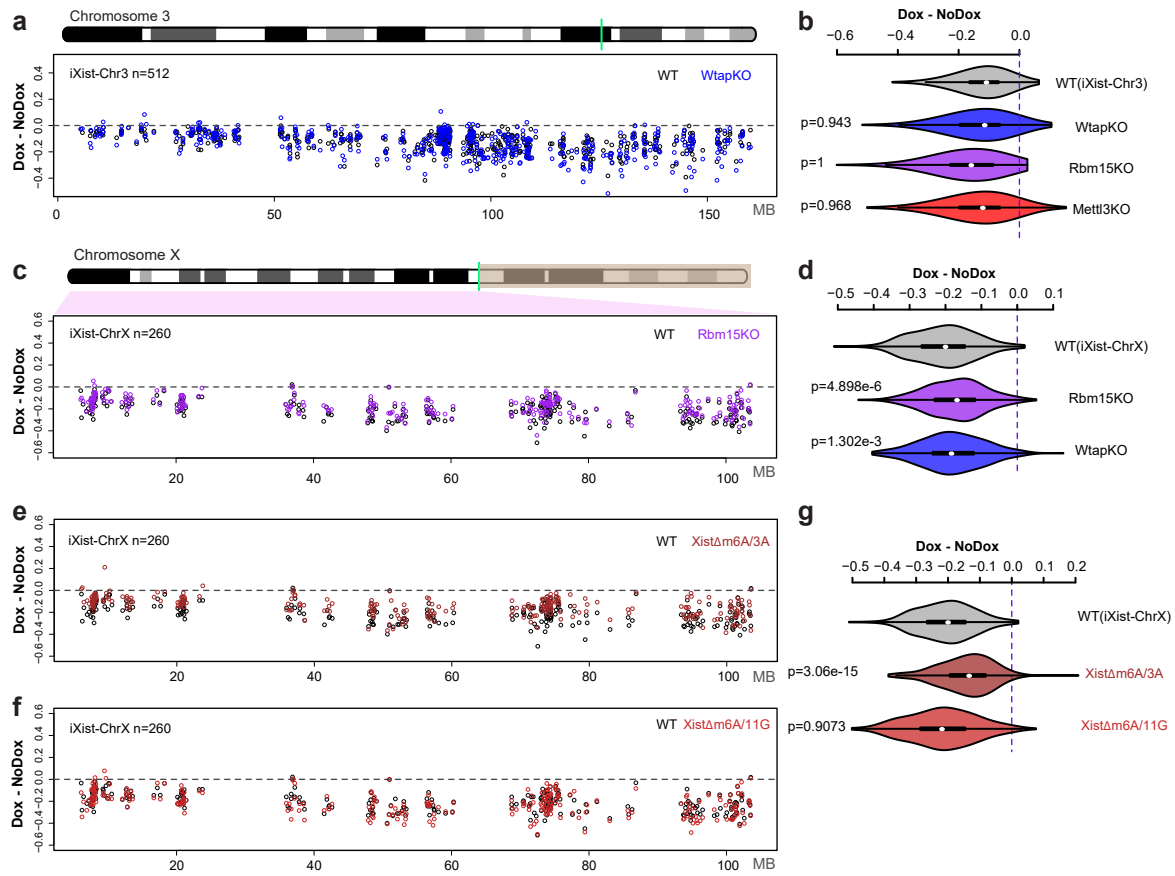


Figure 4

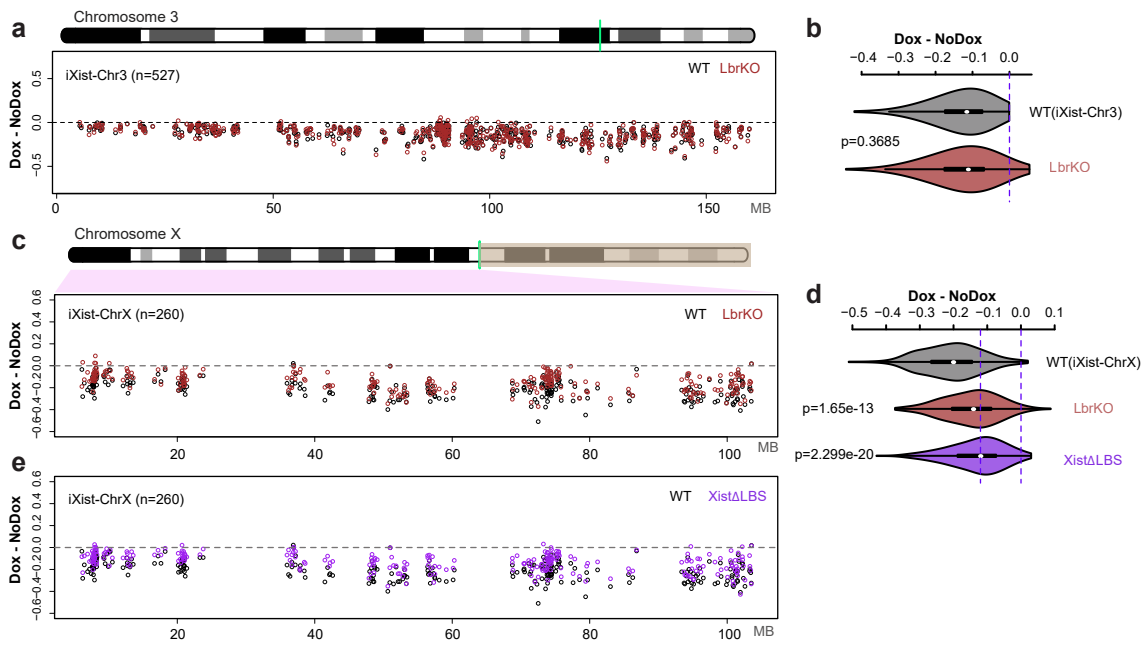


Figure 5

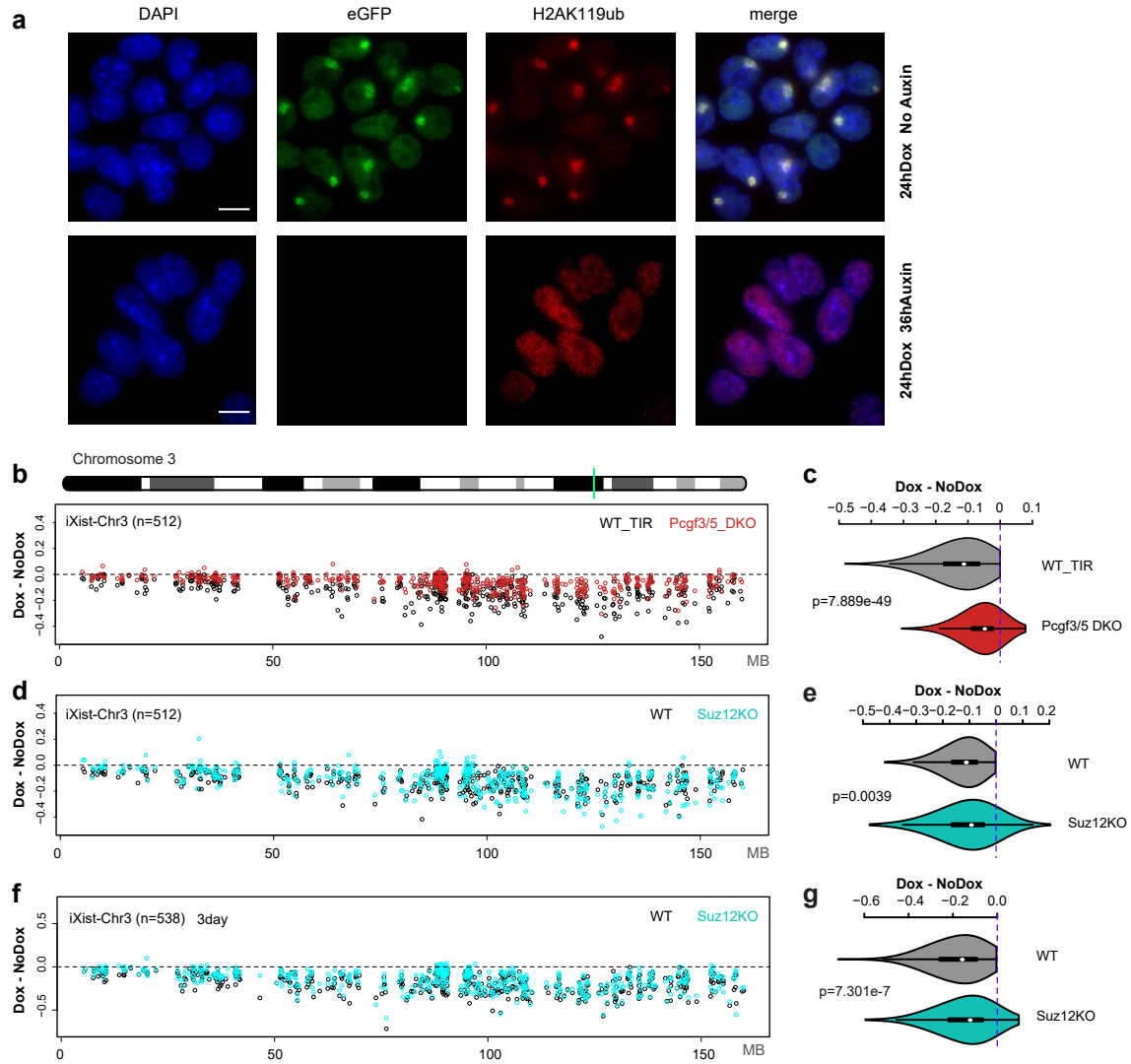


Figure 6

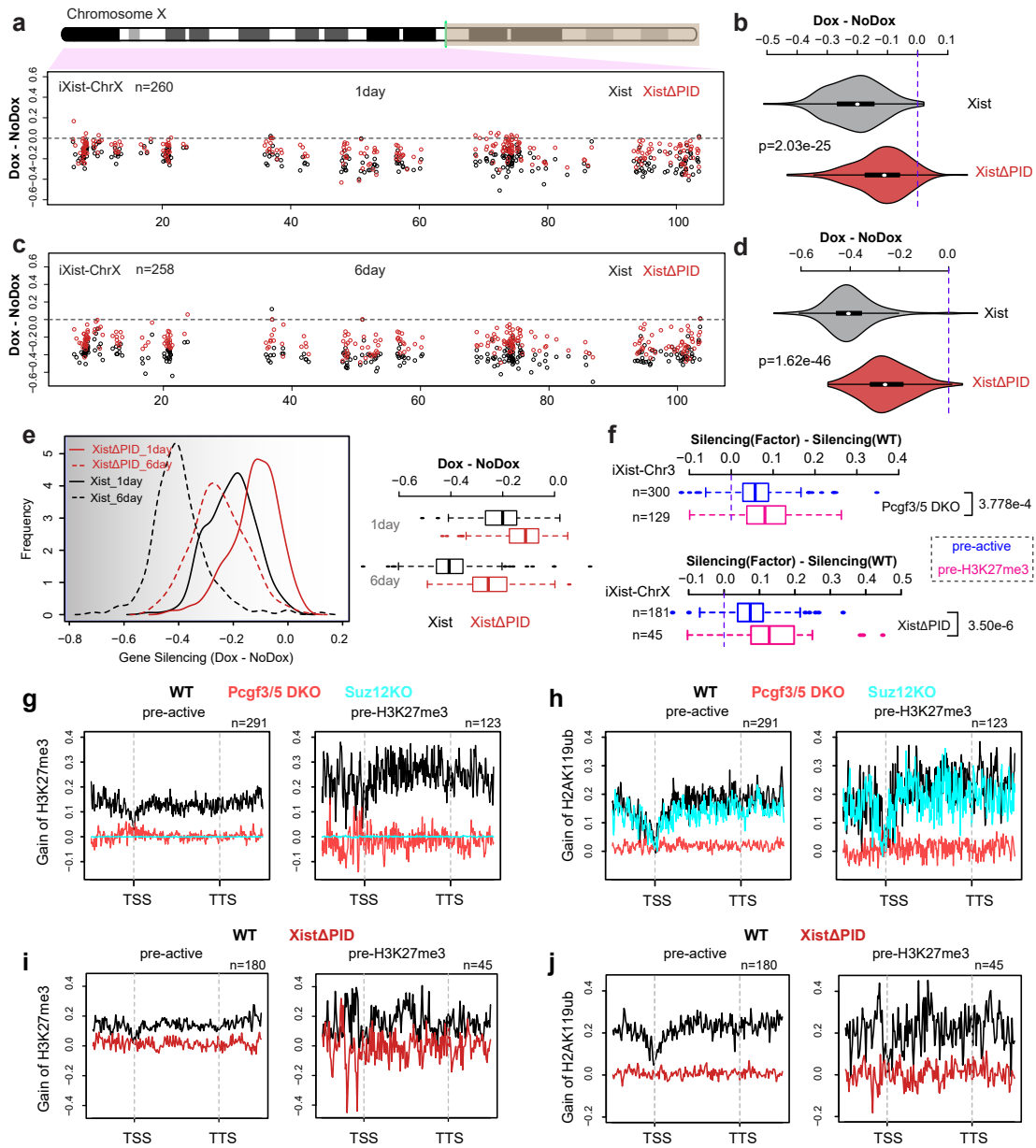
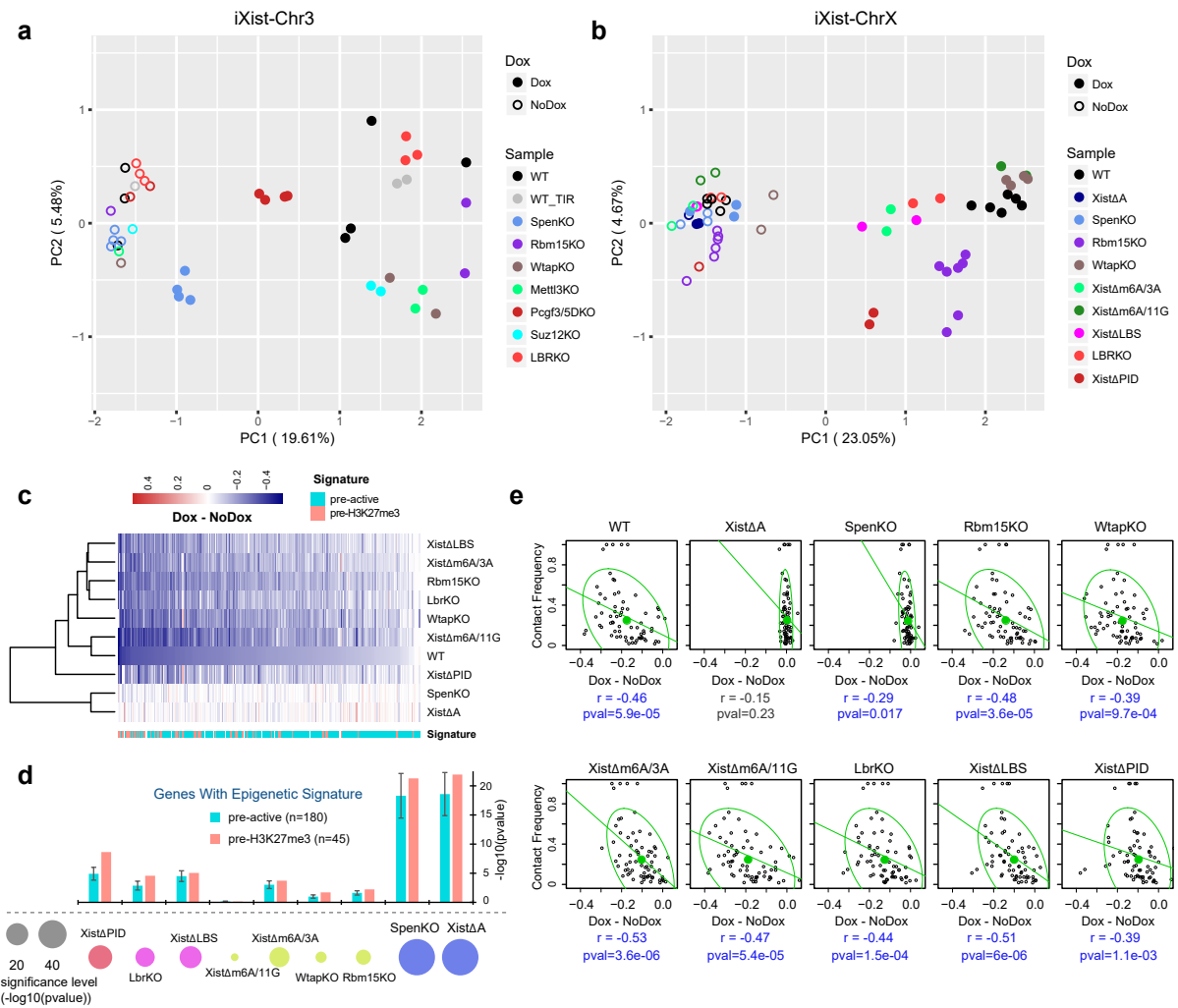


Figure 7



METHODS

Experimental model

ES cells were grown in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies) supplemented with 10% foetal calf serum (Seralab), 2 mM L-glutamine, 0.1 mM non-essential amino acids, 50 μ M β -mercaptoethanol, 100 U/ml penicillin / 100 μ g/ml streptomycin (all from Life Technologies) and 1000 U/ml LIF (made in-house). Mitomycin C-inactivated mouse fibroblasts were used as feeders. All ES cells were grown in feeder-dependent conditions on gelatinised plates at 37°C in a humid atmosphere with 5% CO₂. Xist expression was driven by a TetOn promoter induced by addition of 1.0 to 1.5 μ g/mL of doxycycline (Sigma, D9891) for 1 to 6 days depending on the experiment. ES cell differentiation was achieved by LIF withdrawal from the medium and low density cell plating.

Generation of mESC lines

To generate iXist-ChrX ESC lines, F1 2-1 XX mESC line (129/Sv-Cast/Ei, a gift from J.Gribnau) was consecutively modified using CRISPR-mediated genome engineering to replace one of the endogenous Xist promoters with tetracycline-inducible promoter (tetOP), and to introduce reverse tetracycline-controlled transactivator (rtTA) into the Tigre locus⁴⁶. The construct for Xist promoter modification (pBS_Xist_loxN_Pgkneo-tetOP) contained 2.6kb (5') and 2.8kb (3') homology regions with the insertion of tetOP 41bp upstream of the Xist TSS. The construct included a floxed neomycin resistance gene driven by the mouse Pgk promoter for positive selection of the targeted clones. The construct for targeting the Tigre locus (pBS_TIGRE_CAG_rtTA) contained 1kb homology regions, surrounding CAG promoter-driven rtTA. Both targeting events were facilitated by co-transfection of the targeting plasmids together with specific Xist (Xist_TRE_gRNA1 and gRNA2) or Tigre (TIGRE_gRNA) sgRNAs. Cells were transfected using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. 1 μ g of each sgRNA was lipofected together with 1 μ g of the targeting vector. After 12 hrs transfected cells were passaged to 90 mm gelatinised Petri dishes with feeders, and puromycin (1.5-2.5 μ g/mL) was applied 24 hrs later. Cells were grown under puromycin selection for 2 days and then without puromycin for a further 8-10 days, until colonies were picked. For targeting of the Xist promoter, neomycin selection (Geneticin® (G-418), ThermoFisher) was additionally applied 48 hrs after lipofection, and was maintained until colonies were picked. Selected clones were screened and validated by Sanger sequencing (SS) of amplified genomic fragments (gPCR), by Southern blot hybridization (SB), by Xist RNA FISH (RNA-FISH) and immunofluorescence for histone modifications (IF) after Xist doxycycline induction and karyotyped using 21XMouse MFISH (Zeiss Metasystem), following the manufacturer's instructions. Two clones were selected for further analysis; clone B2 carries tetOP on the M.m.domesticus allele and clone C7 has integration of tetOP on the M.m.castaneus allele.

For the generation of CRISPR/Cas9-mediated KO ES cell lines, cells were transfected with 0.5-1 μ g of each sgRNA, and puromycin selection was applied 24 hrs after transfection for 48 hrs. For CRISPR-assisted homologous recombination, 0.5-1 μ g of sgRNA was lipofected together with 1 μ g of targeting vector. Cells were grown for 10-12 days, and positive clones validated. Full details for all mESC lines used in this study, sgRNA sequences, plasmids, as well as validation methods can be found in Supplementary Table 1. Oligonucleotides used for cloning sgRNAs as well as primers used for PCR and sequencing are listed in Supplementary Table 2.

RNA-FISH

Cells were plated on Superfrost Plus gelatinised slides (VWR) and grown overnight. After Xist induction (1, 3 or 6 days) the slides were washed twice with PBS and cells were permeabilised for 5 min in CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂, 10 mM PIPES) with 0.5% Triton X-100 (Sigma) on ice. Slides were rinsed briefly in PBS and fixed in 4% formaldehyde / PBS for 10 min on ice, followed by two washes in 70% ethanol. Slides were either stored in 70% ethanol at 4°C until use or dehydrated (80%, 95%, 100% ethanol, 3 min each, RT) and air dried immediately before hybridisation with Xist probe. Xist probe was generated from an 18 kb cloned cDNA spanning the whole Xist transcript using a nick translation kit (Abbott Molecular, 7J0001) as previously described⁴⁷. Directly labelled probe (1.5 µl) was co-precipitated with 10 µg salmon sperm DNA, 1/10 volume 3 M sodium acetate (pH 5.2) and 3 volumes of 100% ethanol. After washing in 75% ethanol, the pellet was dried, resuspended in 6 µl deionised formamide (Sigma) and denatured at 75 °C for 7 min before quenching on ice. Probe was diluted in 6 µl 2X hybridisation buffer (5X SSC, 12.5% dextran sulfate, 2.5 mg / mL BSA (NEB)), added to the denatured slides and incubated overnight at 37°C in a humid chamber. After incubation, slides were washed three times with a solution of 2X SSC / 50% formamide followed by three washes with 2X SSC in a water bath at 42°C. Slides were mounted with Vectashield with 4,6-diamidino-2-phenylindole (DAPI) (Vector labs) and sealed with nail varnish. Cells were analysed on an inverted fluorescence Axio Observer Z.1 microscope using a PlanApo 63x/1.4 N.A. oil-immersion objective. Images were acquired using AxioVision software.

Immunofluorescence

Cells were plated on gelatinised slides or coverslips (13 mm diameter cover slips from VWR), at least a day before the experiment. On the day of the experiment, cells on slides were washed with PBS and then fixed with 2% formaldehyde for 15 min followed by 5 min of permeabilisation in 0.4% Triton X-100. Cells were briefly washed with PBS before blocking with a 0.2% PBS-based solution of fish gelatine for three washes of 5 min each. Primary antibody dilutions were prepared in fish gelatine solution with 5% normal goat or normal donkey serum depending on the secondary antibodies used. A complete list of antibodies used in the study is shown in Supplementary Table 3. Cells on slides were incubated with primary antibodies for 2 hours in a humid chamber at room temperature. Slides were washed three times in fish gelatine solution. Secondary antibodies were diluted 1:400 in fish gelatine solution and incubated with cells on slides for 30 min in a humid chamber at 37°C. After incubation, slides were washed twice with fish gelatine and one time with PBS before mounting using Vectashield mounting medium with DAPI. Excess mounting medium was removed and the coverslips were sealed to slides using nail varnish. Cells were analysed on an inverted fluorescence Axio Observer Z.1 microscope using a PlanApo 63x/1.4 N.A. oil-immersion objective. Images were acquired using AxioVision software.

Auxin-inducible degradation

For the degron-mediated degradation of PCGF3-AID-EGFP, indole-3-acetic acid (IAA, chemical analog of auxin, Sigma, I5148) was added to the ES medium to a final concentration of 500 µM for 12 hrs prior to Xist induction for all experiments. IAA was maintained in the media at all times after initial addition. Degradation of PCGF3-AID-EGFP was monitored by western and IF.

Nuclear extract

Cell pellets were washed with PBS and resuspended in 10 packed cell volume (PCV) buffer A (10 mM HEPES-KOH pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, with 0.5 mM DTT, 0.5 mM PMSF, and complete protease inhibitors (Roche) added fresh). After a 10 min incubation at 4°C, cells were collected by centrifugation (1500g, 5 min, 4°C) and resuspended in 3 PCV of buffer A + 0.1% NP-40 (Sigma). After another 10 min incubation at 4°C, nuclei were collected by centrifugation (400g, 5 min, 4°C) and resuspended in 1 PCV buffer C (250 mM NaCl, 5 mM HEPES-KOH (pH 7.9), 26% glycerol, 1.5 mM MgCl₂, 0.2 mM EDTA-NaOH, pH 8.0 with complete protease inhibitors (Roche) + 0.5 mM DTT added fresh). 5 M NaCl was added drop-wise to bring the concentration to 350 mM and the mixture was incubated for 1 hr at 4°C with occasional agitation. After centrifugation (16,100g, 20 min, 4°C), the concentration of the supernatant was quantified using the Bio-Rad Bradford assay and stored at -80°C until use.

Western blot

Protein from nuclear extract was used for all experiments. Samples were separated on a polyacrylamide gel and transferred onto PVDF membrane by semi-dry transfer (15 V for 50 min). Membranes were blocked by incubating them for 1 hr at room temperature in 10 mL TBS, 0.1% Tween (TBST) with 5% w/v Marvell milk powder. Blots were incubated overnight at 4°C with the primary antibody (see Supplementary Table 3), washed 4 times for 10 min with TBST and incubated for 40 min with secondary antibody conjugated to horseradish peroxidase. After washing 4 times for 5 min with TBST, bands were visualised using ECL (GE Healthcare).

Chromatin RNA-seq

Chromatin RNA was extracted from one confluent 15 cm dish of mESCs. Briefly, cells were trypsinised and washed in PBS. Cells were lysed on ice in RLB (10 mM Tris pH 7.5, 10 mM KCl, 1.5 mM MgCl₂, and 0.1% NP40), and nuclei were purified by centrifugation through a sucrose cushion (24% sucrose in RLB). The nuclei pellet was resuspended in NUN1 (20 mM Tris pH 7.5, 75 mM NaCl, 0.5 mM EDTA, 50% glycerol), then lysed with NUN2 (20 mM HEPES pH 7.9, 300 mM, 7.5 mM MgCl₂, 0.2 mM EDTA, 1 M Urea). Samples were incubated for 15 minutes on ice then centrifuged at 2800 g to isolate the insoluble chromatin fraction. The chromatin pellet was resuspended in TRIzol by passing multiple times through a 23 gauge needle. Finally chromatin-associated RNA was purified through standard TRIzol/chloroform extraction followed by isopropanol precipitation. Samples were then treated with Turbo DNase, and 500ng – 1µg of RNA was used for library preparation using the Illumina TruSeq stranded total RNA kit (RS-122-2301). Libraries were quantified by qPCR with KAPA Library Quantification DNA standards (Kapa biosystems, KK4903). Biological replicates for each experiment are listed in Supplementary Table 4. The libraries were pooled and 2X81 paired end sequencing was performed using Illumina NextSeq500 (FC-404-2002).

Native ChIP-seq

For calibrated native ChIP (H3K27me3 and H2AK119ub), 40 million ES cells and 10 million Drosophila SG4 Cells were carefully counted using a LUNA-II Automated Cell Counter (Logos Biosystems), pooled and lysed in RSB (10 mM Tris pH 8, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40) for 5 min on ice with gentle inversion. Nuclei were resuspended in 1 mL of RSB + 0.25 M sucrose + 3 mM CaCl₂, treated with 200 U of MNase (Fermentas) for 5 minutes at 37°C, quenched with 4 µL of 1M EDTA, then centrifuged at 5000 rpm for 5 minutes. The supernatant was transferred to a fresh tube as fraction S1. The chromatin pellet was incubated for 1 h in 300 µL of nucleosome release buffer (10 mM Tris pH 7.5, 10 mM NaCl, 0.2 mM EDTA), carefully passed 5 times through a 27G needle then centrifuged at 5000 rpm for 5 minutes. The supernatant from this S2 fraction was combined with S1 as soluble chromatin extract. For each ChIP reaction, 100 µL of chromatin was diluted in Native ChIP incubation buffer (10 mM Tris pH 7.5, 70 mM NaCl, 2 mM MgCl₂, 2 mM EDTA, 0.1% Triton) to 1 mL and incubated with antibody (see Supplementary Table 3) overnight at 4°C. For H2AK119ub ChIP, all buffers were supplemented with 10 mM of the deubiquitinase inhibitor N-Ethylmaleimide (Sigma, E3876-5G). Samples were incubated for 1 hr with 40 µl protein A agarose beads pre-blocked in Native ChIP incubation buffer with 1 mg/ml BSA and 1 mg/ml yeast tRNA, then washed for a total of 4 times in Native ChIP wash buffer (20 mM Tris pH 7.5, 2 mM EDTA, 125 mM NaCl, 0.1% Triton) and once in TE. The DNA was eluted with 1% SDS and 100 mM NaHCO₃, and was purified using ChIP DNA Clean and Concentrator kit (Zymo Research).

Approximately 20-50 ng of ChIPed DNA was used for library prep using the NEBNext Ultra II DNA Library Prep Kit with NEBNext Single indices (E7645), and then further quantified by qPCR with KAPA Library Quantification DNA standards and SensiMix SYBR (Bioline, UK). The libraries were pooled and 2X 81 paired end sequencing was performed on using Illumina NextSeq500.

m6A-seq

m6A-seq was based on the method by Dominissini and colleagues⁴⁸. Briefly, total RNA was extracted from pre-plated ES cells under reducing conditions using the RNeasy kit (Qiagen) and on-column DNase treatment as per the manufacturers' instructions. RNA was fragmented by incubation for 5 min at 94°C in thin-walled PCR tubes with fragmentation buffer (100 mM Tris-HCl, 100 mM ZnCl₂). Fragmentation was quenched using stop buffer (200 mM EDTA, pH 8.0) and incubation on ice, before ensuring the correct size (50-100 bp) using 1.5% agarose gel electrophoresis. 375 µg of total RNA was incubated with 12.5 µg anti-m6A antibody (Synaptic Systems, 202 003), RNasin (Promega), 2 mM VRC, 50 mM Tris, 750 mM NaCl and 5% Igepal CA-630 in DNA / RNA low-bind tubes for 2 hrs before m6A containing RNA was isolated using 200 ul Protein A agarose beads (Repligen) per IP (pre-blocked with BSA). 6.7 mM m6A (Sigma) was used to elute RNA from the beads. Input and eluate samples were EtOH precipitated, quantitated and pooled as libraries generated using TruSeq Stranded total RNA LT sample prep according to manufacturer's instructions, but excluding the fragmentation step. 50 bp single reads were obtained using Illumina NextSeq500 sequencer (Illumina, FC-404-2005).

m6A HPLC-MS/MS quantification

mRNA was prepared from pre-plated ES cells using two sequential rounds of purification using the Dynabeads mRNA Direct kit (ThermoFisher, 61011) as per the manufacturer's

protocol. RNA samples were hydrolysed with DNA Degradase Plus™ (Zymo Research) supplied with 40 nM deaminase inhibitor erythro-9-Amino- β -hexyl- α -methyl-9H-purine-9-ethanol hydrochloride (Sigma-Aldrich) at 37 °C for 4 hrs. After hydrolysis, the enzymes were removed with Amicon Ultra-0.5 ml 10 K centrifugal filters (Merck Millipore).

The HPLC-MS/MS analysis was carried out with 1290 Infinity LC Systems (Agilent) coupled with a 6495B Triple Quadrupole Mass Spectrometer (Agilent). A ZORBAX Eclipse Plus C18 column (2.1 x 150 mm, 1.8-Micron, Agilent) was used. The column temperature was maintained at 40 °C, and the solvent system was water containing 10 mM ammonium acetate (pH 6.0, solvent A) and methanol (solvent B) with 0.4 ml/min flow rate. The gradient was: 0-5 min; 0 solvent B; 5-8 min; 0-5.63% solvent B; 8-9 min; 5.63% solvent B; 9-16 min; 5.63-13.66% solvent B; 16-17 min; 13.66-100% solvent B; 17-21 min; 100% solvent B; 21-24.3 min; 100-0% solvent B; 24.3-25 min; 0% solvent B. The dynamic multiple reaction monitoring mode (dMRM) of the MS was used for quantification. The source-dependent parameters were as follows: gas temperature 230 °C, gas flow 14 L/min, nebulizer 40 psi, sheath gas temperature 400 °C, sheath gas flow 11 L/min, capillary voltage 1500 V in the positive ion mode, nozzle voltage 0 V, high pressure RF 110 V and low pressure RF 80 V, both in the positive ion mode. The fragmentor voltage was 380 V for all compounds, while other compound-dependent parameters were summarized in the table.

Compound-dependent LC-MS/MS parameters used for nucleosides quantification.

RT: retention time, CE: collision energy, CAE: cell accelerator voltage. All the nucleosides were analysed in the positive mode.

Compound	Precursor Ion (m/z)	Product Ion (m/z)	RT (min)	Delta RT(min)	CE (V)	CAE (V)
rA+Na	290	158	14.9	2	8	4
rG+Na	306	174	9.1	2	10	4
rU+Na	267	135	3.8	2	10	4
rC+Na	266	134	9.05	2	10	4
m6A+Na	304	172	18.4	2	16	4

QUANTIFICATION AND STATISTICAL ANALYSIS

iXist-Chr3 integration site identification

To determine the integration locus of the inducible Xist transgene in Chromosome 3, over 500 million paired-end reads from all ChIP-seq experiments were combined and mapped to the mm10 mouse genome. All read pairs which could not be mapped were then used for further analysis. Paired reads where one mapped to Chr3 and the other to the plasmid

containing the Xist transgene were identified and quantified. The region in Chr3 that contained the highest concentration of these chimeric reads was Chr3:125.8097-125.8145MB. The integration site was also confirmed by using combined reads from previously generated ATAC-seq data ⁴⁹.

Chromatin state segregation by ChromHMM

Using mESC ENCODE ChIP-seq data for various histone modifications and transcription factors, we defined 12 chromatin states (strong enhancer, promoter, gene body, *etc*) by ChromHMM (v1.11) ⁵⁰ following the instructions given at (https://github.com/guifengwei/ChromHMM_mESC_mm10) (state1: CTCF-binding; state2: Intergenic_Region; state3: Heterochromatin; state4: Enhancer; state5: RepressedChromatin; state6: BivalentChromatin; state7: ActivePromoter; state8: StrongEnhancer; state9: TranscriptionTransition; state10: TranscriptionElongation/GeneBody; state11: Weak/poised_Enhancer; state12: LowSignal/RepetitiveElements). ChromHMM was used to determine the chromatin state of each gene based on 2kb flanking either side of the TSS. When multiple transcripts were annotated, we chose the TSS used for the highest number of unique transcripts, or one at random if multiple TSSs were equally frequent. When multiple chromatin states were present in the TSS \pm 2kb region, the predominant state was assigned to the gene.

RNA-seq data analysis

The following analysis strategy was used for Chromatin RNA-seq and mRNA-seq. The RNA-seq data mapping pipeline was similar to our previous study ⁴⁹. Briefly, the raw fastq files of read pairs were first mapped to an rRNA build by bowtie2 (v2.3.2) ⁵¹ and rRNA-mapped reads discarded. The remaining unmapped reads were aligned to the 'N-masked' genome (from mm10 coordinates) with STAR (v2.4.2a) using parameters "--outFilterMultimapNmax 1 --outFilterMismatchNmax 4 --alignEndsType EndToEnd" for all the sequencing libraries ⁵². Unique alignments were retained for further analysis. We made use of 23,005,850 SNPs between Cast and 129S genomes and employed SNPsplit (v0.2.0, Babraham Institute, Cambridge, UK) to split the alignment into distinct alleles (Cast and 129S) using the parameter "--paired". The (allelic) read numbers were counted by the program featureCounts (-t transcript -g gene_id -s 2) ⁵³ and the alignments were sorted by Samtools ⁵⁴. Bigwig files were generated by Bedtools ⁵⁵ and visualized by IGV ⁵⁶ or UCSC Genome Browser. For biallelic analysis, counts were normalized to 1 million mapped read pairs (as CPM) by the edgeR R package. Genes with at least 10 SNP-covering reads across all the samples were further taken to calculate the allelic ratio of $X_i/(X_i + X_a)$ where X_i and X_a indicate inactive and active allele, respectively.

In the iXist-Chr3 XY cell line the inducible Xist transgene expresses from the Cast allele, whereas in the iXist-ChrX XX line the endogenous Xist from the 129S1 allele is induced (see Figure 1A). For some purposes we also used data from ChrRNA-seq performed in a reciprocal cell line iXist-ChrX XX C7H8 in which endogenous Xist from the Cast allele is induced. Silencing is primarily quantified in all cell lines i.e. wild-type Xist, Xist mutants, and Xist-interacting factor KOs by the difference in allelic ratios between uninduced and induced samples, such that

$$Gene\ Silencing\ (z) = \left[\frac{X_i}{X_i + X_a} \right]_{Dox} - \left[\frac{X_i}{X_i + X_a} \right]_{NoDox}$$

For further comparisons, we set thresholds of $z > -0.05$, $-0.05 < z < -0.2$, $z < -0.2$ to represent genes demonstrating weak/no, low, and high silencing respectively. Additionally, original levels of gene expression for each cell line were calculated as FPKM by Cuffnorm (v2.2.1)⁵⁷, and used to further categorize genes into 3 graded expression groups ($E1 < E2 < E3$) with the same number of genes per group.

Data generated in Suz12 KO lines and corresponding wild-type iXist-Chr3 after 3 days in differentiation media was processed with the same pipeline as above. The p-values were calculated by Student's T-test based on the 3 biological replicates with 0.05 as the significance cutoff, and additional Repression Score (RS) analysis was performed as our previous study⁴⁹.

PCA analysis was performed by R function `prcomp` and plotted with the R package `ggplot2`.

Native ChIP-seq calibration and data analysis

For ChIP-seq experiments quantitatively calibrated with *Drosophila* SG4 cells, raw fastq read pairs were mapped to the 'N-masked' mm10 genome concatenated with the dm6 genome sequence by STAR (v2.4.2a)⁵² using the same parameters as for RNA-seq data with the additional parameter "--alignIntronMax 10". Allelic split of mm10-mapped reads was conducted by SNPsplit as for RNA-seq. We then normalized the mm10-mapped reads to spiked-in *Drosophila* library size, and scaled down the WT sample to 10 million reads and the other calibrated samples by the same factor, for processing by Bedtools⁵⁵ into bedGraph files for normalized signal. All the parameters and normalization intermediates can be found in Supplementary Table 5. We calculated absolute gain of H3K27me3 and H2AK119ub upon Xist induction by the formula $(Xi - Xa)_{Dox} - (Xi - Xa)_{NoDox}$ for further analysis and visualization as bigWig files. Custom scripts (`ExtractInfoFrombedGraph_AtBed.py`) were used to extract values from sorted bedGraph files for different regions (e.g. 250 kb windows or ChromHMM annotations), with signal from regions of different sizes comparable by FPKM. Metagene profiles for pre-active and pre-H3K27me3 gene sets were generated by DANPOS (Chen et al., 2013) with "--bin_size 50" and otherwise default parameters.

m6A-seq data analysis

For conventional m6A-seq data, first we removed the rRNA reads computationally by mapping the single-end reads to the mouse rRNA build with Bowtie2⁵¹. The remaining unmapped reads were then aligned to mm10 genome by STAR (v2.4.2a)⁵² with "--single-end" mode. BigWig files were generated by Bedtools⁵⁵ and normalized to 10 million mapped reads (`Generate_BigWig_from_RNA_seq_Bam_mm10.sh`), then visualized in IGV or UCSC.

HiChIP data analysis and virtual 4C

To integrate the data of higher order chromosome structure in mESCs, we downloaded two biological replicates of cohesin HiChIP-seq data in mESCs⁵⁸ and processed data according to the HiC-Pro pipeline⁵⁹. The Xist loci in iXist-Chr3 and iXist-ChrX were used as viewpoints for virtual 4C visualization at 1MB resolution across Chr3 or ChrX. For comparison we also extracted calibrated ChIP-seq signal of H3K27me3 and H2AK119ub gain averaged at 1MB resolution across Chr3 or ChrX. The correlations between Xist virtual 4C and ChIP-seq or

RNA-seq were calculated using these 1MB windows across the chromosome excluding the regions containing Xist locus as well as its flanking 10MB.

Machine Learning

The random forest classifier in the h2o R package (<https://github.com/h2oai/h2o-3>) was implemented to train models (key parameters: ntrees=50, keep_cross_validation_predictions=T, nfolds=5) for predicting the silencing efficiency (“High” versus “Low” silenced on the basis of above or below the median of silencing degree; *i.e.* 1 vs 0 in the training, validation and testing matrices) in both iXist-Chr3 and iXist-ChrX C7H8 line (Xist was expressed in Cast allele under the engineered TetOn promoter). Three types of features based on different characteristics of genes were used as inputs. (1) Gene expression. Original gene expressed level (FPKM), which was calculated from Cuffnorm program (v2.2.1)⁵⁷. Each gene was assigned a value from 0-1 using two different strategies. In the first method, the highest expressing gene was normalized to one and the rest scaled down accordingly. In the second method, genes were ranked based on expression level, the highest expressing gene was assigned 1 and the lowest assigned 0 with the rest evenly distributed between 0-1. (2) Distance from the Xist locus for each gene. Genomic distance (2D) from the Xist locus was calculated as $1 - \left| \frac{Xist_locus - gene_locus}{Xist_locus} \right|$. Additionally, topological distance to Xist locus (3D) was calculated from Xist virtual 4C⁵⁸, described as above, then normalized to the highest value to scale down to range 0-1. (3) Epigenetic state. Given that gene expression is tightly coordinated by epigenetic setting, we extracted and calculated the proportion of each chromatin state from 4kb surrounding the TSS regions covering all the isoforms, characterized by ChromHMM⁵⁰ as above. For genes containing multiple TSSs, the proportion of each chromatin state within a composite of 4 kb TSS regions was used and weighted according to frequency of TSS usage. The epigenetic state encompassed 12 features, each of ranging from 0 to 1, representing the fraction of each chromatin state in the TSS regulatory region.

Correlation between gene silencing and virtual 4C data

The topological distance to Xist locus was calculated from the cohesin HiChIP-seq data⁵⁸, and then scaled down to range 0-1 as described above. The degree of silencing was calculated for each 1MB genomic window as the ratio of expressed reads from this window ($X_i / (X_i + X_a)$). The raw counts for each 1MB window was calculated by featureCounts without strand specificity⁵³. For each factor knockout or Xist region deletion, biological replicates were averaged (Supplementary Table 4), and silencing scores ($\left[\frac{X_i}{X_i + X_a} \right]_{Dox}$ – $\left[\frac{X_i}{X_i + X_a} \right]_{NoDox}$) across the chromosome were calculated as above. Only 1MB genomic windows that contain both valid virtual 4C value and at least 20 allelic reads were considered. The correlation was calculated in R by the ‘Spearman’ method.

Silencing preference on pre-active and pre-H3K27me3 marked genes

We examined the silencing deficiency in gene sets with different pre-existing chromatin epigenetic signatures, *i.e.* pre-active and pre-H3K27me3 marks. Based on the aforementioned ChromHMM annotation, 180 and 45 genes are annotated as pre-active and pre-H3K27me3 marked genes respectively. We employed one-sided Wilcoxon rank-sum

tests to examine significance with the hypothesis that factor-knockout or Xist sequence deletion will cause silencing deficiency. To eliminate the discrepancy in gene number between the two sets, we permuted 180 times by random sampling 45 genes from pre-active gene sets, then performed the statistics. The range of significance level was the standard deviation of these 180 statistical analyses (Figure 7).

DATA AND SOFTWARE AVAILABILITY

Scripts used for analysis are available at <https://github.com/gui Fengwei>.

Original unprocessed gel images in this manuscript have been deposited to Mendeley Data and are available following these links

<https://data.mendeley.com/datasets/52pjcx486/draft?a=17cc1f84-f752-48be-a881-d1fcee12d99b>.

High-throughput sequencing data (ChrRNA-seq, m6A-seq, ChIP-seq) generated in this study have been submitted to the Gene Expression Omnibus (GEO) under GSE119607.

REFERENCES

- 46 Zeng, H. *et al.* An inducible and reversible mouse genetic rescue system. *PLoS genetics* **4**, e1000069, doi:10.1371/journal.pgen.1000069 (2008).
- 47 Moindrot, B. *et al.* A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Rep* **12**, 562-572, doi:10.1016/j.celrep.2015.06.053 (2015).
- 48 Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nat Protoc* **8**, 176-189, doi:10.1038/nprot.2012.148 (2013).
- 49 Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Molecular cell* **68**, 955-969 e910, doi:10.1016/j.molcel.2017.11.013 (2017).
- 50 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
- 51 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 52 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 53 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

- 54 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 55 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 56 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
- 57 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).
- 58 Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature methods* **13**, 919-922, doi:10.1038/nmeth.3999 (2016).
- 59 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).