Title

Defining the sequence requirements for Xist function in X inactivation

Author

Chau, Anthony Chun-Yu

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Defining the sequence requirements for Xist function in X inactivation

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Anthony Chun-Yu Chau

2022

ABSTRACT OF THE DISSERTATION

Defining the sequence requirements for Xist function in X inactivation

by

Anthony Chun-Yu Chau

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2022

Professor Kathrin Plath, Chair

Mammalian genomes encode thousands of long non-coding (lnc) RNAs, many with important functions including the regulation of gene expression, yet, how lncRNAs function remains largely unexplored. The lncRNA *Xist* provides a remarkable model to investigate the function of lncRNAs in gene regulation, as it spreads from its site of transcription on the X chromosome over the entire chromosome *in cis* to induce gene silencing, alter chromatin state, and modulate the three-dimensional chromosome architecture in the process of X-chromosome inactivation (XCI). XCI is fundamentally important for female mammalian development but, despite its critical role, the mechanisms by which *Xist* carries out the various tasks associated with XCI still remain largely unclear. Recently, our and other labs proposed that *Xist* fulfills its different roles during XCI, such as gene silencing, chromatin association, spreading, recruitment of repressive chromatin regulators, membrane-less compartment formation, through

different RNA domains, which in turn recruit different proteins. The 17kb long *Xist* RNA consists of a series of conserved repeats, termed A-F, as well as intervening non-repeat regions. Both repeat and non-repeat regions have been demonstrated to bind proteins, so both types of sequences can form functional domains. However, except for the A-repeat, which is now known to mediate silencing by recruitment of the proteins SPEN and RBM15, the function of nearly all other *Xist* sequences is still unknown. We sought to determine the *Xist* domains required for the initiation and maintenance of XCI in female mESCs. In Chapter 2, we identified the Xist F-repeat DNA sequence as a critical regulatory element of Xist expression during initiation of XCI using CRISPR mediated deletional analysis of Xist sequences. We also identify E2F3 as a potential interactor of the Xist-F-repeat DNA sequence using an in vitro pulldown assay coupled with mass spectrometry analysis. In Chapter 3, we identified the Xist E-repeat as a functional domain of Xist RNA that is required for the formation of a membrane-less silencing compartment on the inactive X. The E-repeat enables this condensate formation through recruitment of PTBP1, MATR3, TDP-43, and CELF1 via its multivalent protein binding sites, resulting in higher order assemblies of these proteins. Collectively, our findings have expanded on our understanding of the diverse molecular mechanisms employed by Xist-protein interactions at both the DNA and RNA level, to reveal the means by which *Xist* integrates different functions through its domains. Our data has revealed new paradigms for regulation of gene expression by lncRNAs and uncovered important insights into the molecular regulation of XCI by Xist.

The dissertation of Anthony Chun-Yu Chau is approved.

Douglas Black

Siavash Kurdistani

Feng Guo

Tracy Johnson

Kathrin Plath, Committee Chair

University of California, Los Angeles

2022

DEDICATION

I would like to dedicate this Ph.D. thesis research to my mother Glorianna Yuen, and my father Dr. Kevin Chau, for their love, guidance, and inspiration.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

ACKNOWLEDGEMENTS

My Ph.D. journey has been the most enlightening chapter of my life. I have had the wonderful opportunity to learn not only the intriguing complexities of biological mechanism, but also the unyielding resilience required to decipher the unknown. I have met many brilliant and passionate scientists during my journey in scientific research. Their guidance has shaped me into the scientist I am today, and I will forever be grateful for their advice, encouragement, and support.

Firstly, I would like to thank Dr. Kathrin Plath for being the best Ph.D. thesis advisor any student could ask for. She has taught me the critical thinking necessary to succeed as an independent scientist, as well as the mental fortitude needed to overcome scientific challenges. Every time I have the privilege of observing her critique and engage with scientific data during meetings and conferences, she inspires me to challenge myself to be more critical, thoughtful, and creative. In addition, she is a fantastic mentor that truly cares about the growth of her students. She takes great care in demonstrating how to effectively communicate scientific data both visually and in writing. She has also given me many opportunities to work in collaborative projects that has significantly contributed to my development and maturity in working productively with my colleagues. Kathrin's approach to scientific research and mentorship will forever inform the way I approach my work in the next chapters of my life.

In addition to Kathrin, there are many other scientists in Plath lab that has made a lasting impact on my graduate education. Dr. Justin Langerman was my first mentor during my time in Plath lab while I was still a rotation student. He demonstrated to me

that the willingness to push myself beyond my comfort zone is something I must become accustomed to if I wanted to succeed. Dr. Martina Roos showed me the focus and dedication necessary in the unrelenting pursuit of scientific discovery. Dr. Amy Pandya-Jones played an instrumental role in teaching me the methodology for studying X inactivation that allowed me to eventual become independent in my graduate research. Amy's positive attitude and kindness are also important traits that I have adopted in all aspects of my life. Dr. Amanda Collier, Dr. Weixian Deng, Dr. Sammy Yu, Dr. Elsie Jacobson, and Dr.Shan Sabri all provided valuable scientific training and advice for my graduate research projects. Dr. Yolanda Markaki, Dr. Tom Allison, Dr. William Tu, and Dr. Boseon Kim were all wonderful bay mates that have made staying in lab for long hours a much more enjoyable experience. I would also like to thank Nivedita Damodaren, Anya Afasizheva, Clara Cano, Shawn Tan, Robin Mckee, Dr. Iris Dror, Dr. Tsotne Chitiashvili, Dr. Jarrett Miller, Dr. Anna Sahakyan, and Dr. Konstantinos Chronis, for being wonderful lab mates. Their collective positive energy has really defined my experience in Plath lab. I will always look back at my Plath lab experience with pride and joy.

I would like to thank my thesis committee members Dr. Douglas Black, Dr. Siavash Kurdistani, Dr. Tracy Johnson, and Dr. Feng Guo, for providing guidance on my thesis research and support for my research fellowship and job applications. The UCLA MBIDP staff members have also been incredibly supportive during my time in the program. I am grateful for the UCLA Whitcome fellowship, which partially funded my graduate work.

I would like to thank my previous research advisors Dr. Auinash Kalsotra and Dr. Sayee Anakk for their mentorship and support throughout the years. During my time working as a research technician in their respective laboratories, I was deeply influenced by their contagious enthusiasm for scientific discovery. They cultivated an exciting and nurturing laboratory environment that really solidified my decision to pursue graduate level research afterwards. Everything that I was able to learn regarding the study of molecular biology using mammalian systems from their laboratories put me in a position to succeed in graduate school afterwards. Furthermore, I would like to thank my first two graduate student mentors Dr. Katie Whalen and Dr. Daniel Wichelecki. Their initial support when I first started undergraduate research is a big reason why I even considered a career in scientific research in the first place.

Most importantly, I would like to thank my family for their unconditional support throughout my life. My father, Dr. Kevin Chau, is a brilliant electrical engineer and the first inspiration of my life for my decision to pursue a career in science. He has made significant contributions to MEMS technology development, and I can only hope that my achievements in my scientific career will be as meaningful as his. My mother, Glorianna Yuen, is an exceedingly ambitious entrepreneur. Her determination and fearless spirit are constant reminders for me to bold and deliberate in my actions. I have experienced many successes but also many failures. There were times when I was uncertain whether I had what it takes to achieve my goals in life. My parents have never stopped believing in my potential despite my many failures and flaws. Without their support I would not be where I am today.

The work described in chapter 3 is published: Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Mancia Leon WR, Damianov A, Chronis C, Papp B, Chen CK, McKee R, Wang XJ, Chau A, Sabri S, Leonhardt H, Zheng S, Guttman M, Black DL, Plath K. A protein assembly mediates Xist localization and gene silencing. Nature. 2020 Nov;587(7832):145-151. doi: 10.1038/s41586-020-2703-0. Epub 2020 Sep 9. Erratum in: Nature. 2020 Oct 2;: PMID: 32908311; PMCID: PMC7644664.

VITA

2013	Bachelor of Science in Biochemistry, Minor in Chemistry
	University of Illinois, Urbana-Champaign
	Urbana, IL
2013	Biochemistry Department Senior Thesis Distinction Award
	University of Illinois, Urbana-Champaign
	Urbana, IL
2015-Present	Ph.D. student
	Molecular Biology Interdepartmental Program
	University of California, Los Angeles
	Los Angeles, CA
2018	Whitcome Pre-Doctoral Fellowship in Molecular Biology
	University of California, Los Angeles
	Los Angeles, CA
2019-2021	NIH F31 Ruth L. Kirschstein National Research Service Award
	Individual Predoctoral Fellowship

PUBLICATIONS

1. Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Mancia Leon WR, Damianov A, Chronis C, Papp B, Chen CK, McKee R, Wang XJ, **Chau A**, Sabri S, Leonhardt H, Zheng S, Guttman M, Black DL, Plath K. **A protein assembly mediates Xist localization and gene silencing**. Nature. 2020 Nov;587(7832):145-151. PubMED PMID: 32908311.
2. Bhate A, Parker DJ, Bebee TW, Ahn J, Arif W, Rashan EH, Chorghade S, **Chau A**, Lee JH, Anakk S, Carstens RP, Xiao X, Kalsotra A. **ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation**. Nature Commun. 2015 Nov 4;6:8768. PubMed PMID: 26531099.
3. **Chau A**, Kalsotra A. **Developmental insights into the pathology of and therapeutic strategies for DM1: Back to the basics**. Dev Dyn. 2015 Mar;244(3):377-90. PubMed PMID: 25504326.
4. Wichelecki DJ, Balthazor BM, **Chau AC**, Vetting MW, Fedorov AA, Fedorov EV, Lukk T, Patskovsky YV, Stead MB, Hillerich BS, Seidel RD, Almo SC, Gerlt JA. **Discovery of function in the enolase superfamily: D-mannonate and d-gluconate dehydratases in the D-mannonate dehydratase Subgroup**. Biochemistry. 2014 Apr 29;53(16):2722-2731. PubMed PMID: 24697546
5. Whalen KL, **Chau AC**, Spies MA. **In silico optimization of a fragment-based hit yields biologically active, high-efficiency inhibitors for glutamate racemase**. ChemMedChem. 2013 Oct;8(10):1681-9. PubMed PMID: 23929705

Chapter 1

Introduction

Long non-coding RNAs regulate gene expression through diverse mechanisms

Mammalian cells express thousands of long non-coding RNAs (lncRNAs), some of which have been characterized to play important roles in gene regulation[1-2]. The mechanism that lncRNAs utilize to regulate gene expression varies broadly. The first point of divergence is that lncRNAs can function either in *cis* or in *trans*, depending on whether they function at genomic regions from which they are transcribed[3]. An example of a trans acting lncRNA is Firre, which has been demonstrated to localize to several trans-chromosomal loci to bring these genomic regions together in proximity with the Firre genomic locus[4]. An example of a cis acting lncRNA is Air, where imprinted expression of Air leads to recruitment of H3K9 histone methyl transferase G9a at the promoter of Slc22a3 and gene silencing of Slc22a3 in cis[5]. lncRNAs can then be further categorized as either transcriptional activators or repressors[2]. A widely studied activating cis acting lncRNA is HOTTIP, which activates expression of several HOXA locus genes through recruitment of the WDR5/MLL complex and deposition of H3K4me3 at target genes[6]. A well-known repressive trans acting lncRNA is HOTAIR, which represses transcription of the HOXD locus in trans through recruitment of Polycomb Repressive Complex 2 and deposition of H3K27me3[7]. Lastly, lncRNA mechanisms of action can be differentiated based on whether the RNA molecule is itself functional, or active transcription of the RNA molecule is required to exert gene regulatory functions[3]. Post-transcriptional mechanism of lncRNA function often include recruitment of chromatin modifiers through RNA sequence motifs to target genes, either in cis or in trans[2] (Fig. 1). Although some lncRNAs have been characterized, the function of the vast majority of lncRNAs is still completely unknown. Therefore, further investigation into the mechanisms of lncRNA mediated gene

regulation could significantly advance our understanding of both RNA biology, and how the non-coding genome plays a role in gene expression and cell fate determination.

The lncRNA Xist is the master regulator of X-chromosome inactivation

Perhaps the most deeply studied lncRNA is the X-inactive specific transcript (Xist)[8]. Xist orchestrates the process of X-chromosome inactivation (XCI) in cells of female placental mammals during early embryonic development. Since female cells have two X chromosomes compared to the one X chromosome found in male cells, one of the two X chromosomes is transcriptionally silenced during embryonic development to achieve dosage compensation of X-linked genes between male and female cells[9-12]. Once expressed, Xist will coat the entire X-chromosome in cis to initiate silencing of genes across the chromosome through an entirely epigenetic mechanism. This involves Xist recruitment of chromatin modifiers that induce the removal of active histone marks and accumulation of repressive histone marks, to facilitate chromatin compaction where RNA polymerase II is excluded from this heterochromatin domain[13]. Xist expression is both necessary and sufficient to induce XCI[14-15]. Deletion of Xist in cultured cells and mice embryos results in a complete failure to initiate XCI upon female embryonic stem cell differentiation and embryonic development respectively. Ectopic expression of Xist inserted at an autosome is sufficient to drive heterochromatin formation and gene silencing on that chromosome in cis[16].Therefore, Xist provides a unique model for understanding multiple aspects of lncRNA biology, including how RNA protein interaction dynamics direct changes in gene expression, and how lncRNA association with chromatin

can induce epigenetic modifications that result in chromosome wide architectural changes during gene silencing.

Xist utilizes 3D genome architecture to coat the entire X-chromosome

One of the most unique aspects of Xist is its ability to spread across the entire X-chromosome to initiate gene silencing. To understand what factors influence Xist spreading patterns once Xist is expressed, Engreitz et al. employed RNA antisense purification followed by next generation sequencing (RAP-seq) to map Xist contact with chromatin over time in high resolution[17]. Briefly, this method a utilizes biotinylated antisense probes that can hybridize to Xist, which allows for purification of chromatin fragments that are Xist associated after crosslinking. They discovered that Xist spreading patterns do not strongly correlate with any sequence specificity. Rather, Xist first spreads to genomic loci that are close to the Xist locus in 3D space. Therefore, spatial 3D proximity to the Xist locus dictated by chromosome conformation is the most important factor that determines where Xist initially spreads during initiation of XCI[17]. In addition, Xist first spreads to gene dense regions, which is expected since the Xist locus lies within an active chromatin compartment[17-18]. However, these gene dense regions are enriched for genes that are not actively transcribed. While Xist can spread to the periphery of gene-dense regions where genes are actively transcribed, the ability of Xist to induce silencing is required for spreading into these regions[17]. Over the course of XCI, Xist will form a silencing compartment on the inactive X as it continuously spreads, silences genes, and repositions these silenced domains into the growing transcriptionally silent compartment. During the repositioning of silenced domains, Xist is then able to pull in other actively

transcribed genomic regions to induce silencing[17,19]. This “proximity transfer” model of Xist spreading that does not rely on sequence motif based affinity explains how Xist efficiently and reproducibly silences the entire X chromosome despite variations in chromosome conformation between cells.

Xist A-repeat and B-repeat mediate gene silencing during XCI

The Xist gene codes for a 17kb nuclear retained RNA that is alternatively spliced and polyadenylated[20-21]. However, it does not code for any protein product. About 50% of the Xist transcript consists of tandem repeat sequences that is well conserved between species of female placental mammals[22]. These repeats, termed repeats A-F, have been shown to work independently from each other to facilitate different aspects of Xist function in XCI (Fig. 2A)[23]. The most well studied Xist domain is the A-repeat located at the very 5’ end of the Xist transcript. The A-repeat is absolutely required for gene silencing, but deletion of the A-repeat does not significantly affect Xist localization on chromatin[23]. This was the first evidence for Xist domains facilitating specific aspects of Xist function (gene silencing but not chromatin localization in this case) independently of other domains. The protein SPEN (also known as SMRT and HDAC-associated repressor protein (SHARP)) is the key direct interactor of the A-repeat that drives Xist induced gene silencing (Fig. 2B) [24-26]. SPEN directly recruits several protein complexes to direct gene silencing, which include the NcoR and SMRT corepressors as well as the NuRD repressor complex[27]. Recruitment of these complexes is mediated by the SPOC domain of SPEN. NcoR and SMRT interacts with HDAC3 to catalyze histone deacetylation, while NuRD is known to displace RNA polymerase II through nucleosome remodeling[28]. Knockdown of SPEN

during XCI resulted in loss of silencing at 80% of X-linked genes, which demonstrates that SPEN is the primary Xist interactor required for gene silencing[27]. Although The Polycomb Complexes (PRC) 1 and 2 do not directly bind the A-repeat, Zylicz et al. observed that deletion of the A-repeat resulted in loss of PRC marks H3K27me3 and H2AK119ub specifically at active genes[29]. This is consistent with the observation from Engreitz et al. that Xist spreading into actively transcribed gene dense regions requires functional gene silencing[17].

Another well characterized repeat domain of Xist is the B-repeat. The B-repeat recruits PRC1 and PRC2 for deposition of repressive histone marks H2AK119ub and H3K27me3 respectively (Fig. 2C) [30]. PRC1/2 do not directly bind the B-repeat. Instead, hnRNPK directly binds the B-repeat and recruits PRC1 through direct interaction with the PRC1 subunit PCGF3/5[30]. While deletion of the B-repeat resulted in significant gene silencing defect across the entire X chromosome in Pintacuda et al.'s study, several other studies have contested that the degree of silencing defect may less severe than previously reported[30-33]. This may be due to the fact that Pintacuda et al.'s study relied on Xist transgene integration at autosomes, while Bousard et al. and Colognori et al. both expressed Xist at its endogenous locus. Nevertheless, the absence of the B-repeat does result in some degree of defective gene silencing. While the gene silencing defect in A-repeat deletion is immediately noticeable, the silencing defect in B-repeat deletion is most obvious slightly later during differentiation of female mouse embryonic stem cells (mESCs)[33]. Therefore, B-repeat mediated gene silencing may be more crucial during the late establishment phase or maintenance phase, but not the early establishment phase of XCI. This is consistent with the observation that deletion of the B-repeat results in a

defect in chromosome compaction when PRC1 and PRC2 cannot be recruited to deposit their respective heterochromatin marks[34]. Due to this defect in chromosome compaction in B-repeat deletion cells, silencing defects are strongest among genes that typically silence late in XCI. This strongly suggests that when PRC1 and PRC2 cannot be recruited to induce chromatin compaction, Xist cannot spread effectively across the entire X-chromosome, leading to silencing defects of genes that are farther away from the Xist locus in 3D space[34]. Together, the A-repeat and B-repeat of Xist are the primary functional domains responsible for inducing gene silencing during XCI initiation.

Xist E-repeat recruits protein interactors to enable formation of a phase-separated silencing compartment

Liquid – liquid phase separation (LLPS) has been intensely studied in recent years as a ubiquitous mechanism underlying a variety of gene regulatory processes[35]. Briefly, LLPS refers to a phenomenon where multivalent macromolecular interactions past a certain concentration threshold result in the formation of membrane-less structures due to decreased solubility of multivalent oligomer assemblies[36-37]. These phase separated compartments are characterized by higher molecular density but weaker molecular motion compared to its surrounding environment. Many proteins that are involved in LLPS have intrinsically disordered regions (IDRs), which often have repetitive elements of low sequence complexity allowing for multivalent interactions, but do not have a specific 3D structure[35]. DNA and RNA molecules can also promote formation of phase separated compartments due to their ability to form multivalent interactions with other macromolecules depending on their sequence elements. Therefore, DNA and RNA

molecules can serve as scaffolds to create dense local protein concentrations that seed the formation of phase separated biomolecular condensates[38-39].

Many biological processes have recently been characterized to involve LLPS, some of which include stress granule formation, nucleolar sub-compartment formation, heterochromatin formation, and even transcriptional regulation[40-45]. In chapter 3 we report evidence for LLPS in establishment of the inactive X silencing compartment during XCI through protein assembly on the Xist E-repeat[45]. We discovered that the E-repeat directly recruits PTBP1, MATR3, CELF1, and TDP-43. Several of these proteins are known to form higher-order assemblies and induce phase separation in the presence of RNA, which prompted us to further investigate this possibility[46-47]. The E-repeat is a 1.4kb sequence consisting of repetitive C/U/G rich sequences, allowing for multivalent binding of its interactors to create a high local concentration of these proteins on the inactive X. Upon deletion of the E-repeat, Xist clouds became dispersed at late time points of differentiation, demonstrating that while ΔE Xist clouds can initially form during the initiation phase of XCI, the E-repeat is required for retaining Xist localization on chromatin at the maintenance phase of XCI[45]. Synthetic recruitment of PTBP1, MATR3, CELF1, or TDP-43 back onto ΔE Xist RNA rescued the ΔE Xist cloud formation phenotype, indicating that recruitment of these proteins is required for Xist localization on chromatin. Using in vitro droplet formation assay, we demonstrated that addition of E-repeat RNA together with PTBP1 purified protein readily formed droplets resembling phase-separated liquids[45]. This supports our hypothesis the multivalent binding of PTBP1 to the E-repeat can result in a phase separated silencing compartment on the inactive X, and that

retention of Xist molecules on the Xi is dependent on the establishment of this membrane-less compartment.

Xist seeds supramolecular complexes to propagate chromosome-wide silencing

One of the most intriguing discoveries regarding Xist mechanism in the past decade is the fact that an Xist "cloud" that forms when it coats the inactive X during XCI consists of only about 100 Xist molecules[34,48]. Previously, the number of Xist molecules estimated to constitute an Xist cloud ranged from 300-2000 RNA molecules[49-50]. However, this was proven to be incorrect using super resolution microscopy[34,48]. This observation is particularly surprising since 100 Xist molecules can only cover about 1% of the X chromosome surface area at a given time, and the X chromosome has about 1000 genes. Based on their super resolution microcopy data, Sunwoo et al. proposed that Xist acts in a "hit-and-run" fashion, where Xist RNPs would each locally spread within a distance of 1-3 Mb to induce histone methylation for gene silencing[48]. This turned out not to be true, as Markaki et al. used super resolution microscopy and live cell imaging to demonstrate that movement of Xist foci are actually locally confined[34]. Each Xist focus consists of two Xist molecules confined to about 200nm displacement distance. While movement of these Xist foci are locally confined, the movement of proteins they recruit are incredibly dynamic. Markaki et al. discovered that Xist direct interactors SPEN, CELF1, and PCG5, concentrate at super stoichiometric levels on the inactive X[34]. These proteins also exhibit much less restricted movement distances and much more rapid binding kinetics than Xist molecules in Xist RNPs. Therefore, Xist induces formation of supramolecular complexes (SMACs), where proteins recruited by Xist form high local

protein concentrations that move dynamically between the locally confined Xist molecules. Furthermore, IDR containing Xist interactors facilitate supramolecular aggregation to increase local protein concentrations and enhance enzymatic activity under increased crowding conditions for efficient gene silencing[34]. This novel XCI model involving Xist seeding of dynamic SMACs explains how locally confined and sub-stoichiometric levels of Xist RNA can propagate gene silencing across the entire X chromosome.

Positive and negative regulators of Xist expression

Xist expression is regulated by a complex network of molecules consisting of both proteins and lncRNAs[51]. The first major regulator of Xist expression discovered is Tsix, which is a 40kb lncRNA transcribed antisense to Xist[52]. Tsix transcription start site is 15kb downstream of Xist, and the length of the Tsix gene completely overlaps with Xist. In male and female mESCs, Tsix expression is always on before XCI is induced[52]. Upon heterozygous Tsix deletion in female mESCs, XCI becomes non-random such that ΔTsix allele will always be the one chosen for inactivation[53]. Therefore, Tsix is a strong negative regulator of Xist and plays a crucial role in XCI allelic choice. While Tsix deletion does result in ectopic Xist expression in male mESCs during differentiation, Xist expression is eventually silenced, indicating that there are other negative regulators that play a role in down regulating Xist expression[54]. Jpx and Ftx are both lncRNAs upregulated at the onset of XCI and identified to be activators of Xist expression[55-56]. Jpx deletion results in loss of Xist clouds and significantly reduced Xist RNA levels[55]. Interestingly, autosomal insertion of Jpx rescues Xist cloud formation, which shows that Jpx acts in trans. Ftx deletion also

results in reduced Xist RNA levels as well as increased CpG island methylation at Xist[56]. This suggests that Ftx expression regulates Xist expression through epigenetic modifications at the Xist locus.

Several protein factors have also been demonstrated to regulate Xist expression[51]. The first X-linked activator of Xist identified was the E3 ubiquitin ligase RNF12[57]. Deletion of Rnf12 resulted in significant reduction in Xist expression, while insertion of additional transgenic copies of Rnf12 resulted in induction of XCI in both male and female mESCs[57-58]. In addition, RNF12 has been shown to activate Xist expression specifically through proteasomal degradation of pluripotency factor REX1[59]. REX1 can directly bind Xist to downregulate its expression, and defective XCI due to Rnf12 deletion can be rescued via deletion of Rex1[59-60]. Since REX1 is a pluripotency factor that directly represses Xist expression, and RNF12 is an Xist activator expressed from the X-chromosome, it is possible that RNF12 is a X-linked dosage dependent regulator of Xist expression. YY1 is another protein factor that plays a crucial role in activating Xist expression[61]. Upon YY1 knockdown in female MEFs, the proportion of cells with Xist clouds were reduced by 80% and Xist RNA levels were reduced by 75%. This result was also reproducible in differentiating female mESCs, demonstrating that YY1 is required for Xist expression during both initiation and maintenance phase of XCI. Furthermore, three YY1 binding sites identified just upstream of the Xist F-repeat DNA have been shown to be required for Xist upregulation during XCI[61]. RIF1 has recently been identified as a potent activator of Xist expression[62]. Similar to YY1, RIF1 also binds Xist DNA and deletion of RIF1 also results in failure to initiate Xist expression upon differentiation of female mESCs. Interestingly, RIF1 initially binds both Xist alleles in the undifferentiated state, but switches

to monoallelic binding only at the future Xi, suggesting that RIF1 plays a role in XCI allelic choice[62]. In chapter 2, we report the identification of the Xist F-repeat as a cis regulatory element required for Xist expression. In addition, we show that the E2F binding sites located within the F-repeat sequence are functional in regulating Xist expression. Using a sophisticated in vitro pulldown assay followed by mass spectrometry analysis, we identified E2F3 as a novel Xist F-repeat interactor that may potentially activate Xist expression. Our results reveal a new regulatory mechanism of Xist expression that further elucidates the requirements for Xist regulation during XCI.

Given that Xist expression is regulated by many cis and trans acting molecules, understanding the Xist genomic locus could provide clues as to how these regulators coordinate Xist expression during initiation of XCI. The genomic locus that encodes Xist and many of its regulators is termed the X inactivation center (XIC)[63-64]. The XIC is organized into two topologically associated domains (TADs) (Fig. 3)[63]. Interestingly, the boundary between the two TADs is located between the Xist and Tsix promoters. The Xist TAD also harbors Xist activators Rnf12, as well as lncRNAs Jpx and Ftx. The primary antagonist of Xist expression Tsix and its activator Xite are found in the Tsix TAD[63]. This spatial partitioning of Xist positive and negative regulators into separate TADs within the XIC is an excellent example of how higher order chromosome organization can shape gene regulatory landscapes.

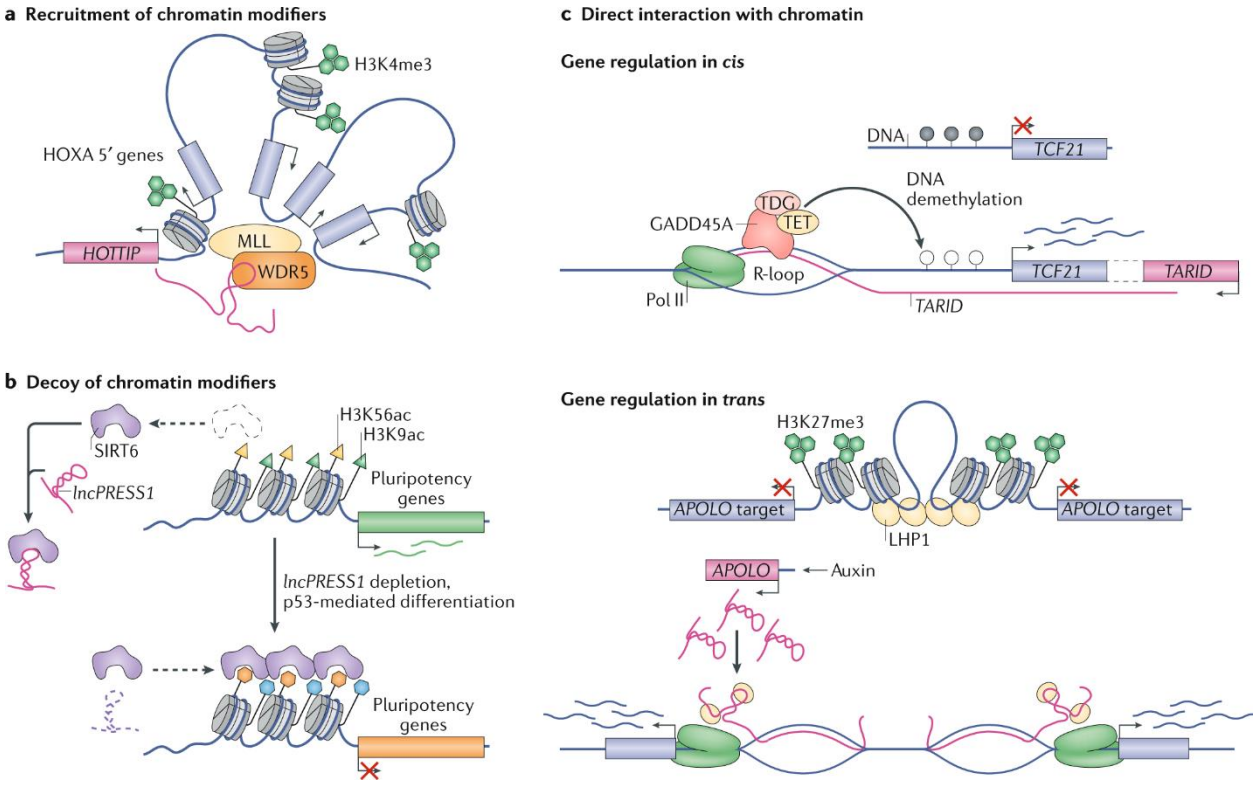


Figure 1-1 Post-transcriptional mechanisms of lncRNA function through recruitment of chromatin modifiers. (Statello et al., 2020)[2]

lncRNAs can function using a wide variety of mechanisms post-transcriptionally, one of which is recruitment of chromatin modifiers to regulate target genes.

A) The lncRNA HOTTIP interacts with WDR5/MLL to promote H3K4 trimethylation at HOXA genes for silencing.

B) The lncRNA lncPRESS1 is able to support embryonic stem cell pluripotency by sequestering the histone deacetylase sirtuin 6 (SIRT6) from promoters of pluripotency genes.

C) lncRNAs can also directly interact with chromatin to form RNA-DNA hybrids (R-loops) for gene activation or repression, in cis or in trans. The lncRNA TARID forms an R-loop upstream of the promoter of TCF21, which results in GADD45A recognition of the R-loop,

GADD45A interaction with TDG and TET1, and subsequent demethylation of the TCF21 promoter for gene activation. The lncRNA APOLO regulates target gene expression by forming R-loops as well. Normally LHP1 represses APOLO target gene expression by formation of chromatin loops and H3K27me3. Expression of APOLO reverses LHP1 repression by forming R-loops at promoter of these target genes, which at as decoys for LHP1 binding.

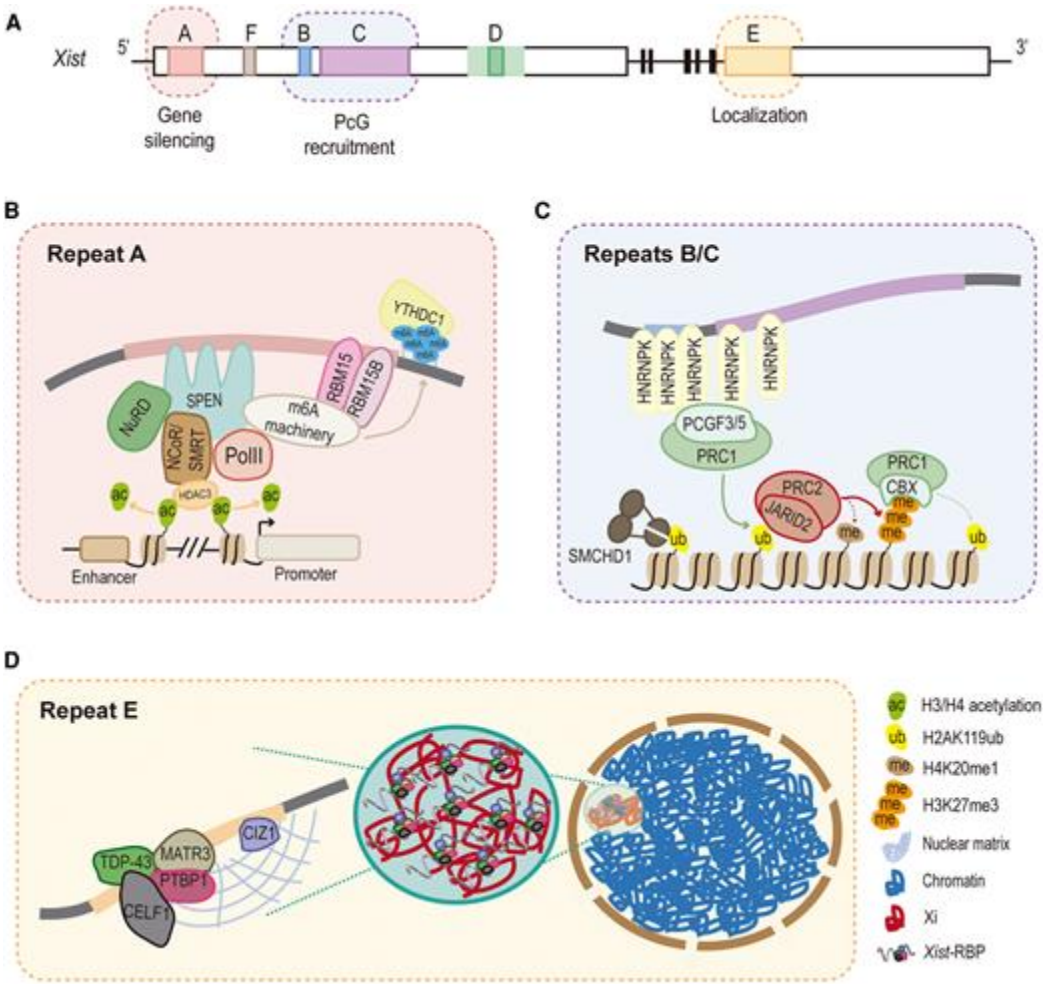


Figure 1-2 Different aspects of Xist function is mediated by its individual repeat domains. (Raposo et al., 2021)[24]

A) The positions of tandem repeat domains A-F on the Xist transcript are shown.

B) The A-repeat directly interacts with SPEN to recruit NCoR/SMRT and NuRD repressor complexes for gene silencing.

C) The B/C-repeats binds HNRNPK to recruit PRC1 and PRC2 to mediate chromatin compaction

D) The E-repeat recruits PTBP1, MATR3, CLEF1, and TDP-43 to establish a phase-separated nuclear silencing compartment

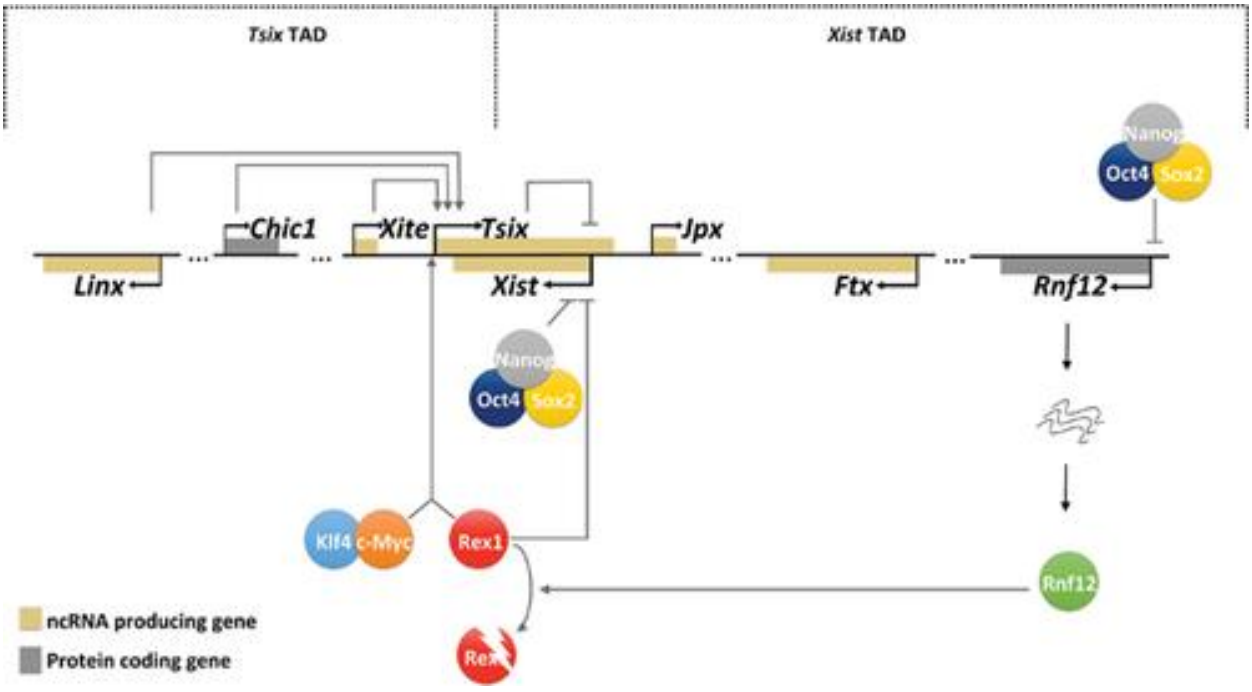


Figure 1-3 The X inactivation center is partitioned between two topologically associated domains. (Vallot et al., 2016)[64]

The X inactivation center consists of two adjacent topologically associated domains, where Xist activators are found in the Xist TAD, and Xist regulators are found in the Tsix TAD. Non-coding genes are bronze and protein coding genes are grey.

References

1. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol. 2015 Jan;22(1):5-7. PMID: 25565026.

2. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021 Feb;22(2):96-118. doi: 10.1038/s41580-020-00315-9. Epub 2020 Dec 22. Erratum in: Nat Rev Mol Cell Biol. 2021 Jan 8;: PMID: 33353982; PMCID: PMC7754182.

3. Perry RB, Ulitsky I. The functions of long noncoding RNAs in development and stem cells. Development. 2016 Nov 1;143(21):3882-3894. PMID: 27803057.

4. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR, Morse M, Engreitz J, Lander ES, Guttman M, Lodish HF, Flavell R, Raj A, Rinn JL. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol. 2014 Feb;21(2):198-206. PMCID: PMC3950333.

5. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science. 2008 Dec 12;322(5908):1717-20. PMID: 18988810.

6. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. Nat Rev Genet. 2020 Feb;21(2):102-117. PMID: 31729473.

7. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007 Jun 29;129(7):1311-23. PMCID: PMC2084369.

8. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature. 1991 Jan 3;349(6304):38-44. PMID: 1985261.

9. Avner P, Heard E. X-chromosome inactivation: counting, choice and initiation. Nat Rev Genet. 2001 Jan;2(1):59-67. PMID: 11253071.

10. Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet. 2002;36:233-78. PMID: 12429693.

11. Galupa R, Heard E. X-chromosome inactivation: new insights into cis and trans regulation. Curr Opin Genet Dev. 2015 Apr;31:57-66. PMID: 26004255.

12. Goodrich L, Panning B, Leung KN. Activators and repressors: A balancing act for X-inactivation. Semin Cell Dev Biol. 2016 Aug;56:3-8. PMID: 27223409.

13. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat Rev Mol Cell Biol. 2016 Dec;17(12):756-770. PMID: 27780979.

14. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. Nature. 1996 Jan 11;379(6561):131-7. PMID: 8538762.

15. Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 1997 Jan 15;11(2):156-66. PMID: 9009199.

16) Lee JT, Jaenisch R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature. 1997 Mar 20;386(6622):275-9. PMID: 9069285.

17. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science. 2013 Aug 16;341(6147):1237973. PMCID: PMC3778663.

18. Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature. 2013 Dec 19;504(7480):465-469. PMCID: PMC3904790.

19. Chaumeil J, Le Baccon P, Wutz A, Heard E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. 2006 Aug 15;20(16):2223-37. PMCID: PMC1553206.

20. Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 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. 1992 Oct 30;71(3):527-42. PMID: 1423611.

21. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 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. 1992 Oct 30;71(3):515-26. PMID: 1423610.

22. Nesterova TB, Slobodyanyuk SY, Elisaphenko EA, Shevchenko AI, Johnston C, Pavlova ME, Rogozin IB, Kolesnikov NN, Brockdorff N, Zakian SM. 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. 2001 May;11(5):833-49. PMCID: PMC311126.

23. Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet. 2002 Feb;30(2):167-74. PMID: 11780141.

24. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. Systematic discovery of Xist RNA binding proteins. Cell. 2015 Apr 9;161(2):404-16. PMCID: PMC4425988.

25. McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature. 2015 May 14;521(7551):232-6. PMCID: PMC4516396.

26. Monfort A, Di Minin G, Postlmayr A, Freimann R, Arieti F, Thore S, Wutz A. Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. Cell Rep. 2015 Jul 28;12(4):554-61. PMCID: PMC4530576.

27. Dossin F, Pinheiro I, Żylicz JJ, Roensch J, Collombet S, Le Saux A, Chelmicki T, Attia M, Kapoor V, Zhan Y, Dingli F, Loew D, Mercher T, Dekker J, Heard E. SPEN integrates transcriptional and epigenetic control of X-inactivation. Nature. 2020 Feb;578(7795):455-460. PMCID: PMC7035112.

28. Bornelöv S, Reynolds N, Xenophontos M, Gharbi S, Johnstone E, Floyd R, Ralser M, Signolet J, Loos R, Dietmann S, Bertone P, Hendrich B. The Nucleosome

Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. Mol Cell. 2018 Jul 5;71(1):56-72.e4. PMCID: PMC6039721.

29. Żylicz JJ, Bousard A, Žumer K, Dossin F, Mohammad E, da Rocha ST, Schwalb B, Syx L, Dingli F, Loew D, Cramer P, Heard E. The Implication of Early Chromatin Changes in X Chromosome Inactivation. Cell. 2019 Jan 10;176(1-2):182-197.e23. PMCID: PMC6333919.

30. Pintacuda G, Wei G, Roustan C, Kirmizitas BA, Solcan N, Cerase A, Castello A, Mohammed S, Moindrot B, Nesterova TB, Brockdorff N. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol Cell. 2017 Dec 7;68(5):955-969.e10. PMCID: PMC5735038.

31. Nesterova TB, Wei G, Coker H, Pintacuda G, Bowness JS, Zhang T, Almeida M, Bloechl B, Moindrot B, Carter EJ, Alvarez Rodrigo I, Pan Q, Bi Y, Song CX, Brockdorff N. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. Nat Commun. 2019 Jul 16;10(1):3129. PMCID: PMC6635394.

32. Bousard A, Raposo AC, Żylicz JJ, Picard C, Pires VB, Qi Y, Gil C, Syx L, Chang HY, Heard E, da Rocha ST. The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. EMBO Rep. 2019 Oct 4;20(10):e48019. PMCID: PMC6776897.

33. Colognori D, Sunwoo H, Wang D, Wang CY, Lee JT. Xist Repeats A and B Account for Two Distinct Phases of X Inactivation Establishment. Dev Cell. 2020 Jul 6;54(1):21-32.e5. PMCID: PMC7362899.

34. Markaki Y, Gan Chong J, Wang Y, Jacobson EC, Luong C, Tan SYX, Jachowicz JW, Strehle M, Maestrini D, Banerjee AK, Mistry BA, Dror I, Dossin F, Schöneberg J, Heard E, Guttman M, Chou T, Plath K. Xist nucleates local protein gradients to propagate silencing across the X chromosome. Cell. 2021 Dec 9;184(25):6174-6192.e32. PMC8671326.

35. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nat Rev Mol Cell Biol. 2017 May;18(5):285-298. PMC7434221.

36. Brangwynne CP. Soft active aggregates: mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. Soft Matter [Internet]. 2011 Mar 22;7(7):3052–9. Available from: https://pubs.rsc.org/en/content/articlehtml/2011/sm/c0sm00981d

37. Brangwynne CP. Phase transitions and size scaling of membrane-less organelles. J Cell Biol. 2013 Dec 23;203(6):875-81. PMCID: PMC3871435.

38. Lin Y, Protter DS, Rosen MK, Parker R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Mol Cell. 2015 Oct 15;60(2):208-19. PMCID: PMC4609299.

39. Zhang H, Elbaum-Garfinkle S, Langdon EM, Taylor N, Occhipinti P, Bridges AA, Brangwynne CP, Gladfelter AS. RNA Controls PolyQ Protein Phase Transitions. Mol Cell. 2015 Oct 15;60(2):220-30. PMCID: PMC5221516.

40. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, Taylor JP. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell. 2015 Sep 24;163(1):123-33. PMCID: PMC5149108.

41. Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV, Brangwynne CP. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. Cell. 2016 Jun 16;165(7):1686-1697. PMCID: PMC5127388.

42. Strom AR, Emelyanov AV, Mir M, Fyodorov DV, Darzacq X, Karpen GH. Phase separation drives heterochromatin domain formation. Nature. 2017 Jul 13;547(7662):241-245. PMCID: PMC6022742.

43. Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, Abraham BJ, Afeyan LK, Guo YE, Rimel JK, Fant CB, Schuijers J, Lee TI, Taatjes DJ, Young RA. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell. 2018 Dec 13;175(7):1842-1855.e16. PMCID: PMC6295254.

44. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, Li CH, Guo YE, Day DS, Schuijers J, Vasile E, Malik S, Hnisz D, Lee TI, Cisse II, Roeder RG, Sharp PA, Chakraborty AK, Young RA. Coactivator condensation at super-enhancers links phase separation and gene control. Science. 2018 Jul 27;361(6400):eaar3958. PMCID: PMC6092193.

45. Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Leon WRM, Damianov A, Chronis C, Papp B, Chen CK, McKee R, Wang XJ, Chau A, Sabri S, Leonhardt H, Zheng S, Guttman M, Black DL, Plath K. Publisher Correction: A protein assembly mediates Xist localization and gene silencing. Nature. 2020 Oct;586(7830):E30. PMID: 33005055.

46. Gallego-Iradi MC, Strunk H, Crown AM, Davila R, Brown H, Rodriguez-Lebron E, Borchelt DR. N-terminal sequences in matrin 3 mediate phase separation into droplet-like structures that recruit TDP43 variants lacking RNA binding elements. Lab Invest. 2019 Jul;99(7):1030-1040. PMCID: PMC6857798.

47. Cerase A, Armaos A, Neumayer C, Avner P, Guttman M, Tartaglia GG. Phase separation drives X-chromosome inactivation: a hypothesis. Nat Struct Mol Biol. 2019 May;26(5):331-334. PMID: 31061525.

48. Sunwoo H, Wu JY, Lee JT. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. Proc Natl Acad Sci U S A. 2015 Aug 4;112(31):E4216-25. PMCID: PMC4534268.

49. Buzin CH, Mann JR, Singer-Sam J. Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. Development. 1994 Dec;120(12):3529-36. PMID: 7529677.

50. Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell. 2006 Mar 3;21(5):617-28. PMID: 16507360.

51. Goodrich L, Panning B, Leung KN. Activators and repressors: A balancing act for X-inactivation. Semin Cell Dev Biol. 2016 Aug;56:3-8. PMID: 27223409.

52. Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet. 1999 Apr;21(4):400-4. doi: 10.1038/7734. PMID: 10192391.

53. Lee JT, Lu N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. Cell. 1999 Oct 1;99(1):47-57. PMID: 10520993.

54. Sado T, Li E, Sasaki H. Effect of TSIX disruption on XIST expression in male ES cells. Cytogenet Genome Res. 2002;99(1-4):115-8. PMID: 12900553.

55. Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell. 2010 Oct 29;143(3):390-403. PMCID: PMC2994261.

56. Chureau C, Chantalat S, Romito A, Galvani A, Duret L, Avner P, Rougeulle C. Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. Hum Mol Genet. 2011 Feb 15;20(4):705-18. PMID: 21118898.

57. Jonkers I, Barakat TS, Achame EM, Monkhorst K, Kenter A, Rentmeester E, Grosveld F, Grootegoed JA, Gribnau J. RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. Cell. 2009 Nov 25;139(5):999-1011. PMID: 19945382.

58. Barakat TS, Gunhanlar N, Pardo CG, Achame EM, Ghazvini M, Boers R, Kenter A, Rentmeester E, Grootegoed JA, Gribnau J. RNF12 activates Xist and is essential for X chromosome inactivation. PLoS Genet. 2011 Jan 27;7(1):e1002001. PMCID: PMC3029249.

59. Gontan C, Achame EM, Demmers J, Barakat TS, Rentmeester E, van IJcken W, Grootegoed JA, Gribnau J. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. Nature. 2012 Apr 29;485(7398):386-90. PMID: 22596162.

60. Gontan C, Mira-Bontenbal H, Magaraki A, Dupont C, Barakat TS, Rentmeester E, Demmers J, Gribnau J. REX1 is the critical target of RNF12 in imprinted X chromosome inactivation in mice. Nat Commun. 2018 Nov 12;9(1):4752. PMCID: PMC6232137.

61. Makhlouf M, Ouimette JF, Oldfield A, Navarro P, Neuillet D, Rougeulle C. A prominent and conserved role for YY1 in Xist transcriptional activation. Nat Commun. 2014 Sep 11;5:4878. PMCID: PMC4172967.

62. Enervald E, Powell LM, Boteva L, Foti R, Blanes Ruiz N, Kibar G, Piszczek A, Cavaleri F, Vingron M, Cerase A, Buonomo SBC. RIF1 and KAP1 differentially regulate the choice of inactive versus active X chromosomes. EMBO J. 2021 Dec 15;40(24):e105862. PMCID: PMC8672179.

63. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012 Apr 11;485(7398):381-5. PMCID: PMC3555144.

64. Vallot C, Ouimette JF, Rougeulle C. Establishment of X chromosome inactivation and epigenomic features of the inactive X depend on cellular contexts. Bioessays. 2016 Sep;38(9):869-80. doi: 10.1002/bies.201600121. Epub 2016 Jul 8. PMID: 27389958.

Chapter 2

Research Project

The Xist F-repeat is required for Xist expression during the initiation of X chromosome inactivation

Chau A, Sun Y, Weixian D, Sha J, Korsakova E, Pandya-Jones A, Mancia Leon WR, Wohlschlegel J, Plath K

Abstract

The lncRNA Xist provides a remarkable model to investigate the function of lncRNAs in gene regulation, as it induces chromosome-wide silencing in cis in the process of X-chromosome inactivation (XCI)[1]. XCI is fundamentally important for female mammalian development, but, despite its critical role, the mechanisms by which Xist carries out the various tasks associated with XCI still remain largely unclear. Xist is thought to fulfill different roles during XCI, such as gene silencing, chromatin association, spreading, and recruitment of repressive chromatin regulators, through its different functional domains [2-4]. Here, we sought to define Xist sequences required for Xist function that have yet to be characterized, through systematic deletional analysis in mouse embryonic stem cells (mESCs) using CRISPR/Cas9. We identified the Xist F-repeat as a transcriptional regulatory element that is essential for Xist expression during the initiation of XCI. Furthermore, we demonstrate that the E2F binding sites within the F-repeat sequence are critical elements for F-repeat function, and identify E2F3 as a potential regulator of Xist expression. Together, our results reveal a novel role for a previously uncharacterized Xist domain in Xist function, and identified a novel transcriptional regulator of Xist expression.

Introduction

Mammalian gene regulation is an incredibly complex process, requiring multiple layers of regulation and involving a broad range of intricate mechanisms. Recently, long noncoding RNAs (lncRNAs) have emerged as an important class of molecules that can function in transcriptional regulation, the most well-known and most intensely studied of them being *Xist*[1-2]. *Xist* presents a unique model to study lncRNA localization, lncRNA-mediated transcriptional repression, chromatin changes and reorganization of 3D chromatin structure. *Xist* expression is both necessary and sufficient to induce chromosome-wide silencing in X-chromosome inactivation (XCI), which is required to achieve dosage compensation of X-linked genes in female cells[3]. Upon expression, *Xist* spreads *in cis* from its site of transcription on the X chromosome, first localizing to sites proximal to the Xist locus in 3D space, then continually spreading to coat the entire X chromosome[4]. Hallmarks of XCI downstream of Xist spreading include gene silencing, depletion of active histone marks, enrichment of repressive histone marks including H3K27me3 and H3K9me2/3, exclusion of RNA polymerase II, DNA methylation, and chromosome compaction[5]. *Xist* contains six repeat domains termed A-F that are highly conserved among placental mammals (Fig 1)[6-8]. These domains are of interest in studying *Xist* function since high sequence conservation suggests functional importance. In addition, the intervening non-repeat regions have been shown to bind proteins as well, indicating that these regions also form functional domains[9]. To date, a clear function is only assigned to the most 5' A-repeat region, which is critical for silencing of the chromosome by *Xist*[10]. It is currently believed that the functional domains of *Xist* act independently of each other, primarily through recruiting proteins for mediating different functions[5,6].

Several *Xist*-protein interactions have been characterized in recent years. The *Xist* A repeat has been demonstrated to interact with SPEN, which recruits SMRT to activate HDAC3 for histone deacetylation and thereby mediate silencing[10]. Silencing also requires the binding of RBM15 to the A repeat[11]. *Xist's* direct interaction with Lamin B receptor (LBR) at non-repeat sequences is required for X inactivation, possibly through facilitating *Xist* spreading across the X chromosome during XCI initiation[9]. HNRNP-K has been shown to directly bind the B and C repeats to recruit PCGF3/5-PRC1 to initiate polycomb-mediated deposition of heterochromatin marks[12]. However, many additional proteins identified as direct interactors of *Xist* require further characterization to understand where they bind on *Xist* and how each protein functions mechanistically in XCI. Importantly, the majority of *Xist* 's sequence has not been functionally characterized.

Previous published studies have described functional roles of specific *Xist* domains, but did not find functions for most Xist sequences[6]. However, these studies were mostly conducted using an inducible *Xist* transgene expression system where Xist is expressed from single X chromosome in male ESCs, resulting in cell death 24hrs after the induction of Xist due to silencing of genes on the sole X chromosome[6]. With this approach, which is used because it is synchronous and does not require induction of differentiation, only early steps in the initiation of XCI can be studied, suggesting that critical steps of XCI and Xist function could be missed. Moreover, the assay is performed without the induction of differentiation which is normally required for XCI to occur in female development. Thus, not all steps of XCI and Xist-function can be assessed in this system. In addition, other studies use often use terminally differentiated cells, where XCI has already occurred[13]. This method would only assess Xist function during the maintenance

phase of XCI, but not the initiation phase. Therefore, we set out to study how each *Xist* domain contributes to XCI in the native biological context of female embryonic stem cell differentiation. We systematically deleted Xist domains using CRISPR/Cas9 to identify sequences required for Xist function.

Results

Identification of functional Xist sequences through CRISPR/Cas9 deletional analysis

To identify additional functional sequences within Xist that have yet to be characterized, we utilized a mESC line (termed F1 2-1) that was derived from embryos of a F1 cross between two mouse strains: CAST and 129. Moreover, the 129 *Xist* allele is tagged with 11X MS2 hairpins downstream of the E repeat (Fig. S1), so that the two *Xist* alleles can be differentiated by RNA FISH, which is particularly important for heterozygous deletions. Our lab previously generated homozygous Xist-MS2 mice indicating that the MS2 hairpins do not interfere with function (not shown). Differentiation of female ESCs is the ideal model to study XCI, since ESC differentiation is the native biological context in which XCI occurs. We first generated a heterozygous deletion of a large region in exon 1 of Xist that contains many putative functional sequences (termed ΔF-D) using CRISPR/Cas9 (Fig. S2). The deleted 6.5 kb region contains the F, B, C, and D repeats and the intervening non-repeat regions on either the 129 or CAST allele using CRISPR/Cas9 ($X_{cas}X_{129}^{\Delta F\text{-}D\ -\ 11x\text{-}MS2}$ or $X_{cas}^{\ \Delta F\text{-}D}\ X_{129}^{11x\text{-}MS2}$ respectively) (Fig. 1A). We hypothesized that a large deletion in Xist encompassing several repeat and non-repeat regions would yield a non-functional *Xist* allele, leading to skewing of XCI choice towards choosing the WT Xist allele for Xist upregulation. The A repeat was not included in the deletion since that region is known to be required for the induction of silencing by Xist through the recruitment of the protein SPEN. We differentiated ΔF-D heterozygous mESCs for 6 days through withdrawal of LIF and addition of retinoic acid to induce X inactivation. By performing RNA fluorescence in situ hybridization (FISH) using

fluorescent probes against Xist and MS2, we found that these *Xist* mutant cells inactivate the mutant allele exclusively during differentiation (Fig. 1B,C).The absolute skewing of XCI choice in these cells towards the WT Xist allele indicates that there are functional sequences within the deleted region that is required for Xist function.

To determine the individual Xist functional sequences within the ΔF-D deletion, we generated six additional heterozygous Xist domain deletion mESCs (Fig. 1A). Each Xist domain deletion mESC line harbors the deletion only on the MS2+ 129 allele. We differentiated these Xist mutants for 6 days through LIF withdrawal and addition of retinoic acid, and performed RNA FISH to determine whether any of the Xist mutants exhibited an XCI choice skewing phenotype similar to that in the ΔF-D heterozygous mutant describe previously (Fig. 1D,E). The Xist Δ5' deletion recapitulated the XCI choice skewing phenotype, which indicates that the functional element responsible for this phenotype is within the 5' end of the ΔF-D deleted sequence. At day 6 of differentiation, over 90% of Δ5' heterozygous Xist mutant clones have chosen the WT allele for XCI regardless of whether the Δ5' deletion is on the MS2+ or MS2- allele (Fig. 1 F,G). Considering that only the ΔF-D and Δ5' Xist mutant mESCs exhibit skewing in XCI choice, the Xist sequence that is responsible for this phenotype must be deleted only in these two genotypes and not in any of the other Xist domain deletion mutants. By comparing the deleted regions among all the Xist mutants, we were able to identify the F-repeat region as the element that is only deleted in ΔF-D and Δ5' Xist mutant mESCs. Therefore, the Xist F-repeat must contribute to Xist function in XCI.

The F-repeat is required for Xist cloud formation

We next sought to confirm that deletion of only the Xist F-repeat domain is sufficient to recapitulate the XCI skewing phenotype when introduced as a heterozygous deletion. To generate a mESC line of this genotype, we utilized CRISPR/Cas9 mediated homology directed repair (HDR) to delete the Xist F-repeat in F1 2-1 female mESC line (Fig. S4). As expected, we observed that there is significantly reduced Xist cloud formation on the mutant Xist allele at both day 2 and day 6 of retinoic acid induced differentiation (Fig. 2A,B,C,D). No mutant Xist clouds can be observed at day 2 of differentiation. About 25% of Xist clouds observed at day 6 of differentiation come from the ΔF MS2+ allele compared to about 80% of MS2+ Xist clouds in WT F1 2-1 at that time point. This severe reduction of ΔF Xist cloud formation demonstrates that the Xist F-repeat is a crucial element required for Xist cloud formation.

To understand why there is such a reduction in Xist cloud formation in the absence of the Xist F-repeat, we performed quantitative real-time PCR (qPCR) to measure the Xist RNA levels in ΔF Xist vs WT Xist. Since the F-repeat deletion is only introduced on the MS2+ 129 Xist allele for $X_{cas}X_{129}^{\Delta F\text{ - }11x\text{-}MS2}$ mESC, we used qPCR primers that specifically bind the MS2 tag of Xist from the 129 Xist allele to specifically quantify only the ΔF Xist RNA levels. By measuring ΔF Xist RNA levels by qPCR at day 0, day 2, and day 6 of differentiation, we observed that ΔF Xist RNA levels are significantly lower compared to WT Xist RNA levels at each time point in differentiation (Fig. 2E). Therefore, the reduction in Xist cloud formation in ΔF mESC is most likely due to decreased steady state ΔF Xist RNA levels.

Induced expression of ΔF Xist RNA does results in Xist cloud formation and X-linked gene silencing

Since our ΔF female mESC data suggests that decreased mutant Xist RNA levels of ΔF mESC is the reason for the reduction in ΔF Xist cloud formation, we next asked whether forced expression of ΔF Xist could result in Xist cloud formation and XCI. To assess whether induced expression of ΔF Xist results in proper XCI, we deleted the Xist F-repeat region in PSM33 male mESC (Fig. S5). This mESC line expresses the reverse tetracycline transactivator (rtTA) and harbors a tetracycline responsive element incorporated at the endogenous Xist promotor, allowing for induced Xist expression with the addition of doxycycline. After 24hrs of doxycycline induction, ΔF PSM33 mESCs form Xist clouds that are morphologically indistinguishable from WT PSM33 Xist clouds based on Xist RNA FISH (Fig. 3A). The proportion of cells that have an Xist cloud after 24hr doxycycline induction is also the same between ΔF clones and WT PSM33 (Fig. 3B). This demonstrates that ΔF Xist RNA can form Xist clouds when expressed at adequate levels. One of the hallmarks of XCI is the enrichment of heterochromatic marks H3K27me3 and H2AK119ub[14]. In order to assess whether ΔF Xist can induce heterochromatin formation, we performed immunofluorescence combined with Xist RNA FISH for H3K27me3 and H2AK119ub. After 24hrs of doxycycline induction, H3K27me3 and H2AK119ub enrichment is present on the X chromosome in ΔF PSM33 mESCs (Fig. 3C, D, E, F). The proportion of Xist positive cells that have enrichment of H3K27me3 and H2AK119ub is similar between ΔF and WT PSM33, indicating that heterochromatin formation occurs when ΔF Xist expression is forcibly induced. We conclude that ΔF Xist RNA can form Xist clouds and initiate heterochromatin formation when induced to express.

To determine whether ΔF Xist RNA can silence genes across the entire X chromosome, we performed total mRNA sequencing on a 24hr doxycycline induction time

course in ΔF PSM33 and WT PSM33. We extracted RNA at time points 0hr, 12hr, and 24hr for sequencing library preparation and bioinformatics analysis. To visualize how X linked gene expression differed between the two genotypes at each induction time point, we clustered the samples using principal components analysis (PCA) based on X-linked gene expression information (Fig. 3G). ΔF samples did not cluster separately from WT samples to any significant extent, and all samples clustered together primarily based on time point of induction. Therefore, this indicates that the silencing kinetics of X-linked genes is not significantly different between ΔF PSM33 and WT PSM33. Next, we plotted Xist expression over time to determine if there are any differences in steady state Xist RNA levels during the time course (Fig. 3H). Xist RNA levels between ΔF PSM33 and WT PSM33 were virtually identical. Since transcriptional output of ΔF Xist and WT Xist is the same due to the same doxycycline concentration used for induction between samples, the equal amount of ΔF Xist and WT Xist RNA at each timepoint demonstrates that there is no difference in RNA stability between ΔF Xist and WT Xist. To examine X-linked gene silencing more closely, we plotted the ratio of X-linked gene reads to autosomal reads for each sample (Fig. 3I), as well as a heatmap representation of normalized X-linked gene expression to visualize overall X-linked gene silencing over time (Fig. 3J). Both plots show that overall X-linked gene silencing occurs in ΔF PSM33 similar to that of WT PSM33. Lastly, we utilized the DESeq2 R package to call differentially expressed X-linked genes (Fig. 3K)[15]. Comparison of X-linked gene expression between ΔF PSM33 and WT PSM33 at each timepoint resulted in less than 5 genes that are differentially expressed ($|log2FC| > 1$, adjusted p-value < 0.05), indicating that there is no defect in X-linked gene silencing in ΔF PSM33. Together, these findings show that when ectopically expressed at

sufficiently high levels, ΔF Xist can form Xist clouds, induce heterochromatin formation, and silence X-linked genes.

The Xist F-repeat DNA element can act as a transcriptional enhancer

Thus far, we have shown that Xist cloud formation is impaired in the absence of the Xist F-repeat in female mESC due to decreased Xist RNA levels. While we have demonstrated that ectopic expression of ΔF Xist can result in Xist cloud formation, silencing of X-linked genes with no defects in RNA stability, we cannot conclude that the F-repeat RNA element is not functional in XCI. Therefore, to differentiate whether the F-repeat DNA element or the RNA element is functional, we generated a female mESC line with an inverted F-repeat DNA sequence using CRISPR HDR. We reasoned that an inversion of the F-repeat DNA to its reverse complementary sequence would result in transcription of Xist RNA molecules where the F-repeat sequence is altered specifically at the RNA level. This should abolish any RNA binding protein interaction at the F-repeat RNA sequence that would normally occur during XCI, while DNA binding protein interaction at the F-repeat DNA sequence should not be affected. If inversion of the F-repeat DNA results in a phenotype similar to deletion of the F-repeat, this would indicate that RNA binding protein interaction at the F-repeat RNA element is required for Xist function. Conversely, if the F-repeat DNA sequence inversion does not result in any phenotype, this would suggest that the F-repeat DNA element is functional (since inversion of the DNA sequence does not affect DNA binding protein interaction at that region). We inverted the F-repeat sequence at the MS2+ 129 Xist allele only in F1 2-1 mESC and differentiated these clones for 6 days. Xist clouds from the mutant allele were observed that were morphologically identical to WT Xist clouds at day 6 of differentiation

(Fig. 4A). The ratio of MS2- vs MS2+ Xist clouds in these F-repeat inversion heterozygous clones were the same as that of WT F1 2-1, indicating there is no defect in Xist cloud formation that could affect XCI allelic choice as seen in F-repeat deletion clones (Fig. 4B). Altogether, this F-repeat inversion experiment suggests that the F-repeat DNA element is functional, but not the RNA element.

Given that the F-repeat DNA element may be functional, and deletion of the F-repeat in female mESC results in lower mutant Xist RNA levels, we sought to determine whether the F-repeat DNA sequence can act as a transcriptional enhancer for Xist. To accomplish this, we cloned the Xist F-repeat sequence at the minimal promotor region of a luciferase reporter construct. As a comparison, we also cloned another luciferase reporter construct with a YY1 sites containing Xist sequence that have been well characterized as an enhancer of Xist expression[16]. These luciferase constructs were transfected into either mouse embryonic fibroblasts (MEFs) or HEK293T cells, and luciferase signal was measured 24hrs after transfection. Insertion of the Xist F-repeat sequence significantly increased luciferase expression compared to Xist YY1 sites insertion or minimal promoter alone (Fig. 4C). Since inclusion of the F-repeat sequence at the promoter of a reporter gene can enhance expression of its downstream reporter, this demonstrates that it is possible the F-repeat can act as a transcriptional enhancer for Xist expression in its native context. This result is particularly striking since the level of transcriptional enhancement based on the luciferase assay indicates that the F-repeat is even stronger than the Xist YY1 sites sequence, which is widely accepted as a critical enhancer of Xist expression.

To further corroborate the F-repeat sequence's role as a transcriptional enhancer, we searched for additional evidence of this by analyzing publicly available sequencing datasets that examined chromosomal accessibility or enrichment of epigenetic marks on the X chromosome during female mESC differentiation. Upon examining ATAC-seq data from Gjaltema et al.'s female mESC differentiation, we find that there is a strong ATAC peak at the Xist F-repeat (Fig. 4D)[17]. This indicates that the F-repeat region needs to be accessible in the biological context where Xist expression is induced through differentiation of female mESC. Since accessibility of genomic regions usually correlates with those regions being functional in gene regulation, this suggests that the F-repeat DNA may be a gene regulatory element as well. In addition, we also examined Gjaltema et al.'s H3K27ac Cut&Tag data on differentiating female mESCs. Briefly, Cut&Tag is a next generation sequencing method for profiling protein chromatin interaction genome wide[18]. We find a strong enrichment for H3K27ac at the F-repeat DNA sequence (Fig. 4E). Since H3K27ac is an epigenetic mark associated with active enhancers, this further supports our hypothesis that the F-repeat is a transcriptional enhancer region of Xist.

E2F sites within the Xist F-repeat plays a functional role in regulating Xist expression

Upon closer examination of the F-repeat sequence, we find that there are two consensus binding motifs TTGGCGGGCTTT for the E2F family of transcription factors. To determine whether these E2F transcription factor binding sites are functional in regulating Xist expression in XCI, we introduced a minimal 3 bp mutation to each binding site, only on the MS2+ 129 Xist allele (Fig. 5A). This specific minimal mutation has been shown previously to abolish biding of E2F family of transcription factors[19]. Surprisingly,

we find that these heterozygous E2F sites mutation female mESCs exhibited skewing in XCI allelic choice similar to that seen in ΔF female mESCs at day 6 of differentiation (Fig. 5B, C). Only 31-33% of Xist clouds in the E2F sites mutation clones are coming from the MS2+ mutant allele, compared to 78% of WT F1 2-1 Xist clouds coming from the MS2+ allele. The severity of the skewing phenotype in the E2F sites mutation clones is close to that of ΔF female mESCs (31% vs 25% of total Xist clouds come from the MS2+ mutant Xist allele respectively). This indicates that the E2F sites within the F-repeat constitute the majority of F-repeat functionality.

Having identified the F-repeat as a transcriptional enhancing element and the E2F binding sites within the F-repeat as critical for F-repeat function, we next sought to identify transcriptional activators that bind at the F-repeat DNA. To globally profile DNA binding proteins that can interact with the F-repeat DNA sequence with high specificity, we turned to the Protein-nucleic acid affinity quantification by mass spectrometry in nuclear extracts (PAQMAN) assay[20]. Briefly, this is a quantitative protein DNA binding assay that allows for global profiling and calculation of apparent affinities between proteins and a DNA sequence of interest. 10 Affinity purifications are prepared using nuclear extract applied to a serial dilution titration series of DNA oligonucleotide concentrations. The eluted proteins from each affinity purification are then subjected to isobaric labeling and mass spectrometry analysis to quantify apparent binding affinities proteome wide. The binding specificity of each protein can be determined by plotting the amount of protein found in each pulldown as a fraction of the amount found in the highest concentration pulldown as function of DNA oligonucleotide concentration. This plot is then fitted to the Hill-like curve, which describes protein ligand interactions as a function of ligand concentration. A high

correlation to the Hill-like curve will indicate specificity of the protein for the DNA oligonucleotide. By performing the PAQMAN assay using an F-repeat DNA oligonucleotide and nuclear extract from female mESCs, we identified specific interactors of the F-repeat, one of which is E2F3 (Fig. 5D). The high R^2 value of 0.92 indicates high binding specificity of E2F3 to the F-repeat DNA. Other protein interactors discovered that have an R^2 of >0.85 include XRCC5, XRCC6, CTCF, TFAP2C, ZFP384, and NFIC (Fig. 5E). These proteins had low R^2 values when a scrambled DNA oligonucleotide was used instead of the F-repeat sequence, demonstrating that these interactors were not identified only because they have high affinity for any DNA sequence. As an example of a non-specific interactor that was found in our F-repeat affinity pulldowns, we also plotted the curve for CBX3 that has a low R^2 value of 0.33 (Fig. 5F).

Since we have demonstrated that the E2F binding sites within the F-repeat sequence are functional in XCI and identified E2F3 as an interactor of F-repeat DNA, E2F3 could potentially act as a transcriptional regulator of Xist through binding of the F-repeat DNA.

Discussion

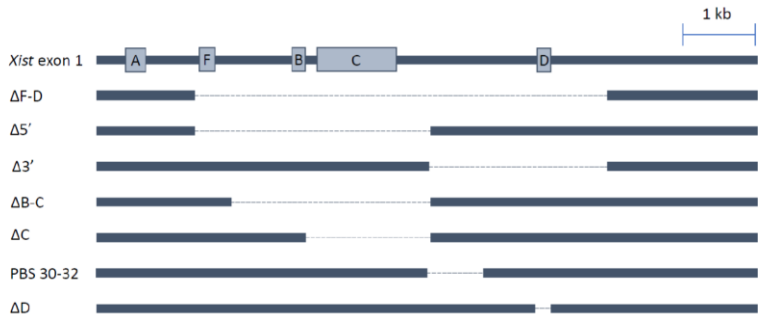
Through CRISPR/Cas9 deletional analysis of Xist sequences in female mESC, we unexpectedly identified the F-repeat region as a cis-regulatory element for Xist expression. When the F-repeat is deleted on only one of the two Xist alleles, a significant proportion of cells chose the WT Xist allele for upregulation, resulting in skewing in XCI choice away from the ΔF Xist allele. Our data demonstrates that this is most likely due to a defect in Xist expression since the absence of the F-repeat results in significantly lower Xist RNA levels, and artificial induction of ΔF Xist expression results in the typical hallmarks of XCI such as Xist cloud formation, heterochromatin formation, and X-linked gene silencing. Therefore, the reduction in Xist clouds upon F-repeat deletion in female mESCs is not due to a defect in Xist RNA localization or gene silencing. Unlike all the other Xist repeat domains that have already been previously characterized, this is the first Xist repeat that has been shown to function at the DNA level rather than the RNA level.

In addition, we further identified the E2F binding sites within the F-repeat sequence as critical components of F-repeat function. Minimal mutations introduced at the E2F binding sites of only one Xist allele resulted in an XCI allelic choice skewing phenotype that was almost as severe as the XCI skewing phenotype observed in heterozygous ΔF female mESCs. This data, along with our identification of E2F3 as an F-repeat interactor using the PAQMAN assay, strongly suggests E2F3 as a transcriptional regulator of Xist expression. We will need to follow up with knockdowns and functional assays to confirm the potential functional role of E2F3 in Xist expression regulation.

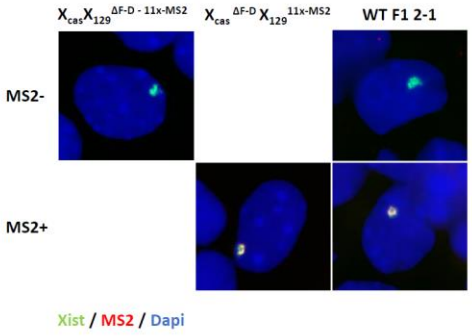
Currently, the gene regulatory networking governing monoallelic upregulation of Xist expression is poorly understood. While prior work has demonstrated certain

transcriptional activators such as YY1, RIF1, and RNF12 play a role in Xist upregulation, details regarding how each factor can synergistically work together to induce Xist expression at the correct developmental time point remains unclear. To further complicate matters, other lncRNAs Jpx and Ftx act as positive regulators of Xist expression as well [21-22]. Our discovery of the F-repeat and E2F3 regulatory aspect of Xist is a significant contribution to our current understanding of Xist regulation. Due to the proximity of the Xist YY1 binding sites to the Xist F-repeat, it is possible that YY1 and F-repeat/E2F3 function are interdependent. RIF1 also binds at the P2 promoter region encompassing the YY1 sites and F-repeat, raising the possibility that all three factors could bind cooperatively to regulate Xist expression[16,23]. Further investigation is required to understand how the F-repeat and E2F3 fit in to the overall Xist expression regulatory network.

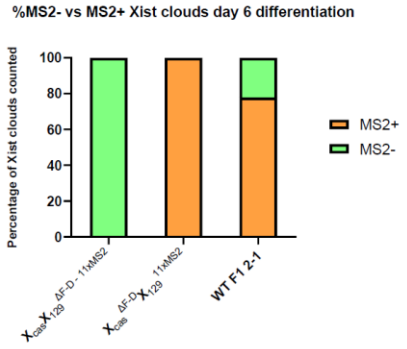
A


1 kb
Xist exon 1
A
F
B
C
D
ΔF-D
Δ5'
Δ3'
ΔB-C
ΔC
PBS 30-32
ΔD

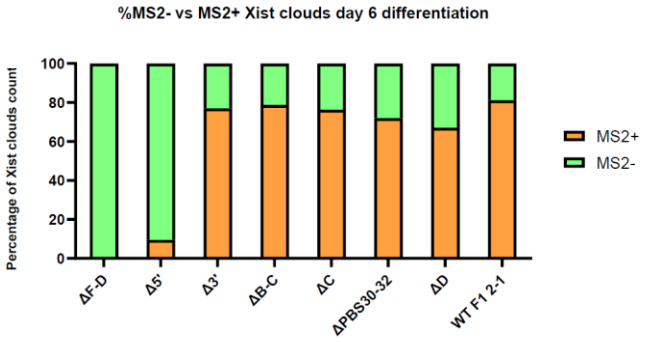
B


WT F1 2-1
MS2-
MS2+
Xist / MS2 / Dapi

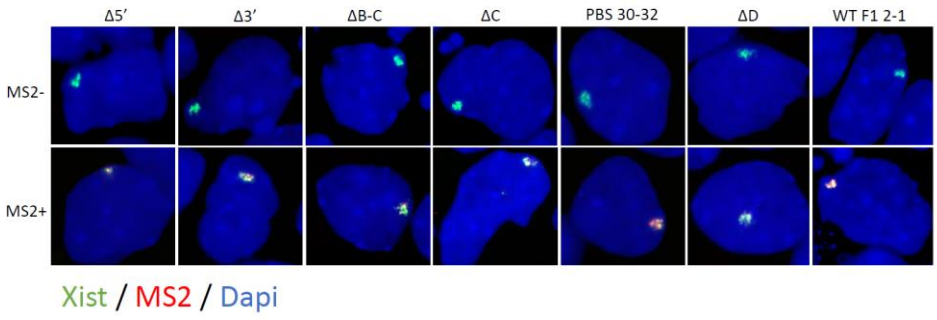
C


%MS2- vs MS2+ Xist clouds day 6 differentiation
Percentage of Xist clouds counted
100
80
60
40
20
0
MS2+
MS2-
WT F1 2-1

D


%MS2- vs MS2+ Xist clouds day 6 differentiation
Percentage of Xist clouds count
100
80
60
40
20
0
ΔF-D
Δ5'
Δ3'
ΔB-C
ΔC
ΔPBS30-32
ΔD
WT F1 2-1
MS2+
MS2-

E


Δ5'
Δ3'
ΔB-C
ΔC
PBS 30-32
ΔD
WT F1 2-1
MS2-
MS2+
Xist / MS2 / Dapi

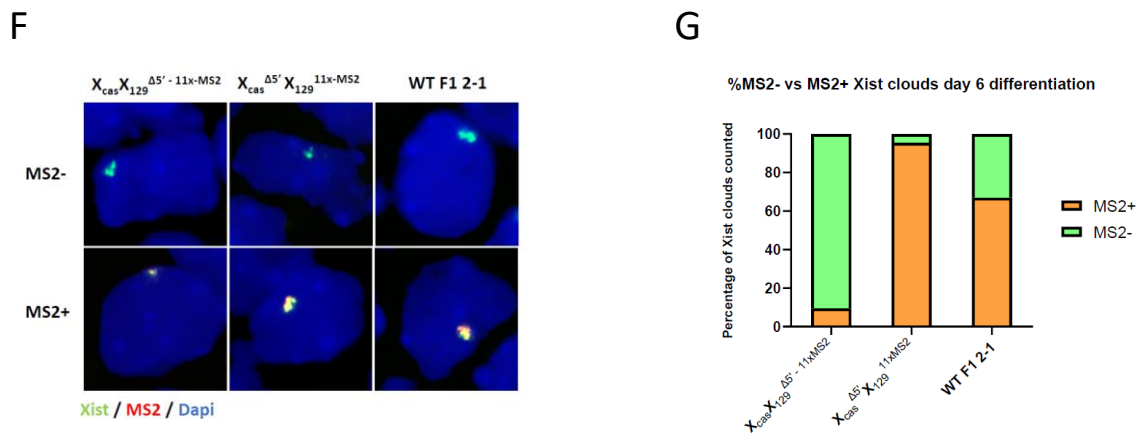


Figure 2-1. Identification of functional Xist sequences through CRISPR/Cas9 deletional analysis

A) Schematic of mutant Xist mESC lines generated using CRISPR/Cas9

B) Representative fluorescence in situ hybridization microscopy images of ΔF-D heterozygous mutant Xist female mESC and WT F1 2-1 female mESC at day 6 of RA differentiation. Xist signal is shown in green, MS2 signal in red, and DAPI in blue.

C) Percentage of Xist positive nuclei in ΔF-D heterozygous mutant Xist female mESC and WT F1 2-1 female mESC that have an MS2+ Xist cloud or MS2- Xist cloud at day 6 of RA differentiation. 100 cells counted for each genotype.

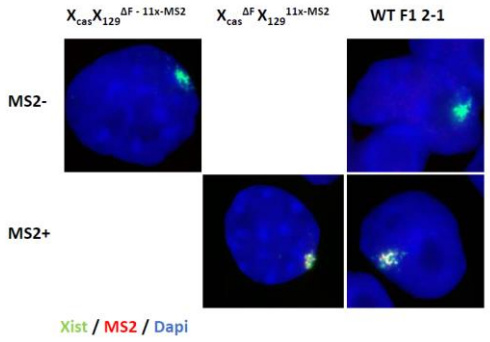
D) Percentage of Xist positive nuclei in the indicated mutant Xist female mESC that have an MS2+ Xist cloud or MS2- Xist cloud at day 6 of RA differentiation. 100 cells counted for each genotype.

E) Representative fluorescence in situ hybridization microscopy images of the indicated heterozygous mutant Xist female mESC at day 6 of RA differentiation. Xist signal is shown in green, MS2 signal in red, and DAPI in blue.

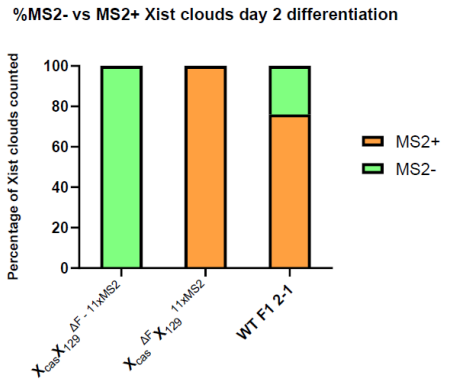
F) Representative fluorescence in situ hybridization microscopy images of Δ5' heterozygous mutant Xist female mESC and WT F1 2-1 female mESC at day 6 of RA differentiation. Xist signal is shown in green, MS2 signal in red, and DAPI in blue.

G) Percentage of Xist positive nuclei in Δ5' heterozygous mutant Xist female mESC and WT F1 2-1 female mESC that have an MS2+ Xist cloud or MS2- Xist cloud at day 6 of RA differentiation. 100 cells counted for each genotype.

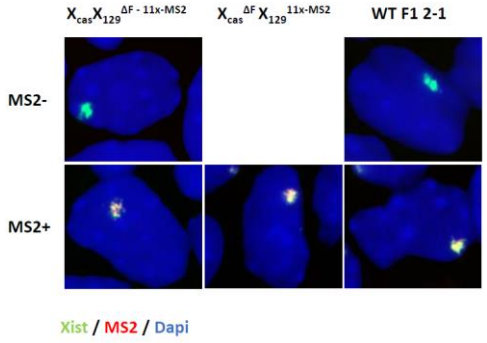
A


$X_{cas}X_{129}^{\Delta F - 11x\text{-}MS2}$
$X_{cas}^{\Delta F}X_{129}^{11x\text{-}MS2}$
WT F1 2-1
MS2-
MS2+
Xist / MS2 / Dapi

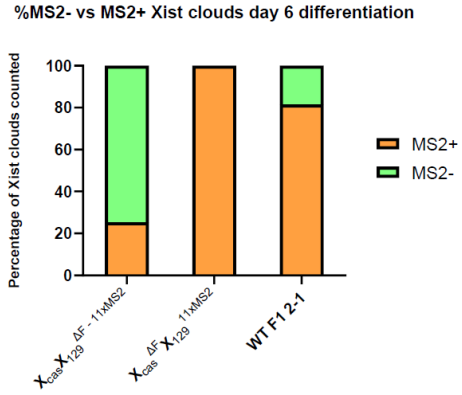
B


%MS2- vs MS2+ Xist clouds day 2 differentiation
Percentage of Xist clouds counted
100
80
60
40
20
0
MS2+
MS2-
$X_{cas}X_{129}^{\Delta F - 11xMS2}$
$X_{cas}^{\Delta F}X_{129}^{11xMS2}$
WT F1 2-1

C


$X_{cas}X_{129}^{\Delta F - 11x\text{-}MS2}$
$X_{cas}^{\Delta F}X_{129}^{11x\text{-}MS2}$
WT F1 2-1
MS2-
MS2+
Xist / MS2 / Dapi

D


%MS2- vs MS2+ Xist clouds day 6 differentiation
Percentage of Xist clouds counted
100
80
60
40
20
0
MS2+
MS2-
$X_{cas}X_{129}^{\Delta F - 11xMS2}$
$X_{cas}^{\Delta F}X_{129}^{11xMS2}$
WT F1 2-1

E

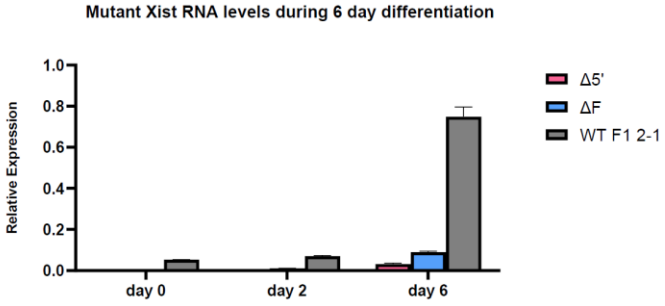

Mutant Xist RNA levels during 6 day differentiation
Relative Expression
1.0
0.8
0.6
0.4
0.2
0.0
day 0
day 2
day 6
Δ5'
ΔF
WT F1 2-1

Figure 2-2. Heterozygous deletion of Xist F-repeat results in skewing of XCI choice and reduction in mutant Xist RNA levels

A) Representative fluorescence in situ hybridization microscopy images of ΔF heterozygous mutant Xist female mESC and WT F1 2-1 mESC at day 6 of RA differentiation. Xist signal is shown in green, MS2 signal in red, and DAPI in blue.

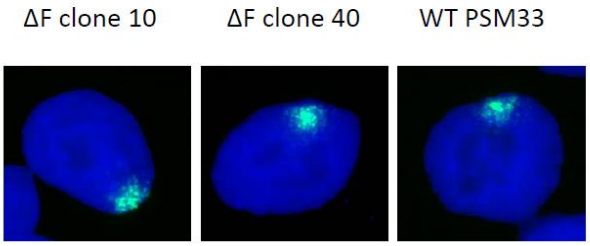
B) Percentage of Xist positive nuclei in ΔF heterozygous mutant Xist female mESC and WT F1 2-1 female mESC that have an MS2+ Xist cloud or MS2- Xist cloud at day 6 of RA differentiation. 100 cells counted for each genotype.

C) Representative fluorescence in situ hybridization microscopy images of ΔF heterozygous mutant Xist female mESC and WT F1 2-1 mESC at day 2 of RA differentiation. Xist signal is shown in green, MS2 signal in red, and DAPI in blue.

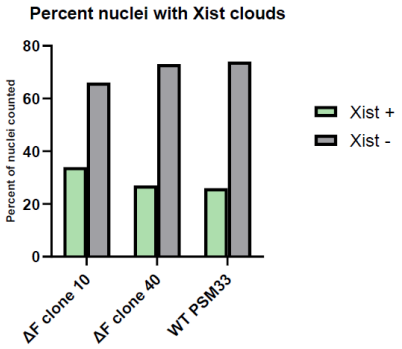
D) Percentage of Xist positive nuclei in ΔF heterozygous mutant Xist female mESC and WT F1 2-1 female mESC that have an MS2+ Xist cloud or MS2- Xist cloud at day 2 of RA differentiation. 100 cells counted for each genotype.

E) qPCR quantification of MS2+ Xist RNA levels of Δ5', ΔF, of WT F1 2-1 mESC at the indicated timepoint in RNA differentiation

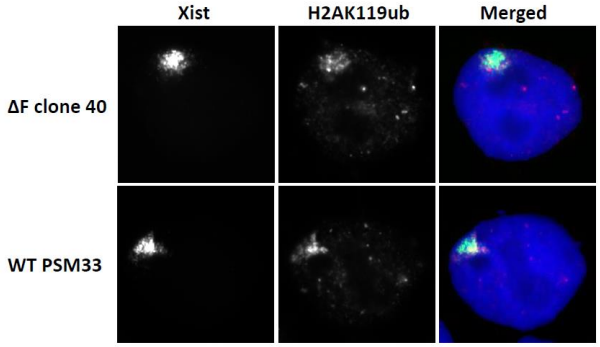
A


ΔF clone 10
ΔF clone 40
WT PSM33

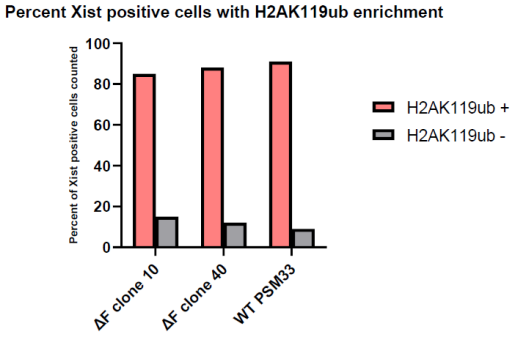
B


Percent nuclei with Xist clouds
Percent of nuclei counted
80
60
40
20
0
ΔF clone 10
ΔF clone 40
WT PSM33
Xist +
Xist -

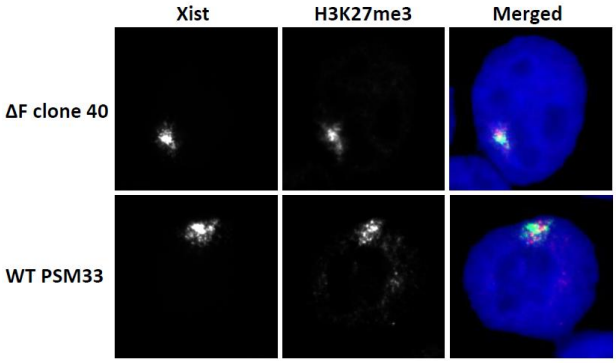
C


Xist
H2AK119ub
Merged
ΔF clone 40
WT PSM33

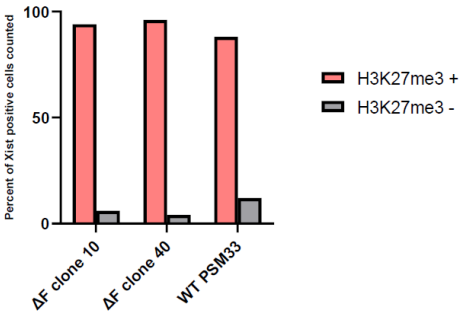
D


Percent Xist positive cells with H2AK119ub enrichment
Percent of Xist positive cells counted
100
80
60
40
20
0
ΔF clone 10
ΔF clone 40
WT PSM33
H2AK119ub +
H2AK119ub -

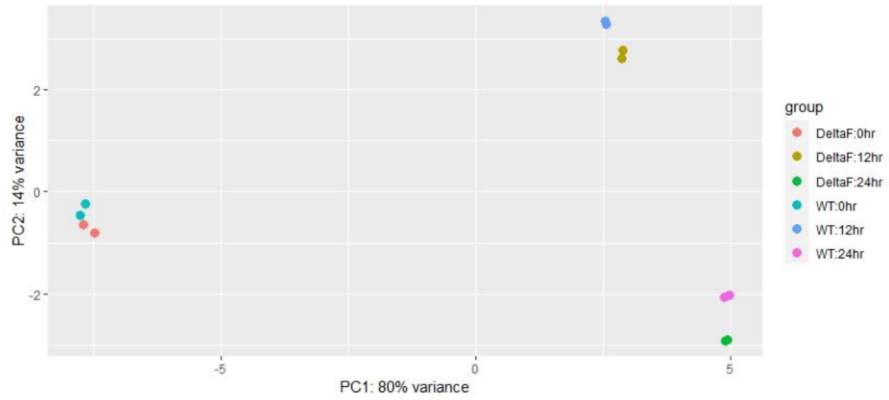
E


Xist
H3K27me3
Merged
ΔF clone 40
WT PSM33

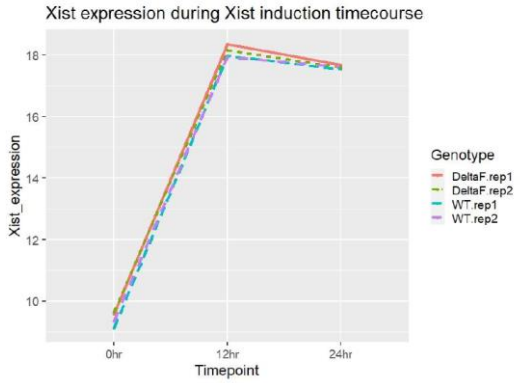
F


Percent Xist positive cells with H3K27me3 enrichment
Percent of Xist positive cells counted
100
50
0
ΔF clone 10
ΔF clone 40
WT PSM33
H3K27me3 +
H3K27me3 -

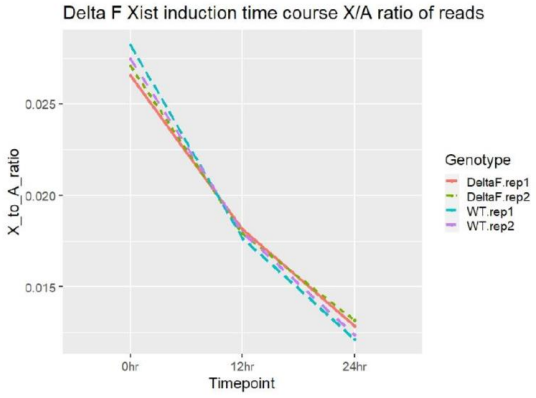
G


PC2: 14% variance
PC1: 80% variance
group
DeltaF:0hr
DeltaF:12hr
DeltaF:24hr
WT:0hr
WT:12hr
WT:24hr

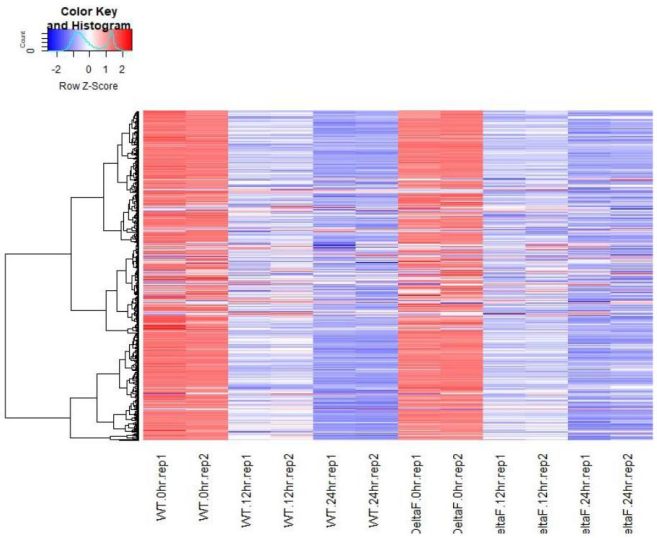
H


Xist expression during Xist induction timecourse
Xist_expression
Timepoint
0hr
12hr
24hr
Genotype
DeltaF.rep1
DeltaF.rep2
WT.rep1
WT.rep2

I


Delta F Xist induction time course X/A ratio of reads
X_to_A_ratio
Timepoint
0hr
12hr
24hr
Genotype
DeltaF.rep1
DeltaF.rep2
WT.rep1
WT.rep2

J


Color Key and Histogram
Row Z-Score
WT.0hr.rep1
WT.0hr.rep2
WT.12hr.rep1
WT.12hr.rep2
WT.24hr.rep1
WT.24hr.rep2

K

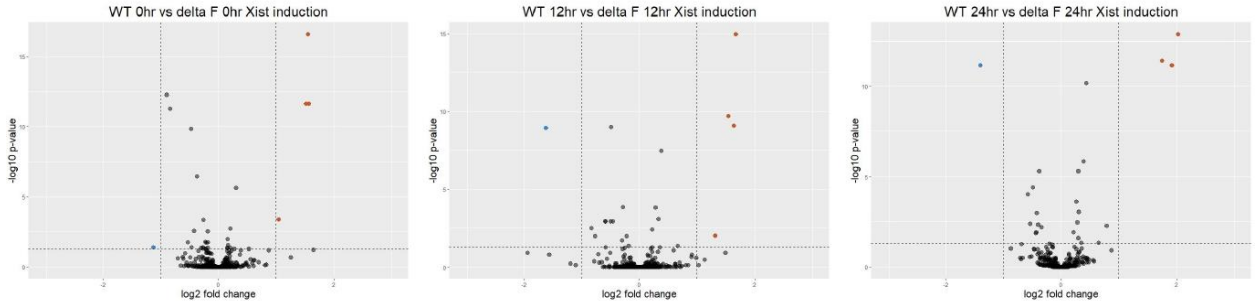


Figure 2-3. Xist F-repeat deletion in male mESC dox inducible model does not result in defects in Xist cloud formation or X-linked gene silencing

A) Representative fluorescence in situ hybridization microscopy images of ΔF mutant Xist male mESC at 24hrs dox induction. Xist signal is shown in green and DAPI in blue.

B) Percentage of nuclei that have Xist clouds in ΔF male mESC compared to WT male mESC after 24hr dox induction. 100 cells counted for each genotype.

C) Representative immunofluorescence combined with fluorescence in situ hybridization microscopy images of ΔF mutant Xist male mESC at 24hrs dox induction. Xist signal is shown in green, H2AK119ub in red, and DAPI in blue.

D) Percentage of Xist positive nuclei that have enrichment of H2AK119ub signal on the X chromosome in ΔF male mESC compared to WT male mESC after 24hr dox induction. 100 cells counted for each genotype.

E) Representative immunofluorescence combined with fluorescence in situ hybridization microscopy images of ΔF mutant Xist male mESC at 24hrs dox induction. Xist signal is shown in green, H3K27me3 in red, and DAPI in blue.

F) Percentage of Xist positive nuclei that have enrichment of H3K27me3 signal on the X chromosome in ΔF male mESC compared to WT male mESC after 24hr dox induction. 100 cells counted for each genotype.

G) Principal components analysis plot of ΔF mutant Xist male mESC and WT male mESC 24hr dox induction time course RNA-seq data using only X-linked genes.

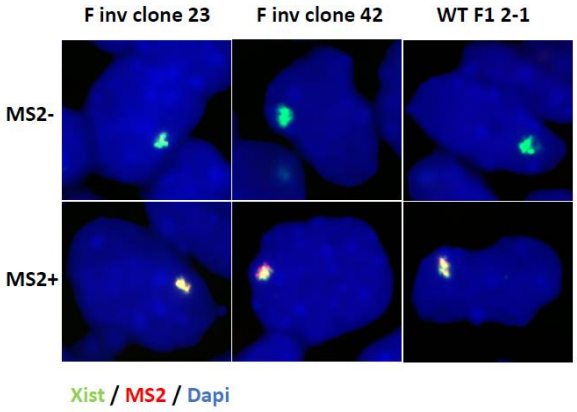
H) Xist expression during 24hr dox induction time course of ΔF mutant Xist male mESC and WT male mESC

I) Assessing X-linked gene silencing during 24hr dox induction time course of ΔF mutant Xist male mESC and WT male mESC by plotting RNA-seq reads from the X-chromosome divided by RNA-seq reads from autosomes

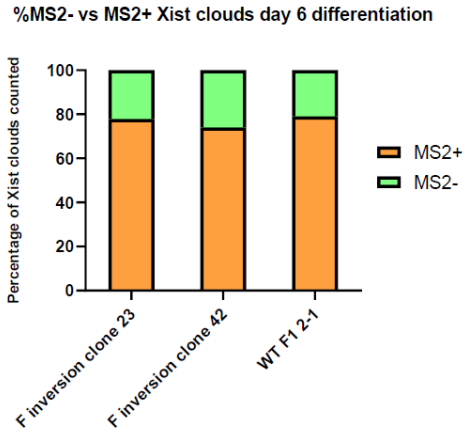
J) Heatmap representation of X-linked gene expression during 24hr dox induction time course of ΔF mutant Xist male mESC and WT male mESC

K) Differential gene expression analysis comparing X-linked gene expression between ΔF mutant Xist male mESC and WT male mESC at each indicated time point using DESeq2 package.

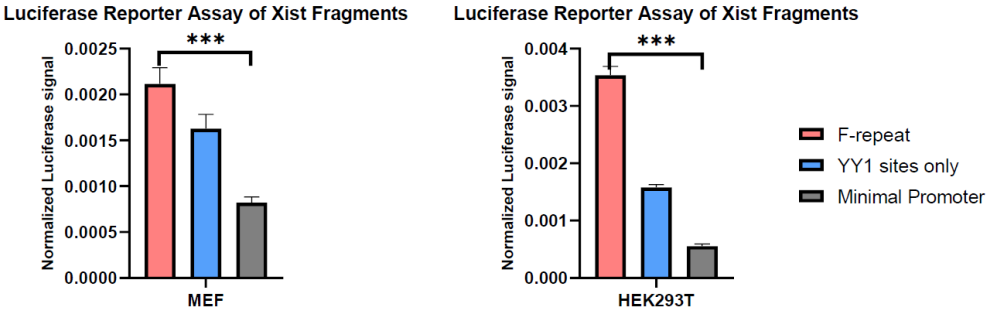
A


F inv clone 23
F inv clone 42
WT F1 2-1
MS2-
MS2+
Xist / MS2 / Dapi

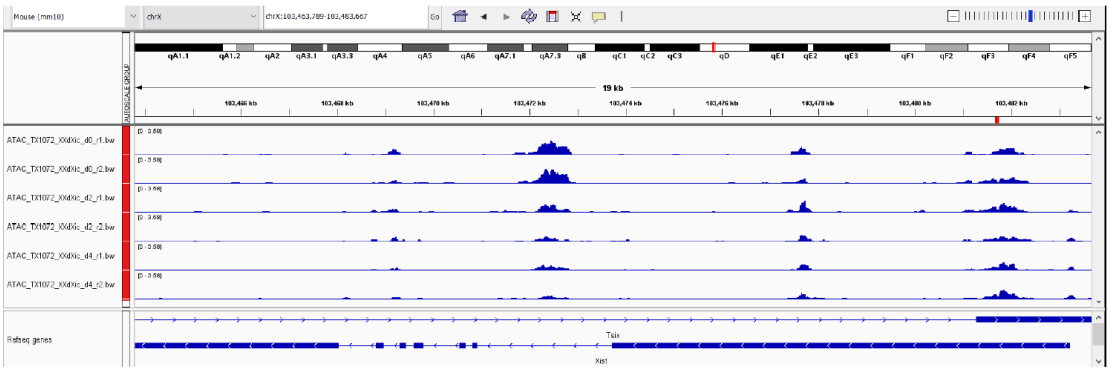
B


%MS2- vs MS2+ Xist clouds day 6 differentiation
Percentage of Xist clouds counted
100
80
60
40
20
0
MS2+
MS2-
F inversion clone 23
F inversion clone 42
WT F1 2-1

C


Luciferase Reporter Assay of Xist Fragments
Normalized Luciferase signal
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
***
MEF
Luciferase Reporter Assay of Xist Fragments
Normalized Luciferase signal
0.004
0.003
0.002
0.001
0.000
***
HEK293T
F-repeat
YY1 sites only
Minimal Promoter

D



E

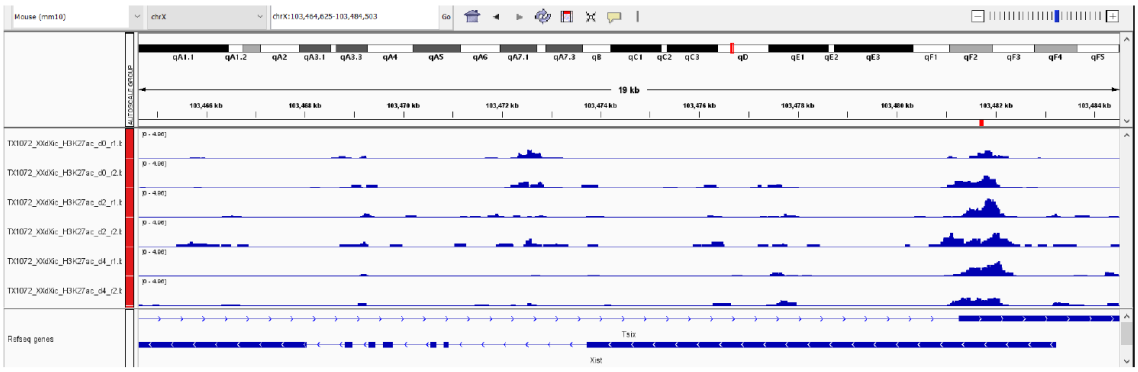


Figure 2-4. Xist F-repeat DNA element can enhance transcription

A) Representative fluorescence in situ hybridization microscopy images of F-repeat inversion heterozygous mutant Xist female mESC and WT F1 2-1 female mESC at day 6 of RA differentiation. Xist signal is shown in green, MS2 in red, and DAPI in blue.

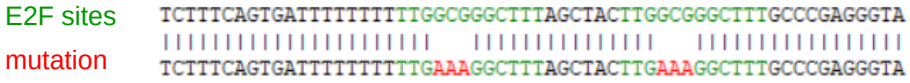
B) Percentage of Xist positive nuclei in F-repeat inversion heterozygous mutant Xist female mESC and WT F1 2-1 female mESC that have an MS2+ Xist cloud or MS2- Xist cloud at day 6 of RA differentiation. 100 cells counted for each genotype.

C) Luciferase assay showing quantification of firefly luciferase signal normalized to Renilla luciferase signal. Luciferase constructs were transfected into MEFs or HEK293T cells, and luciferase signal was measured 24hrs after transfection

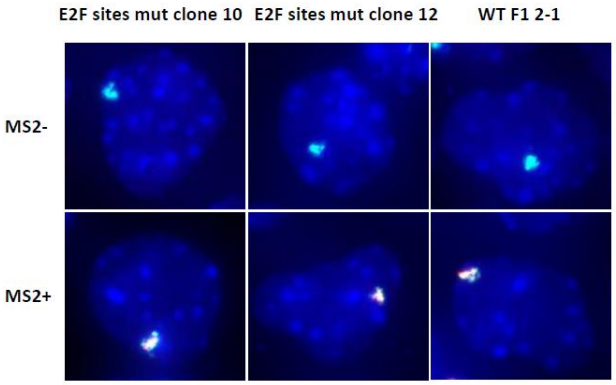
D) Genome browser tracks of ATAC-seq data using female mESC differentiation from Gjaltema et al. 2022. Xist F-repeat region is highlighted in red

E) Genome browser tracks of H3K27ac Cut&Tag data using female mESC differentiation from Gjaltema et al. 2022. Xist F-repeat region is highlighted in red

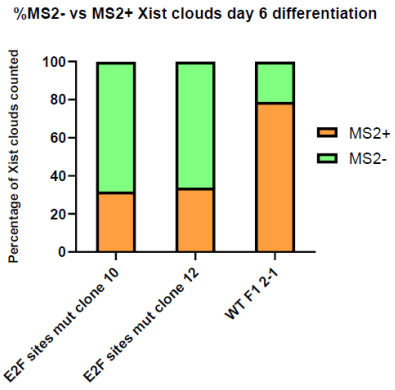
A



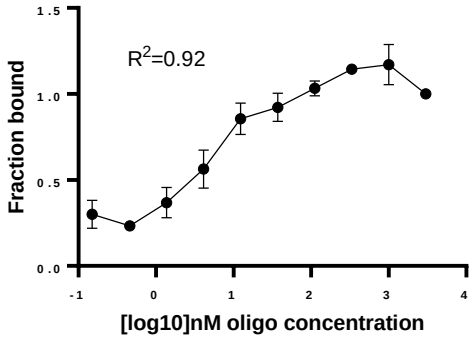
B



C



D



E

Candidate	F-repeat R^2	Scrambled R^2
XRCC5	0.99	0.79
XRCC6	0.96	0.74
CTCF	0.96	0.53
E2F3	0.92	N/A
TFAP2C	0.90	0.39
ZFP384	0.89	N/A
NFIC	0.86	0.46

F

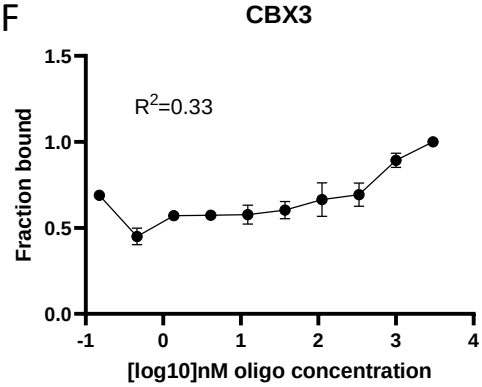


Figure 2-5. E2F sites within the Xist F-repeat plays a functional role in regulating Xist expression

A) Partial sequence of the Xist F-repeat is shown, with the two E2F sites highlighted in green. The minimal mutation introduced to abolish E2F transcription factor binding is highlighted in red.

B) Representative fluorescence in situ hybridization microscopy images of E2F sites heterozygous mutant Xist female mESC and WT F1 2-1 female mESC at day 6 of RA differentiation. Xist signal is shown in green, MS2 in red, and DAPI in blue.

C) Percentage of Xist positive nuclei in E2F sites heterozygous mutant Xist female mESC and WT F1 2-1 female mESC that have an MS2+ Xist cloud or MS2- Xist cloud at day 6 of RA differentiation. 100 cells counted for each genotype.

D) PAQMAN assay binding curve for F-repeat DNA oligonucleotide and transcription factor E2F3

E) List of protein interactors that have an R^2 value of >0.85 from PAQMAN assay performed using F-repeat DNA oligonucleotide. R^2 value for each interactor calculated when using scrambled oligonucleotide sequence for PAQMAN assay is also shown

F) PAQMAN assay binding curve for F-repeat DNA oligonucleotide and DNA binding protein CBX3

Methods

Cell culture

All mouse ES cell lines were cultured in knockout DMEM (Life Technologies) supplemented with 15% FBS (Omega), 2 mM l-glutamine (Life Technologies), 1× NEAA (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma), 1× penicillin/streptomycin (Life Technologies), and 1,000 U ml−1 murine LIF (homemade) on 0.3% gelatinized plates (porcine skin gelatin, Sigma) pre-plated with irradiated male DR4 feeders (homemade from day 14.5 embryos, with appropriate protocols in place ensuring the ethical treatment of animals, approved by the UCLA Institutional Animal Care and Use Committee, known as the Chancellor's Animal Research Committee (ARC), 2007-180-41). ES cells were maintained as small colonies and passaged with trypsin and single-cell dissociation at 80% confluency. Mycoplasma tests (Lonza) are routinely conducted on cells cultured in the laboratory. Additionally, DAPI staining of the cells used in the study did not indicate any mycoplasma contamination.

Female ES cell differentiation

Female wild-type F1 2-1 were trypsinized to single cells and counted. Cells were seeded in 2 ml of MEF medium (DMEM (Invitrogen) supplemented with 10% FBS (Omega), 2 mM l-glutamine (Life Technologies), 1× NEAA (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma) and 1× penicillin/streptomycin (Life Technologies)) at a density of 20,000–200,000 cells per 4 cm^2 (depending on the experiment) on tissue culture plates for RNA collection or onto 18 mm sterile glass coverslips for immunofluorescence (IF)/FISH experiments, both of which were pre-coated with sterile Geltrex (Thermo Fisher, diluted 1:400). At 24 h post-seeding, the culture medium was changed and supplemented

with 1 μM all-*trans* retinoic acid (Sigma), which was changed daily thereafter until the cells were collected for analysis.

Female MEF culture

Female MEFs were maintained in MEF medium (DMEM (Invitrogen) supplemented with 10% FBS (Omega), 2 mM l-glutamine (Life Technologies), 1× NEAA (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma) and 1× penicillin/streptomycin (Life Technologies))

Male ES cell culture

Male ES cells were maintained as described in the section 'Cell culture'. To express *Xist*, ES cells were trypsinized to single cells and counted. Cells were seeded in 2 ml of mouse embryonic cell media at a density of 20,000–200,000 cells per 4cm2 (depending on the experiment) on tissue culture plates for RNA collection or onto 18 mm sterile glass coverslips for IF/FISH experiments, both of which were pre-coated with sterile Geltrex (Thermo Fisher, diluted 1:400). For *Xist* expression, doxycycline (Sigma) was added to a final concentration of 2 μg ml^{-1} for 6–24 h, depending on the experiment.

RNA FISH

FISH against *Xist* RNA and MS2 tag was performed using DNA probes. In undifferentiated ES cells, the DNA probe against *Xist* additionally detects *Tsix*.

DNA probe preparation

DNA probes were synthesized using the CGH Bioprime Array Kit (Thermo Fisher) according to the manufacturer's instructions. In brief, a 40 μl solution containing 100 ng of template DNA was denatured in the presence of 1× random primers at 95 °C for 5 min and snap-cooled on ice. Five microlitres of nucleotide mix, 5 μl of 488, 555-, or 594- dUTP or dCTP chromatide fluorophore (Life Technologies) and 5U Klenow exo-enzyme were

then added and incubated in the dark at 37 °C for 6 h, after which an additional 5U of Klenow exo-enzyme was added. The reaction was incubated at 37 °C overnight, quenched with 10 µl stop solution, and then purified over a Chromatide-100 column or AMPure beads as described in the section 'RNA probe preparation'. The eluate was precipitated in the presence of 100 mg yeast tRNA (Life Technologies) and sodium acetate (Sigma). The final DNA probe mix was then prepared as in the section 'RNA probe preparation' to yield 400 µl of probe solution in formamide/hybridization buffer. The MS2 DNA template for DNA probe preparation was PCR-amplified from genomic DNA purified from wild-type F1 2-1 MS2*129* female ES cells. For *Xist*, the DNA probe was synthesized using a full-length mouse *Xist* cDNA plasmid (p15A-31-17.9kb *Xist*, unpublished).

RNA FISH procedure for epifluorescence microscopy

Culture medium was changed 10 min before cell collection to remove dead cells and stimulate transcription. Upon collection, culture medium was aspirated, and coverslips were gently rinsed twice with cold 1×PBS. Coverslips were then transferred to a new culture dish containing 1× PBS, which was then aspirated, and the cells were fixed in 4% paraformaldehyde (PFA) (Electron Microscopy Sciences) in 1× PBS for 10 min at room temperature (RT) under standard laboratory safety practices. After fixation, the cells were permeabilized in 0.5% Triton X-100 (Acros) in 1× PBS with 2 mM vanadyl ribonucleoside complex (NEB) for 10–20 min on ice. Coverslips were then stored in 70% ethanol at −20 °C for 1 h or until samples from all time points had been collected. Before hybridization with the probe, the coverslips with cells were brought back to 4 °C and serially dehydrated by 5-min incubations in ice-cold 80%, 95% and 100% ethanol. Coverslips were removed

from 100% ethanol and allowed to air dry before incubation with probe for 48 h at 37 °C in a sealed chamber humidified with 2× SSC/50% formamide. Coverslips were washed for 3 × 5 min in 50% formamide/2× SSC, 3 × 5 min in 2× SSC and 3 × 5 min in 1× SSC before mounting. A 1:10,000 dilution of DAPI (0.5 mg ml−1) was included in all penultimate 1× SSC washes. All washes were conducted at 42 °C, cells were protected from light. All procedures were performed, and used reagents disposed of, according to standard laboratory safety procedures. The *Xist* RNA FISH probe used in our study covers the ~17.9kb exonic regions of *Xist*. The MS2 tag is ~1.1kb long and the MS2 FISH probe was designed to cover the entirety of the tag. Differences in the length of sequence targeted by these probes made the *Xist* probe signal much brighter than the MS2 probe signal when visualized microscopically. Consequently, in our RNA FISH experiments, we used both probes to differentiate between the *cas* (detected by the *Xist* probe only) and the *129* allele (detected by the *Xist* and MS2 probes) in both wild-type and mutant cells.

Immunofluorescence staining

The cell culture medium was changed 10 min before collection. Upon collection, culture medium was aspirated, and coverslips were gently rinsed twice with cold 1× PBS. Coverslips were then transferred to a new culture dish containing cold 1× PBS. For CSK treatment, coverslips were gently treated with 1 ml (added dropwise) ice-cold CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM MgCl2, 10 mM PIPES pH 6.8) and incubated on ice for 30 s before aspiration. Coverslips were then similarly treated with 1 ml ice-cold CSK-Trt Buffer (CSK+0.5% Triton X-100) for 30 s, followed with a second ice-cold CSK treatment. Coverslips were then processed as described in Kraus et al.[24].

Immunofluorescence staining combined with RNA FISH

Where immunostaining and RNA FISH were combined, immunostaining preceded FISH. Combined staining for epifluorescence microscopy. The immunostaining protocol was followed as outlined above, but coverslips were not mounted. Instead, after the last round of washes (omitting DAPI in the penultimate wash), coverslips were re-fixed in 4% PFA in 1× PBS for 10 min at RT and then dehydrated through a 70–85–95–100% ice-cold ethanol series before overnight incubation with probe as described above in the section 'RNA FISH procedure for epifluorescence microscopy'.

Microscopy

Epifluorescence imaging

Cells with immunofluorescence and RNA FISH stainings were imaged using a Zeiss AxioImager M1 microscope with a 63× objective and acquired with AxioVision software. Epifluorescence images shown are sections and were analysed, merged and quantified using ImageJ or Adobe Photoshop.

Quantitative RT–PCR

For PCR with reverse transcription (RT–PCR), cells were collected in 1 ml Trizol (Thermo Fisher), after culture medium removal and washing with PBS. RNA was purified over RNAeasy columns (Qiagen). Total RNA (1 μg) was used in a reverse-transcription (RT) reaction with SuperScript III and an appropriate strand-specific reverse primer, according to the manufacturer's instructions (Thermo Fisher). One-twentieth of the RT reaction was used in a quantitative PCR reaction, using either 480 SYBR Green LightCycler PCR mix (Roche) or SYBR Green Master Mix (Applied Biosystems) and appropriate primers, in triplicate reactions. RT–qPCR experiments were normalized against *U6* snRNA.

Plasmid construction and cell line generation

CRISPR guide plasmid cloning

CRISPR guide sequences were synthesized as single stranded DNA from IDT with the appropriate overhangs for cloning into pSg2 plasmid that can express the guide RNA under a U6 promoter. Synthesized forward and reverse guide oligonucleotides were phosphorylated, annealed together, and ligated into BplI and NheI digested pSg2 plasmid.

CRISPR HDR template plasmid cloning

HDR template plasmids were created using pCR2.1-Puro vector. 1 kb upstream and 1 kb downstream homology arms were PCR-amplified from mouse genomic DNA using primers modified for In-Fusion cloning (Clontech) and Phusion polymerase (Thermo Fisher) according to the manufacturer's instructions. For F-repeat inversion and E2F sites mutation HDR templates, the mutations were encoded at the 3' end of the upstream homology arm and synthesized using Genewiz FragmentGene service. The upstream homology arm was integrated at the EcoRV site and the downstream homology arm at the KpnI site, using In-Fusion cloning reaction (Takara Bio), into a vector containing a floxed puromycin resistance cassette (PCR2.1-*loxP*-pGK-Puro-pA-*loxP*). Positive recombinants were identified by restriction digest and sanger sequencing.

CRISPR paired guide deletions of Xist domains in mESC

The ΔF-D, Δ5', Δ3', ΔB-C, ΔC, PBS 30-32, ΔD deletions were generated in female wild-type F1 2-1 MS2*129* ES cells derived from an F1 cross of mice from pure bred *129* and *castaneous* background, and then targeted to contain a 11× tandem repeat of the MS2 hairpin located 1.2 kb downstream of the E-repeat via homologous recombination. The wild-type F1 2-1 MS2*129* female ES cells also harbour an M2-reverse tetracycline

TransActivator (M2rtTA) cassette within the Rosa26 locus that confers neomcyin resistance on the cells. 125,000 cells were transfected with 1ug of Cas9 expressing plasmid, 750ng of each 5' and 3' CRISPR guide expressing plasmid, using Lipofectamine 3000 transfection reagent (Thermo Fisher) according to the manufacturer's protocol. 24hrs post transfection, the media was changed and cells were selected for 24hrs with 10ug/mL blasticidin. Selected cells were then trypsinized and seeded onto 10cm plates with irradiated DR4 feeders at various dilutions. One hundred clones were picked and expanded. PCR analysis and sanger sequencing of genomic DNA was performed to confirm deletions and determine which Xist allele the deletion is on. We ensured that all targeted cells maintained two X chromosomes throughout the targeting process.

CRISPR HDR generation of mutant Xist mESC

The ΔF, F-repeat inversion, and E2F sites mutation mESC were generated in female wild-type F1 2-1 MS2*129* ES cells derived from an F1 cross of mice from pure bred *129* and *castaneous* background, and then targeted to contain a 11× tandem repeat of the MS2 hairpin located 1.2 kb downstream of the E-repeat27 via homologous recombination. The wild-type F1 2-1 MS2*129* female ES cells also harbour an M2-reverse tetracycline TransActivator (M2rtTA) cassette within the Rosa26 locus that confers neomcyin resistance on the cells. 125,000 cells were transfected with 500ng of CRISPR guide expressing plasmid, 1ug of Cas9 expressing plasmid, and 1ug of linearized homology repair template plasmid using Lipofectamine 3000 transfection reagent (Thermo Fisher) according to the manufacturer's protocol. 24hrs post transfection cells were trypsinized and seeded onto 10cm plates with irradiated DR4 feeders at various dilutions. The cells were selected with 1ug/mL puromycin for seven days. One hundred clones were picked

and expanded. PCR analysis and sanger sequencing of genomic DNA was performed to confirm mutation and determine which Xist allele the mutation is on. Positive clones were expanded in culture, then transfected with a Cre-recombinase plasmid using Lipofectamine 3000 according to the manufacturer's protocol (Thermo Fisher), to delete the floxed puromycin resistance cassette. Transfected cells were serially diluted, 100 clones were picked, expanded and replica-plated for growth in the presence or absence of puromycin. Puromycin sensitive clones were expanded, and PCR analysis was performed to confirm the genotype. We ensured that all targeted cells maintained two X chromosomes throughout the targeting process.

Bulk RNA-seq

Cells were washed with 1XPBS twice, and RNA was harvested using Trizol reagent (Life Technologies). RNA was isolated using Qiagen RNeasy kit according to manufacturer's protocol, and RNA-seq libraries were prepared with TruSeq Stranded mRNA Library Prep Kit (Illumina). Libraries were sequenced single end 100bp on the Illumina NovaSeq S1 at a depth of about 30 million reads per library.

RNA-seq data analysis

RNA-seq reads were mapped using STAR alignment package with default parameters. Read counts for each gene were generated using featureCounts from the subread package. Only genes with at least 1 count per million in at least 2 samples were kept. Regularized log transformation (rlog) from DESeq2 package was used to normalize gene expression data before plotting as heatmap[15]. Differentially expressed genes between samples were called using the DESeq2 package[15].

Luciferase construct cloning

Xist F-repeat sequence or Xist YY1 sites sequence were synthesized as phosphorylated, double stranded DNA oligonucleotides with appropriate overhangs for In-fusion cloning (Takara Bio) into EcoRV linearized Promega pGL4.27[luc2P/minP/Hygro] vector.

Luciferase assay

MEFs or HEK293T cells were grown in white 96-well flat bottom tissue culture treated plates. Cells were transfected with either Xist F-repeat pGL4.27 firefly luciferase plasmid, Xist YY1 sites pGL4.27 firefly luciferase plasmid, or minimal promoter pGL4.27 firefly luciferase plasmid. All transfections also included co-transfection of pGL4.75[hRluc/CMV] renilla luciferase plasmid for normalization purposes. Each transfection was performed in triplicate wells. 24hrs post transfection, cells were lysed and prepared for luciferase signal detection using the Dual-Glo Luciferase Assay System (Promega) according to manufacturer's protocol. Firefly luciferase and renilla luciferase signals were detected using Promega plate reader. Firefly luciferase signal of each well was normalized to renilla luciferase signal, and normalized firefly luciferase signal from triplicate wells were averaged.

Nuclei extract preparation for PAQMAN assay

mESC were harvested using trypsin and washed once with 1XPBS. Cells were resuspended in 5 volumes of cold buffer A (10mM HEPES pH 7.9, 1.5mM MgCl2, 10mM KCl) and incubated on ice for 10 min for osmotic lysis. Cells were then pelleted and resuspended in 2 volumes of buffer A containing protease inhibitors and 0.15% NP40. Nuclei was released from lysed cells using a dounce homogenizer (30-40 strokes with a type B pestle (tight)). Nuclei were pelleted by centrifugation and washed once with 10 volumes of 1XPBS. To prepare nuclei lysate, nuclei were pelleted and resuspended in 2

volumes of buffer C (20mM HEPES pH7.9, 2mM MgCl2, 420mM NaCl, 20% glycerol, 0.2mM EDTA, 0.1% NP40, 0.5mM DTT, 1X protease inhibitors). Nuclei suspension was then incubated at 4 degrees C for 1hr on rotating wheel for lysis. Lysed nuclei suspension was then pelleted, and the supernatant (nuclear extract) was transferred into a 7000 MWCO Slide-A-Lyzer cassette for dialysis overnight in dialysis buffer (150mM NaCl, 50mM Tris, pH7.5, proteinase inhibitors). Dialyzed nuclear extract was stored at -80 with 10% glycerol.

DNA oligonucleotide preparation for PAQMAN Assay

DNA oligonucleotides were custom synthesized with forward strand biotinylated at the 5' end. Oligonucleotides were first diluted to 100 uM in TE buffer. Next, oligonucleotides were annealed with a 1.5X molar excess of the reverse strand in 2X Annealing Buffer (100mM NaCl, 20mM Tris pH8.0, 2mM EDTA) by first heating to 95 degrees C for 10 minutes, then slowly cooled to room temperature. 300ul of the following oligonucleotide concentrations were prepared in DNA binding buffer (1M NaCl, 10mM Tris pH8.0, 1mM EDTA, 0.05% NP40): 0.15nM, 0.46nM, 1.37nM, 4.1nM, 12.3nM, 37nM, 111.1nM, 333nM, 1000nM, 3000nM.

PAQMAN assay

Aliquot 200ul streptavidin sepharose beads for each pulldown. Centrifuge and remove supernatant, then add 150ul of corresponding diluted oligonucleotide. Incubate mixtures for overnight at 4 degrees on a rotator to allow binding of biotinylated oligonucleotides on streptavidin sepharose beads. After oligonucleotide immobilization, wash beads once with 200ul DNA binding buffer and twice with 200ul protein incubation buffer (150mM NaCl, 50mM Tris pH8.0, 0.25% NP40, 1mM TCEP, proteinase inhibitors). Prepare

nuclear extract solution by add 100ug nuclear extract to a final volume of 150ul in protein incubation buffer per pulldown. Apply 150ul of nuclear extract to each pulldown and incubate at 4 degrees for 4hrs on rotator. After incubation, remove unbound proteins by centrifugation and washing with 200ul washing buffer (150mM NaCl, 100mM TEAB). Repeat for 5 washes total. To elute, add 30ul of on-beads digestion buffer to each pulldown and incubate at room temperature in the dark for 30 mins. Add 1ul of 0.4ug/uL trypsin and 1ul of 0.1ug/uL Lys-C to each pulldown, and incubate at 37 degrees overnight on benchtop shaker. Isobaric labeling was performed using 10-plex TMT (Thermo Fisher). 0.8mg TMT reagent for each reporter mass was resuspended in 100ul anhydrous acetonitrile. 10ul of resuspended TMT reagent was added to each sample, and reactions were incubated at room temperature for 1hr in the dark. Reactions were quenched with 100mM Tris pH8.0 for 30 min. All pulldown samples (of the same bait oligonucleotide) were pooled and acidified with trifluoroacetic acid and desalted for mass spectrometry analysis using C18 StageTip method[25].

Supplemental Figures

Figure S1. Xist sequence with repeat-domains and MS2 tag location

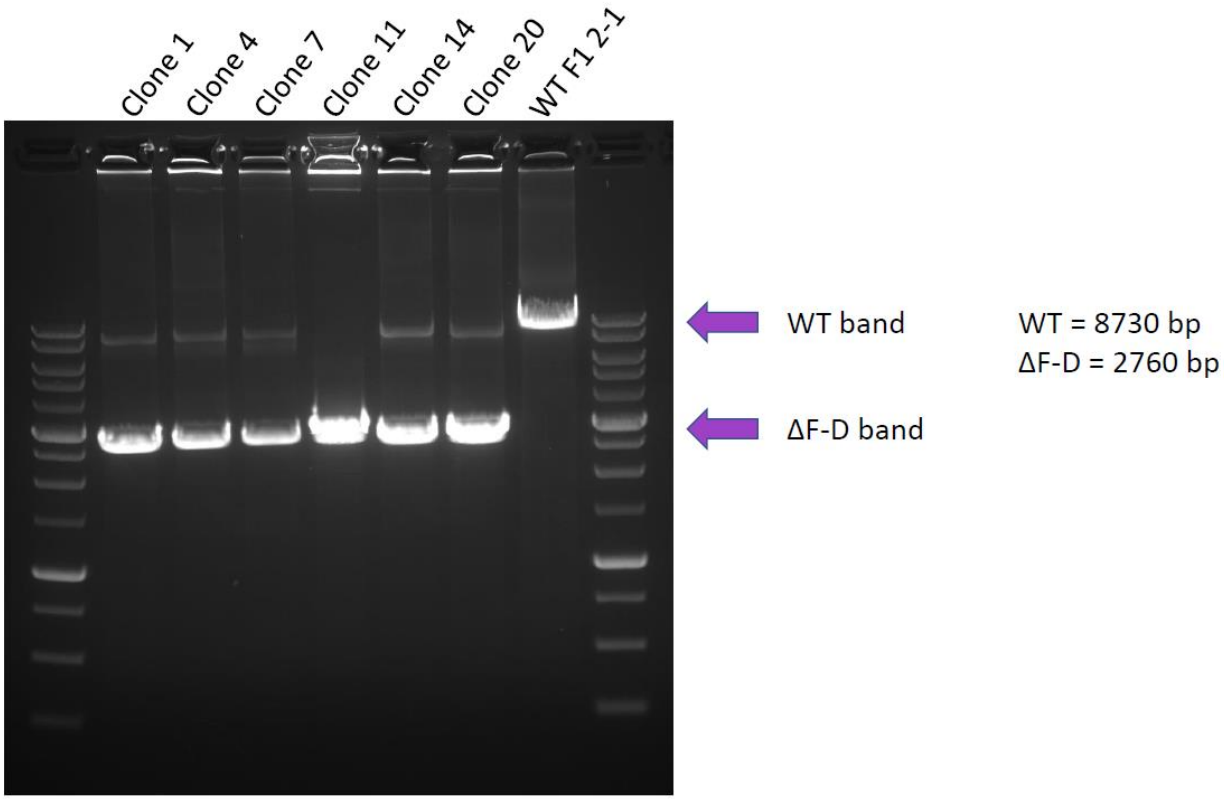


Figure S2 – PCR genotyping of of ΔF-D female mESC

Clones 7 and 20 were used for downstream experiments

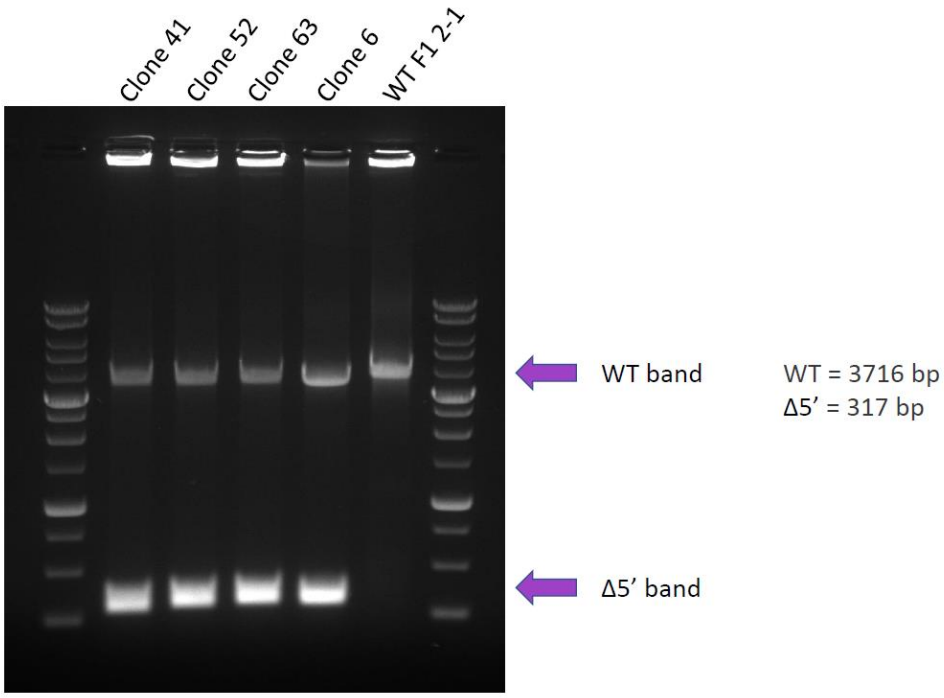


Figure S3 – PCR genotyping of of Δ5' female mESC

Clones 63 and 52 were used in downstream experiments

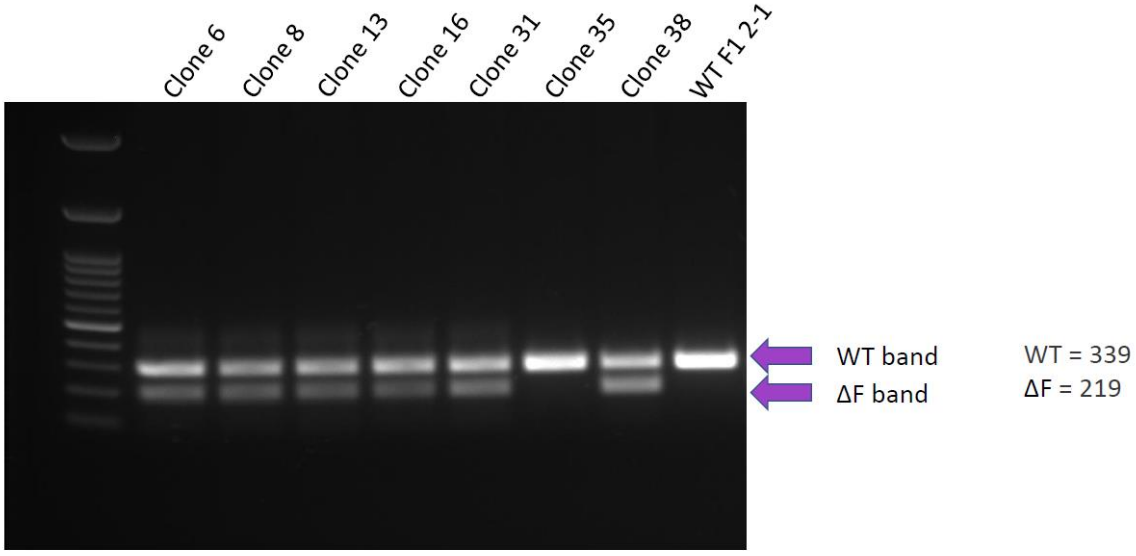


Figure S4 – PCR genotyping of of ΔF female mESC

Clones 31 and 38 were used in downstream experiments

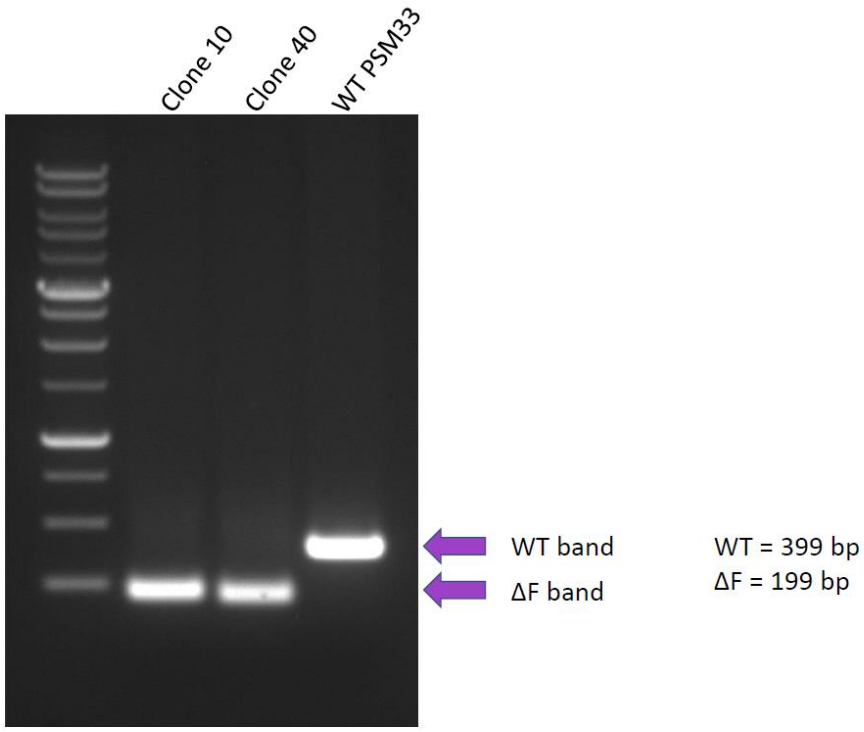


Figure S5 – PCR genotyping of of ΔF male mESC

Clones 10 and 40 were used in downstream experiments

Acknowledgements

We thank members of the Plath for discussions and reading of the manuscript. A.C was supported by the UCLA Whitcome Fellowship and the NIH F31 Ruth L. Kirschstein National Research Service Award Individual Predoctoral Fellowship; K.P. by Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research (BSCRC) at UCLA, the David Geffen School of Medicine at UCLA, and the Jonnson Comprehensive Cancer Center at UCLA, the NIH (R01 GM115233), and a Faculty Scholar grant from the Howard Hughes Medical Institute; W.D by the UCLA Whitcome Fellowship; A.P.-J. was supported by postdoctoral fellowships from the Helen Hay Whitney Foundation and NIH (F32 GM103139)

Author contributions

K.P., A.C., A.P.J. conceptualized the project and A.C. performed the experiments unless stated otherwise. S.Y., E.K., and W.M. helped to create ES cell deletion lines. A.P.J helped with Xist FISH staining's. W.D., J.S., J.W., helped with performing PAQMAN assay and data analysis. A.C. and K.P. administered the project and A.C. and K.P. wrote the manuscript.

References

1. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145-66. PMCID: PMC3858397.

2. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011 Aug 28;477(7364):295-300. PMCID:PMC3175327.

3. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. Nature. 1996 Jan 11;379(6561):131-7. PMID:8538762.

4. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X-chromosome. Science. 2013 Aug 16;341(6147):1237973. PMCID:PMC3778663.

5. Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome Inactivation. Annu Rev Genet. 2002 Jun 1136:233-78. PMID:12429693

6. Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet. 2002 Feb;30(2):167-74. PMID:11780141.

7. Sarma K, Levasseur P, Aristarkhov A, Lee JT. Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. Proc Natl Acad Sci USA. 2010 Dec 21;107(51):22196-201. PMCID:PMC3009817.

8. Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. The human Xist gene: analysis of a 17kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell. 1992 Oct 30;71(3)527-42. PMID:1423611.

9. Chen CK, Blanco M, Jackson C, Aznauryan E, Ollikainen N, Surka C, Chow A, Cerase A, McDonel P, Guttman M. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science. 2016 Oct 28;354(6311):468-472. PMID: 27492478.

10. McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature. 2015 May 14;521(7551):232-6. PMCID: PMC4516396.

11. Patil DP, Chen CK, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature. 2016 Sep 15;537(7620):369-373. PMCID:PMC5509218.

12. Pintacuda G, Wei G, Roustan C, Kirmizitas BA, Solcan N, Cerase A, Castello A, Mohammed S, Moindrot B, Nesterova TB, Brockdorff N. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol Cell. 2017 Dec 7;68(5):955-969.e10. PMCID:PMC5735038.

13. Colognori D, Sunwoo H, Kriz AJ, Wang CY, Lee JT. Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for

Spreading along the Inactive X. Mol Cell. 2019 Apr 4;74(1):101-117. PMCID: PMC6469964.

14. Wutz A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet. 2011 Jul 18;12(8):542-53. PMID: 21765457.

15. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. PMCID: PMC4302049.

16. Makhlouf M, Ouimette JF, Oldfield A, Navarro P, Neuillet D, Rougeulle C. A prominent and conserved role for YY1 in Xist transcriptional activation. Nat Commun. 2014 Sep 11;5:4878. PMCID: PMC4172967.

17. Gjaltema RAF, Schwämmle T, Kautz P, Robson M, Schöpflin R, Ravid Lustig L, Brandenburg L, Dunkel I, Vechiatto C, Ntini E, Mutzel V, Schmiedel V, Marsico A, Mundlos S, Schulz EG. Distal and proximal cis-regulatory elements sense X chromosome dosage and developmental state at the Xist locus. Mol Cell. 2022 Jan 6;82(1):190-208. PMID: 34932975.

18. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019 Apr 29;10(1):1930. PMCID: PMC6488672.

19. Zhu L, Zhu L, Xie E, Chang LS. Differential roles of two tandem E2F sites in repression of the human p107 promoter by retinoblastoma and p107 proteins. Mol Cell Biol. 1995 Jul;15(7):3552-62. PMCID: PMC230592.

20. Makowski MM, Gräwe C, Foster BM, Nguyen NV, Bartke T, Vermeulen M. Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry. Nat Commun. 2018 Apr 25;9(1):1653. PMCID: PMC5916898.
21. Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. Cell. 2010 Oct 29;143(3):390-403. PMCID: PMC2994261.
22. Chureau C, Chantalat S, Romito A, Galvani A, Duret L, Avner P, Rougeulle C. Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. Hum Mol Genet. 2011 Feb 15;20(4):705-18. PMID: 21118898.
23. Enervald E, Powell LM, Boteva L, Foti R, Blanes Ruiz N, Kibar G, Piszczek A, Cavaleri F, Vingron M, Cerase A, Buonomo SBC. RIF1 and KAP1 differentially regulate the choice of inactive versus active X chromosomes. EMBO J. 2021 Dec 15;40(24):e105862. PMCID: PMC8672179.
24. Kraus F, Miron E, Demmerle J, Chitiashvili T, Budco A, Alle Q, Matsuda A, Leonhardt H, Schermelleh L, Markaki Y. Quantitative 3D structured illumination microscopy of nuclear structures. Nat Protoc. 2017 May;12(5):1011-1028. doi: 10.1038/nprot.2017.020. Epub 2017 Apr 13. PMID: 28406495.
25. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2007;2(8):1896-906. PMID: 17703201.

Chapter 3

Research Project

A protein assembly mediates Xist localization and gene silencing

This chapter is adapted from the following publication with minor changes

Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Mancia Leon WR, Damianov A, Chronis C, Papp B, Chen CK, McKee R, Wang XJ, Chau A, Sabri S, Leonhardt H, Zheng S, Guttman M, Black DL, Plath K. A protein assembly mediates Xist localization and gene silencing. Nature. 2020 Nov;587(7832):145-151. PubMED PMID: 32908311.

Abstract

Nuclear compartments have diverse roles in regulating gene expression, yet the molecular forces and components that drive compartment formation remain largely unclear1. The long non-coding RNA *Xist* establishes an intra-chromosomal compartment by localizing at a high concentration in a territory spatially close to its transcription locus2 and binding diverse proteins3–5 to achieve X-chromosome inactivation (XCI)6,7. The XCI process therefore serves as a paradigm for understanding how RNA-mediated recruitment of various proteins induces a functional compartment. The properties of the inactive X (Xi)-compartment are known to change over time, because after initial *Xist* spreading and transcriptional shutoff a state is reached in which gene silencing remains stable even if *Xist* is turned off8. Here we show that the *Xist* RNA-binding proteins PTBP19, MATR310, TDP-4311 and CELF112 assemble on the multivalent E-repeat element of *Xist*7 and, via self-aggregation and heterotypic protein–protein interactions, form a condensate1 in the Xi. This condensate is required for gene silencing and for the anchoring of *Xist* to the Xi territory, and can be sustained in the absence of *Xist*. Notably, these E-repeat-binding proteins become essential coincident with transition to the *Xist*-independent XCI phase8, indicating that the condensate seeded by the E-repeat underlies the developmental switch from *Xist*-dependence to *Xist*-independence. Taken together, our data show that *Xist* forms the Xi compartment by seeding a heteromeric condensate that consists of ubiquitous RNA-binding proteins, revealing an unanticipated mechanism for heritable gene silencing.

Results

Although many *Xist*-interacting proteins have a defined function during the initiation of XCI3,5,6,13,14, the induction of X-linked gene silencing is largely unaffected when the *Xist*-interacting RNA-binding proteins (RBPs) PTBP1, MATR3, TDP-43 or CELF1 are depleted (Extended Data Fig. 1a–c), raising the question of what role(s) these proteins have in XCI (Supplementary Note 1). Notably, in addition to their known functions in RNA processing9–12, these RBPs can form higher-order assemblies—particularly when concentrated by RNAs containing multivalent protein-binding sites15–17. Because *Xist* contains several highly repetitive sequences7, we explored interactions between *Xist*, PTBP1, MATR3, CELF1 and TDP-43 that might create a higher-order assembly within the Xi and thereby contribute to the formation of the Xi compartment.

We first examined whether the depletion of PTBP1, MATR3, CELF1 or TDP-43 affects *Xist* localization. Short interfering RNA (siRNA)- mediated knockdown of each factor during XCI initiation in female differentiating embryonic stem cells (ES cells) revealed considerable nuclear dispersal of *Xist* and defects in the *Xist*-dependent accumulation of the Xi mark H3K27me3 (histone 3 lysine 27 trimethylation)18,19, with only small changes in *Xist* transcript or splicing levels (Extended Data Figs. 1d, e, 2a–f). PTBP1 knockdown in ES cells expressing *Xist* from an inducible cDNA transgene lacking introns resulted in similar dispersal of the *Xist* signal in RNA fluorescence in situ hybridization (FISH) experiments (Extended Data Fig. 2g). These findings demonstrate that these four RBPs mediate *Xist* localization on the forming Xi, independently of their RNA-processing activities.

To determine where on *Xist* these factors bind, we performed crosslinking immunoprecipitation followed by sequencing (CLIP–seq) analysis during XCI initiation. We identified a marked accumulation of PTBP1, MATR3 and CELF1 reads over the E-repeat of *Xist*, which comprises more than fifty elements that are rich in C, U and G nucleotides, and that are predicted to serve as PTBP1-, MATR3- and CELF1-binding sites20–22 (Extended Data Fig. 3a–c). We confirmed homomeric binding of recombinant (r)PTBP1 to the E-repeat RNA using an electrophoretic mobility shift assay (Extended Data Fig. 3d). Chromatin immunoprecipitation followed by sequencing (ChIP–seq) analysis of PTBP1 revealed a peak that is primarily located over the genomic E-repeat region and appears upon induction of *Xist* expression (Extended Data Fig. 3a), indicating that PTBP1 engages *Xist* co-transcriptionally. The *Xist* CLIP–seq profiles of PTBP1 and PTBP2 (the neural homologue of PTBP1) in differentiated cells were markedly similar to that of PTBP1 during XCI initiation, and TDP-43 in the embryonic mouse brain displayed strongest binding at the 3′ end of the E-repeat, where multiple (GU)*n* tracts presumably serve as binding motifs11 (Extended Data Fig. 3a, c). Together, these data show that the E-repeat serves as a multivalent binding platform for PTBP1, MATR3, CELF1 and TDP-43; that binding of TDP-43 and PTBP1 to the E-repeat persists after XCI initiation is complete; and that members of the same protein family can replace PTBP1 on *Xist*.

Next, we asked whether recruitment of PTBP1, MATR3, CELF1 or TDP-43 by *Xist* could be detected by microscopy within the Xi during XCI initiation and after transition to the *Xist*-independent phase of XCI after day 3 of differentiation8. We observed an accumulation of CELF1 on the Xi, which increases in intensity from day 3 to day 7 of

female ES cell differentiation, and noted a mesh-like pattern of PTBP1 localization within the Xi territory of some cells at day 7 of differentiation (Extended Data Fig. 4a–e). Although MATR3, TDP-43 and PTBP1 did not enrich in the Xi in most cells, they were not depleted (Extended Data Fig. 4e–h, Supplementary Note 2). We therefore conclude that PTBP1, MATR3 and TDP-43 are present—and CELF1 gradually concentrates—within the Xi-territory; findings that are consistent with the time-dependent formation of a spatially concentrated protein assembly.

If PTBP1, MATR3, CELF1 and TDP-43 control the accumulation of *Xist* on the Xi, loss of the E-repeat should disrupt XCI by reducing *Xist* enrichment within the X-chromosome territory. In support of this hypothesis, it has been shown that *Xist* exon 7—which contains the E-repeat—is required for persistent localization of *Xist* on the Xi in differentiating ES cells23. We tested this possibility by deleting the E-repeat within *Xist* on the 129 allele—which also contains 11 copies of an MS2-RNA tag within *Xist*—in a polymorphic 129 × Cas female mouse ES cell line, yielding the *Xist*ΔE,MS2(129)/WT(Cas) genotype (ΔE ES cells) (Fig. 1a, Extended Data Fig. 5). RNA FISH experiments revealed that the number of cells containing an *Xist*-coated X chromosome increased gradually until differentiation day 4 in both wild-type and ΔE cells (Fig. 1b, c). The proportion of ΔE cells with an *Xist* enrichment then declined compared to wild-type cells, with a significant reduction of approximately 50% by day 7 (Fig. 1c). This reduction was specific to the *Xist*ΔE,MS2(129) allele (hereafter denoted *Xist*ΔE)—as revealed by RNA FISH experiments against the MS2 tag (Fig. 1d)—and there was no significant difference in the abundance or the half-life of the *Xist*MS2 allele in ΔE cells

compared to the wild type (Extended Data Fig. 6a, b). RNA FISH experiments against an intronic *Xist* sequence labelled the nascent transcription

site and not the *Xist* cloud (Extended Data Fig. 6c), indicating that the *Xist*MS2 RNA that coats the Xi in wild-type and ΔE cells is processed. The loss of the *Xist* accumulation in ΔE cells over the X territory is therefore not a consequence of decreased *Xist* abundance, splicing defects or reduced RNA stability.

A closer inspection of *Xist* localization at differentiation day 3 showed that *Xist*ΔE enriched over the X chromosome, with aggregation measurements (see Methods) revealing only a modest defect in the localization of *Xist*ΔE compared to wild-type *Xist* (Fig. 1e, f). RNA antisense purification followed by sequencing (RAP–seq)2 revealed highly correlated patterns of *Xist* association across the X chromosome for wild-type *Xist* and *Xist*ΔE (Fig. 1g), indicating that the E-repeat is not involved in the initial transfer of *Xist* across the X chromosome. However, at day 7, we found significant dispersal of *Xist*ΔE within the nucleus compared to the wild type, often localizing to the nuclear lamina (Fig. 1f, h, Extended Data Fig. 6d–j). Super-resolution three-dimensional structured illumination microscopy (3D-SIM) imaging additionally revealed a significant increase in the number of individual *Xist*MS2 foci in ΔE cells compared to wild-type cells at differentiation day 7; a difference that was not seen on day 3 (Fig. 1i). As this increase occurred without an increase in the number of *Xist*ΔE transcripts compared to wild-type *Xist* transcripts (Extended Data Fig. 6a), these data support a model in which the E-repeat is required for the integration of multiple *Xist* transcripts into individual *Xist* foci and for stabilizing these foci within the X-chromosome compartment.

Consistent with defects in *Xist* localization, we observed lower H3K27me3 enrichment and reduced chromatin compaction over the Xi territory at differentiation day 7 in ΔE cells, despite normal establishment at day 3 (Extended Data Fig. 7a–e). The *Xist*ΔE localization phenotype arises as the enrichment of the PRC2 complex on the Xi decreases18,19 (Extended Data Fig. 7f), suggesting that it is associated with a reorganization of the X-chromosome compartment (Supplementary Note 3). Together, these results reveal a transition in the mechanisms that enrich *Xist* on the X chromosome during XCI initiation—switching from a largely E-repeat-independent phase to an E-repeat-dependent phase.

Because the control of *Xist* localization switches upon transition to the *Xist*-independent-phase of XCI initiation8, we addressed whether X-linked gene silencing was affected by loss of the E-repeat. We examined nascent transcripts from five X-linked genes that are subject to XCI: *Gpc4*, *Rnf12* (also known as *Rlim*), *Mecp2*, *Chic1* and *Atrx* (Fig. 2a, b). We observed little difference in the extent of gene silencing between wild-type and ΔE cells early during differentiation (Fig. 2c, d, Extended Data Fig. 8). However, at later stages of differentiation (days 4–7), cells expressing *Xist*ΔE failed to maintain silencing of these five genes (Fig. 2c, d, Extended Data Fig. 8). Moreover, RNA Pol II—which was excluded from the *Xist*ΔE-marked territory during early differentiation—intermingled with the *Xist*ΔE foci at later times (Fig. 2e, Extended Data Fig. 7g–i). The E-repeat is therefore essential for sustaining *Xist* coating, silencing of X-linked genes and exclusion of RNA Pol II beyond the initial wave of transcriptional shutoff. Thus, the *Xist*-independent state of XCI initiation8 is not established in the absence of the E-repeat,

demonstrating that the E-repeat is involved in generation of the epigenetic memory for gene silencing by *Xist*.

To evaluate whether there is a causal relationship between the E-repeat-binding RBPs, *Xist* localization and gene silencing, we synthetically fused PTBP1, MATR3, CELF1 or TDP-43 to the MS2-coat protein (MCP) in order to recruit these proteins to *Xist*ΔE via the 11×MS2-tag (Fig. 3a, Extended Data Fig. 9a–d). Continued expression of the MCP–PTBP1, MCP–MATR3, MCP–TDP-43 or MCP–CELF1 fusion proteins during differentiation in ΔE cells rescued *Xist* localization, silencing of *Gpc4* and *Atrx*, and H3K27me3 enrichment on the *Xist*ΔE,MS2(129)-associated X chromosome at differentiation day 7 (Fig. 3b–e, Extended Data Fig. 9e, f). These data demonstrate that the E-repeat controls *Xist* localization, gene silencing and heterochromatin formation through interaction with the proteins PTBP1, MATR3, TDP-43 and CELF1.

Next, we addressed whether PTBP1, MATR3, TDP-43 and CELF1 act together to control these processes. Making use of a known direct interaction between PTBP1 and MATR320, we found that MCP–MATR3 harbouring a mutant PTBP1–RRM interaction (PRI; where RRM is RNA recognition motif) sequence (denoted MATR3(mutPRI))20 partially rescued H3K27me3 enrichment, but was unable to rescue the *Xist* localization and gene silencing defects observed upon loss of the E-repeat (Fig. 3b, d, e, Extended Data Fig. 9c–f). Similar results were observed with the converse mutation in PTBP1, a tyrosine-to-glutamine substitution at position 247 (Y247Q)20 that prevents the interaction of PTBP1 with MATR3 (Fig. 3b–e, Extended Data Fig. 9c–e). These findings are supported by co-immunoprecipitation experiments demonstrating that PTBP1, MATR3, CELF1 and TDP-43 co-precipitate one another in the presence of RNA, whereas only

PTBP1 and MATR3 robustly interact after Rnase treatment (Fig. 3f, Extended Data Fig. 9g). These data indicate that a specific direct interaction between PTBP1 and MATR3 is critical for XCI, and consequently show that these proteins act non-redundantly in the XCI process. Furthermore, the finding that CELF1 enriches on the Xi in ΔE rescue cells that express MCP–PTBP1, MCP–MATR3 or MCP–TDP-43 (Fig. 3g, h) indicates that each of these RBPs can initiate the formation of a heteromeric protein assembly within the Xi.

The protein CIZ1 was previously suggested to anchor *Xist* to chromatin through the E-repeat24,25. Although PTBP1 and MATR3 can interact with CIZ1, the expression of MCP–CIZ1 in ΔE cells did not rescue *Xist* cloud formation or X-linked gene silencing (Fig. 3f, Extended Data Figs. 9g, 10a–e). Moreover, the Xi accumulation of CIZ1 that was observed in wild-type cells was not detected in ΔE cells expressing MCP–PTBP1, MCP–MATR3 or MCP–TDP-43 (Extended Data Fig. 10f, Supplementary Note 4). This indicates that the rescue of *Xist*ΔE phenotypes by PTBP1, MATR3, TDP-43 and CELF1 is independent of CIZ1 and that distinct functional complexes assemble on the E-repeat. A bivalent MCP–GFP–MCP fusion protein was also unable to rescue the *Xist* localization and silencing defects in ΔE cells (Extended Data Fig. 10g–k)—consistent with linkages formed by the four factors not simply tethering *Xist* transcripts together.

We next aimed to define additional specific activities conferred by the recruited proteins that could facilitate the compartmentalization of *Xist* and downstream events in XCI. By using a mutated version of MATR3 lacking both zinc finger domains (MATR3(ΔZfn)), we found that rescue of the *Xist*ΔE phenotypes by MATR3 is independent of its zinc fingers (Fig. 3b, d, e, Extended Data Fig. 9c–f). We also noted that the expression of PTBP1 lacking RRMs 3 and 4 (denoted PTBP1(ΔC)) in ΔE cells rescued

defects resulting from loss of the E-repeat; however, closer inspection of the *Xist*ΔE clouds binding MCP–PTBP1(ΔC) revealed dispersed *Xist* foci within the nucleus (Fig. 3b–e, Extended Data Fig. 9c–e). This finding implicates binding valency as a functional parameter of the PTBP1–*Xist* assembly.

The formation of an *Xist* territory containing PTBP1 was of interest given that PTBP1 can undergo liquid–liquid de-mixing in vitro when incubated at high concentrations with a binding RNA15. We therefore asked whether rPTBP1 forms liquid droplets upon interaction with the *Xist* E-repeat. The addition of 3.2 μM E-repeat RNA to 60 μM rPTBP1 produced aggregate-like assemblies15,16, whereas the addition of lower concentrations of RNA (0.1–0.5 μM) resulted in droplets that resembled phase-separated liquids and could fuse with each other (Fig. 4a, b, Extended Data Fig. 11a, b, Supplementary Note 5). By contrast, smaller droplets were produced by a control RNA (containing five short CU tracts that could bind PTBP1), with no observed aggregation at the highest RNA concentration (Fig. 4a). These findings indicate that the multivalent binding of PTBP1 to the E-repeat strongly promotes its condensation. A lack of droplet formation at near-physiological concentrations of rPTBP1 (20–40 μM) suggested that additional proteins promote the E-repeat-induced condensation of PTBP1 in vivo (Fig. 4c, Extended Data Fig. 11c), which is consistent with the interdependence of the functions of MATR3, PTBP1, CELF1 and TDP-43, as described above. We tested this idea by adding 20 μM rCELF1 to solutions containing 0.5 μM E-repeat RNA and varying concentrations of rPTBP1. Whereas aggregates formed at high rPTBP1 concentrations, lowering the concentration resulted in decreased aggregate sizes until, at 20 μM, they

resolved into small spherical structures (Extended Data Fig. 11d, e). These observations are consistent with the formation of a higher-order protein condensate in the Xi that forms via the multivalent binding of several RBPs to the E-repeat, and suggest that the involvement of multiple RBPs lowers the concentration of each factor that is required for condensate induction.

The in vitro data suggested that the self-assembly of E-repeat-interacting proteins is required for their function in XCI. To explore this idea further, we assessed whether the self-assembly of TDP-43 affected XCI. TDP-43 forms higher-order complexes that undergo liquid–liquid phase separation, and this activity is reduced by several mutations: S48E, W334G, W385G and W412G11. Unlike wild-type TDP-43, the fusion protein MCP–TDP-43(EGGG)—containing these four mutations—did not rescue phenotypes associated with *Xist*ΔE (Figs. 3b, 4d, e, Extended Data Fig. 11f–h); this suggests that self-association of TDP-43 permits the few available TDP-43 sites to support recruitment of multiple TDP-43 monomers. Similar results were obtained using a MATR3(S85C) mutant, which has previously been shown to impair both droplet formation and TDP-43 recruitment in comparison with wild-type MATR317 (Fig. 3b, 4d, e, Extended Data Fig. 11f–h). We therefore conclude that, through high-density binding to the E-repeat, *Xist* concentrates PTBP1, MATR3, TDP-43 and CELF1, which use homo- and heterotypic interactions to establish a physical condensate that compartmentalizes *Xist* and enforces X-linked gene silencing.

Our results suggested that the condensate established by the E-repeat is crucial for the *Xist*-independent phase of XCI after day 3 of differentiation8, leading to the hypothesis that this condensate—containing PTBP1, MATR3, TDP-43 and CELF1—can

be retained in the Xi in the absence of *Xist*. To test this idea, we confirmed that CELF1 is enriched in the Xi in primary female mouse embryonic fibroblasts carrying a conditional *Xist* allele (Fig. 4f). Upon *Xist* deletion, loss of the H3K27me3 Xi-accumulation closely followed the loss of *Xist* over time (Fig. 4f, g, Extended Data Fig. 11i–k). Notably, CELF1 enrichment

remained in 25–40% of cells even after the enrichment of *Xist* or H3K27me3 in the Xi became undetectable (Fig. 4g, Extended Data Fig. 11i–k). CELF1 enrichment was dependent upon PTBP1, MATR3 and TDP-43, as their depletion in the absence of *Xist* resulted in fewer cells with CELF1 Xi-accumulation (Extended Data Fig. 11l–q).

We conclude that the protein condensate seeded by the E-repeat is stable without *Xist* and is critical for the enforcement of silencing during the *Xist*-independent phase of XCI in differentiating ES cells. Our findings uncover a mechanism for the persistence of a functional RNA-seeded nuclear compartment, and reveal an unanticipated mechanism for RBP-mediated gene regulation and epigenetic memory (Fig. 4h, I, Supplementary Note 6).

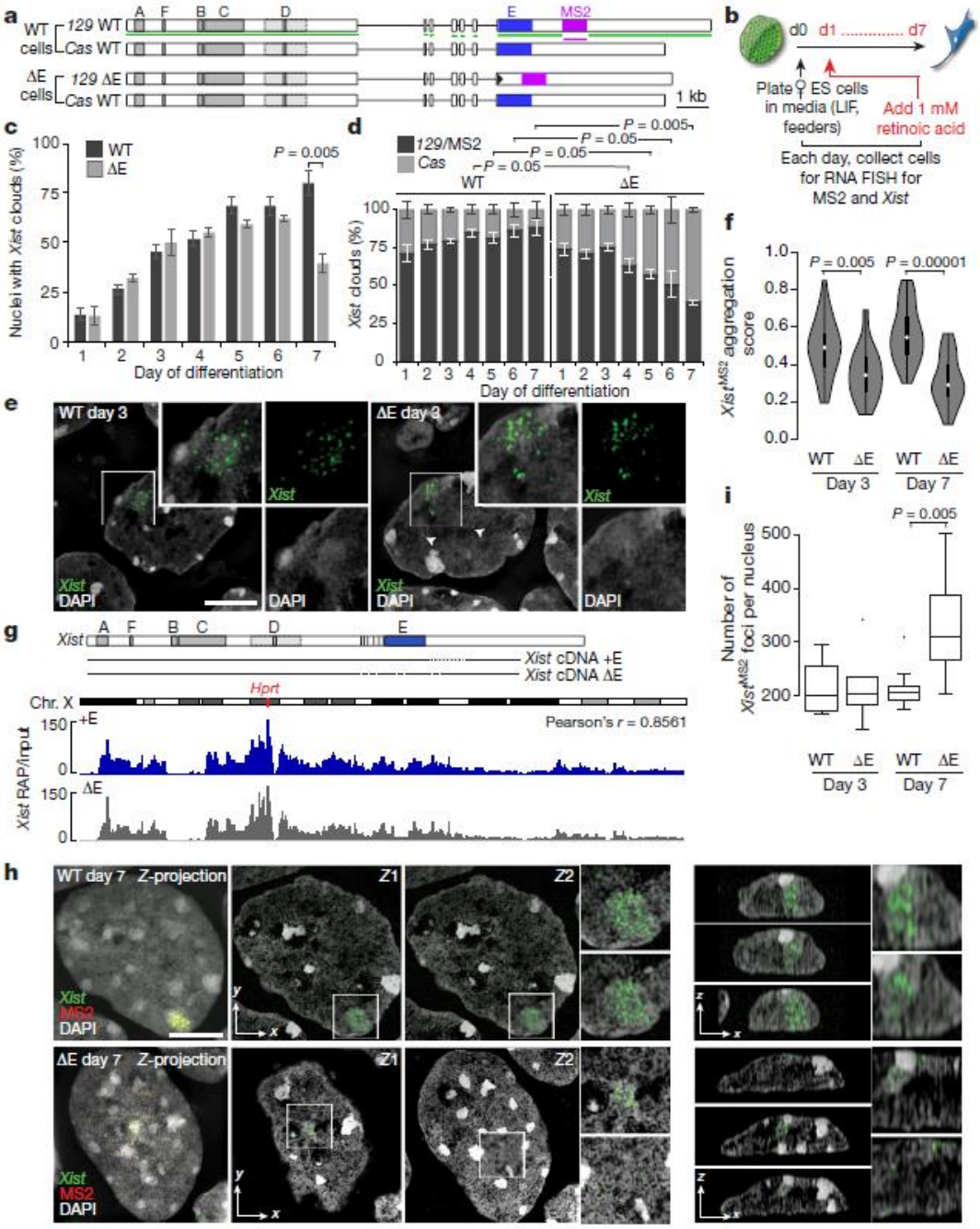


Fig. 3-1. The E-repeat mediates *Xist* sequestration and controls the number of *Xist* foci.

a) *Xist* alleles in female wild-type and ΔE ES cells. Green (*Xist*) and magenta (MS2) lines indicate FISH probes. WT, wild type.

b) Schematic of the experimental procedure.

c) The percentage of nuclei with an *Xist* cloud (*n* = 100) at the indicated days of wild-type and ΔE ES cell differentiation. Data are mean ± s.e.m across 3 replicates; two-tailed Student's *t*-test.

d) The allelic origin of *Xist* clouds (*n* = 50). Data are mean ± s.e.m. across 3 replicates; two-tailed Student's *t*-test.

e) Sections from 3D-SIM images showing *Xist* RNA FISH signals from the *Xist*MS2(129*)* and *Xist*ΔE,MS2(129*)* alleles at differentiation day 3 in wild-type and ΔE cells, respectively. Arrowheads indicate *Xist*ΔE,MS2 foci located away from the *Xist* cloud. The insets show the enlargement of the marked regions. Right, the same images as in the inset but with the DAPI and *Xist* signals separated. Scale bar, 5 μm.

f) Violin plots showing aggregation scores of *Xist*MS2 clouds (*n* = 30) from one replicate in **d**. Violin plots depict median (white) and interquartile range (black), trimmed (grey) to represent data minimum and maximum values. Two-sample Kolmogorov–Smirnov test.

g) Top, tet-inducible *Xist* cDNA transgenes inserted into the *Hprt* locus in male ES cells used in RAP-seq experiments. Dashed lines indicate deleted regions. Bottom: RAP–seq profile of *Xist* containing the E-repeat (+E) and *Xist*ΔE across the X chromosome after 6 h of doxycycline treatment in ES cells. Data are from one experiment.

h) Left, 3D-SIM Z-projection of co-localizing *Xist* and MS2 RNA FISH signals from wild-type or ΔE cells at differentiation day 7, merged with DAPI. Scale bar, 5 μm. The next two panels show *Xist* and DAPI signals from two different *Z*-planes, and the smaller panels to the right show enlargements of the *Xist* signal from each *Z*-plane. Right, *Y*-plane sections

through cells shown in the left panels, highlighting *Xist* localization relative to the nuclear lamina. Enlargements of the areas containing *Xist* are shown on the right.

i) Box plot of the distribution of the number of *Xist* RNA foci from the wild-type *Xist* allele or the *Xist*ΔE,MS2(129) allele (*n* = 10). Horizonal lines denote the median, whiskers indicate 1.5 times the interquartile range, dots represent outliers. Two-sample Kolmogorov–Smirnov test.

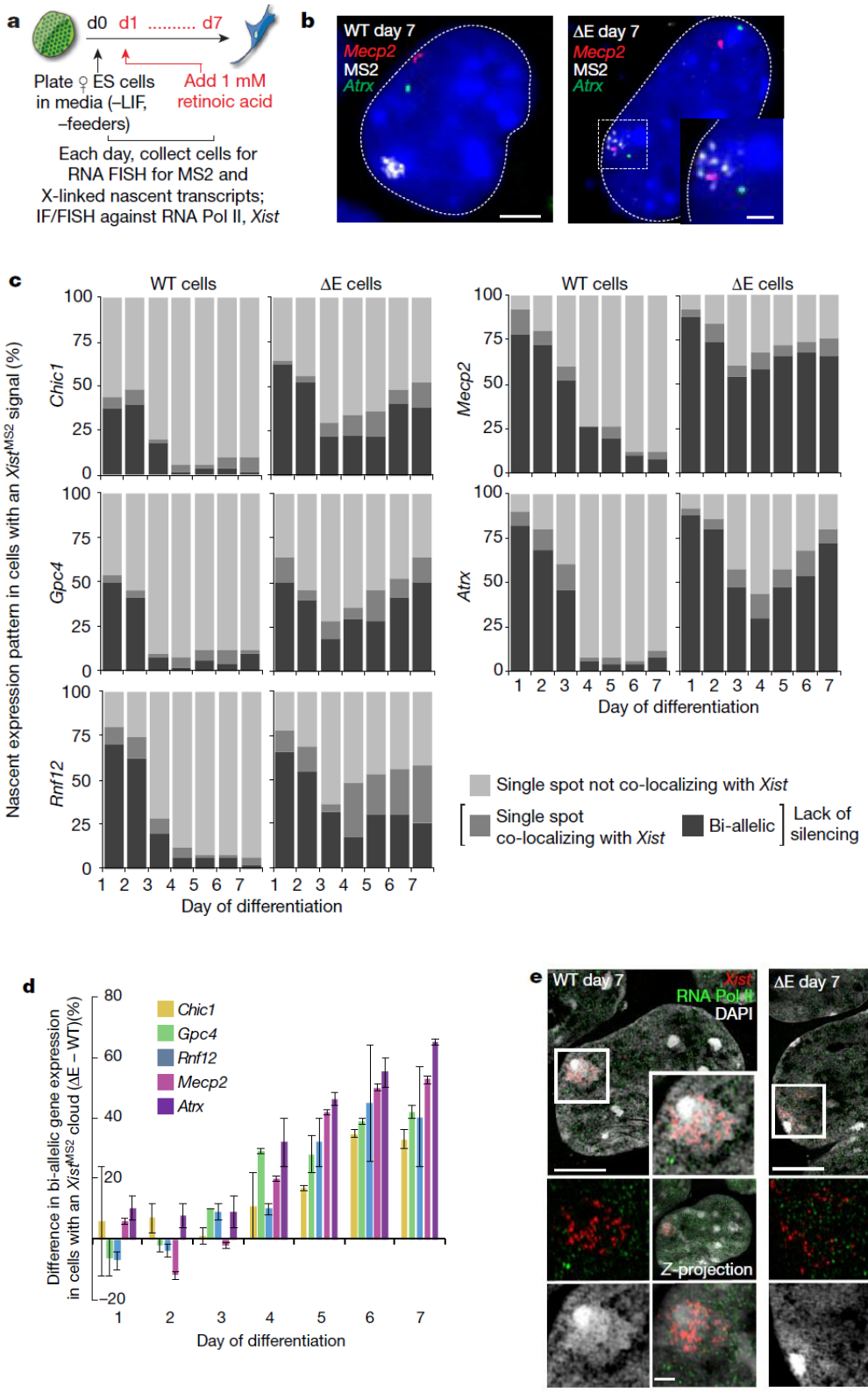

a
d0 d1 ..... d7
Plate ♀ ES cells in media (–LIF, –feeders)
Add 1 mM retinoic acid
Each day, collect cells for RNA FISH for MS2 and X-linked nascent transcripts; IF/FISH against RNA Pol II, Xist
b
WT day 7
Mecp2
MS2
Atrx
ΔE day 7
Mecp2
MS2
Atrx
c
WT cells
ΔE cells
Nascent expression pattern in cells with an Xist^MS2 signal (%)
Chic1
Gpc4
Rnf12
Mecp2
Atrx
Day of differentiation
Single spot not co-localizing with Xist
Single spot co-localizing with Xist
Bi-allelic
Lack of silencing
d
Difference in bi-allelic gene expression in cells with an Xist^MS2 cloud (ΔE – WT)(%)
Chic1
Gpc4
Rnf12
Mecp2
Atrx
Day of differentiation
e
WT day 7
Xist
RNA Pol II
DAPI
Z-projection
ΔE day 7
Xist
RNA Pol II
DAPI
Z-projection

Fig. 3-2. he E-repeat establishes heritable gene silencing.

a) Schematic of the experimental procedure.

b) Epifluorescence images showing the predominant nascent expression pattern of X-linked genes *Mecp2* (red) and *Atrx* (green) in wild-type (mono-allelic expression from the active X chromosome, Xa) and ΔE cells (bi-allelic expression) with a *Xist*-MS2 signal (white) at differentiation day 7. Inset, an enlargement of the boxed region, highlighting the fainter, dispersed *Xist*ΔE-MS2 signal. Scale bars, 5 μm (main), 1 μm (inset)

c) Quantification of nascent expression patterns of the indicated X-linked genes in wild-type and ΔE cells displaying an *Xist*MS2-coated X chromosome (*n* = 50), across 7 days of differentiation. Results were replicated three times.

d) The mean percentage difference in bi-allelic nascent gene expression (lack of silencing) between ΔE cells and wild-type cells with an *Xist*MS2 cloud, across 7 days of differentiation. Data are mean ± s.e.m.

e) 3D-SIM sections through wild-type and ΔE cells expressing *Xist*MS2 at differentiation day 7, stained for RNA Pol II (green) and DAPI (grey) and probed for *Xist* (red). The inset shows a magnification of the indicated region. The four small images below each large image are as follows, clockwise from top left: same as inset but without DAPI; *z*-projection of the whole nucleus; z-projection of the inset; same as inset showing only DAPI. Scale bars, 5 μm (main); 1 μm (insets).

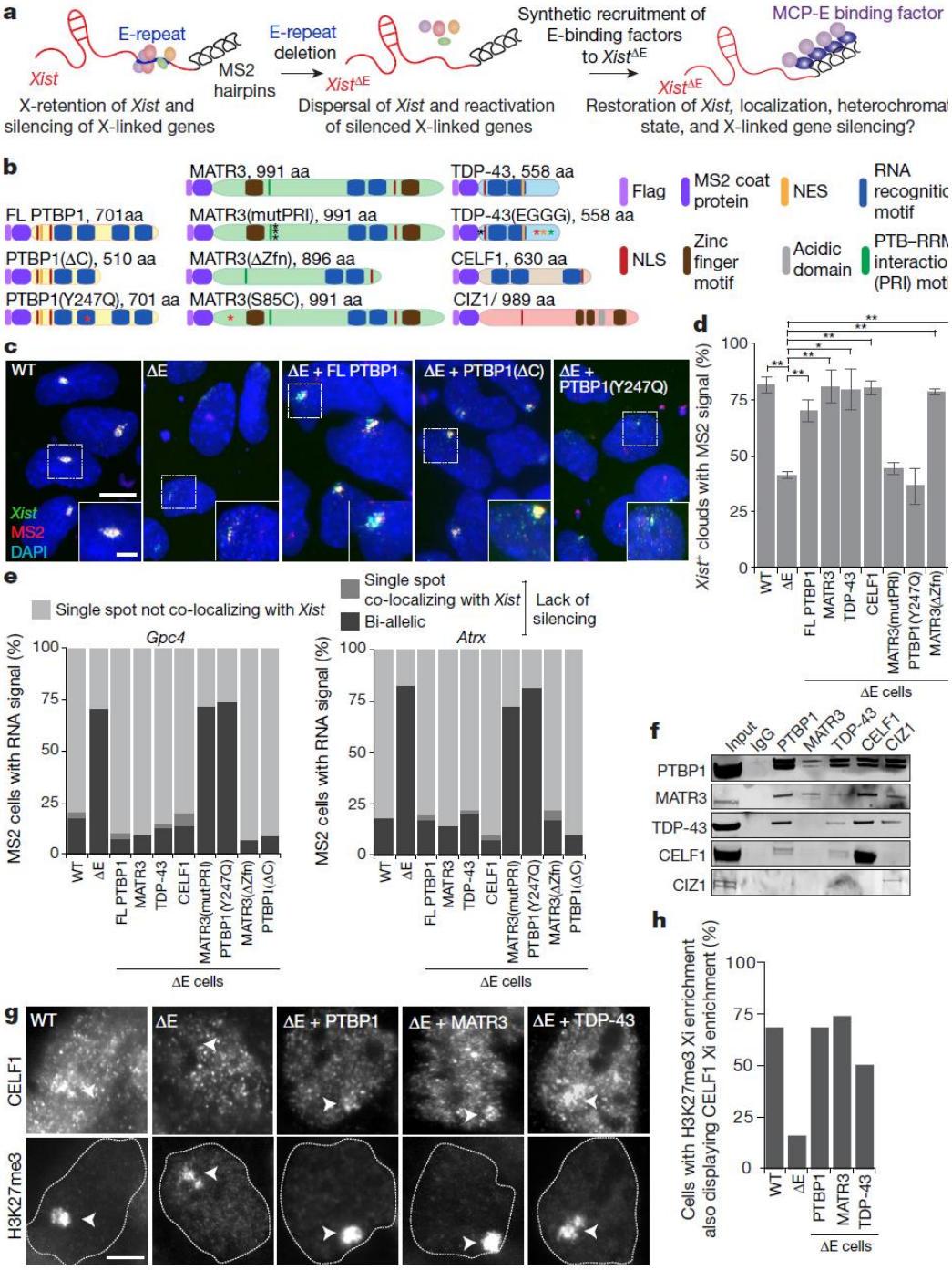


Fig. 3-3. PTBP1, MATR3, TDP-43 and CELF1 confer gene silencing and *Xist* sequestration functions on the E-repeat.

a) MCP-fusion protein rescue approach for *Xist*ΔE,MS2.

b) Illustration of Flag-tagged MCP-fusion proteins and mutants. Point mutations are indicated with asterisks. Length of fusions includes the Flag and MCP sequences.

c) Representative epifluorescence images from RNA FISH experiments against *Xist*-MS2 in day 7 differentiated wild-type, ΔE, and ΔE lines expressing variants of MCP–PTBP1 fusion proteins. The inset shows the enlargement of the marked region. Scale bars, 10 μm.

d) Histogram showing the proportion of nuclei with an *Xist* FISH signal (n = 80) that also displayed a co-localizing MS2 signal at differentiation day 7 in wild-type, ΔE, or ΔE lines expressing the indicated MCP fusion proteins. Data are mean ± s.e.m from two independent experiments; *P < 0.05, **P < 0.005, two-tailed Student's t-test.

e) Quantification of nascent *Gpc4* or *Atrx* expression pattern in cells expressing *Xist*-MS2 (n = 50) at differentiation day 7. The experiment was repeated twice with similar results.

f) Immunoprecipitation of PTBP1, MATR3, CELF1, TDP-43 and CIZ1 from ES cell extracts (no RNase) and detection of co-precipitated proteins by immunoblotting, using the same antibodies. Similar results were obtained from three independent trials; for source data see Supplementary Fig. 1.

g) Representative epifluorescence images showing the CELF1 pattern in wild-type, ΔE or ΔE-rescue cell lines expressing MCP–PTBP1, MCP–MATR3 or MCP–TDP-43 at day 7 of differentiation. Arrowheads indicate the Xi marked by H3K27me3-enrichment. Scale bars, 5 μm.

h) Histogram showing the percentage of wild-type, ΔE or ΔE-rescue cell lines with H3K27me3 Xi enrichment that also display a co-localizing accumulation of CELF1 at differentiation day 7 (n = 50, from one experiment).

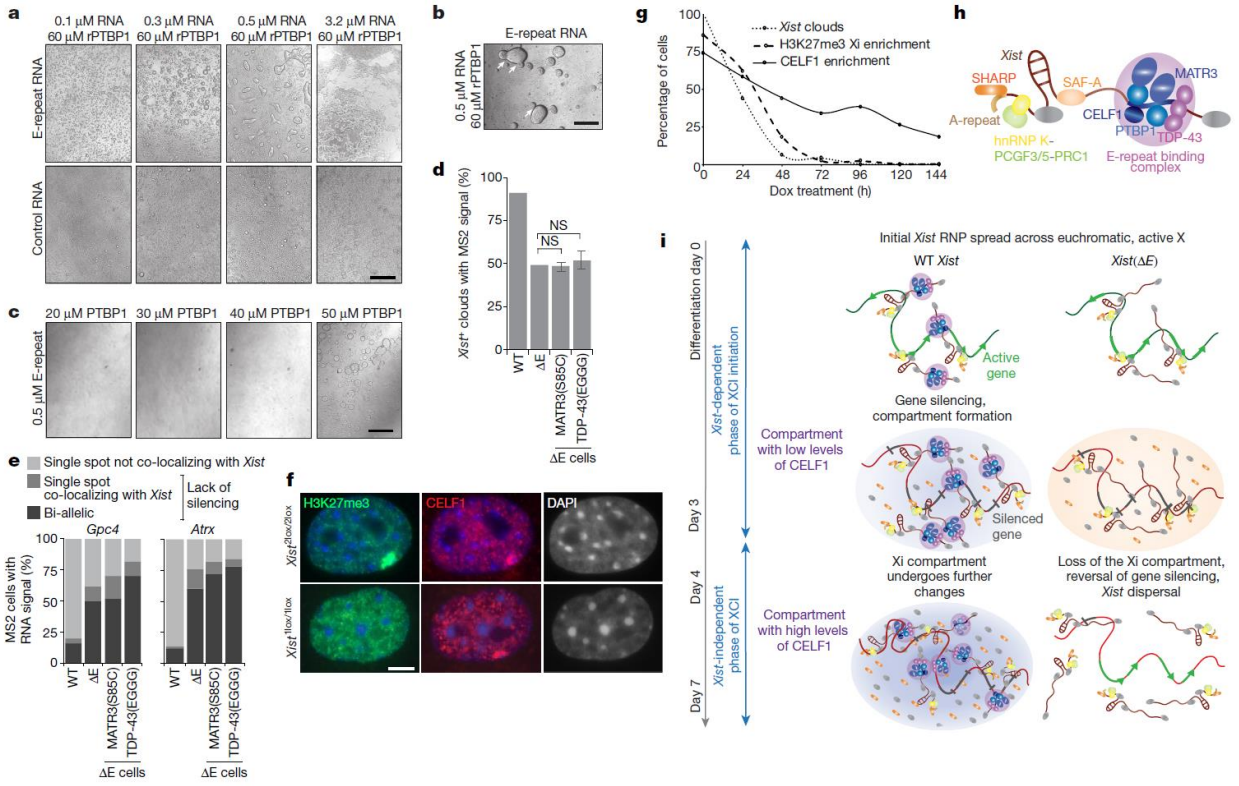


Fig. 3-4. Self-association of E-repeat-binding RBPs is critical for formation of the Xi compartment.

a) Bright-field images of droplets formed with rPTBP1 and different concentrations of E-repeat RNA or control RNA. Scale bar, 100 μm.

b) Bright-field image of droplets undergoing fusion (arrows). Scale bar, 50 μm.

c) Bright-field images of droplets formed with 0.5 μM E-repeat RNA and decreasing amounts of rPTBP1. Scale bar, 100 μm. For **a**–**c**, images were taken after 40 min.

d) The percentage of nuclei with an *Xist* FISH signal (n = 100) at differentiation day 7 that also displayed a co-localizing MS2 signal in wild-type, ΔE or ΔE lines expressing the indicated MCP fusion proteins. Data are mean ± s.e.m. from two independent experiments; two-tailed Student's t-test.

e) Histograms showing nascent *Gpc4* or *Atrx* expression patterns in cells described in **d** expressing *Xist*-MS2 (*n* = 50, from one experiment).

f) Representative images showing H3K27me3, CELF1 and DAPI staining in *Xist*2lox/2lox, *Rosa26*M2rtTA/tetO-Cre mouse embryonic fibroblasts (MEFs) before (2lox/2lox) or 96 h after (1lox/1lox) the addition of doxycycline to induce Xist excision. Scale bar, 5 μm.

g) Graph showing the percentage of MEFs described in **f** that that show Xi-enrichment of H3K27me3 or CELF over a 144 h time course of doxycycline (dox) treatment (*n* = 50, from one experiment).

h) Illustration of the E-repeat-bound *Xist* ribonucleoprotein complex (RNP) with PTBP1, CELF1, MATR3 and TDP-43 binding to the E-repeat and undergoing additional protein–protein interactions. Other *Xist*-interactors are indicated. **i**, Model of Xi-compartment formation via protein condensation (Supplementary Note 6). Wild-type *Xist*: upon differentiation, the *Xist* RNP spreads across the X chromosome and induces the formation of a higher-order assembly by recruiting additional protein molecules into the *Xist* territory through extensive homo- and heterotypic protein–protein interactions (purple oval). We postulate that the condensate, in addition to E-repeat-interactors, integrates other proteins (grey) including SHARP3 (orange). The assembly changes over time as indicated by low and high CELF1 levels and the increased purple coloring of the oval. *Xist*ΔE: without the E-repeat, *Xist* localization and X-linked gene silencing initiate normally (middle), potentially through non-E-repeat-dependent protein condensation events (orange oval); however, they cannot be reinforced later, despite the *Xist*-independence of XCI at this point.

Methods

Cell culture

All mouse ES cell lines were cultured in knockout DMEM (Life Technologies) supplemented with 15% FBS (Omega), 2 mM l-glutamine (Life Technologies), 1× NEAA (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma), 1× penicillin/streptomycin (Life Technologies), and 1,000 U ml−1 murine LIF (homemade) on 0.3% gelatinized plates (porcine skin gelatin, Sigma) pre-plated with irradiated male DR4 feeders (homemade from day 14.5 embryos, with appropriate protocols in place ensuring the ethical treatment of animals, approved by the UCLA Institutional Animal Care and Use Committee, known as the Chancellor's Animal Research Committee (ARC), 2007-180-41). For 3D-SIM microscopy experiments, ES cells were maintained in 2i culture conditions without feeders, before differentiation26. No differences in results upon cell differentiation were observed between the ES cell propagation conditions. ES cells were maintained as small colonies and passaged with trypsin and single-cell dissociation at 80% confluency. Mycoplasma tests (Lonza) are routinely conducted on cells cultured in the laboratory. Additionally, DAPI staining of the cells used in the study did not indicate any mycoplasma contamination.

Female ES cell differentiation

Female wild-type F1 2-1 MS2*129* (and derivatives thereof)27 were trypsinized to single cells and counted. Cells were seeded in 2 ml of MEF medium (DMEM (Invitrogen) supplemented with 10% FBS (Omega), 2 mM l-glutamine (Life Technologies), 1× NEAA (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma) and 1× penicillin/streptomycin (Life

Technologies)) at a density of 20,000–200,000 cells per 4 cm2 (depending on the experiment) on tissue culture plates for western blotting or onto 18 mm sterile glass coverslips for immunofluorescence (IF)/FISH experiments, both of which were pre-coated with sterile 0.3% gelatin (porcine skin gelatin, Sigma) or Matrigel (Corning, diluted 1:100). At 24 h post-seeding, the culture medium was changed and supplemented with 1 μM all-*trans* retinoic acid (Sigma), which was changed daily thereafter until the cells were collected for analysis

Female MEF culture

Female MEFs (*Xist*2lox/2lox, *Rosa26*M2rtTA/tetO-Cre-recombinase)28 were maintained in MEF medium. To delete *Xist*, cells were treated with 2 μg ml−1 doxycycline (Sigma) for up to 144 h to induce expression of Cre-recombinase.

Male ES cell culture

Male ES cells were maintained as described in the section ‘Cell culture’. To express *Xist*, ES cells were trypsinized to single cells and counted. Cells were seeded in 2 ml of mouse embryonic cell media at a density of 20,000–200,000 cells per 4cm2 (depending on the experiment) on tissue culture plates for western blotting and RNA collection or onto 18 mm sterile glass coverslips for IF/FISH experiments, both of which were pre-coated with sterile 0.3% gelatin (porcine skin gelatin, Sigma) or Matrigel (Corning, diluted 1:100 in cold DMEM media). For knockdown experiments, siRNAs were added upon plating (see section ‘siRNA treatments’). For *Xist* expression, doxycycline (Sigma) was added to a final concentration of 2 μg ml−1 for 6–24 h, depending on the experiment.

RNA FISH

FISH against *Xist* RNA was performed using both RNA and DNA probes. FISH against the MS2-insert, *Atrx*, *Gpc4*, *Mecp2*, *Rnf12* and *Chic1* was performed using DNA probes. In undifferentiated ES cells, the DNA probe against *Xist* additionally detects *Tsix*.

RNA probe preparation

Strand-specific RNA probes were generated using a T3 in vitro transcription kit (Promega) in the presence of Chromatide AlexaFluor-UTP (ThermoFisher). Six transcription templates (about ~700 nt) were generated from *Xist* exon 1 (Primers UCLA 1416–1429, Supplementary Table 1), and used in transcription reactions containing 0.5 mM ATP, CTP, GTP, 0.1 mM UTP and 0.05 mM Chromatide Alexa Fluor 488-UTP (Life Technologies) along with 1× T3 transcription buffer supplemented with 10 mM DTT, 500U RNase inhibitor, 170U T3 RNA polymerase and 5 µg of pooled template DNA in a final volume of 500 µl at 37 °C overnight in the dark. The transcription reaction was treated with 15U RNase-Free DNase for 15 min at 37 °C before probe purification. To purify the probes, 1/3 of the transcription reaction was loaded on a pre-spun (700*g*, 5 min) Chromaspin-100 column (Clontech) and centrifuged (700*g*, 5 min). The eluates were combined and precipitated with 100% EtOH in the presence of 100 mg tRNA and 1/10 volume of sodium acetate (Sigma). We sometimes also purified RNA probes using a 2.5× volume of AMPure beads (Thermo Fisher 09-981-123, reconstituted according to ref. 29), which were washed twice on a magnet with 80% ethanol before elution of the probes from the beads with 50 µl water, followed by ethanol precipitation. The RNA pellet was washed twice in 70% ethanol, resuspended in 400 µl of RNase-free water, to which 1 ml EtOH was added for storage at −20 °C. To make the final probe mix, 1/7 of the Probe/EtOH

solution was added to 90 µl salmon sperm DNA (Sigma), 90 µl mouse Cot1 DNA (Life Technologies), 40 µl 3 M RNase-free sodium acetate (Sigma), 40 µl 10 mg ml−1 tRNA (Life Technologies) and 1 ml EtOH. After vigorous shaking, the solution was centrifuged at maximum speed for 10 min. The pellet was washed once with 70% EtOH and then once with 100% EtOH, allowed to dry completely, and then resuspended in 200 µl deionized formamide (VWR) and 200 µl 2× hybridization buffer (20% dextran sulfate (Sigma), 4× SSC (Ambion), 0.1 M NaH2PO4). Probes were stored at −80 °C and denatured at 95 °C for 5 min before use.

DNA probe preparation

For 3D-SIM and Airyscan experiments, FISH probes were labelled by nick translation as described previously30 using p15 cDNA plasmid as template and home-labelled Atto488, Cy3- or Texas Red-conjugated dUTPs31. For all other experiments, DNA probes were synthesized using the CGH Bioprime Array Kit (Thermo Fisher) according to the manufacturer's instructions. In brief, a 40 µl solution containing 100 ng of template DNA was denatured in the presence of 1× random primers at 95 °C for 5 min and snap-cooled on ice. Five microlitres of nucleotide mix, 5 µl of 488, 555-, or 594- dUTP or dCTP chromatide fluorophore (Life Technologies) and 5U Klenow exo-enzyme were then added and incubated in the dark at 37 °C for 6 h, after which an additional 5U of Klenow exo-enzyme was added. The reaction was incubated at 37 °C overnight, quenched with 10 µl stop solution, and then purified over a Chromatide-100 column or AMPure beads as described in the section 'RNA probe preparation'. The eluate was precipitated in the presence of 100 mg yeast tRNA (Life Technologies) and sodium acetate (Sigma). The final DNA probe mix was then prepared as in the section 'RNA probe preparation' to yield

400 µl of probe solution in formamide/hybridization buffer. The MS2 DNA template for DNA probe preparation was PCR-amplified from genomic DNA purified from wild-type F1 2-1 MS2*129* female ES cells (see Supplementary Table 1 for primers). For *Xist*, the DNA probe was synthesized using a full-length mouse *Xist* cDNA plasmid (p15A-31-17.9kb *Xist*, unpublished). The intron probe in Extended Data Fig. 6c was against intron 1, which is the longest intron within the gene. We were unable to get probes against other introns to work well in this assay, presumably owing to their short length and labile nature. Probes against X-linked genes were synthesized using BACs RP23-467J21 (*Gpc4*), RP23-265D6 (*Atrx*), WIBR1-2150D22 (*Chic1*), WIBR1-2704K12 (*Rnf12*) and W11-894A5 and W11-1189K18 (*Mecp2*) (all obtained from CHORI-BACPAC). Note that the use of two BACs for the *Mecp2* probe sometimes resulted in nascent FISH signals from one X chromosome that appeared as doublets (see Fig. 2b, wild-type panel).

RNA FISH procedure for epifluorescence microscopy

Culture medium was changed 10 min before cell collection to remove dead cells and stimulate transcription. Upon collection, culture medium was aspirated, and coverslips were gently rinsed twice with cold 1×PBS. Coverslips were then transferred to a new culture dish containing 1× PBS, which was then aspirated, and the cells were fixed in 4% paraformaldehyde (PFA) (Electron Microscopy Sciences) in 1× PBS for 10 min at room temperature (RT) under standard laboratory safety practices. After fixation, the cells were permeabilized in 0.5% Triton X-100 (Acros) in 1× PBS with 2 mM vanadyl ribonucleoside complex (NEB) for 10–20 min on ice. Coverslips were then stored in 70% ethanol at −20 °C for 1 h or until samples from all time points had been collected. Before hybridization

with the probe, the coverslips with cells were brought back to 4 °C and serially dehydrated by 5-min incubations in ice-cold 80%, 95% and 100% ethanol. Coverslips were removed from 100% ethanol and allowed to air dry before incubation with probe for 48 h at 37 °C in a sealed chamber humidified with 2× SSC/50% formamide. For RNA probes, coverslips were washed for 3 × 5 min in 50% formamide (Fisher)/2× SSC (Ambion) and 3 × 5 min in wash buffer II (10 mM Tris, 0.5 M NaCl, 0.1% Tween-20), before a 45 min incubation with 25 µg ml−1 RNaseA (Thermo Fisher) in wash buffer II at 37 °C. After RNaseA treatment, coverslips were washed for 2 × 5 min in wash buffer II, 2 × 5 min in 50% formamide/2× SSC, 3 × 5 min in 2× SSC and 3 × 5 min in 1× SSC before briefly drying excess 1× SSC off and mounting with Vectashield mounting media lacking DAPI (Vector Labs). Coverslips were sealed with Biotium Covergrip coverslip sealant (Thermo Fisher). For DNA probes, coverslips were washed for 3 × 5 min in 50% formamide/2× SSC, 3 × 5 min in 2× SSC and 3 × 5 min in 1× SSC before mounting. A 1:10,000 dilution of DAPI (0.5 mg ml−1) was included in all penultimate 1× SSC washes. All washes were conducted at 42 °C, cells were protected from light. All procedures were performed, and used reagents disposed of, according to standard laboratory safety procedures. The *Xist* RNA FISH probe used in our study covers the ~17.9kb exonic regions of *Xist*. The MS2 tag is ~1.1kb long and the MS2 FISH probe was designed to cover the entirety of the tag (see Fig. 1a). Differences in the length of sequence targeted by these probes made the *Xist* probe signal much brighter than the MS2 probe signal when visualized microscopically. Consequently, in our RNA FISH experiments, we used both probes to differentiate between the *cas* (detected by the *Xist* probe only) and the *129* allele (detected by the *Xist* and MS2 probes) in both wild-type and ΔE cells. Using the *Xist* probe allowed for better detection of the

extent of dispersal of the *Xist*ΔE transcripts, which was important for our aggregation score calculations (see Extended Data Fig. 1e and section '*Xist* aggregation analysis'). In the RNA FISH assay for the nascent transcripts of X-linked genes (Fig. 2b, c, d, Extended Data Fig. 8), the presence of two nuclear nascent transcript foci (or spots) is indicative of bi-allelic expression of the respective X-linked gene as is observed in undifferentiated ES cells that do not express *Xist* and have not yet initiated XCI (see Extended Data Fig. 8a, b). In cells expressing *Xist*, one focus on the X chromosome lacking *Xist* indicates silencing (see Fig. 2b, wild-type cell). Conversely, we interpreted a single focus co-localizing with the X chromosome expressing *Xist* (or MS2) as a lack of silencing. Cells expressing *Xist* with bi-allelic X-linked gene expression were also considered to have defective silencing (see Fig. 2b, ΔE cell). RNA FISH procedure for 3D-SIM and improved-resolution microscopy. All coverslips were processed according to ref. 32.

Immunofluorescence staining

The cell culture medium was changed 10 min before collection. Upon collection, culture medium was aspirated, and coverslips were gently rinsed twice with cold 1× PBS. Coverslips were then transferred to a new culture dish containing cold 1× PBS. If cells were CSK treated (MS2-CP-GFP expressing wild-type F1 2-1 MS2*129* ES cells; Extended Data Fig. 9b), then coverslips were gently treated with 1 ml (added dropwise) ice-cold CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM MgCl2, 10 mM PIPES pH 6.8) and incubated on ice for 30 s before aspiration. Coverslips were then similarly treated with 1 ml ice-cold CSK-Trt Buffer (CSK+0.5% Triton X-100) for 30 s, followed with a second ice-cold CSK treatment. Coverslips were then processed as described in ref. 33. See Supplementary Table 2 for antibody information.

Immunofluorescence staining combined with RNA FISH

Where immunostaining and RNA FISH were combined, immunostaining preceded FISH. Combined staining for epifluorescence microscopy. The immunostaining protocol was followed as outlined above, but coverslips were not mounted. Instead, after the last round of washes (omitting DAPI in the penultimate wash), coverslips were re-fixed in 4% PFA in 1× PBS for 10 min at RT and then dehydrated through a 70–85–95–100% ice-cold ethanol series before overnight incubation with probe as described above in the section 'RNA FISH procedure for epifluorescence microscopy'. Combined staining for 3D-SIM and improved-resolution microscopy. All coverslips were processed according to refs. 32,33.

Plasmid construction and cell line generation

***Xist*ΔE targeting construct**.

To create the targeting vector pCR2.1-Puro-*Xist*ΔE, 3 kb upstream and 1.2 kb downstream of the mouse *Xist* E-repeat were PCR-amplified from mouse genomic DNA using primers WRM163-166, modified for In-Fusion cloning (Clontech) using Kapa polymerase (Kapa biosystems) according to the manufacturer's instructions. The upstream homology arm was integrated at the EcoR1 site and the downstream homology arm at the BamH1 site, in a four-piece InFusion cloning reaction, into a vector containing a floxed puromycin resistance cassette (PCR2.1-*loxP*-pGK-Puro-pA-*loxP*). Positive recombinants were identified by restriction digest with HindIII and sequencing.

E-repeat deletion in wild-type F1 2-1 MS2129 ES cells

The *Xist* E-repeat was deleted in female wild-type F1 2-1 MS2*129* ES cells derived from an F1 cross of mice from pure bred *129* and *castaneous* background, and then targeted to contain a 11× tandem repeat of the MS2 hairpin located 1.2 kb downstream of the E-repeat27 via homologous recombination. The wild-type F1 2-1 MS2*129* female ES cells also harbour an M2-reverse tetracycline TransActivator (M2rtTA) cassette within the Rosa26 locus that confers neomcyin resistance on the cells. Cells obtained from half of a confluent T75 flask of wild-type F1 2-1 MS2*129* female ES cells were electroporated with 40 µg of PciI linearized PCR2.1-Puro *Xis*tΔE targeting plasmid (800 V, 0.2 ms, 4 mm cuvette, Biorad X-Cell electroporation module) and plated at varying dilutions on 10 cm plates of confluent irradiated DR4 feeders. Then, 36 h after plating, the cells were selected with 1 µg ml−1 puromycin for 10 days. One hundred clones were picked, expanded and subjected to Southern blot analysis using a Sac1 digest and an external probe (amplified using primers WRM193/194 (Supplementary Table 1) as outlined in Extended Data Fig. 5. The positive clone number 35 was expanded in culture, then transfected with a Cre-recombinase plasmid using Lipofectamine 2000 according to the manufacturer's protocol (Thermo Fisher), to delete the floxed puromycin resistance cassette. Transfected cells were serially diluted, 100 clones were picked, expanded and replica-plated for growth in the presence or absence of puromycin. Sub-clone number 96 was sensitive to puromycin. PCR analysis of genomic DNA confirmed the deletion of the puromycin cassette with primers APJ439/440 (Supplementary Table 1). Subsequent Southern blot analysis and sequencing of wild-type *Xist* and *Xist*ΔE PCR amplicons from genomic DNA (intron 6 to exon 7 using APJ248/631 (Extended Data Fig. 5, Supplementary Table 1)) showed that

the ΔE targeting construct integrated on the *129* allele of *Xist* upstream of the MS2 tag, preserving the 3′ splice site of intron 6, to yield the heterozygous E-repeat deletion ES cell line *Xist*ΔE,MS2(129)/WT(Cas) (ΔE ES cells) (Extended Data Fig. 5 and data not shown). Sequencing of the exon 6–exon 7 RT–PCR amplicon (obtained from cDNA of differentiated ES cells) derived from the *129*MS2 *Xist* transcript revealed the use of a cryptic 3′ splice site downstream of the *loxP* site (Extended Data Fig. 5). The use of the cryptic splice site extended the E-repeat deletion within the *Xist* transcript (as initially designed) by 42 nt and removed the *loxP* site and additional vector sequences present in the genomic DNA from mature *Xist*ΔE transcripts, resulting in a scar-less ligation of the 3′ terminus of exon 6 to nucleotide 1479 of exon 7 (Extended Data Fig. 5 and data not shown). We ensured that ΔE ES cells maintained two X chromosomes throughout the targeting process and differentiated equally to wild type as judged by changes in morphology, and loss of NANOG and *Tsix* expression upon induction of differentiation (Extended Data Figs. 5, 8).

Engineering of wild-type and ΔE ES cells with a FLP-FRT recombination platform for rescue experiments

Wild-type F1 2-1 MS2*129* ES cells and ΔE ES cells as described in the section ‘E-repeat deletion in wild-type F1 2-1 MS2*129* ES cells’ (half of a confluent T75 flask) were electroporated with 40 µg of Fsp1 linearized Flp-IN homing plasmid (that integrates a FRT landing site downstream of the Col1A locus and carried a puromycin resistance cassette for selection34) at 800 V, 0.2 ms, 4 mm cuvette using a Biorad X-Cell electroporation module before being serially diluted on 10 cm plates, pre-coated with irradiated DR4 feeders. At 36 h after plating, the cells were selected with 2 µg ml−1 puromycin for 10

days after which 200 clones were picked and expanded. Genomic DNA was isolated and digested with EcoRI, before being subjected to Southern blot analysis with the Col1A Xba/Pst1 3′ probe. Positive clones 1–61 (wild-type) and 137 (ΔE) were used for all subsequent experiments (Extended Data Fig. 9a).

Generation of Flp-In plasmids encoding Flag–MS2–CP fusion proteins

The MS2 coat protein (MCP) coding sequence was PCR-amplified with a forward primer encoding a 3×Flag tag downstream of a Kozak-ATG start signal from the pHMM vector (Addgene, 67717). The reverse primer contained an in-frame Nhe1 site (primers APJ526/570 (Supplementary Table 1)) such that any fragment ligated into the site would be expressed in frame with the MCP protein, separated by a three-amino acid (Gly-Leu-Gly) linker. The Flag–MCP–Nhe1 fragment was inserted into the EcoRI site of the pBS32 vector using Infusion cloning. This vector is similar to the pgkATGfrt vector described in ref. 34 except that the tet-inducible promoter was replaced with a CAGGS promoter, enabling constitutive expression of the fusion protein. The coding sequence for each protein (GFP, PTBP1, MATR3, TDP-43 and CELF1) was PCR-amplified from cDNA with infusion overhangs, or synthesized (Genewiz) and ligated into the Nhe1 site of the pBS32–Flag–MCP parent plasmid using InFusion cloning (Clontech). The PTBP1(Y247Q), MATR3(mutPRI) and MATR3(ΔZfn) mutants were generated using primer-directed mutagenesis. The wild-type PRI sequence (GILGPPP) was mutated to create the mutant PRI sequence (GAAAPPA)20. The coding sequences for the CELF1, MATR3(S85C), TDP43(EGGG) and MS2CP–GFP–MS2CP fusions were synthesized (Genewiz). All plasmids were verified by sequencing. The ΔC-terminal PTBP1 fragment that is fused to Flag–MCP in our rescue system comprises the first 299 amino acids of

PTBP1—which includes the first two RRMs as well as the MATR3 interaction site—followed by 68 amino acids that are out of frame, and do not encode a functional linker region. A premature stop codon terminates the protein at residue 367.

Generating wild-type and ΔE ES cells expressing Flag–MCP–fusion proteins via Flp-In recombination

A total of 33 μg of the pBS32 plasmid DNA encoding the Flag–MCP fusion proteins and 26 μg of plasmid encoding the flpase FlpO were electroporated into wild-type ES cells carrying the FRT homing site (clone 1–61) for the GFP fusion and ΔE ES cells with the FRT homing site (clone 137) for all other fusion constructs (1/2 of a confluent T75 flask of ES cells per electroporation) (Extended Data Fig. 9a). Cells were plated on confluent irradiated DR4 feeders in a 10 cm dish and, 36 h after plating, selected with 170 μg ml−1 hygromycin for 14 days, after which all colonies were picked and expanded. The resulting clones were tested for protein expression by immunoblot of lysates (RIPA buffer in 1× SDS lysis buffer (Thermo Fisher)) using an anti-Flag antibody as well as antibodies against the respective fusion protein (Supplementary Table 2). Immunostaining confirmed nuclear localization of all fusion proteins that failed to rescue the phenotypes associated with loss of the E-repeat. All clones used maintained two X chromosomes, as determined by FISH against *Tsix* in undifferentiated cells. For all rescue experiments, at least two clones were analysed, which revealed that the data are robust. Owing to space limitations, often the results from only one rescue clone per protein or mutant are shown.

Generation of tet-inducible XistΔTsix V6.5 male ES cells

Tet-On *Xist* male V6.5 ES cells carrying a tet-inducible promoter in place of the endogenous *Xist* promoter and a M2rtTA trans-activator as well as puromycin resistance

in the R26 locus2 (1/2 of a confluent T75 flask) were electroporated with 40 µg of Not1 linearized paa2Δ1.7 plasmid DNA35 (800 V, 0.2 ms, 4 mm cuvette using a Biorad X-Cell electroporation machine) and plated on confluent irradiated DR4 feeders, to stop *Tsix* expression. Then, 36 h after plating, the cells were selected with Neomycin/G418 for 10 days after which 100 clones were picked and subjected to Southern blot analysis as described in ref. 35 (data not shown). Positive clone 70 was used for the PTBP1 ChIP–seq experiments.

siRNA treatments

Silencer Select siRNAs (Thermo Fisher) against PTBP1 (s72337), MATR3 (s69629), CELF1 (s64632) and TDP-43 (s106686) were diluted to 20 nM in 1× siRNA buffer (60 mM KCl, 6 mM HEPES pH 7.5 0.2 mM MgCl2), aliquoted and stored at −80 °C until further use. Under sterile conditions at RT, 2.5 µl of 20 nM siRNA were added to 80 µl of fresh Opti-MEM solution (Gibco). siRNA MAX transfection reagent (1.6 µl, Life Technologies) was added to 80 µl Opti-MEM solution and subsequently added to the siRNA/opti-MEM solution after 5 min of incubation. The resulting solution was mixed by pipetting and left to incubate at RT for 20 min. The solution was then added to 200,000 cells in 0.8 ml of culture medium and plated in 1 well of a 12 well plate on 18-mm gelatinized coverslips and left overnight at 37 °C. For female ES cells undergoing differentiation, cells were plated in MEF medium and after 24 h, the culture medium was changed (with the addition of 1 µM all-*trans* retinoic acid, Sigma) and a second round of siRNA treatment was performed. Knockdown efficiency was assessed by immunoblotting (Supplementary Table 2).

Immunoblotting

Cells were collected by trypsinization, pelleted (1,000*g*, 5 min), resuspended in 500 µl 1× PBS to wash, and re-pelleted. The washed cell pellet was lysed in 5 pellet volumes of RIPA buffer and 40U benzonase (Novogen) and incubated at 4 °C overnight. The lysate was centrifuged at maximum speed to pellet the remaining insoluble material and the supernatant was transferred to a new tube and mixed with 4× Novex sample buffer containing 5% 14.3 M β-mercaptoethanol (Sigma) to a final concentration of 1×. The samples were then denatured for 5 min at 95 °C and loaded onto a 4–12% Novex Bis-Tris acrylamide gel with 1× MES running buffer (Life Technologies) run at 120 V for 1.5–2 h. The gels were transferred to a protran BA-85 nitrocellulose membrane (Whatman) using a Novex XCell II transfer system for 2 h at 30 V, 4 °C (or overnight at 4 °C at 10 V) in transfer buffer (25 mM Tris-HCl, 192 mM glycine, 20% methanol). Membranes were probed with primary antibody (Supplementary Table 2) in 1× Odyssey blocking buffer (LI-COR) overnight at 4 °C, washed for 3 × 5 min in PBS+0.2% Tween-20 (Thermo Fisher) and then incubated with appropriate secondary antibodies (1:10,000 dilution, Odyssey 700 and 800 nm antibodies) in the dark at room temperature for 30 min before being washed again and scanned on a LI-COR infrared imaging system.

Co-immunoprecipitation

For co-immunoprecipitation experiments, rabbit IgG and antibodies against PTBP1, MATR3, CELF1, CIZ1 and TDP-43 (Supplementary Table 2) were crosslinked to ProteinG-Dynabeads (Thermo Fisher) using the protocol provided by Abcam (http://www.abcam.com/protocols/ cross-linking-antibodies-to-beads-protocol) with minor modifications. In brief, 20 µl of bead slurry was isolated on a magnet and washed for 3

× 5 min in 5 volumes of 1× PBS. Beads were then washed once in 5 volumes of binding buffer (100 µl, 1× PBS containing 1 mg ml−1 of BSA (NEB)) for 10 min and incubated in 100 µl binding buffer supplemented with 5 µg of rabbit IgG or antibodies against PTBP1, MATR3, CELF1, CIZ1 or TDP-43. Samples were rotated for 1 h at 4 °C. Beads were then washed in binding buffer for 5 min, followed by an additional 5-min wash in 1× PBS. Next, the antibody was crosslinked by incubating in 100 µl of 1× PBS solution containing 0.2 M triethanolamine (Sigma) and 6.5 mg ml−1 dimethyl pimelimidate (DMP) (Sigma) pH 8.5 for 30 min with rotation at room temperature. Beads were then washed in 250 µl 0.2 M triethanolamine in 1× PBS for 5 min. DMP incubation and wash steps were repeated twice more before samples were quenched in 100 µl of 50 mM ethanolamine in 1× PBS for 5 min. The quenching step was repeated, and excess non-crosslinked antibody removed with 2 × 10 min incubations in fresh 1 M glycine pH 3.0. Beads were washed in 1× PBS for 3 × 5 min before use in immunoprecipitations. Immunoprecipitations were performed under non-denaturing conditions according to the Abcam protocol (http://www.abcam.com/ps/pdf/protocols/immunoprecipitation%20protocol%20(ip).pdf). Plates (4 × 15cm) of confluent wild-type F1 2-1 MS2*129* female ES cells were lysed by pipetting in 3 ml of lysis buffer (10 M Tris-HCl pH8, 137 mM NaCl, 1% NP40, 2 mM EDTA) supplemented with 1× Complete EDTA-free Protease Inhibitors (Roche) and incubated for 1 h on ice with or without RNase (10 µg ml−1 RNase A) (Thermo Fisher). Lysate was centrifuged at 4 °C at maximum speed in a tabletop microfuge for 15 min to pellet insoluble material. The supernatant was transferred to new tubes and precleared with 20 µl of washed ProteinG Dynabeads per 1 ml of lysate with rotation at 4 °C for 1 h. Then, 500 µl of lysate was added to each crosslinked antibody-proteinG Dynabead prep and

rotated at 4 °C overnight. The next day, crosslinked antibody–proteinG Dynabeads were isolated on a magnet and washed for 4 × 5 min in ice-cold wash buffer (10 mM Tris-HCl pH 7.4, 1 mM EDTA, 1 mM EGTA, 150 mM NaCl, 1% TritonX-100) supplemented with 1× Complete EDTA-free Protease Inhibitors. The co-purified proteins were eluted by boiling in 1× NuPage Protein Loading buffer (Thermo Fisher) supplemented with 5% β-mercaptoethanol, at 95 °C for 5 min. Samples were assessed by immunoblotting. The input represents 4% of lysate added per immunoprecipitate. 1/4 of the eluate was loaded per lane.

In vitro RNA transcription

For several in vitro experiments (Droplet assays, electrophoretic mobility shift assays (EMSAs)), RNAs encoding the E-repeat and other sequences were obtained by in vitro transcription (IVT). Templates for IVT were amplified from DNA using KAPA polymerase according to the manufacturer's instructions (KAPA Biosystems), and then gel-purified and concentrated over AMPure beads (homemade)29. See Supplementary Table 1 for primer information. RNA was transcribed and UREA-PAGE purified as described in ref. 36. For biotinylated RNAs, Biotin–UTP (Ambion) comprised 18% of the total UTP.

Droplet assays

rPTBP1 purification

Recombinant 6×-His tagged PTBP1 was expressed by IPTG induction from plasmid pQE-80L-PTBP4 (human PTBP1, isoform 4) (Douglas Black Lab) in BL21 bacterial cultures and purified using Ni-NTA agarose (Invitrogen) according to the manufacturer's instructions. The purified protein was dialysed and stored in buffer DG (20 mM HEPES-

KOH pH 7.9, 80 mM K glutamate, 20% glycerol, 2.2 mM MgCl2, 1 mM DTT and 0.1 mM PMSF) at a stock concentration of 36 mg ml−1.

rCELF1 purification

Recombinant 6×-His-tagged CELF1 was expressed by IPTG induction from plasmid pET28a-CELF1 (human) in Rosetta bacterial cultures and purified over His-Trap and Superdex 200 gel filtration columns. Purified protein was concentrated and stored in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 10% glycerol at a stock concentration of 5 mg ml−1. pET28a-CELF1 was constructed via In-fusion, using a fragment encoding the CELF1 coding region (see Supplementary Table 1 for primers) into the PET-28a plasmid. The CELF1 coding region was amplified from a fragment synthesized by Genewiz with primers modified for the pET-28a plasmid. The sequence of the plasmid was verified before use.

Droplet assays

Droplets (10 µl) were assembled in 1.5 ml Eppendorf tubes as described in ref. 37. In brief, 5 µl of a 2× buffer containing 200 mM NaCl, 40 mM imidazole, 2 mM DTT and 20% glycerol was supplemented with the E-repeat or control IVT RNA (varying concentrations), rPTBP1 (to a maximum concentration of 60 µM) and/or rCELF1 (maximum concentration of 38 µM) and water to 10 µl (final volume). The solution was mixed by pipetting and transferred to one well of an 8-well glass chamber slide (Ibidi) that had been pre-coated with 3% BSA, washed 3 times with RNase-Free water and dried. Droplets were imaged at 10×–20× magnification.

Electrophoretic mobility shift assays

EMSAs were performed as described in ref. 38 except that 40,000 counts per million of 5′ end labelled RNA was used per condition.

Quantitative RT–PCR and actinomycin D treatment

In several experiments we determined the levels of *Xist* by RT–PCR. For experiments with actinomcyin D treatment, the drug was dissolved in DMSO at 1 mg ml−1 and added to the culture medium to a final concentration of 1 μg ml−1. For PCR with reverse transcription (RT–PCR), cells were collected in 1 ml TRIzol (Thermo Fisher), after culture medium removal and washing with PBS. RNA was purified over RNAeasy columns (Qiagen). Total RNA (1 μg) was used in a reverse-transcription (RT) reaction with SuperScript III and an appropriate strand-specific reverse primer, according to the manufacturer's instructions (Thermo Fisher). One-twentieth of the RT reaction was used in a quantitative PCR reaction, using either 480 SYBR Green LightCycler PCR mix (Roche), SsoAdvanced Universal SYBR mix (Bio-Rad) or SYBR Green Master Mix (Applied Biosystems) and appropriate primers (see Supplementary Table 1), in triplicate reactions. RT–qPCR experiments were normalized against *Gapdh, Snrnp27* or *Rrm2* transcripts.

Crosslinking and immunoprecipitation of RNA and high-throughput sequencing (iCLIP–seq) for MATR3 and PTBP1

PTBP1, PTBP2 and TDP-43 iCLIP in differentiated cells was obtained from published datasets39,40. PTBP1 and MATR3 iCLIP experiments in ES cells were performed as described in ref. 41. For iCLIP–seq, all washes were conducted for 5 min per wash, at 4 °C with ice cold buffers. Three confluent 15-cm plates of male tetO-*Xist* V6.5 (pSM33) ES cells2 were used per immunoprecipitation upon 6 h of induction of *Xist* expression with 2

μg ml−1 doxycycline, and crosslinking was performed at 100 mJ cm−2 at 4 °C in a Stratalinker 1800 (Stratagene). Crosslinked cells were collected by scraping in cold 1× PBS and pelleted at 700*g* for 2 min. Cell pellets were lysed in ice cold lysis buffer (20 mM HEPES-KOH pH 7.5 (Sigma), 150 mM NaCl (Sigma), 0.6% Triton X-100 (Sigma), 0.1% SDS (Sigma), 1 mM EDTA (Gibco), and 0.5 mM DTT (Sigma)) and sonicated in a bioruptor (Diagenode) for 2 × 15 min (30 s on, 30 s off) on high setting at 4 °C. Sonicated lysates were cleared by centrifugation at 20,000*g*, 5 min, 4 °C, supernatants transferred to 15 ml Falcon tubes and diluted in 5 volumes of buffer containing 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.5 mM DTT, 1.25× complete protease inhibitors EDTA-free (Roche), 50 μg ml−1 yeast tRNA (Life Technologies) and 400 U RNaseOUT (Life Technologies). Samples were briefly mixed and rotated overnight at 4 °C. To prepare beads for pulldown, a magnet was used to isolate beads from 200 μl of proteinG–Dynabead slurry, which were then washed 3 times in WB150 (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.1% Triton-X100) and incubated overnight at 4 °C with 50 μg anti-MATR3 (Abcam, ab151714) or anti-PTBP1 (Abcam, ab5642) in 700 μl WB150. Beads were washed three times in WB750 (20 mM HEPES-KOH pH 7.5, 750 mM NaCl, 0.1% Triton-X100) and once with WB150 (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.1% Triton-X100) before incubation with lysate. After overnight incubation in lysate, beads were collected at the bottom of the Falcon tube with a magnet and the supernatant was removed. Beads were then transferred to a 1.5 ml Eppendorf tube with 1 ml of WB150, washed five times in WB750 and twice in PNK buffer (20 mM HEPES-KOH pH 7.5, 10 mM MgCl2, 0.2% Tween-20). The immunoprecipitated RNA was fragmented in 100 μl of 1× MNase buffer (NEB) containing 5.0 μg of yeast tRNA that was pre-warmed to 37 °C in

a thermomixer (Eppendorf) set to shake for 15 s on/15 s off at 750 rpm (or minimum speed required to prevent settling of the beads). 1× MNase buffer (50 µl) containing 60 gel units per ml (6 Kunz units per ml) of micrococcal nuclease (NEB M0247S) were added and incubated for exactly for 5 min. The reaction was stopped by the addition of 500 µl of EGTA buffer (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 20 mM EGTA, 0.1% TritonX-100). The beads were then washed four times in EGTA buffer and twice in cold PNK buffer. The fragmented RNA was dephosphorylated in 100 µl of 1x FastAP buffer (Fermentas) containing 0.15 U µl−1 of fast alkaline phosphatase (Thermo Scientific, EF0651) and 0.2 U µl−1 of RNaseOUT (Life Technologies, 10777-019), incubated in a thermomixer for 90 min at 37 °C, 15 s shaking/20 s rest. Beads were washed four times in WB750 and twice in cold PNK buffer. The dephosphorylated RNA was then ligated to a 3′biotinylated linker RNA in 40 µl of buffer containing 1 mM ATP, 25% PEG4000 (Sigma, 202398), 0.5 U µl−1 T4 RNA ligase1 (NEB M0204S), 0.5 U µl−1 RNaseOUT, and 6.0 µM L3 linker (Supplementary Table 1). The ligation reaction was incubated in a thermomixer overnight at 16 °C, 15 s on/4 min off at a speed that prevents beads from settling. The next day, beads were washed four times in WB150 and twice with cold PNK buffer. The RNA was then 5′ end labelled in 24 µl PNK wash buffer with 16 µl of 1×PNK buffer (NEB) containing 150 µCi of γ[32P]ATP, 10U PNK and 1U µl−1 of RNaseOUT. The reaction was incubated in a thermomixer for 20 min at 37 °C set to shake for 15sec on/20 s off. The beads were then washed three times with WB150. The immunoprecipitated complexes were eluted off the Dynabeads in 50 µl of buffer (100 mM Tris-HCl pH 7.5, 0.6% SDS, 5 mM EDTA, 50 mM DTT and 50 ng µl−1 yeast tRNA) incubated for 10 min at 85 °C, shaking continuously at 900 rpm. The elute was transferred to a new tube and the beads

were rinsed with 1,200 µl of buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1.25× complete protease inhibitors (Roche), 50 ng µl−1 yeast tRNA and 0.1% Triton X-100) which was added to the first eluate. The combined eluates were centrifuged for 5 min at 4 °C at maximum speed and the supernatant transferred to a new tube to prevent carry over of any remaining Dynabeads. To prevent contamination with IgG heavy chain, which co-migrates with many proteins of interest, the biotinylated RNA–protein complexes were bound to monomeric avidin beads. To do this, 10 µl packed monomeric avidin agarose beads (Thermo Fisher) were washed three times with WB150. Beads were pelleted after each wash by spinning in a swing bucket rotor at 1,000*g*, 4 °C (use of the swing bucket rotor helps prevent loss of agarose beads). One packed bead volume was mixed with an equal volume of WB150 and 15 µl of the bead slurry was added to each combined eluate and rotated at 4 °C for 4 h. The beads were then pelleted as above and washed three times with WB150. After the final wash, the remaining 5–20 µl of supernatant was carefully removed with a p10 pipette. The complexes were eluted from the avidin beads by incubation in 30 µl of buffer (10 mM Tris-HCl pH 7.5, 10% glycerol, 2.2% SDS, 5 mM EDTA) at 85 °C for 10 min in a thermomixer shaking at 900 rpm. After centrifugation to pellet the beads, the supernatant was transferred to a new tube and mixed with 5 µl of 1× LDS sample buffer (Life Technologies) with 300 mM DTT. Samples were incubated at 90 °C for 10 min and then loaded on a pre-run (75 v, 10 min) NuPAGE Bis-Tris Gel (Life Technologies NP0307) with 1× MOPS running buffer and run for 10–15 min at 75 V and then 120 V until each sample is satisfactorily separated. The gel was then incubated in transfer buffer (25 mM Bis-Tris, 25 mM bicine, 1 mM EDTA pH 7.2, 20% methanol) for 5 min and then transferred onto a protran BA-85 nitrocellulose membrane using a semi-dry

transfer apparatus (Biorad 170-3940) for 75 min at 400 mA (not exceeding 15 V). After completion of the transfer, the membrane was briefly washed in milli-Q water, wrapped in plastic film and exposed on a phosophoimager screen for 1 h. The regions of interest were then excised from the membrane and transferred to an Eppendorf tube. The RNA was eluted from the membrane by incubation in 300 µl of buffer (100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA and 2 µg µl−1 proteinase K) for 30 min at 55 °C in a thermomixer, shaking continuously. 300 µl of pre-warmed buffer (100 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM EDTA, 7 M urea and 2 µg µΛ−1 proteinase K) was then added to the tube and incubated for a further 30 min at 55 °C. The supernatant was then transferred to a new tube and extracted with an equal volume of phenol:chloroform (5:1, pH 4.5). The separated aqueous phase was precipitated with 0.5 µl of Glycblue (Life Technologies), 60 µl sodium acetate pH 5.4 and 600 µl isopropanol overnight at −20 °C. The next day, the RNA was pelleted by centrifugation at 4 °C for 30 min at maximum speed. The pellet was then washed with 1 ml 75% EtOH before air-drying for 2 min and dissolved in 5.70 µl RNase-free water and left on ice for 5–10 min before being reverse transcribed. To do this, 0.5 µl of 10 mM dNTPs and 0.5 µl of 2 µM RT primer (Supplementary Table 1) were added to the RNA, mixed by pipetting and denatured for 5 min at 70 °C before snap-cooling on ice. The RT primers contain an 11 nt unique molecule identifier (UMI) used in sequence analysis (see section ‘CLIP–seq analysis’). The sample was then equilibrated at 25 °C in a PCR machine before 3.5 µl of RT mix were added (2 µl 5× first strand buffer, 0.5 µl 100 mM DTT, 0.5 µl 100 U µl−1 Superscript III (Life Technologies) and 0.5 µl 40 U µl−1 RNaseOUT (Life Technologies)) and incubated for 5 min at 25 °C and then for 20 min at 42 °C, and then 20 min at 48 °C. The

reverse transcription reaction was then transferred to a new Eppendorf tube containing 100 µl TE, 11 µl 3 M sodium acetate and 2.5 volumes of 100% EtOH. The cDNA was precipitated overnight at −20 °C, pelleted and washed as described above, dissolved in 5 µl RNase-free water and then mixed with 7.5 µl of formamide containing 10 mM EDTA, bromophenol blue and xylene cyanol tracking dyes. For size determination, ladder was prepared as follows: 2 µl GeneScan 500LIZ size marker (Life Technologies 4322682), 3 µl H2O, 15 µl formamide containing 10 mM EDTA with no tracking dyes. The samples and ladder were denatured for 5 min at 85 °C and then loaded on a pre-run 5.5% (19:1 bisacrylamide:acrylamide) urea-PAGE gel (1× TBE, 7.5 M urea) for 20 min at 21 V. The gel was then scanned and a gel slice in the range of 70–120 nt was excised, chopped into 1-mm cubes and the cDNA eluted in 700 µl of TE buffer rotating at RT overnight. The next day, the cDNA was precipitated overnight as described above. The washed pellet was then dissolved in 6.7 µl of RNase free water and left on ice for 5–10 min before being transferred to a PCR tube. Subsequently, the RNA was circularized by addition of 1.5 µl of: 0.8 µl of CicrLigase II buffer, 0.4 µl 50 mM MnCl2 and 0.3 µl of 100 U µl−1 Cicr-Ligase II ssDNA ligase (Epicentre CL9021K) and incubated in a PCR machine at 60 °C for 60 min. The circularized cDNA was then digested with BamHI by addition of 30 µl of: 4 µl 10× FastDigest Buffer, 0.9 µl of 10 µM cut_oligo and 25.1 µl RNase-free water. This mix was incubated for 4 min at 95 °C after which the temperature was decreased by 1 °C each minute until 37 °C after which 2 µl of FastDigest BamH1 (Thermo Scientific FD0054) was added and incubated at 37 °C for a further 30 min. The sample was transferred to an Eppendorf tube and pelleted as described above. The pelleted DNA was dissolved in 12 µl RNase free water. 2 µl was used to prepare a 42 µl PCR mix, containing 1× PFU

buffer, 0.2 mM dNTPs, 0.2 μM P3 and P5 solexa primers, and 0.5 U PFU polymerase. A negative control containing water instead of cDNA was also prepared. The 42 μl reaction was then split into 4 × 10 μl reactions and PCR-amplified for 20, 24, 28 and 32 cycles (94 °C/3′; 94 °C/30 s; 63.5 °C/15 s; 72 °C/30 s; with final extension at 72 °C for 7 min). The PCR amplicons were run on a 2% agarose gel in 0.5× TBE/EtBr and the number of cycles required to produce 50–200 ng of PCR product from the remaining 10 μl of cDNA template was calculated. The PCR reaction was repeated using 10 μl of the remaining ssDNA template, run on a 2% gel as before and the 150–210 bp size range was excised and purified using Zymoclean Gel DNA recovery kit (Zymo Research D4007). DNA concentration was determined by qubit using the dsDNA Broad Range assay and prepared for sequencing on an Illumina HiSeq2000 machine using a single end 100 bp protocol.

Enhanced crosslinking and immunoprecipitation of RNA and high-throughput sequencing (eCLIP–Seq) for CELF1

eCLIP experiments against CELF1 (anti-CELF1 (ab129115) were performed as described in ref. 42 with a few modifications. As with iCLIP, male tet-inducible -*Xist* V6.5 ES cells (pSM33)2 were induced with 2 μg ml−1 doxycycline for 6 h before crosslinking at 100 mJ cm−2 at 4 °C in a Stratalinker 1800 (Stratagene). Cells were then processed according to the eCLIP protocol for both input and immunoprecipitated samples until DNA was obtained. We then followed the iCLIP protocol from the gel-purification of the cDNA through to amplification and purification of the DNA library. eCLIP samples were sequenced on an Illumina HiSeq2000 machine using the single end 50 bp protocol.

CLIP–seq analysis

CLIP–seq results were mapped using TopHat and processed with the publicly available fastq-tools, fastx-toolkit, Samtools, Bedtools, Deep-Tools and UCSC scripts43–45. The first 11 bases of each sequenced read correspond to a UMI, composed of a library-specific barcode (3 nt) flanked by four degenerate nucleotides. The UMI permitted removal of PCR duplicates from the total sequenced reads with the fastq-uniq command-line tool. The Fastx-toolkit was then used to clip 3′ adaptor sequences. Sequences shorter than 20 nt were discarded. Reads were then de-multiplexed and mapped to the iGenome mm9 reference genome by TopHat with high stringency settings. Library-depth normalized counts were generated, and data were converted to bigwig format to visualize tracks in IGV. Peaks were called using CLIPper (see ref. 45 and https://github.com/YeoLab/clipper) using the–superlocal option. Scripts are available at https://www.github.com/ShanSabri/iCLIP.

ChIP–seq

Plates (3 × 15 cm) of male tetOXist-Δ*Tsix* V6.5 ES cells (around 100 million cells) were used to prepare chromatin for PTBP1 ChiP–seq. Cells were induced for 0 h or for 20 h with 2 μg ml−1 doxycycline to induce *Xist* before collection by trypsinization. Cells were then pelleted by centrifugation at 700*g* for 5 min at RT and resuspended in a total volume of 10 ml PBS. The wash step was repeated twice before resuspending in 10 ml of 1× PBS and transferring to a 50 ml Falcon tube to which disuccinimidyl glutarate (DSG; Pierce) in DMSO was added for crosslinking to a final concentration of 2 mM and incubated for 10 min at room temperature with gentle mixing. Cells were then pelleted, and the supernatant was discarded. Cells were re-suspended in 10 ml ES cell medium and incubated for 10

min with 1% formaldehyde (16% methanol free, Pierce) at room temperature with gentle. The reaction was quenched by addition of freshly made 0.125 M glycine (Sigma) for 5 min at room temperature. Cells were pelleted and supernatant was discarded. Cells were washed twice in 50 ml PBS with protease inhibitors (Complete EDTA free, Roche) before being pelleted and flash-frozen in liquid nitrogen and stored at −80 °C. The frozen pellets were processed for ChIP–seq as described in ref. 46.

ChIP–seq analysis

Reads were mapped using BowTie to the iGenome mm9 reference genome. Duplicate reads were removed, and length extended to 49 nt. Normalized reads count were generated across 50-nt bins. Tracks were visualized in IGV in bigWig format.

RNA affinity purification

RAP was performed as described in ref. 2. For the RAP–seq experiment, we used male T20 ES cells47 carrying a homing site in the *Hprt* locus on the single X chromosome as well as a tetracycline-inducible transactivator in the R26 locus. The *Hprt* homing site includes a bidirectional, tetracycline-inducible promoter for expression of a control gene (*EGFP*) and of the *Xist* cDNA transgene introduced later by site-specific recombination, as well as a *loxP* site neighbouring the tet-promoter and linked to a truncated neomycin-resistance gene that lacks a promoter and translation initiation codon34. We integrated two different *Xist* cDNA transgenes into the homing site by electroporation of the respective *Xist* cDNA encoding plasmid and a Cre expression plasmid. The *Xist* transgene plasmid contained a promoter-less *Xist* sequence followed by a polyadenylation signal and a PGK promoter and translation initiation codon linked to a *loxP* site. Site-specific recombination of the *loxP* sites in the *Xist* cDNA plasmid and the

homing site linked the translation initiation codon and *Pgk1* promoter to the *neo* gene, which restored the antibiotic resistance marker. A single copy of the *Xist* cDNA transgene was thus integrated under the control of the inducible promoter. In this study, we used an approximately 14.5kb *Xist* cDNA with either a 4122 nucleotide deletion between BstEii sites within the *Xist* cDNA, deleting the E-repeat and surrounding sequences, or a 1237 nucleotide deletion ending at a similar region within *Xist* and not including the E-repeat, which was generated by deleting internal sequences within the cDNA by SnaBI digestion and re-ligation (Fig. 1g and data not shown). Cells were induced with 2 μg ml−1 dox for 6 h, before fixation for RAP—seq. RAP libraries were sequenced on the Illumina platform. Adaptor trimming was performed using cutadapt version 1.15 in paired-end mode with the following parameters: "-a AGATCGGAAGAGC -A AGATCGGAAG AGC". Read mapping was performed using bwa version 0.7.17-r1188 and samtools version 1.4 with the following command: "bwa mem -t 30 -T 0 ${INDEX} ${R1} ${R2} | samtools view -b - | samtools sort -O sam -T tmp -n - | samtools fixmate -O sam - - | samtools sort -O bam -T tmp - > ${BAM}". Read filtering was performed using samtools version 1.4 with the following command: "samtools view -b -q 30 ${BAM} | samtools rmdup - - | samtools view -b -L ${BED} - | samtools sort -O bam -T ${tmp} - > ${FILT_BAM}". The ${BED} variable is a path to a .bed file containing all chromosomes except for the *Xist* locus (chrX:103460216-103483359). Library-depth-normalized tracks were generated using bedtools v2.27.1 with the following commands: "bamToBed -i ${FILT_BAM} | sort -u -k 1,1 -k 2,2 -k 3,3 > ${FILT_BED} ; genomeCoverageBed -split -bga -scale `echo 1000000000/$(wc -l ${FILT_BED} | awk '{print $1}') | bc` -i ${FILT_BED} -g ${CHROM_SIZES} > ${BG} ; bedGraphToBigWig ${BG} ${CHROM_SIZES} ${BW}".

Sample library-depth-normalized tracks were divided by the library-depth-normalized input (SRR850637 from ref. 2) in R v.3.5 using rtracklayer 1.42.2 to import the tracks as Rle and export the divided tracks as bigwig. Input-normalized tracks were smoothed using deeptools v.3.4.3 and bedtools v.2.27.1 with the following command: “multiBigwigSummary bins -b ${NORM_BW} -out res.npz -bs 1000–outRawCounts ${NORM_BG}; bedGraphToBigWig ${NORM_BG} ${CHROM_SIZES} ${SMOOTHED_NORM_BW}”.

Microscopy

Epifluorescence imaging

Cells with immunofluorescence and RNA FISH stainings were imaged using a Zeiss AxioImager M1 microscope with a 63× objective and acquired with AxioVision software. Epifluorescence images shown are sections and were analysed, merged and quantified using ImageJ or Adobe Photoshop.

3D-structured illumination microscopy (3D-SIM)

3D-SIM superresolution imaging was performed on a DeltaVision OMX V3 system (Applied Precision, GE Healthcare) equipped with a 100 A/~1.40 NA Plan Apo oil immersion objective (Olympus), Cascade II:512 EMCCD cameras (Photometrics) and 405, 488 and 593 nm diode lasers. Image stacks were acquired with a *z*-distance of 125 nm and with 15 raw images per plane (five phases, three angles). The raw data were computationally reconstructed with the soft-WoRx 6.0 software package (Applied Precision) using a wiener filter set at 0.002 and channel-specifically measured optical transfer functions using immersion oil with different refractive indices as described in ref. 33. Images from the different channels were registered using alignment parameters

obtained from calibration measurements with 0.2-µm diameter TetraSpeck beads (Invitrogen) as described in ref. 48.

Improved confocal microscopy

Improved confocal laser scanning microscopy was performed on a LSM880 platform equipped with 100×/1.46 NA or 63×/1.4 NA plan Apochromat oil objectives and 405/488 diode and 594

helium–neon lasers using the Airyscan detector (Carl Zeiss Microscopy). An appropriate magnification was used in order to collect image stacks from a region that encompassed the nucleus of interest thereby optimizing imaging time and reducing photobleaching. The pixel size and *z*-optical sectioning were set to meet the Nyquist sampling criterion in each case. Airyscan raw data were linearly reconstructed using the ZEN 2.3 software.

Quantitative image analysis

All image analysis steps were performed using Fiji or ImageJ49,50, or IMARIS (Oxford Instruments).

***Xist* aggregation analysis**

To quantify the aggregation of *Xist* clouds, images were taken as *Z*-stacks and transformed in a maximum intensity projection image to detect the entire *Xist* FISH signal in one plane. The background was removed using a rolling ball radius of 50 pixels. *Xist* RNA cloud areas were measured by creating a binary mask over the *Xist* RNA FISH signal for each analysed *Xist* cloud. The edges of each *Xist* cloud signal were determined by selecting a central pixel and all associated pixels of same intensity value (• }5 units). The ImageJ FracLac51 plugin was then used to calculate the area of a circle encompassing each cloud signal. The ratio of the *Xist* cloud area over its bounding circle

area approximates the compaction of the *Xist* RNA cloud. Significant differences between wild-type and ΔE ES cell or siRNA-treated samples were tested with the non-parametric two-sample Kolmogorov–Smirnov test.

Imaris measurements

Raw *z*-stack 3D-SIM images were converted to an Imaris-compatible format using the Imaris File Converter module. Before analysis, all images were adjusted to ensure identical intensity/brightness levels. Using the Imaris MeasurementPro module, 50 linear 3D distances between 100 randomly chosen *Xist* foci were measured per cell, across 5 cells per sample.

CELF1 intensity plot profiles

Airyscan image stacks were imported into ImageJ and converted to 16-bit composites. The 3D-stacks were reduced into 2D images and 2 μm intensity line plots were used to extract the intensity profiles over the Xi enriched signal in the CELF1 channel. The same line-plot was used in a random nucleoplasmic region to select for the average nuclear CELF-1 intensities. The ratio of the top 10% intensities of the signals were plotted after diving over the nucleoplasmic signal.

Xi DAPI intensities quantification

Wide-field image stacks were generated from 3D-SIM raw data of H3K27me3 and DAPI stained cells by average projection of five consecutive phase-shifted images from each plane of the first angle and subjected to an iterative 3D deconvolution using soft-WoRx 6.0 software. The reconstructed image stacks were imported to ImageJ and converted to 8-bit files. In order to measure the Xi underlying DAPI intensity, binary masks from the H3K27me3 channel were created to define the Xi territory of day 7 wild-type and

ΔE ES cell nuclei. A threshold was carefully applied selecting the border of the H3K27me3-enriched region that demarcates the Xi territory. Subsequently, the grey values of the corresponding masked region in the DAPI channel were extracted and plotted.

Segmentation of Xist RNA foci from 3D-SIM datasets

The 32-bit reconstructed 3D-SIM image stacks were imported into ImageJ. Grey values were shifted to the positive range and converted to 16-bit composites after subtracting the mode grey value to remove background noise. Segmentation of *Xist* RNA foci was performed by using the TANGO plugin52 on ImageJ according to the pipelines described in ref. 33. In brief, nuclear masks were created by using the nucleus processing chain. *Xist* foci were segmented by first pre-filtering with a TopHat filter with a radius of 1 pixel in all three dimensions (*xyz*), followed by a Laplace of Gaussian filter with a radius of 1 pixel (*x*, *y*, *z*). Segmentation of foci was performed using the spot detector 3D with Otsu auto-thresholding. Segmented objects were post-filtered with a size and edge filter of 5 pixels per spot and a signal-to-noise ratio greater than 2.

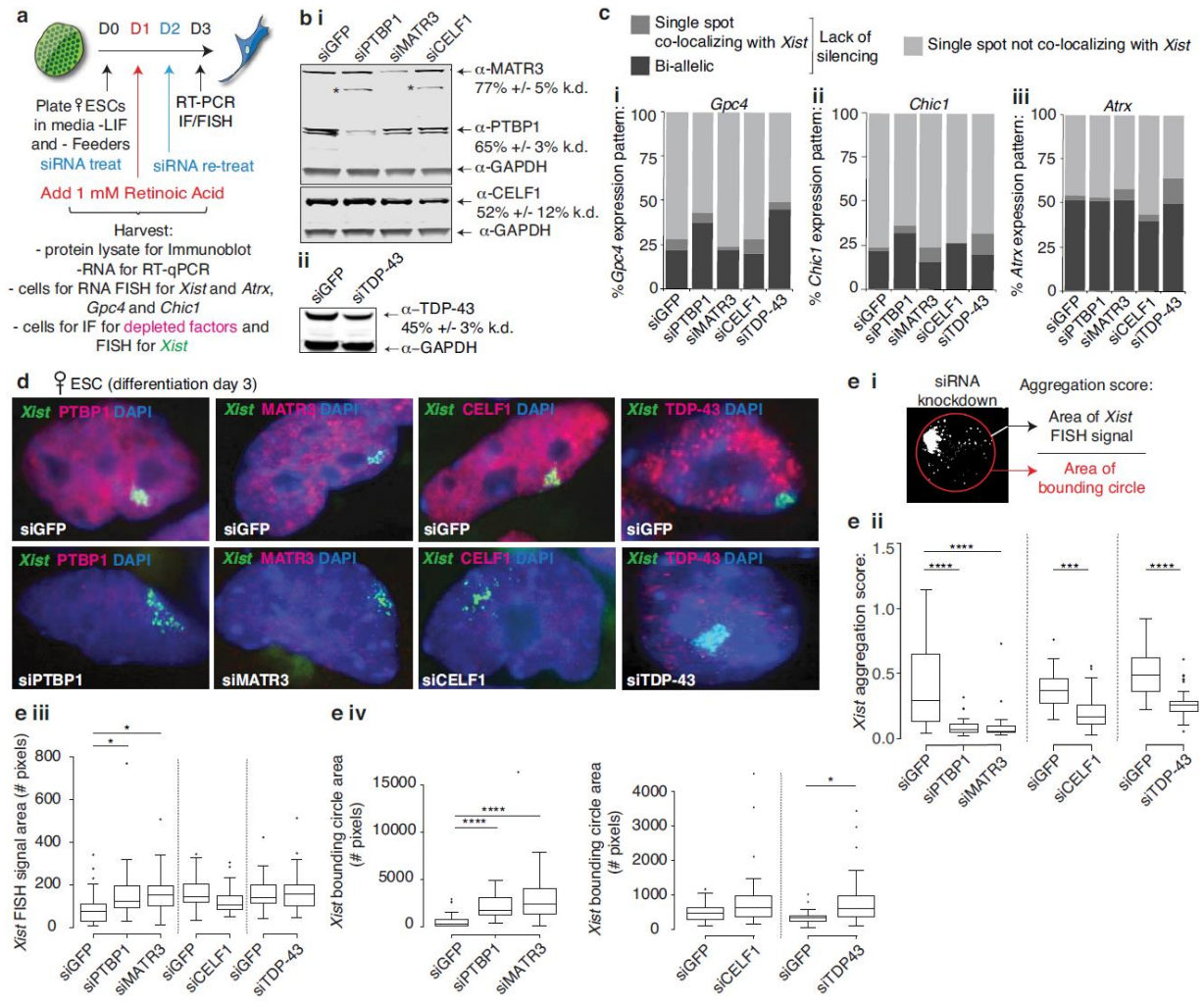
Amira reconstructions

3D reconstructions were performed using Amira 2.3 (Mercury Computer Systems). Image stacks were imported into Amira as separate channels. *Xist* FISH or antibody stainings were reconstructed as surface renderings, while DAPI was reconstructed as volume rendering using the Volren module that enables visualization of intensity in colour maps.

Data availability

All genomic data for *Xist* interactions and chromatin association have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE137305. Reagents are available upon request.

Supplemental Information



Extended Data Fig. 1 Depletion of PTBP1, MATR3, CELF1 and TDP-43 does not strongly affect gene silencing during the *Xist*-dependent stage of XCI initiation.

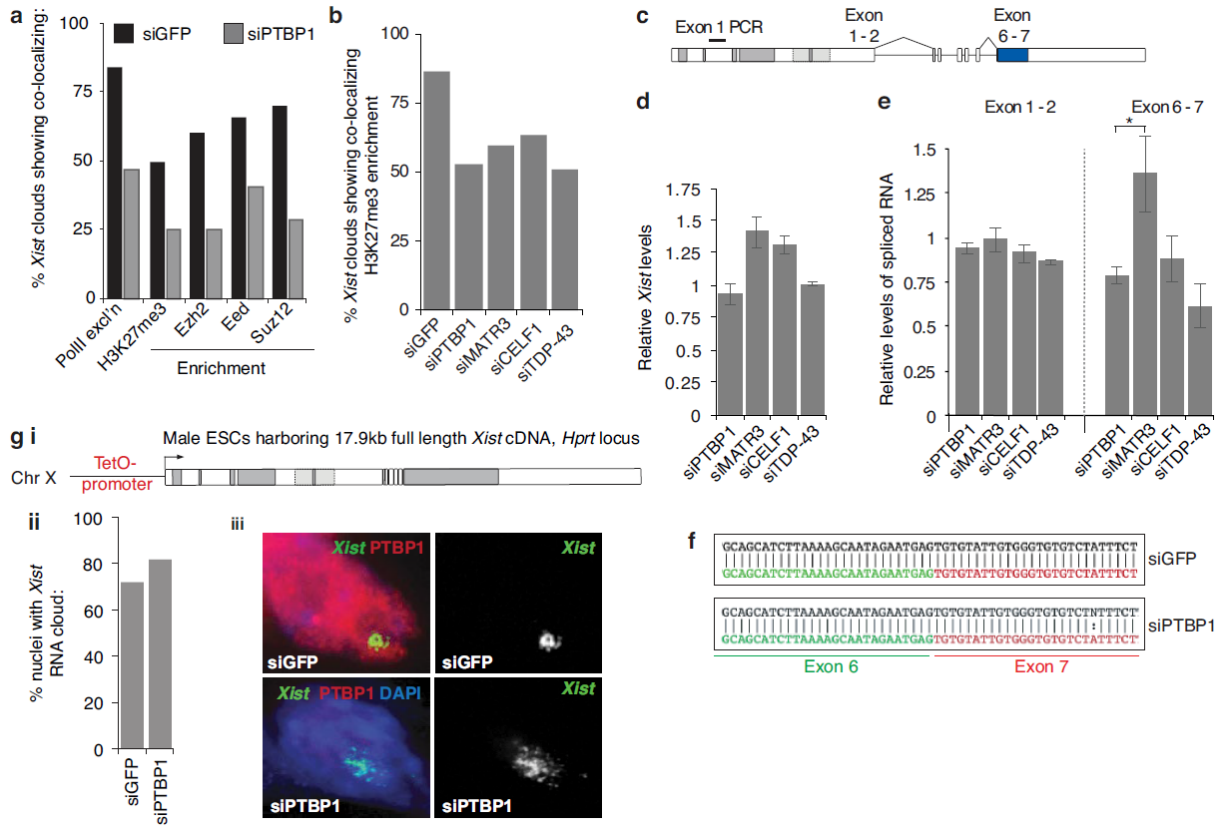
a) Experimental schematic.

b) (i), Immunoblot confirming the siRNA-mediated knockdown of PTBP1, MATR3, and CELF1, normalized to GAPDH. Asterisks indicate non-specific bands. (ii), As in (i), except for TDP-43. Error represents the s.e.m. from three independent experiments. For source data see Supplementary Fig. 1.

c) (i), Graph showing nascent transcription patterns of the X-linked gene *Gpc4* after 3 days of differentiation and knockdown of the indicated factor (spot refers to nascent transcription event on one chromosome) (*n* = 50 from 1 experiment). (ii), Same as (i) but for *Chic1*. (iii), Same as (i) but for *Atrx*.

d) Representative images of siRNA-treated differentiating cells immunostained for indicated proteins (red), probed for *Xist* (green) and DAPI stained (blue).

e) (i), Schematic for aggregation score calculation. (ii), Box plots showing *Xist* aggregation scores upon depletion of indicated proteins. Independent siGFP controls were used for CELF1 and TDP-43 experiments. (iii), Box plots showing the *Xist* mask values used to calculate the aggregation scores in (ii). (iv), Box plots showing the bounding circle area values encompassing the *Xist* mask used to calculate the *Xist* aggregation scores in (ii). For box plots in (ii)–(iv): (*n* = 25): **P* < 0.05, ****P* < 0.0005, *****P* < 0.00005; two-tailed Kolmogorov–Smirnov test from one replicate in **b**. Horizontal lines denote the median, whiskers indicate 1.5× the interquartile range, dots represent outliers.



Extended Data Fig. 2 Depletion of PTBP1, MATR3, CELF1 and TDP-43 affects *Xist* localization during XCI initiation without strongly altering *Xist* processing.

a) Proportion of *Xist*-positive cells with co-localizing exclusion of RNA Pol II or enrichment of H3K27me3 or the PRC2 components EZH2, EED, SUZ12 on the Xi, in female ES cells differentiated for 3 days and treated with siGFP or siPTBP1 (*n* = 50 from one experiment).

b) Percentage of *Xist*-positive cells with H3K27me3 Xi-enrichment in day 3 differentiated female ES cells treated with siGFP, siPTBP1, siMATR3, siTDP-43 or siCELF1 (*n* = 100, from one experiment). The siPTBP1 sample is independent from that in **a**.

c) *Xist* splicing events assessed below.

d) Histogram showing *Xist* abundance (exon 1 PCR above) upon siRNA-mediated knockdown of indicated RBPs in female ES cells at differentiation day 3.

e) As in **d**, except for the abundance of spliced *Xist* exon 1–2 and exon 6–7 amplicons upon knockdown. For **d** and **e**, samples were normalized against siGFP and *Snrnp27* RNA and assessed in triplicate from three independent experiments. Error bars represent s.e.m; *$P < 0.05$, two-tailed Student's *t*-test.

f) Snapshot of expected spliced exon 6 (green) to exon 7 (red) sequence. Correct exon 6–7 ligation occurs after 72 h of siGFP or siPTBP1 treatment in differentiating female ES cells (black sequence) in two independent experiments.

g) (i), A tet-inducible full-length *Xist* cDNA transgene was inserted into the X-linked *Hprt* locus in male ES cells. (ii), Percentage of cells with an *Xist* cloud after 48 h of siPTBP1, and dox treatment starting at 24 h of siRNA treatment, in cells described in (i) ($n = 80$, from one experiment). (iii), Representative RNA FISH images of *Xist*, co-immunostained for PTBP1 and DAPI labelled, in cells described and treated as in (i), (ii). Note *Xist* dispersal upon PTBP1 knockdown, despite the absence of *Xist* splicing.

	# Reads in dataset	Window	# Reads in window	% of total mapped reads	% of reads in *Xist*
PTBP1 CLiP	3,328,965	1st kb of *Xist*	1084	0.0325	0.897
		E repeat (1435nt)	111514	3.349	92.292
		Xist (22860nt)	120827	3.629	
MATR3 CLiP	5,020,287	1st kb of *Xist*	1425	0.028	1.058
		E repeat	124617	2.482	92.579
		Xist	134605	2.681	
CELF1 CLiP	3,897,448	1st kb of *Xist*	2	$5x10^{-5}$	0.004
		E repeat	46147	1.184	99.99
		Xist	46149	1.184	

Extended Data Fig. 3 PTBP1, MATR3, CELF1 and TDP-43 directly bind the *Xist* E-repeat, comprising a tandem array of 20–25nt C/U/G-rich elements.

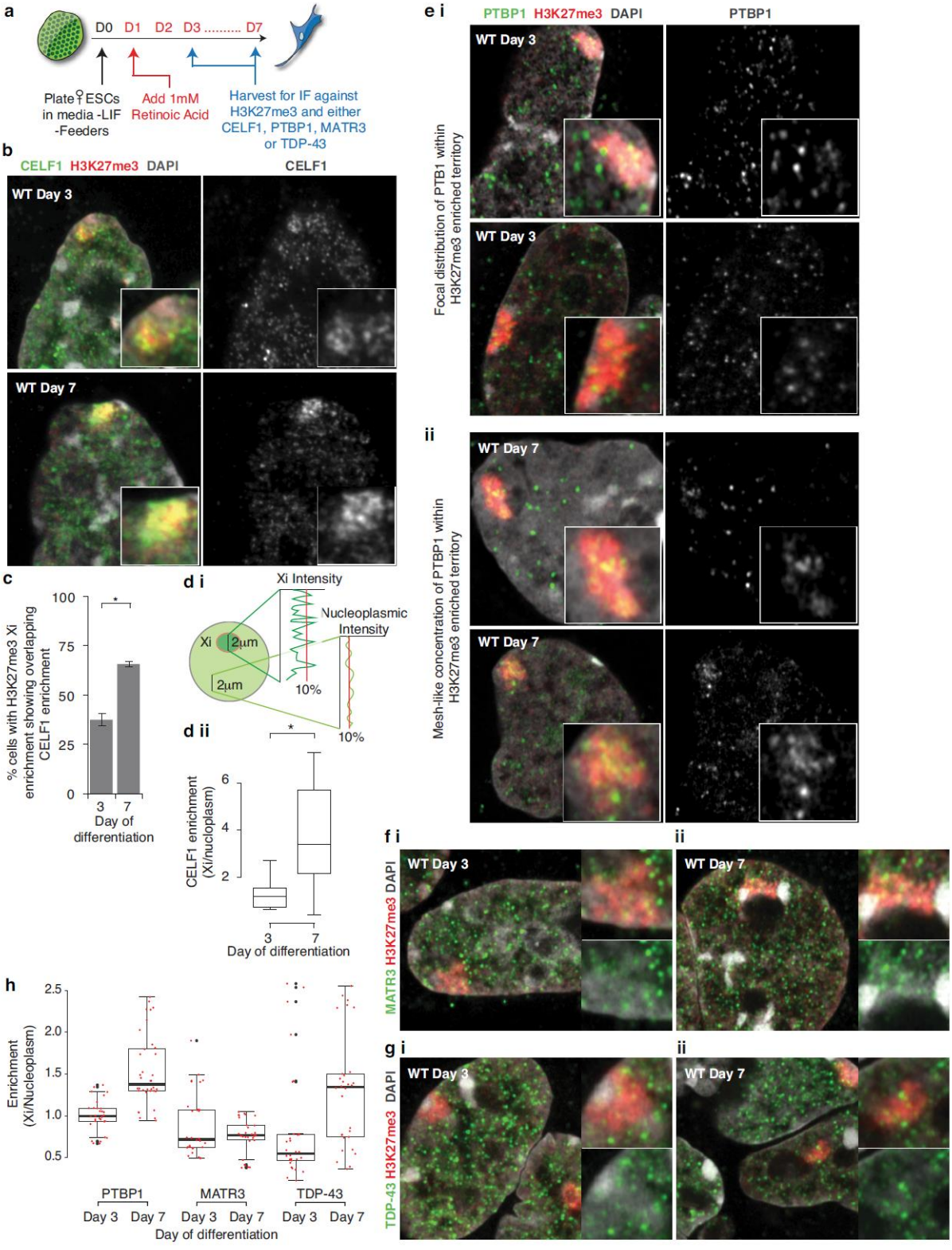
a) (i), Top, diagram of the *Xist* genomic locus. The IVT E-repeat RNA used in **d** is indicated. Bottom, PTBP1, MATR3 and CELF1 i/eCLIP–seq profiles across the *Xist* locus in male tetO-*Xist* ES cells after 6 h of dox induction. CELF1 input profile is shown, read counts indicated on left. (ii), PTBP1 ChIP–seq profiles across the *Xist* locus before or after

20 h of dox treatment in male tetO-*Xist* ES cells. (iii), PTBP1, PTBP2 and TDP-43 iCLIP–seq profiles across the *Xist* locus in the female mouse brain.

b) Table of mapping statistics for PTBP1, MATR3 and CELF1 i/eCLIP–seq data in **a**. Note that *Xist* is overexpressed in this experiment, which influences the number of reads mapping to the locus.

c) (i), The first 1,500 nt of exon 7 of *Xist* are shown, capturing the E-repeat. The sequence remaining after splicing of the *Xist*ΔE transcript is underlined and italicized. The C/U/G tandem repeats within the 5′ half of the E-repeat are indicated (pink-full and blue-truncated repeats) as are the CU-tracts (green) in the 3′ half. Potential TDP-43 sites are indicated in orange. (ii), Alignment of the 25 full C/U/G-tandem repeats (pink) from (i). Brown tracts encode putative PTBP1/MATR3 binding sites, red tracts putative CELF1/TDP-43 binding sites. (iii), Alignment of the nine truncated C/U/G-tandem repeats (blue) from (i). Orange coloured nucleotides are variable within each truncated repeat unit.

d) Left: EMSA of IVT E-repeat RNA (see **a**) and either none, or increasing amounts of rPTBP1 (0, 1.95 nM, 3.9 nM, 7.8 nM, 15.6 nM, 31.3 nM, 62.5 nM, 125 nM, 250 nM, 500 nM, 1 μM and 2 μM). Right, quantification of the bound RNA fraction (dissociation constant, *K*d ≈ 200 nM, from two independent experiments, with s.e.m shown). For source data see Supplementary Fig. 1.


a
D0 D1 D2 D3 ..... D7
Plate ♀ ESCs in media -LIF -Feeders
Add 1mM Retinoic Acid
Harvest for IF against H3K27me3 and either CELF1, PTBP1, MATR3 or TDP-43
b
CELF1 H3K27me3 DAPI
CELF1
WT Day 3
WT Day 7
c
% cells with H3K27me3 Xi enrichment showing overlapping CELF1 enrichment
Day of differentiation
d i
Xi Intensity
Nucleoplasmic Intensity
Xi
2µm
10%
d ii
CELF1 enrichment (Xi/nucloplasm)
Day of differentiation
e i
PTBP1 H3K27me3 DAPI
PTBP1
Focal distribution of PTB1 within H3K27me3 enriched territory
ii
Mesh-like concentration of PTBP1 within H3K27me3 enriched territory
f i
ii
MATR3 H3K27me3 DAPI
g i
ii
TDP-43 H3K27me3 DAPI
h
Enrichment (Xi/Nucleoplasm)
PTBP1
MATR3
TDP-43
Day 3
Day 7
Day of differentiation

Extended Data Fig. 4 CELF1 and PTBP1 localize within the *Xist*-coated territory.

a) Experimental schematic

b) Left, confocal-Airyscan sections of wild-type ES cells at differentiation day 3 and 7, immunostained for CELF1 and H3K27me3. Inset, enlargement of the Xi territory. Right, CELF1 staining in greyscale.

c) Histogram showing the proportion of H3K27me3-marked Xi's with a co-localizing CELF1 enrichment. Error bars indicate s.e.m. (n = 50 from 3 coverslips across 2 independent differentiations); *P < 0.05, two-tailed Student's t-test.

d) (i), Intensity values for CELF1 fluorescence were recorded across a 2 μm line over the Xi (identified on the basis of the H3K27me3 Xi staining) or within the nucleoplasm of the same nucleus in z-stack projections. (ii), Box plot showing the distribution of the ratio between the top 10% CELF1 Xi intensity values compared to the top 10% intensity values from the nucleoplasm

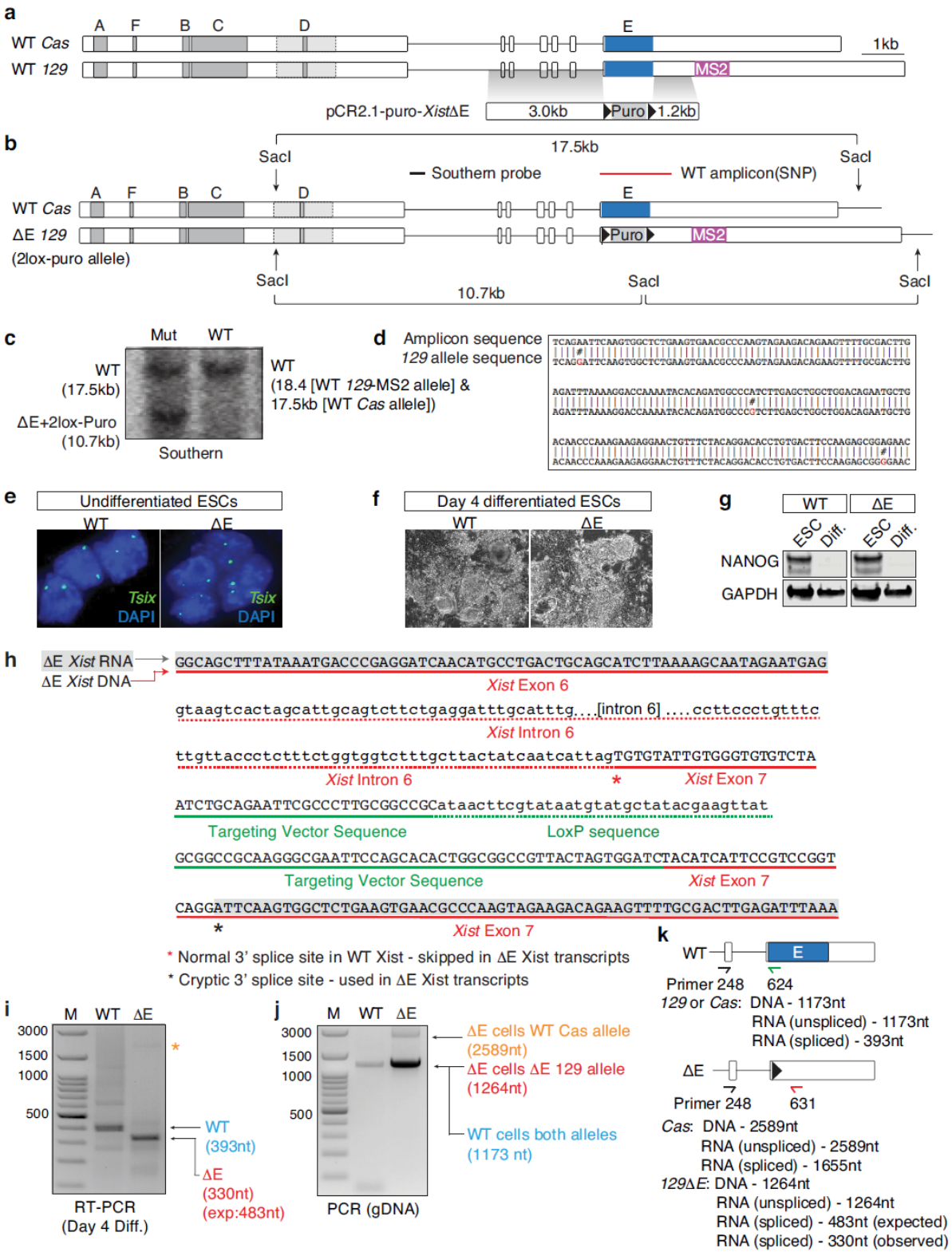
(n = 12, from one experiment); *P < 0.05, two-sample Kolmogorov–Smirnov test.

e) (i), Left, As in **b**, but showing PTBP1 immunostaining at differentiation day 3. Right, PTBP1 staining in greyscale. (ii), As in (**i**), except at differentiation day 7. Note that these images highlight a mesh-like PTBP1 concentration within the Xi observed in a small fraction of cells, distinct from that observed in the nucleoplasm of these cells or from the pattern within the Xi at day 3.

f) (i), As in **e** (i), but showing MATR3 immunostaining and Xi-zoom ins. (ii), As in **e** (ii), except showing MATR3 immunostaining.

g) As in **f**, except for TDP-43.

h) As in **d** (ii), except showing data for PTBP1, MATR3 and TDP-43, (*n* = 5, from one experiment). Red dots, data points for the top 10% Xi/Nucleoplasmic intensity values from 5 cells. For box plots in **d** (ii) and **h**, horizontal lines denote the median, whiskers indicate 1.5× the interquartile range, dots represent outliers.


a
WT Cas
WT 129
A F B C D E
1kb
MS2
pCR2.1-puro-XistΔE
3.0kb
Puro
1.2kb
b
SacI
17.5kb
SacI
Southern probe
WT amplicon(SNP)
WT Cas
ΔE 129
(2lox-puro allele)
Puro
MS2
SacI
SacI
SacI
10.7kb
c
Mut
WT
WT
(17.5kb)
ΔE+2lox-Puro
(10.7kb)
Southern
WT
(18.4 [WT 129-MS2 allele] &
17.5kb [WT Cas allele])
d
Amplicon sequence
129 allele sequence
e
Undifferentiated ESCs
WT
ΔE
Tsix
DAPI
f
Day 4 differentiated ESCs
WT
ΔE
g
WT
ΔE
ESC
Diff.
NANOG
GAPDH
h
ΔE Xist RNA
ΔE Xist DNA
GGCAGCTTTATAAATGACCCGAGGATCAACATGCCTGACTGCAGCATCTTAAAAGCAATAGAATGAG
Xist Exon 6
gtaagtcactagcattgcagtcttctgaggatttgcatttg....[intron 6] ....ccttccctgtttc
Xist Intron 6
ttgttaccctctttctggtggtctttgcttactatcaatcattagTGTGTATTGTGGGTGTGTCTA
Xist Intron 6
*
Xist Exon 7
ATCTGCAGAATTCGCCCTTGCGGCCGCataacttcgtataatgtatgctatacgaagttat
Targeting Vector Sequence
LoxP sequence
GCGGCCGCAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCTACATCATTCCGTCCGGT
Targeting Vector Sequence
Xist Exon 7
CAGGATTCAAGTGGCTCTGAAGTGAACGCCCAAGTAGAAGACAGAAGTTTTGCGACTTGAGATTTAAA
*
Xist Exon 7
* Normal 3' splice site in WT Xist - skipped in ΔE Xist transcripts
* Cryptic 3' splice site - used in ΔE Xist transcripts
i
M WT ΔE
3000
1500
1000
500
*
WT
(393nt)
ΔE
(330nt)
(exp:483nt)
RT-PCR
(Day 4 Diff.)
j
M WT ΔE
3000
1500
1000
500
ΔE cells WT Cas allele
(2589nt)
ΔE cells ΔE 129 allele
(1264nt)
WT cells both alleles
(1173 nt)
PCR (gDNA)
k
WT
E
Primer 248
624
129 or Cas: DNA - 1173nt
RNA (unspliced) - 1173nt
RNA (spliced) - 393nt
ΔE
Primer 248
631
Cas: DNA - 2589nt
RNA (unspliced) - 2589nt
RNA (spliced) - 1655nt
129ΔE: DNA - 1264nt
RNA (unspliced) - 1264nt
RNA (spliced) - 483nt (expected)
RNA (spliced) - 330nt (observed)

Extended Data Fig. 5 | ΔE ES cells undergo differentiation similar to wild-type ES cells and splicing of *Xist*-intron 6 proceeds in the absence of the E-repeat.

a) Homologous recombination strategy used to delete the *Xist* E-repeat in female ES cells.

b) Southern blot strategy with a 5′ external probe for identification of deletion clones.

c) Southern blot (described in **b**) on targeted ES cells with a *loxP*-flanked puromycin cassette in place of the E-repeat on one *Xist* allele.

d) Sequencing analysis (black) of the wild-type *Xist*-PCR amplicon in ΔE cells (red line in **b**). *129*-allele SNPs are shown in red and do not match those in PCR amplicon, confirming E-repeat deletion on the *Xist*MS2(129) allele.

e) *Tsix* RNA FISH on undifferentiated wild-type and ΔE ES cells confirms the presence of two *Tsix* nascent transcription units, used as a proxy to confirm targeted cells maintain two X chromosomes.

f) Bright-field images of wild-type and ΔE cells at day 4 of differentiation, showing that differentiating cells are morphologically similar.

g) Immunoblot of differentiation day 2 wild-type and ΔE cell lysate, showing equal loss of NANOG expression.

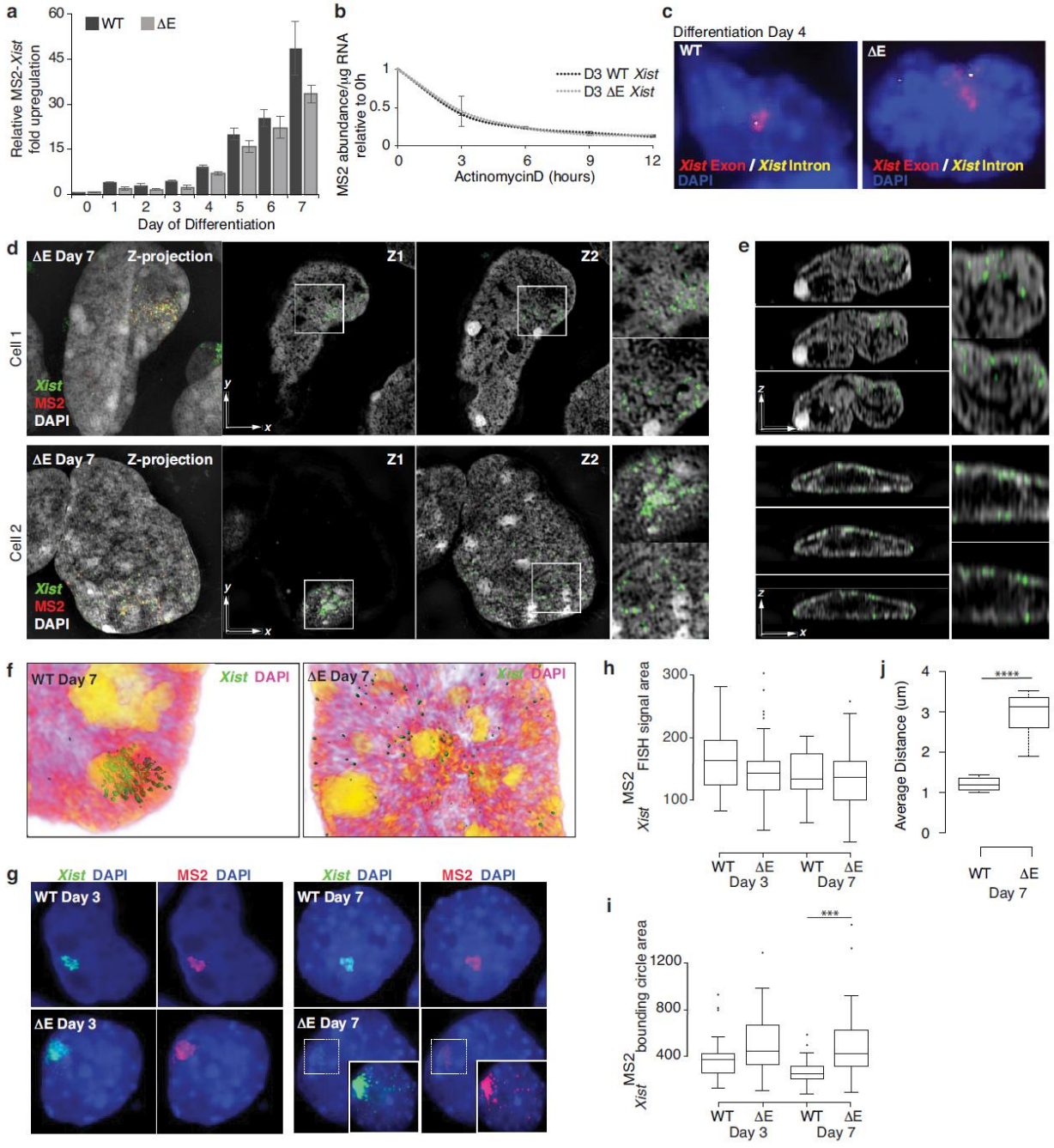
h) Sequence of genomic and cDNA amplicons of the *Xist*ΔE allele after puromycin cassette removal, confirming correct targeting and the use of a cryptic splice site in ΔE cells.

i) Exon 6–7 RT–PCR amplicons generated from RNA isolated from day 4 differentiated wild-type (primers APJ248/624) or ΔE (primers APJ248/631) cells. The ΔE PCR amplicon was shorter than expected. Sequencing revealed a cryptic 3′ splice site downstream of

the *loxP* site that extended the E-repeat deletion within the *Xist* transcript (but not the *Xist* genomic DNA) by 42 nt (see (**h**)).

j) PCR amplicons from wild-type or ΔE genomic DNA using the same primers as in **i**. The intron 6-containing products can be amplified, indicating non-detection of intron 6-containing *Xist* transcripts is not due to amplification problems.

k) Schematic outlining primers used to assess *Xist* DNA and RNA in **i** and **j**. For **c**, **g**, **i** and **j**, see Supplementary Fig. 1 for source data.


a
WT
ΔE
Relative MS2-Xist fold upregulation
Day of Differentiation
b
MS2 abundance/µg RNA relative to 0h
D3 WT Xist
D3 ΔE Xist
ActinomycinD (hours)
c
Differentiation Day 4
WT
ΔE
Xist Exon / Xist Intron
DAPI
d
ΔE Day 7
Z-projection
Z1
Z2
Cell 1
Cell 2
Xist
MS2
DAPI
e
f
WT Day 7
ΔE Day 7
Xist DAPI
g
Xist DAPI
MS2 DAPI
WT Day 3
ΔE Day 3
WT Day 7
ΔE Day 7
h
MS2 Xist FISH signal area
WT ΔE WT ΔE
Day 3 Day 7
i
MS2 Xist bounding circle area
***
j
Average Distance (um)
****

Extended Data Fig. 6 Loss of the E-repeat does not affect *Xist* abundance, splicing or stability.

a) RT–qPCR quantification of the fold upregulation of *Xist*MS2 RNA during differentiation of wild-type or ΔE cells normalized against undifferentiated samples and an internal control (*Rrm2*).

b) RT–qPCR measurements of *Xist*MS2 RNA half-life (upon actinomycin D treatment) at day 3 of differentiation in wild-type or ΔE cells, calculated as MS2 transcript copy number per µg of total RNA. For **a** and **b**, error bars represent the s.e.m. (n = 3, measured in triplicate). Differences were not significant by two-tailed Student's t-test.

c) Epifluorescence images of differentiation day 4 wild-type and ΔE cells probed for exonic regions of *Xist* (red) or *Xist* intron 1 (yellow), and DAPI stained, indicating that the *Xist*ΔE transcripts within the cloud are spliced.

d) Same as Fig. 1h, except two additional *Xist*ΔE-expressing nuclei are shown. Scale bar, 5 µm.

e) Same as Fig. 1h, except for the nuclei in **d**. Note aberrant localization of *Xist*ΔE at the nuclear lamina.

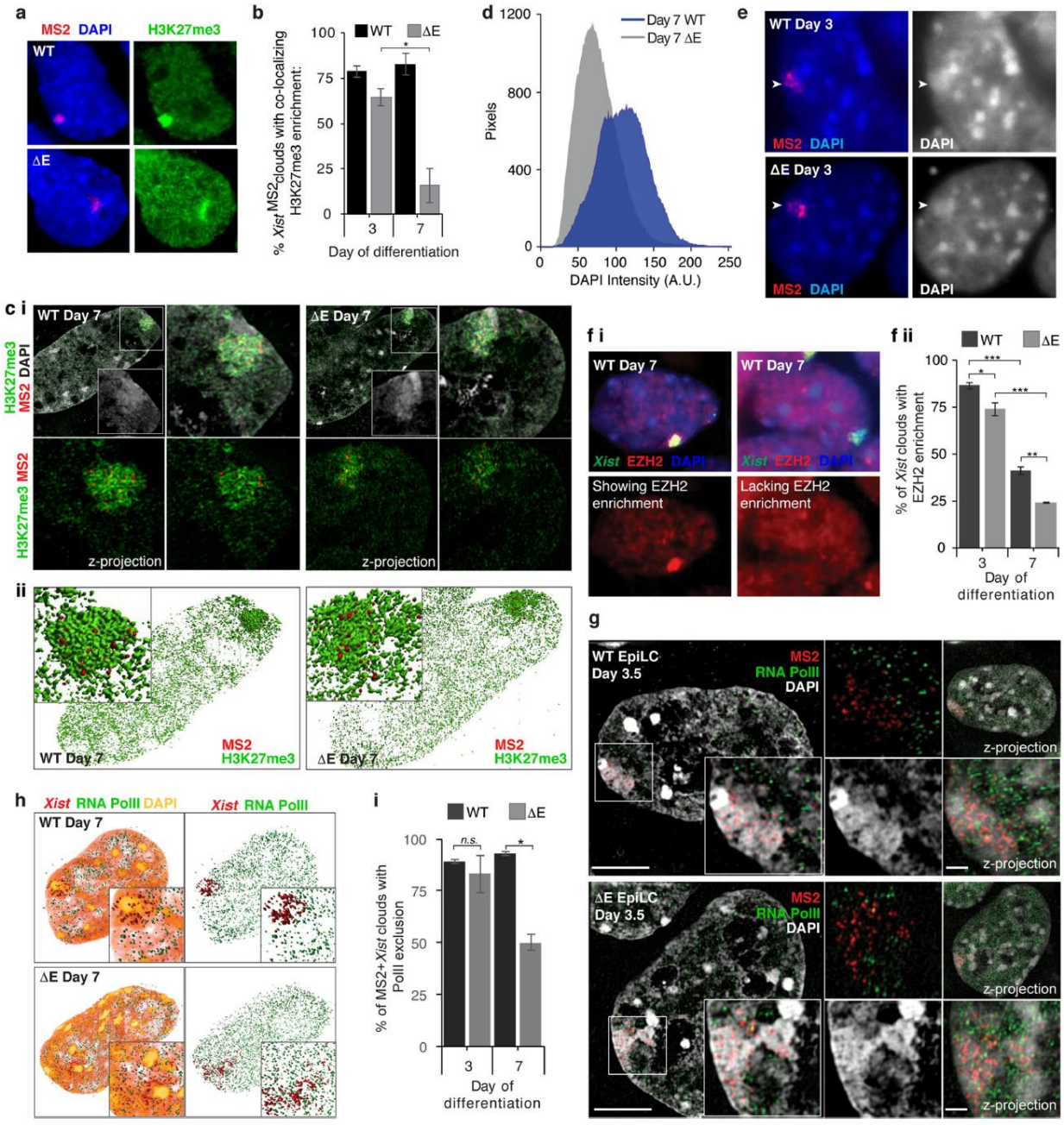
f) 3D Amira reconstructions of the cells shown in Fig. 1h**.**

g) Representative epifluorescence images of RNA FISH against *Xist* and MS2 with DAPI staining for comparison to super-resolution images in **d**, **e** and Fig. 1e, h. Inset, enhanced image of the marked area.

h) Box plot showing the distribution of the area (in pixels) covered by the *Xist* RNA FISH signal, used to calculate the *Xist* aggregation score in Fig. 1f (n = 30, from one experiment).

i) Same as **h** except showing distribution of the bounding circle area, (n = 30, from one experiment); ***P < 0.00005, two-sample Kolmogorov–Smirnov test.

j) Box plot of the average distance between *Xist* foci within *Xist*-MS2 clouds in differentiation day 7 wild-type and ΔE ES cells, as measured by IMARIS. 50 measurements were made per cell, 5 cells per sample; ****P < 0.000005, two-sample Kolmogorov–Smirnov test. For **h**–**j**, horizontal lines denote the median, whiskers indicate 1.5× the interquartile range, dots represent outliers.



Extended Data Fig. 7 | The *Xist*ΔE-coated X chromosome displays decreased DAPI staining and less compact H3K27me3 accumulation at differentiation day 7.

a) Epifluorescence images of cells immunostained for H3K27me3 and probed for MS2. **b)** Quantification of *Xist*MS2 RNA FISH clouds with a co-localizing accumulation of

H3K27me3 at day 3 or 7 of differentiation in wild-type or ΔE cells (*n* = 60/coverslip, 3 coverslips over 2 experiments); **P* = 0.05, two-tailed Student's *t*-test.

c) (i), Top left, 3D-SIM section of wild-type and ΔE cells at differentiation day 7 stained for H3K27me3 and DAPI and probed for MS2. Inset, DAPI staining of marked region. Right, magnification of inset area with (top) or without DAPI (bottom). Bottom left, *Z*-stack projection of inset without DAPI. (ii), 3D Amira reconstruction of images in (i).

d) Graph showing the number of pixels with indicated DAPI fluorescence intensity from *Xist*MS2-expressing X chromosome in wild-type and ΔE cells, masked by H3K27me3 enrichment (*n* = 10, from one experiment).

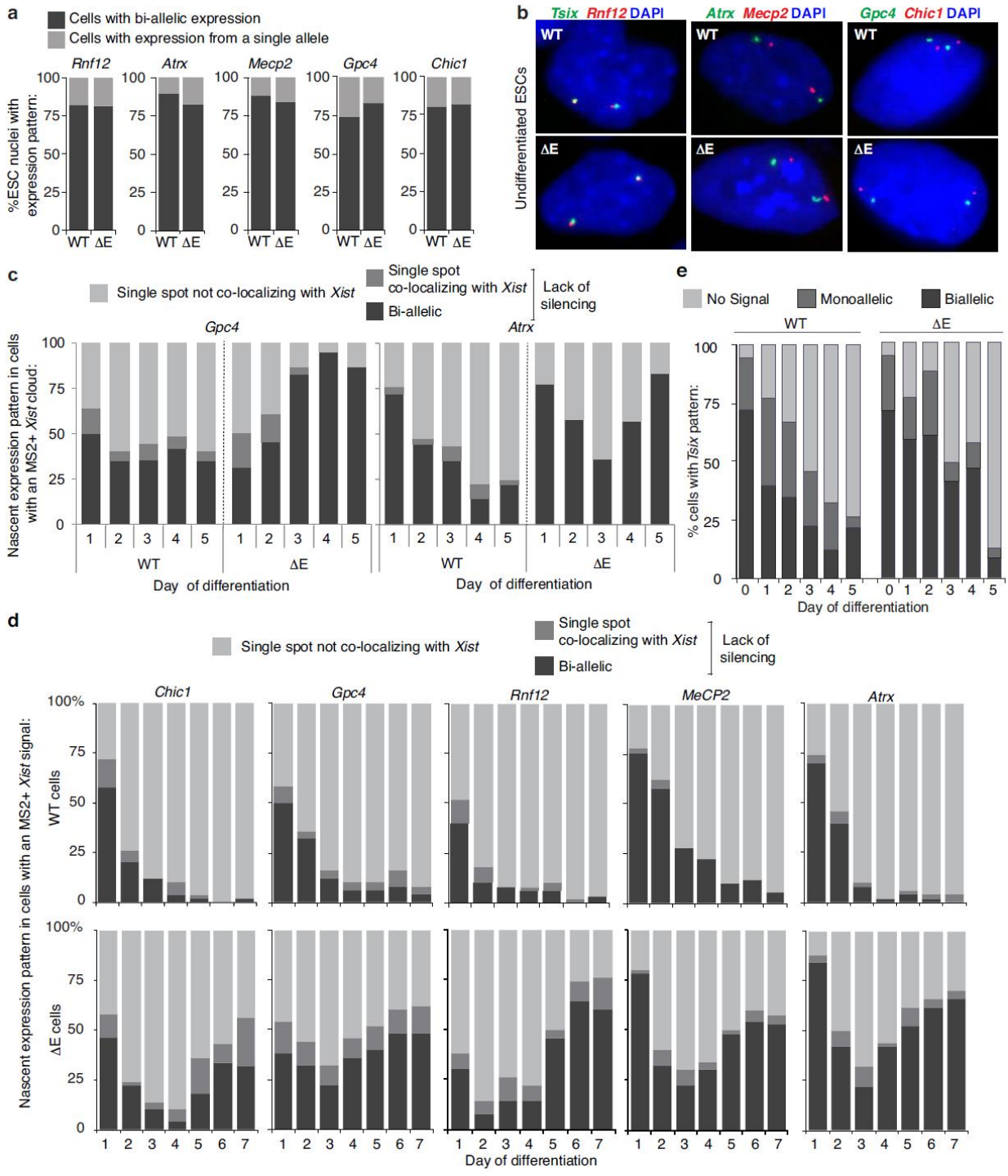
e) Epifluorescence images of wild-type and ΔE cells probed for MS2. Arrowheads point to the *Xist* cloud and highlight the DAPI-bright staining for the X-territory.

f) (i), Epifluorescence images of wild-type cells stained for EZH2 and *Xist*, with (left) and without (right) EZH2 Xi-enrichment at differentiation day 7. (ii), Histogram of the percentage of *Xist* clouds with co-localized EZH2 enrichment (*n* = 60 per coverslip, 3 coverslips from 2 experiments), **P* < 0.05, ***P* < 0.005, ****P* < 0.0005, two-tailed Student's *t*-test.

g) 3D-SIM sections through day 3.5 differentiated wild-type or ΔE ES cells (EpiLC differentiation), immunostained for RNA Pol II and probed for *Xist*, showing exclusion of RNA Pol II from the X-territory. Inset, signals derived from marked area. Small images: top left, same as inset without DAPI; bottom left, same as inset with only DAPI; top right, *Z*-stack projection of the cell; bottom right: *Z*-stack projection of the *Xist*-coated X chromosome. Scale bar, 5 μm; inset, 1 μm.

h) 3D Amira reconstruction of cells in Fig. 2e. Inset, enlargement of the *Xist*MS2-expressing X. Right, same as left without DAPI.

i) Quantification of RNA Pol II exclusion from *Xist*MS2-coated territory (*n* = 50 per coverslip, 2 coverslips from 1 experiment), **P* = 0.05, two-tailed Student's *t*-test.


a
Cells with bi-allelic expression
Cells with expression from a single allele
Rnf12
Atrx
Mecp2
Gpc4
Chic1
%ESC nuclei with expression pattern:
WT ΔE
b
Tsix Rnf12 DAPI
Atrx Mecp2 DAPI
Gpc4 Chic1 DAPI
Undifferentiated ESCs
WT
ΔE
c
Single spot not co-localizing with Xist
Single spot co-localizing with Xist
Bi-allelic
Lack of silencing
Gpc4
Atrx
Nascent expression pattern in cells with an MS2+ Xist cloud:
Day of differentiation
e
No Signal
Monoallelic
Biallelic
% cells with Tsix pattern:
Day of differentiation
d
Single spot not co-localizing with Xist
Single spot co-localizing with Xist
Bi-allelic
Lack of silencing
Chic1
Gpc4
Rnf12
MeCP2
Atrx
WT cells
ΔE cells
Nascent expression pattern in cells with an MS2+ Xist signal:
Day of differentiation

Extended Data Fig. 8 | Loss of the E-repeat prevents continued gene silencing in differentiating ES cells.

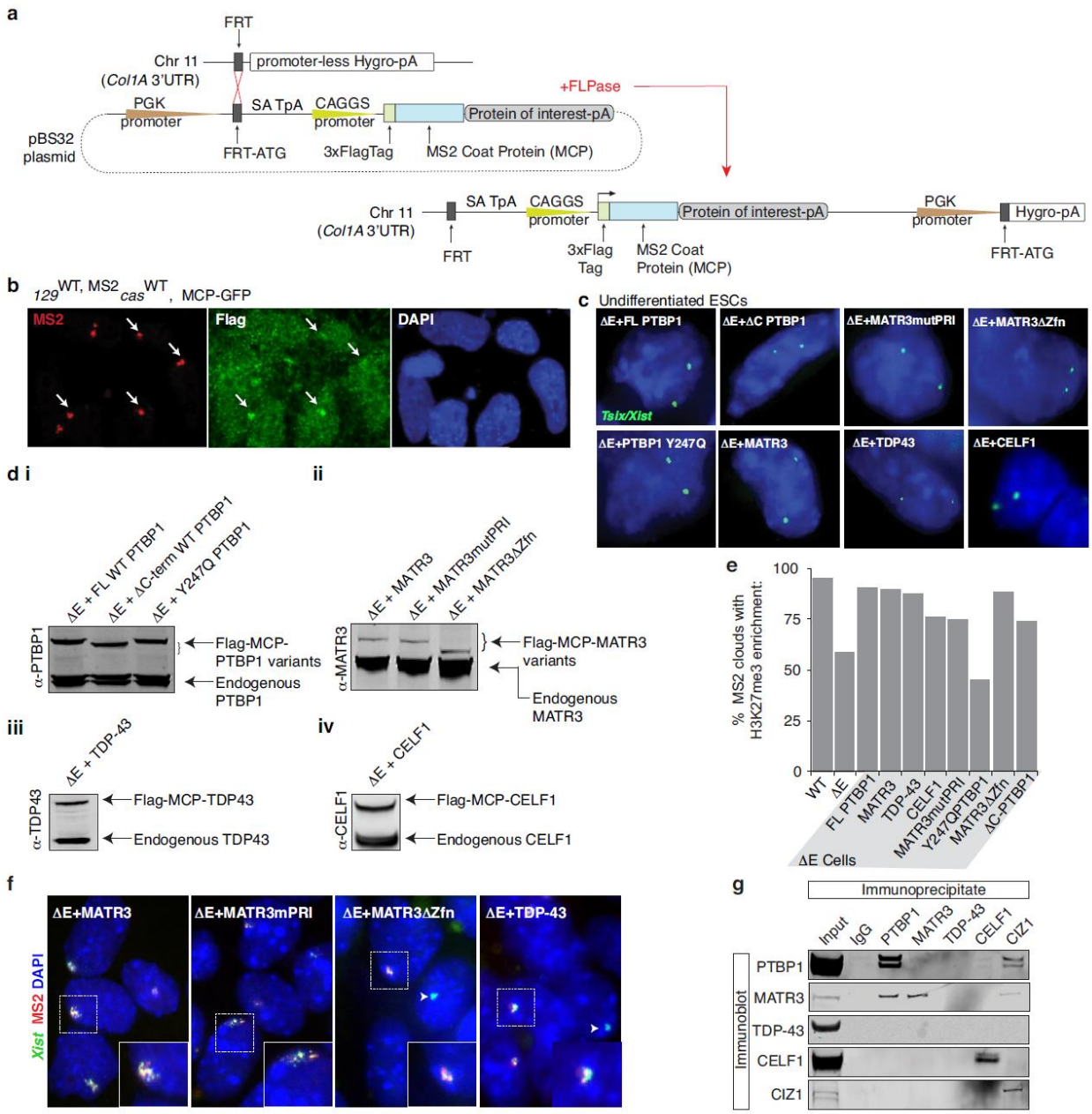
a) Histograms of nascent transcription pattern of indicated X-linked genes (*Rnf12* (*Rlim*), *Atrx*, *Mecp2*, *Gpc4* and *Chic1*) in undifferentiated wild-type and ΔE ES cells, demonstrating that heterozygous deletion of the E-repeat does not interfere with X-linked gene expression in undifferentiated ES cells (*n* = 60, from one experiment).

b) Representative epifluorescence images of cells counted in **a**. *Tsix*, the antisense transcript of *Xist*, was also detected here to identify both X chromosomes. Co-localized foci appear yellow.

c) Histograms of nascent expression patterns of the X-linked genes *Gpc4* and *Atrx* in wild-type and ΔE cells displaying an *Xis t*MS2-coated X chromosome (*n* = 50), across 5 days of differentiation. These data were derived from an independent differentiation from that shown in Fig. 2c.

d) Histograms of nascent expression patterns of indicated X-linked genes in wild-type and ΔE cells displaying an *Xis t*MS2-coated X chromosome (*n* = 50), across 7 days of differentiation derived from an independent differentiation from that shown in **c** and Fig. 2c.

e) Histogram of nascent expression patterns of the X-linked gene *Tsix* in wild-type and ΔE cells across 5 days of differentiation. Note that these data were not scored relative to *Xist*MS2 expression (that is, the monoallelic Tsix signal can be derived from either the *129* or *cas* allele (*n* = 70, except for the ΔE cells at day 5 with only 47 cells counted).



Extended Data Fig. 9 | A site-specific recombination-based approach to rescue phenotypes associated with loss of the E-repeat.

a) Flp-In approach taken to constitutively express Flag-tagged MCP fusion proteins in ES cells (Methods). The Flag–MCP–GFP fusion protein was only expressed in wild-type ES cells. All other rescue constructs were expressed in ΔE ES cells.

b) Flag–MCP–GFP fusion protein recruitment to *Xist*MS2 in wild-type cells at differentiation day 7 shown with representative epifluorescence images. Arrows indicate MS2+*Xist*129 clouds with co-localizing Flag–MCP–GFP enrichment.

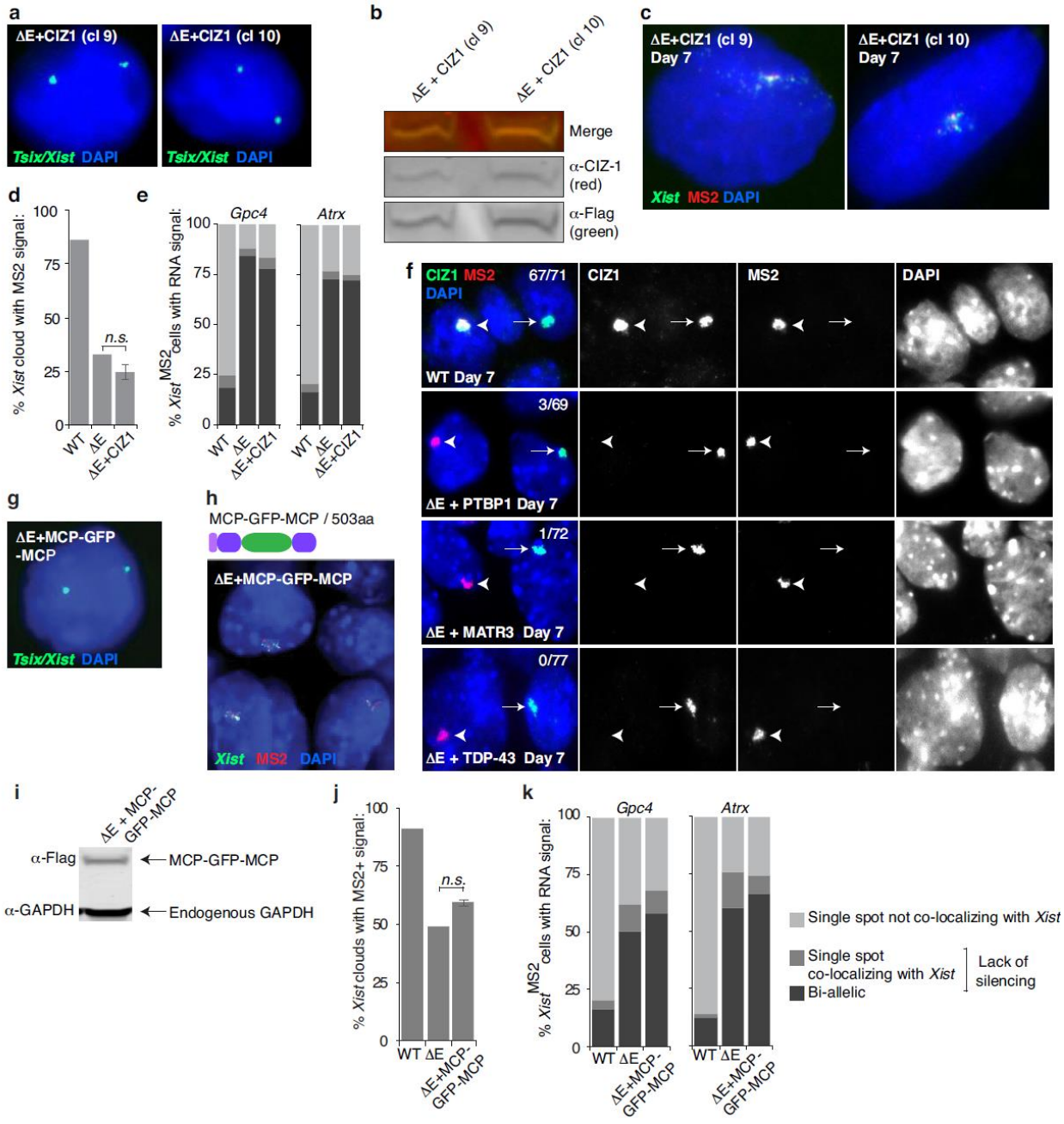
c) *Tsix* expression was used as a proxy to confirm presence of two X chromosomes in rescue ES cell lines.

d) (i), PTBP1-probed immunoblot on lysates from undifferentiated ΔE ES cells expressing full-length MCP–PTBP1 or MCP-PTBP1 mutants. (ii), As in **d** (i) except for MATR3 immunoblot for various MATR3 rescue lines. (iii), TDP-43-probed immunoblot on lysates from undifferentiated ΔE ES cells expressing MCP–TDP-43. (iv), CELF1-probed immunoblot on lysates from undifferentiated ΔE ES cells expressing MCP–CELF1.

e) Histogram of the percentage of *Xist*MS2 clouds that also show enrichment of H3K27me3 in wild-type or ΔE cells, or ΔE cells expressing the indicated MCP-fusion protein at differentiation day 7 (*n* = 80, from one experiment).

f) Representative epifluorescence images of RNA FISH against *Xist* (green) and MS2 (red) in day 7 differentiated ΔE cell lines expressing the indicated variants of MCP fusion proteins. Inset, enlargement of the marked area. Arrowheads indicate wild-type *Xist* clouds in ΔE cells, derived from the *cas* allele.

g) Immunoprecipitation of PTBP1, MATR3, CELF1, TDP-43 and CIZ1 from ES cell nuclear extracts (RNase treated) and detection of co-precipitated proteins with the same antibodies by immunoblotting (to accompany Fig. 3f). For images in **d** and **g**, see Supplementary Fig. 1 for source data.



Extended Data Fig. 10 | Expression of MCP–CIZ1 or MCP–GFP–MCP does not rescue phenotypes due to loss of the E-repeat.

a) RNA FISH images of *Tsix* transcripts for detection of two X chromosomes. Two ΔE MCP–CIZ1 ES cell clones (9 and 10) are shown.

b) Immunoblot result for undifferentiated ΔE ES cell clones expressing MCP–CIZ1.

c) Representative epifluorescence images of day 7 differentiated MCP–CIZ1-expressing ΔE clones, probed for *Xist* and MS2.

d) Proportion of *Xist* clouds also displaying a co-localizing MS2 signal at differentiation day 7. The results for both CIZ1 rescue clones from one experiment were merged and the error bars represent s.e.m (n = 120), P: not significant, two-tailed Student's t-test.

e) Quantification of nascent *Gpc4* or *Atrx* expression patterns in wild-type, ΔE, or ΔE cells expressing MCP–CIZ1 (clone 9) displaying *Xist*MS2 expression, at differentiation day 7 (n = 50, from one experiment). See **k** for legend.

f) Representative epifluorescence images of in wild-type, ΔE or indicated ΔE rescue cell lines at differentiation day 7 immunostained for CIZ1 and probed for MS2. Arrowheads indicate rescued cloud from the ΔE *Xist*MS2 allele. Fraction of MS2+*Xist* clouds showing CIZ1 enrichment is given.

g) RNA FISH images of *Tsix* transcripts in ΔE MCP– GFP–MCP ES cells to demonstrate the presence of two X chromosomes.

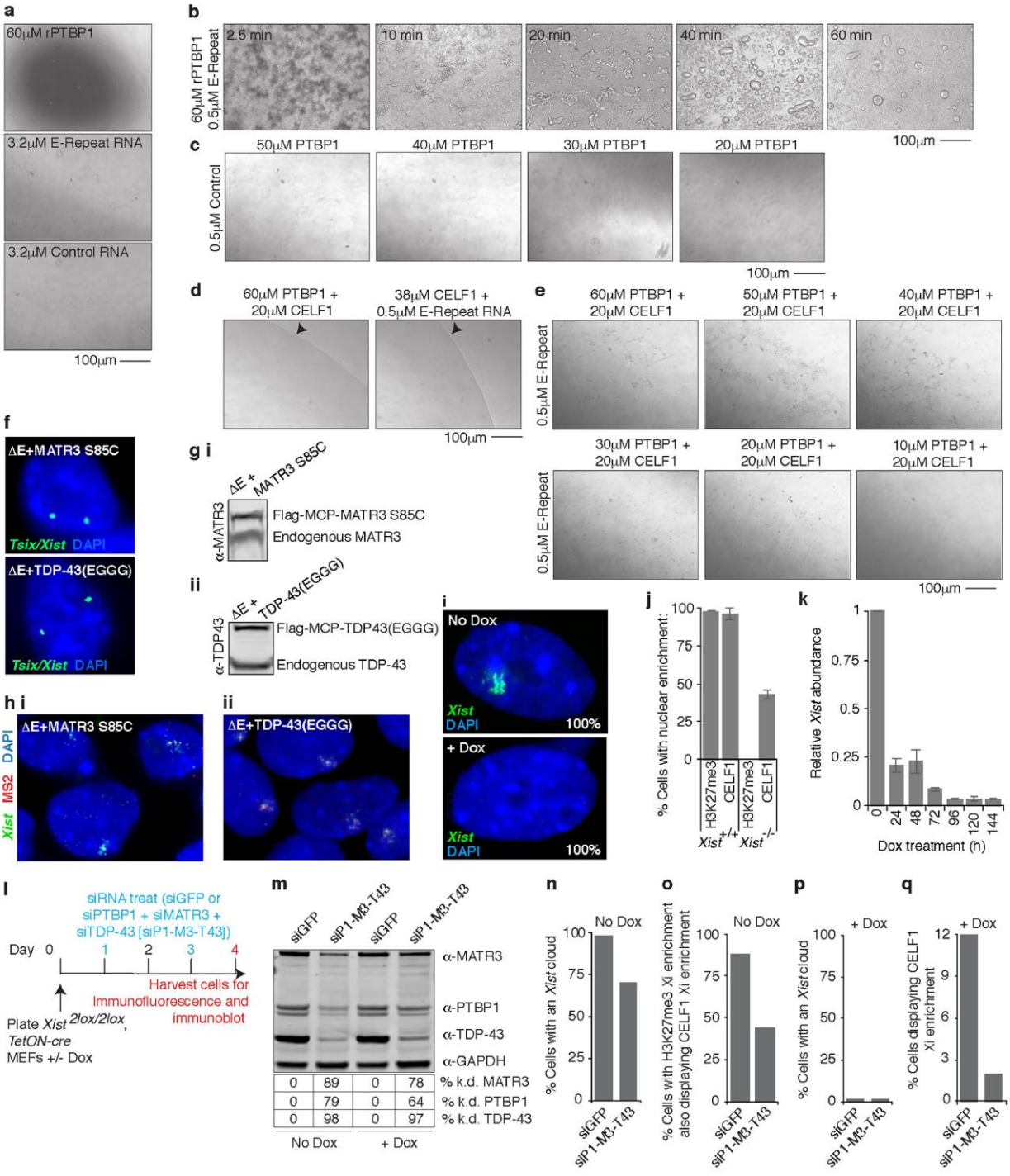
h) Representative epifluorescence images of day 7 differentiated MCP–GFP–MCP-expressing ΔE ES cells probed for *Xist* and MS2, and illustration of Flag-tagged MCP–GFP–MCP fusion protein (see Fig. 3b for key).

i) Immunoblot against the Flag-tag and GAPDH using lysates from undifferentiated MCP–GFP–MCP ΔE ES cells.

j) Histogram showing the proportion of nuclei with *Xist* FISH signal that also displayed a co-localizing MS2 signal at differentiation day 7 for indicated cell lines (n = 100); P: not

significant, two-tailed Student's *t*-test, using 2 independent MCP–GFP–MCP expressing clones from one experiment.

k) Quantification of nascent *Gpc4* or *Atrx* expression patterns in cells displaying *Xist*MS*2* expression at differentiation day 7 (*n* = 50, from one experiment). For images in **b** and **i**, see Supplementary Fig. 1 for source data.


a
60μM rPTBP1
3.2μM E-Repeat RNA
3.2μM Control RNA
100μm
b
60μM rPTBP1 0.5μM E-Repeat
2.5 min
10 min
20 min
40 min
60 min
100μm
c
0.5μM Control
50μM PTBP1
40μM PTBP1
30μM PTBP1
20μM PTBP1
100μm
d
60μM PTBP1 + 20μM CELF1
38μM CELF1 + 0.5μM E-Repeat RNA
100μm
e
0.5μM E-Repeat
60μM PTBP1 + 20μM CELF1
50μM PTBP1 + 20μM CELF1
40μM PTBP1 + 20μM CELF1
30μM PTBP1 + 20μM CELF1
20μM PTBP1 + 20μM CELF1
10μM PTBP1 + 20μM CELF1
100μm
f
ΔE+MATR3 S85C
Tsix/Xist DAPI
ΔE+TDP-43(EGGG)
Tsix/Xist DAPI
g i
ΔE + MATR3 S85C
α-MATR3
Flag-MCP-MATR3 S85C
Endogenous MATR3
ii
ΔE + TDP-43(EGGG)
α-TDP43
Flag-MCP-TDP43(EGGG)
Endogenous TDP-43
i
No Dox
Xist DAPI
100%
+ Dox
Xist DAPI
100%
j
% Cells with nuclear enrichment:
H3K27me3
CELF1
Xist+/+
Xist-/-
k
Relative Xist abundance
Dox treatment (h)
h i
DAPI
MS2
Xist
ΔE+MATR3 S85C
ii
ΔE+TDP-43(EGGG)
l
siRNA treat (siGFP or siPTBP1 + siMATR3 + siTDP-43 [siP1-M3-T43])
Day 0 1 2 3 4
Harvest cells for immunofluorescence and immunoblot
Plate Xist 2lox/2lox, TetON-cre MEFs +/- Dox
m
siGFP
siP1-M3-T43
siGFP
siP1-M3-T43
α-MATR3
α-PTBP1
α-TDP-43
α-GAPDH
0 89 0 78 % k.d. MATR3
0 79 0 64 % k.d. PTBP1
0 98 0 97 % k.d. TDP-43
No Dox
+ Dox
n
No Dox
% Cells with an Xist cloud
siGFP
siP1-M3-T43
o
No Dox
% Cells with H3K27me3 Xi enrichment also displaying CELF1 Xi enrichment
siGFP
siP1-M3-T43
p
+ Dox
% Cells with an Xist cloud
siGFP
siP1-M3-T43
q
+ Dox
% Cells displaying CELF1 Xi enrichment
siGFP
siP1-M3-T43

Extended Data Fig. 11 | CELF1 enhances droplet formation of PTBP1 with the E-repeat in vitro and mutations in MATR3 and TDP-43 that abrogate their self-association do not rescue ΔE phenotypes.

a) Images showing lack of droplets with 60 μM rPTBP1, 3.2 μM E-repeat or control RNA at 40 min.

b) Droplets formed from 60 μM rPTBP1 and 0.5 μM E-repeat RNA over time.

c) Same as **a** except with 0.5 μM control RNA and different concentrations of rPTBP1 (40 min).

d) Same as **a** except with 60 μM rPTBP1 and 20 μM rCELF1, or 38 μM rCELF1 with 0.5 μM E-Repeat RNA. Arrowheads indicate solution boundary with sample on left.

e) Bright-field images showing aggregate-like formations of 20 μM rCELF1, 0.5 μM E-repeat RNA with varied concentrations of rPTBP1.

f) RNA FISH images of *Tsix* transcripts in indicated ES cell lines to show presence of both X chromosomes.

g) (i), MATR3 immunoblot on extracts from ΔE ES cells expressing MCP–MATR3(S85C). (ii), TDP-43 immunoblot on ΔE ES cells expressing MCP–TDP-43(EGGG).

h) Epifluorescence images of day 7 differentiated ΔE cells expressing MCP–MATR3(S85C) (i) or MCP–TDP- 43(EGGG) (ii) probed for *Xist* and MS2.

i) MEFs (*Xist*2lox/2lox, R26M2rtTA/tetO-Cre) probed for *Xist*, before or after dox treatment (96 h). Percentage of cells with displayed *Xist* pattern is given (*n* = 50, two biological replicates).

j) Histogram showing percentage of MEFs with H3K27me3 or CELF1 Xi-enrichment under conditions described in **i**. Error bars represent s.e.m, (*n* = 50, from two biological replicates).

k) Histogram showing relative *Xist* abundance over time of dox treatment for cells in **i** (see Fig. 4g.).

l) Experimental schematic for knockdown experiment in **m**–**q**.

m) Immunoblot showing knockdown of indicated factors in the experiment described in **l**.

n) Percentage of MEFs (no dox) with an *Xist* cloud for indicated knockdowns (*n* = 50, from one experiment).

o) Percentage of MEFs (no dox) with an Xi-enrichment of H3K27me3 that show a co-localizing accumulation of CELF1 (*n* = 50, from one experiment).

p) Same as **n** except with dox treatment.

q) Percentage of MEFs with CELF1 enrichment (*n* = 50, from one experiment). For images in **g** and **m**, see Supplementary Fig. 1 for source data.

Acknowledgements

We thank members of the Plath and Black laboratories for discussions and reading of the manuscript. A.P.-J. was supported by postdoctoral fellowships from the Helen Hay Whitney Foundation and NIH (F32 GM103139); K.P. by Eli and Edythe Broad Centerof Regenerative Medicine and Stem Cell Research (BSCRC) at UCLA, the David Geffen School of Medicine at UCLA, and the Jonnson Comprehensive Cancer Center at UCLA, the NIH (R01 GM115233), and a Faculty Scholar grant from the Howard Hughes Medical Institute; D.L.B. by the NIH (R01 GM049662 and R01 MH109166 (to K.P. and D.L.B.)); M.G. was funded by the New York Stem Cell Foundation, Searle Scholars Program and the Pew-Steward Scholars Program. M.G. is a NYSCF-Robertson Investigator. Y.M., B.P. and S.Z. were supported by the NIH (NICHD 5R03HD095086 to Y.M., R03HD088380 to B.P., R01NS104041 and R01MH116220 to S.Z.). Y.M. and H.L. were supported by the Deutsche Forschungsgemeinschaft (SFB1064/A17 and LE721/18-1). T.C., A.C. and S.S. are supported by graduate fellowships from the Boehringer Ingelheim Foundation (to T.C.); the UCLA Whitcome Fellowship (to A.C.); and the UCLA Broad Stem Cell Research Center – Rose Hills Foundation training award and the UCLA Dissertation Year Fellowship (to S.S.).

Author contributions

K.P., A.P.-J., Y.M. and D.L.B. conceptualized the project and A.P.-J. performed the experiments unless stated otherwise. Y.M. and T.C. performed experiments for 3D-SIM imaging and acquired and analysed 3D-SIM data, overseen by H.L. Y.M. acquired high-resolution images and performed image analysis on immunostained cells. J.S. performed all aggregation measurements, helped with EMSAs and analysed RAP–seq data. R.M., W.M. and A.C. helped to create ES cell deletion lines. S.Z. performed the initial PTBP1/2 iCLIP–seq experiments, A.D. helped A.P.-J. with iCLIP–seq experiments, S.S. and J.S. analysed CLIP–seq data, B.P. and C.C. performed and analysed CHIP–seq experiments, X.-J.W. purified rPTBP1 and rCELF1, and C.-K.C. performed RAP–seq experiments. A.P.-J., J.S., Y.M., T.C. and K.P. analysed data, A.P.-J., Y.M., J.S., M.G. and K.P. interpreted the data and contributed towards methodology and model creation, K.P., D.L.B, M.G. and H.L. acquired funding to support the project, A.P.-J. and K.P. administered the project and A.P.-J. and K.P. wrote the manuscript, including edits from all authors.

References

1. Strom, A. R. & Brangwynne, C. P. The liquid nucleome – phase transitions in the nucleus at a glance. *J. Cell Sci*. 132, jcs235093 (2019).

2. Engreitz, J. M. et al. The *Xist* lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973 (2013).

3. McHugh, C. A. et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232–236 (2015).

4. Minajigi, A. et al. A comprehensive *Xist* interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* 349, aab2276 (2015).

5. Chu, C. et al. Systematic discovery of *Xist* RNA binding proteins. *Cell* 161, 404–416 (2015).

6. Galupa, R. & Heard, E. X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. *Annu. Rev. Genet*. 52, 535–566 (2018).

7. Brockdorff, N. Local tandem repeat expansion in *Xist* RNA as a model for the functionalisation of ncRNA. Noncoding *RNA* 4, 28 (2018).

8. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* 5, 695–705 (2000).

9. Keppetipola, N., Sharma, S., Li, Q. & Black, D. L. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit. Rev. Biochem. Mol. Biol*. 47, 360–378 (2012).

10. Coelho, M. B., Attig, J., Ule, J. & Smith, C. W. J. Matrin3: connecting gene expression with the nuclear matrix. *WIREs RNA* 7, 303–315 (2016).

11. Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A. & Patel, B. K. Molecular mechanisms of TDP-43 misfolding and pathology in amyotrophic lateral sclerosis. *Front. Mol. Neurosci*. 12, 25 (2019).

12. Beisang, D., Bohjanen, P. R. & Vlasova-St. Louis, I. A. in *Binding Protein* (ed. Abdelmohsen, K.) Ch. 8 (InTech, 2012).

13. Pintacuda, G. et al. hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish Polycomb-mediated chromosomal silencing. *Mol. Cell* 68, 955–969.e10 (2017).

14. 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 (2015).

15. Li, P. et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336–340 (2012).

16. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol*. 18, 285–298 (2017).

17. Gallego-Iradi, M. C. et al. N-terminal sequences in Matrin 3 mediate phase separation into droplet-like structures that recruit TDP43 variants lacking RNA binding elements. *Lab. Invest*. 99, 1030–1040 (2019).

18. Plath, K. et al. Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300, 131–135 (2003).

19. Silva, J. et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 Polycomb group complexes. *Dev. Cell* 4, 481–495 (2003).

20. Coelho, M. B. et al. Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. *EMBO J*. 34, 653–668 (2015).

21. Han, A. et al. De novo prediction of PTBP1 binding and splicing targets reveals unexpected features of its RNA recognition and function. *PLoS Comput. Biol*. 10, e1003442 (2014).

22. Marquis, J. et al. CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding. *Biochem. J*. 400, 291–301 (2006).

23. Yamada, N. et al. *Xist* exon 7 contributes to the stable localization of *Xist* RNA on the inactive X-chromosome. *PLoS Genet*. 11, e1005430 (2015).

24. Ridings-Figueroa, R. et al. The nuclear matrix protein CIZ1 facilitates localization of *Xist* RNA to the inactive X-chromosome territory. *Genes Dev*. 31, 876–888 (2017).

25. Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. Repeat E anchors *Xist* RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl Acad. Sci. USA* 114, 10654–10659 (2017).

26. Kunath, T. et al. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* 134, 2895–2902 (2007).

27. Jonkers, I. et al. *Xist* RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. *Mol. Cell. Biol*. 28, 5583–5594 (2008).

28. Pasque, V. et al. X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency. *Cell* 159, 1681–1697 (2014).

29. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res*. 22, 939–946 (2012).

30. Cremer, M. et al. in *The Nucleus* Vol. 463 (ed. Hancock, R.) 205–239 (Humana Press, 2012).

31. Henegariu, O., Bray-Ward, P. & Ward, D. C. Custom fluorescent-nucleotide synthesis as an alternative method for nucleic acid labeling. *Nat. Biotechnol*. 18, 345–348 (2000).

32. Markaki, Y., Smeets, D., Cremer, M. & Schermelleh, L. Fluorescence in situ hybridization applications for super-resolution 3D structured illumination microscopy. *Methods Mol. Biol*. 950, 43–64 (2013).

33. Kraus, F. et al. Quantitative 3D structured illumination microscopy of nuclear structures. *Nat. Protoc*. 12, 1011–1028 (2017).

34. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* 44, 23–28 (2006).

35. Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development* 128, 1275–1286 (2001).

36. Pandya-Jones, A. & Black, D. L. Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15, 1896–1908 (2009).

37. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and maturation of phaseseparated liquid droplets by RNA-binding proteins. *Mol. Cell* 60, 208–219 (2015).

38. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol*. 20, 1250–1257 (2013).

39. Vuong, J. K. et al. PTBP1 and PTBP2 serve both specific and redundant functions in neuronal pre-mRNA splicing. *Cell Rep*. 17, 2766–2775 (2016).

40. Rogelj, B. et al. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep*. 2, 603 (2012).

41. Damianov, A. et al. Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. *Cell* 165, 606–619 (2016).

42. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514 (2016).

43. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10, R25 (2009).

44. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111 (2009).

45. Lovci, M. T. et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol*. 20, 1434–1442 (2013).

46. Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* 168, 442–459.e20 (2017).

47. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet*. 30, 167–174 (2002).

48. Demmerle, J. et al. Strategic and practical guidelines for successful structured illumination microscopy. *Nat. Protoc*. 12, 988–1010 (2017).

49. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682 (2012).

50. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675 (2012).

51. Karperian, A. *FracLac for ImageJ* http://rsb.info.nih.gov/ij/plugins/fraclac/FLHelp/Introduction.htm (1999–2013).

52. Ollion, J., Cochennec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* 29, 1840–1841 (2013).

Chapter 4

Conclusion

Our initial goal for using female mESC differentiation as a XCI model was to study Xist function in its native context of ESC differentiation induced X-inactivation. Previous published work heavily relied on either autosomal transgene insertion or artificial Xist induction models, which does not allow for complete assessment of all steps of XCI and[1-5]. Indeed, using a doxycycline inducible Xist model, we did not observe an XCI phenotype upon F-repeat deletion or E-repeat deletion (data not shown). However, using female mESC differentiation, we were able to characterize the function of both F-repeat and E-repeat, both of which previously had no defined functions. Therefore, it is important that in female mESC differentiation, we can model sequence and signaling requirements for differentiation induced Xist upregulation, as well as accurately capture the transition of initiation phase to the maintenance phase of XCI.

Xist F-repeat is a novel cis-regulatory region required for transcriptional activation of Xist expression

In Chapter 2, we identified the Xist F-repeat as a transcriptional regulatory element critical for Xist upregulation during initiation of XCI. Unlike all the other Xist repeat domains that have already been previously characterized, this is the first Xist repeat domain that has been shown to function at the DNA level rather than the RNA level[1,2,6]. Based on our luciferase data, the Xist F-repeat sequence being an even stronger transcriptional enhancing element than the Xist YY1 sites sequence is particularly striking, as YY1 is universally accepted as a critical regulator of Xist expression. The observation that heterozygous deletion of the F-repeat results in an XCI choice allelic skewing phenotype upon suggests this region may play a role in XCI choice. In addition, we

identified E2F3 as a potential regulator of Xist expression. Since E2F transcription factors are tightly linked to cell cycle, it is possible that cell cycle plays a role in Xist upregulation as well[7]. This discovery adds another dimension to the complexity of Xist regulation. The region where the F-repeat is located at is termed the P2 region, as that region was previously characterized as an alternative promoter for Xist[8]. The P2 regions contains several important binding sites for other Xist regulators such as YY1 and RIF1[9-10]. Therefore, it is possible that E2F3 could work synergistically with YY1 and RIF1 to upregulate Xist expression during XCI initiation. Further experiments will need to be performed to dissect how each of these regulators contribute to Xist regulation, and whether they exhibit any cooperative binding at this region. In summary, our characterization of the Xist F-repeat reveals a novel regulatory mechanism for Xist expression which could contribute to our understanding of the developmentally regulated signaling pathways involved in Xist upregulation during XCI initiation.

Xist E-repeat seeding of condensate formation as a new model for Xist function in X-inactivation

In Chapter 3, our characterization of the Xist E-repeat defines a new model for how Xist establishes the Xi-domain during XCI initiation. Upon induction of differentiation, Xist is upregulated, assembles with proteins across the RNA, and spreads along the X-chromosome[11]. Previous studies have elucidated functions for several of these proteins. SAF-A/hnRNP-U mediates the chromatin attachment of Xist[12], while the proteins SHARP[13], bound at the A-repeat, and PRC1 recruited via hnRNP-K binding to the B-repeat[14], silence transcription. We now define a function for the Xist E-repeat that recruits

the RNA binding proteins PTBP1, MATR3, TDP-43 and CELF1 during the initiation of XCI. These factors each carry multiple RRMs allowing for the simultaneous engagement of distinct repeat motifs within in the long E-repeat sequence, whose multivalency will increase the avidity of binding to a single transcript. Together, these multivalent RNA-protein and protein-protein interactions will form a higher-order Xist–protein network. We propose that increasing Xist abundance with differentiation and, likely, compaction of the X-chromosome[18] concentrates the Xist binding factors within the confined nuclear region of the Xi and result in condensation of PTBP1, MATR3, TDP-43 and CELF1 around the nucleating Xist molecules.

After day 3 of differentiation and XCI initiation, the condensate formed by the E-repeat binding proteins is critical for XCI and sustained silencing of X-linked genes during the Xist independent phase of XCI initiation. At this point, the E-repeat has led to coalescence of Xist transcripts into the Xist granules and compartmentalization of the Xist-coated Xi. By binding and concentrating the factors needed for establishment of the Xi-domain, Xist enforces its own cis-limited spread and resultant gene silencing. The formation of the E-repeat dependent condensate can allow Xist to remain associated solely on the X chromosome from which it is expressed. In the absence of the E-repeat, the loss of gene silencing and the dissociation of Xist from the X-chromosome domain only occur after transcriptional shutoff and heterochromatin formation. We suggest that gene silencing, loss of active transcriptional regulators, and heterochromatinization alter the interaction of Xist with the Xi and induce a transition from an E-repeat-independent mode of association to one that is E-repeat-dependent. Our model also suggests a mechanism for the epigenetic memory that perpetuates the silent state after the inducing

molecule (Xist) has been deleted. We propose that continued gene silencing upon deletion of Xist after day 3 of differentiation is mediated by the E-repeat-seeded protein condensate. This is consistent with our finding that CELF1 enrichment on the Xi can be maintained in the absence of Xist. We hypothesize that the condensate integrates additional Xist-interacting proteins, such as SHARP, via specific protein interactions (Fig. 5h top/bottom right). Weak interactions between these different proteins might permit them to diffuse within the Xi-domain. In this way, Xist-interactors, such as SHARP, maintain association with the multi-molecular assembly independently of direct Xist interaction. Such a model could explain how 100-200 Xist granules (foci)[15,16] can silence >1000 genes across the 167Mb of X chromosome DNA.

A full understanding of how the condensed silencing domain controls gene silencing will involve determining all of its components and their stoichiometry within the Xi as well further biophysical characterization of the condenstate. It will be particularly interesting to examine individual genes within the compartment and whether silenced or escaper genes exhibit different interactions with the condensate or locations within it. Our silencing domain model may also explain the finding that Xist-dependent silencing can only be triggered within a defined developmental window upon onset of XCI. PTBP1, MATR3 and CELF1 are highly expressed in ESCs and decline in abundance upon differentiation[17]. In differentiated cells, the lower levels of these RBPs may be insufficient to multimerize on Xist and condense into a silenced compartment.

Our work provides a new way of thinking about the mechanism of XCI, where the condensation of Xist with its interacting proteins drives the compartmentalization needed for sustained gene regulation. Our results also reveal how RBPs, known for their roles in

RNA processing, mediate lncRNA localization and exert control over gene regulation via mechanisms independent of their previously described RNA processing activities.

References

1. Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet. 2002 Feb;30(2):167-74. PMID:11780141.

2. Pintacuda G, Wei G, Roustan C, Kirmizitas BA, Solcan N, Cerase A, Castello A, Mohammed S, Moindrot B, Nesterova TB, Brockdorff N. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol Cell. 2017 Dec 7;68(5):955-969.e10. PMCID: PMC5735038.

3. Nesterova TB, Wei G, Coker H, Pintacuda G, Bowness JS, Zhang T, Almeida M, Bloechl B, Moindrot B, Carter EJ, Alvarez Rodrigo I, Pan Q, Bi Y, Song CX, Brockdorff N. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. Nat Commun. 2019 Jul 16;10(1):3129. PMCID: PMC6635394.

4. Bousard A, Raposo AC, Żylicz JJ, Picard C, Pires VB, Qi Y, Gil C, Syx L, Chang HY, Heard E, da Rocha ST. The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. EMBO Rep. 2019 Oct 4;20(10):e48019. PMCID: PMC6776897.

5. Colognori D, Sunwoo H, Kriz AJ, Wang CY, Lee JT. Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. Mol Cell. 2019 Apr 4;74(1):101-117. PMCID: PMC6469964.

6. Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Leon WRM, Damianov A, Chronis C, Papp B, Chen CK, McKee R, Wang XJ, Chau A, Sabri S, Leonhardt H, Zheng S, Guttman M, Black DL, Plath K. Publisher Correction: A protein assembly

mediates Xist localization and gene silencing. Nature. 2020 Oct;586(7830):E30. PMID: 33005055.

7. Dimova DK, Dyson NJ. The E2F transcriptional network: old acquaintances with new faces. Oncogene. 2005 Apr 18;24(17):2810-26. PMID: 15838517.

8. Johnston CM, Nesterova TB, Formstone EJ, Newall AE, Duthie SM, Sheardown SA, Brockdorff N. Developmentally regulated Xist promoter switch mediates initiation of X inactivation. Cell. 1998 Sep 18;94(6):809-17. PMID: 9753327.

9. Makhlouf M, Ouimette JF, Oldfield A, Navarro P, Neuillet D, Rougeulle C. A prominent and conserved role for YY1 in Xist transcriptional activation. Nat Commun. 2014 Sep 11;5:4878. PMCID: PMC4172967.

10. Enervald E, Powell LM, Boteva L, Foti R, Blanes Ruiz N, Kibar G, Piszczek A, Cavaleri F, Vingron M, Cerase A, Buonomo SBC. RIF1 and KAP1 differentially regulate the choice of inactive versus active X chromosomes. EMBO J. 2021 Dec 15;40(24):e105862. PMCID: PMC8672179.

11. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES, Plath K, Guttman M. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science. 2013 Aug 16;341(6147):1237973. PMCID: PMC3778663.

12. Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K, Nakagawa S. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. Dev Cell. 2010 Sep 14;19(3):469-76. PMID: 20833368.

13. McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES,

Hess S, Plath K, Guttman M. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature. 2015 May 14;521(7551):232-6. PMCID: PMC4516396.

14. Pintacuda G, Wei G, Roustan C, Kirmizitas BA, Solcan N, Cerase A, Castello A, Mohammed S, Moindrot B, Nesterova TB, Brockdorff N. hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol Cell. 2017 Dec 7;68(5):955-969.e10. PMCID: PMC5735038.

15. Smeets D, Markaki Y, Schmid VJ, Kraus F, Tattermusch A, Cerase A, Sterr M, Fiedler S, Demmerle J, Popken J, Leonhardt H, Brockdorff N, Cremer T, Schermelleh L, Cremer M. Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. Epigenetics Chromatin. 2014 Apr 28;7:8. PMCID: PMC4108088.

16. Sunwoo H, Wu JY, Lee JT. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. Proc Natl Acad Sci U S A. 2015 Aug 4;112(31):E4216-25. PMCID: PMC4534268.

17. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell. 2017 Jan 26;168(3):442-459.e20. PMCID: PMC5302508.