UNIVERSITY OF CALIFORNIA

Los Angeles

# Investigating the Genome-Wide Localization of the *Xist* lncRNA and its Roles in X-Chromosome Dampening and the Formation of the Inactive X-Chromosome Compartment

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Ying Xuan Shawn Tan

2021

# ABSTRACT OF THE DISSERTATION

Investigating the Genome-Wide Localization of the *Xist* lncRNA and its Roles in X-Chromosome Dampening and the Formation of the Inactive X-Chromosome Compartment

by

Ying Xuan Shawn Tan

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2021

Professor Kathrin Plath, Chair

Despite decades of biochemical, proteomic and genetic characterizations of the *Xist* long non-coding RNA, the master regulator of X-chromosome inactivation (XCI), the spatiotemporal kinetics of *Xist* expression and how it coordinates the dynamic recruitment of protein effectors to direct robust gene silencing across the X-chromosome remain unexplored. Furthermore, most studies of *Xist* are conducted in the mouse system, so its role in mediating X-chromosome dampening (XCD), a unique mode of X-chromosome dosage compensation in humans, is still an open question. Therefore, in this thesis, I sought to plug this lack of understanding by (i) elucidating the molecular mechanism by which the chromosome-wide silencing compartment on the inactive X-chromosome (Xi) is formed and maintained across XCI, (ii) determining if *XIST* is responsible for XCD in human embryonic development, and (iii) characterizing the link between the genome-wide localization patterns of *Xist* and its effect on gene silencing.

In Chapter 2, we employed sophisticated super-resolution microscopy and single particle tracking techniques in living mouse embryonic stem cells (ESCs) undergoing XCI to precisely interrogate the spatiotemporal kinetics of *Xist* upregulation and its association with key protein partners that direct various aspects of XCI, and found that *Xist* first forms distinct foci that are locally restricted to about 50 sites across the Xi. Each site contains two *Xist* molecules, which seed the spatial concentration of integral protein interactors into what we term as a supra-molecular complex (SMC), made up of CIZ1 and CELF1 for localization of *Xist* across the Xi, PCGF5 and other components of Polycomb Repressive Complex 1 (PRC1) for inducing heterochromatinization and Xi compaction, and the transcriptional repressor SPEN, which is aided by its intrinsically disordered regions (IDRs) to promote phase separation-mediated aggregation of more SPEN molecules for the faithful and robust maintenance of X-linked gene silencing.

In Chapter 3, we demonstrated the requirement of *XIST* in regulating XCD, as CRISPR-Cas9-mediated ablation of *XIST* in various human pluripotent stem cell (hPSC) lines that model XCD *in vitro* resulted in the loss of transcriptional downregulation on the dampened X-chromosome (Xd). By applying a high throughput genomics method that maps the chromatin binding of *XIST*, we discovered increased *XIST* enrichment on the Xi relative to the Xd, suggesting that *XIST* may be unable to exert complete silencing due to its reduced accumulation on the Xd. Moreover, we characterized a novel function of *XIST*, which can surprisingly propagate to certain autosomal regions beyond the X-chromosome territory, as well as repress some of the autosomal genes that it targets, resulting in a sex

imbalance of the *XIST*-enriched autosomal gene expression levels that is also observed in human pre-implantation embryos *in vivo*.

In Chapter 4, we built on our findings from Chapter 3 and showed that like in hPSCs in which the Xd exhibits a lower enrichment of *XIST* binding, early differentiating mouse ESCs also display diminished levels of *Xist* accumulation on the Xi, accompanied by reduced X-linked gene silencing, compared to the higher levels of *Xist* enrichment in late differentiated cells with complete XCI. Additionally, we characterized the distribution of *Xist* to autosomes in the mouse system, and illustrated how *Xist* spreading to the A- and B-compartments of the genome can be used to predict its differential ability to trigger gene silencing. Furthermore, we highlighted the remarkable ability of *Xist* to enact partial X-linked gene repression in a mouse somatic cell line, which has surpassed the narrow developmental window surrounding pluripotency that was previously established in the XCI field for *Xist*-mediated silencing.

Collectively, our findings gathered from these research projects have greatly expanded on the current knowledge about the molecular landscape established by *Xist* in setting up the Xi compartment, revealed insights on how *XIST* controls both XCI and XCD, and examined the genome-wide localization of *Xist*/*XIST* and its consequent gene silencing dynamics. In addition to addressing several open questions in the XCI field, our knowledge gained from this body of work on the *Xist* lncRNA may be extended to characterize other lncRNAs that contribute to the establishment and/or maintenance of gene regulatory nuclear compartments.

The dissertation of Ying Xuan Shawn Tan is approved.

Douglas Black

Siavash Kurdistani

Steven Jacobsen

Tracy Johnson

Kathrin Plath, Committee Chair

University of California, Los Angeles

2021

# DEDICATION

In dedication to my beloved Heavenly Father, the Author and Perfector of my faith, for guiding me through this journey of learning and discovery.

# TABLE OF CONTENTS

# LIST OF FIGURES

**Chapter 1**



**Chapter 2**

**Chapter 3**

**Chapter 4**

# ACKNOWLEDGMENTS

Embarking on a career in research overseas has truly been one of the most remarkable and rewarding experiences in my life, and I cannot be more thankful for the tremendous amount of help, advice and encouragement I received along this journey.

First of all, I am grateful to my mentor, Kathrin Plath, who has always been very hands-on, patient and approachable. Not only have I learned so much from her through the numerous intellectually stimulating conversations and scientific discussions we had, but she has also challenged me to think critically and independently about my research. I have also benefited immensely from Kathrin's approach of encouraging her students to work in collaborative teams, as this makes it easy to brainstorm ideas, seek advice and troubleshoot experiments with many talented post-docs in the lab.

In particular, I sincerely thank Iris Dror for her hard work in all the bioinformatic analyses she has provided for my work and the numerous fun hikes we did together, Amy Pandya-Jones for her guidance during the initial stage of my PhD journey and her constant encouragement when the going gets tough, Yolanda Markaki for working closely with me and imparting her microscopy expertise, and Konstantinos Chronis for teaching me various genomics techniques and running marathons with me. I also want to give a shout-out to the wonderful company of graduate student friends I have made in the Plath lab, especially Nivedita Damodaren, Tsotne Chitiashvili and Clara Cano for being my

close trio of buddies whom I can comfortably confide in, commiserate and hang out with both within and outside of the lab.

I have also been very blessed to have other outstanding mentors beyond the Plath lab, including my thesis committee: Douglas Black, Siavash Kurdistani, Steven Jacobsen and Tracy Johnson for their advice and guidance throughout my graduate work, the head of the Gene Regulation home area Michael Carey for his care and encouragement since I started my graduate school interview process, my past mentors Bruno Reversade and Lena Ho for inspiring me to pursue this exciting path of research and discovery, and not to forget the endless support from the amicable staff of the UCLA MBIDP, especially Ashley TerHorst, Stephanie Cuellar and Helen Houldsworth.

Last but not least, I want to extend my heartfelt thanks to my close-knit family of church friends and mentors, who truly embodied the concept of “home away from home” for me throughout my PhD journey. Thank you for constantly keeping me in your prayers, especially during the “dark valley” episodes of my life when curveballs flew in my direction and caused me to become disillusioned, and reminding me to keep my faith in God. Thank you for all the love and care you have selflessly showered upon me, and I will forever cherish the memorable times we have shared together as one body united by Christ.

# VITA

## EDUCATION

**University of California, Los Angeles** 2021

Ph.D. Candidate, Molecular Biology (Gene Regulation)

**Johns Hopkins University** 2013

B.S., General Honors, Molecular and Cellular Biology

B.S., Departmental Honors, Cognitive Science

## RESEARCH

**University of California, Los Angeles** 2014-present

Graduate Student Researcher; Advisor: Kathrin Plath

**Agency for Science, Technology and Research (A*STAR), Singapore** 2013-2014

Research Officer; Advisor: Bruno Reversade

**Johns Hopkins University School of Medicine** 2012-2013

Undergraduate Student Researcher; Advisor: Richard L. Huganir

## FELLOWSHIPS AND GRANTS

Eli and Edythe Broad Stem Cell Research Center (BSCRC) Pre-doctoral Training Program 2019

A*STAR National Science Scholarship (Ph.D.) 2014

## PUBLICATIONS

Markaki, Y. *et al.* Xist-seeded nucleation sites form local concentration gradients of silencing proteins to inactivate the X-chromosome. 2021. *In Review. Cell.* 2021

Dror, I. *et al*. XIST controls X chromosome dampening and autosomal genes in early human development. 2021. *In Review. Cell*. 2021

Ho, L. *et al.* ELABELA is an endogenous growth factor that sustains hESC self-renewal via the PI3K/AKT pathway. 2015. *Cell Stem Cell*. 17(4), 435-447. 2015

**PRESENTATIONS**

Tan, S.Y.X. *et al.* Investigating the unique spreading of Xist, the master regulator of X chromosome inactivation, on autosomes. BSCRC Stem Cell Club. Los Angeles, CA. March 2020. *Oral Presentation*. 2020

Tan, S.Y.X. *et al.* The role of Xist, the master regulator of X chromosome inactivation, on autosomes. 16th UCLA Annual Stem Cell Symposium. Los Angeles, CA. February 2020. *Poster*. 2020

Tan, S.Y.X. *et al*. The interplay of Xist, DNA methylation and the Xist-mediated silencing compartment in the maintenance of X chromosome inactivation. UCLA Molecular Biology Interdepartmental Program (MBIDP) Annual Retreat. Lake Arrowhead, CA. September 2019. *Oral Presentation*. 2019

Tan, S.Y.X. *et al*. Spatiotemporal dynamics of the lncRNA Xist in X chromosome inactivation. UCLA MBIDP Graduate Student Recruitment. Los Angeles, CA. January 2019. *Oral Presentation*. 2019

Tan, S.Y.X. *et al*. Spatiotemporal dynamics of Xist cloud formation and maintenance. UCLA MBIDP Student Seminar Series. Los Angeles, CA. October 2018. *Oral Presentation.* 2018

# Chapter 1

Introduction

**Long non-coding RNAs play important roles in gene regulation**

Gene regulation is a complicated process that involves a variety of mechanisms acting on multiple levels. Long non-coding RNAs (lncRNAs) represent a class of RNAs that contain more than 200 nucleotides, and are not translated into proteins (Cabili et al., 2011; Derrien et al., 2012). These RNAs are rapidly emerging as an important class of molecules that transcriptionally regulate gene expression, as loss-of-function studies have implicated lncRNAs in controlling diverse gene regulatory outputs not only in various human cell types (Loewer et al., 2010; Guttman et al., 2011; Sun et al., 2013; Alvarez-Dominguez et al., 2014), but also in other organisms like C. elegans, mouse and zebrafish (Ponting et al., 2009; Nam and Bartel, 2012; Pauli et al., 2012; Delas et al., 2017).

Mammalian genomes encode thousands of lncRNAs (Mattick and Rinn, 2015), some of which act as functional RNAs while others appear to constitute non-functional by-products of genomic regulatory elements like enhancers (Martens et al., 2004; Ebisuya et al., 2008; Latos et al., 2012; Mele and Rinn, 2016). Although the exact percentage of lncRNAs that are functional remains an open question, many of them have been shown to exert essential roles in several biological processes (Guttman and Rinn, 2012; Rinn and Chang, 2012).

Thus far, lncRNAs have been characterized to function in a myriad of gene regulatory contexts, including spatial chromosomal organization (Engreitz et al., 2013; Lu et al., 2020), recruitment of chromatin-modifying proteins for engaging chromatin (Long et al., 2020), and creation of stable nuclear compartments (Kanduri, 2011; Schertzer et

al., 2019; Yao et al., 2019; Pandya-Jones et al., 2020; Valsecchi et al., 2021). However, a lot is still not known regarding how lncRNAs target specific genomic sites and their modes of action. Deciphering the detailed molecular mechanisms by which lncRNAs function is of paramount importance for understanding how they direct cell-type specific gene expression programs that determine cell-fate identities (Percharde et al., 2018) and transitions in development and disease progression.

### *Xist* lncRNA orchestrates X-chromosome inactivation

*Xist* (X-inactive specific transcript), one of the first lncRNAs to be discovered (Brown et al., 1991), is also one of the most well-studied examples of a lncRNA that controls gene regulation. *Xist* directs X-chromosome inactivation (XCI), a crucial developmental process in which one of the two X-chromosomes in female placental mammals is transcriptionally shut-down to enact X-linked dosage balance with males who have only one X-chromosome (Avner and Heard, 2001; Plath et al., 2002; Payer and Lee, 2008; Augui et al., 2011; Deng et al., 2014; Gendrel and Heard, 2014; Galupa and Heard, 2015).

*Xist* serves as an excellent model to investigate the molecular mechanisms of lncRNA function for the following reasons: It is transcribed from the X-chromosome at the onset of XCI and localizes within the nucleus (Brown et al., 1992), creating a silent nuclear compartment (Chaumeil et al., 2006) that distributes across the entire X-chromosome territory to mediate X-linked gene silencing (Clemson et al., 1996; Engreitz et al., 2013; Simon et al., 2013). Importantly, *Xist* ablation results in an inability to initiate XCI (Penny et al., 1996; Marahrens et al., 1997), disruption of *Xist* accumulation on chromatin impairs

X-linked gene silencing (Beletskii et al., 2001; Hasegawa et al., 2010), deletion of one of the most conserved domains of *Xist* also abrogates its silencing capacity (Wutz et al., 2002), and ectopic expression of *Xist* at different sites on the X-chromosome and even on autosomes can induce chromosome-wide repression *in cis* (Lee and Jaenisch, 1997; Wutz and Jaenisch, 2000; Hall et al., 2002a; Kelsey et al., 2015). Collectively, these data highlight the experimental power of *Xist* as a functional lncRNA that is necessary and sufficient for the transcriptional shut-off of X-linked genes.

### *Xist* sets up a heterochromatinized environment for gene silencing

The *Xist* lncRNA does not act on its own to orchestrate XCI, but recruits a plethora of RNA binding proteins (RBPs), transcriptional repressors and chromatin-modifying complexes to form a ribonucleoprotein (RNP) complex that suppresses gene expression on the future inactive X-chromosome (Xi) (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015; Moindrot et al., 2015; Monfort et al, 2015). Some of these protein interactors have been characterized to bind directly to specific conserved domains of the *Xist* RNA (Figure 1). These domains reside within blocks of tandem repeats dubbed A–F repeats (Brockdorff et al., 1992; Brown et al., 1992; Nesterova et al., 2001; Wutz et al., 2002), which comprise unique multimeric short lengths of nucleotides that have distinct functions in XCI, thereby establishing a modular repeat-to-function correlation that better encapsulates the holistic role of *Xist* (Pintacuda et al., 2017b).

XCI is a complex, multi-step process that is triggered by *Xist*, which sets up a stable heterochromatinized environment on the Xi to maintain long-term gene silencing over the

lifetime of the individual. Xi heterochromatin formation consists of the initial erasure of active chromatin modifications such as H3K27ac and H4ac (Keohane et al., 1996; Wakefield et al., 1997; Keohane et al., 1998; Csankovszki et al., 2001; Boggs et al., 2002), which enables accumulation of repressive histone marks like H2AK119Ub and H3K27me3 mediated by Polycomb group (PcG) proteins PRC1 and PRC2, respectively (Plath et al., 2003; Silva et al., 2003; Almeida et al., 2017; Pintacuda et al., 2017a), deposition of the histone variant macroH2A (Costanzi et al., 2000; Mermoud et al., 1999; Rasmussen et al., 2000), and methylation on CpG islands of X-linked genes (Norris et al., 1991; Csankovszki et al., 2001).

Moreover, *Xist* induces major structural remodeling of the Xi, including its folding into a unique three-dimensional (3D) conformation compared to the active X-chromosome (Xa) (Nora et al., 2012; Giorgetti et al., 2016; Wang et al., 2018; Gdula et al., 2019), accompanied by the depletion of topologically associated domains (TADs) (Nora et al., 2012; Giorgetti et al., 2016), chromosome compaction (Chaumeil et al., 2006; Naughton et al., 2010; Nozawa et al., 2013; Smeets et al., 2014), and re-positioning of the Xi to the periphery of the nucleus (Barr and Bertram, 1949; Dyer et al., 1989; Chen et al., 2016), which is closely linked with transcriptional repression (Shevelyov and Nurminsky, 2012; Amendola and van Steensel, 2014).

### *Xist* exploits phase separation for its localization and gene silencing

Liquid-liquid phase separation (LLPS) is a phenomenon that characterizes the partitioning of locally saturated molecules into two distinct phases, and it has been used to describe

the dynamic interaction of nucleic acids and proteins that is integral to the formation of membrane-less cellular structures like stress granules, Cajal bodies and paraspeckles that control various aspects of gene regulation (Dundr and Misteli, 2010; Shevtsov and Dundr, 2011; Banani et al., 2017; Shin and Brangwynne, 2017).

LLPS is favored by proteins harboring intrinsically disordered regions (IDRs), which often contain repetitive amino acid sequences with polar residues that facilitate weak, multivalent protein-protein interactions (Mittag and Forman-Kay, 2007; Uversky, 2015; Banani et al., 2017; Turoverov et al., 2019). Additionally, repetitive nucleotide sequences in RNAs can also act as architectural platforms to recruit and participate in heterotypic interactions with proteins (Jain and Vale, 2017; Cid-Samper et al., 2018; Maharana et al., 2018). Many RBPs harbor IDRs that promote self-aggregation, and together with RNA, help set up high local protein concentrations that give rise to the formation of phase-separated condensates (Lin et al., 2015; Banani et al., 2016). For instance, the lncRNA *Neat1* carries repeat sequences that bind specific proteins to mediate phase separation in paraspeckle formation (Yamazaki et al., 2018).

*Xist* is suspected to play a role in phase separation for Xi formation due to the following reasons: A substantial number of proteins within the *Xist* interactome contains IDRs, some of which are able to participate in RNA-mediated phase separation *in vitro* (Fackelmayer et al., 1994; Li et al., 2012; Banani et al., 2016; Lin et al., 2017; Ying et al., 2017). Several of these *Xist* interactors are also known to contribute to the formation of phase-separated organelles like stress granules and paraspeckles (Cerase et al., 2019).

Moreover, some *Xist* RBPs can bind to and interact with certain repetitive segments of the *Xist* RNA (Cirillo et al., 2016; Lu et al., 2016). Taken together, these findings hint at the propensity of *Xist* to exploit phase separation with its protein partners in mediating compartmentalization and executing chromosome-wide silencing.

In fact, recent work from our lab illustrated the physiological relevance of the *Xist* lncRNA in driving the assembly of a phase-separated compartment on the Xi to ensure the long-term maintenance of gene silencing (Pandya-Jones et al., 2020). Specifically, the E-repeat sequence of *Xist* serves as a scaffold to seed a multi-molecular protein network consisting of the RBPs PTBP1 (Keppetipola et al., 2012), MATR3 (Coelho et al., 2016), TDP-43 (Prasad et al., 2019), and CELF1 (Beisang et al., 2012) (Figure 2a). This higher-order RNA-protein aggregate increases dynamically over time, exhibiting features of LLPS that is required for the persistence of *Xist* localization and transcriptional silencing after *Xist* has first coated the Xi (Figure 2b) (Pandya-Jones et al., 2020).

### *Xist* nucleates the super-stoichiometric enrichment of protein interactors for XCI

Despite our current knowledge of phase separation harnessed by *Xist* for XCI, the precise molecular mechanisms that underpin the formation of this *Xist* E-repeat seeded Xi silencing compartment remains incompletely understood, especially with regard to whether and how certain *Xist* interactors that exert critical known functions in XCI are incorporated within this higher-order protein assembly. Therefore, in Chapter 2 of this thesis, we set out to explore the spatial and temporal dynamics of how *Xist* integrates

various other protein effectors that regulate X-chromosome structure and function within the Xi compartment, using super-resolution microscopy approaches in living cells.

We discovered that the Xi compartment constitutes about 50 locally-restricted granules, each made up of two *Xist* RNA molecules that nucleate key *Xist* interactors with diverse functions, including SPEN, PCGF5, CELF1 and CIZ1, which we collectively term as supra-molecular complexes (SMCs). SPEN is an integral transcriptional repressor of XCI (described in the next section). PCGF5 is a component of PRC1 that has dual roles in gene silencing and chromatin compaction (Almeida et al., 2017; Bousard et al., 2019; Colognori et al., 2019; Nesterova et al., 2019), and is recruited to the *Xist* B- and C-repeats via the RBP hnRNP-K (Pintacuda et al., 2017a). CELF1 and CIZ1 have known roles in constraining *Xist* localization on the Xi, and are recruited by the *Xist* E-repeat (Ridings-Figueroa et al., 2017; Sunwoo et al., 2017; Pandya-Jones et al., 2020). These *Xist*-SMCs are short-lived complexes that concentrate rapidly recycling protein factors within the Xi by raising the affinity of protein binding, culminating in the macromolecular aggregation of heterochromatin-inducing proteins over time to achieve Xi compaction and concomitant gene silencing.

Furthermore, our finding that *Xist* is able to fulfill its transcriptional silencing role with merely ~100 copies of the RNA concurs with previous reports that indicate the presence of 50–200 *Xist* foci residing on the Xi of female somatic cells (Smeets et al., 2014; Sunwoo et al., 2015), but how such a small number of *Xist* molecules can localize

across the vast 167Mb of DNA across the X-chromosome to achieve and maintain robust silencing of more than 1000 genes remained elusive.

Thus, our work in Chapter 2 resolves this puzzle by proposing a novel XCI model in which *Xist*-SMCs within the ~50 locally restricted sites anchor the accumulation of *Xist*-interacting proteins to super-stoichiometric levels, thereby promoting PcG-mediated chromosome compaction and IDR-dependent protein interactions (discussed below) on the Xi. In this way, sub-stoichiometric amounts of *Xist* can still enforce effective chromosome-wide silencing of a relatively greater number of X-linked genes, even in the absence of *Xist* or DNA methylation (Csankovszki et al., 2001; Gendrel et al., 2012). Gene regulation via few *Xist* molecules, coupled with macromolecular crowding of *Xist*-interacting proteins, may represent a more general mechanism by which other lncRNAs, which tend to be lowly expressed compared to other RNAs, establish and maintain nuclear compartments.

### ***Xist*** **recruits the transcriptional repressor SPEN to mediate gene silencing**

The sole protein factor identified from all the genetic and proteomic screens performed thus far (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015; Moindrot et al., 2015; Monfort et al, 2015) is SPEN, also known as SMRT- and HDAC-associated repressor protein (SHARP). SPEN is a large 400kDa protein containing four RNA recognition motifs (RRMs), a nuclear receptor interaction domain (RID), and a highly conserved Spen paralog and ortholog C-terminal (SPOC) domain, which helps to recruit SMRT, the transcriptional co-repressor that activates histone deactylases (HDACs) to mediate gene

silencing (Guenther et al., 2001; Shi et al., 2001; You et al., 2013). Indeed, both SMRT and HDAC3 are required for RNA polymerase II (RNA Pol II) exclusion from the Xi for transcriptional shut-down of X-linked genes (McHugh et al, 2015).

The role of SPEN in XCI was unveiled when it was shown to bind directly to the *Xist* A-repeat (Chu et al., 2015; Lu et al., 2016), which was first identified as the conserved domain on the 5' end of *Xist* that is needed for transcriptional repression (Wutz et al., 2002). SPEN ablation not only abrogated gene silencing (McHugh et al., 2015; Dossin et al., 2020), but also led to a decreased chromosome-wide enrichment of the inactive histone modifications H2AK119Ub and H3K27me3, implying that SPEN contributes to the recruitment of PcG proteins to the Xi (McHugh et al., 2015; Monfort et al., 2015). However, an A-repeat-deficient *Xist* can still recruit both PRC1 and PRC2, and set up a transcriptionally inert milieu on the Xi marked by the absence of RNA Pol II and general transcription factors (Kohlmaier et al., 2004; Chaumeil et al., 2006; Schoeftner et al., 2006; Zylicz et al., 2019). Zylicz et al. subsequently showed that even though *Xist* RNA missing the A-repeat can recruit PRC1 and PRC2, both H2AK119Ub and H3K27me3 accumulation across active genic regions become severely attenuated, indicating that active transcription hinders the enrichment of PcG factors across the Xi (Zylicz et al., 2019). This explains why the loss of SPEN results in lower levels of PcG-dependent H2AK119Ub and H3K27me3 deposition, due to defective gene silencing.

In addition, a recent study described a new mode of action of SPEN as an extension of its silencing ability by demonstrating its binding to the nucleosome

remodeling and deacetylation (NuRD) complex, which itself is a transcriptional repressor typically found at promoter regions to evict RNA Pol II from transcription start sites (Bornelov et al., 2018), during early XCI (Dossin et al., 2020). Hence, SPEN contributes to the epigenetic silencing of the Xi via both the SPEN-SMRT-HDAC3 and SPEN-NuRD axes for enhancer and promoter deacetylation, respectively (McHugh et al., 2015; Zylicz et al., 2019; Dossin et al., 2020).

Finally, another interesting aspect of SPEN is that it contains IDRs (Cerase et al., 2019) that can engage in homotypic interactions (Shin and Brangwynne, 2017) to increase its local protein concentration within the Xi. The *Xist* A-repeat is also found to be able to bind to and recruit multiple SPEN molecules (Lu et al., 2016), thereby enabling multivalent, heterotypic interactions with this critical transcriptional repressor to potentially create a phase-separated RNA-protein assembly for enforcing and sustaining X-linked gene silencing.

Despite these well-characterized roles of SPEN in suppressing X-linked gene expression especially in the initiation phase of XCI, it is unclear whether SPEN is required for stable long-term gene silencing during the maintenance phase of XCI, as well as its molecular mechanisms for doing so. Therefore, in Chapter 2 of this thesis, we address these questions by leveraging on live, single-cell tracking of tagged *Xist* and SPEN molecules, and found that the enrichment of SPEN during the progression of XCI closely mirrors the kinetics of X-linked gene silencing, as the lower SPEN levels observed in early XCI is linked to incomplete silencing, which becomes gradually reversed upon rising

SPEN levels within the Xi across time. Furthermore, we demonstrate the requirement of the IDRs of SPEN for its accumulation within the X-chromosome territory, lending further support to the LLPS model of XCI in generating and efficiently maintaining the silenced Xi compartment (Cerase et al., 2019; Pandya-Jones et al., 2020).

### *XIST* controls X-chromosome dampening

The epigenetic paradigm of X-linked dosage compensation by XCI has been intensively studied in the mouse, where two kinds of XCI are established: imprinted and random XCI (Figure 3a). Imprinted XCI begins at the four- to eight-cell stage of the female mouse pre-implantation embryo, in which the paternal X-chromosome (Xp) is always silenced but not the maternal X-chromosome (Xm) (Takagi and Sasaki, 1975; Wang et al., 2016; Borensztein et al., 2017). As development proceeds to generate the blastocyst, trophectoderm cells that give rise to placental and other extra-embryonic tissues keep their imprinted XCI status, while inner mass cells that contribute to the formation of the embryo proper transiently reactivate the inactive Xp to yield two active X-chromosomes, which undergo a second round of XCI that now becomes a random process, i.e. either the Xp or Xm has an equal chance of being silenced (Mak et al., 2004; Okamoto et al., 2004; Williams et al., 2011). Random XCI is then sustained for the rest of the organism's lifetime, such that all cells of the female adult mouse display mosaicism, expressing either the paternal or maternal copy of X-linked genes (Lyon, 1962; Wu et al., 2014).

On the other hand, in human pre-implantation embryos, the first round of XCI is not imprinted like in the mouse, but both the Xp and Xm stay active from zygotic gene

activation to blastocyst formation (Okamoto et al., 2011; Petropoulos et al., 2016). Instead, genes from both X-chromosomes are downregulated by about half each to align X-linked gene expression levels with that of male embryos, in a recently uncovered dosage compensation process called X-chromosome dampening (XCD) (Petropoulos et al., 2016; Sahakyan and Plath, 2016; Sahakyan et al., 2017). Subsequently during post-implantation development, both embryonic and extra-embryonic cell lineages undergo random XCI (Figure 3b) (Moreira de Mello et al., 2010; Deng et al., 2014). XCD has never been observed in the mouse or in any other mammal, so the molecular basis for this important developmental process is currently poorly understood.

Therefore, in Chapter 3 of this thesis, we aim to unravel the mechanisms that undergird XCD, especially the role *XIST* plays in this process, as several findings point to a strong correlation between *XIST* expression and XCD, including XCD induction being closely tied to *XIST* upregulation from the four-cell to the blastocyst stage (Petropoulos et al., 2016), human female germline cells that exhibit XCD also express *XIST* (Chitiashvili et al., 2020), and imaging data revealed *XIST* spreading on both the dampened X-chromosomes (Xd) in human female pre-implantation embryos and germline cells (Okamoto et al., 2011; Petropoulos et al., 2016; Vallot et al., 2017; Chitiashvili et al., 2020). We found that X-linked gene dampening is abrogated via CRISPR-Cas9-mediated deletion of *XIST* in various naïve human pluripotent stem cell (hPSC) lines that successfully recapitulate XCD *in vitro*, thereby confirming the necessity of *XIST* in controlling both XCD and XCI.

### *Xist* localization across the X-chromosome is governed by 3D spatial proximity

To determine how the upregulation of *Xist* leads to its physical coating across the entire length of the X-chromosome, Engreitz et al. used a high-resolution genomics approach called RNA Antisense Purification, which involves the pull-down of *Xist* RNA with antisense oligonucleotide probes in crosslinked cell lysates, followed by deep sequencing (RAP-seq), to map the chromatin regions bound by *Xist*, and found that *Xist* first spreads to genomic areas on the X that are in close 3D spatial proximity to its locus (Engreitz et al., 2013). Ectopic expression of an *Xist* transgene from a different site on the X-chromosome led to an altered re-wiring of its long-range intra-chromosomal contacts, giving rise to a different *Xist* spreading pattern that confirms the proximity-guided localization technique employed by *Xist* (Engreitz et al., 2013). This also likely explains why *Xist* enriches *in cis* and not onto other chromosomes, as intra-chromosomal DNA domains tend to be closer to one another in physical space, better known as a chromosome territory (Cremer and Cremer, 2001), than inter-chromosomal regions.

In Chapter 3 of this thesis, to further dissect the mechanistic similarities and differences between XCD and XCI orchestrated by the same lncRNA *XIST/Xist*, we applied RAP-seq in various hPSC and human somatic cell lines, and found a lower level of *XIST* enrichment on the Xd relative to the Xi that may account for the diminished silencing in cells undergoing XCD compared to XCI. We made a similar observation from the RAP-seq datasets of mouse cells in Chapter 4, illustrating the overall reduced accumulation of *Xist* on the Xi in early differentiating cells that manifest X-linked gene silencing defects, as opposed to the greater amount of *Xist* binding in late differentiated

cells with complete XCI. These results also shed insight into the elusive uncoupling of *XIST/Xist* expression from complete gene repression, as *Xist* expression has always been thought to be closely associated with X-linked gene silencing.

In addition, we discovered the unanticipated localization of *XIST/Xist* beyond the X-chromosome territory and onto autosomal regions in both human and mouse cells. This finding came as a surprise because of the prevailing view that *Xist* spreading is restricted *in cis* to the X-chromosome from which it is transcribed (Jonkers et al., 2008). Previous studies have demonstrated the ability of *XIST/Xist* to implement silencing in the context of its ectopic expression from an autosome (Lee and Jaenisch, 1997; Wutz and Jaenisch, 2000; Hall et al., 2002a; Jiang et al., 2013; Kelsey et al., 2015), or its distribution to certain autosomal domains in X to autosome (X;A) translocation events (Hall et al., 2002b; Popova et al., 2006; Yang et al., 2011; Loda et al., 2017), but to our knowledge, autosomal spreading by *XIST/Xist* transcribed from its endogenous locus has never been reported.

Moreover, in Chapters 3 and 4 of this thesis, we delineated several well-annotated genomic features that characterize the genome-wide localization phenotypes of *XIST/Xist* in multiple cell types, and uncovered that the spreading pattern of *XIST/Xist* highly correlates with its ability to exert gene silencing. For instance, the distribution of *XIST/Xist* into the transcriptionally active A-compartment versus the repressive B-compartment appears to predict whether X-linked gene silencing is efficient or deficient, respectively. Furthermore, we illuminated the potential biological relevance of the propagation of *XIST/Xist* to chromosomes *in trans*, as *XIST/Xist* can mediate the downregulation of

autosomal genes that it targets in both cultured human and mouse cells *in vitro*, leading to a sex imbalance of the expression levels of a subset of autosomal genes that is also detected in human pre-implantation embryos *in vivo*.

**The repressive role of *Xist* can be modulated in different developmental contexts**

One early seminal study described an under-investigated characteristic of the function of *Xist*, that the capability of discharging its gene silencing role is highly dependent on its developmental context. Specifically, *Xist* can only initiate transcriptional repression within a narrow temporal window at the onset of XCI, between pluripotency and 48 hours of mouse embryonic stem cell (ESC) differentiation (Wutz and Jaenisch, 2000). Beyond this strict developmental timeframe, even cancer cells that erroneously trigger their pluripotency program can no longer induce silencing, except for a small number of cancer cells that display slight repression of genes close to the *XIST* locus (Hall et al., 2002a; Chow et al., 2007; Minks et al., 2013). In another study that used male *Xist* transgenic mice in which silencing of the sole X-chromosome led to cell mortality, only a specific subset of hematopoietic progenitor cells could temporarily re-establish XCI (Savarese et al., 2006), further supporting the prevailing view that *Xist* cannot enforce silencing in differentiated/somatic cells. Lastly, from a phase separation perspective, sharp regulatory transitions like this narrow *Xist*-mediated silencing window are often characteristic of molecular phase-transition events (Hyman et al., 2014).

However, a recent study challenged this long-held view by highlighting the unprecedented ability of *XIST* to enact robust chromosome-wide silencing of one of the

three copies of chromosome 21 that has a tetracycline-inducible *XIST* cDNA transgene incorporated in Down syndrome patient-derived human induced pluripotent stem cell (iPSC)-differentiated neural cells (Czerminski and Lawrence, 2020). To address these conflicting findings, in Chapter 4 of this thesis we assessed the extent of X-linked gene silencing in male *Xist*-inducible ESCs differentiated for up to 12 days, and found that silencing is gradually attenuated past 48 hours of differentiation. We also over-expressed *Xist* in fully-differentiated male *Xist*-inducible mouse embryonic fibroblasts (MEFs) that have surpassed this critical *Xist*-mediated silencing window, and showed that partial repression of X-linked genes can still occur, which is reminiscent of XCD in human pre-implantation embryos. Nonetheless, the extent of this partial X-linked gene suppression cannot be enhanced with longer duration of *Xist* induction in these fibroblastic cells, unlike in differentiated neural stem cells (Czerminski and Lawrence, 2020).

In summary, differentiated cells appear to retain a certain degree of epigenetic plasticity to respond to *XIST/Xist*, overturning the prevailing mechanism that *Xist* only functions within a narrowly defined timeframe surrounding pluripotency (Wutz and Jaenisch, 2000). This remarkable pliability of *XIST/Xist* to modulate gene regulation in various developmental settings holds great promise in pursuing further research to better understand and harness the exciting therapeutic prospects of this lncRNA for treating disorders caused by chromosomal abnormalities, such as duplications, trisomies and aneuploidies (Theisen and Shaffer, 2010).

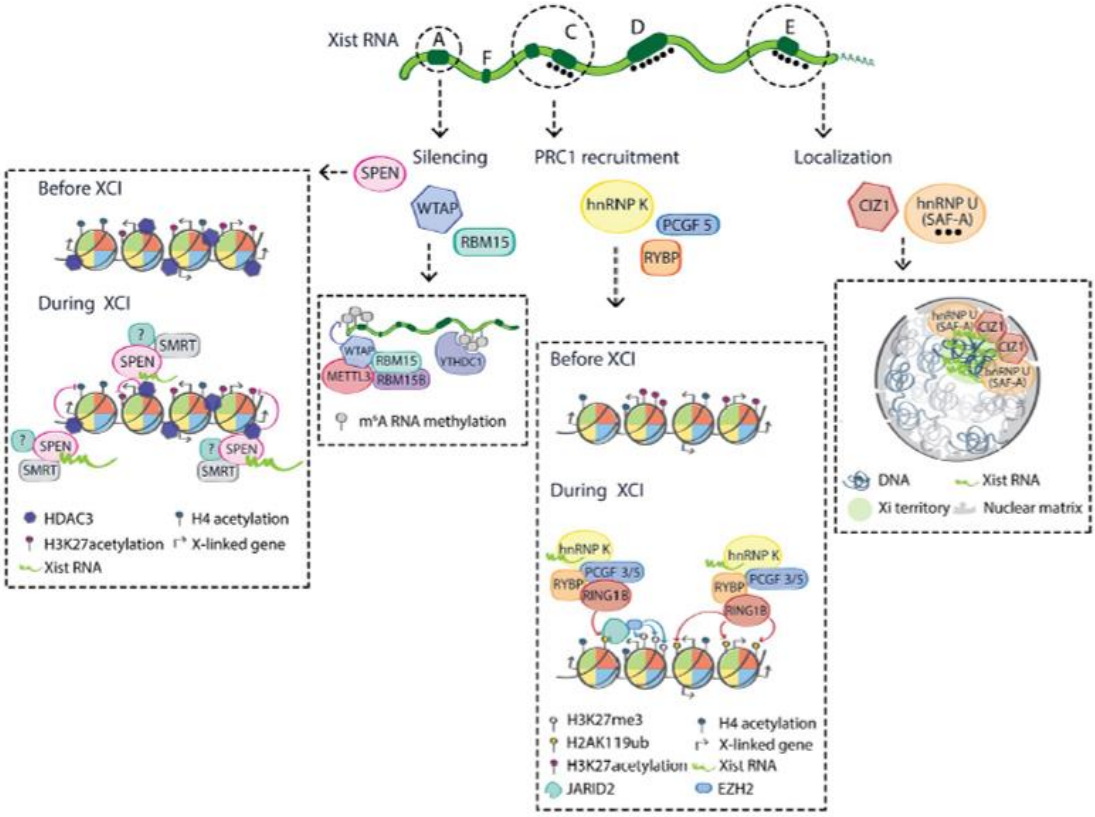



**Figure 1. *Xist* recruits various protein partners via its A–F repeat domains for X-inactivation (Loda and Heard, 2019).**

The *Xist* lncRNA harbors conserved domains of tandem repeat sequences termed A–F repeats, each consisting of multimeric short lengths of nucleotides, which bind and recruit specific protein interactors that mediate different functions in X-chromosome inactivation. Notably, the A-repeat binds the silencing factor SPEN to activate HDAC3 via the SMRT transcriptional co-repressor to induce X-linked gene silencing. The B- and C-repeats bind hnRNP-K to recruit PCGF5 and other components of the Polycomb Group proteins PRC1 to mediate H2AK119Ub deposition, followed by accumulation of another repressive histone mark H3K27me3 via PRC2. The E-repeat binds CIZ1 and hnRNP-U to enable the anchoring and localization of *Xist* across the X-chromosome.

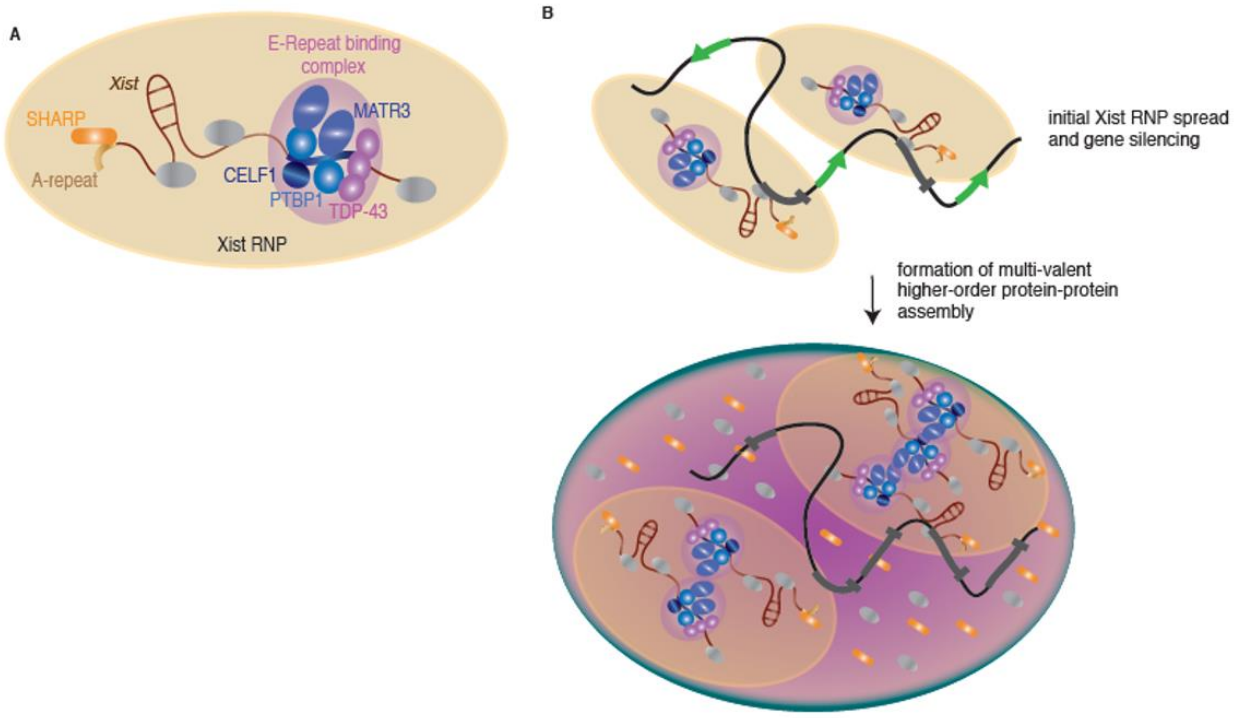



**Figure 2. The *Xist* E-repeat seeds a multi-valent protein assembly for the maintenance of X-linked gene silencing (Pandya-Jones et al., 2020).**

(A) The E-repeat of *Xist* scaffolds the recruitment of the RNA-binding proteins PTBP1, MATR3, TDP-43 and CELF1 to form the *Xist* ribonucleoprotein (RNP) complex. (B) This *Xist* RNP spreads across the X-chromosome to initiate X-linked gene silencing, which is stably maintained over time with the dynamic accumulation of more E-repeat binding proteins, forming a higher-order, macromolecular protein aggregate that exhibit properties of phase separation.

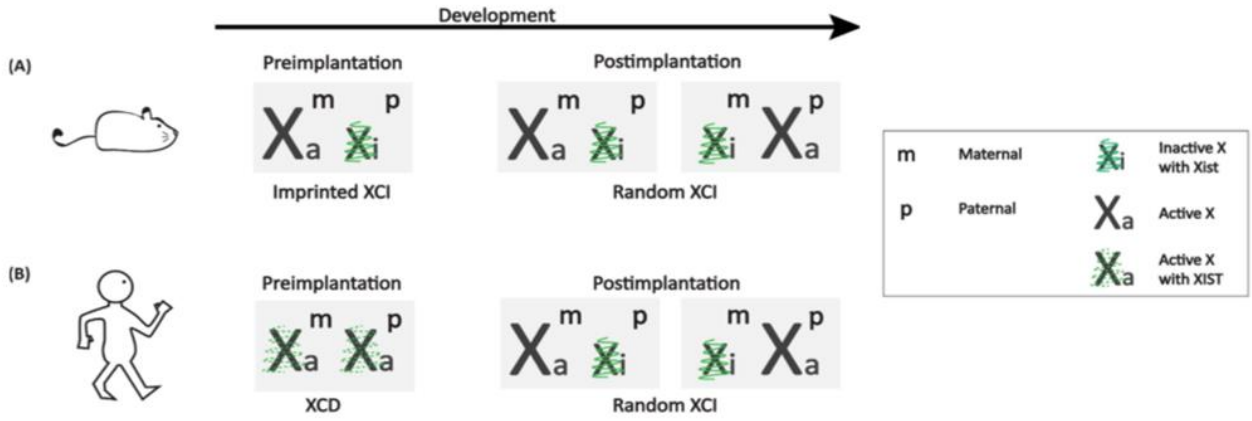



**Figure 3. X-chromosome dosage compensation mechanisms differ between female mouse and human (Sahakyan et al., 2018).**

(A) In mouse, all cells of the female pre-implantation embryo undergo imprinted X-chromosome inactivation (XCI), in which the paternally inherited X (Xp) is always chosen to be silenced while the maternally inherited X (Xm) stays active. Upon implantation, trophectoderm cells that give rise to extra-embryonic tissues like the placenta maintain their imprinted XCI status, while inner mass cells that contribute to the formation of the embryo proper transiently reactivate the inactive Xp, before the second round of XCI takes place, in which either the Xp or Xm has an equal chance of being silenced, in a process called random XCI. (B) In human, all cells of the female pre-implantation embryo undergo X-chromosome dampening (XCD), in which both the Xp and Xm have their X-linked gene expression reduced by about half, but upon implantation, they go through a similar random XCI process as in the mouse.

## REFERENCES


Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., Schermelleh, L., Nesterova, T.B., Koseki, H., and Brockdorff, N. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science *356*, 1081–1084.

Alvarez-Dominguez, J.R., Hu, W., Yuan, B., Shi, J., Park, S.S., Gromatzky, A.A., van Oudenaarden, A., and Lodish, H.F. (2014). Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. Blood *123*, 570–581.

Amendola, M., and van Steensel, B. (2014). Mechanisms and dynamics of nuclear lamina–genome interactions. Curr. Opin. Cell Biol. *28*, 61–68.

Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X inactivation centre. Nat. Rev. Genet. *12*, 429–442.

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

Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K. (2016). Compositional control of phase-separated cellular bodies. Cell *166*, 651–663.

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

Barr, M.L., and Bertram, E.G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. Nature *163*, 676–677.

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

Beletskii, A., Hong, Y.K., Pehrson, J., Egholm, M., and Strauss, W.M. (2001). PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. Proc. Natl. Acad. Sci. *98*, 9215–9220.

Boggs, B.A., Cheung, P., Heard, E., Spector, D.L., Chinault, A.C., and Allis, C.D. (2002). Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat. Genet. *30*, 73–76.

Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.B., Vassilev, I., Galupa, R., Servant, N., Barillot, E., Surani, A., Chen, C.J., and Heard, E., (2017). Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. *24*, 226–233.

Bornelöv, S., Reynolds, N., Xenophontos, M., Gharbi, S., Johnstone, E., Floyd, R., Ralser, M., Signolet, J., Loos, R., Dietmann, S., Bertone, P., and Hendrich, B. (2018). The nucleosome remodeling and deacetylation complex modulates chromatin structure at sites of active transcription to fine-tune gene expression. Mol. Cell *71*, 56–72.

Bousard, A., Raposo, A.C., Zylicz, J.J., Picard, C., Pires, V.B., Qi, Y., Gil, C., Syx, L., Chang, C.Y., Heard, E., da Rocha, S.T. (2019). The role of Xist-mediated Polycomb recruitment in the initiation of X chromosome inactivation. EMBO Rep. *20*, e48019.

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

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

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., and Tartaglia, G.G. (2019). Phase separation drives X-chromosome inactivation: a hypothesis. Nat. Struct. Mol. Biol. *26*, 331–334.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. *20*, 2223–2237.

Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016) Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science *354*, 468–472.

Chitiashvili, T., Dror, I., Kim, R., Hsu, F.M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K., and Clark, A. (2020). Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. Nat. Cell Biol. *22*, 1436–1446.

Chow, J.C., Hall, L.L., Baldry, S.E.L., Thorogood, N.P., Lawrence, J.B., and Brown, C.J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. Proc. Natl. Acad. Sci. *104*, 10104–10109.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

Cid-Samper, F., Gelabert-Baldrich, M., Lang, B., Lorenzo-Gotor, N., Ponti, R.D., Severijnen, L.W.F.M., Bolognesi, B., Gelpi, E., Hukema, R.K., Botta-Orfila, T., and Tartaglia, G.G. (2018). An integrative study of protein-RNA condensates identifies scaffolding RNAs and reveals players in Fragile X-associated tremor/ataxia syndrome. Cell Rep. 25, 3422–3434.

Cirillo, D., Blanco, M., Armaos, A., Buness, A., Avner, P., Guttman, M., Cerase, A., and Tartaglia, G.G. (2016). Quantitative predictions of protein interactions with long noncoding RNAs. Nat. Methods *14*, 5–6.

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

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

Colognori, D., Sunwoo, H., Kriz, A.J., Wang, C.Y., and Lee, J.T. (2019). Xist deletional analysis reveals an interdependency between Xist RNA and Polycomb complexes for spreading along the inactive X. Mol. Cell *74*, 101–117.

Costanzi, C., Stein, P., Worrad, D.M., Schultz, R.M., and Pehrson, J.R. (2000). Histone macroH2A1 is concentrated in the inactive X chromosome of female preimplantation mouse embryos. Development *127*, 2283–2289.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. *2*, 292–301.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J. Cell Biol. *153*, 773–784.

Czerminski, J.T., and Lawrence, J.B. (2020). Silencing trisomy 21 with XIST in neural stem cells promotes neuronal differentiation. Dev. Cell *52*, 294-308.

Delas, M.J., Sabin, L.R., Dolzhenko, E., Knott, S.R., Maravilla, E.M., Jackson, B.T., Wild, S.A., Kovacevic, T., Stork, E.M., Zhou, M., Erard, N., Lee, E., Kelley, D.R., Roth,

M., Barbosa, I.A., Zuber, J., Rinn, J.L., Smith, A.D., and Hannon, G.J. (2017) LncRNA requirements for mouse acute myeloid leukemia and normal differentiation. Elife *6*: e25607.

Deng, X., Berletch, J.B., Nguyen, D.K., and Disteche, C.M. (2014). X chromosome regulation: diverse patterns in development, tissues and disease. Nat. Rev. Genet. *15*, 367–378.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M., Davis, C.A., Shiekhattar, R., Gingeras, T.R., Hubbard, T.J., Notredame, C., Harrow, J., and Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

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

Dundr, M., and Misteli, T. (2010). Biogenesis of nuclear bodies. Cold Spring Harb. Perspect. Biol. *2*, a000711.

Dyer, K.A., Canfield, T.K., and Gartler, S.M. (1989). Molecular cytological differentiation of active from inactive X domains in interphase: implications for X chromosome inactivation. Cytogenet. Genome Res. *50*, 116–120.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. Nat. Cell Biol. *10*, 1106–1113.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., Plath, K., and Guttman, M. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Fackelmayer, F.O., Dahm, K., Renz, A., Ramsperger, U., and Richter, A. (1994). Nucleic-acid-binding properties of hnRNP-U/SAF-A, a nuclear-matrix protein which binds DNA and RNA in vivo and in vitro. Eur. J. Biochem. *221*, 749–757.

Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. Curr. Opin. Genet. Dev. *31*, 57–66.

Gdula, M.R., Nesterova, T.B., Pintacuda, G., Godwin, J., Zhan, Y., Ozadam, H., McClellan, M., Moralli, D., Krueger, F., Green, C.M., Reik, W., Kriaucionis, S., Heard, E., Dekker, J., and Brockdorff, N. (2019). The non-canonical SMC protein SmcHD1

antagonises TAD formation and compartmentalisation on the inactive X chromosome. Nat. Commun. *10*, 30.

Gendrel, A.V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., Giannoulatou, E., Heard, E., Stancheva, I., and Brockdorff, N. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev. Cell *23*, 265–279.

Gendrel, A.V., and Heard, E. (2014). Noncoding RNAs and epigenetic mechanisms during X chromosome inactivation. Annu. Rev. Cell Dev. Biol. *30*, 561–580.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., and Dekker, J. (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575–579.

Guenther, M.G., Barak, O., and Lazar, M.A. (2001). The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. Mol. Cell Biol. *21*, 6091–6101.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn,

J.L., Root, D.E., and Lander, E.S. (2011). LincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Guttman, M., and Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346.

Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., and Lawrence, J.B. (2002a). An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. *99*, 8677–8682.

Hall, L.L., Clemson, C.M., Byron, M., Wydner, K., and Lawrence, J.B. (2002b). Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum. Mol. Genet. *11*, 3157–3165.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. Dev. Cell 19, 469–476.

Hyman, A.A., Weber, C.A., and Julicher, F. (2014). Liquid-liquid phase separation in biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.

Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. Nature *546*, 243–247.

Jiang, J., Jing, Y., Cost, G.J., Chiang, J.-C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., Pearl, J.R., Rebar, E.J., Byron, M., Gregory, P.D., Brown, C.J., Urnov, F.D., Hall, L.L., and Lawrence, J.B. (2013). Translating dosage compensation to trisomy 21. Nature *500*, 296–300.

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

Kanduri, C. (2011). Kcnq1ot1: a chromatin regulatory RNA. Semin. Cell Dev. Biol. *22*, 343–350.

Kelsey, A.D., Yang, C., Leung, D., Minks, J., Dixon-McDougall, T., Baldry, S.E.L., Bogutz, A.B., Lefebvre, L., and Brown, C.J. (2015). Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. Genome Biol. *16*, 208.

Keohane, A.M., O'Neill, L.P., Belyaev, N.D., Lavender, J.S., and Turner, B.M. (1996). X-inactivation and histone H4 acetylation in embryonic stem cells. Dev. Biol. *180*, 618–630.

Keohane, A.M., Lavender, J.S., O'Neill, L.P., and Turner, B.M. (1998). Histone acetylation and X inactivation. Dev. Genet. *22*, 65–73.

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

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol. *2*:E171.

Latos, P.A., Pauler, F.M., Koerner, M.V., Şenergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., Aumayr, K., Pasierbek, P., and Barlow, D.P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469–1472.

Lee, J.T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275–279.

Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., Russo, P.S., Jiang, Q.X., Nixon, B.T., and Rosen, M.K. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

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

Lin, Y., Currie, S.L., and Rosen, M.K. (2017). Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. J. Biol. Chem. *292*, 19110–19120.

Loda, A., Brandsma, J.H., Vassilev, I., Servant, N., Loos, F., Amirnasr, A., Splinter, E., Barillot, E., Poot, R.A., Heard, E. and Gribnau, J. (2017). Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. Nat. Commun. *8*, 690.

Loda, A., and Heard, E. (2019). *Xist* RNA in action: Past, present, and future. PLoS Genet. *15*: e1008333.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., Manos, P.D., Datta, S., Lander, E.S., Schlaeger, T.M., Daley, G.Q., and Rinn, J.L. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. *42*, 1113–1117.

Long, Y., Hwang, T., Gooding, A.R., Goodrich, K.J., Rinn, J.L., and Cech, T.R. (2020). RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. Nat. Genet. *52*, 931–938.

Lu, J.Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., Hong, Y., Percharde, M., Guo, L., Wu, Z., Liu, L., Liu, W., Yan, P., Ramalho-Santos, M., Sun, Y., and Shen, X. (2020). Genomic repeats categorize genes with distinct functions for orchestrated regulation. Cell Rep. *30*, 3296–3311.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., Mesirov, J.P., Cech, T.R., and Chang, H.Y. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. Cell *165*, 1267–1279.

Lyon, M.F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. Am. J. Hum. Genet. *14*, 135–148.

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillen-Boixet, J., Franzmann, T.M., Jahnel, M., Marrone, L., Chang, Y.T., Sterneckert, J., Tomancak, P., Hyman, A.A., and Alberti, S. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science *360*, 918–921.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science *303*, 666–669.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. *11*, 156–166.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571–574.

Mattick, J.S. and Rinn, J.L. (2015). Discovery and annotation of long noncoding RNAs. Nat. Struct. Mol. Biol. *22*, 5–7.

McHugh, C.A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M.J., Shishkin, A.A., Su, J., Lander, E.S., Hess, S., Plath, K., and Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Melé, M., and Rinn, J. L. (2016). 'Cat's cradling' the 3D genome by the act of lncRNA transcription. Mol. Cell *62*, 657–664.

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

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., and Lee, J.T. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*:aab2276.

Minks, J., Baldry, S.E., Yang, C., Cotton, A.M., and Brown, C.J. (2013). XIST-induced silencing of flanking genes is achieved by additive action of repeat a monomers in human somatic cells. Epigenet. Chromatin *6*, 23.

Mittag, T., and Forman-Kay, J.D. (2007). Atomic-level characterization of disordered protein ensembles. Curr. Opin. Struct. Biol. *17*, 3–14.

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. Cell Rep. *12*, 562–572.

Monfort, A., Minin, G.D., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. (2015). Cell Rep. *12*, 554–561.

Moreira de Mello, J.C., Araújo, E.S.S.d., Stabellini, R., Fraga, A.M., Souza, J.E.S.d., Sumita, D.R., Camargo, A.A., and Pereira, L.V. (2010). Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. PLoS ONE *5*:e10947.

Nam, J.W., and Bartel, D.P. (2012). Long noncoding RNAs in C. elegans. Genome Res. *22*, 2529–2540.

Naughton, C., Sproul, D., Hamilton, C., and Gilbert, N. (2010). Analysis of active and inactive X chromosome architecture reveals the independent organization of 30nm and large-scale chromatin structures. Mol. Cell *40*, 397–409.

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., and Zakian, S.M. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res. *11*, 833–849.

Nesterova, T.B., Wei, G., Coker, H., Pintacuda, G., Bowness, J.S., Zhang, T., Almeida, M., Bloechl, B., Moindrot, B., Carter, E.J., Rodrigo, I.A., Pan, Q., Bi, Y., Song, C.X., and Brockdorff, N. (2019). Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. Nat. Commun. *10*, 3129.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Bluthgen, N., Dekker, J., and Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Norris, D.P., Brockdorff, N., and Rastan, S. (1991). Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. Mamm. Genome *1*, 78–83.

Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nat. Struct. Mol. Biol. *20*, 566–573.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644–649.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.P., Renard, J.P., Duranthon, V., and Heard, E. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W.R., Damianov, A., Chronis, C., Papp, B., Chen, C.K., McKee, R., Wang, X.-J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, D.L., and Plath, K. (2020). A protein assembly mediates Xist localization and gene silencing. Nature *587*, 145–151.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., and Schier, A.F. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. *22*, 577–591.

Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. *42*, 733–772.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

Percharde, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-nucleolin partnership regulates early development and ESC identity. Cell *174*, 391–405.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *165*, 1012–1026.

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., and Brockdorff, N. (2017a). hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish Polycomb-mediated chromosomal silencing. Mol. Cell *68*, 955–969.

Pintacuda, G., Young, A.N., and Cerase, A. (2017b) Function by structure: spotlights on Xist long non-coding RNA. Front Mol Biosci*. 4*, 90.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. *36*, 233–278.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell *136*, 629–641.

Popova, B.C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T.B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. Proc. Natl. Acad. Sci. *103*, 7706–7711.

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

Rasmussen, T.P., Mastrangelo, M.A., Eden, A., Pehrson, J.R., and Jaenisch, R. (2000). Dynamic relocalization of histone macroH2A1 from centrosomes to inactive X chromosome during X inactivation. J. Cell Biol. *150*, 1189–1198.

Ridings-Figueroa, R., Stewart, E.R., Nesterova, T.B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., Bageghni, S.A., Albadrani, G., Mansfield, W., Roulson, J.-A., Brockdorff, N., Ainscough, J.F.X., and Coverley, D. (2017). The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. Genes Dev. *31*, 876–888.

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Sahakyan, A., and Plath, K. (2016). Transcriptome encyclopedia of early human development. Cell *165*, 777–779.

Sahakyan, A., Plath, K., and Rougeulle, C. (2017). Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372.

Sahakyan, A., Yang, Y., and Plath, K. (2018). The role of Xist in X-chromosome dosage compensation. Trends Cell Biol. *28*, 999–1013.

Savarese, F., Flahndorfer, K., Jaenisch, R., Busslinger, M., and Wutz, A. (2006). Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. Mol. Cell. Biol. *26*, 7167–7177.

Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J. *25*, 3110–3122.

Shevelyov, Y.Y., and Nurminsky, D.I. (2012). The nuclear lamina as a gene-silencing hub. Curr. Issues Mol. Biol. *14*, 27–38.

Shevtsov, S.P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. Nat. Cell Biol. *13*, 167–173.

Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.C., Hon, M., and Evans, R.M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. Genes Dev. *15*, 1140–1151.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*:eaaf4382.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H.F.M., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell *4*, 481–495.

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465–469.

Smeets, D., Markaki, Y., Schmid, V.J., Kraus, F., Tattermusch, A., Cerase, A., Sterr, M., Fiedler, S., Demmerle, J., Popken, J., Leonhardt, H., Brockdorff, N., Cremer, T., Schermelleh, L., and Cremer, M. (2014). Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. Epigenet. Chromatin *7*, 8.

Sun, L., Goff, L.A., Trapnell, C., Alexander, R., Lo, K.A., Hacisuleyman, E., Sauvageau, M., Tazon-Vega, B., Kelley, D.R., Hendrickson, D.G., Yuan, B., Kellis, M., Lodish, H.F., and Rinn, J.L. (2013). Long noncoding RNAs regulate adipogenesis. Proc. Natl. Acad. Sci. *110*, 3387–3392.

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

Sunwoo, H., Wu, J.Y., and Lee, J.T. (2015). 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. *112*, E4216–4225.

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature *256*, 640–642.

Theisen, A., and Shaffer, L.G. (2010). Disorders caused by chromosome abnormalities. Appl. Clin. Genet. *3*, 159–174.

Turoverov, K.K., Kuznetsova, I.M., Fonin, A.V., Darling, A.L., Zaslavsky, B.Y., and Uversky, V.N. (2019). Stochasticity of biological soft matter: Emerging concepts in

intrinsically disordered proteins and biological phase separation. Trends Biochem. Sci. *44*, 716–728.

Uversky, V.N. (2015). The multifaceted roles of intrinsic disorder in protein complexes. FEBS Lett. *589*, 2498–2506.

Vallot, C., Patrat, C., Collier, A.J., Huret, C., Casanova, M., Ali, T.M.L., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J., and Rougeulle, C. (2017). XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. Cell Stem Cell *20*, 102–111.

Valsecchi, C.I.K., Basilicata, M.F., Georgiev, P., Gaub, A., Seyfferth, J., Kulkarni, T., Panhale, A., Semplicio, G., Manjunath, V., Holz, H., Dasmeh, P., and Akhtar, A. (2021). RNA nucleation by MSL2 induces selective X chromosome compartmentalization. Nature *589*, 137–142.

Wakefield, M.J., Keohane, A.M., Turner, B.M., and Graves, J.A. (1997). Histone underacetylation is an ancient component of mammalian X chromosome inactivation. Proc. Natl. Acad. Sci. *94*, 9665–9668.

Wang, C.Y., Jégu, T., Chu, H.P., Oh, H.J., and Lee, J.T. (2018). SMCHD1 merges chromosome compartments and assists formation of super-structures on the inactive X. Cell *174*, 406–421.

Wang, F., Shin, J., Shea, J.M., Yu, J., Boskovic, A., Byron, M., Zhu, X., Shalek, A.K., Regev, A., Lawrence, J.B., Torres, E.M., Zhu, L.J., Rando, O.J., and Bach, I. (2016). Regulation of X-linked gene expression during early mouse development by Rlim. Elife *5*: e19127.

Williams, L.H., Kalantry, S., Starmer, J., and Magnuson, T. (2011). Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass. Dev. Camb. Engl. *138*, 2049–2057.

Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., Smallwood, P.M., Erlanger, B., Wheelan, S.J., and Nathans, J. (2014). Cellular resolution maps of X chromosome inactivation: Implications for neural development, function, and disease. Neuron *81*, 103–119.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *4*, 695–705.

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

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., and Hirose, T. (2018). Functional domains of NEAT1

architectural lncRNA induce paraspeckle assembly through phase separation. Mol. Cell *70*, 1038–1053.

Yang, C., Chapman, A.G., Kelsey, A.D., Minks, J., Cotton, A.M., and Brown, C.J. (2011). X chromosome inactivation: molecular mechanisms from the human perspective. Hum. Genet. *130*, 175–185.

Yao, R.W., Wang, Y., and Chen, L.L. (2019). Cellular functions of long noncoding RNAs. Nat. Cell Biol. *21*, 542–551.

Ying, Y., Wang, X.J., Vuong, C.K., Lin, C.H., Damianov, A., and Black, D.L. (2017). Splicing activation by Rbfox requires self-aggregation through its tyrosine-rich domain. Cell *170*, 312–323.

You, S.H., Lim, H.W., Sun, Z., Broache, M., Won, K.J., and Lazar, M.A. (2013). Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. Nat. Struct. Mol. Biol. *20*, 182–187.

Zylicz, J.J., Bousard, A., Zumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D., Cramer, P., and Heard, E. (2019). The implication of early chromatin changes in X chromosome inactivation. Cell *176*, 182–197.

# Chapter 2

Research Project

## Xist-seeded nucleation sites form local concentration gradients of silencing proteins to inactivate the X-chromosome

Yolanda Markaki[1*] Johnny Gan Chong[1†], Christy Luong[1†], Shawn Y.X. Tan[1†], Yuying Wang[1], Elsie C. Jacobson[1], Davide Maestrini[2], Iris Dror[1], Bhaven A. Mistry[2,3], Johannes Schöneberg[4], Abhik Banerjee[5,6], Mitchell Guttman[5], Tom Chou[2*], Kathrin Plath[1,7*]

1. Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA 90095, USA

2. Departments of Computational Medicine and Mathematics, University of California Los Angeles, Los Angeles, CA 90095, USA

3. Claremont McKenna College, Claremont, CA, USA

4. Departments of Pharmacology & Chemistry and Biochemistry, University of California San Diego, San Diego, CA, USA

5. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

6. Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

7. Molecular Biology Institute, Jonsson Comprehensive Cancer Center, Brain Research Institute, Graduate Program in the Biosciences, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA 90095, USA

† Equal contribution

*Correspondence: kplath@mednet.ucla.edu, tomchou@ucla.edu and gmarkaki@mednet.ucla.edu

## ABSTRACT

The long non-coding RNA Xist exploits numerous effector proteins to progressively induce gene silencing across the X chromosome and form the inactive X (Xi)-compartment. The mechanism underlying formation of the chromosome-wide Xi-compartment is poorly understood. **Here, we** find that formation of the Xi-compartment is induced by ~50 locally confined granules, where two Xist RNA molecules nucleate supra-molecular complexes (SMCs) of interacting proteins. Xist-SMCs are transient structures that concentrate rapidly recycling proteins in the X by increasing protein binding affinity. We find that gene silencing originates at Xist-SMCs and propagates across the entire chromosome over time, achieved by Polycomb-mediated coalescence of chromatin regions and aggregation, via its intrinsically disordered domains, of the critical silencing factor SPEN. Our results suggest a new model for X chromosome inactivation, in which Xist RNA induces macromolecular crowding of heterochromatinizing proteins near distinct sites which ultimately increases their density throughout the chromosome. This mechanism enables deterministic gene silencing without the need for Xist ribonucleoprotein complex-chromatin interactions at each target gene.

**One Sentence Summary:** Spatial regulation of gene expression in RNA-nucleated compartments

Non-coding RNAs are known to seed membraneless bodies in the cytosol and nucleus such as stress granules, the histone locus body, P-bodies, splicing speckles, paraspeckles or Cajal bodies (*1-3*). Their function relies on the recruitment of effector proteins and often on interactions between low-complexity domains that support Liquid-Liquid Phase Separation (LLPS) (*2, 4-10*) A subset of nuclear RNAs, including the long non-coding RNAs (lncRNAs) Xist, Kcnq1ot1, Rox1/2 and Airn, act by regulating chromatin structure and function through the formation of nuclear compartments (*10-21*). The mechanisms underlying compartment formation by this class of RNAs and how they exploit effector proteins to regulate gene expression and chromatin states within the compartment are still poorly understood. Here, we utilize Xist RNA as a model to interrogate these mechanisms. Our work reveals a spatial organization mechanism by which few RNA molecules can regulate a broad nuclear compartment through the recruitment and local concentration of dynamic effector proteins and provides a quantitative framework for studying such compartments.

Xist is transcribed from, coats and silences one of the two X chromosomes during the development of female mammals in a process referred to as X chromosome inactivation (XCI) (*13, 14, 16-18, 21-26*). Xist localization across the X progressively induces changes in gene expression and alters the three-dimensional chromatin structure (*27-29*) through the recruitment of silencing proteins (*30-32*) and the buildup of epigenetic modifications on the chromosome (*33-39*). The prevailing view is that Xist recruits effector proteins to form ribonucleoprotein complexes that spread across the X and silence genes in a stoichiometric manner (*34, 35, 40-42*). However, super-resolution microscopy has

revealed that Xist distributes in few diffraction-limited foci on the inactive X chromosome (Xi) (*43, 44*). How a small number of Xist foci recruit interacting proteins to deterministically silence genes across an entire chromosome remains unexplored. Here, we analyzed the distribution and dynamics of Xist and its interacting proteins through quantitative and live-cell super-resolution microscopy, and functionally tested key relationships, to derive fundamental insights into this problem. We performed these analyses during the initiation of XCI in female embryonic stem cells (ESCs), at an early time point when Xist RNA has initially established its territory over the X and silencing initiates at first genes, and at a later timepoint when silencing of most genes has been completed. This critical comparison allowed us to observe changes in Xist and/or protein distributions, mediated by few Xist foci, that lead to robust gene silencing over time.

To explore how Xist RNA orchestrates the progressive formation of the Xi-compartment we first examined when Xist coats the X relative to gene silencing in our ESC to epiblast-like cell differentiation system (*45, 46*) (**fig. S1A**). RNA FISH showed that Xist demarcated the X-chromosome and formed a large 'Xist cloud' by day 2 (D2) whereas silencing of *Atrx* and *Mecp2,* two X-linked genes known to be silenced late during XCI initiation, occurred by day 4 (D4) (**Fig. 1A** and **fig. S1, B** and **C**). Thus, consistent with prior data (*47*), silencing progresses after Xist initially coats the X chromosome. We therefore chose to examine the X at D2 and D4 of differentiation to understand the mechanisms leading to progressive Xi-compartment formation. We refer to the X state at D2 as the "pre-XCI" state with the "pre-Xi", and at D4 as the "post-XCI" state with the "Xi".

To determine if the transition from the initial localization of Xist on D2 to the completion of silencing on D4 can be explained by a change in Xist foci number, we used super-resolution three-dimensional Structured Illumination Microscopy (3D-SIM). Previous reports have shown that about 50-200 Xist foci are present in the Xi of female somatic cells (*41, 43*), but whether this number changes during XCI initiation is unknown. To assess Xist foci, we labelled the RNA in living cells by exploiting the MS2 hairpin-MS2 Coat Protein (MCP) interaction (*48*). Briefly, we tagged the endogenous Xist RNA of one of the two X-chromosomes in female mouse ESCs with 24 MS2-repeats and co-expressed MCP-GFP (**Fig. 1B**). MCP-GFP was recruited to Xist ($Xist^{MS2\text{-}GFP}$) and the X-linked gene *Atrx* was efficiently silenced (**fig. S1, D** and **E**), demonstrating the functionality of the $Xist^{MS2\text{-}GFP}$ allele. Quantitative 3D-SIM analysis of $Xist^{MS2\text{-}GFP}$ shows that the Xist territory consists, on average, of 74 diffraction-limited foci on the pre-Xi and 60 foci on the Xi (**Fig. 1, C** and **D**). These distributions were confirmed by RNA FISH with Xist probes (**fig. S2, A** and **B, Text S1**). We found that the doubling of the X chromosome with DNA replication in S-phase is accompanied by the doubling of the number of Xist foci, from ~50 foci in G1 to ~100 in G2, which was confirmed in somatic cells that maintain the Xi (**Fig. 1E** and **fig. S2, C** and **D, Text S2**). Thus, the range of Xist foci number in the cell population is largely due to cell cycle differences. Moreover, these findings show that the number of Xist foci is correlated with the length of the X-chromosome. We confirmed this correlation using cell lines that have abnormal Xis of different length (*49*) (**fig. S2, E** and **F**). Taken together, these data show that transition from the pre-Xi to the Xi is induced by a set number of ~50 Xist foci.

We next investigated whether the amount of Xist RNA in each focus changes during the transition from pre- to post-XCI. We found that Xist foci are stable assemblies that maintain their integrated fluorescence density and volume during differentiation (**Fig. 1F** and **fig. S2B**). Consistent with this finding, the Xist locus is constitutively transcribed during differentiation (**fig. S3, A** and **B**). To estimate the number of Xist molecules in each focus, we utilized a novel fluorescence quantification standard together with quantitative 3D-SIM. Specifically, we transiently expressed nanocages consisting of 60 GFP molecules (cage$^{60GFP}$) (*50*) as internal fluorescence standard in Xist$^{MS2\text{-}GFP}$ cells. Integrated density measurements showed that the amount of fluorescence within one Xist$^{MS2\text{-}GFP}$ focus on the pre-Xi and the Xi, respectively, corresponds to that of one cage$^{60GFP}$ (**Fig. 1G, fig. S3, C** and **D**). Fluorescence fluctuation spectroscopy measurements have shown that ~30 MCP-GFP molecules are bound to 24 copies of the MS2-repeat at a given time (*51*). This number denote that, throughout XCI initiation, each focus contains two molecules of Xist. Thus, only ~100 Xist molecules orchestrate the initiation of gene silencing across the entire X-chromosome.

The X chromosome contains ~1000 genes that are subject to silencing (*52*), therefore, the limited number of Xist foci suggests that they should be highly diffusive if they are to directly regulate target genes across the entire chromosome. We therefore investigated the mobility of Xist foci in the pre- and post-XCI states by performing live-cell 3D-SIM of Xist$^{MS2\text{-}GFP}$, followed by single-particle tracking of individual foci (**movies S1**

and **S2**). This experiment allows for near single-molecule tracking, as each focus contains two Xist transcripts. Unexpectedly, we found that Xist foci exhibited restricted motion and did not undergo fission or fusion (**Fig. 2, A** and **B**). In 90% of cases, the displacement of Xist foci over time was less than 200 nm and their movement was characterized as diffusion in a local confining potential (**Fig. 2, C** to **E** and **Text S3**). The confined motion of Xist foci is highly correlated with the motion of chromatin (*53-57*) both on the pre-Xi and the Xi (**Fig. 2E**). We conclude that Xist foci are tethered to chromosomal locations with high affinity, which constrains each of them locally and limits their movement to that of the Brownian motion of the bound chromatin. Thus, progression of XCI is mediated through ~50 sites where two Xist molecules are locally confined.

To investigate whether the 'wiggling' of Xist foci around their centers is confined within a specific chromatin environment, we introduced a histone H2B-Halo transgene into $Xist^{MS2\text{-}GFP}$ ESCs and performed live-cell 3D-SIM (**Fig. 2F** and **movie S3**). H2B signals were segmented into seven intensity levels that correspond to chromatin density classes, where class 1 represents DNA-free space (interchromatin channels, IC) and classes 2 to 7 increasing chromatin densities (*43*) (**Fig. 2G**). Xist foci covered predominantly classes 1 to 3 (**fig. S4A**). Over time, the chromatin densities underlying each focus footprint never surpassed class 3, and the centers of mass (centroids) of Xist foci remained within chromatin class 2 (**Fig. 2H**). These data are consistent with our finding that chromatin density increases in linear increments (**fig. S4B**). We infer that Xist foci are spatially confined to the periphery of dense chromatin domains, facing the interchromatin channels, and stably maintain their positions relative to chromatin over

time. In agreement with these observations, RNA antisense purification (RAP) of Xist from pre- and post-XCI stages followed by DNA sequencing of the associated chromatin (*40*) showed that Xist localizes to gene-rich, open chromatin regions of the A-compartment (*28, 40, 58*) (**Fig. 2I**). The Xist localization patterns are very similar between the pre-Xi and Xi (Pearson's correlation r=0.83) (**Fig. 2I**). We identified 65 and 63 highly correlated peaks of Xist enrichment on the pre-Xi and Xi, respectively (**Fig. 2J** and **fig. S4, C** and **D**), similar to the number of Xist foci detected by 3D-SIM. Taken together these findings confirm that the localization of Xist relative to chromatin persists as gene silencing proceeds.

The discovery of only two Xist molecules in ~50 confined locations indicates that Xist and effector proteins cannot distribute across the entire chromosome space to induce XCI in a stoichiometric manner with gene targets. We reasoned that to effect gene silencing over the entire chromosome, hundreds of Xist-recruited proteins form diffuse clouds localized about Xist foci. These overlapping diffuse clouds can span the entire X and thereby more frequently interact with silencing sites. We addressed this hypothesis by first examining how Xist-interacting proteins accumulate relative to Xist foci, during the initiation of XCI.

We focused on SPEN, PCGF5, CELF1 and CIZ1, four proteins that bind to distinct repeat sequences of Xist RNA and differ in their function in XCI (*59, 60*). SPEN binds the A-repeat sequence of Xist and is the key transcriptional repressor of XCI that activates

HDAC3 to induce histone deacetylation and gene silencing (*30, 31, 42, 61-63*). PCGF5 is recruited to the Xi via binding of hnRNP-K to the Xist B/C-repeat sequences (*64*), and is a component of the polycomb complex PRC1 that contributes to the silencing of X-linked genes (*35, 63-67*) and has a critical role in chromatin compaction genome-wide (*68-71*). CELF1 and CIZ1 both bind to the E-repeat sequence of Xist and are critical for restricting the localization of Xist in the X-territory (*72-75*).

We imaged antibody-stained and stably expressed Halo-fusion proteins together with $Xist^{MS2\text{-}GFP}$ by 3D-SIM to simultaneously detect Xist and two effector proteins at differentiation D2 and D4 (**fig. S5A**). Endogenous proteins detected by antibody staining or Halo-fused transgenes displayed similar distributions (**fig. S5B**). We found that the interrogated proteins formed distinctive assemblies in proximity to Xist foci on the pre-Xi as well as the Xi (**Fig. 3A**). Moreover, protein assemblies appeared larger in the pre-Xi and Xi-territory than in other nuclear accumulations, indicating that Xist induces the *de novo* formation of unique protein complexes. To quantitatively define the protein aggregates induced by Xist, we extracted the spatial coordinates of thousands of diffraction-limited segmented protein foci throughout nuclei (**fig. S5, C** to **E** and **Table S1**). We measured the nearest-neighbor distances between pairs of different Xist interactors that were either associated with Xist foci or found in the remainder or the nucleus, which includes the active X-chromosome (Xa) (**Fig. 3B**). All investigated pairs of SPEN, CELF1, PCGF5 and CIZ1 foci were, on average, within ~150-200 nm of each other when associated with Xist foci but separated by >350 nm in the nucleus both pre- and post-XCI (**Fig. 3C**). Therefore, upon distributing across the pre-Xi, foci comprising

two Xist molecules immediately recruit arrays of XCI-effector proteins and bring them closer to each other than elsewhere in the nucleus. Hence, large multi-protein assemblies, that are not typically found outside the Xi, form around Xist foci. We refer to these Xist-nucleated proteinaceous nanostructures as Xist-associated supra-molecular complexes (Xist-SMCs).

To explore whether protein integration into Xist-SMCs is the main mechanism of protein recruitment in the Xi, we probed the distribution of additional XCI effectors (**fig. S6A**), including the PRC1 component RYBP (*76*); EZH2 (*37, 39, 41, 77*) belonging to the Polycomb repressive complex PRC2; hnRNP-K which binds the B/C-repeats of Xist to recruit PCGF5 (*64*) (reviewed in (*59*)); PTBP1 and MATR3, which function with CELF1 in the maintenance of silencing and Xist localization (*74*). Nearest-neighbor measurements showed a diffraction-limited particle for each of these proteins in a near 1:1 ratio within 200 nm from the center of Xist foci (**Fig. 3D** and **fig. S6B**). These results corroborate the *de novo* assembly of a multi-protein cloud around Xist foci at the onset of XCI and identify the macromolecular crowding of many direct and indirect Xist interactors in Xist-SMCs, likely involving 100s to 1000s of protein molecules. Notably, our results revealed little differences between Xist-SMCs in the pre-Xi and Xi. Thus, SMCs around Xist foci contain all interrogated proteins before gene silencing completes.

We next investigated whether the pre-Xi to Xi transition is associated with changes in the concentration of proteins within Xist-SMCs compared to the nucleus (**Fig. 3E** and

**fig. S6C, Tables S2** and **S3**). Integrated density and volume particle measurements showed that the levels of CIZ1, CELF1, PCGF5, EZH2 and RYBP are significantly higher in Xist-SMCs than in nuclear foci. For CIZ1, CELF1, PCGF5 levels are stable from D2 to D4, whereas EZH2 and RYBP levels in Xist-SMCs change along with nuclear changes. Conversely, the concentrations of MATR3, PTBP1 and hnRNP-K are similar in Xist-SMCs and nuclear assemblies throughout differentiation, suggesting that they can fulfill their function in XCI at baseline concentration. Overall, these results show that the intense protein accumulations in Xist-SMCs are relatively stable throughout progression of XCI initiation. However, we found that SPEN levels within Xist-SMCs increase from the pre-Xi to the Xi, reaching higher levels than in nuclear assemblies only in the Xi. Thus, the completion of gene silencing is linked to the presence of more SPEN molecules in Xist-SMCs.

We next explored the mechanism that leads to the increased concentration of SPEN in Xist-SMCs with the pre- to post-XCI transition. SPEN contains intrinsically disordered regions (IDRs), which often mediate weak, multivalent interactions (*2, 44, 78-80*). We thus investigated whether the IDRs are required for the progressive accumulation of SPEN. We stably expressed SPEN-Halo with an IDR deletion (ΔIDR SPEN) in $Xist^{MS2\text{-}GFP}$ cells and confirmed that it does not interfere with gene silencing (**Fig. 3F** and **fig. S7, A** and **B**). We found that ΔIDR SPEN and wild type (WT) SPEN accumulated similarly in SMCs on the pre-Xi but, unlike WT SPEN, ΔIDR SPEN levels did not increase on SMCs in the Xi (**Fig. 3, F** and **G**, **fig. S7C)**. These results show that the time-dependent increase

of SPEN in Xist-SMCs is driven by IDRs and suggest that increased protein aggregation is associated with the completion of gene silencing on the Xi.

Our analyses show that Xist nucleates SMCs, leading to macromolecular crowding of proteins at topologically confined locations. We next explored the kinetic behavior of protein components of Xist-SMCs. We reasoned that the accumulation of proteins in Xist-SMCs should reveal both long-lived binding events, allowing for a topological retention in the SMC, as well as rapidly exchanging constituents, facilitating their access and deposition across the X chromosome. In such a model, transient Xist-SMCs structures would allow SPEN to regulate genes across the entire X chromosome.

To investigate kinetic behavior of protein exchange in the Xi, we stably expressed Halo or mCherry SPEN, PCGF5, CIZ1, CELF1 or PTBP1 fusions, and performed Fluorescence Recovery After Photobleaching (FRAP) over the Xist $^{MS2\text{-}GFP}$ territory or other nuclear regions of the same size. We also examined Xist dynamics for comparison. We observed a slow exchange of photobleached Xist $^{MS2\text{-}GFP}$ (**fig. S8A**), comparable to previous findings of ectopically expressed Xist (*81*). By fitting measured FRAP curves to a kinetic model with a single-exponential, we inferred a slow dissociation rate (0.05/min) resulting in an average Xist lifetime of ~20 min, consistent with a single dominant type of high-affinity interaction between Xist and chromatin (**Fig. 3H, fig. S8B** and **Text S4**). We found that CIZ1 exhibits a ~18 min recovery time in the Xi, similar to Xist and much longer than that of the other interacting proteins. The lifetime of CIZ1 in the Xi is longer than in

other nuclear regions, suggesting that Xist recruitment reinforces CIZ1 binding to chromatin (**Fig. 3I** and **fig. S8C**). The tight kinetic and spatial (**Fig. 3D** and **movie S4**) relationship between Xist and CIZ1 suggests that CIZ1 and Xist molecules form a stable core of Xist-SMCs.

Kinetic modelling of the SPEN, PCGF5, CELF1 and PTBP1 FRAP curves yielded rapid exchange rates compared to CIZ1 and Xist and two types of binding sites (**Fig. 3J** and **fig. S9**). Using two-exponential fits we inferred parameters for short-lived ($f_1$) and long-lived ($f_2$) bound fractions within and outside of the Xi. For all four proteins rapid binding occurred within seconds and the more stable interactions lasted several minutes (**Fig. 3J**). SPEN is the most dynamic protein, with kinetic rates characteristic of transcription factors (*82, 83*). It is likely that the more dynamic fraction of SPEN ($f_1$), which has similar characteristics inside and outside the Xi (~2s), represents fast-exchanging chromatin-engaging molecules that act to control gene expression. Conversely, the longer-lived fraction ($f_2$) may represent SPEN molecules associated with Xist-SMCs. In line with this hypothesis, FRAP experiments showed a >50% reduction of ΔIDR SPEN of the long-lived Xi binding fraction compared to WT SPEN and a more rapidly dissociating fast population (**fig. S10** and **Text S5**). Notably, recruitment to the Xi extends the long-lived binding rates and/or increases the fraction of long-lived binding events for the interrogated proteins indicating that the Xi forms a unique nuclear compartment where proteins exhibit distinct kinetic behaviors (**Fig. 3J** and **fig. S9B**). These data are also consistent with their increased accumulation in the Xist territory.

The kinetic assays revealed that proteins with short residence times aggregate around a slowly exchanging Xist-CIZ1 core. These findings denote that Xist-SMCs are rapidly exchanging ‘transient complexes’ that form local, high affinity concentration platforms, thereby increasing the typical residence times of proteins within the Xi. The accumulation in Xist-SMCs primes an increased amount of protein in the Xi, while the rapid binding and dissociation on Xist-SMCs allows proteins, particularly SPEN, to probe and spatially modulate targets further than the locations where two Xist molecules are confined. The presence of a large number of protein molecules in the Xi, without an interaction with Xist, may allow distinct protein species (such as SPEN or PRC1) to explore unique targets and with individual binding rates, which is reflected by the distinct residence time of each tested protein. Such a mechanism is likely thermodynamically favorable and critical for deterministic silencing.

To validate that the formation of Xist-SMCs yields an influx of proteins in the Xi, we next examined the protein population in the Xi but outside Xist-SMCs (referred to as Xi-fraction). We focused on the main two heterochromatinizing proteins SPEN and PCGF5. Using quantitative 3D-SIM, we detected protein assemblies formed in the X-territory, outside Xist-SMCs, which exhibit lower protein concentration than Xist-SMCs but a higher concentration than nuclear assemblies, as assessed by measurements of their integrated fluorescent density and volume (**fig. S11**). PCGF5 levels in the Xi fraction are significantly higher than within nuclear assemblies on both the pre-Xi and Xi, whereas

the level of SPEN in the Xi-fraction increases gradually during this transition. These data suggest that accumulation at Xist-SMCs leads to enrichment of constituent proteins across local neighborhoods in the X, through dynamic feeding from Xist-SMCs.

We next explored the relationship between gradual gene silencing and Xist-SMCs. It is well established that a subset of genes silences soon after Xist coating, whereas other genes, including *Atrx* and *Mecp2,* become silenced later (*29, 34*), yet the mechanism underlying these distinct silencing kinetics is unknown. We examined whether the formation of Xist-SMCs primes early gene silencing events. To test this idea, we investigated whether gene silencing originates at locations proximal to Xist-SMCs by measuring Xist enrichment over gene silencing half-times (*29*). Genes silencing with faster kinetics display more Xist binding on the pre-Xi compared to genes that become silenced later (**fig. S12, A** and **B**). These results suggest that rapid silencing kinetics are favorable in proximity to Xist-SMCs. Despite these differences, both fast and slow silencing genes are dependent on SPEN for gene silencing (*30, 31, 42, 61-63*) (**fig. S12C**). Therefore, we next interrogated whether the ablation of the dynamic association of SPEN with Xist-SMCs interferes with the progression of silencing.

To this end, we exploited an approach (*42*), in which the SPOC domain of SPEN, which is required for transcriptional repression, was tethered to Xist through the BglG/Bgl stem loop (SL) interactions and the endogenous copies of SPEN were depleted (*42, 84*). BglG-BglSL tethering of the SPOC domain to Xist ($Xist^{SPOC}$) is expected to prevent the

rapid exchange from SMCs, normally observed for SPEN, and immobilize it primarily at the core of SMCs, where Xist is localized (**Fig. 4A**). If SPEN dynamics control gene inactivation kinetics, then $Xist^{SPOC}$ should result in more efficient silencing of genes that are typically silenced early in XCI than later silencing genes. Indeed, we found that genes normally silenced early during XCI are silenced more effectively by $Xist^{SPOC}$ than genes that are normally silenced late during XCI (**Fig. 4B** and **fig. S12D**). This result suggests that SPEN-mediated silencing originates at regions that are proximal to Xist-SMCs and expansion of silencing to other target genes occurs progressively and requires the rapid kinetic behavior of the WT SPEN protein. We next addressed the mechanism underlying this expansion.

Gradual gene silencing occurs without major changes in Xist-SMC protein levels except for the IDR-dependent increase in SPEN. IDRs are known to adopt more protein-protein interactions upon changes in the local environment (*85-88*), such as an increase in chromatin density that promotes macromolecular crowding (*89, 90*). These observations suggest that the progression of gene silencing beyond targets in the spatial vicinity of Xist-SMCs, as well as the IDR-dependent concentration increase of SPEN, may be mechanistically linked to a conformational change in the X-territory. We therefore investigated whether higher-order chromatin changes occur during the pre- to post-XCI transition.

Measuring the volume and sphericity of the X chromosome upon X-chromosome painting, we found that the conformation of the pre-Xi is similar to that of the active X-chromosome (Xa) (**Fig. 4, C** and **D**). At D4, the Xi reached the distinct compact and spherical organization known for the silent X in somatic cells (*91*) (**Fig. 4, C** and **D**). We extended this result by assessing the conformation of seven unique genomic loci across the X through DNA FISH. We observed a moderate change in the higher-order chromosome configuration of the pre-Xi and a dramatic difference of the Xi compared to the Xa (**Fig. 4, E** to **G** and **fig. S13, A** to **C**). These changes in chromosomal structure increase the concentration of Xist-SMCs within the X space (**Fig. 4H, Table S4**). Accordingly, minimal distances of Xist foci or protein assemblies in Xist-SMCs are significantly reduced on the Xi compared to the pre-Xi (**fig. S13D**). Thus, increased chromosome compaction allows more genes to come in closer proximity to the Xist-SMC neighborhood, primes the IDR-dependent SPEN increase in the Xi, and the extension of silencing to more genes over time. However, the mechanisms controlling chromatin compaction on the Xi are not known.

PCGF5 exhibited the highest particle density in the X-territory among all interrogated proteins (**Fig. 4H** and **fig. S13D**), suggesting that PCGF5-containing Polycomb complexes are significantly more concentrated on the pre-Xi than the other proteins, in agreement with its early occupancy over the chromosome upon induction of Xist (*34, 35*). Given the importance of PRC1 in controlling chromatin compaction and long-range chromatin contacts to regulate gene expression during development (*68-71*),

we interrogated whether the extensive concentration of PRC1 on the pre-Xi is critical for inducing the transition of the pre-Xi to a compact Xi.

To test this hypothesis, we functionally perturbed PRC1 recruitment to the X-chromosome by deleting the B-repeat of Xist (ΔB-Xist) in ESCs (*63, 65-67, 92*). Chromosome territory measurements (by X-paints) showed that the volumetric occupancy of an Xi formed by ΔB-Xist is smaller than that formed by the full-length RNA (FL-Xist) and similar to the Xa (**Fig. 4I** and **fig. S14A**). Consistent with this result, 3D-SIM measurements revealed that the distribution of Xist foci was significantly impaired on a ΔB-Xist Xi compared to a FL-Xist Xi, with larger focus-to-focus distances and an expansion of the Xist-territory, as well as a lack of the characteristic DAPI density of the Xi (**Fig. 4J** and **fig. S14, B** to **D**). Moreover, the distribution of ΔB-Xist foci on the Xi is similar to that observed on the FL-Xist-coated pre-Xi (**fig. S13D**). These results uncover a role of the B-repeat, and in turn PRC1, in driving the compaction of the X and the concentration of Xist-SMCs. Consistent with this result, we found that genes that are affected in their silencing by the absence of PRC1, through deletion of the B and C-repeats of the RNA (*65*), are normally more likely to be silenced late during XCI initiation and exhibit lower Xist enrichment (**Fig. 4, K** and **L**). These results extend to the architectural protein structural-maintenance of chromosomes hinge domain containing 1 (SMCHD1) that depends on PRC1 for its recruitment to the Xi and controls the compartmentalization of the Xi (*93, 94*) (**Fig. S14, E** and **F**). We found that genes with high Xist binding, i.e. close to SMCs, and silenced by Xist$^{SPOC}$ domain are more likely to be SCMHD1-independent, whereas genes controlled by SCMHD1 tend to be inaccessible

for silencing by Xist$^{SPOC}$. Thus, SPEN-mediated silencing originates at regions that are proximal to Xist-SMCs and progressive expansion of silencing to other target genes requires the compaction of the X by PRC1 and SCMHD1. Thus, in the absence of PRC1 recruitment to the X, SPEN can still initiate silencing of genes that are exposed to higher SPEN concentrations due to proximity to ~50 Xist-SMCs, explaining the differential effect on silencing upon SPEN and PRC1 ablation. We conclude that macromolecular crowding, chromatin compaction and silencing by SPEN are interdependent mechanisms of heterochromatin formation.

Taken together, our work reveals a fundamentally new model for how Xist establishes a repressive compartment and orchestrates deterministic transcriptional silencing along the entire X-chromosome (**Fig. 4M**). Our model arises from the key observation that XCI is mediated by a limited number of locally-confined Xist clusters. Through expression, diffusion, sequestration, and degradation (see **Text S6**), Xist becomes localized and tightly bound to chromatin at 50 sites across the X-territory where it strongly interacts with CIZ1 to form the stable core of protein concentration hubs. These Xist/CIZ1 hubs induce macromolecular crowding of multiple XCI effector proteins in their vicinity, prior to chromosome-wide gene silencing and heterochromatinization of the X. In this way, Xist nucleates the formation of SMCs where silencing initiates (**Fig. 4M**, pre-Xi). The high Xist-seeded concentration of PRC1 progressively induces chromatin compaction, altering the local environment of Xist-SMCs and enhancing IDR-dependent protein-protein interactions. Through this process the concentration of SPEN across the

Xi also progressively increases, genes move closer to Xist-SMCs and silencing expands across the entire X (**Fig. 4M**, Xi).

The dense seeding of Xist-SMCs across the X-territory and the local binding and unbinding due to transient interactions enriches proteins, and not Xist-ribonucleoprotein complexes, over the X and allows them to spatially probe genomic targets. Consequently, XCI results from locally-confined Xist-mediated macromolecular crowding and supra-molecular aggregation of a large number of protein molecules. Through this process a relatively small number of confined Xist molecules can induce the robust and precise silencing of a much larger number of genes. The dramatic increase in macromolecular concentrations across the X-territory and the sharp change in protein particle density at the boundary of the Xi-territory define the membrane-free chromosome-wide condensate known as the Xi-compartment. The inhomogeneous distribution of Xist-SMCs and altered kinetics of proteins in the Xi, may suggest that Xist induces Polymer-Polymer Phase Separation (PPPS) rather than LLPS in the Xi (*95, 96*).

Finally, our model of how few Xist molecules can establish a chromosome-wide repressive compartment has implications for the regulation of gene expression by other lncRNAs. LncRNAs are typically expressed at low numbers, raising the question of how they can effectively regulate genes. The spatial organization of the X chromosome by few Xist molecules and IDR-based aggregation of protein effectors likely represents a general mechanism through which lncRNAs establish gene-regulatory nuclear compartments.

Other lncRNAs have also been found to nucleate spreading of Polycomb complexes (*20*), suggesting that Polycomb-mediated compaction is a common mechanism in the organization of an efficient repressive nuclear compartment.

## REFERENCES


1. M. Dundr, T. Misteli, Biogenesis of nuclear bodies. *Cold Spring Harb Perspect Biol* **2**, a000711 (2010).
2. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
3. S. P. Shevtsov, M. Dundr, Nucleation of nuclear bodies by RNA. *Nat Cell Biol* **13**, 167-173 (2011).
4. X. Li, X. D. Fu, Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet* **20**, 503-519 (2019).
5. S. Jain *et al.*, ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164**, 487-498 (2016).
6. T. Yamazaki *et al.*, Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol Cell* **70**, 1038-1053 e1037 (2018).
7. T. Hirose *et al.*, NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* **25**, 169-183 (2014).
8. A. Hubstenberger *et al.*, P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol Cell* **68**, 144-157 e145 (2017).
9. D. C. Tatomer *et al.*, Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis. *J Cell Biol* **213**, 557-570 (2016).
10. C. I. K. Valsecchi *et al.*, RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature*, (2020).

11. R. W. Yao, Y. Wang, L. L. Chen, Cellular functions of long noncoding RNAs. *Nat Cell Biol* **21**, 542-551 (2019).

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

13. A. Wutz, Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12**, 542-553 (2011).

14. R. Galupa, E. Heard, X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annu Rev Genet* **52**, 535-566 (2018).

15. A. V. Gendrel, E. Heard, Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu Rev Cell Dev Biol* **30**, 561-580 (2014).

16. A. Cerase, G. Pintacuda, A. Tattermusch, P. Avner, Xist localization and function: new insights from multiple levels. *Genome Biol* **16**, 166 (2015).

17. N. Brockdorff, J. S. Bowness, G. Wei, Progress toward understanding chromosome silencing by Xist RNA. *Genes & development* **34**, 733-744 (2020).

18. K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, B. Panning, Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet* **36**, 233-278 (2002).

19. C. Kanduri, Kcnq1ot1: a chromatin regulatory RNA. *Semin Cell Dev Biol* **22**, 343-350 (2011).

20. M. D. Schertzer *et al.*, lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. *Mol Cell* **75**, 523-537 e510 (2019).

21. T. Jegu, E. Aeby, J. T. Lee, The X chromosome in space. *Nat Rev Genet* **18**, 377-389 (2017).

22. N. Brockdorff *et al.*, The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-526 (1992).

23. C. J. Brown *et al.*, The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-542 (1992).

24. G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137 (1996).

25. Y. Marahrens, J. Loring, R. Jaenisch, Role of the Xist gene in X chromosome choosing. *Cell* **92**, 657-664 (1998).

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

27. L. Giorgetti *et al.*, Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575-579 (2016).

28. C. Y. Wang, T. Jegu, H. P. Chu, H. J. Oh, J. T. Lee, SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell* **174**, 406-421 e425 (2018).

29. E. S. L. Barros de Andrade *et al.*, Kinetics of Xist-induced gene silencing can be predicted from combinations of epigenetic and genomic features. *Genome Res* **29**, 1087-1099 (2019).

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

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

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

33. M. Escamilla-Del-Arenal, S. T. da Rocha, E. Heard, Evolutionary diversity and developmental regulation of X-chromosome inactivation. *Hum Genet* **130**, 307-327 (2011).

34. J. J. Zylicz *et al.*, The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* **176**, 182-197 e123 (2019).

35. M. Almeida *et al.*, PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081-1084 (2017).

36. A. V. Gendrel *et al.*, Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev Cell* **23**, 265-279 (2012).

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

38. C. Costanzi, J. R. Pehrson, Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393**, 599-601 (1998).

39. J. Silva *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).

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

41. H. Sunwoo, J. Y. Wu, J. T. Lee, 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* **112**, E4216-4225 (2015).

42. F. Dossin *et al.*, SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* **578**, 455-460 (2020).

43. D. Smeets *et al.*, 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* **7**, 8 (2014).

44. A. Cerase *et al.*, Phase separation drives X-chromosome inactivation: a hypothesis. *Nat Struct Mol Biol* **26**, 331-334 (2019).

45. E. G. Schulz *et al.*, The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* **14**, 203-216 (2014).

46. K. Hayashi, M. Saitou, Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat Protoc* **8**, 1513-1524 (2013).

47. J. Chaumeil, P. Le Baccon, A. Wutz, E. Heard, A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* **20**, 2223-2237 (2006).

48. E. Bertrand *et al.*, Localization of ASH1 mRNA particles in living yeast. *Mol Cell* **2**, 437-445 (1998).

49. B. P. Chadwick, Characterization of chromatin at structurally abnormal inactive X chromosomes reveals potential evidence of a rare hybrid active and inactive isodicentric X chromosome. *Chromosome Res* **28**, 155-169 (2020).

50. Y. Hsia *et al.*, Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* **535**, 136-139 (2016).

51. B. Wu, J. A. Chao, R. H. Singer, Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophys J* **102**, 2936-2944 (2012).

52. T. Tukiainen *et al.*, Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244-248 (2017).

53. T. Nozaki *et al.*, Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Mol Cell* **67**, 282-293 e287 (2017).

54. B. Chen *et al.*, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013).

55. V. Levi, Q. Ruan, M. Plutz, A. S. Belmont, E. Gratton, Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophys J* **89**, 4275-4285 (2005).

56. J. R. Chubb, S. Boyle, P. Perry, W. A. Bickmore, Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* **12**, 439-445 (2002).

57. T. Germier *et al.*, Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophys J* **113**, 1383-1394 (2017).

58. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

59. A. Loda, E. Heard, Xist RNA in action: Past, present, and future. *PLoS Genet* **15**, e1008333 (2019).

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

61. B. Moindrot *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).

62. A. Monfort *et al.*, Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell Rep* **12**, 554-561 (2015).

63. T. B. Nesterova *et al.*, Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat Commun* **10**, 3129 (2019).

64. G. Pintacuda *et al.*, hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol Cell* **68**, 955-969 e910 (2017).

65. A. Bousard *et al.*, The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep* **20**, e48019 (2019).

66. S. T. da Rocha *et al.*, Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell* **53**, 301-316 (2014).

67. D. Colognori, H. Sunwoo, A. J. Kriz, C. Y. Wang, J. T. Lee, Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Mol Cell* **74**, 101-117 e110 (2019).

68. D. J. Grau *et al.*, Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev* **25**, 2210-2221 (2011).

69. N. J. Francis, R. E. Kingston, C. L. Woodcock, Chromatin compaction by a polycomb group protein complex. *Science* **306**, 1574-1577 (2004).

70. S. Boyle *et al.*, A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev* **34**, 931-949 (2020).

71. R. S. Illingworth, Chromatin folding and nuclear architecture: PRC1 function in 3D. *Curr Opin Genet Dev* **55**, 82-90 (2019).

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

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

74. A. Pandya-Jones *et al.*, A protein assembly mediates Xist localization and gene silencing. *Nature*, (2020).

75. M. Yue *et al.*, Xist RNA repeat E is essential for ASH2L recruitment to the inactive X and regulates histone modifications and escape gene expression. *PLoS Genet* **13**, e1006890 (2017).

76. L. Tavares *et al.*, RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678 (2012).

77. J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).

78. T. Mittag, J. D. Forman-Kay, Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **17**, 3-14 (2007).

79. K. K. Turoverov *et al.*, Stochasticity of Biological Soft Matter: Emerging Concepts in Intrinsically Disordered Proteins and Biological Phase Separation. *Trends Biochem Sci* **44**, 716-728 (2019).

80. V. N. Uversky, The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett* **589**, 2498-2506 (2015).

81. K. Ng *et al.*, A system for imaging the regulatory noncoding Xist RNA in living mouse embryonic stem cells. *Mol Biol Cell* **22**, 2634-2645 (2011).

82. Z. Liu, R. Tjian, Visualizing transcription factor dynamics in living cells. *J Cell Biol* **217**, 1181-1191 (2018).

83. D. Mazza, A. Abernathy, N. Golob, T. Morisaki, J. G. McNally, A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res* **40**, e119 (2012).

84. J. Chen *et al.*, High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc Natl Acad Sci U S A* **106**, 13535-13540 (2009).

85. A. P. Minton, Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol* **10**, 34-39 (2000).

86. C. Tan, S. Saurabh, M. P. Bruchez, R. Schwartz, P. Leduc, Molecular crowding shapes gene expression in synthetic cellular nanosystems. *Nat Nanotechnol* **8**, 602-608 (2013).

87. M. Kato *et al.*, Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767 (2012).

88. S. K. Maji *et al.*, Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**, 328-332 (2009).

89. K. Richter, M. Nessling, P. Lichter, Experimental evidence for the influence of molecular crowding on nuclear architecture. *J Cell Sci* **120**, 1673-1680 (2007).

90. A. Bancaud *et al.*, Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J* **28**, 3785-3798 (2009).

91. K. Teller *et al.*, A top-down analysis of Xa- and Xi-territories reveals differences of higher order structure at >/= 20 Mb genomic length scales. *Nucleus* **2**, 465-477 (2011).

92. N. Brockdorff, Polycomb complexes in X chromosome inactivation. *Philos Trans R Soc Lond B Biol Sci* **372**, (2017).

93. C. Y. Wang, D. Colognori, H. Sunwoo, D. Wang, J. T. Lee, PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. *Nat Commun* **10**, 2950 (2019).

94. N. Jansz *et al.*, Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway. *Cell Rep* **25**, 1912-1923 e1919 (2018).

95. L. Frank, K. Rippe, Repetitive RNAs as Regulators of Chromatin-Associated Subcompartment Formation by Phase Separation. *J Mol Biol* **432**, 4270-4286 (2020).

96. F. Erdel *et al.*, Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation. *Mol Cell* **78**, 236-249 e237 (2020).

97. I. Jonkers *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).

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

99. B. Panning, J. Dausman, R. Jaenisch, X chromosome inactivation is mediated by Xist RNA stabilization. *Cell* **90**, 907-916 (1997).

100. Q. L. Ying, A. G. Smith, Defined conditions for neural commitment and differentiation. *Methods Enzymol* **365**, 327-341 (2003).

101. M. Cremer *et al.*, Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biol* **463**, 205-239 (2008).

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

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

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

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

106. E. H. Finn, G. Pegoraro, S. Shachar, T. Misteli, Comparative analysis of 2D and 3D distance measurements to study spatial genome organization. *Methods* **123**, 47-55 (2017).

107. J. Y. Tinevez *et al.*, TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80-90 (2017).

108. M. Dundr, T. Misteli, Measuring dynamics of nuclear proteins by photobleaching. *Curr Protoc Cell Biol* **Chapter 13**, Unit 13 15 (2003).

109. S. Lee, D. Cook, M. Lawrence, plyranges: a grammar of genomic data transformation. *Genome Biol* **20**, 4 (2019).

110. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

111. S. M. Gartler, L. Goldstein, S. E. Tyler-Freer, R. S. Hansen, The timing of XIST replication: dominance of the domain. *Hum Mol Genet* **8**, 1085-1089 (1999).

112. Bernd A. Berg and Robert C. Harris, From data to probability densities without histograms. *Computer Physics Communications* **179**, 443-448, (2008).

113. Joshua C. Chang, Pak-Wing Fok, and Tom Chou, Bayesian uncertainty quantification for bond energies and mobilities using path integral analysis. *Biophysical Journal* **109**, 966-974, (2015).

114. Joshua C. Chang, Van M. Savage, and Tom Chou, A path-integral approach to bayesian inference for inverse problems using the semiclassical approximation. *Journal of Statistical Physics* **157**, 582-602, (2014).

115. Ramses van Zon and Jeremy Schofield, Constructing smooth potentials of mean force, radial distribution functions, and probability densities from sampled data. *Journal of Chemical Physics* **132**, 154110, (2010).

116. Jose Braga, James G. McNally, and Maria Carmo-Fonseca, A Reaction-Diffusion Model to Study RNA Motion by Quantitative Fluorescence Recovery after Photobleaching. *Biophysical Journal* **92**, 2694-2703, (2007).

117. James G. McNally, Quantitative frap in analysis of molecular binding dynamics in vivo. In Fluorescent Proteins, *Methods in Cell Biology* **85**, 329-351, (2008).

118. Minchul Kang, Charles A. Day, Emmanuele DiBenedetto, and Anne K. Kenworthy, A quantitative approach to analyze binding diffusion kinetics by confocal frap. *Biophysical Journal* **99**, 2737-2747, (2010).

119. C. A. Brackley and D. Marenduzzo. Bridging-induced microphase separation: photobleaching experiments, chromatin domains and the need for active reactions. *Briefings in Functional Genomics* **19**, 111-118, (2020).

120. A. J. Wollman *et al.*, Transcription factor clusters regulate genes in eukaryotic cells. *Elife* **6**, (2017).

121. K. Clauss *et al.*, DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Res* **45**, 11121-11130 (2017).

122. J. Liu, C. A. Shively, R. D. Mitra, Quantitative analysis of transcription factor binding and expression using calling cards reporter arrays. *Nucleic Acids Res* **48**, e50 (2020).

123. P. Mier *et al.*, Disentangling the complexity of low complexity proteins. *Brief Bioinform* **21**, 458-472 (2020).

124. Brian D. Hendrich, Robert M. Plenge, and Huntington F. Willard, Identification and characterization of the human XIST gene promoter: implications for models of X chromosome inactivation. *Nucleic Acids Research* **25**, 2661-2671, (1997).

125. Carolina Eliscovich, Adina R. Buxbaum, Zachary B. Katz, and Robert H. Singer. mRNA on the Move: The Road to Its Biological Destiny. *Journal of Biological Chemistry* **288**, 20361-20368, (2013).

**ACKNOWLEDGMENTS**

We thank David Baker (UW) for sharing the ct-60 gene fused to GFP, Joost Gribnau (Oncode Institute) for the MS2-targeting construct, Alexander Shishkin for the SPEN entry clone, Irina Solovei for DNA from flow sorted mouse X-chromosomes, and Yi-Yun Ho for help optimizing the chromosome barcoding experiments. We thank Amy Pandya-Jones for help with the establishment of the ΔB-Xist cell lines and Tsotne Chitiashvili for help with FISH and providing XIST-488 probes. We thank Edith Heard and Francois Dossin for sharing the gene expression profiles after Bgl-GFP-SPOC rescue in SPEN depleted cells. We thank Douglas Black and Emilie Marcus and all members of the Plath laboratory for critical reading of the manuscript and helpful discussions. We also thank the David Geffen School of Medicine (DGSOM) at UCLA, the Jules Stein Eye Institute, David Williams, the Department of Biological Chemistry for supporting the imaging approaches.

**Funding**

The imaging core was supported by the NIH (R01GM115233). Y.M. was supported by the NIH (R03HD095086), K.P. was supported by an Innovation Award from and facilities of the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA, DGSOM, and the Jonsson Comprehensive Cancer Center at UCLA, the NIH (R01GM115233, 1R01MH109166, R21HD094172), the Keck Foundation, and a Faculty Scholar grant from the Howard Hughes Medical Institute. D.M. and T.C. were supported by the NSF (DMS-1814364) and NIH (R01HL146552), and A.B. was supported by the NIH (F30HL136080) and the USC MD/PhD Program.

## Author Contributions

Y.M. and K.P. conceived the project and Y.M. performed the experiments unless stated otherwise. S.T., C.L., Y.M. and J.C.G. created engineered cell lines and J.C.G., J.W. and Y.M. performed 3D-SIM imaging, data analyses and visualization. K.P. created the ΔB-Xist cell lines. C.L. and Y.M. performed and analyzed FRAP experiments. J.C.G and Y.M. synthesized probes and performed RNA/DNA FISH experiments. T.C. developed all modeling included in this study. E.J. analyzed gene silencing kinetics and correlation of RAP-seq data to Hi-C and expression data. D.M. performed the fitting of FRAP experiments and extracted parameters, and inferred the confining potential of Xist granules, under supervision of T.C. B.M. derived the X-chromosome configurations, which was also overseen by T.C. S.T. performed RAP-seq experiments and I.D. analyzed the data. A.B. cloned the WT- and ΔIDR-SPEN constructs. J.S. wrote the code for subtracting the developmental motion in live-cell 3D-SIM experiments. Y.M., K.P. and T.C. interpreted the data and contributed towards methodology and model creation. K.P. supervised all experimental work. K.P., Y.M., T.C. and M.G. acquired funding to support the project and Y.M., K.P. and T.C. wrote the manuscript with edits from all authors.

## Competing interests

The authors declare no competing interests.

## Data and materials availability

Raw and processed RAP-seq data generated in this study is available on GEO, accession number GSEXXXXX. All data and material derived from this study are available to researchers upon request.

**Supplementary Materials**

Materials and Methods

Figures S1-S14

Tables S1-S5

Movies S1-S4

References (*1-125*)

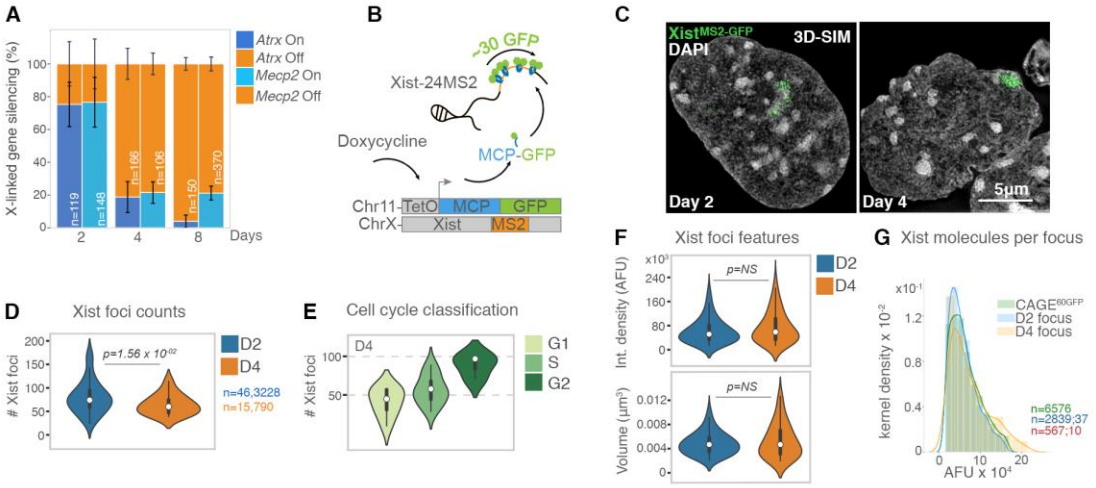



**Fig. 1 About 50 Xist foci are maintained during progression of gene silencing**

**A,** Percentage of cells, from two RNA FISH experiments, with indicated nascent transcription patterns of *Atrx* and *Mecp2* at the indicated day of differentiation. Error bars indicate standard deviation. *n* is the number of cells analyzed.

**B,** Illustration of live-cell Xist labelling strategy.

**C,** 3D-SIM projections showing $Xist^{MS2\text{-}GFP}$ signals (green) and DAPI counterstaining (grey) at the indicated differentiation day.

**D,** Violin plots of the 3D-SIM quantification of $Xist^{MS2\text{-}GFP}$ foci number at differentiation D2 and D4. Dots denote the median, thick black bars the interquartile range, thin bars show upper/lower values. *n* denotes the number of Xist foci measured (first number) and the number of cells analyzed (second number). Mann-Whitney-Wilcoxon (MWW) *p*-value is given.

**E,** As in **D** at differentiation D4 after scoring for cell cycle stages.

**F,** Violin plots of the 3D-SIM quantification of $Xist^{MS2\text{-}GFP}$ foci showing integrated density of fluorescence (AFU) and volume ($\mu m^3$) of Xist foci at D2 and D4.

**G,** Histograms depicting integrated fluorescence of $Xist^{MS2\text{-}GFP}$ foci and cages of 60 GFP molecules ($cage^{60GFP}$) (AFU) per pixel kernel density, detected by 3D-SIM.

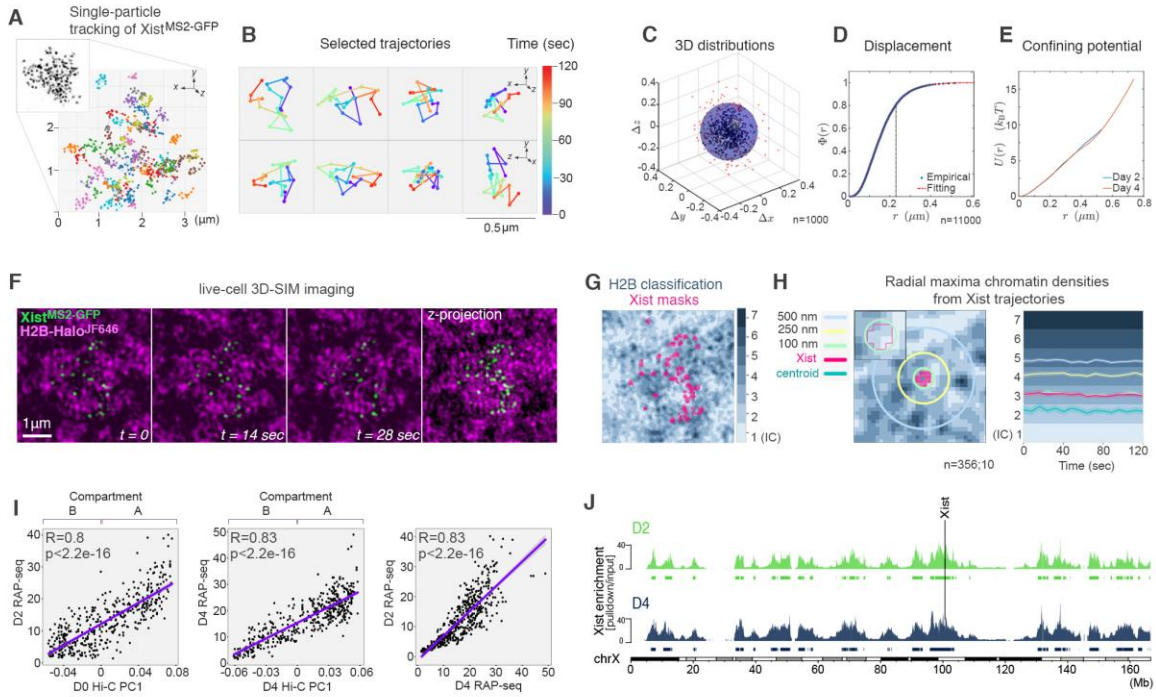



**Fig. 2 Xist foci are locally confined**

**A,** Trajectories of Xist$^{MS2\text{-}GFP}$ foci imaged for 2 minutes (5sec/ frame). Inset: Projection of one frame from live-cell 3D-SIM experiment showing an Xist$^{MS2\text{-}GFP}$ cluster at D4.

**B,** Selected trajectories from **A** showing the displacement of Xist foci over time (color-gradient) in top (xyz, top) and side (zyx, bottom) views.

**C,** Coordinates of D4 Xist foci displacements derived from ~100 trajectories, each centered about their centers of mass. *n* denotes the number of foci analyzed.

**D,** Cumulative distribution function $\Phi(r)$ of the number of displacement positions at D4 with distance from origin <*r.* The distance marked with the dashed line at *r*=0.22μm

corresponds to the radius of the shaded sphere in **C** where ~80% of all distances lie (see **Text S3**). *n* denotes the number of foci analyzed from ~800 trajectories.

**E,** Effective spherically symmetric confining potential inferred from the spatial distribution of displacement distances of Xist foci at differentiation D2 and D4. We assume an equilibrium Boltzmann distribution over an effective potential energy well that is a function of *r* (see **Text S3**).

**F,** Image sequence from *t*=0 to *t*=28sec and z-projection (from *t*=0) of $Xist^{MS2\text{-}GFP}$ (green) and $H2B\text{-}Halo^{JF646}$ (magenta), based on live-cell 3D-SIM. Bar: 1µm.

**G,** Segmentation of $H2B\text{-}Halo^{JF646}$ from live-cell 3D-SIM images into seven density classes and overlay of Xist foci masks.

**H,** Left: Schematic for the assessment of the chromatin landscape around one Xist focus. Mask (bright pink) of one Xist focus showing distances (circles) from its centroid (dark green). Inset: magnification showing the outline of the Xist mask. Right: Plot of Xist trajectories showing the average radial maxima of chromatin density reached at each timepoint, where Xist foci never surpass density class 3 and the Xist centroid remains within class 2. Light shaded areas show 95% confidence interval. *n* denotes the number of foci and cells analyzed.

**I,** Correlation of Xist enrichment at D2 and D4 of differentiation, based on RAP-seq data, to the first principal component of Hi-C data (A-compartment = positive values, B-compartment = negative values). Far right panel shows the correlation between D2 and D4 Xist RAP-seq data. Pearson correlation r-coefficients and the associated *p*-values are given.

**J,** Xist enrichment tracks along the entire X chromosome, defined based on RAP-seq data for Xist over the input, at differentiation D2 (green) and D4 (blue), with peak calls below. The Xist locus is indicated.

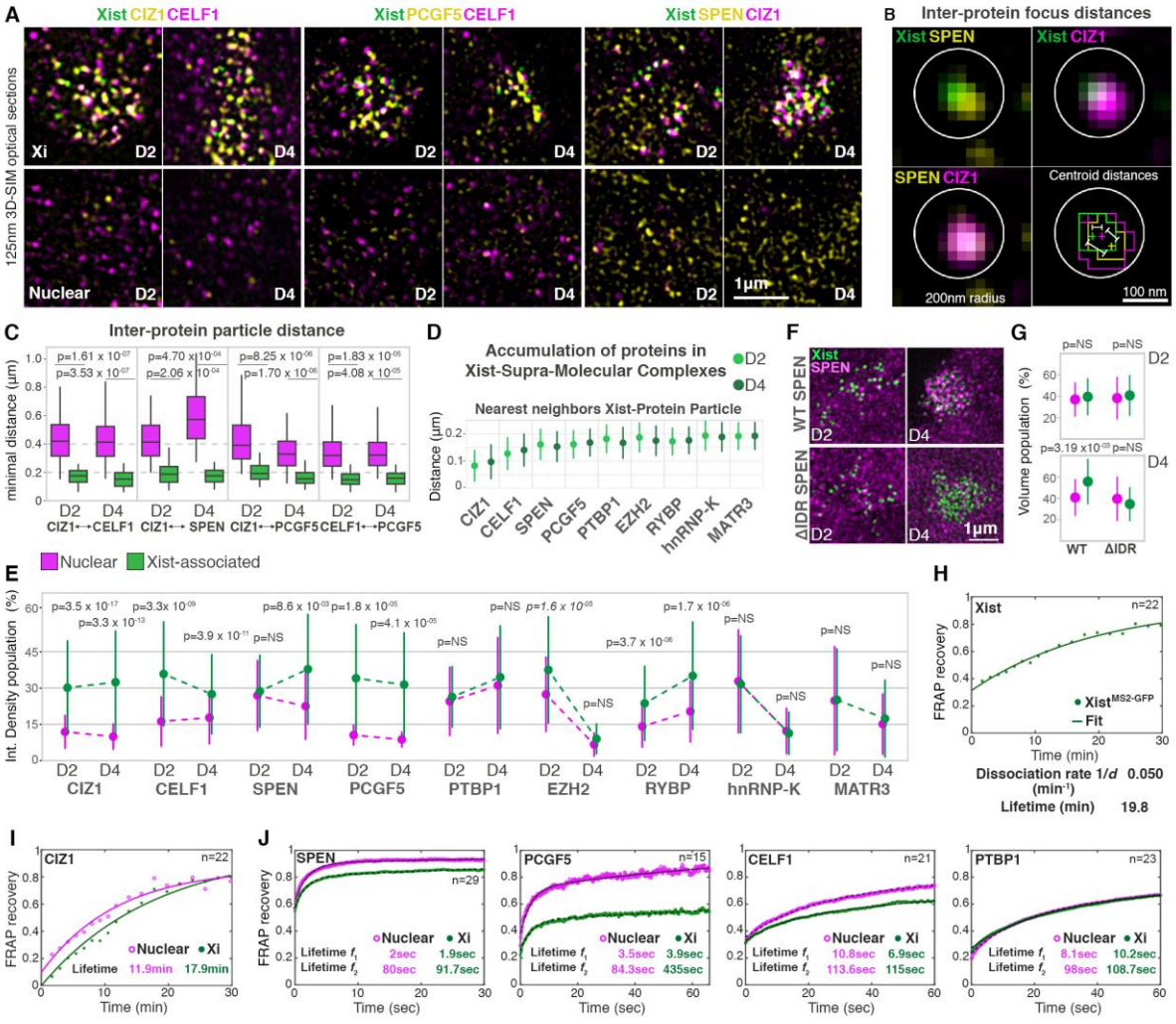



**Fig. 3 Xist seeds SMCs of interacting proteins and alters their kinetics**

**A,** 125nm 3D-SIM optical sections showing detection of indicated Halo protein-fusions labelled with JF549 (yellow) and immunodected proteins (magenta) in Xist$^{MS2\text{-}GFP}$ cells at D2 and D4 of differentiation. Chromatin counterstained with DAPI (grey). Top panels show the Xist-demarcated X-territory (Xi); bottom panels a nuclear region (Nuclear). Note the distinctive enrichment of all pairs of interactors around Xist foci.

**B,** Overview of inter-protein particle distance measurements between proteins associated with one Xist focus, based on data in **A**. Overlay: Xist (green), SPEN (yellow) and CIZ1 (magenta). Bottom right panel shows mask outlines after image segmentation, depiction of protein and Xist foci centroids (crosses) and measurement of inter-particle distances as annotated in **C** and **D** are shown. Circle denotes a 200nm radius.

**C,** Boxplots showing the distribution of nearest-neighbor distances between the indicated pairs of protein particles in nuclear and Xist-associated fractions (obtained within 250nm from Xist foci centroids) on differentiation D2 and D4. Boxes delineate the upper/lower quartiles, horizontal lines denote the median. *n* denotes the number of protein particles measured; the number of cells analyzed. Mann-Whitney-Wilcoxon (MWW) *p*-value for the comparison between the nuclear and Xist-associated fractions is given.

**D,** Point-plots showing the median distance (dot) of the nearest protein particle centroids to Xist particle centroids on D2 and D4. The bars denote the standard deviation.

**E,** Point-plots showing the integrated density of fluorescence of indicated protein particles in Xist-associated (green) and nuclear (magenta) fractions, on D2 and D4, from two experiments. Dots denote the median, bars the standard deviation. The medians at D2 and D4 are connected by dotted lines to visualize any changes. Data are normalized to the highest signal observed across the entire population. MWW *p*-values for the comparison between the nuclear and Xist-associated fractions are given.

**F,** 3D-SIM projections of the Xist-territory in wild-type (WT) SPEN-Halo $^{JF646}$ and ΔIDR-SPEN-Halo$^{JF646}$ together with Xist$^{MS2\text{-}GFP}$ on D2 and D4.

**G,** Particle volume measurements in Xist-associated (green) and nuclear (magenta) fractions, on D2 and D4, comparing ΔIDR-SPEN and WT-SPEN.

**H,** $Xist^{MS2\text{-}GFP}$ FRAP curve and fitting at D4. Dissociation rate and lifetime inferred from fitting are indicated.

**I,** FRAP recovery curve and fitting showing the nuclear and Xist-associated populations of CIZ1-mCherry at D4. Parameters from fitting with a single exponential are given.

**J,** FRAP recovery curves and fitting (dashed black lines) showing the nuclear and Xist-associated populations of SPEN-$Halo^{TMR}$, PCGF5-$Halo^{TMR}$, CELF1-mCherry and PTBP1-$Halo^{TMR}$ at differentiation D4. Lifetimes for two subpopulations ($f_1$, $f_2$) extracted from bi-exponential fitting are indicated.

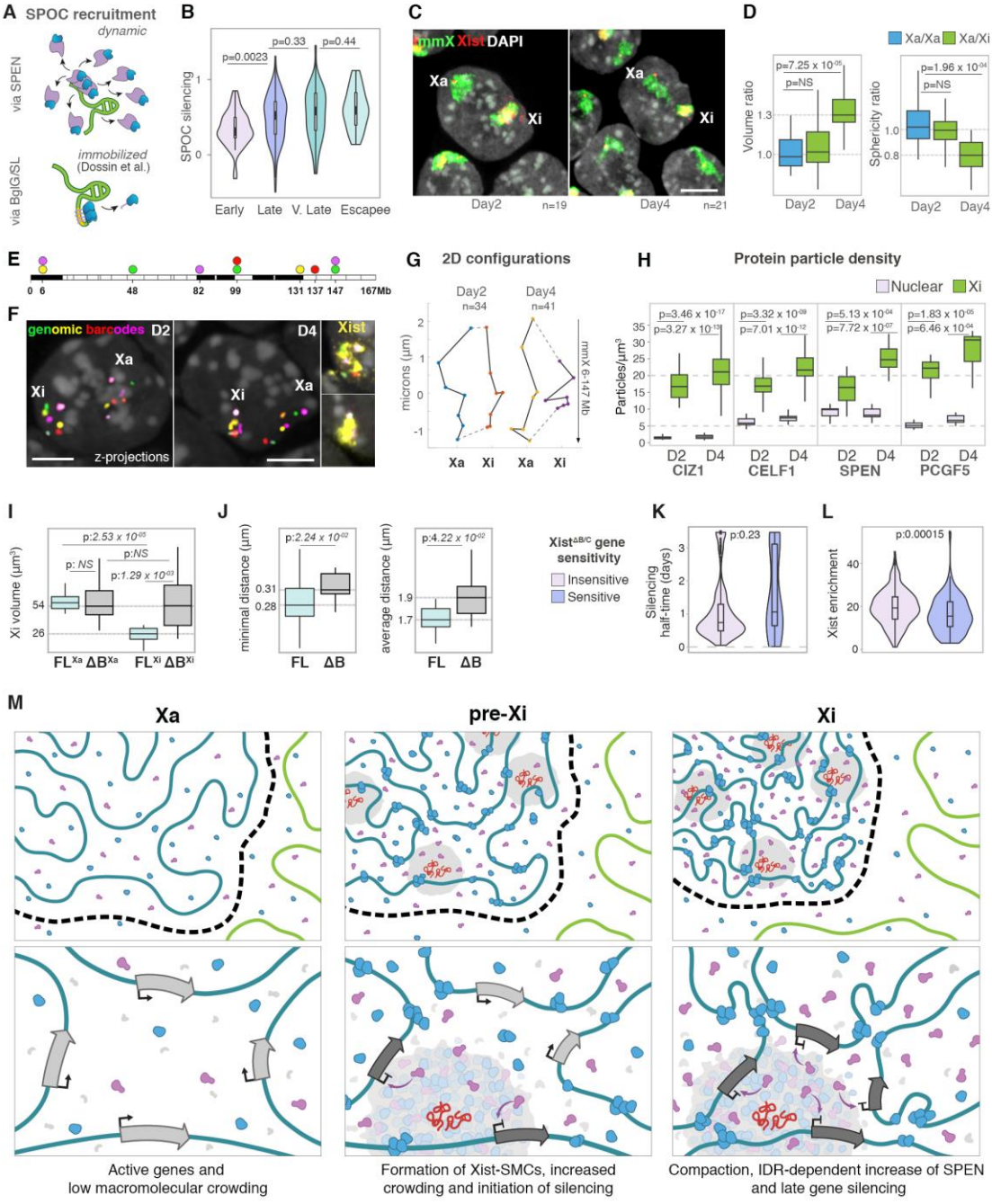



**Fig. 4 PRC1-mediated compaction is required for late gene silencing by SPEN**

**A,** Schematic showing recruitment of the SPEN silencing domain SPOC to Xist under physiological conditions (via the full length SPEN protein) or aptamer tethering (via BglG/SL, Xist$^{SPOC}$).

**B,** Violin plots showing the silencing of X-linked genes on Xist$^{SPOC}$ upon depletion of endogenous SPEN. X-linked genes are classified according to their silencing half-times during normal XCI. 0 indicates complete silencing by Xist$^{SPOC}$ and 1 complete lack of silencing. Wilcoxon *p*-values are given.

**C,** Projections of confocal stacks of D2 and D4 of differentiation after DNA-RNA FISH using X-chromosome paints (mmX, green) and Xist probes (Xist, red). DAPI counterstaining shown in grey. Bar: 5µm.

**D,** Boxplots showing the ratio of volumes (left) and sphericity (right) for Xa/Xa (blue) at day 2 and Xa/Xi (green) at D2 and D4.

**E,** Schematic of the spectral barcoding strategy applied to X-chromosome mapping through sequential DNA FISH rounds in **F**.

**F,** DNA FISH images of the spectrally barcoded genomic regions on days 2 and 4 of differentiation, as described in **E**. Overlay with Xist RNA FISH signals (far right, yellow) to score for the Xi. Bars: 5µm.

**G,** 2D configuration plots of average coordinates of genomic barcodes along the X-chromosome from **F,** extracted with 95% confidence.

**H,** Boxplots showing the distribution of the density of indicated protein particles (per µm$^3$) in the Xi and in nuclear regions on D2 and D4.

**I,** Boxplots showing the volume ($\mu m^3$) of the Xi territory after RNA-DNA FISH with X-chromosome paints and Xist probes in ESCs expressing tet-inducible full-length Xist (FL-Xist) or a deletion-mutant of the B-repeat (ΔB-Xist) after 18hrs of doxycycline induction.

**J,** Boxplots showing the minimal (left) and average (right) distance between Xist foci in FL- or ΔB-Xist expressing female cells at D4 of differentiation.

**K,** Violin plots showing the silencing half-times (in days) of genes sensitive or insensitive to Xist B/C-repeat deletion. Wilcoxon *p*-value is given.

**L,** Violin plots showing Xist RAP-seq enrichment (day 2) of B/C-repeat sensitive/insensitive genes. Wilcoxon *p*-value is given.

**M,** Model of XCI. Top panels illustrate part of the X chromosome (blue chromatin) in active (Xa), early silencing (pre-Xi) and inactive (Xi) states and surrounding autosome (green chromatin). Bottom panels show a zoom into the chromatin neighborhood where two Xist molecules (red) nucleate one Xist-SMC. Left: Xa, active genes (light grey arrows). Middle: pre-Xi, macromolecular crowding of Xist-effector proteins (grey, blue, purple proteins) around Xist 'cores' leads to the formation of SMCs and initiation of gene silencing (dark grey arrows) in their vicinity. The rapid binding kinetics of proteins creates a local concentration gradient across the entire X-territory and an early Polycomb (blue proteins) deposition across the X. Right: Xi, Polycomb-mediated compaction increases the density of Xist-SMCs, promoting IDR-dependent crowding of SPEN (purple proteins) within Xist-SMCs that leads to an increase in the occupancy of SPEN across the X-territory and completion of gene silencing.

## METHODS

Plasmid construction for engineered cell lines

Plasmids containing the 24xMS2 repeats (#31865) and MS2-Coat-Protein-GFP (MCP-GFP) coding sequence (#27121) were obtained from Addgene. The pBglII5k plasmid was used for targeting the 24xMS2 repeats into Xist (described in (*97*)) and contains homology arms for insertion into exon 7 of Xist, downstream of the E-repeat sequence, and a floxed neomycin resistance cassette. The 24xMS2 repeats were excised from plasmid #31865 by restriction digest with BglII and BamHI and cloned into the pBglII5k plasmid by infusion cloning yielding the pBglII5k-24xMS2 plasmid (which replaces the 16xMS2 repeat array originally contained in the pBglII5k plasmid). The coding region for MCP-GFP was amplified by PCR and introduced under control of a tetracycline-inducible promoter (tetO) into the pBS31 plasmid (pgkATGfrt) (*98*) by infusion cloning yielding pBS31-MCP-GFP. A reverse tetracycline TransActivator (rtTA3) cassette containing the PGK promoter and a BGH polyA element was amplified by PCR from the MXS_PGK::rtTA3-bGHpA plasmid (#62446, Addgene) and introduced into the unique AscI site of pBS31-MCP-GFP, downstream of the tetO-MCP-GFP-polyA insert, by infusion cloning, resulting in the pBS31-MCP-GFP-rtTA3 plasmid. For deletion of the B-repeat of Xist the plasmid was constructed from PCR-amplified 5' and 3' homology regions and a loxP-flanked hygroTK cassette that replaces the B-repeat sequence (chrX: 103480156-103480430).

Genetic engineering strategy for integrating MS2 repeats into the Xist locus in ESCs

The pBglII5k-24xMS2 plasmid was electroporated into the polymorphic *Mus musculus/Mus castaneus* F1 2-1 ES cell line (*99*) after linearization with XhoI. The cell culture was exposed to neomycin selection 36 hours post-electroporation. Colonies were picked and expanded for screening by genotyping PCR and RNA FISH with Xist and MS2 probes. The loxP-flanked neomycin resistance cassette was removed from targeted clones by transient expression of the Cre-recombinase. Subsequently, a FRT-recombination site-containing landing pad was targeted into the *Col1A* locus (on chromosome 11) in F1 2-1$^{24xMS2\text{-}Xist}$ ESCs as described in (*98*). MCP-GFP-rtTA3 expression cassette was then inserted into the FRT site by electroporation of a FlpO-recombinase-encoding plasmid and the pBS31-MCP-GFP-rtTA3 plasmid. The resulting ESC line was denoted as Xist$^{MS2\text{-}GFP}$.

Genetic engineering strategy for deletion of the B-repeat of Xist

F1 2-1 ES cell line (*99*) or male ESCs expressing Xist under a TetO promoter (*40*) were electroporated with linearized plasmid harboring homology arms for targeting into the B-repeat region or XIST and replacing it with a loxP-flanked hygroTK cassette for antibiotic selection. The loxP-flanked hygroTK resistance cassette was removed from targeted clones by transient expression of the Cre-recombinase. Genotyping and confirmation of deletion of the B-repeat on the 129 allele in F1 2-1 ESCs were performed by Southern blotting (not shown).

Establishment of transgenic ESC lines

For the integration of transgenes expressing various mCherry or Halo protein fusions under the control of the endogenous *Rosa26* promoter of Xist$^{MS2\text{-}GFP}$ ESCs, we employed a parent plasmid harboring homology arms for targeting into the *Rosa26* locus and a loxP-flanked puromycin cassette for antibiotic selection (R26 plasmid). A splice-acceptor (SA) and splice-donor (SD) coding sequence synthesized by Genewiz was inserted into the R26 plasmid after MluI/MfeI restriction digest by infusion reaction. The resulting R26-SA/SD plasmid was used as the parent plasmid for insertion of all protein fusions in three-piece infusion reactions. Coding sequence for CIZ1 was amplified from a donor plasmid described in (*74*). Coding sequences for histone H2B and mCherry were amplified from a H2B-mCherry plasmid (Addgene, #20972) and the Halo cDNA was obtained from plasmid Halo-EasyFusion (Addgene, #112852). The coding sequences for the PTBP1, PCGF5, CELF1 were synthesized (Genewiz). To generate the Spen-ΔIDR-Halo plasmid the full-length Spen Entry Clone (Sp22) was modified using Polymerase Incomplete Primer Extension-based mutagenesis with primers designed to delete amino acids 639-3460. Sp22 and the Spen-ΔIDR entry clone, respectively, were inserted into the PyPP-CAG-Halo-V5-IRES-Puro destination vector using Gateway LR Recombination, generating PyPP-CAG-Halo-full-length-Spen-V5 and PyPP-CAG-Halo-Spen-ΔIDR-V5, respectively, both also containing an IRES-puromycin resistance cassette. These plasmids enable constitutive expression of Spen variants with an N-terminal Halo tag and a C-terminal V5 tag and contain a polyoma episomal origin of replication for efficient propagation in mammalian cell culture. All plasmids were verified by restriction digests and sequencing.

Targeting of ESC lines

ESC lines were grown on DR4 feeders. All targetings were performed by electroporation using the GenePulserII (Biorad). Approximately $2x10^7$ cells and 50µg of DNA were resuspended in 400µl PBS in 4mm diameter cuvettes and pulsed twice for 0.2msec at 800V. Antibiotics were added to the growth media 24-36 hours after electroporation. Puromycin was used at 1.5µg/ml, hygromycin at 130µg/ml and G418 at 400µg/ml. The culture medium containing the respective antibiotics was exchanged every 2 to 3 days. Once adequate colony growth was observed (1-2 weeks), 100-200 colonies were picked under a stereoscope, dissociated by trypsinization and seeded in 96-well plate replicates. One replicate plate was used for genomic DNA extraction and subsequent genotyping PCR. All positive clones used in this study were screened to ensure gene silencing by Xist and normal Xist distribution across the X-territory upon induction of differentiation (**fig. S1, D** and **E**). Additionally, we confirmed that the 24xMS2-repeat unit was introduced into the 129 allele (**fig. S1E**).

Creation and delivery of the cage$^{60GFP}$ expression plasmid

The gene encoding ct-60 (cage$^{60GFP}$) (*50*) was amplified by PCR. The fragment was introduced under control of the CACGS promoter into the pBS32 plasmid by infusion reaction yielding pBS32-cage$^{60GFP}$ and positive clones were confirmed by restriction digests and sequencing. The pBS32 plasmid was derived from the pBS31 plasmid upon replacement of the tetO promoter with a CAGGS promoter. To visualize both Xist$^{Ms2-GFP}$

and cage$^{60GFP}$, Xist$^{MS2\text{-}GFP}$ ESCs were differentiated into EpiLCs to induce Xist expression and doxycycline was added to induce MCP-GFP expression. Expression of the cage$^{60GFP}$ was achieved by transient transfection of the pBS32-cage$^{60GFP}$ plasmid into differentiating cells by Lipofectamine3000 (Thermo Fisher) 24 hours prior to imaging, according to the manufacturer's instructions.

Cell culture

Female mouse F1 2-1 ESCs and its engineered derivatives were grown on 0.5% gelatin-coated flasks seeded with irradiated DR4 feeders (obtained from day 14.5 embryos with appropriate animal protocols in place). Cultures were maintained in mouse ESC medium containing knockout medium DMEM (Life Technologies), 15% FBS (Omega), 2mM L-glutamine (Life Technologies), 1x NEAA (Life Technologies), 0.1mM β-Mercaptoethanol (Sigma), 1x Penicillin/Streptomycin (Life Technologies), and 1000U/mL mouse LIF (homemade) in 5% CO2, 37 ºC incubators.

For all differentiation experiments, cells were adjusted for 3 passages to feeder-free conditions in the presence of LIF and two inhibitors, CHIR99021 (3μM) and PD0325901 (0.4μM) (2i+LIF). Epiblast-like (EpiLC) differentiation was performed as described in (*46*). Briefly, cells were maintained for 3 passages in serum-free 2i+LIF N2B27 media containing 1x N2 supplement and 1x B27 supplement (Thermo Fischer), 2mM L-glutamine (Life Technologies), 1x NEAA (Life Technologies), 0.1mM β-Mercaptoethanol (Sigma), 0.5 x Penicillin/Streptomycin (Life Technologies) prior to EpiLC differentiation.

To induce differentiation, cells were dissociated and seeded at a density of $2x10^5$ cells/ml in N2B27 media containing 20 ng/ml Activin A (Peprotech) and 12 ng/ml bFGF (Peprotech).

For experiments extending beyond day 4 of differentiation, we applied a protocol previously described in (*100*). Briefly, at day 4 of differentiation, EpiLCs were dissociated with accutase (Life Technologies) and seeded on geltrex-coated coverslips at a density of $5x10^5$ cells/$cm^2$. Cells were then grown in N2B27 media supplemented with EGF and FGF (10 ng/ml each), on geltrex-coated coverslips for 4 more days (d8 of differentiation). At this developmental stage, cells have lost Tsix expression as observed in **Figure S1C**. Media was exchanged daily.

C127 cells were purchased from ATCC and human fibroblasts containing abnormal X-chromosomes (GM3827, GM00735, GM06960, GM07213) were obtained from Coriell. These cell lines were cultured in DMEM (Life Technologies), 15% FBS (Omega), 2mM L-glutamine (Life Technologies) and 1x Penicillin/Streptomycin (Life Technologies).

FISH Probe synthesis

Probes for DNA and RNA FISH experiments and X chromosome paints were labelled by home-made Nick Translation and fluorescent dUTPs as described in (*101*). DNA from flow sorted mouse X-chromosomes was a gift from Irina Solovei. To create mouse Xist

probes, we used a full-length mouse Xist cDNA plasmid (p15A-31-17.9kb Xist). Human XIST probes were created from a full-length XIST cDNA construct. For assessing X-linked gene silencing, *Atrx* probes were synthesized using BAC RP23-265D6 and *Mecp2* probes using fosmid WI-894A5. For the chromosome barcoding experiment, we used BACs RP23-53H15, RP23-83J1, RP23-451D5, RP24-81K23, RP24-374B8, RP23-401G5, RP23-104K18. To create an intronic probe against the first intron of Xist, the corresponding region was amplified from the Xist-encoding BAC RP23-223G18 and was labelled by Nick Translation. RNA or DNA FISH probes were used at a concentration of 0.1μg/cm$^2$. For multispectral chromosome barcoding experiments, individual BACs were labelled separately, pooled in a 1:1 ratio and used at a concentration of 0.2μg/cm$^2$. Nick Translation products were labelled with Atto488-dUTP, Alexa Fluor 568-dUTP, Cy3-dUTP, Cy5-dUTP, Texas Red-dUTP and chromosome paints were labelled with Atto448-dUTPs or Cy3-dUTPs. A summary of the probes used in each experiment can be found in **Table S5**. All BACs and fosmids used in this study were purchased from CHORI-BACPAC.

Halo labelling

For FRAP experiments of Halo-fused proteins, 5μM of TMR Halo ligand (Promega) was added to the culture medium for 30min following a 30min incubation in media without added ligand to wash-off unbound ligand. For fixed and live-cell 3D-SIM imaging, 1μM JF549 or JF646 Halo ligands (Promega) were introduced to the media for 15min, washed-

off twice with PBS and exchanged with fresh medium which was incubated for another 15 min. Live-cell imaging or fixation was done as described in the corresponding sections.

Immunofluorescence staining

Immunodetection was performed as described in (*102*). For combined Halo ligand and antibody detection, cells were labelled with the Halo ligands and fixed followed by immunofluorescence staining. For the 4-color 3D-SIM imaging where we detect combinations of proteins together with $Xist^{MS2\text{-}GFP}$ (**Fig. 3A** and **fig. S5A)** we used CIZ1-Halo and CELF1 antibody staining, SPEN-Halo and CIZ1 antibody staining, PCGF5-Halo and CIZ1/CELF1 antibody staining. Halo transgenes were detected with the Halo ligand JF549 and primary antibodies with secondary antibodies conjugated to AlexaFluor647. In **Figures 3, D** and **E** we used Halo transgenes for detection of CIZ1, WT-/ΔIDR-SPEN, PCGF5, and PTBP1 the Halo ligand JF549 and antibody stainings for RYBP, EZH2, hnRNP-K with secondary antibodies conjugated to CF568 dye. We compared the localization of the CIZ1-Halo fusion protein and the endogenous (antibody-stained) CIZ1 protein and show the same trend (**fig. S5B**).

RNA/DNA FISH

RNA and DNA FISH experiments were conducted as previously described (*103*). For sequential RNA and DNA FISH experiments with X chromosome paints (mmX paints) and Xist probes, RNA FISH preceded DNA FISH.

For the detection of genomic regions across the X chromosome with spectral barcoding, cells were seeded in ibidi chambers with a gridded bottom and DNA FISH was performed first. 5-color optical z-stacks of 0.35µm were acquired on a confocal Zeiss LSM880 system. Spatial coordinates of the acquired positions were recorded on the ZEN software and saved. Following samples were equilibrated with 50% formamide in 2xSSC pH 7.2 solution for 3 hours at 37$^{o}$C followed by RNA FISH with Cy3-labelled Xist probes. Specimens were returned to the microscope stage and saved spatial coordinates were revisited to acquire the Xist RNA signal and discriminate between the Xi and Xa. Z-stacks from sequential rounds were superimposed using ImageJ/Fiji and alignment of the two sequential rounds was performed with the affine transformation of the StackReg plugin based on the DAPI channel. Although hybridization of RNA usually precedes DNA FISH, we have found that Xist RNA is remarkably stable during the sequential process. Since the sequential hybridization for this experiment was only necessary for the scoring of the Xi, without the need for harsh probe strip-off steps, RNA FISH was performed last.

Cell cycle analysis of Xist foci

To discriminate between different cell cycle stages, we used a combination of EdU pulse labelling, to detect S-phase cells, and anti-histone H3-phospho-Serine10 (Active Motif, #39253), to detect G2/M phase cells, while G1 cells remained marker-free. A 10mM EdU stock solution was diluted 1:1000 in growth media and cells were pulsed for 20 minutes prior to fixation. RNA FISH with Xist RNA probes and detection of EdU by click-iT reaction

with CF dye Azide 568 (Biotium, #92082) were combined with immunodetection of phospho-histone H3 Serine 10 as described in (*103*). For the assessment of Xist foci features and number throughout the cell cycle in EpiLCs (at day 4 of differentiation), we used the $Xist^{MS2\text{-}GFP}$ cell line and detected $Xist^{MS2\text{-}GFP}$ signals after addition of doxycycline.

Super-resolution microscopy

3D-Structured Illumination Microscopy (3D-SIM) was performed on a DeltaVision OMX-SR system (Cytiva, Marlborough, MA, USA) equipped with a 60x/1.42 NA Plan Apo oil immersion objective (Olympus, Tokyo, Japan), sCMOS cameras (PCO, Kelheim, Germany) and 405, 488, 642 nm diode lasers and a 568 nm DPSS laser. Image stacks were acquired on the OMX AcquireSR software package 4.4.9934 with a z-steps of 125nm and with 15 raw images per plane (five phases, three angles). Raw data were computationally reconstructed with the soft-WoRx 7.0.0 software package (Cytiva, Marlborough, MA, USA) using a Wiener filter set at 0.001 to 0.002 (up to 0.006 for DAPI) and optical transfer functions (OTFs) measured specifically for each channel using immersion oil with different refractive indices (RIs) as described in (*102, 104*). Images from different channels were registered using alignment parameters obtained from a calibration slide of 100 nm gold grid holes and a second calibration for axial alignment using 100nm diameter Tetraspeck beads (Invitrogen) as described in (*104*).

Live-cell imaging

Wide-field and confocal scanning microscopy (for FRAP experiments) or 3D-SIM live-cell imaging (4D-SIM) were performed at 37$^{o}$C (for 3D-SIM in conjunction with an objective heater), with 5% $CO_2$, controlled humidity and 10% $O_2$, having equilibrated the system and immersion oils for at least five hours prior to acquisitions. This equilibration was particularly important for obtaining artifact-free 3D-SIM datasets and minimize stage drift. Cells were differentiated in geltrex-coated chambers fitted with a high precision glass (ibidi) with daily exchange of media. To induce MCP-GFP expression, doxycycline was added to the cells two hours prior to acquisitions at a concentration of 1µg/ml. Imaging was performed in media containing no phenol red and supplemented with ProlongLive Antifade reagent (Thermo Fisher). For live-cell 3D-SIM imaging, typically 1µm to 2µm stacks of 125nm z-sections were acquired in 1- or 2-color 3D-SIM imaging to obtain 240-500 raw images per frame in 5-8 second intervals depending on exposure times and z-depth. Photobleaching over time was corrected by using histogram matching on the BleachCorrection plugin in ImageJ/Fiji.

Quantitative 3D-SIM analyses

For image segmentation, 32-bit raw datasets were imported into ImageJ/Fiji (*102*) and converted to 16-bit tiff composite stacks. The segmentation of Xist and protein foci was performed as previously described in (*102*) using the TANGO suite (*105*). Image segmentation pipelines, adjustment of thresholds and creation of seeds were performed in high-throughput batch-processing and without manual intervention. Specifically, raw datasets without filtering or subtraction of signals were imported into the segmentation

pipeline. Resulting masks of segmented particles were inspected by overlays over the raw data to ensure that the majority of signals was contained in the area to be analyzed. Nuclear masks were created using the DAPI channel as the segmentation volume. For each channel, a duplicate was generated and filtered with a 3D Gaussian filter with standard deviation of 1 (σ=1) and a Tophat filter with a radius of two pixels in xy and a one-pixel radius in z. The filtered image was segmented using the 3D Suite's Watershed method. Seed threshold and image threshold for watershed were calculated by equations Mean+StdDev*2*seed multiplier and (Mean+StdDev*2*seed multiplier)/image multiplier (Signal-to-Noise Ratio, SNR) respectively, where seed multiplier and image multiplier were determined and inspected manually to ensure the inclusion of all the regions of interest (ROIs) and the removal of background noise. Object features and distance measurements were performed using the 3D ImageJ Suite's "Measure 3D", "Quantif 3D" and "Distance" option plugins for ImageJ/Fiji.

For the assessment of the cage$^{60GFP}$ versus Xist signals, cells expressing the cage$^{60GFP}$ plasmid were typically imaged in the same Field of View (FOV) as cells with the Xist$^{MS2\text{-}GFP}$ signal, allowing us to obtain data that could be directly compared. When cells expressed both entities, since the cages are located in the cytosol, nuclear masks from the DAPI channel were created and Xist$^{MS2\text{-}GFP}$ signals were measured inside the masked regions, whereas the signal from the GFP-expressing cages was measured outside the nuclear masks.

To extract global nuclear protein particle features (in and outside the Xi), masks of the protein signals of interest were created by filtering raw data with a 3D Gaussian blur followed by automatic thresholding to include all signals and exclude nucleoli. ROIs within a 4μm radius of Xist centroids were selected for features extraction to limit computation time to ~1 hour per nucleus. Nearest neighbor centroid distances and all distances between ROIs within each channel and across different channels were extracted using the 3D ImageJ Suite for minimal distance and average distance analysis, respectively. Distance averaging was performed in Python. Assignment of Xist-associated signals was based on a proximity threshold to Xist centroids with a radius of 250nm. Signals 500nm away from Xist centroids, resulting in a 'rim' around the Xi due to the scattering of many Xist foci throughout the Xi, were defined as the nuclear fraction. To assign the Xi territory coordinates of protein signals within a 250nm radius of a Xist foci were selected, overlapping (double-called) pixels were removed and multiplied by voxel dimensions. The comparison of protein features, such as integrated density of fluorescence and volume, was performed by measurements acquired in the same laser line (568nm) for all proteins detected either with the Halo ligand JF549 or primary and secondary antibodies conjugated to CF568 dye. For each experiment, ROIs with integrated density and volume values below the 10th percentile or above the 90th percentile of the dataset were removed as outliers.

Creation of X-chromosome masks and measurements of X-territory volume and sphericity

Confocal optical stacks were imported to ImageJ/Fiji and converted to 16-bit tiffs. Using the Yen method, an automatic threshold was set to create 3D masks for the X chromosome territories. Assignment of the Xi was based on RNA FISH signals from the Xist channel. Masks were imported into 3D Suite and the volume and sphericity measurements of the X chromosomes (Xa and Xi) were extracted. Sphericity is defined as the length of the object over its width, with a maximum value of one. For the creation of mmX masks from 3D-SIM stacks to allocate Xist foci inside and outside the X-territory, a 3D Gaussian blur with standard deviation of 5 ($\sigma=5$) was used to filter the channel with the mmX paint followed by automatic threshold by the Yen method. After creating X-chromosome masks, Xist foci inside the masked region or in the remaining nucleus were analyzed as described in the 'Quantitative 3D-SIM analyses' section.

Analysis of X chromosome conformation from genomic barcoding

Confocal optical stacks were imported into Fiji/ImageJ and smoothed with a 3D Gaussian blur with a standard deviation of 1 ($\sigma=1$) and background removal using the "Subtract Background" plugin with a rolling ball radius of 10 pixels. Xa and Xi (scored by the presence of Xist RNA) were identified and saved as separate stacks. Subsequently, each probe signal centroid was extracted using the "3D Object Counter" plugin. The 3D Object Counter generated a list of coordinates of probe signals for each channel. To assign signals to multi-spectral barcodes consisting of two labels, a nearest neighbor search between the two corresponding channels was applied based on all spatial coordinates in

each channel. Once pairs of signals were assigned to the multi-spectral barcodes the coordinates obtained in the shortest wavelength were used. In cases were two adjacent signals were detected per probe, potentially due to the presence of transcripts or DNA replication, only one of the signals was used. The coordinates of individual barcodes for the Xa and Xi at days 2 and 4 of EpiLC differentiation were reoriented in 3D space to compute spatial statistics across all cells. To obtain configurations of chromosomal backbones, for each set of probe coordinates, principal component analysis (PCA) was performed in the x and y axes. The z axis was unused as the segmentation resolution in that axis is significantly lower, contributing to large variations in the z coordinate (*106*) and confounding the reorientation method used which is highly sensitive to anisotropic error. The principle component is assumed to be the “backbone” of the chromosome: the expected orientation of a chromosome if initially stretched out along that component’s direction before entropically relaxing into an equilibrium configuration. Each set of probes are rotated in order to align its corresponding principle components with the y-axis and translated such that the probes’ centroid is aligned with the coordinate origin. Probes of the same loci were then statistically compared to locate their local spatial centroid and 95% confidence interval for Xa day 2, Xi day 2, Xa day 4, and Xi day 4 separately. Ellipsoids encompassing the 95% confidence interval were plotted around each loci centroid. In order to quantify the relative compaction between Xa and Xi from day 2 to day 4, the pairwise distances of 3D coordinates (x,y,z) between each barcode location were measured and averaged over all cells. Averages of Xa distances were subtracted from those of Xi at day 2 and the same was done for day 4 in order to measure the

absolute change between chromosomes. A heatmap of this change was plotted where large negative numbers indicate a higher compaction.

Single-Particle Tracking (SPT) and extraction of trajectories

Individual Xist particles from live-cell 3D-SIM data were extracted by using TrackMate (*107*) an ImageJ plugin. DoG Detector with a 0.2μm diameter was used to define the particles and the Simple LAP Tracker with 0.25 max linking distance, 0.3 gap-closing max distance and 2 gap-closing max frame gap were used to track the particles. Trajectories that were not possible to track for over 10 consecutive frames were not used. Over 850 trajectories from 30 cells were analyzed and approximately 50% of all Xist granules without manual intervention were possible to track per nucleus for an average of two minutes. Data extracted from the software were fed into downstream confinement analyses (see **Text S3**).

Live-cell 3D-SIM image registration: subtracting background developmental motion

Two types of motion are captured simultaneously in live-cell 3D-SIM microscopy: a) the developmental motion of the cell and the nucleus, and b) the individual motion of Xist granules within the nucleus. To specifically extract (b) from live-cell 3D-SIM images, a custom Python/Jupyter tool was created that implemented the following algorithm: 1) A set of Xist granules were tracked using TrackMate. 2) For each timestep *t*, the individual displacement vectors $\boldsymbol{x}_{i,t}$ of each Xist granule were calculated. 3) For each time-step, individual displacement vectors were averaged to obtain $\boldsymbol{X}_t$, an approximation of the

developmental motion in that time-step. 4) This developmental motion was then subtracted from each Xist granule's displacement vector to arrive at an approximation to granule *i* 's motion, $\boldsymbol{x}_{i,t} - \boldsymbol{X}_t$ .

Segmentation of H2B density classes and assignment of the maximal radial distances of chromatin density classes to Xist granules

Histone H2B-Halo$^{JF646}$ intensities were extracted on ImageJ/Fiji plugin using the "getValue" macro command, that iterates over every pixel in the image to get the intensity value of each pixel, generating a list of all the pixel intensities and their corresponding coordinates. The list of intensities was imported to Python. Then, seven intensity/density classes of equal variance were determined. The 3D Suite was used to create Xist masks, while Xist trajectories were extracted from TrackMate to obtain spatial coordinates (centroids) from each time point. The matrices were paired within the radius of one pixel and chromatin density classes were measured under the masks. Radial distances were measured at all pixels within the respective 100, 250, 500 nm radius of the Xist centroid and the maximal intensity value within that range was defined. Averaged values were then plotted in a line graph as a function of time. To extract the nearest neighbors in chromatin density maps, neighboring intensities for each H2B pixel were determined as the average intensity of all adjacent pixels and stored in an array. A strip plot was used to plot the averaged intensity values where each value was assigned to one of the 7 classes based on the class of the origin pixel.

Statistical analyses of imaging data and visualization

Data analysis and visualization were performed using Python. All violin plots, boxplots, bar plots and point-plots were generated using Seaborn and Matplotlib. NumPy and SciPy were used for mathematical computation and Pandas for data manipulation and analysis. Unless stated otherwise, all graphs show the median as the central point or the central line, and bars on point plots represent the standard deviation. Point plots of protein integrated intensity and volume in **Figures 3**, **S7** and **S11** show the percentage of the maximum absolute value in each group. Statistical differences were analyzed by the two-sided Wilcoxon's or Mann-Whitney rank-sum test.

FRAP experiments

FRAP experiments with z-sectioning for $Xist^{MS2\text{-}GFP}$ and CIZ1-mCherry were performed on an LSM880 equipped with an Airyscan on a Plan-Apochromat 63x1.4NA oil immersion objective, an image size of 67.5µm x 67.5µm with a pixel size of 0.085µm. Z-optical stacks of 0.5µm were obtained through a 15µm z-depth. Bleaching was performed in ROIs demarcating the Xist territory or corresponding nuclear (control) regions at full laser power and 4 iterations with a pixel dwell time of 4.04 µsec. The first post-bleach frame was acquired immediately after bleaching. Time series were acquired every 1.3 minutes up to 10 frames and every 2 minutes thereafter for a total of 30 minutes with an Argon ion 488nm laser or a DPSS 561nm laser set to 1% laser power.

Single-plane FRAP experiments for all other proteins were performed on the OMX-SR platform in widefield mode and an image size of 512x512 pixels with a pixel size of

0.08μm. In these experiments we employed transgenic cells lines carrying mCherry-tagged CIZ1 and CELF1 and carrying Halo-tagged SPEN, PCGF5 and PTBP1, respectively (**fig. S8** and **S9**). Images were acquired for $Xist^{MS2\text{-}GFP}$ in the 488nm channel (95MHz- 6% amplitude, 20msec) and for all mCherry- or Halo-fused-TMR proteins in the 568nm channel (272MHz, 6% amplitude, 50-100ms exposure). Bleaching in ROIs demarcating the Xist territory or a corresponding nuclear (control) regions was performed by using the 568nm laser line in the Ring-TIRF/PK photokinetics module with a bleach spot of 1μm for one iteration for 0.1 seconds.

FRAP time series from z-stacking were projected and all FRAP data were analyzed as described (*108*). In brief, ~2μm user-defined ROIs identified the bleached region and data intensities were measured through time after normalization for fluorescence decay. To correct for drift, images were registered using the Correct3DDrift plugin. For compiling figures, FRAP time series were bleach-corrected using the BleachCorrect ImageJ/Fiji plugin. FRAP curves for Xist and all proteins were fit to single or double exponential models derived from mass-action kinetics. Squared errors were minimized to obtain best-fit detachment rates, binding site densities, and freely diffusing fractions. Details are given in **Text S4**.

RNA-Antisense purification (RAP-seq)

F1 2-1 female mouse ESCs were seeded on geltrex-coated plates and differentiated for 2 or 4 days. $5x10^6$ cells were collected per conditions after dissociation by accutase and RAP-seq was performed as previously described (*40*). In brief, reads were trimmed using trim_galore (v0.4.1) with default parameters to remove the standard Illumina adaptor sequence. Bowtie2 (v2.2.9) was used to align reads to the mouse genome (mm9) with the default parameters. Reads with mapping quality less than 30 were removed using samtools (v1.2). Picard MarkDuplicates (v2.1.0) was used to remove PCR duplicates. Bedtools intersect (v2.26.0) was used to count reads in sliding windows (100Kb every 25Kb) along the X chromosome. Xist localization across the X-chromosome was defined by calculating the Xist enrichment scores (pulldown/input) in the sliding windows. Unmappable regions were masked.

Hi-C compartments

To compare Hi-C compartmentalization and Xist enrichment, Hi-C PC1 values for the inactive X were downloaded from GSE99991(*28*). PC1 values at day 0 were correlated with Xist enrichment at day 2 and PC1 values at day 4 with Xist enrichment at day 4. In addition, Xist enrichment at day 2 and day 4 were correlated. Datasets were intersected with plyranges (v1.4.4) (*109*). Pearson's correlation coefficients and p-values (t-test, r≠=/=0) were calculated in R (v3.6.0), and plots were made with ggplot2 (v3.3.2) and ggpubr (v0.4.0).

Xist-tethered SPOC silencing

To identify genes that were repressed by SPOC tethered to Xist in the absence of SPEN, we used allele-specific RNA-seq count data from (*42*). SPOC silencing values for each gene were calculated as described in (*42*). Briefly, we filtered out genes that were skewed or not silenced under control conditions. We then calculated a silencing index under normal conditions ($silencing\_index_{DOX} = 1 - (allelic\_ratio_{DOX}/allelic\_ratio_{control})$) and after degrading SPEN and expressing SPOC tethered to Xist ($silencing\_index_{SPOC} = 1 - (allelic\_ratio_{SPOC}/allelic\_ratio_{control})$). To quantify the silencing defect in cells expressing only $Xist^{SPOC}$, we calculated the SPOC_[silencing_index (SPOC_[silencing_index = 1 – (silencing_indexSPOC + aux/silencing_indexDOX). To determine whether SPOC was more effective at repressing early or late silencing genes, we downloaded the proseq-estimated silencing half-times from (*29*) and categorized genes with a half-time range of (0 - 0.4] days as early silencing genes, (0.4 – 1 days] as late silencing genes (1- 2] days as very late silencing genes, and >2 as escapee genes. Wilcoxon *p*-values were calculated in R. SPEN dependence index and Bgl-GFP [silencing] indices were calculated in the same way.

SMCHD1 sensitivity

To identify genes with a silencing defect as a result of SMCHD1 KO, we downloaded RNA-seq allelic counts from GSE99991 and classified genes as SMCHD1 sensitive or insensitive following the method in (Wang et al). Briefly, %Xi was determined by finding the percentage of reads from the Xi out of the total allele specific reads in each gene. We only included active genes with >13 allele specific reads in all samples, that are normally

silenced during X inactivation. SMCHD1 sensitive genes were defined as having %Xi in SMCHD1$^{-/-}$ 3-fold greater than %Xi in WT.

Xist BC-repeat deletion

To identify genes sensitive to Xist BC repeat, we downloaded RNA-seq count data from GSE123743(*65*). We found the log2 fold change of genes with and without dox-induced Xist expression for 2 days using DESeq2 (*110*), for both full length and ΔBC Xist. We then found the difference between the log2 fold changes (Δfc) for full length and ΔBC Xist. We considered genes to be ΔBC sensitive if Δfc was greater than 0.5.

Δfc = log2(Dox 2 days/No dox) ΔBC - log2(Dox 2 days/No dox) Full length

We compared the silencing rate of ΔBC sensitive and insensitive genes by intersecting with proseq-identified silencing half-times. We compared the Xist enrichment of ΔBC sensitive and insensitive genes by intersecting with RAP-seq day 2 Xist enrichment using plyranges (*109*). Xist was excluded from these comparisons. Wilcoxon *p*-values were calculated in R.

Antibodies and dilutions

Endogenous CELF1 was detected with monoclonal rabbit anti-CUG-BP1 antibody ab129115 (1/800; Abcam); hnRNP-K with polyclonal rabbit antibody A300-678A (1/800); SPEN (Sharp) with polyclonal rabbit antibody A301-119A (1/1000); MATR3 with polyclonal rabbit antibody IHC-00081 (1/200, all from Bethyl); RYBP (DEDAF) with polyclonal rabbit antibody AB3637 (1/1000; Sigma); Ezh2 with monoclonal rabbit antibody #5246 (1/500; Cell Signaling Technology); CIZ1 with a polyclonal rabbit antibody NB100-74624 (1/800; Novus Biologicals), and histone H3 phospho-Serine 10 with the polyclonal rabbit anti-histone H3-phospho-Serine10 #39253 (1/1000; Active Motif). Two secondary antibodies were used, including high cross-absorbed donkey anti-rabbit IgG CF568 antibody SAB4600076 (1/400; Sigma) and high cross-absorbed goat anti-rabbit IgG Alexa Fluor 647 antibody A21245 (1/400; Life Technologies).

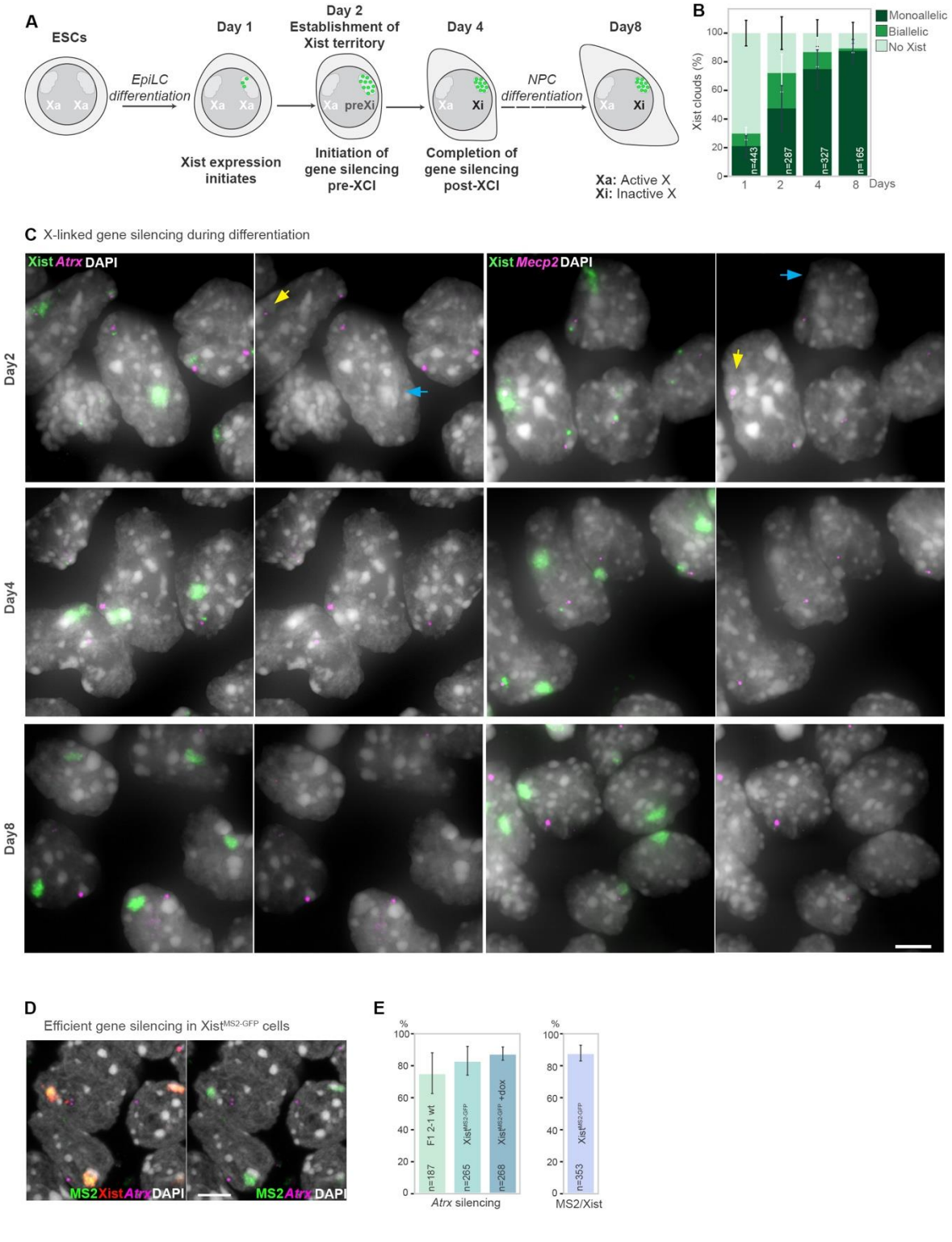

A
ESCs
Xa Xa
EpiLC differentiation
Day 1
Xa Xa
Xist expression initiates
Day 2
Establishment of Xist territory
Xa preXi
Initiation of gene silencing pre-XCI
Day 4
Xa Xi
Completion of gene silencing post-XCI
NPC differentiation
Day8
Xa Xi
Xa: Active X
Xi: Inactive X
B
Monoallelic
Biallelic
No Xist
Xist clouds (%)
100
80
60
40
20
0
n=443
n=287
n=327
n=165
1 2 4 8 Days
C X-linked gene silencing during differentiation
Xist Atrx DAPI
Xist Mecp2 DAPI
Day2
Day4
Day8
D
Efficient gene silencing in Xist MS2-GFP cells
MS2 Xist Atrx DAPI
MS2 Atrx DAPI
E
%
100
80
60
40
20
0
F1 2-1 wt
Xist MS2-GFP
Xist MS2-GFP +dox
n=187
n=265
n=268
Atrx silencing
n=353
Xist MS2-GFP
MS2/Xist

**Figure S1. Progressive gene silencing after Xist localization on the X chromosome**

**A,** Schematic of ESC differentiation protocol used throughout this study.

**B,** Quantification of the proportion of cells with Xist signals on one or both X-chromosomes at the indicated differentiation day, based on RNA FISH from two experiments. Error bars indicate standard deviation. *n* is the number of cells analyzed.

**C,** Wide-field projections of RNA FISH experiments with Xist RNA (green) and *Atrx* or *Mecp2* (magenta) probes at days 2, 4 and 8 of differentiation, with DAPI counterstaining in grey. The second column of images shows only the DAPI and X-linked transcript channels. Yellow and blue arrows indicate the presence or lack of gene expression, respectively. Bar: 5μm.

**D,** RNA FISH of $Xist^{MS2\text{-}GFP}$ cells at D4 of differentiation with Xist (red), MS2 (green) and *Atrx* (magenta) probes. Chromatin counterstaining by DAPI is shown in grey. The second image lacks the Xist FISH signal. Bar: 5μm.

**E,** Left: Demonstration of effective silencing of *Atrx* in differentiated unmodified (F1 2-1) cells and $Xist^{MS2\text{-}GFP}$ cells at D4 of differentiation, before and after induction of MCP-GFP expression with doxycycline. Right: Quantification of the proportion of $Xist^{MS2\text{-}GFP}$ cells with Xist-MS2 clouds. Error bars indicate the standard deviation.

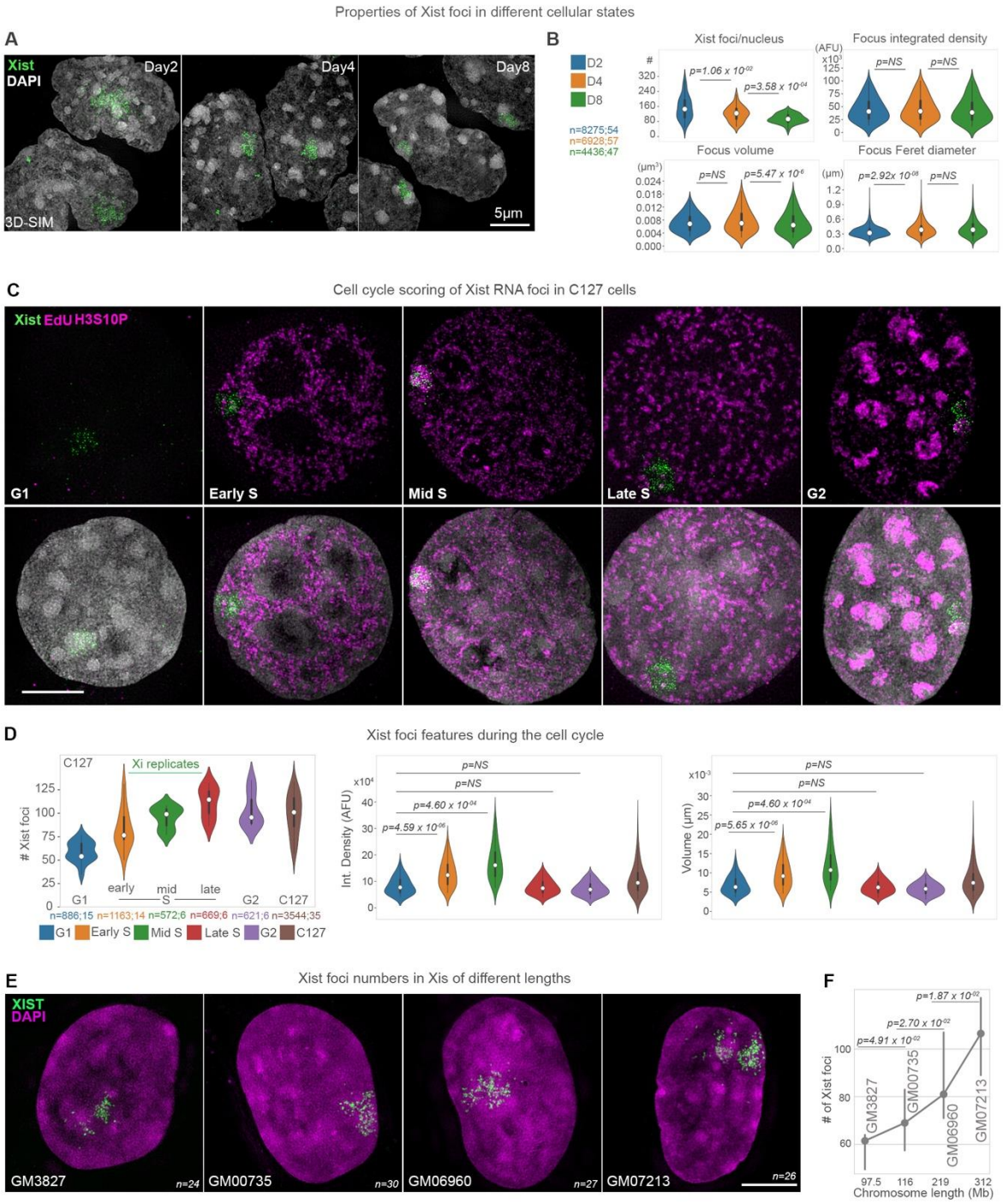

Properties of Xist foci in different cellular states
A
Xist
DAPI
Day2
Day4
Day8
3D-SIM
5µm
B
D2
D4
D8
n=8275;54
n=6928;57
n=4436;47
Xist foci/nucleus
p=1.06 x 10-02
p=3.58 x 10-04
Focus integrated density
p=NS
p=NS
Focus volume
p=NS
p=5.47 x 10-6
Focus Feret diameter
p=2.92x 10-08
p=NS
C
Cell cycle scoring of Xist RNA foci in C127 cells
Xist EdU H3S10P
G1
Early S
Mid S
Late S
G2
D
Xist foci features during the cell cycle
C127
Xi replicates
# Xist foci
G1
early
mid
late
S
G2
C127
n=886;15 n=1163;14 n=572;6 n=669;6 n=621;6 n=3544;35
G1 Early S Mid S Late S G2 C127
Int. Density (AFU)
p=NS
p=NS
p=4.60 x 10-04
p=4.59 x 10-06
Volume (µm)
p=NS
p=NS
p=4.60 x 10-04
p=5.65 x 10-06
E
Xist foci numbers in Xis of different lengths
XIST
DAPI
GM3827
n=24
GM00735
n=30
GM06960
n=27
GM07213
n=26
F
p=1.87 x 10-02
p=2.70 x 10-02
p=4.91 x 10-02
# of Xist foci
GM3827
GM00735
GM06960
GM07213
97.5
116
219
312
Chromosome length (Mb)

**Figure S2. Features of Xist foci during differentiation, across cell types and X chromosome lengths**

**A,** 3D-SIM projections showing Xist RNA FISH signals (green) and DAPI counterstaining (grey) at the indicated differentiation day. Bar: 5µm.

**B,** Violin plots of the 3D-SIM quantification of Xist foci number (#), integrated density of fluorescence (AFU), volume ($\mu m^3$), and Feret diameter (µm) at differentiation days 2, 4, and 8 of 2 replicates. White dots denote the median, thick black bars the interquartile range, thin bars show upper/lower values. *n* denotes the number of Xist foci measured; the number of cells analyzed. Mann-Whitney-Wilcoxon (MWW) *p*-values are given by comparing populations indicated by horizontal black lines. NS, Not significant.

**C,** 3D-SIM projections of immuno-RNA FISH experiments detecting Xist, EdU and H3 phospho-serine10 in C127 cells to score for the cell cycle stages G1, early S-phase (Early S), mid S-phase (Mid S), late S-phase (Late S) and G2. DAPI counterstaining in grey is additionally shown in the bottom panels. Stainings for EdU (to detect S-phase cells) and anti-H3 phospho-serine10 antibodies (to detect cells in G2) are detected in the same channel shown in magenta. Bar: 5µm.

**D,** Quantitative features of Xist RNA foci in different cell cycle stages from **C**. The C127-labelled violin plot (brown) depicts the data for the entire cell population. The *p*-value is derived from a MWW test comparing plots as indicated by black horizontal lines. NS, Not Significant.

**E,** 3D-SIM projections of RNA FISH experiments in human cell lines harboring abnormal Xi-chromosomes. XIST RNA probes are shown in green and DAPI counterstaining in magenta. Bar: 5µm. *n* denotes the number of cells used for quantification in **F**.

**F,** Graph showing the average number of XIST granules in each human cell line from **E** (y axis), ordered by the length of the abnormal Xi-chromosome in Mb (x axis). Dots denote the median, bars the 95% confidence interval. *p*-values were derived from a MWW test.

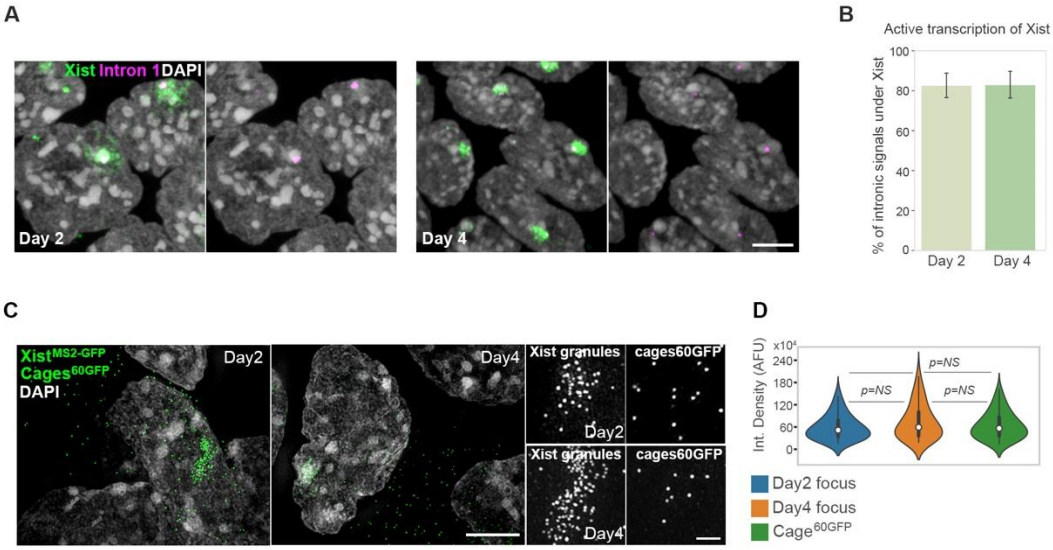



**Figure S3. Stable Xist levels during ESC differentiation**

**A,** RNA FISH with Xist (green) and intron 1 of Xist (magenta) probes to detect nascent transcripts of Xist at differentiation D2 and D4. Chromatin is counterstained with DAPI (grey). The second images show intron 1 and DAPI signals. Bar: 5μm.

**B,** Bar graphs showing the percentages of cells exhibiting a signal for the nascent transcript of Xist under Xist signal-filled nuclear regions (i.e. the pre-Xi/Xi) from experiment in **A**. Error bars denote the standard deviation.

**C,** 3D-SIM projections showing visualization of Xist$^{MS2\text{-}GFP}$ signals (green) at D2 and D4 after addition of 1μg/ml doxycycline for 2hrs to induce MCP-GFP expression. Cells were transfected with plasmids expressing cages$^{60GFP}$ (green) 24hrs prior to fixation. Chromatin is counterstained with DAPI (grey). Bar: 5μm. Bottom panels: magnifications of Xist$^{MS2\text{-}GFP}$ or cages$^{60GFP}$ signals. Bar: 1μm.

**D,** Comparison of integrated density of fluorescence of Xist granules at D2 and D4 and cage$^{60GFP}$ from **C** after image segmentation. The *p*-values from the MWW test were non-significant (NS).

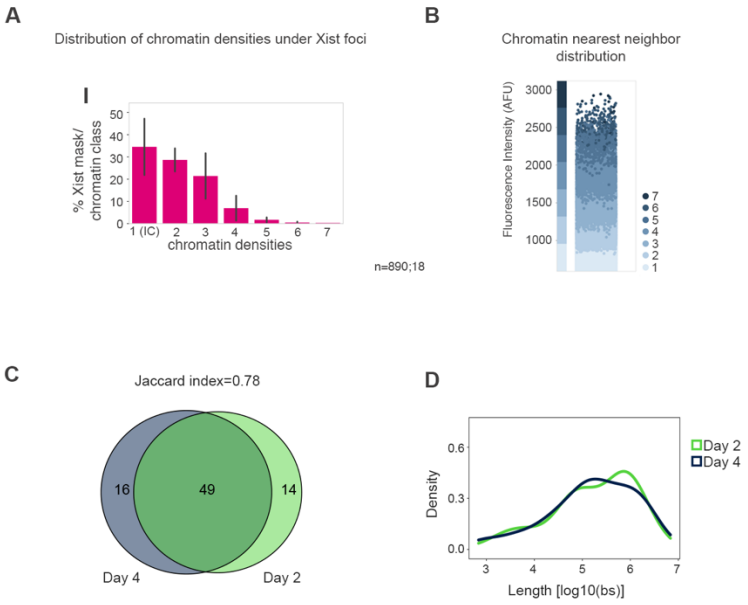



**Figure S4. Persistent localization of Xist in low chromatin density during XCI initiation**

**A,** Quantification of the percentages of chromatin density classes under Xist masks. Error bars denote the standard deviation. *n* is the number of Xist foci analyzed; number of cells.

**B,** Graph indicating the transition between chromatin densities. For each H2B-Halo$^{JF646}$ pixel, the intensity of all neighboring pixels was determined and plotted across the 7 density classes (y axis). The data show that the neighbors (placed in the bin of their associated chromatin density and color-coded by the chromatin class of the pixel of origin) will be found in the same or the directly incrementing density class to the pixel of origin.

**C,** Overlap of Xist RAP-seq peaks across the X-chromosome at D2 and D4 of differentiation.

**D,** Density plot of the length ($log_{10}$(number of bases (bs))) of Xist RAP-seq peaks on the X-chromosome at D2 and D4 of differentiation.

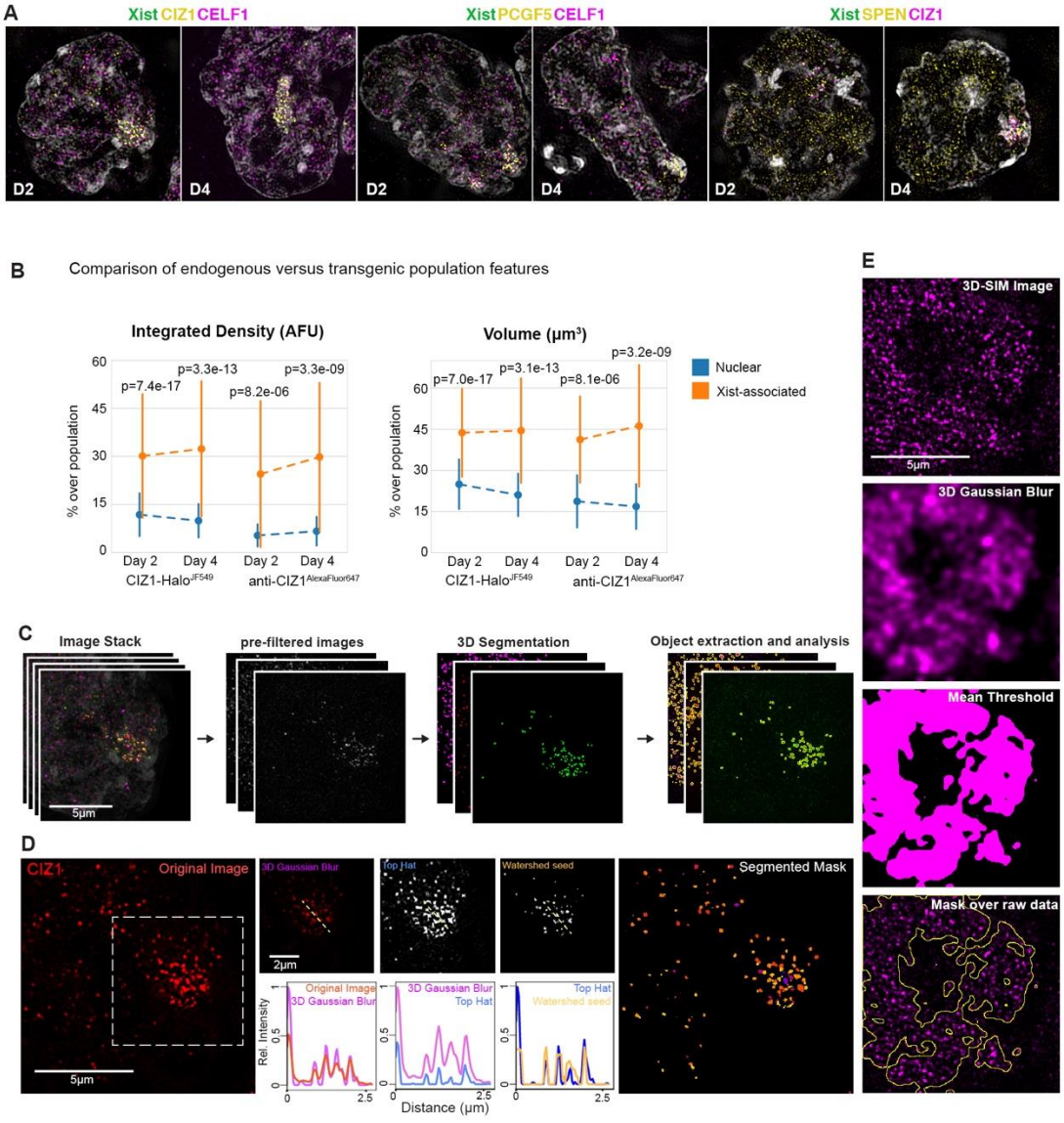



**Figure S5. Quantitative 3D-SIM analysis pipelines**

**A,** 125nm 3D-SIM optical sections showing detection of Halo protein-fusions labelled with JF549 (yellow) and immunodected proteins (magenta) in Xist$^{MS2\text{-}GFP}$ cells at day 2 (D2) and 4 (D4) of differentiation. Data are whole nucleus images of insets shown in **Figure 3A**.

**B,** Comparison of the CIZ1-Halo$^{JF549}$ pattern to the localization of the endogenous CIZ1 detected with anti-CIZ1 primary and AlexaFluor647-conjugated secondary antibodies showing the same trend in both endogenous and transgenic protein populations after 3D-SIM quantitative analyses. *p*-values are derived from a MWW test.

**C,** Overview of image segmentation pipeline of 3D-SIM data for assignment of objects (particles) for downstream measurements of features.

**D,** Outline of image segmentation pipeline with the example of CIZ1 signals, showing the output of pre- and post-filtering steps for assigning protein particles. 3D Gaussian and TopHat kernels are applied for contrast enhancement and a 3D-seeded watershed algorithm is used to obtain segmented particles. Histograms indicate the intensity profiles obtained under the dashed line during the processing steps, showing the increased resolving and separation of individual particle signals.

**E,** Pipeline for the creation of masks to include all protein signals quantified in 3D-SIM particles analyses.

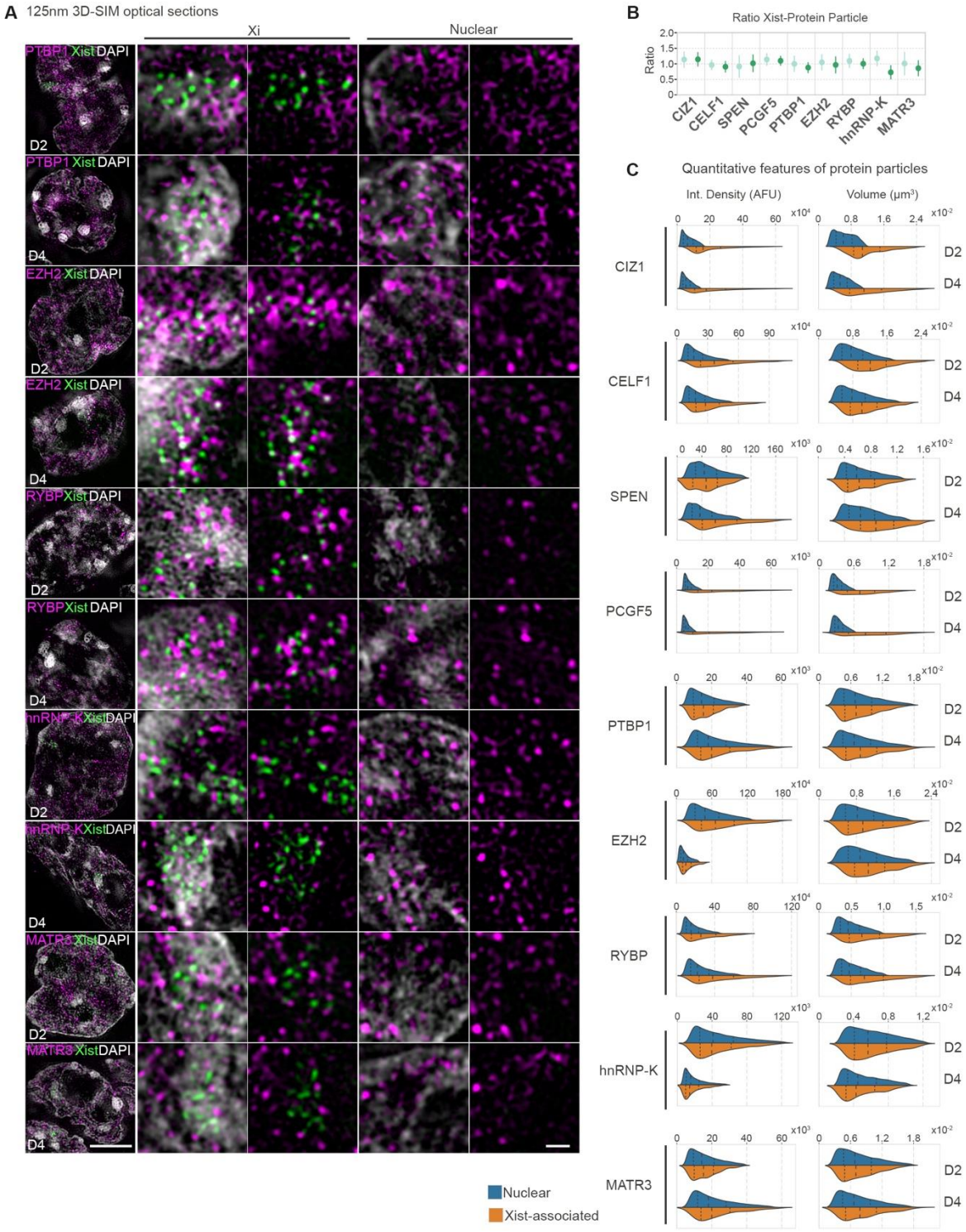

A
125nm 3D-SIM optical sections
Xi
Nuclear
PTBP1 Xist DAPI
D2
PTBP1 Xist DAPI
D4
EZH2 Xist DAPI
D2
EZH2 Xist DAPI
D4
RYBP Xist DAPI
D2
RYBP Xist DAPI
D4
hnRNP-K Xist DAPI
D2
hnRNP-K Xist DAPI
D4
MATR3 Xist DAPI
D2
MATR3 Xist DAPI
D4
B
Ratio Xist-Protein Particle
Ratio
2.0
1.5
1.0
0.5
0
CIZ1
CELF1
SPEN
PCGF5
PTBP1
EZH2
RYBP
hnRNP-K
MATR3
C
Quantitative features of protein particles
Int. Density (AFU)
Volume (µm³)
CIZ1
CELF1
SPEN
PCGF5
PTBP1
EZH2
RYBP
hnRNP-K
MATR3
D2
D4
Nuclear
Xist-associated

**Figure S6. Quantitative features of proteins in nuclear and Xist-associated fractions**

**A**, Left panels: 3D-SIM 125nm optical sections of cells at days 2 (D2) and 4 (D4) of differentiation showing $Xist^{MS2\text{-}GFP}$ signals (green) and immunodetected or Halo-fused proteins labelled with primary and secondary CF568-conjugated antibodies or JF549 Halo ligands (magenta) (see methods). DAPI counterstaining shown in grey. Right panels: magnifications of the Xi and nuclear regions with or without DAPI. Bar: 5μm; Magnifications: 1μm.

**B,** Point-plots showing the average ratio of the number of segmented protein particles to Xist foci within 250nm.

**C,** Violin plots of integrated densities and volumes of nuclear and Xist-associated protein particles from experiment shown in **Figures S6A** and **3E,** at days 2 and 4 of differentiation. Long-dashed black lines denote the median, short- dashed lines denote the upper/lower quartiles.

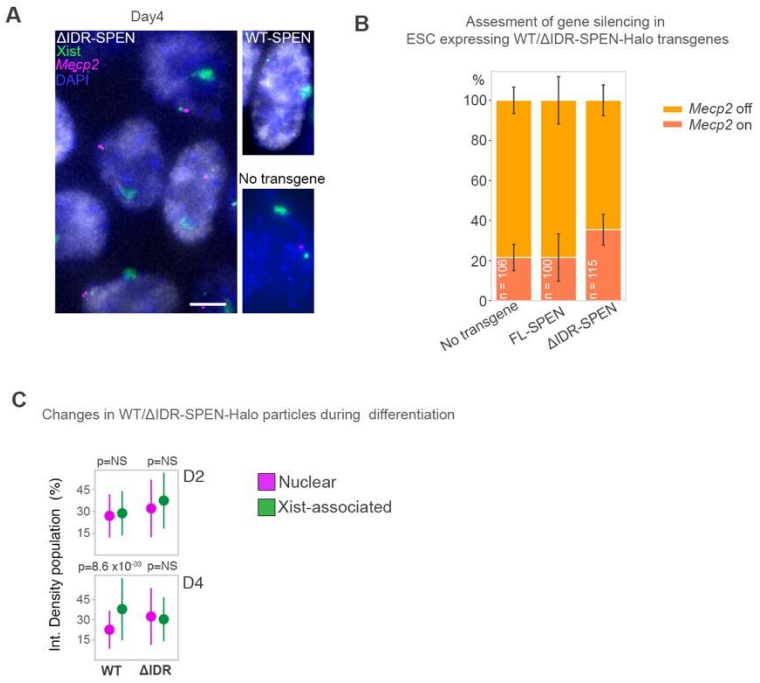



**Figure S7. Assessment of gene silencing and protein levels in differentiated cells expressing the WT or ΔIDR-SPEN transgene**

**A,** Left: RNA FISH experiment with Xist probes (green), and *Mecp2* probes (magenta) on Xist$^{MS2\text{-}GFP}$ cells expressing the ΔIDR-SPEN-Halo transgene at day 4 of differentiation. ΔIDR-SPEN-Halo signals are detected with JF549 ligand (grey) and chromatin is counterstained with DAPI (blue). Right: Same RNA FISH experiment for Xist$^{MS2\text{-}GFP}$ cells expressing the WT-SPEN-Halo transgene (top) and lacking a transgene (no transgene, bottom). Bar: 5μm.

**B,** Demonstration of effective X-linked gene silencing capacity of $Xist^{MS2\text{-}GFP}$ cells expressing ΔIDR-SPEN-Halo or WT-SPEN-Halo transgenes or cells without transgene expression.

**C,** Particle integrated density measurements, comparing data for ΔIDR-SPEN and WT-SPEN at D4 of differentiation.

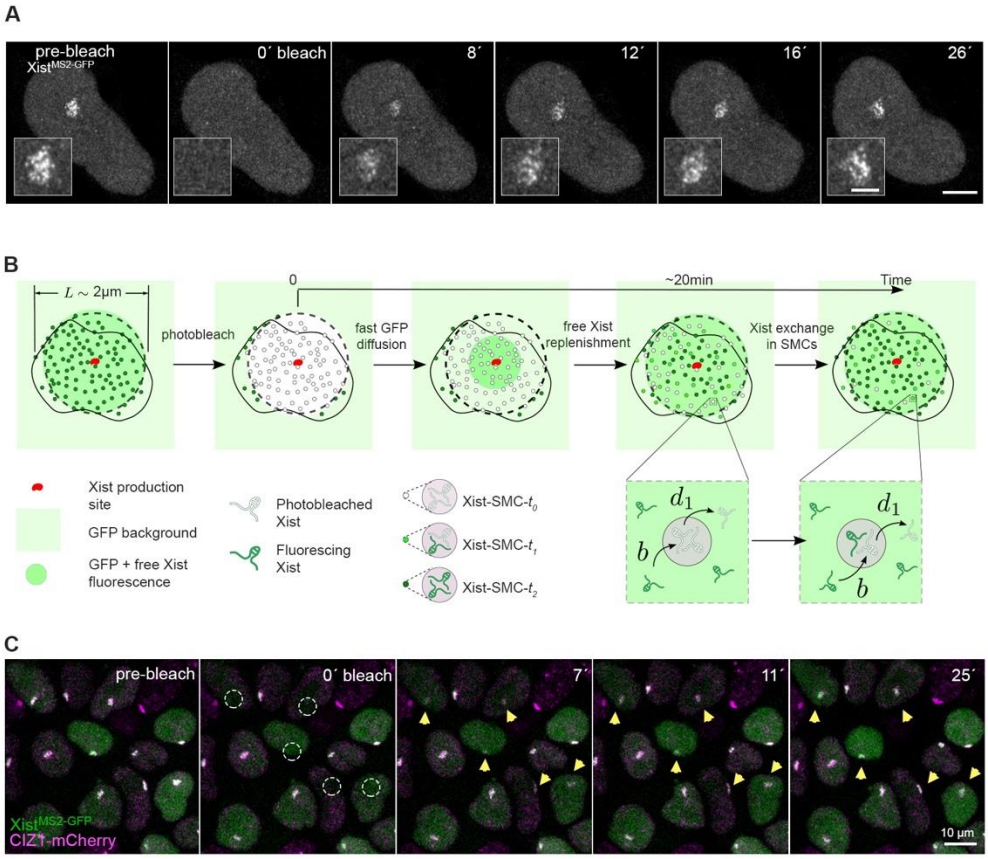


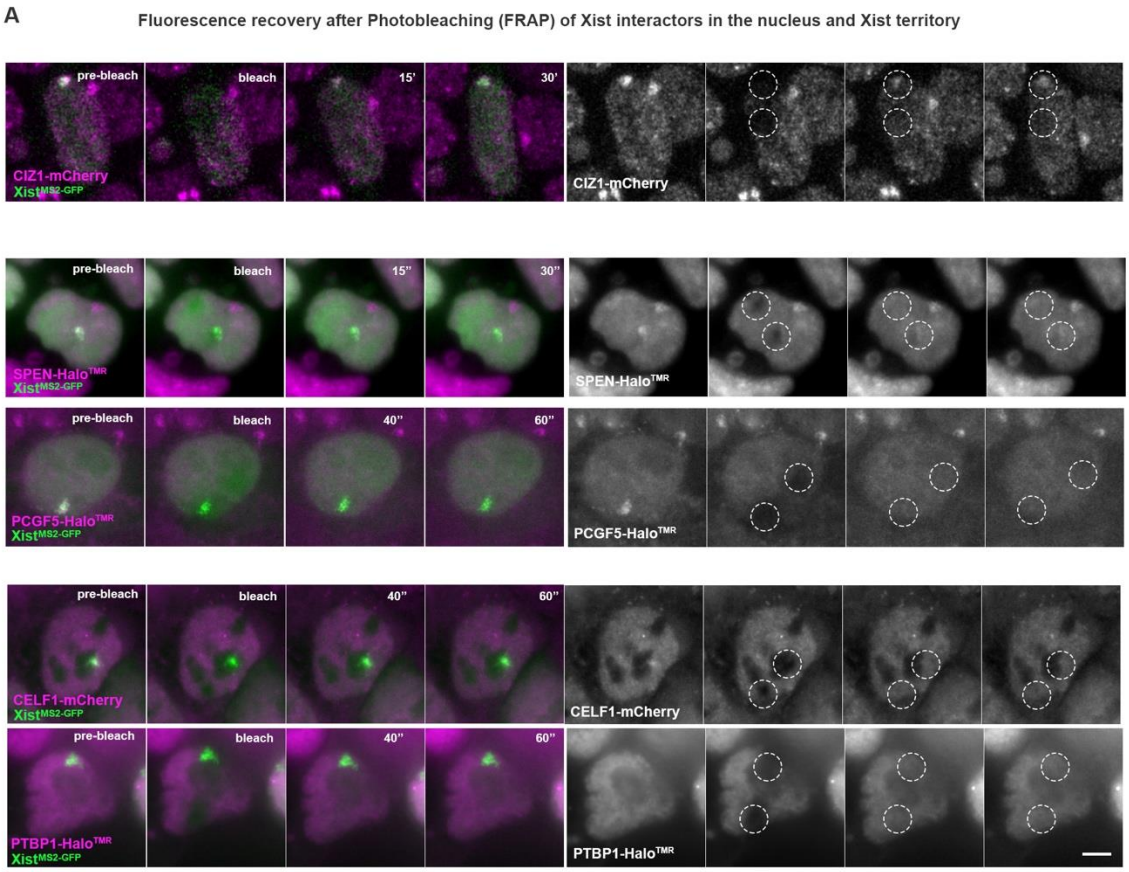

**Figure S8. Xist and CIZ1 exhibit similar kinetics**

**A,** Image sequence from FRAP experiment of Xist$^{MS2\text{-}GFP}$ at differentiation day 4, showing Airyscan optical sections. Insets show magnifications of the Xist territory. Bar: 5μm; inset: 2μm.

**B,** Schematic of the model for the Xist FRAP process. The expression and replenishment of Xist from its expression site, and of proteins from other chromosomes the nuclear fraction is assumed to be fast (free MCP-GFP replenishment is even faster; we assume is almost instantaneous just after *t*=0). The exchange of photobleached Xist with fluorescing Xist is fast in the Xi-territory outside of Xist-SMCs (free pool) and slow within

Xist-SMCs. Xist-SMCs with zero, one, and two fluorescing Xist molecules are denoted Xist-SMC-0, -1, and -2, respectively. Binding of Xist to sites in the SMCs occurs at rate *b*, while dissociation occurs at rate $d_1$, which sets the timescale for FRAP recovery. The FRAP curves for Xist were fit with a single exponential. Besides the dissociation rates *d*, the percentage of fluorescence coming from freely diffusing and SMC-associated compartments, 1-$f_1$ and $f_1$, were also inferred. See **Text S4** for modeling details.

**C,** Image sequence showing projections from FRAP experiment of $Xist^{MS2\text{-}GFP}$ cells expressing CIZ1-mCherry at D4. $Xist^{MS2\text{-}GFP}$ is shown in green and CIZ1-mCherry in magenta. Dashed circles indicate bleached Xist-territories and yellow arrows monitor recoveries of bleached regions. Bar: 10μm.



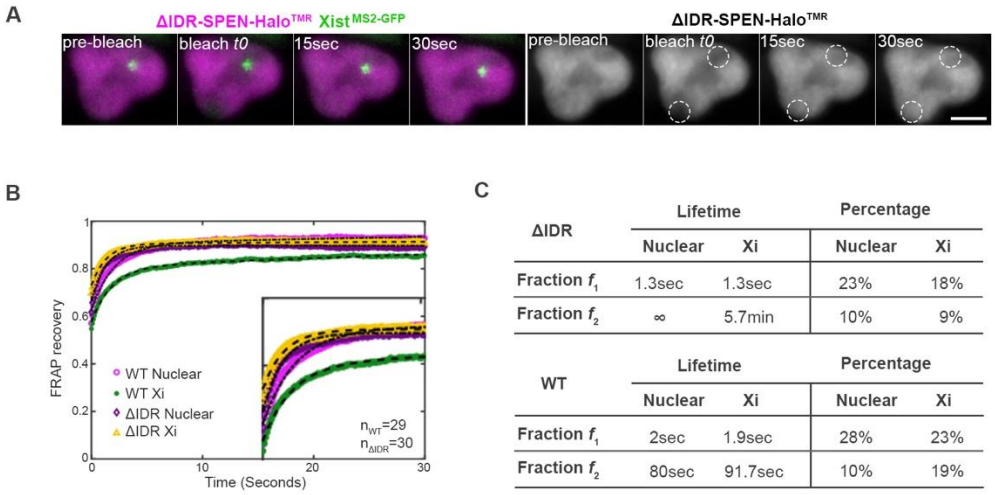

B

| | SPEN | | PCGF5 | | CELF1 | | PTBP1 | |
|---|---|---|---|---|---|---|---|---|
| | Nuclear | Xi | Nuclear | Xi | Nuclear | Xi | Nuclear | Xi |
| Fraction 1 ($f_1$) | 28 % | 23 % | 33 % | 22 % | 33 % | 18 % | 32 % | 30 % |
| Fraction 2 ($f_2$) | 10 % | 19 % | 29 % | 52 % | 31 % | 49 % | 44 % | 42 % |

**Figure S9. FRAP experiments of Xist interactors in nuclear and Xi-regions**

**A,** Time series from FRAP experiments showing bleaching (dashed circles) of nuclear or the Xi regions demarcated by Xist$^{MS2\text{-}GFP}$ signals, for CIZ1-mCherry, CELF1-mCherry, SPEN-Halo, PCGF5-Halo and PTBP1-Halo transgenes at day 4 of differentiation. Bar: 5µm.

**B,** FRAP parameters extracted from fitting for indicated proteins in the nucleus and the Xi, respectively, showing the percentage of slow ($f_1$) and fast ($f_2$) detaching fractions inferred for each protein.




| ΔIDR | Lifetime | | Percentage | |
|---|---|---|---|---|
| | Nuclear | Xi | Nuclear | Xi |
| Fraction $f_1$ | 1.3sec | 1.3sec | 23% | 18% |
| Fraction $f_2$ | ∞ | 5.7min | 10% | 9% |

| WT | Lifetime | | Percentage | |
|---|---|---|---|---|
| | Nuclear | Xi | Nuclear | Xi |
| Fraction $f_1$ | 2sec | 1.9sec | 28% | 23% |
| Fraction $f_2$ | 80sec | 91.7sec | 10% | 19% |

**Figure S10. Abolishment of Xi-immobility and increased detachment rates upon deletion of the IDR domain of SPEN**

**A,** Time-sequence from FRAP experiment of Xist$^{MS2\text{-}GFP}$ cells expressing the ΔIDR-SPEN-Halo transgene at D4 of differentiation. Bar: 5μm.

**B,** FRAP recovery curves comparing WT and ΔIDR-SPEN-Halo recoveries in nuclear and Xist territories. The first 12 seconds of this graph are shown as a magnified inset.

**C,** FRAP parameters extracted from fitting for indicated proteins in the nucleus and the Xi, respectively, showing the lifetime and percentage of slow ($f_1$) and fast ($f_2$) detaching fractions inferred for each protein. Top panel are parameters inferred for ΔIDR and bottom panel for WT SPEN for comparison. Note, that the data for WT SPEN are the same shown in **Figures 3** and **S9**, for comparison purposes.

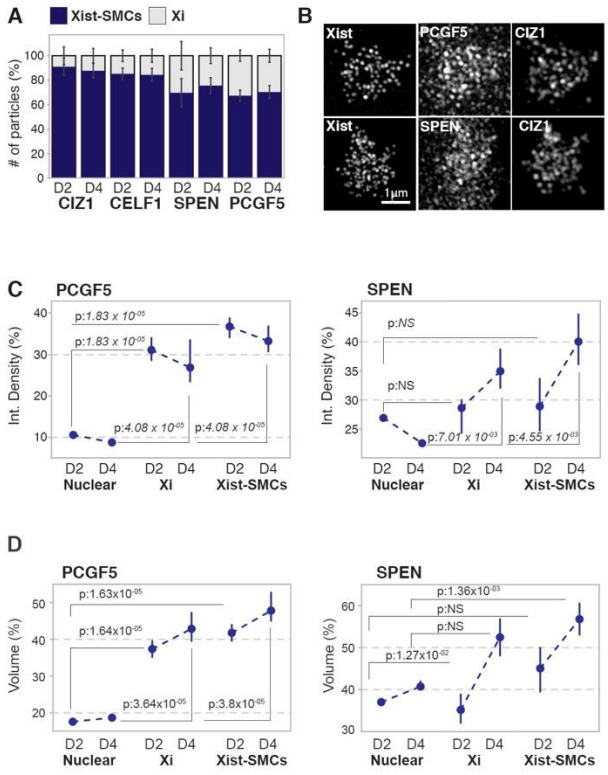



**Figure S11. Influx of proteins from Xist-SMCs regulates protein levels in the X chromosome**

**A,** Bar graphs showing the percentage of protein particles detected in the pre-Xi/Xi outside Xist-SMCs (labelled as Xi) or in Xist-SMCs, at D2 and D4. Error bars indicate standard deviation.

**B,** 3D-SIM projections showing magnifications of the Xi region for Xist, PCGF5 and CIZ1 (top) and Xist, SPEN and CIZ1 (bottom). Note the extended population of PCGF5 and SPEN in the Xi compared to Xist and CIZ1.

**C,** Point-plots showing the integrated densities of protein particles for PCGF5 (left) and SPEN (right) within the nuclear fraction, the Xi (outside Xist-SMCs) or in Xist-SMCs, at D2 and D4. Dots denote the median, bars the 95% confidence interval. The medians at days 2 and 4 are connected by dotted lines to visualize any changes. Data are normalized to the highest signal observed across the entire population. MWW *p*-values for the comparison between the nucleus and the Xi or Xist-SMCs at day 2 or 4 are given.

**D,** Same as **C** expect showing volume measurements.

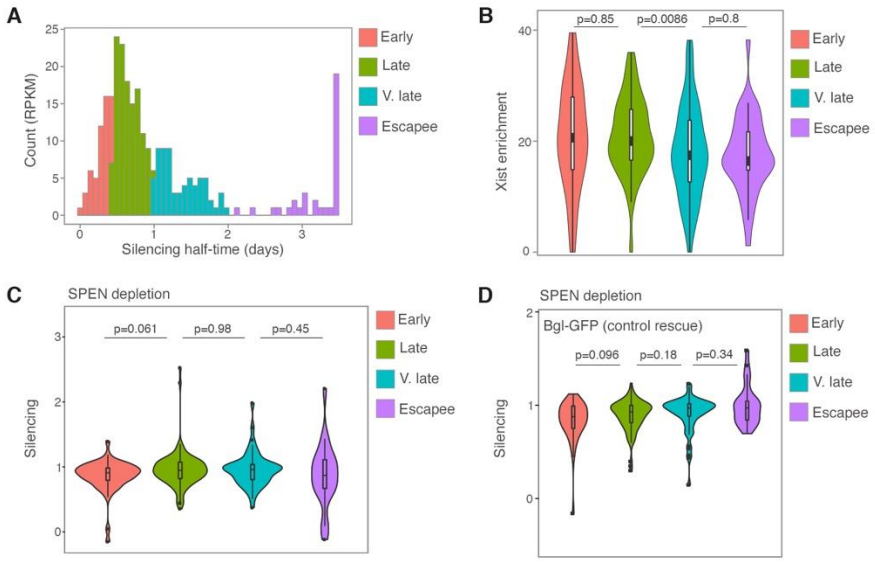



**Figure S12. SPEN-dependence of X-linked gene silencing and Xist enrichment for genes with differing silencing kinetics**

**A,** Distribution of silencing half-times of X-linked genes as determined in (*29*) and classification in early, late, and very late (v. late) silenced genes as well as XCI escapees.

**B,** Violin plots showing Xist enrichment (D2) for gene groups described in **A.** Wilcoxon *p*-value is given.

**C,** Violin plots showing silencing profiles of the gene groups defined in **A** in SPEN depleted cells. 0 indicates complete silencing and 1 complete lack of silencing. Wilcoxon *p*-value is given.

**D,** Violin plots showing the effect on gene silencing upon SPEN depletion and GFP is recruited to Xist via the BglG/SL tethering. X-linked genes are classified according to their silencing half-times during normal XCI as given in **A**. 0 indicates complete silencing and

1 complete lack of silencing by Xist$^{GFP}$ (Bgl-GFP). Wilcoxon *p*-value is given. This figure serves as control for the Xist$^{SPOC}$ (Bgl-GFP-SPOC) experiment in **Figure 4B**.

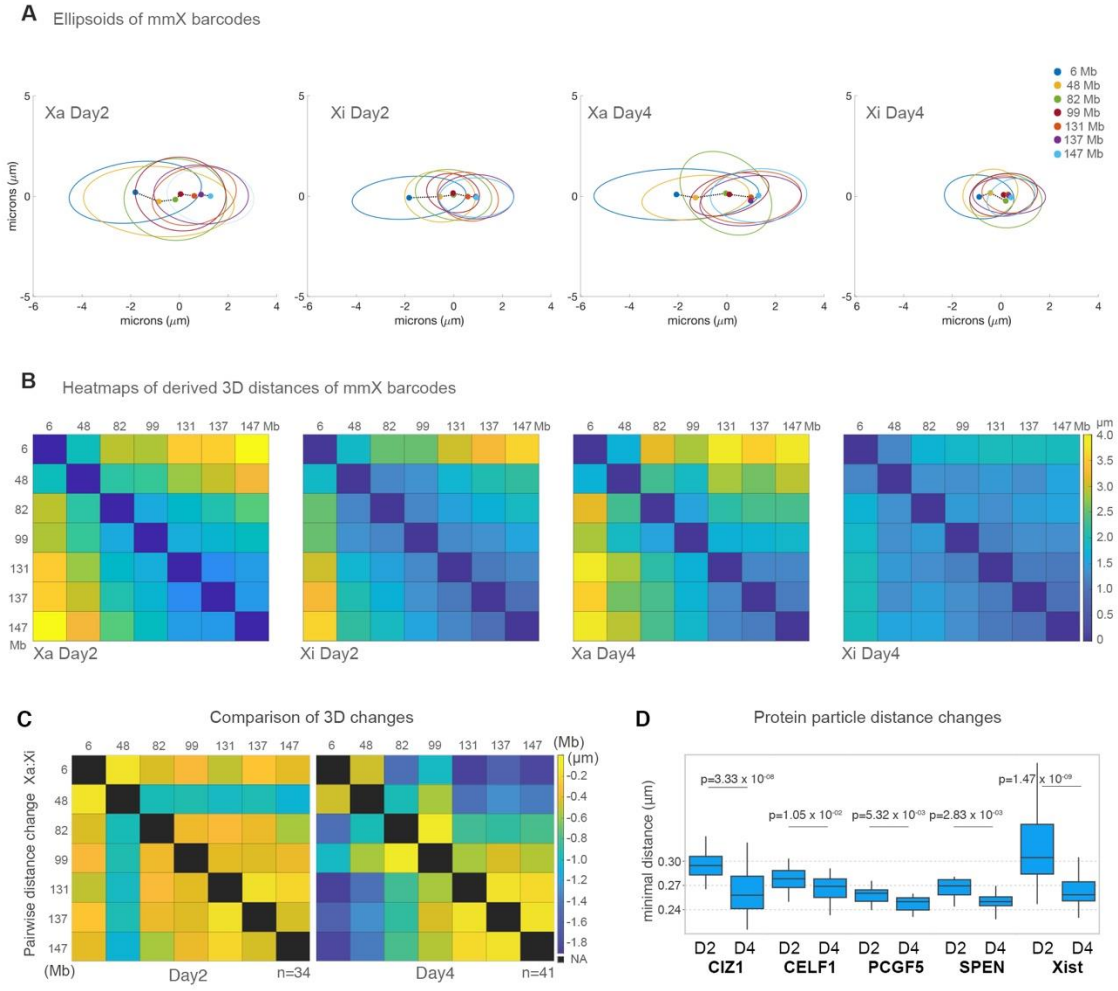



**Figure S13. Progressive compaction of the Xi and concentration of Xist-SMCs during differentiation**

**A,** Ellipsoids of 2D coordinates of X-chromosome (mmX) barcodes on the Xa and Xi at D2 and D4 of differentiation, showing a radius of 95% confidence in the allocated positions. The position along the X-chromosome for each probed location is given in megabases (Mb).

**B,** Heatmaps showing the average 3D spatial distances of genomic barcodes across the X chromosome on the Xa and Xi at D2 and D4 of differentiation.

**C,** Heatmaps as in **B,** showing changes in distances between probes on the Xa and Xi on D2 and D4.

**D,** Nearest neighbor Xist foci and protein particle distances in the Xi at D2 and D4.

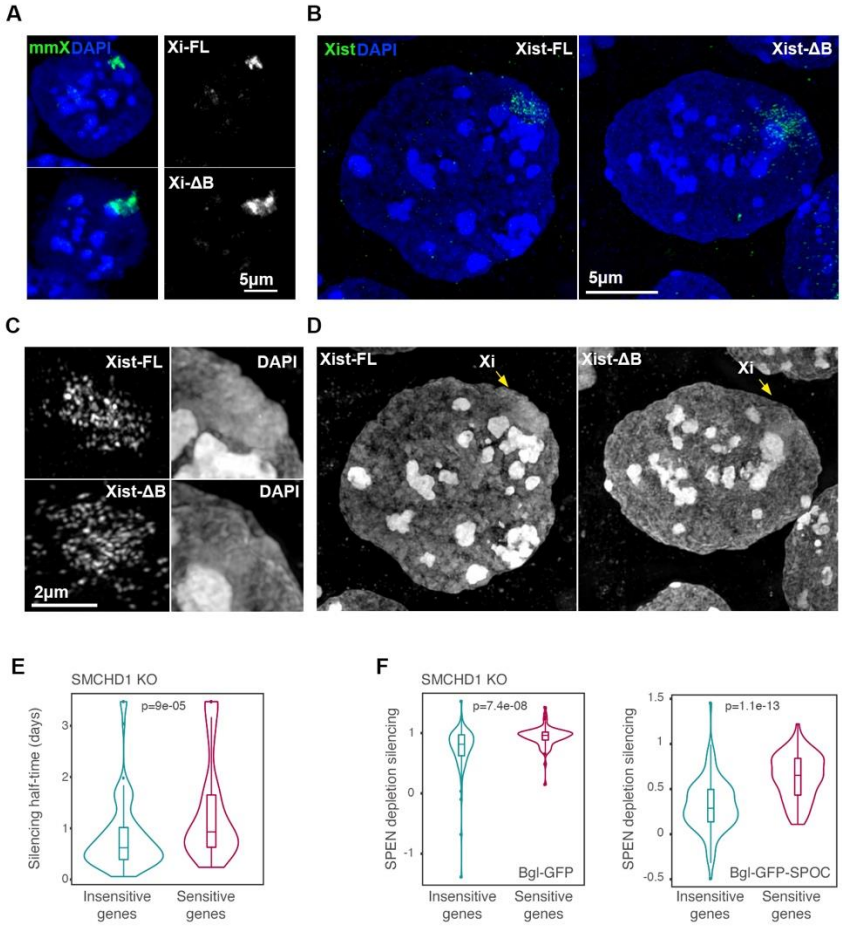



**Figure S14. Gene silencing originates at SMCs and proceeds via PRC1-mediated chromosome compaction**

**A,** Projection of DNA FISH with X-chromosome paints (mmX) in ESCs expressing tet-inducible full-length Xist (FL-Xist) or a deletion-mutant of the B-repeat (ΔB-Xist) after 18hrs of doxycycline induction. DAPI counterstaining is shown in blue.

**B,** 3D-SIM projections of RNA FISH with Xist probes (green) in WT (FL-Xist) or a deletion-mutant of the B-repeat (ΔB-Xist) in female cells at D4 of differentiation. DAPI counterstaining is shown in blue.

**C,** Magnified insets from **B** showing Xist or DAPI signals. Note the characteristic compaction of the Xi territory evident in WT cells, which is not present in ΔB-Xist cells.

**D,** As in **C** showing only the DAPI channel. The Xi regions are indicated by arrows.

**E,** Violin plots showing the silencing half-times (in days) of SCMHD1-sensitive and insensitive X-linked genes. Wilcoxon *p*-value is given.

**F,** Violin plots showing the effect on gene silencing when WT-SPEN is depleted and GFP or the SPOC-domain of SPEN is recruited to Xist via the Bgl tethering, for X-linked genes sensitive and insensitive to SCMHD1 deletion, respectively. 0 indicates complete silencing and 1 complete lack of silencing by $Xist^{GFP}$ or $Xist^{SPOC}$. Wilcoxon *p*-value is given.

| Protein pairs | number of cells | number of particles |
|---|---|---|
| **CIZ1-CELF1, D2** | 18 | 22392 |
| **CIZ1-CELF1, D4** | 17 | 8111 |
| **CIZ1-SPEN, D2** | 9 | 15992 |
| **CIZ1-SPEN, D4** | 8 | 3601 |
| **CIZ1-PCGF5, D2** | 13 | 11756 |
| **CIZ1-PCGF5, D4** | 15 | 3285 |
| **CELF1-PCGF5, D2** | 12 | 16783 |
| **CELF1-PCGF5, D4** | 11 | 15276 |

**Table S1.**

A list of the number of cells and protein particles analyzed in **Figure 3C** is given.

| Protein | number of cells | number of particles |
|---|---|---|
| **CIZ1, D2** | 47 | 22508 |
| **CIZ1, D4** | 35 | 13736 |
| **CELF1, D2** | 23 | 25581 |
| **CELF1, D4** | 31 | 37904 |
| **SPEN, D2** | 16 | 19788 |
| **SPEN, D4** | 10 | 8741 |
| **PCGF5, D2** | 12 | 12902 |
| **PCGF5, D4** | 11 | 22580 |
| **PTBP1, D2** | 11 | 21905 |
| **PTBP1, D4** | 9 | 11815 |
| **EZH2, D2** | 16 | 33851 |
| **EZH2, D4** | 13 | 12734 |
| **RYBP, D2** | 14 | 16552 |
| **RYBP, D4** | 15 | 11688 |
| **hnRNP-K, D2** | 12 | 22662 |
| **hnRNP-K, D4** | 13 | 11688 |
| **MTR3, D2** | 17 | 47314 |
| **MTR3, D4** | 18 | 89464 |
| **ΔIDR SPEN, D2** | 16 | 49283 |
| **ΔIDR SPEN, D4** | 20 | 57429 |

**Table S2.**

A list of the number of cells and protein particles analyzed in **Figures 3, D** to **E, 4H** is given.

| Protein | Location | Feature | D2/D4 *p*-value |
|---|---|---|---|
| **CIZ1** | Nuclear | Adjusted_IntDen | 5.98e-07 |
| **CIZ1** | Xist_Associated | Adjusted_IntDen | 1.56e-01 |
| **CIZ1** | Nuclear | Adjusted_Volume | 5.50e-08 |
| **CIZ1** | Xist_Associated | Adjusted_Volume | 3.00e-01 |
| **CELF1** | Nuclear | Adjusted_IntDen | 6.18e-03 |
| **CELF1** | Xist_Associated | Adjusted_IntDen | 1.94e-06 |
| **CELF1** | Nuclear | Adjusted_Volume | 1.64e-01 |
| **CELF1** | Xist_Associated | Adjusted_Volume | 4.01e-06 |
| **SPEN** | Nuclear | Adjusted_IntDen | 9.83e-02 |
| **SPEN** | Xist_Associated | Adjusted_IntDen | 2.12e-02 |
| **SPEN** | Nuclear | Adjusted_Volume | 1.11e-01 |
| **SPEN** | Xist_Associated | Adjusted_Volume | 3.05e-03 |
| **PCGF5** | Nuclear | Adjusted_IntDen | 2.77e-05 |
| **PCGF5** | Xist_Associated | Adjusted_IntDen | 1.70e-01 |
| **PCGF5** | Nuclear | Adjusted_Volume | 2.98e-02 |
| **PCGF5** | Xist_Associated | Adjusted_Volume | 4.38e-03 |
| **PTBP1** | Nuclear | Adjusted_IntDen | 3.12e-03 |
| **PTBP1** | Xist_Associated | Adjusted_IntDen | 7.53e-03 |
| **PTBP1** | Nuclear | Adjusted_Volume | 3.80e-01 |
| **PTBP1** | Xist_Associated | Adjusted_Volume | 2.47e-01 |
| **EZH2** | Nuclear | Adjusted_IntDen | 2.83e-06 |
| **EZH2** | Xist_Associated | Adjusted_IntDen | 2.83e-06 |
| **EZH2** | Nuclear | Adjusted_Volume | 8.18e-02 |
| **EZH2** | Xist_Associated | Adjusted_Volume | 1.10e-01 |
| **RYBP** | Nuclear | Adjusted_IntDen | 2.55e-06 |
| **RYBP** | Xist_Associated | Adjusted_IntDen | 3.13e-06 |
| **RYBP** | Nuclear | Adjusted_Volume | 5.51e-02 |

| | | | |
|---|---|---|---|
| **RYBP** | Xist_Associated | Adjusted_Volume | 1.06e-01 |
| **hnRNP-K** | Nuclear | Adjusted_IntDen | 1.25e-05 |
| **hnRNP-K** | Xist_Associated | Adjusted_IntDen | 1.25e-05 |
| **hnRNP-K** | Nuclear | Adjusted_Volume | 1.16e-05 |
| **hnRNP-K** | Xist_Associated | Adjusted_Volume | 1.52e-05 |
| **MTR3** | Nuclear | Adjusted_IntDen | 3.40e-07 |
| **MTR3** | Xist_Associated | Adjusted_IntDen | 3.49e-05 |
| **MTR3** | Nuclear | Adjusted_Volume | 3.79e-02 |
| **MTR3** | Xist_Associated | Adjusted_Volume | 1.20e-01 |

**Table S3.**

A list of the p-values derived from a Mann-Whitney Wilcoxon (MWW) test comparing the integrated density and volume of Xist-associated or nuclear fractions of Xist interactors in days 2 and 4 of differentiation is given.

| Protein | Location | MWW *p*-value |
|---|---|---|
| **CELF1** | Nuclear | 6,84e-04 |
| **CELF1** | Xi | 6,84e-04 |
| **CIZ1** | Nuclear | 2,19e-02 |
| **CIZ1** | Xi | 3,45e-04 |
| **PCGF5** | Nuclear | 5,32e-03 |
| **PCGF5** | Xi | 1,96e-02 |
| **SPEN** | Nuclear | 2,85e-01 |
| **SPEN** | Xi | 2,52e-04 |

**Table S4.**

A list of the p-values derived from a Mann-Whitney Wilcoxon (MWW) test comparing the change in density of Xi or nuclear fractions in days 2 and 4 of differentiation is given.

**Table S5.**

A list of probe labelling with specific fluorophores in each experiment is given.

| Figures | Fluorescently labelled probes used in this study |
|---|---|
| **Fig. S1, B and C** | Xist-Atto488, Mecp2-Cy3, Atrx-Cy3 |
| **Fig. S1D** | MS2-Atto488, Xist-Cy3, Atrx-Cy5 |
| **Fig. S3A** | Xist-Atto488, Xist Intron1-Cy3 |
| **Fig. S2A-D** | Xist-Atto488 |
| **Fig. S2E** | XIST-Atto488 |
| **Fig. 4F** | Color-coded as: Green-Atto488, Yellow-Cy3, Red-Texas Red, Purple-Cy5 |
| **Fig. 4I, Fig. S13E** | mmX-Atto488 |
| **Fig. 4J, Fig. S13F-H** | Xist-Atto488 |
| **Fig. S7A** | Xist-488, Mecp2-Cy5 |

**Movie S1.**

**Xist foci are locally confined within a radius of ~250nm**

Live-cell 3D-SIM imaging of Xist-MCP-GFP signals at day 4 of differentiation. Projections of optical stacks acquired through time are shown.

**Movie S2.**

**Xist foci are persistent through time**

Xist foci from **Movie S1** after magnification with bilinear pixel size interpolation for clarity.

**Movie S3.**

**Xist foci are locally confined at chromatin borders facing the IC space**

Live-cell 3D-SIM imaging of $Xist^{MS2\text{-}GFP}$ (green) and H2B-$Halo^{JF647}$ (magenta) at day 4 of differentiation. Projections of optical stacks acquired through time are shown.

**Movie S4.**

**Xist and CIZ1 exhibit a highly correlated motion in the Xi-territory**

Live-cell 3D-SIM imaging of $Xist^{MS2\text{-}GFP}$ (green) and CIZ1-$Halo^{JF647}$ (magenta) at day 4 of differentiation. Projections of optical stacks acquired through time are shown.

# Chapter 3

Research Project

## XIST controls X chromosome dampening and autosomal genes in early human development

Iris Dror[#1], Tsotne Chitiashvili[#1], Anna Sahakyan[#1], Shawn Y.X. Tan[1], Fangyuan Ding[2,3], Jarrett Miller[1] and Kathrin Plath*[1]

1. Department of Biological Chemistry, Molecular Biology Institute, Jonsson Comprehensive Cancer Center, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA 90095, USA

2. Division of Biology and Biological Engineering, Caltech, Pasadena, CA, USA

3. Department of Biomedical Engineering, University of California, Irvine, CA, USA

# co-first authors

* correspondence: kplath@mednet.ucla.edu

**ABSTRACT**

Female human pre-implantation embryos and naïve human pluripotent stem cells (hPSCs) equalize X-linked gene expression with males via X-chromosome dampening (XCD), a unique strategy of dosage compensation in mammals. The mechanisms controlling XCD are unknown. Here, we show that the long non-coding RNA XIST, which mediates X-chromosome inactivation (XCI), is required for XCD. XIST employs similar principles and protein partners, including SPEN, to execute XCD and XCI, but displays a lower accumulation and different distribution on the dampened versus the inactive X. Unexpectedly, XIST also spreads to specific autosomal regions and induces the downregulation of autosomal developmental genes in female naïve hPSCs and pre-implantation embryos. Thus, XIST balances X-linked gene expression but causes imbalances in autosomal gene expression between male and female cells in early human development. Together, our results show that the XIST-SPEN-axis can induce distinct gene expression outputs on the X-chromosome and transiently regulate autosomal genes in humans.

## HIGHLIGHTS

• XIST controls X chromosome dampening in female naïve hPSCs

• SPEN requirement, X chromosome-wide localization of XIST, and escape from gene repression are common principles of XCD and XCI

• XIST localizes to and represses a small number of autosomal genes in naïve hPSCs

• XCD and autosomal gene repression are unique functions of the XIST lncRNA in early human development

## INTRODUCTION

Different dosage compensation strategies have evolved to solve the problem posed by the imbalance of sex chromosomes (Cline and Meyer, 1996; Ferrari et al., 2014; Mahadevaiah et al., 2020; Samata and Akhtar, 2018; Straub and Becker, 2007). In placental female mammals, X chromosome inactivation (XCI) transcriptionally silences one of the two X chromosomes, thereby equalizing X-linked gene expression with XY males (Augui et al., 2011; Avner and Heard, 2001; Brockdorff et al., 2020; Deng et al., 2014; Galupa and Heard, 2015; Gendrel and Heard, 2014; Minkovsky et al., 2012; Payer and Lee, 2008; Plath et al., 2002). XCI occurs at random either on the paternally or maternally inherited X chromosome and initiates early in embryonic development, during the transition from naïve to primed pluripotency, upon implantation of the blastocyst (Lyon, 1961; Mohammed et al., 2017). XCI and the establishment of the inactive X chromosome (Xi)-compartment are fundamentally important for mammalian development and homeostasis (Marahrens et al., 1998, 1997; Schulz and Heard, 2013; Yang et al., 2016).

Although XCI occurs in all placental mammals during post-implantation development, the regulation of X chromosome gene dosage in pre-implantation embryos differs significantly between the species (Okamoto et al., 2011). For example, mouse pre-implantation embryos undergo an imprinted form of XCI, silencing specifically the paternally inherited X chromosome (Takagi and Sasaki, 1975). The silencing of the paternal X chromosome is maintained in extra-embryonic lineages, whereas pluripotent

epiblast cells arising in the blastocyst reactivate the paternal Xi and subsequently undergo random XCI (Mak et al., 2004; Okamoto et al., 2004; Williams et al., 2011). In contrast, in human pre-implantation embryos, both the paternally and maternally inherited X chromosomes remain active (Okamoto et al., 2011; Petropoulos et al., 2016). Yet, genes on both X chromosomes are downregulated by reducing transcription by about half, in a process referred to as X-chromosome dampening (XCD) (Petropoulos et al., 2016; Sahakyan et al., 2017b; Sahakyan and Plath, 2016). Thus, X chromosome dosage compensation in human development occurs by two different and sequential mechanisms, first XCD and later, upon implantation, XCI. Recently it was shown that X chromosome dosage compensation in the human germ line is also achieved by XCD (Chitiashvili et al., 2020), suggesting that certain aspects of human development uniquely require two active, yet dosage compensated, X chromosomes. Yet, since XCD does not occur in mice, which typically serve as model system for mammalian X chromosome dosage compensation, a mechanistic understanding of XCD is lacking.

The long non-coding RNA (lncRNA) XIST (X-inactive specific transcript) is necessary and sufficient for the induction of XCI (Borensztein et al., 2017; Brockdorff et al., 1991; Brown et al., 1991; Marahrens et al., 1997; Penny et al., 1996; Wutz and Jaenisch, 2000). During the initiation of XCI, XIST becomes induced on the future Xi, spreads in cis to cover the entire X chromosome (Engreitz et al., 2013; Simon et al., 2013), interacts with a large number of effector proteins (Chu et al., 2015; McHugh et al., 2015), and induces gene silencing and the formation of the chromosome-wide Xi-compartment (Augui et al., 2011; Avner and Heard, 2001; Brockdorff et al., 2020; Deng

et al., 2014; Galupa and Heard, 2015; Gendrel and Heard, 2014; Minkovsky et al., 2012; Payer and Lee, 2008; Plath et al., 2002). Intriguingly, imaging approaches have shown that, similar to the Xi, the dampened X chromosomes (Xd) in female human pre-implantation embryos and the female germline are associated with XIST RNA (Chitiashvili et al., 2020; Okamoto et al., 2011; Petropoulos et al., 2016; Vallot et al., 2017), revealing a puzzling stable uncoupling of XIST expression from complete gene silencing. In female human pre-implantation embryos, XCD is gradually induced from the 4-cell to the blastocyst stage, which correlates with the upregulation of XIST expression (Petropoulos et al., 2016). Similarly, female primordial germ cells with XCD express XIST, and XIST turns off during meiotic differentiation, which is accompanied by an increase in X-linked transcript levels (Chitiashvili et al., 2020). Moreover, dampening is not observed on the male X in pre-implantation embryos and the germ line, which also express XIST at a much lower level such that the XIST cloud is often not detectable (Chitiashvili et al., 2020; Okamoto et al., 2011; Petropoulos et al., 2016). Despite the correlation between XIST expression and XCD, it is unclear if XIST mediates XCD as, in the past, XIST expression has been functionally linked to the completely inactive status of the X chromosome.

In this study, we set out to explore the role of XIST in XCD. Recent reports have described culture conditions for the establishment and maintenance of naïve hPSCs that capture the state of the pluripotent epiblast cells of human pre-implantation embryos, including features such as XCD and XIST expression (Guo et al., 2017, 2016; Sahakyan et al., 2017a, 2017b; Takashima et al., 2014; Theunissen et al., 2016, 2014; Vallot et al., 2017). Naïve human embryonic stem cells (hESCs) can be derived directly from the

blastocyst or converted from developmentally advanced primed hESCs (Guo et al., 2016; Nichols and Smith, 2009; Takashima et al., 2014; Theunissen et al., 2014). Similarly, naïve human induced pluripotent stem cells (hiPSCs) can be generated by reprogramming from fibroblasts (Liu et al., 2017). Here, we generate and exploit diverse naïve human pluripotent stem cell (hPSC) lines to explore, for the first time, the mechanisms underlying XCD and the role of XIST in XCD.

Applying a CRISPR-Cas9-based loss-of-function approach, we found that transcriptional repression on the Xd is lost in the absence of XIST, which establishes XIST as regulator of both XCD and XCI. Using genomic approaches, we discovered a differential localization and accumulation of XIST on the dampened versus inactive X chromosomes, in addition to shared characteristics. XIST displays a lower enrichment on the Xd compared to the Xi, suggesting that XIST's repressive function is attenuated in early human development due its decreased accumulation in the X-territory. Unexpectedly, we found an additional and new function of XIST. Although XIST is predominantly enriched on the X chromosome in cells with XCD, we identified a remarkable spreading of XIST beyond the X-territory to specific autosomal genes. We provide evidence that the autosomal localization of XIST follows similar principles as on the X chromosome and that XIST transcriptionally downregulates its autosomal target genes in female hPSCs. We show that the female-to-male imbalance of the autosomal genes targeted by XIST is also observed in human pre-implantation embryos. Together, our study uncovers XIST as master regulator of both XCD and XCI and establishes a new

role for XIST in the regulation of autosomal genes, which leads to gene expression differences between male and female human pre-implantation embryos.

## RESULTS

### XIST expression in naïve hPSCs correlates with XCD in *cis*

Here, we set out to evaluate the role of XIST in the regulation of XCD. Transcriptomics studies have shown that 5iLAF-cultured naïve hPSCs (Theunissen et al., 2014) capture the naïve pluripotent state of epiblast cells of human pre-implantation embryos (Huang et al., 2014; Sahakyan et al., 2017a; Theunissen et al., 2016). Accordingly, these cells are considered a model for studies of XCD (Sahakyan et al., 2017a). Yet, a notable difference between 5iLAF-cultured hPSCs and the human pre-implantation embryo is that XIST is expressed from only one of the two X chromosomes in naïve hPSC cultures (Sahakyan et al., 2017a). Although this observation suggests that further culture improvements are needed (An et al., 2020), the monoallelic expression of XIST in naïve hPSCs provides the unique opportunity to explore whether the XIST-associated chromosome is regulated by XCD. We therefore established several female naïve hPSC lines in 5iLAF culture medium to examine the relationship between XCD and monoallelic XIST expression. To this end, we derived female naïve hESC lines, from the primed hESC lines UCLA1 and H9, and a naïve hiPSC line, through transcription factor-mediated reprogramming from fibroblasts. Next, we examined XIST expression and the dosage of X-linked genes through RNA fluorescent in situ hybridization (FISH) and RNA-seq approaches (Table S1).

Consistent with prior reports, RNA FISH for XIST showed that the RNA is expressed from one of the two X chromosomes in most cells, forming a cloud-like signal around its transcription locus (Figures 1a,b). To evaluate X chromosome expression

dynamics in relation to XIST, we took advantage of the fact that the conversion of primed hESCs to the naïve state occurs through an intermediate state in which both X chromosomes are active and XIST is repressed (naïve hPSCs$^{pre\text{-}XIST}$) (Sahakyan et al., 2017a) (Figure S1a). Based on bulk RNA-seq data, we found that the ratio of reads aligned to X-linked or autosomal genes (X/A ratio) negatively correlates with XIST levels (Figure 1c). The lowered X/A ratio in XIST-expressing naïve cells is due to a decrease of X-linked gene expression in the absence of dramatic changes in autosomal gene expression (Figure 1d). These changes bring X-linked gene expression in female naïve hPSCs closer to that of the male naïve hESC line WIN1, which carries a single XIST-negative X chromosome and does not display XCD (Figures 1a,b,e, S1a-c) (Okamoto et al., 2011; Petropoulos et al., 2016; Sahakyan et al., 2017a). Notably, female naïve hPSCs do not completely reach the X chromosome expression level of male naïve hESCs (Figure 1e), consistent with the idea that dampening of X-linked gene expression may only be occurring on one of the two X chromosomes due to monoallelic XIST presence. Single cell (sc) RNA-seq analysis of naïve H9 hESCs corroborated the inverse correlation between XIST expression and the X/A ratio (Figure 1f, S1d). Together, these data show that X chromosome dosage compensation occurs in female naïve hPSCs and that higher XIST levels are associated with stronger repression of X-linked genes.

To confirm that the dosage compensation process observed in naïve hPSCs is not achieved by XCI, we examined X-linked gene expression at allelic resolution. Reads from the X-chromosomes were split according to allele-specific single nucleotide polymorphisms (SNPs) and assigned to a given X chromosome (X1 = Xa and X2 = Xi in

primed hPSCs with nonrandom XCI, see methods). The allelic analysis of scRNA-seq data demonstrated the biallelic expression of the majority of X-linked genes in individual naïve H9 cells (Figures 1g, S1e). Similarly, nascent RNA FISH for the three X-linked genes *SMS*, *SMARCA1* and *GPC3* that are subject to XCI (Tukiainen et al., 2017) (Figures S1f), revealed biallelic transcription in naïve hPSCs (Figures S1g, h). Together, these findings demonstrate that both X chromosomes are active in female naïve hPSCs, in agreement with previous studies (Okamoto et al., 2011; Petropoulos et al., 2016; Sahakyan et al., 2017a).

We next explored whether the XIST-positive X chromosome displays XCD compared to the XIST-negative X chromosome. We first performed allele-specific analysis of population RNA-seq data for different naïve preparations of UCLA1 hESCs and found that when XIST expression is skewed towards one of the two X chromosomes in the population, X-linked gene expression is biased towards the other X (Figures 1h, S1i). Second, applying quantitative RNA FISH, we observed that cells with biallelic expression of the X-linked genes *THOC2 and UTX* (*KDM6A*) display a 25% smaller nascent transcript signal on the XIST-positive versus XIST-negative X chromosome (Figures 1i,j). Qualitatively, we made similar observations for the X-linked genes *SMS*, *SMARCA1* and *GPC3* (Figures S1g, h). Moreover, the lncRNA XACT, which associates with the X chromosome in *cis* (Vallot et al., 2017, 2013), displays a less intense cloud-like signal when transcribed on the same X chromosome as XIST (Figures 1k,l, S1j). Together, these results show that dampening occurs specifically on the XIST-coated X chromosome and additionally suggest that XIST regulates XCD in naïve hPSCs.

**XIST is required for XCD**

To interrogate the function of XIST in XCD further, we first compared the nature of XIST transcripts in naïve hESCs with an Xd and somatic cells with an Xi and found that the transcriptional start site, 3' end, and exon/intron structure are the same in XCD and XCI (Figures S2a,b). These findings suggested that the differential function of XIST on the Xd and Xi is not explained by the predominant expression of alternative isoforms and defined the promoter region of XIST in naïve hPSCs for our subsequent deletion approach.

To directly test if XIST is required for XCD, we applied CRISPR/Cas9-based genome editing to homozygously excise a 2kb region containing the promoter region and 5' end of XIST in the primed H9 hESC line (Figures 2a, S2c,d). We converted two independently derived homozygous knockout (KO) clones (clones C7 and C18) and control wildtype (WT) H9 hESCs to the naïve state in 5iLAF medium (Figure 2a). As expected, XIST was detectable by RNA-seq and RNA FISH in naïve control hESCs but not in targeted clones (Figures 2b-d, S2e). The induction of naïve pluripotency markers (Collier et al., 2017; Theunissen et al., 2016) confirmed the successful conversion to naïve pluripotency (Figure 2e). Moreover, we performed scRNA-seq of control H9 hESCs and the deletion clone C18 and confirmed that naïve hESCs with and without XIST reach the same expression state with two active X chromosomes (Figures S2f,g). Accordingly, most WT and KO cells belong to the same large cluster (cluster 0 with 91%, and 96% of WT and KO cells, respectively; Figure 2f).

Since the naïve pluripotent state could be established and maintained in the absence of XIST, we were able to interrogate the gene expression data for whether the absence of XIST results in the loss of XCD. Based on bulk and scRNA-seq data, we found a significant increase of the X/A ratio in naïve hESCs lacking XIST (Figures 2g,h, S2h,i), which, at the global level, is due to the upregulation of X-linked genes compared to autosomal genes (Figures 2i, S2j). Thus, in the absence of XIST, the global level of X-linked transcripts is significantly increased, whereas global autosomal gene expression levels remain constant. Together, these data demonstrate that X chromosome dosage in naïve hESCs is controlled by XIST. These findings establish XIST as regulator of not only XCI but also XCD in humans.

**Genes subject to XCD are affected by XIST deletion more than XCD escapees**

A key characteristic of XIST-mediated XCI is that a subset of X-linked genes escapes silencing (Berletch et al., 2015; Carrel and Willard, 2005; Deng et al., 2014; Navarro-Cobos et al., 2020; Tukiainen et al., 2017). Therefore, we explored whether X-linked genes can also escape XCD. Typically, allele-specific expression based on SNPs or the presence of biallelic nascent expression of X-linked genes by RNA-FISH are used to define XCI escapees, but these approaches cannot easily discern XCD escapees due to the inherent biallelic expression of X-linked genes. Consequently, we defined XCD escapees as those genes with higher expression in female versus male naïve hPSCs. Since most X-linked genes are skewed to higher expression in female cells because our

hPSC lines only have one dampened X chromosome, we only classified genes with the most significant gene expression differences as escapees (see methods, Table S2).

This analysis identified 157 (H9), 122 (UCLA1), and 67 (iPSCs) X-linked genes with significantly higher transcript level in 5iLAF-cultured female hPSCs compared to the naïve male hESC line WIN1 (Table S2), defining 8-25% of X-linked genes as escapees (Figure S3a). XCD escapees are distributed over the entire X-chromosome (Figure S3b). 64 escapee genes were conserved in two female hPSC lines and 15 in all three (Figure S3c). XCD escapees also overlapped with genes that were unchanged or even upregulated upon transition from the pre-XIST to the XIST-positive naïve state (Figure S3d). Similarly, XCD escapees defined in 5iLAF-cultured naïve cells overlapped with those defined in naïve hESCs that were derived directly from blastocysts with a different naïve culture method (t2iLGö) (Guo et al., 2016; Takashima et al., 2014) (Figures S3e,f, Table S2). We conclude that XCD escape occurs in female naïve hPSCs regardless of cell line, origin of derivation (from primed hESCs, fibroblasts, blastocyst), and culture condition. Notably, even upon removing the XCD escapees from our analysis, X-linked genes were biallelically expressed in female naïve hESCs (Figure S3g), providing further evidence that naïve cells achieve dosage compensation by XCD rather than by XCI.

Next, we determined whether the XCD escapees defined in female hPSCs capture the endogenous regulation in female pre-implantation embryos. Here, we compared the expression of X-linked genes between female and male epiblast cells from pre-

implantation blastocysts at day 7 of development (E7), taking advantage of previously published scRNA-seq data (Petropoulos et al., 2016). Due to XCD on both X chromosomes in female embryos, female and male epiblast cells express a comparable overall dose of X-linked genes (Petropoulos et al., 2016). Yet, we identified 86 genes as XCD escapees (Figures S3a,b, Table S2), of which 7 overlapped with the 15 XCD escapees conserved among our three female 5iLAF-cultured hPSC lines (Figure S3h). Thus, escape from dosage compensation is a conserved feature of XCD *in vivo* and *in vitro*.

XCD escapees are, on average, more highly expressed than the remaining X-linked genes, even in cells that lack XCD (like in the male naïve hESC line WIN1, pre-XCI naïve UCLA1 and male embryos) (Figures S3i,j), suggesting that the higher gene expression level plays a role in the ability of a gene to escape XCD. Consistent with the regulation of both XCD and XCI by XIST, we also found significant overlap among XCI and XCD escapees (Figure S3k). For instance, the X-linked gene *UTX* (*KDM6A*) is a well-known XCI escapee that also escapes XCD in naïve UCLA1 hESCs and epiblast cells (Figure S3l). Taken together, a conserved set of genes can escape both XCI and XCD consistent with the regulation of both processes by the XIST lncRNA.

Interestingly, our population- and scRNA-seq data showed that genes subject to and escaping XCD were upregulated in the absence of XIST, yet genes subject to XCD displayed significantly higher upregulation (Figures 2j, S2k,l). This result is consistent with

the observation that the escapee *KDM6A* is repressed by 25% on the XIST-expressing X-chromosome compared to the other X-chromosome in the same cell (Figure 1i). In summary, these results demonstrate that a subset of highly expressed genes escape XCD at least partially, yielding higher expression level of these genes in female embryos and naïve hPSCs compared to their male counterparts.

**XIST is differentially distributed on the Xd versus Xi**

We previously showed that Xist localizes broadly across the entire Xi in mouse somatic cells. To interrogate whether XIST spreads across the Xd in naïve hPSCs, we performed RNA antisense purification (RAP)-seq, a biochemical method that enables high-resolution mapping of RNA localization on chromatin (Engreitz et al., 2013). We applied RAP-seq to the female naïve H9 and UCLA1 hESC lines and the naïve hiPSC line. Additionally, since XIST has not been previously mapped on the human Xi, we also mapped the localization of the RNA in human female fibroblasts in duplicate, enabling a comparison of XIST localization on the Xd and Xi. We designed biotinylated antisense probes that hybridize to the spliced human XIST transcript to purify the RNA and its associated genomic DNA from crosslinked cell lysates (see methods, Table S3), and sequenced the genomic DNA that co-purified with XIST RNA. Confirming that the RAP-seq approach specifically captured XIST interactions, we found that the XIST transcription locus with its accumulating nascent transcripts was among the most highly enriched genomic region (Figure S4a).

Examining the RAP-seq data, we found strong XIST enrichment on both the Xd and the Xi; with >38% of the sequencing reads originating from the X chromosome versus ~5% from the input DNA samples in naïve hPSCs and fibroblasts (Figures 3a,b). Moreover, XIST localized across the entire Xd and Xi (Figures 3c, S4b). Indeed, >94% of all 100kb windows along the Xd and Xi are enriched for XIST more than 2-fold (>89% 3-fold, >79% 4-fold, and >68% 5-fold). Thus, XIST displays a broad localization pattern across the Xd and Xi despite the clear differences in X-linked gene regulation.

Although XIST is enriched across the entire X chromosome, we found that the RNA is differentially distributed across the Xd versus the Xi (Figures 3c,d). Consistent with this observation, XIST enrichment across the X was more highly correlated between naïve hPSC and fibroblast data sets, respectively, than between these cell types (Figure S4c). Comparing XIST distribution across the X, we identified a 80 Mb region across the center of the chromosome that was preferentially bound by XIST on the Xi in fibroblasts and a 40 Mb region, at the end of the X, that was preferentially bound by XIST on the Xd in naïve hPSC (Figures 3d, S4d). The switch from the Xi- to Xd-biased XIST localization occurred at the macro-satellite repeat locus *DXZ4* (Figures 3d, S4d), which is known to partition the Xi in somatic cells into two spatial super domains (Bonora et al., 2018; Darrow et al., 2016; Deng et al., 2015; Froberg et al., 2018; Giorgetti et al., 2016; Horakova et al., 2012; Rao et al., 2014; Wang et al., 2016) (Figure S4e). The distribution of XIST over the first 30Mb is more similar between the Xd and Xi (Figure 3d), which may be related to the fact that many XCI escapers are concentrated in this region (Tukiainen et al., 2017). Consistent with this, intra-chromosomal interactions in this chromosomal region are

similar between the Xa and Xi in somatic cells than elsewhere on the chromosome (Figure S4e). Together, these results suggest that chromatin structure plays a role in controlling the precise localization of XIST along the Xd and Xi and that the differential distribution of the RNA may play a role in the differential gene regulation.

To further characterize the distribution of XIST across the Xd and Xi, we correlated the enrichment of XIST with diverse genomic features (Table S4). On the Xd, XIST enrichment correlated best with gene density, SINE elements, mammalian-wide interspersed repeats (MIRs), and the LINE element L2 (Figure 3e). MIRs and L2 elements are often located in gene rich regions (Medstrand et al., 2002) and both carry enhancer characteristics including high levels of H3K27ac and DNase I hypersensitivity (Cao et al., 2019), consistent with the association of XIST localization with gene density. The positive correlation with gene density and MIR elements also applied to XIST's association with the Xi, albeit it was much weaker compared to the Xd (Figure 3e). XIST enrichment on both the Xd and Xi was negatively correlated with Long Terminal Repeats (*LTRs)* retrotransposons, particularly with ERVL-MaLR elements (Figure 3e). Closer inspection of XIST localization at candidate escapees in female naïve hPSCs revealed a reduction of XIST levels over the gene bodies of some XCD escapees. For example, the XCD escapee *KDM6A* is in a genomic region with XIST depletion relative to neighboring intergenic regions (Figure 3f). A quantitative analysis confirmed that genes subject to XCD tend to be more enriched for XIST than genes that escape XCD in naïve hPSCs (Figures 3g, S4f) or in epiblast cells (Figure 3h). Interestingly, for XCI escapees this difference appears more pronounced (Figures 3f,i). We conclude that XIST preferentially

localizes to gene-rich regions on the Xd in naïve hPSCs as well as the Xi in fibroblasts. Genes that escape from XCD or XCI are associated with reduced XIST localization on the Xd and Xi, respectively, albeit to larger effect size on the Xi.

These data reveal that although XIST distribution differs between the Xd and Xi, its localization follows similar principles, highlighting important parallels between XCD and XCI. The X chromosome-wide localization of XIST on the Xd in naïve hPSCs reinforces the role of XIST as master regulator of XCD. Yet, we noted that the overall level of XIST enrichment differs dramatically between the Xi and Xd (Figures 3a,b), as discussed below.

**XIST spreads beyond the X chromosome territory in naïve hPSCs**

We surprisingly found that the proportion of reads that aligned to the X in the XIST pulldowns was reproducibly larger in fibroblasts than in naïve hPSCs (Figure 3a). Consequently, more genomic regions showed higher enrichment of XIST RNA on the Xi versus Xd (Figure 3b). Specifically, whereas 41-46% of genomic windows on the Xi displayed an at least 20-fold XIST enrichment, only 2-17% of windows were similarly enriched for XIST on the Xd (Figure 3b). Consequently, XIST accumulates on the Xi at a dramatically higher overall level compared to the Xd (Figure S4g). Thus, the inability of XIST to induce complete gene silencing in naïve hPSCs is associated with decreased overall levels of the RNA on the X chromosome.

Given the lower overall accumulation of XIST on the Xd, we hypothesized that XIST spreads beyond the X chromosome in naïve hPSCs. In support of this idea, XIST forms a less compact cloud-like signal on the Xd than the Xi (Okamoto et al., 2011; Petropoulos et al., 2016; Sahakyan et al., 2017a; Vallot et al., 2017) (Figure 4a). Moreover, DNA FISH with an X-chromosome paint combined with RNA FISH revealed that XIST spreads beyond the Xd territory in naïve UCLA1 hESCs, whereas it is restricted to the Xi-territory in fibroblasts (Figure 4b). To further characterize the unique localization pattern of XIST in naïve hPSCs, we inspected the XIST RAP-seq data for autosomal binding of the RNA.

Remarkably, consistent with the imaging data, we found that the reduced accumulation of XIST on the Xd in naïve hPSCs was accompanied by localization of XIST to autosomes, which is not the case in female fibroblasts (Figure 4c). More than 53% of reads from the XIST pulldown aligned to autosomes in naïve hPSCs, compared to only 18% in fibroblasts (Figure S5a). It is worth noting that the steady state level of XIST transcripts is slightly lower in naïve hPSCs (with an Xd) compared to somatic cells (with an Xi) (Figure S5b). This result indicates that the association of XIST with autosomes in naïve hPSCs cannot simply be explained by a saturation of all binding sites on the X chromosome.

Interestingly, in the three female naïve hPSC samples, only 6-10% of 100 Kb autosomal windows are enriched for XIST (Figure 4c), indicating that only a small portion

of the autosomal genome is targeted by the RNA. Taken together, these results show that the reduced accumulation of XIST in the X chromosome territory in naïve hPSCs is accompanied by the localization of the RNA to a small set of autosomal regions. Moreover, autosomal targeting of XIST is a feature uniquely associated with the naïve pluripotent state.

**XIST spreads to specific autosomal regions in naïve hPSCs**

To explore the unprecedented localization of XIST to autosomal regions in naïve hPSCs, we defined XIST-enriched autosomal regions using the MACS2 broad peak calling algorithm (Zhang et al., 2008) (Table S5). We identified thousands of XIST peaks in naïve hPSCs (6873 in H9, 9835 in UCLA1, and 9325 in iPSCs) (Figure 4d), targeting all autosomes (Figures 4e, S5c/d). In contrast, in fibroblasts, where we expected no localization of XIST to autosomal regions, we observed 79 and 100 significant autosomal peaks in the two RAP-seq replicates (q-value<0.05, Figures 4d/e, S5c/d). Close examination of these peaks revealed that most fall into highly repetitive genomic regions such as those around centromeres and telomeres (Figure S5d), and therefore likely represent false alignment rather than real XIST localization.

771 of the autosomal XIST peaks were conserved in all three different hPSC lines (UCLA1, H9 and iPSCs) (Figure 4f). None of these conserved autosomal XIST peaks were detected in fibroblasts (Figure S5e). In addition, 2355 autosomal peaks overlapped in 2 out of the 3 naïve hPSC datasets (Figure S5e). The conservation of peaks correlates

with the level of XIST enrichment, such that peaks present in all three naïve hPSC lines showed a higher level of XIST than those detected in only a subset of lines (Figures 4g, S5f). Conserved autosomal XIST peaks, i.e. those overlapping among all three naïve hPSC lines, are on average 52 kilobases long (Figure S5g) and present on all autosomes except chromosome (chr) 13 (Figures 4h, S5h). Proportionally, chr11 (106 peaks, 4.9% of the chromosome), chr20 (57 peaks, 3.3% of the chromosome), and chr1 (101 peaks, 3.2% of the chromosome) carry the most peaks and also contain large clusters of XIST peaks (Figures 4f,h). For comparison, only two conserved autosomal peaks were identified in the fibroblast replicates (Figure S5d, Table S5).

Taken together, the identification of conserved genomic targets of XIST on autosomes reveals that the autosomal localization of XIST in naïve hPSCs is not a stochastic process. Our results identify a unique phenomenon whereby XIST can spread beyond the X chromosome it is expressed from, onto other chromosomes. These data describe for the first time a stable association of XIST RNA with a chromosome *in trans*.

**XIST spreads to regulatory regions on the autosomes**

To get insights into the localization and function of XIST on autosomes, we explored the enrichment of diverse genomic and epigenomic features associated with autosomal XIST peaks. Similar to observations on the X chromosome, we found that XIST accumulation on autosomes was best correlated with MIR and L2 repetitive elements (Figure 5a, Table S4), suggesting that similar mechanisms govern the localization of XIST on the X and on

autosomes. Consistent with the association of MIRs and L2 elements with gene regulatory regions (Cao et al., 2019; Petri et al., 2019), we found that XIST localizes to autosomal regions containing gene bodies. For example, conserved autosomal XIST peaks extend across 492 autosomal protein-coding genes, representing 2.6% of all autosomal protein-coding genes (Figure 5b, Table S6). These genes are significantly enriched for differentiation and developmental gene ontology terms (p-value < 0.000005, Figure 5c) and include various transcription factor-encoding genes such as *ARNTL*, *CAMTA1*, *CTCFL*, *ESRRB*, *ETV5*, *FLI1*, *GLI2*, *HIVEP3*, *HOXA10*, *HOXA9*, *IRX6*, *KLF6*, *PKNOX2*, *TEAD1*, *ZNF423*, and *ZNF618.* Some of these genes are known to play a critical role in pluripotent stem cells, as for example Estrogen Related Receptor Beta (ESRRB) (Festuccia et al., 2012; Martello et al., 2012). Other examples of XIST targets on autosomes include the gene encoding Transcriptional-regulating factor 1 (*TRERF1*), a progesterone receptor coactivator on chr6 (Gizard et al., 2006), Transforming Growth Factor alpha (*TGFA*) on chr2, and Cluster of Differentiation 5 and 6 (*CD5* and *CD6*) involved in the activation and differentiation of lymphocytes on chr11 (Gimferrer et al., 2003) (Figure 5d).

Autosomal genes targeted by XIST tend to be significantly more lowly expressed than those not associated with XIST, in both female and male naïve hPSCs (Figure 5e). This result suggests that the gene expression state is one factor that guides XIST localization to specific autosomal regions. To further explore the nature of genes under autosomal XIST peaks, we mined ENCODE ChIP-seq data across a large variety of cell types (see methods). We found that the binding sites of multiple regulators involved in

transcriptional repression, including the Polycomb Repressive Complex 2 (containing SUZ12 and EZH2), NRSF and CtBP2, are enriched within autosomal XIST peaks (Figure 5f Table S4), consistent with the role of XIST-associated genes in developmental regulation and their lower overall expression.

Our results reveal the unexpected and remarkable localization of XIST to a small subset of developmental regulators on autosomes in female naïve hPSCs. In mouse and human, Xist/XIST can silence autosomal genes when ectopically expressed from an autosome (Hall et al., 2002a; Kelsey et al., 2015) or when spreading into autosomal regions on X:A translocations (Hall et al., 2002b; Yang et al., 2011). For example, integration of XIST into one of the three chr21 copies in primed human iPSCs derived from Down Syndrome patients enables the silencing of the XIST-associated chromosome and corrected gene expression to nearly the normal disomic chr21 levels (Jiang et al., 2013). Although it was known that XIST is inherently able to regulate autosomal genes, the physiological relevance of this capacity was unclear. Our observations raise the interesting possibility that XIST may transcriptionally regulate autosomal developmental genes during early human development and in naïve hPSCs.

**XIST mediates the repression of autosomal genes**

To explore this exciting possibility, we compared gene expression levels between female naïve hPSCs with XIST (UCLA1, H9, and iPSCs) and the male naïve hESC line WIN1, which lacks XIST. We found that autosomal genes targeted by XIST had significantly

lower expression in female versus male naïve hPSCs, whereas genes that do not overlap XIST peaks showed similar expression in male and female hPSCs (Figure 6a). Similarly, XIST was more highly enriched across genes that are more lowly expressed in female versus male naïve hPSCs, compared to genes with similar expression in female and male cells (Figure 6b).

We identified 87 autosomal genes that are located within conserved XIST peaks in naïve hPSC and more lowly expressed in all three female naïve hPSC lines compared to WIN1 (Table S6). Among these genes is *SPON1* (*SPONDIN1*), which is encoded on chr11 and known to promote neuronal differentiation (Gyllborg et al., 2018) (Figure 6c). Consistent with lowered expression of *SPON1* in female cells, RNA FISH for the nascent transcripts of *SPON1* in female and male naïve hESCs revealed a higher proportion of female cells with monoallelic expression of this gene compared to male cells (Figure 6d). 82 additional genes were more lowly expressed in two of the three female naïve hPSC lines compared to male WIN1 hESCs (Table S6), including *HUNK* (Hormonally Up-Regulated Neu-Associated Kinase) on chr21 which has a role in the control of proliferation and *differentiation of epithelial cells* (Gardner et al., 2000) (Figure 6e). For both *HUNK* and *SPON1*, we found a negative correlation between their expression level and XIST transcript levels (Figure 6f). This result held true for other genes that overlapped with autosomal XIST peaks. Specifically, XIST targets have lower expression in cells where XIST is highly expressed and vice versa, whereas genes found in XIST-depleted autosomal regions show no correlation to XIST transcript levels (Figure 6g). Overall,

these data strongly support the conclusion that XIST mediates the transcriptional repression of a small number of autosomal genes in naïve hPSCs.

To directly interrogate the role of XIST in the regulation of autosomal genes, we examined the regulation of XIST-targeted autosomal genes upon loss of XIST, using the scRNA-seq expression data generated for naïve WT and KO H9 hESCs (Figure 2). We found a small, yet significant, correlation between the degree by which an autosomal XIST target is upregulated in XIST KO cells and the enrichment of XIST around that gene in WT cells (Figure 6h). In addition, genes associated with autosomal XIST peaks were significantly upregulated upon XIST deletion compared to genes in XIST-depleted regions (Figure 6i). These data show that XIST spreading beyond the X chromosome results in the downregulation of autosomal genes in naïve hPSCs, providing evidence for a role of endogenous XIST expression in controlling autosomal gene expression.

An important question is whether the regulation of autosomal genes by XIST is recapitulated during human pre-implantation development. Intriguingly, both *SPON1* and *HUNK* were more lowly expressed not only in female compared to male hPSCs, but also in female versus male epiblast cells of human pre-implantation embryos (Figure 6j). At a global level, we found that genes in XIST-targeted autosomal regions display significantly lower expression in female compared to male epiblast cells in E6 and E7 pre-implantation embryos (Figure 6k). Interestingly, cells from E5 embryos did not show this sex-specific difference (Figure 6k), consistent with the prior report that XIST is strongly upregulated in

female embryos only between E5 and E6 (Petropoulos et al., 2016). Similarly, autosomal genes with significantly lower expression in female compared to male epiblast cells *in vivo*, are associated with higher XIST enrichment compared to genes that showed no significant differences between female and male epiblasts (Figure 6l).

Taken together, these findings show that the sex-specific expression of XIST-enriched autosomal genes observed in naïve hPSCs extends to the human embryo and point to a novel role of XIST in downregulating autosomal genes during human embryonic development. This new function of XIST results in expression differences of a subset of autosomal genes between female and male embryos and hPSCs.

**SPEN is required for XCD and the repression of autosomal XIST targets**

To further explore the mechanism by which XIST regulates XCD and autosomal genes in naïve hPSCs, we examined whether these processes share additional features with the XCI. First, we explored if XCD is associated with changes in chromosome organization, which is a hallmark of XCI (Pandya-Jones and Plath, 2016). Indeed, combining RNA FISH for XIST with DNA FISH for specific genomic locations on the X chromosome, we found that several high-order interactions differ between the XIST-coated Xd and the Xa, yet in a different way compared to the Xi (Figures S6a,b). Second, we began to explore the role of effector proteins of XIST. Xist interacts with various proteins to establish gene silencing and the Xi compartment during XCI in mouse cells (Chu et al., 2015; Dossin et al., 2020; McHugh et al., 2015; Minajigi et al., 2015; Monfort et al., 2015; Nesterova et al., 2019;

Pandya-Jones et al., 2020), allowing us to explore whether known interactors of XIST participate in XCD and autosomal gene regulation. We initially assessed the localization of the protein Cip1-interacting zinc finger protein 1 (CIZ1), which binds to the E-repeat sequence of Xist and restricts the localization of XIST to the X-territory during XCI (Pandya-Jones et al., 2020; Ridings-Figueroa et al., 2017; Sunwoo et al., 2017). Immunostaining for CIZ1 in female naïve UCLA1 hESCs revealed a strong enrichment on the Xd in most cells (Figure 7a), showing that XIST exploits CIZ1 for both XCI and XCD.

Next, we tested the role of SPEN. SPEN, also known as SMRT and HDAC associated repressor protein (SHARP), binds the 5' A-repeat sequence of mouse Xist and is the key epigenetic repressor of the XCI process in mouse embryos and mouse ESCs (Dossin et al., 2020; Lu et al., 2016; McHugh et al., 2015; Monfort et al., 2015; Nesterova et al., 2019). SPEN mediates gene silencing on the Xi by interacting with co-repressors and activating HDAC3 (Dossin et al., 2020; McHugh et al., 2015). To examine whether SPEN is required for gene repression on the Xd and on autosomes, we knocked down SPEN in female naïve UCLA1 hESCs by RNAi (Figures 7b, S6c). As a control, we also depleted SPEN in male naïve WIN1 hESCs (Figures 7b, Fig S6c). SPEN knockdown resulted in a strong increase of global transcript levels from the X chromosome relative to autosomes in female cells and a much smaller increase in male cells (Figures 7c,d, S6d-f). The female-biased upregulation of X-linked genes was not associated with major changes in steady-state XIST levels (Figure S6g) or in the expression of naïve pluripotency-related genes (Figure S6h) and affected genes subject to XCD more strongly

than XCD escapees (Figure 7e). Thus, similar to XCI, SPEN is necessary for XCD, suggesting that XIST mediates XCD through its interaction with SPEN. The slight upregulation of X-linked genes upon SPEN depletion in male hESCs (Figures 7d,e) suggests that SPEN also regulates gene expression on the active X chromosome, consistent with a broader role of SPEN in gene regulation.

Exploring SPEN's role in the regulation of XIST-targeted autosomal genes, we found that genes overlapping autosomal XIST peaks were weakly, yet significantly, upregulated upon SPEN depletion in female naïve hESCs (Figure 7f). Similarly, autosomal genes that were upregulated upon SPEN knockdown showed higher XIST enrichment compared to those that did not change expression (Figure 7g) and significantly overlapped with genes upregulated upon deletion of XIST (Figure 7h), demonstrating that SPEN plays a role in autosomal gene regulation together with XIST. Interestingly, a similar upregulation of these genes also occurred upon SPEN knockdown in male hESCs (Figures 7f-h), suggesting that SPEN presence at these genes may contribute to the recruitment of XIST to specific autosomal regions. Overall, these results indicate that XIST-targeted autosomal regions are regulated by both XIST and SPEN, demonstrating that similar mechanisms are exploited in the repression of X-linked and autosomal genes by XIST in naïve hPSCs and in the XCI process.

## DISCUSSION

The role of XIST in the unique XCD process of female human pre-implantation embryos and naïve hPSCs has been a mystery. We uncovered that XIST is required for this process (Figure 7i). Importantly, we also made the unexpected discovery that XIST regulates specific autosomal genes during this stage of development, leading to differential gene expression between females and males (Figure 7i).

Our findings reveal that XIST can generate functionally different outputs of gene expression as human development proceeds: first dampening and later silencing, and give first insights into why XIST mediates dampening in one developmental stage and silencing in another. The remarkable plasticity of XIST's ability to regulate gene expression provides the exciting possibility that nuclear compartments with a specific gene expression output can be rationally engineered. To our knowledge, XIST is the first lncRNA in which the same transcript can produce two distinct regulatory outputs during development. Prior work uncovered opposing gene regulatory roles for the lncRNA locus Haunt, yet these distinct roles arise from the function of the lncRNA transcript on one hand and an enhancer-like function of the genomic locus on the other hand (Yin et al., 2015). It is notable that Xist is normally not expressed in undifferentiated naïve mouse ESCs, yet the ectopic induction of Xist in these cells enables XCI and not XCD, suggesting that it is not simply the developmental state that determines the gene regulatory output of XIST. Therefore, it will be important to decipher whether the sequence of the RNA, which differs between mouse and human even within the

conserved repeat sequence regions (Nesterova et al., 2001), contributes to the differential ability of XIST in human versus mouse naïve hPSCs and pre-implantation embryos.

Our study reveals that XIST executes XCD and XCI through similar principles. Specifically, on the Xd and Xi, the same isoform of XIST transcript is expressed, XIST localizes over the entire X chromosome, recruits similar protein effectors (CIZ1 and SPEN) and requires SPEN for transcriptional repression (Figure 7i). In addition, a subset of X-linked genes escapes the regulation by XIST in both cases. These findings support our conclusion that XCD and XCI are regulated by XIST and suggest that a large number of effector proteins interacting with the RNA are conserved between XCD and XCI.

Our high-resolution mapping of XIST's chromatin association showed that XIST localizes over the entire Xd, as seen on the Xi. However, despite this extensive spread, we found striking differences in XIST accumulation on the Xd and Xi. First, XIST enriches most strongly in different regions on the Xd and Xi. The spatial organization of the X chromosome may contribute to the distinct enrichment pattern of XIST, since the differences in XIST localization are linked to the macro-satellite repeat *DXZ4* that partitions the Xi into two spatial superdomains (Bonora et al., 2018; Darrow et al., 2016; Deng et al., 2015; Froberg et al., 2018; Giorgetti et al., 2016; Horakova et al., 2012; Rao et al., 2014; Wang et al., 2016). This conclusion is consistent with prior observations that Xist exploits the spatial organization of the X chromosome for the initial spread across the X chromosome during XCI initiation (Engreitz et al., 2013). Second, we observed reduced

overall levels of XIST on the Xd compared to the Xi. Since reduced accumulation of XIST in the X-territory is associated with XCD, our results suggest a model where the concentration of XIST within the X chromosome is the critical determinant for the magnitude of XIST-mediated gene repression (Figure 7i).

The reduced concentration of XIST on the Xd versus Xi may reduce the recruitment of effector proteins to the Xd, leading to XCD instead of XCI. Xist and its binding partners have been proposed to execute the maintenance of XCI via phase separation (Cerase et al., 2019; Pandya-Jones et al., 2020), which might be hindered at lower concentrations. Alternatively, proteins involved in regulating XIST localization and function may be present at different levels or differentially modulated in cells with XCD and XCI. For instance, the absence or decreased level of proteins such as PTBP1, MATR3, TDP-43, SAF-A, CIZ1, and CELF1, all involved in regulating the localization of mouse Xist (Hasegawa et al., 2010; Pandya-Jones et al., 2020), or the modulation of their function, as for instance through posttranslational modifications, may alter XIST's ability to regulate gene expression.

Consistent with prior observations (Okamoto et al., 2011; Petropoulos et al., 2016; Sahakyan et al., 2017a; Vallot et al., 2017), our imaging data revealed a dispersed cloud of XIST surrounding the Xd in human pre-implantation embryos and naïve hPSCs. Our genomic approach uncovered specific autosomal regions to which XIST localizes in cells with XCD. These results support the possibility that XIST association with chromatin

beyond the X chromosome territory results in a reduced local concentration on the Xd in naïve hPSCs. It is conceivable that XIST is lost from the X chromosome and associates with chromatin regions on chromosomes in spatial proximity. Alternatively, autosomal regions may effectively compete for the capture of XIST and thereby reduce the accumulation of the RNA on the X chromosome. Regardless of the mechanism, the repression of autosomal genes by XIST requires SPEN and autosomal XIST localization occurs at genic regions that contain L2 and MIR transposable elements, as seen on the X. These observations indicate that autosomal gene repression by XIST and the localization of the RNA follow similar principles as on the X chromosome.

Future efforts that modulate the expression of XIST on the Xd and define all the XIST-interacting proteins, together with their posttranslational modifications will help to uncover the mechanisms that are critical for XCD versus XCI and lead to a quantitative framework of how RNA abundance, sequence elements, protein state and developmental stage affect the function and localization of XIST.

The finding that XIST localizes to specific autosomal regions was surprising as, in the mouse system, Xist has always been shown to be confined to the X chromosome from which it is transcribed (Jonkers et al., 2008). Multiple studies have reported that the ectopic expression of Xist from an autosome is sufficient to induce autosomal gene silencing in undifferentiated mouse ESCs. However, to our knowledge, autosomal gene regulation by endogenously expressed XIST has not been reported. Intriguingly,

autosomal target genes of XIST are involved in the regulation of developmental processes. By comparing gene expression levels between females and male naïve hPSCs and human pre-implantation embryos, we discovered 87 autosomal genes that are more lowly expressed in female cells compared to males and also regulated by XIST. Although the differences in autosomal gene expression are relatively small between male and female cells, or in female cells with and without XIST, they are consistently found in several comparisons including epiblast cells. Further studies are necessary to define the consequences of the female-specific downregulation of autosomal genes. Regardless, the small but consistent downregulation of autosomal genes by XIST provides an unexpected and exciting new role for XIST during early human development.

Intriguingly, XCD resembles the X chromosome dosage compensation mechanism in *Caenorhabditis elegans* (Crane et al., 2015; Ercan et al., 2009; Strome et al., 2014), which also reduces the activity of genes on both X chromosomes in hermaphrodites. In this case, a specialized condensin complex binds to and partially represses transcription from each of the two X chromosomes. Our findings demonstrate that partial X chromosome repression on the Xd and in *C.elegans* are achieved by different regulatory processes, further adding to the excitement about X chromosome dosage compensation mechanisms.

In summary, with the demonstration that XIST mediates XCD, our work enables future studies of the role of dosage compensation by XCD in human pre-implantation

embryos and the human germ line. Moreover, it raises the question of whether the autosomal gene regulation by XIST is critical for early human development and the differentiation of the germ line.

**Acknowledgments**

We are grateful to Robin Mckee, Yolanda Markaki, Amanda Collier, Anna Afasizheva and Konstantinos Chronis for help with experiments. We thank the Guttman Lab at Caltech for designing the sequences for the XIST probes for RAP-seq experiments; Austin Smith and Rudolf Jaenisch for kindly providing the HNES1/3 and WIN1 hESC lines, the Human Embryonic and Induced Pluripotent Stem Cell Core, the Microscopy Core of the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Center (BSCRC) for help with imaging, and the Technology Center for Genomics and Bioinformatics at the UCLA Johnson Comprehensive Cancer Center as well as the Next Generation Sequencing Core of BSCRC for help with genomics approaches. T.C. was supported by a Boehringer Ingelheim PhD Fellowship; A.S. by the Ruth L. Kirschstein NRSA F31 Fellowship (GM115122), an Iris Cantor-UCLA Women's Health Center Executive Advisory Board award (NCATS UCLA CTSI Grant Number UL1TR000124, together with K.P.), the Philip Whitcome Pre-Doctoral Fellowship, and the Mangasar M. Mangasarian Scholarship, and a UCLA Dissertation Year Fellowship; S.Y.X.T. by a A*STAR National Science Scholarship and a PhD Fellowship from the BSCRC; and K.P. by an Innovation Award from the BSCRC, the David Geffen School of Medicine and the Jonsson Comprehensive Cancer Center at UCLA, the NIH (R01HD098387), and a Faculty Scholar grant from the Howard Hughes Medical Institute.

## Author contributions

A.S., I.D. and K.P. conceptualized the project. A.S. and T.C. generated cell lines, conducted imaging experiments and performed bulk RNA-seq; T.C. generated single cell RNA-seq data, the XIST KO, and performed SPEN knockdown experiments; A.S. and S.Y.X.T. conducted RAP-seq experiments; F.D. analyzed FISH data quantitatively; and J.M. examined the 3D-structure of the X. I.D. conducted all computational analyses. I.D. and K.P. designed the experiments, interpreted the data, contributed towards methodology and model creation, and wrote the manuscript.

## Declaration of interests

Authors declare no competing interests. Correspondence and requests for materials should be addressed to kplath@mednet.ucla.edu.

## Figure Legends

### Figure 1: XIST expression in female naïve hPSCs correlates with XCD *in cis*

**A)** Representative RNA FISH images for XIST RNA (green) in female naïve hESC lines (UCLA1 and H9), a female naïve iPSC line, and in the male naïve hESC line WIN1, as indicated. DAPI staining (blue) marks nuclei. Scale bar = 10 microns.

**B)** Quantification of RNA FISH data as described in (A), showing the proportion of cells with different number of XIST clouds (none, monoallelic or biallelic).

**C)** Scatter plot comparing the X/A ratio, based on the sum of read counts from the X-linked and autosomal genes, respectively, and XIST expression, based on bulk RNA-seq data from different naïve conversions or subclones of the indicated naïve hPSC cell lines (5 replicates for naïve UCLA1, 2 replicates for H9, 2 replicates for the iPSC line and 2 replicates the pre-XIST UCLA1 state, Table S1). Pearson correlation (R) and p-value are given. 95% confidence intervals are indicated in gray.

**D)** Density plot of expression differences between the indicted XIST-positive female naïve hPSC line (H9, UCLA1, iPSCs) and an intermediate state arising during the conversion from primed-to-naïve hPSCs, in which XIST is not yet expression (UCLA1$^{pre\text{-}XIST}$), for X-linked genes (X, blue line) and autosomal genes (A, grey line). Dashed vertical line represents no change in gene expression. Wilcoxon test p-values for the difference between gene expression from the X and A are given (* P<0.05, ** P<0.01, *** P<0.001).

**E)** As in (D), except for gene expression changes between each indicated female naïve hPSC line and the male naïve hESC line WIN1.

**F)** Scatter plot of the X/A ratio (sum of read counts) vs XIST expression in individual cells of the female naïve hESC line H9, based on scRNA-seq data. Pearson correlation (R) and p-value are given.

**G)** Density plot showing the proportion of reads aligned to one of the two X chromosomes (X1) in individual cells of the female naïve hESC line H9, based on allelic analysis of scRNA-seq data. The dashed line serves as reference for cells in which both X chromosomes (X1 and X2) are active.

**H)** Scatter plot of the proportion of reads aligned to X1 (excluding XIST locus) vs reads aligned to XIST locus on the X1, for five replicates of female naïve UCLA1 hESCs, based on bulk RNA-seq. Pearson correlation (R) and p-value are given. 95% confidence intervals are indicated in gray.

**I)** Representative RNA FISH image of *XIST* RNA (green) and nascent transcription spots of the X-linked genes *THOC2* (yellow) and *UTX* (red) in female naïve UCLA1 hESCs. Scale bar = 10 microns, DAPI demarcates the nucleus. Quantification of the nascent transcription pattern for *THOC2* and *UTX* is shown on the right (bi = biallelic, mono = monoallelic, none = no signal detected). XIST is expressed monoallelically in nearly all cells as shown in (A).

**J)** Box plot of the raw signal intensity of the nascent transcription focus of *THOC2* and *UTX* on the X with the XIST cloud (XIST+) and away from the XIST cloud (XIST-), based on the data shown in (I). 82 individual cells were analyzed. Wilcoxon p-values are given (* P<0.05, ** P<0.01).

**K)** As in (I), except for the lncRNA XACT (red) and XIST (green) for female naïve UCLA1 and H9 hESCs. Quantification of the number of XACT clouds is shown on the right.

**L)** As in (J), except for the XACT signal close to and away from the XIST cloud for the experiment shown in (K). 15 individual cells were analyzed and the average of the top 10% of intensities of the XACT signal per X plotted. Wilcoxon test p-values: * $P<0.05$, *** $P<0.001$).

**Figure 2: XIST deletion in naïve hPSCs results in the loss of XCD**

**A)** Scheme of XIST deletion approach. XIST was deleted in primed H9 hESCs, resulting in the KO clones C7 and C18. Primed WT and XIST KO H9 cells were then converted to the naïve state with 5iLAF.

**B)** Representative RNA FISH images of XIST RNA (green) in female naïve WT H9 hESCs and the indicated XIST KO clones. The bottom row shows one representative nucleus, boxed in the image above. Scale bar = 20 microns.

**C)** Quantification of the percentage of cells from (B) with no, monoallelic, or biallelic XIST clouds.

**D)** XIST expression level in the naïve WT and XIST KO H9 based on bulk RNA-seq data. Error bars correspond to the standard deviation (SD) of replicates.

**E)** As in (D), except for the expression level of naïve pluripotency-related markers.

**F)** UMAP of scRNA-seq data from naïve WT and XIST KO (clone 18) H9 hESCs, colored by cell line (left) or cluster (right). Far right-top: Proportion of WT and KO cells in each cluster; bottom: proportion of cells in each cluster for the WT and KO cell lines.

**G)** X/A ratio (sum read counts) for naïve WT and KO H9 cells, based on bulk RNA-seq data. Error bars indicate the SD of replicates.

**H)** Boxplots of the X/A ratio (sum read counts) for individual naïve WT and KO H9 cells, based on scRNA-seq and separated by the clusters defined in (F). Wilcoxon p-values (* P<0.05, *** P<0.001) are indicated on top. The number of cells in each cluster is given below.

**I)** Density plot of the log2 fold change of gene expression between the indicated naïve XIST KO and WT H9 hESC lines, for X-linked (blue) and autosomal (grey) genes, based on bulk RNA-seq data. Wilcoxon p-value: *** P<0.001, for the difference in autosomal and X-linked gene expression.

**J)** Density plot of the log2 fold change in expression between naïve XIST KO (merged from both clones) and WT H9 hESCs, for genes that escape (red) or are subject (blue) to XCD. Wilcoxon p-value between both distributions: *** P<0.001.

**Figure 3: XIST localizes across the entire dampened X chromosome**

**A)** Percentage of DNA reads aligned to each chromosome for input genomic DNA (grey), and for the XIST pulldown, by RAP-seq, from two female XiXa fibroblast replicates (R1 and R2, green), female naïve H9 hESCs (light blue), female naïve UCLA1 hESCs (blue),

and female naïve iPSCs (dark blue) with an Xd and Xa. Error bars indicate the SD of all input replicates.

**B)** Cumulative distribution plot of XIST enrichment scores across all 100kb windows on the X chromosome, for the XIST-pulldowns described in (A). Solid vertical line marks no enrichment (no difference between pulldown and input); dashed line the two-fold enrichment; and the dotted line 20-fold enrichment of pulldown over input.

**C)** XIST enrichment along the X chromosome for the cell lines described in (A). The enrichment score was averaged across the two fibroblast XIST RAP-seq replicates. Unmappable regions are masked in white. The genomic *XIST* locus is marked.

**D)** Differential XIST localization (comparing the relative distribution of XIST on the X using diffBind as described in the methods) along the X chromosome between naïve hPSCs (the Xd) and fibroblasts (the Xi). Color represents log10(p-values) indicating the confidence of the differences in XIST binding in each region, and is positive in genomic regions where XIST localization is higher on the Xd in naïve hPSCs, and negative when XIST localization is higher on the Xi in fibroblasts. Only regions with significant differences (p-value<0.01) are presented. *XIST*, *DXZ4*, and *FIRRE* genomic loci are marked.

**E)** Scatter plots showing the linear regression line for the comparison of XIST enrichment scores for each genomic region (1Mb windows every 250kb) along the X chromosome in the indicated cell line with the density of genes, L2, MIR, SINE and ERVL-MaLR elements, respectively, within the same genomic intervals. Pearson correlation (R) and p-value are given (ns≥0.05, * P<0.05, ** P<0.01, *** P<0.001).

**F)** As in (C), for the genomic locus of *KDM6A* on the X chromosome.

**G)** Boxplot of XIST enrichment scores at genes subject to XCD or escaping XCD in female naïve hPSCs. Escapees and genes subject to XCD were defined by comparing X-linked gene expression for each female naïve hPSC line (H9, UCLA1, iPSCs) against the male line WIN1 (indicated on the top). The XIST enrichment data from the respective female line were then plotted (ie for the H9 vs WIN1 comparison XIST enrichment from the H9 line). Wilcoxon p-values: NS≥0.05, ** P<0.01. The number (n) of genes in each group is shown.

**H)** As in (G), except that XIST enrichment is plotted for genes subject to XCD or escaping XCD in female epiblast cells at embryonic day (E) 5, 6 and 7. Genes subject to or escaping XCD were defined by comparing X-linked gene expression levels between female and male epiblast cells from pre-implantation embryos at E5, E6, and E7, respectively (top). XIST enrichment is plotted based on the RAP-seq data from naïve iPSCs (top; dark blue), UCLA1 (middle; blue), and H9 (bottom; light blue). Wilcoxon p-values: NS≥0.05, * P<0.05. The number (n) of genes in each group is shown.

**I)** Box plots of XIST enrichment in fibroblasts for genes that escape or are subject to XCI. Wilcoxon p-value: *** P<0.001.

**Figure 4: XIST spreads to specific autosomal regions in female naïve hPSCs**

**A)** Representative RNA FISH images for XIST RNA (green) in female naïve UCLA1 hESCs and female human fibroblasts, with a zoom in of the indicated cell to display the more dispersed distribution of XIST in naïve hPSCs. Scale bar = 20 microns

**B)** Representative RNA/DNA FISH images (projections) for XIST (green) and X-chromosome paint (pink) in female naïve UCLA1 hESCs and fibroblasts, revealing the spread of XIST beyond the X-territory in naïve hPSCs. Scale bar = 10 microns.

**C)** Cumulative distribution plot showing XIST enrichment across all 100kb windows on autosomes for female naïve H9 hESCs (light blue), female naïve UCLA1 hESCs (blue), and female naïve iPSCs (dark blue) and two fibroblast replicates (light and dark green). The solid vertical line marks a log2 XIST enrichment of 0, and the dashed line a two-fold enrichment of XIST.

**D)** Number of significant XIST peaks on autosomes in the indicated cell lines.

**E)** Number of XIST peaks on each of the autosomal chromosome, colored as in (C).

**F)** Plot showing the location of XIST peaks that are conserved among naïve hPSCs along each autosome.

**G)** Boxplot of the XIST enrichment scores for autosomal peaks in the naïve hPSCs (averaged over the three hPSC lines), divided into peaks that were identified in only one, two or three hPSC line(s). Those peaks present in all three lines represent the conserved peaks of XIST in naïve hPSCs. Wilcoxon p-values: NS≥0.05, *** P<0.001. The number of peaks (n) in each group is given.

**H)** Percentage of the base pairs of each autosome covered by conserved XIST peaks in naïve hPSCs.

**Figure 5: Features of XIST-enriched autosomal regions**

**A)** Scatter plots showing the linear regression line between XIST enrichment scores along autosomes (1Mb windows every 250kb) and the density of L2 and MIR elements, respectively, within the same genomic interval, for female naïve H9 hESCs (light blue), female naïve UCLA1 hESCs (blue), and female naïve iPSC (dark blue). Pearson correlation (R) and p-value are given.

**B)** Percentage of autosomal genes in indicated categories associated with conserved autosomal XIST peaks or any autososomal XIST peaks detected in female naïve H9 hESCs (light blue), female naïve UCLA1 hESCs (blue), and female naïve iPSC (dark blue).

**C)** Most highly enriched gene ontologies (p-value<0.000005) and their enrichment values (-log10(p-value)) for genes associated with a conserved autosomal XIST peak in naïve hPSCs, colored by the odds ratio (log2(observed/expected)).

**D)** Examples of XIST enrichment over autosomal regions in the indicated cell lines.

**E)** Box plots of expression of genes overlapping (+) or not overlapping (-) conserved autosomal XIST peaks or autosomal XIST peaks identified in each of the cell lines indicated above. The transcript levels are given for each cell line indicated on the right. Wilcoxon p-value: *** P<0.001.

**F)** Enrichment (Fisher's exact odd ratios) of the binding sites of indicated regulators in conserved autosomal XIST peaks in hPSCs vs the rest of the autosomes, based on ENCODE data from diverse cell types and tissues.

**Figure 6: XIST mediates the downregulation of gene expression on autosomes in female hPSCs**

**A)** Density plots of changes in autosomal gene expression between the indicated female naïve hPSC line (H9, UCLA1 or iPSCs) and male naïve WIN1 hESCs, for genes that overlap (blue) or do not overlap (black) with autosomal XIST peaks in each of the respective female line. Wilcoxon test p-values indicate that genes under XIST peaks are more lowly expressed in female than male hPSCs.

**B)** Boxplots showing XIST enrichment across all autosomal genes that are significantly more lowly expressed (down) in the indicated female naïve hPSC line (H9, UCLA1 and iPSCs) compared to male naïve WIN1 hESCs, and those that do not differ (no change). The XIST enrichment scores are plotted from the corresponding female naïve hPSC RAP-seq data set in each comparison. Wilcoxon p-value: *** P<0.001. The number of genes (n) in each group is given. Dashed vertical line represents no change in gene expression.

**C)** XIST enrichment in indicated cell lines across the genomic region containing the autosomal gene *SPON1* (chr11). Below, RNA-seq read counts are shown for male (blue) and female naïve hPSCs (red, average of the three female lines and replicates). Zoom in into the gene locus is shown below.

**D)** Representative RNA FISH image for nascent transcription foci of *SPON1* (pink) in female (UCLA1) and male (WIN1) naïve hESCs. Quantification of the number of cells with no, monoallelic, or biallelic *SPON1* signals is shown on the right.

**E)** As in (C), except for the gene *HUNK* (chr21).

**F)** Scatter plot of the transcript level of *HUNK* or *SPON1* compared to that of XIST in each replicate of indicated naïve hPSC lines (as explained in Figure 1C), based on bulk RNA-seq data. Pearson correlation (R) and p-value are given for each gene. 95% confidence intervals are indicated in gray.

**G)** Boxplot of the Pearson correlation values between XIST expression and the expression of all genes that overlap (+) or do not overlap (-) conserved autosomal XIST peaks or all XIST peaks in the indicated cells lines (indicated on top). Wilcoxon p-value: *** P<0.001. The number of genes (n) in each group is given.

**H)** Scatter plot showing the linear regression line between the changes in transcript levels of autosomal genes between naïve XIST KO and WT H9 hESCs (based on scRNA-seq data of cluster 0 cells) and XIST enrichment around those gene (based on H9 RAP-seq data). Pearson correlation (R) and p-value are given. 95% confidence intervals are indicated in gray.

**I)** Box plots of the transcriptional changes between naïve XIST KO and WT H9 hESCs, based on scRNA-seq data (cluster 0 cells), for three XIST enrichment quantiles in naïve H9 hESCs, with low (1), middle (2), and high (3) XIST enrichment.

**J)** Box plots of normalized read counts for *HUNK* and *SPON1* transcripts in female (red) and male (blue) epiblast cells at different stages of pre-implantation development, based on scRNA-seq data.

**K)** Box plots of the mean normalized expression per cell for all genes under indicated autosomal XIST peaks (right) in female (red) and male (blue) epiblast cells at different

stages of pre-implantation development (given on top), based on scRNA-seq data. Wilcoxon p-values: NS≥0.05, * P<0.05, *** P<0.001.

**L)** Box plots of XIST enrichment scores from the indicated cell lines (right) around genes that are significantly more lowly expressed (Down) or have no changes in gene expression between female and male epiblast cells at indicated developmental stages (top). Wilcoxon p-values: NS≥0.05, ** P<0.01, *** P<0.001.

**Figure 7: SPEN is required for XIST-mediated gene regulation in female naïve hPSCs**

**A)** Immunostaining for CIZ1 (pink) along with RNA FISH for XIST (yellow) in female human fibroblasts and naïve UCLA1 hESCs, and quantification of the proportion of cells with CIZ1 accumulation under the XIST cloud.

**B)** Schematic representation of the SPEN depletion experiment.

**C)** X/A ratios (sum of read counts) in female naïve UCLA1 and male naïve WIN1 hESCs upon treatment with siCTRL (control) and si*SPEN*. Error bars = SD of replicates.

**D)** Density plot of the change in gene expression upon treatment with si*SPEN* in female naïve UCLA1 (left) and male naïve WIN1 (right) hESCs, respectively, for X-linked (blue) and autosomal (grey) genes. Wilcoxon p-values: ** P<0.01, *** P<0.001.

**E)** As in (D), except that box plots depict the transcriptional changes for genes that escape XCD or are subject to XCD (defined based on the comparison of transcript levels between

UCLA1 and WIN1). Wilcoxon p-value: *** P<0.001. The number of genes (n) in each group is given.

**F)** As in (D), except for autosomal genes that overlap with autosomal XIST peaks (+, blue) or not (-, black). The XIST peaks here belong to the corresponding UCLA1 line. The Wilcoxon test indicates that autosomal genes with XIST peaks are upregulated upon si*SPEN*. Wilcoxon p-value: *** P<0.001.

**G)** XIST enrichment around autosomal genes that are significantly upregulated (Up) or do not change (NC) upon *SPEN* depletion in female UCLA1 and male WIN1, respectively. RAP-seq enrichment scores are from the corresponding UCLA1. Wilcoxon p-value: *** P<0.001. The number of genes (n)in each group is given.

**H)** Hypergeometric test results for significant overlap between autosomal genes that are upregulated (Up) or do not change (NC) upon *SPEN* depletion in female UCLA1 or male WIN1 (right) or upon deletion of XIST (KO, top). Asterisks represent significant p-values (*** P<0.0005).

**I)** A model of XIST localization and function. A nuclear area containing the X chromosome and surrounding autosomes is shown for human female somatic cells (Xi, left), female naïve hPSCs (Xd, middle), and female naïve hPSCs without XIST (Xa, right). Zoom-ins show the corresponding gene expression state. Left: During XCI, XIST spreads *in cis* over the entire X and interacts with SPEN to mediate complete silencing of most genes on the Xi. Autosomal genes are not targeted and regulated by XIST. Middle: During XCD, XIST spreads *in cis*, across the entire X chromosome, and *in trans*, additionally targeting specific autosomal regions. Together with SPEN, XIST reduces the transcript levels of

both X-linked genes and autosomal targets, but does not induce complete silencing as seen on the Xi. The lower level of XIST in the Xd compared to the Xi (indicated by the lighter shading across the X-compartment and the lower number of XIST RNA molecules) may be mechanistically linked to the differential gene expression output on the Xd versus the Xi. Right: In the absence of XIST in naïve female hPSCs, X-linked genes and autosomal genes normally targeted by XIST are expressed at the level seen in male cells, identifying XIST as master regulator of XCD and defining a new function of XIST in the control of autosomal gene expression.

## Supplemental Figure Legends

### Figure S1: Characterization of the relationship between XIST and XCD in naïve hPSCs [Related to Figure 1]

**A)** XIST expression in indicated female naïve hPSC lines (UCLA1$^{pre\ XIST}$, H9, UCLA1 and iPSC), and male naïve WIN1 hESCs, based on bulk RNA-seq data. For this, 5 replicates for naïve UCLA1, 2 replicates for H9, iPSCs, pre-XIST UCLA1, and WIN1, respectively, were exploited (Table S1). Median expression of all replicates is shown, with error bars corresponding to the standard deviation (SD) between replicates.

**B)** X/A ratios for the cell lines described in (A), based on the median expression (RPKM) of X-linked and autosomal genes. Median X/A ratios of all replicates is shown, with error bars corresponding to the standard deviation (SD) between replicates.

**C)** As in (B), except for the formation of the X/A ratios based on mean expression values.

**D)** Scatter plot of the X/A ratio (based on the mean expression of X and autosomal genes) vs XIST expression in individual cells of the female naïve hESC line H9, based on scRNA-seq data. Pearson correlation (R) and p-value are given. Linear regression line, with 95% confidence intervals, is shown in gray.

**E)** Density plot showing the proportion of reads aligned to one of the two X chromosomes (X1) in individual cells of the female naïve hESC line H9, selecting only cells with detectable XIST expression, based on allelic analysis of scRNA-seq data. The dashed line serves as reference for cells in which both X chromosomes (X1 and X2) produce equal levels of transcripts.

**F)** Boxplots showing the transcript levels of the X-linked genes *GPC3*, *SMARCA1* and *SMS* in individual female and male epiblast cells of E7 pre-implantation embryos, based on scRNA-seq data.

**G)** Representative RNA FISH image of XIST (green) and the nascent transcription foci of *GPC3* (pink) in female naïve UCLA hESCs and the quantification of the nascent transcription patterns of *GPC3*. Note the smaller *GPC3* transcription spot on the XIST-coated X chromosome. Scale bar = 10 microns.

**H)** Representative RNA FISH images for the nascent transcription foci of the X-linked genes *SMARCA1* or *SMS* in male WIN1 and female UCLA1 naïve hESCs. Quantification of the RNA FISH patterns are given below. Scale bar = 10 microns.

**I)** Boxplots of the percentage of reads aligned to X1 and X2 in five different replicates of naïve and two of primed UCLA1 hESCs, based on the allelic analysis of bulk RNA-seq data. Red dots represent reads with SNPs overlapping the XIST transcript. These data show that X1 is the active X chromosome and X2 the inactive X chromosome in primed UCLA1, which display non-random XCI.

**J)** Top: Schematic of how the XACT signal was quantified. Bottom: Average of the top 10% of XACT cloud signal intensity values per cell, measured either close to XIST cloud (green) or away from it (grey). Error bars correspond to the SD.

**Figure S2: Characterization of XIST KO in naïve hESCs [Related to Figure 2]**

**A)** IGV plot showing RNA-seq reads in two replicates of WT female naïve UCLA1 hESCs at the *XIST* locus. In the gene model, exons are indicated with thicker lines. These data show that all expected exons of XIST are transcribed in naïve hESCs.

**B)** IGV plot of RNA-seq data across the *XIST* locus, from the genome-wide 5' (top four tracks) and 3' (bottom four tracks) end mapping of transcripts in female naïve UCLA hESCs and the female human somatic cell line K562. A zoom-in into the genomic region containing the *XIST* transcriptional start site is shown below. These data reveal the 5' and 3' ends for the XIST transcript in naïve hPSCs and somatic cells.

**C)** Schematic representation of the XIST deletion strategy. CRISPR/Cas9 was used to homozygously excise a ~2kb region from the promoter region and 5' end of the XIST transcript. The location of the sequences used for PCR genotyping and the length of the expected fragments are also given.

**D)** PCR genotyping of the XIST KO. Upon deletion, XP-Frw and XE-Rev primers shown in (C) can only amplify 1.5 kb region from genomic DNA, while the extension time is too short to for the WT amplicon. The XP-Frw and WT-Rev primer set amplifies 942 bp region in WT genomic DNA. As the WT-Rev complementary region is not present upon XIST deletion, no PCR band is obtained for in hESCs carrying the homozygous XIST deletions. The PCR genotyping data are shown for primed WT and XIST KO (clones 7 and 18) H9 hESCs.

**E)** Box plots of XIST transcript levels in each individual cell of naïve WT and KO (clone 18) H9 hESCs, for each of the clusters shown in Figure 2F. Wilcoxon p-values: ** P<0.01, *** P<0.001.

**F)** As in (E), expression level of naïve pluripotency-related marker genes is shown. Wilcoxon p-values: * P<0.05, ** P<0.01.

**G)** Density plot showing the proportion of reads aligned to one of the two X chromosomes (X1) in individual cells of the naïve WT and XIST KO (clone 18) H9 hESCs, based on allelic analysis of scRNA-seq data. The dashed line serves as reference for cells in which both X chromosomes (X1 and X2) produce equal levels of transcripts.

**H)** X/A ratios based on the median expression (RPKM) of X-linked and autosomal genes in naïve WT and XIST KO (clones 7 and 18) H9 hESCs, based on bulk RNA-seq data. Median X/A ratios of all replicates is shown, with error bars corresponding to the standard deviation (SD) between replicates

**I)** As in (A), except for the X/A ratio being derived from mean expression data.

**J)** Density plot of the change in expression between naïve XIST KO (clone 18) and WT H9 hESCs (comparing the average expression of a gene between KO and WT cells of cluster 0 described in Figure 2F), for X-linked (blue) and autosomal (grey) genes, based on scRNA-seq data. Wilcoxon p-value: *** P<0.001.

**K)** Density plot of the change in expression between naïve XIST KO (clone7 or 18) and WT H9 hESCs, for genes that escape (red) or are subject to (blue) XCD in WT H9 hESCs, based on bulk RNA-seq data. Wilcoxon p-value: *** P<0.001.

**L)** Same as in (K), except for scRNA-seq data, using the average gene expression of all single cells in cluster 0.

**Figure S3: Defining genes that escape XCD [Related to Figures 2, 3 and 7]**

**A)** Proportion of genes that escape (red) or are subject (grey) to XCD, defined by comparing the transcription state of the indicated female naïve hPSC line (H9, UCLA1, or iPSCs) to male naïve WIN1 hESCs or female to male epiblast cells from human pre-implantation embryos at E7.

**B)** Location of XCD escapee genes, as defined in (A), along the X chromosome. *XIST* locus is marked.

**C)** Graph showing the overlap between XCD escapees defined in different 5iLAF-cultured naïve female hPSC lines as described in (A).

**D)** Hypergeometric test for significant overlap between genes that escape or are subject to XCD in female naïve UCLA1 hESCs (defined against WIN1) and genes that are more highly expressed (Up), not changed (NC) or more lowly expressed in naïve UCLA1 compared to naïve UCLA1 in the pre-XIST state. Asterisks = *** $P<0.0005$.

**E)** Representative RNA FISH images of XIST RNA (green) for female naïve HNES3 hESCs. Quantification of the number of cells in which the XIST cloud was detected in female (HNES3) and male (HNES1) naïve hESC lines. These HNES lines were derived directly from the blastocyst in the t2iLGö culture medium. Scale bar = 20 microns.

**F)** Similar to (D), except that the overlap between genes defined as escape or subject to XCD in indicated female to male comparisons.

**G)** Boxplots of the proportion of reads aligned to X1 for genes escaping or subject to XCD (defined based on the comparison of female naïve H9 hESCs vs male naïve WIN1 hESCs), based on scRNA-seq data from naïve WT H9 hESCs.

**H)** As in (C), additionally adding XCD escapees defined in female epiblast cells from human E7 pre-implantation embryos.

**I)** Boxplots of the expression level of genes escaping or subject to XCD in the naïve hPSC lines indicated on the right, based on bulk RNA-seq data. Escapees and genes subject to XCD were defined based on the comparisons given on the top. Wilcoxon p-values: * P<0.05, ** P<0.01, *** P<0.001.

**J)** As in (I), except that genes escaping and subject to XCD were defined by comparing female and male epiblast cells from E7 embryos and that the expression level is plotted for all cells of female E3 and E4 embryos and specifically the epiblast cells of E5-E7 embryos, based on scRNA-seq data.

**K)** As in (D), except for the comparison of genes that escape XCD (as defined by the indicated comparisons on the top) and XCI (left).

**L)** Boxplots of the expression of the XCD escape *KDM6A* in female and male epiblast cells of E5, E6, and E7 embryos. *KDM6A* was defined as escapee in UCLA1 and E7 epiblast cells.

**Figure S4: Comparison of XIST localization on the Xd and Xi [Related to Figure 3]**

**A)** Boxplot of XIST enrichment scores for all 100Kb windows (every 25Kb) along the X chromosome in the indicated cell lines. Enrichment at in the genomic region containing the *XIST* locus is marked with red circles.

**B)** XIST enrichment along the X chromosome for two fibroblast replicates. Unmappable regions are masked in white. The genomic *XIST* locus is marked.

**C)** Heatmap showing Pearson correlation between XIST enrichment within 100Kb windows (every 25Kb) along the X chromosome in indicated cell lines.

**D)** Differential enrichment of XIST in each 100Kb window (every 25Kb) of the X chromosome between the indicated female naïve hPSC line and the two fibroblast replicates is shown along the entire X chromosome. For this comparison, the XIST localization across the X was normalized to only reads that aligned to the X chromosome, to highlight the differential distribution of XIST along the Xd and Xi. Red color represents genomic regions in which XIST was significantly more enriched on the Xd and blue genomic regions in which XIST was significantly more enriched on the Xi, based on log10(p-values). Only regions with significant differences (p-value<0.01) are presented. The genomic loci of *XIST*, *DXZ4*, and *FIRRE* are marked.

**E)** Top: Same as (D), but comparing all naïve replicate to all fibroblast replicates using diffBind as described in the methods. The genomic *DXZ4* locus is marked. Middle: Normalized Hi-C contacts on the Xi (top half) and Xa (bottom half) in the human female somatic cell line GM12878. Bottom: Normalized Hi-C contact differences between the Xi and Xa.

**F)** Boxplots of the average XIST enrichment scores from the three female naïve hPSCs (H9, UCLA1 and iPSC) in X-linked genes subject to or escaping XCD, defined by comparing female (HNES3) vs male (HNES1) naïve hESC bulk RNA-seq data. Wilcoxon p-value: ** P<0.01.

**G)** As in (D), except that the data were normalized to read aligned to the autosomes as well as the X chromosome. The global normalization of the data shows that XIST more strongly enriched in the Xi compared to the Xd, across nearly the entire chromosome.

**Figure S5: Characterization of the XIST enrichment on autosomes [Related to Figure 4]**

**A)** Percentage of reads from the XIST-pulldown (based on RAP-seq) aligning to autosomes for the indicated cell lines.

**B)** Boxplots showing expression of all genes in indicated female naïve hPSC lines and diverse female somatic tissues (extracted from the human Epigenome Roadmap). The red dots represent the expression of XIST in each sample.

**C)** Number of significant peaks on each autosome in indicated cell lines, normalized by chromosome size.

**D)** Display of all autosomal XIST peaks in (from outside to inside) naïve iPSCs H9 (dark blue), UCLA1 (blue), H9 (light blue), fibroblast R2 (dark green) and R1 (light green).

**E)** Diagram showing the intersection between autosomal XIST peaks detected in indicated cell lines.

**F)** Boxplots of autosomal XIST peak scores in each female naïve hPSC line, for peaks that were identified in only one, two or three hPSC lines. Wilcoxon p-values: NS≥0.05, * P<0.05, ** P<0.01, *** P<0.001. The number of peaks (n) in each group is given below.

**G)** Density plot showing the length (in base pairs) of conserved autosomal XIST peaks in female naïve hPSCs.

**H)** Number of XIST peaks in each of the autosomes, for peaks identified in only one of the naïve hPSC lines (top), two (middle), or in all three (bottom).

**Figure S6: Characterization of the role of SPEN in XCD [Related to Figure 7]**

**A)** Schematic representation of the sequential RNA+DNA FISH approach on the female naïve hESC line H9 and female human fibroblasts. RNA FISH targeting XIST was applied first to identify the Xd and Xi, respectively, in each cell. Then, three sequential rounds of DNA FISH were performed, each targeting three different genomic loci along the X chromosome.

**B)** Analysis of the FISH data described in (A). Boxplots of the pairwise distances between the DNA probes in each hybridization round were measured in 20 and 16 naïve and somatic cells respectively, and are given for the Xd and Xa in naïve hESCs or the Xi and Xa in fibroblasts (somatic). Asterisks represent significant p-values (NS≥0.05, ** P<0.005, *** P<0.0005). These data show that, as expected, the Xi is dramatically different from the Xa in somatic cells. Moreover, the distances of round 3 probes (UTX, XACT and GPC3) are significantly smaller on the Xd than the Xa in hPSCs.

**C)** Relative transcript levels of *SPEN* in female naïve UCLA1 and male naïve WIN1 hESCs upon treatment with siCtrl or si*SPEN*, based on qPCR. Error bars correspond to the SD in replicates.

**D)** X/A ratios based on median gene expression values in bulk RNA-seq data for the samples described in (C). Error bars correspond to the SD in replicates.

**E)** As in (D), except that the X/A ratio was formed based on mean expression values.

**F)** Density plot of the change in expression upon si*SPEN* in female compared to male naïve hESCs, for X-linked (purple) and autosomal (grey) genes. Dashed line represents similar effect of si*SPEN* to gene expression in female and male hESCs. Wilcoxon p-value: *** P<0.001. This shows that the upregulation of X-linked genes upon treatment siSPEN is higher in females.

**G)** XIST expression in female naïve UCLA1 and male naïve WIN1 hESCs upon treatment with siCtrl or siSPEN, based on bulk RNA-seq data.

**H)** As in (G), except for naïve pluripotency-related marker genes.

## Summary of Tables

**Table S1. Summary of the genomics data used in this study [Related to all Figures].**

Summary of the cell lines used in this study and of the corresponding bulk RNA-seq, scRNA-seq and RAP-seq data.

**Table S2. Expression status of genes in naïve hPSCs [Related to all Figures].**

Changes in transcript levels are given for all genes between naïve female HNES3 vs male HNES1 hESCs, naïve female H9 vs male WIN1 hESCs, naïve female UCLA1 vs male WIN1 hESCs, naïve female iPSCs vs male WIN1 hESCs, and naïve female UCLA1 vs pre-XIST UCLA1 hESCs. X-linked genes were classified as XCD escape/subject in all female vs male comparisons. Autosomal genes were classified as upregulated in female (Up), have not significant differences (NS) or downregulated in female (Down) in all female to male comparisons. For naïve female UCLA1 vs pre-XIST UCLA1 hESCs comparison, X-linked and autosomal genes were classified as Up/NS/Down corresponding to genes upregulated, have no significant differences or downregulated in naïve female UCLA1 expressing XIST respectively. In addition, the XCD status in female E7 epiblast cells and the XCI status in female human somatic cells are given.

**Table S3. Sequences of XIST RAP-seq probes, gRNAs, and primers used in this study [Related to Figures 2 and 3].**

The sequences of the probes used to capture XIST in RAP-seq experiments, of primers for PCR validation of the XIST KO, and of gRNAs used for the XIST KO, are given.

**Table S4. Genomic features and chromatin states correlated with XIST enrichment on the X chromosome and autosomes [Related to Figures 3 and 5]**

This table give the Pearson correlation values between the XIST enrichment along the X chromosome and autosomes in naïve H9 and UCLA1 hESCs and naïve hiPSCs, and different genomic features (using 1Mb windows every 250kb). For the X chromosome, these data are also given for the XIST enrichment on the Xi in female human fibroblasts. In addition, the table contains the features enriched in conserved autosomal XIST peaks (using LOLA) for naïve-conserved peaks, or peaks detected in naïve H9 and UCLA1 hESCs and naïve hiPSCs.

**Table S5. Genomic coordinates of autosomal XIST peaks detected in female naïve hPSCs and female fibroblasts [Related to Figure 4].**

Coordinates of autosomal XIST peaks detected in naïve hPSCs (H9, UCLA1, and iPSCs) or in each of the two fibroblast replicates. For each cell line/replicate, 1 indicates that a given peak was detected in the sample, 0 that the peak was not detected. Specificity_tag indicates if the peak was classified as conserved autosomal peak in naïve hPSCs or somatic conserved (detected in both fibroblast replicated and not in hPSCs).

**Table S6. Genes overlapping with XIST naïve-conserved autosomal peaks [Related to Figures 5 and 6].**

For autosomal genes, this table show if the gene overlapped with conserved autosomal XIST peaks. In addition, this table indicates if the gene is expressed/not expressed in naïve hPSCs, and, for those that are expressed, identified them as more lowly expressed in female compared to male hPSCs.

## STAR+METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited Data | | |
| Raw and analyzed data | This paper | GSE163439 |
| Antibodies | | |
| Rabbit Anti-CIZ1 | Novus Biologicals | NB100-74624 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Recombinant Human FGF basic Protein | PeproTech | 100-18B |
| Recombinant Activin A | Cell Guidance Systems | GFH6 |
| Recombinant human LIF | Cell Guidance Systems | GFH200 |
| PD0325901 | Cell Guidance Systems | SM26 |
| SB431542 | Cell Guidance Systems | SM33 |
| WH 4 023 | Cell Guidance Systems | SM16 |
| Y27632 | Cell Guidance Systems | SM02 |
| IM12 | Cell Guidance Systems | SM04 |
| Critical Commercial Assays | | |
| CytoTune-iPS 2.0 Sendai Reprogramming Kit | ThermoFisher | A16517 |
| Human X chromosome Paints | MetaSystems Probes | D-0323-100-OR |

| | | |
|---|---|---|
| TruSeq Stranded mRNA Library Prep | Illumina | 20020594 |
| RNAeasy | Qiagen | 74104 |
| KAPA HiFi hot start Taq | Kapa biosystems | Kk2502 |
| Quick-DNA Miniprep Plus Kit | Zymo Research | D4069 |
| Experimental Models: Cell Lines | | |
| Normal Human Dermal Fibroblasts | Lonza | cc-2511 |
| UCLA1 | Diaz Perez et. al. 2012 | N/A |
| H9 | WiCell Research Institute | N/A |
| H9_XIST_KO_c7 | This paper | N/A |
| H9_XIST_KO_c18 | This paper | N/A |
| iPSCs | This paper | N/A |
| WIN1 | Whitehead Institute | N/A |
| HNES1 | University of Cambridge | N/A |
| HNES3 | University of Cambridge | N/A |
| Oligonucleotides | | |
| Biotin labeled RNA capture probes for human XIST | Eurofins | Separate file |
| Primers | IDT | Separate file |
| gRNAs | IDT | Separate file |
| Recombinant DNA | | |
| PX459 plasmid | Addgene | 62988 |
| Software and Algorithms | | |
| R v3.6.3 | https://cran.r-project.org/ | https://cran.r-project.org/ |

| | | |
|---|---|---|
| Trim Galore v0.4.1 | Babraham Institute, Bioinformatics | http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/<br>https://github.com/FelixKrueger/TrimGalore |
| HISAT2 v2.2.0 | (Kim et al., 2019) | https://daehwankimlab.github.io/hisat2/ |
| Samtools v1.9 | (Li et al., 2009) | http://samtools.sourceforge.net/ |
| HTSeq-count v0.6.1p1 | (Anders et al., 2015) | https://htseq.readthedocs.io/en/release_0.11.1/count.html |
| DESeq2 v1.24.0 | (Love et al., 2014) | http://bioconductor.org/packages/release/bioc/html/DESeq2.html |
| CellRanger v3.0.2 | 10x Genomics | https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome |
| Seurat v3.1.5 | (Butler et al., 2018; Stuart et al., 2019) | https://satijalab.org/seurat/ |
| Scater v1.12.2 | (McCarthy et al., 2017) | https://bioconductor.org/packages/scater/ |
| CrossMap v0.2.9 | (Zhao et al., 2014) | http://crossmap.sourceforge.net/ |
| Bowtie2 v2.2.9 | (Langmead and Salzberg, 2012, p. 2) | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |
| Picard v2.1.0 | ("Picard Tools - By Broad Institute," n.d.) | http://broadinstitute.github.io/picard/<br>https://github.com/broadinstitute/picard |
| Bedtools v2.26.0 | (Quinlan and Hall, 2010) | http://bedtools.readthedocs.io/en/latest/ |
| MACS v2.2.7.1 | (Zhang et al., 2008) | https://github.com/macs3-project/MACS |
| Igvtools v2.3.80 | (Robinson et al., 2011) | http://software.broadinstitute.org/software/igv/ |
| DiffBind v2.12.0 | (Ross-Innes et al., 2012) | https://bioconductor.org/packages/release/bioc/html/DiffBind.html |
| LOLA v1.14.0 | (Sheffield and Bock, 2016) | http://code.databio.org/LOLA |
| topGO v2.36.0 | (Alexa et al., 2006) | https://bioconductor.org/packages/release/bioc/html/topGO.html |
| Fiji v2.0.0 | (Schindelin et al., 2012) | https://imagej.net/Fiji |
| Juicer_tools v1.5.3 | (Durand et al., 2016) | https://github.com/aidenlab/Juicebox |
| ggplot2 v3.3.0 | https://ggplot2.tidyverse.org/ | https://ggplot2.tidyverse.org/ |
| Other | | |

| | | |
|---|---|---|
| RNA-seq for UCLA1 | (Sahakyan et al., 2017a) | GEO: GSE87239 |
| UCLA1 Single-nucleotide polymorphisms | (Sahakyan et al., 2017a) | (Sahakyan et al., 2017a) |
| H9 Single-nucleotide polymorphisms | (Collier et al., 2017) | (Collier et al., 2017) |
| Single-cell RNA-seq in human preimplantation embryos | (Petropoulos et al., 2016) | (Petropoulos et al., 2016)<br>E-MTAB-3929 |
| XCI status | (Tukiainen et al., 2017) | (Tukiainen et al., 2017) |
| Human siSPEN ID s22831 | Thermo Fisher | 4427037 |
| Human siSPEN ID s22829 | Thermo Fisher | 4427037 |
| Human XIST BAC | Bacpac resource center | RP11-13M9 |
| Human XACT BAC | Bacpac resource center | RP11-35D3 |
| Human GPC3 BAC | Bacpac resource center | RP11-678F20 |
| Human THOC2 BAC | Bacpac resource center | RP11-121P4 |
| Human UTX BAC | Bacpac resource center | RP11-256P2 |
| Human SMS BAC | Bacpac resource center | RP11-147O5 |
| Human SMARCA1 BAC | Bacpac resource center | RP11-137A15 |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kathrin Plath (kplath@mednet.ucla.edu)

All genomics data (RAP-seq and RNA-seq) have been uploaded to a private GEO site that can be accessed by the reviewers with GSE163439 and the secure token: gzqtksgirzwhnip.

**Cell lines and culturing conditions**

Human ESC/hiPSC lines used in this study include UCLA1 (46, XX), iPSC (46, XX), H9 (46, XX), WIN1 (46, XY), HNES1 (46, XY) and HNES3 (46, XX). Primed hESCs/hiPSCs were cultured on inactivated mouse embryonic fibroblasts (MEFs) in hESC media, composed of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 0.1mM 2-Mercaptoethanol (GIBCO, 21985-023), 10ng/mL recombinant human FGF basic (R&D systems, 233-FB), 1x Penicillin-Streptomycin (GIBCO, 15140-122) in DMEM/F12 media (Sigma, D8437). Primed hESCs and iPSCs were split every 5-6 days using Collagenase type IV (GIBCO, 17104-019). All human ESC and iPSC studies received the approval of the UCLA Embryonic Stem Cell Research Oversight (ESCRO) Committee (2008-015 and 2007-009).

HNES1 and HNES3 hESC lines were kindly provided by Austin Smith and were derived from human blastocysts in t2iLGö culture medium. Those cell lines were maintained on an irradiated MEF cell layer in N2B27 supplemented with 1 μM PD0325901, 10 ng/ml human LIF, 2 μM Gö6983 (Tocris Bio-Techne, 2285), and 2 μM

XAV939 (Tocris Bio-Techne, 3748). ROCK inhibitor (10μM; Y-27632, Millipore) was added for 24 hours after passaging with StemPro Accutase.

Somatic cell lines used in this study included normal human dermal fibroblasts – NHDFs (46,XX) (Lonza lot #472033). NHDFs were cultured in a fibroblast media composed of 10% FBS (Life Technologies, 10099141), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 0.1mM 2-Mercaptoethanol (GIBCO, 21985-023) in DMEM (Sigma, D6429).

The female naïve hiPSC line was generated by reprogramming NHDF fibroblasts (Lonza, lot #472033) to pluripotency using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (ThermoFisher) following manufacturer's instructions for feeder-dependent reprogramming of fibroblasts. Briefly, 100,000 cells were plated into a well of a 6-well plate and 24hours later subjected to an overnight transduction with the CytoTuneTM vectors. Cell culture medium was renewed daily with fibroblast medium for one week. Seven days post transduction cells were harvested by TrypLE Express Enzyme (Thermofisher) and plated on a monolayer of irradiated MEFs in fibroblast medium. After one day, the medium was changed to naïve (5iLAF) human embryonic stem cell medium (see below). Mycoplasma tests (Lonza, LT07-418) were performed routinely for all cell lines used in this study to confirm the lack of such pathogens.

**Conversion of primed hPSC to the naïve pluripotent state**

A detailed description of the conversion of the primed hESC line UCLA1 to its naïve state has been given in Sahakyan et al. (Sahakyan et al., 2017a). Briefly, the primed hESC line UCLA1 was obtained from the Human Embryonic and Induced Pluripotent Stem Cell Core of the BSCRC at UCLA and plated on irradiated MEFs as single cells in the presence of 10uM ROCK inhibitor Y-27632 in DMEM/F12 medium supplemented with 5% KnockOut Serum Replacement (KSR), 15% heat inactivated FBS, penicillin/streptomycin, nonessential amino acids, GlutaMAX, 0.1mM β-mercaptoethanol, and 10ng/ml FGF2. After two days, the culture medium was switched to the 5iLAF naïve hESC medium (1:1 mixture of DMEM/F12 and Neurobasal supplemented with N2, B27, penicillin/streptomycin, nonessential amino acids, GlutaMAX, 0.5% KSR, 0.1mM β-mercaptoethanol, 50ug/ml bovine serum albumin, 20ng/ml rhLIF, 20ng/ml Activin A, 8ng/ml FGF2, 1uM MEK inhibitor PD0325901 0.5uM B-Raf inhibitor SB590885, 1uM GSK3-β inhibitor IM-12, 1uM Src inhibitor WH-4-023, and 10uM ROCK inhibitor Y-27632). Established naïve UCLA1 hESCs were maintained by passaging every 5-6 days with StemPro Accutase and re-plating onto freshly plated irradiated MEFs. UCLA1 and H9 hESCs were converted and cultured either in 5% CO2, 5% O2 (Hypoxia) at 37C or in 5% CO2 and atmospheric oxygen level (Normoxia) at 37C.

**Bulk RNA-seq**

For bulk RNA-seq, cells were washed with 1xPBS and dissociated with accutase. Harvested cells were then lysed using Trizol reagent (Life Technologies #15596018) and RNA was isolated using Qiagen RNAeasy kit (Cat #74104) according to manufacturer's

instructions. RNA-seq libraries were prepared with the TrueSeq Stranded mRNA Library Prep Kit (Illumina 20020594) according to manufacturer's instructions. RNA sequencing of the naïve UCLA1 R1, R2, and R5, or early naïve UCLA1 hESCs (pre-XIST state) was done in Sahakyan et al,. (Sahakyan et al., 2017a). Briefly, cells were harvested, washed with 1xPBS, and collected in Trizol (ThermoFisher). RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina) as described in (Sahakyan et al., 2017a).

**Processing of bulk RNA-seq data**

RNA-seq reads were trimmed using trim_galore (https://github.com/FelixKrueger/TrimGalore) with default parameters to remove the standard Illumina adaptor sequences. Reads were then mapped to the human genome (hg38 assembly) using HISAT2 (Kim et al., 2019) with default parameters. Reads with mapping quality less than 30 were removed using samtools (Li et al., 2009). Read counts for each gene were calculated using HTSeq using the following parameters: --format=bam --order=pos --stranded=reverse --minaqual=0 --type=exon --mode=union --idattr=gene_name (Anders et al., 2015). Genes with low read count were removed (keeping genes with counts-per-million >/= 0.5 in at least two samples). Regularized log transformation (rlog) of each gene in each sample was calculated using DESeq2 (Love et al., 2014). Rlog values were used as normalized gene expression data unless noted otherwise. In addition, RPKM (Reads Per Kilobase per Million mapped reads) values were calculated for each gene. Differential gene expression analysis was done using DESeq2

(Love et al., 2014). For the analysis of gene expression based on data derived by ENCODE, RNA-seq data (RPKMs) were downloaded from the NIH Roadmap Epigenomics Mapping Consortium (including both coding and non coding RNAs) (Davis et al., 2018). Genes with RPKM=0 across all cell types were removed. Cell types were filtered to select only female samples resulting in Blood, Breast, Lung, Brain, and Ovary.

**scRNA-seq**

For scRNA-seq, female naïve hESCs (H9) (WT or XIST KO) were dissociated with accutase for 5 minutes. Dissociated cells were resuspended in 1xPBS+0.04% BSA and passed through a 40micron strainer to avoid cell clumps. Cell concentration was adjusted to 800-1200cells/ul before loading cells onto the 10X Genomics Chromium instrument. scRNA-seq libraries were generated using the Chromium single cell 3' reagent kit V3 following manufacturer's instructions and library fragment size distribution was determined by BioAnalyzer. Individual libraries were designed to target 10,000 cells. Afterwards libraries were pooled and sequenced on the Illumina Novaseq 6000 platform.

**Processing of single cell expression data**

Reads were aligned using the CellRanger (https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome) count function against the hg38 (GRCh38-3) genome assembly. To account for mouse feeder cells, we also aligned the reads against the mouse mm10 genome assembly. Aligned reads were processed further through the Seurat pipeline (Butler et al., 2018; Stuart et al., 2019) and filtered based on

multiple criteria. 1) An initial filtering for genes detected in at least three cells and cells with at least 200 detected genes. 2) To remove mouse feeder cells, the number of genes detected when aligning to the human and mouse genome was calculated (nFeatures) and cells with log2(nFeatures human/nFeatures mouse)>1 were kept for downstream analysis. 3) Mitochondrial RNA was quantified per cell. A quality control metric was calculated using R function calculateQCMetrics (scater package) (McCarthy et al., 2017)). The R function isOutlier (scater package) (McCarthy et al., 2017)) was applied to identify outliers based on the library size (nmads>3), number of genes (nmads>3), and the percentage of mitochondrial genes in each cell (nmads>1). The data were then normalized and scaled, and highly variable genes were detected, using the sctransform function (Seurat package) (Butler et al., 2018; Stuart et al., 2019) with default parameters. Principal Component Analysis was done using the Seurat function RunPCA on the most highly variable genes. The top 20 principal components were used to find the 20 nearest neighbors of each cell (FindNeighbors function from Seurat package (Butler et al., 2018; Stuart et al., 2019)), following by application of the FindClusters function to identify cell clusters. Clusters were visualized by uniform manifold approximation and projection (UMAP) dimensional reduction technique using the Seurat RunUMAP function.

To explore the expression changes of genes between naïve WT H9 and XIST KO H9 hESCs, the average gene expression for each scRNA-seq cluster was obtained using AverageExpression function in Seurat (Butler et al., 2018; Stuart et al., 2019), and the log fold change between the WT and XIST KO was calculated for each gene.

Single cell expression data for human male and female pre-implantation embryos was obtained from Petropoulos et al., (Petropoulos et al., 2016). The read counts matrix was normalized and scaled using sctransform function (Seurat package (Butler et al., 2018; Stuart et al., 2019)) using default parameters. This was used to explore gene expression in each cell. In addition, gene expression fold change between female and male at each developmental stage and tissue was directly obtained from Petropoulos et al., supplemental table 1 (Petropoulos et al., 2016).

**Definition of genes escaping XCD**

The result of the differential gene expression analysis between male and female hPSCs or epiblast cells from pre-implantation embryos was used to define genes that escape or are subject to XCD. An X-linked gene was defined as XCD escapee in hPSCs if it was significantly (log2FoldChange>0.5 and padj<0.01) upregulated in the comparison of female naïve hPSCs (H9, UCLA1, and iPSC, respectively) vs male naïve hESCs (WIN1), or in the female naïve hESC line HNES3 vs the male naïve hESC line HNES1. For defining escapees in epiblast cells of the pre-implantation embryo, the log2 fold change was calculated for female to male E7 epiblast cells. The definition of epiblast cells and male/female status was obtained from Petropoulos et al. (Petropoulos et al., 2016). An X-linked gene was defined as XCD escapee when the gene displayed log2 fold change > 1 in female cells compared to males. Published assignments (Tukiainen et al., 2017)

were used for genes escaping XCI or subject to XCI. XIST was excluded from downstream analysis of XCD subject and escapee characterization.

**Variant calling**

For UCLA1 hESCs, a list of known X-linked Single-Nucleotide Polymorphisms (SNPs) was obtained from Sahakyan et al., (Sahakyan et al., 2017a). For H9 hESCs, a list of known X-linked SNPs in H9 was obtained from Collier et al., (Collier et al., 2017). The genotype coordinates (in hg19) in the original VCFs were transferred to the hg38 genome built through the liftOver function in CrossMap (Zhao et al., 2014).

**Haplotype phasing**

Since the SNP data in both UCLA1 and H9 hESCs were taken form an unphased genotypes, a read overlapping a known SNP could not be directly assigned to one of the two parental X chromosomes (haplotypes). To directly explore gene expression from each X chromosome, phasing haplotypes analysis, which identify the variant alleles that are co-located on the same chromosome, was done. Through this process, SNPs were assigned to either the X1 or X2. Both primed UCLA1 and H9 hESCs display non-random XCI (Collier et al., 2017; Sahakyan et al., 2017a). Consequently, we assigned the X1 to be the Xa of the primed state and X2 the Xi. This allowed us to explore expression of all SNPs between the two X chromosomes. To assign a SNP to the X1 or X2, RNA-seq data from primed UCLA1 and H9 were used. Due to the non-random silencing of one of the two X chromosomes, all mono-allelicaly expressed alleles are assigned to the Xa, while

the not expressed allele is assigned to the Xi in primed hESCs. To prevent confusion, we refer to the primed Xa as X1 and the primed Xi as X2. Specifically, for haplotype phasing we followed the following steps. Step 1: For each SNP we determine allelic coverage in primed UCLA1 and H9. To this end we used SAMtools mpileup tool (Li et al., 2009) to generate SNP coverage pileups, using min-BQ = 20, and max-depth = 1000000. For each SNP, the read coverage of the reference and alternative allele was calculated. 2) We then assigned each allele to the primed Xa (X1) and Xi (X2), by focusing on X-linked SNPs with mono allelic expression (for which the total number of reads > 5 and the absolute ratio of reference to alternative allele < 0.2 (80% of reads aligned to either the reference or alternative allele). In cases where the reference allele overlapped with more than 80% of reads, the reference allele was assigned to the Xa while the alternative allele was assigned to the Xi, and vice versa. This approach provided a list of alleles detected on the Xa (X1) and Xi (X2) for UCLA1 (39 X-lined phased SNPs) and H9 (524 X-linked phased SNPs).

Haplotype phasing is particularly important for SNP analysis of scRNA-seq data, since for assessing the bi- or mono-allelical presence of a SNP, a larger number of RNA-seq reads needs to overlap that SNP. However, in our scRNA-seq data, only a handful of SNP had more than five overlapping RNA-seq reads. More so, these SNPs with many overlapping reads were often located in genes that escape XCD, likely because XCD escapers tend to be more highly expressed. Consequently, it is difficult to assess the expression state of a X chromosome (active or inactive) based on individual SNPs. The phasing of X-linked SNPs described above enables the pooled analysis of all RNA reads

coming from the X1 or X2 alleles (regardless of the number of reads overlapping a specific SNP)

**Determination of allelic X-linked expression in RNA-seq data**

SAMtools mpileup tool (Li et al., 2009) was used to generate SNP coverage pileups for bulk RNA-seq data, or for each single cell in scRNA-seq data, using min-BQ = 20, and max-depth = 1000000. The number of reads covering X1 and X2, along with the ratio of reference to alternative allele coverage, were calculated for each SNP. For the phased SNP analysis in scRNA-seq data, the number of reads aligned to X1 and X2 was also calculated. When calling SNP coverage in the scRNA data, only cells with more than five read overlapping all phased SNPs were kept.

**RAP-seq**

RNA anti-sense purification followed by DNA sequencing (RAP-seq) for human XIST was adapted from Engreitz et al. (Engreitz et al., 2013). Human XIST probes for RAP-seq were initially generously designed by (as previously described for mouse Xist) and gifted from Guttman Lab at Caltech (https://www.guttmanlab.caltech.edu/). After washing with PBS, 10-30 million cells were harvested from confluent cell cultures and incubated with freshly-made 10ml 2mM DSG in 1x PBS at room temperature for 45 minutes. Cells were subsequently crosslinked further with 10ml 3% formaldehyde for 10 minutes, and the reaction was stopped by the addition of 2ml 2.5M glycine. Cells were pelleted at 4˚C and subjected to lysate preparation as described by Engreitz et al. (Engreitz et al., 2013).

Briefly, the fixed cell pellets were lysed in NP-40 containing cell lysis buffer either with glass dounce homogenizer (UCLA1, iPSCs and NHDFs) or without (H9), with further lysis of nuclei in a buffer with NP-40, sodium deoxycholate and N-lauroylsarcosine. Chromatin was solubilized by sonication and segmented by TURBO DNase digestion, followed by XIST RNA pulldown from 5 million cells using 1ug (UCLA1, iPSCs, NHDF R1) or 5ug (NHDF R2) and Streptavidin C1 beads. For H9, we generated a new set of 162 custom, non-overlapping 90nt long biotinylated oligonucleotides (Eurofins) and used 50pmol. In addition, for each cell line we also performed RAP in non-crosslinked cells (refer to as input) as a control for XIST enrichment. DNA was eluted by Rnase H digestion, and crosslinking was reversed via proteinase K digestion of eluted DNA at 60˚C. The DNA libraries were prepared using NEBNext Ultra End Rpair/dA-Tailing Module (NEB) and TruSeq DNA adapters (Illumina) ligated using Quick Ligase (NEB). Libraries were amplified by KAPA HiFi Polymerase (Roche), pooled, and sequenced on the Illlumina HiSeq platform to generate 50bp single-end reads (UCLA1, iPSCs and NHDFs) or 50 bp pair-end reads (H9).

**RAP-seq alignment**

DNA sequencing reads were trimmed using trim_galore (https://github.com/FelixKrueger/TrimGalore) with default parameters to remove the standard Illumina adaptor sequence. Bowtie2 (Langmead and Salzberg, 2012) was used to align reads to the human genome (hg38) with the default parameters. Reads with mapping quality less than 30 were removed using Samtools (Li et al., 2009), and Picard

MarkDuplicates (“Picard Tools - By Broad Institute,” n.d.) were used to mark PCR duplicates.

**Calculation of RAP-seq enrichment**

To calculate RAP-seq enrichment, we first defined genomic windows across the chromosomes using four different approaches: 1) We used bedtools makewindows (Quinlan and Hall, 2010) to create coordinates of genomic intervals in either 100Kb windows every 25Kb (for enrichment analysis), 2) 1Mb every 250Kb (for feature enrichment), 3) or non-overlapping 100kb windows (for differential analysis) along the genome. 4) We also defined genomic regions that encompass 50kb from start sites (TSSs) and over the gene body to capture the XIST level around gene bodies. This was used for comparing gene expression and XIST enrichment. We then used bedtools intersect (Quinlan and Hall, 2010) to count RAP-seq reads in each respective genomic region as defined by the four approaches, from each sample. To account for differences in sequencing depth, the read counts in each genomic region were normalized to the sum of all reads in that sample. To explore XIST enrichment on autosomes, we also calculated scaled enrichment score by normalizing only autosomal windows to the sum of all autosomal reads.

RAP-seq enrichment scores in each genomic region were calculated using the ratio of the normalized read counts in the RAP-seq XIST pulldown to the input of each sample. A region was defined as an unmappable region using the inputs of all samples.

Specifically, the R function isOutlier (scater package (McCarthy et al., 2017)) was used to identify outliers (having less or more than expected read counts) based on the minimal number of reads across all input samples (nmads>4). Genomic regions identified as outliers were removed from all downstream analysis.

The normalized enrichment ratios were used in all further computational analysis. We note that while the read counts of genomic intervals were defined in 100Kb or 1Mb windows every 25Kb and 250Kb, respectively, along the genome, the enrichment scores were assigned to the 25Kb or 250Kb windows in the center of the 100Kb and 1Mb windows. This was done to prevent overlapping windows in the downstream analysis.

**RAP-seq peak calling**

Peak calling was performed using MACS2 callpeak (Zhang et al., 2008) with max-gap=1000 using the input of each XIST pulldown sample as a control sample. For visualization, MACS2 bdgcmp was also used to generate fold-enrichment tracks. The normalized bedgraph files were then converted to tdf format using igvtools toTDF (Robinson et al., 2011). To compare peak scores between samples, we used bedtools merge (Quinlan and Hall, 2010) to compile a list of all detected peaks using either the narrow or broad peaks of all samples. We then used bedtools intersect (Quinlan and Hall, 2010) to intersect between the merged peaks, and the peaks identified in each sample.

**Differential analysis of RAP-seq data**

Two approaches were used to explore the differential of XIST localization on the X chromosome. First, MACS2 (Zhang et al., 2008) was used to call peaks in each of XIST pulldown, using the input as control, as described above. The R package DiffBind (Ross-Innes et al., 2012) was then used to identify differential peaks comparing all three naïve female hPSCs to the two Fibroblast cells, using the following filtering parameters: filter=10, filterFun=sum and keeping only X-chromosome peaks. Second, DESeq2 (Love et al., 2014) was used on the pulldown read counts in each 100Kb window. Regions with less than 10 reads in all samples were removed. Significant differences were defined as 100kb regions with p-value<0.01. Similar to first approach, only reads on the X chromosome where selected. We note that XIST is more enriched on the X-chromosome in fibroblast cells compared to the naïve cells, and filtering out autosomal reads allowed us to identify differences in the relative enrichment in somatic and naïve hPSCs.

To define conserved autosomal XIST peaks in naïve female hPSCs, broad XIST peaks were used, filtering out non-significant peaks (using p-value < 0.05, and fold enrichment of >2 for all samples, and >3 for iPSCs). Short length peaks (<500bps) were also removed. We note that the right arm of chromosome 8 is likely duplicated in the naïve UCLA1 hESC line, as noted by the increased reads in the UCLA1 input. Therefore, XIST peaks found in this regions (chr8, peak start >121000000) were removed. Conserved autosomal XIST peaks in naïve hPSCs were defined as peaks identified in all three naïve hPSC lines (H9, UCLA1 and iPSCs) and in none of the two fibroblast replicates.

Conserved XIST peaks in somatic cells were defined as peaks identified in both fibroblast replicates but none of the naïve hPSC samples.

**Features analysis for XIST localization on the X and on autosomes**

Two different approaches were used to identify genomic features associated with XIST enrichment. First, we correlated XIST enrichment scores in genomic intervals with interval scores of various genomic features, including gene density and repeats. For this, the gene annotation file was obtained from UCSC ("TxDb.Hsapiens.UCSC.hg38.knownGene," n.d.), and DNA repeat annotations were downloaded from the UCSC Genome Browser, track RepeatMasker, table rmsk (Karolchik et al., 2004). The number of annotated genes and repeats in a given genomic interval was calculated and pearson correlation was used to calculate the correlation between the XIST enrichment score and feature count in each genomic interval (1Mb windows every 250Kb). The XIST locus (and an additional 10Mb at either end) was removed for these analyses, similar to Engreitz et al. (Engreitz et al., 2013). Second, we calculated the enrichment of different features with conserved autosomal XIST peaks of naïve hPSCs. In addition to the features described above, we took advantage of the LOLA Core database (Sheffield and Bock, 2016). The R package LOLA (Sheffield and Bock, 2016) was used to compute the enrichment of the different features within the conserved autosomal XIST peaks.

**Gene Ontology Analysis**

Enrichment for gene ontologies was performed for genes overlapping with conserved, naïve hPSC-specific autosomal XIST peaks using the R package topGO (Alexa et al., 2006) with a Fisher test.

**Immunofluorescence staining**

For immunofluorescence staining, naïve hPSCs were seeded on feeder-coated coverslips and NHDF were seeded directly on the glass coverslips. After 24-48 hours, cells were fixed with 4% PFA for 10 minutes and washed with 1xPBS, afterwards permeabilized with 0.5% Triton X-100 in 1xPBS, and then blocked with 1% BSA in 1xPBS with 0.05% Tween-20 for 40 minutes. Primary antibody incubation was conducted 1% BSA for 1h at RT. Samples were again washed with 3xPBS-tween and incubated with fluorescent secondary antibodies at 1:400 for 45 min, then washed and counterstained with DAPI for 5 min and mounted using Vectashield. The secondary antibodies used in this study were all from Life Technologies used at 1:400 dilution. Images were taken using LSM 880 Confocal Instrument (Zeiss) or Zeiss Axio Imager M1. For image processing Fiji (ImageJ) was used.

**RNA FISH**

For RNA FISH, hPSCs were seeded on a feeder-coated coverslips 24-48 hours before fixation, which keeps hESC colony size small. NHDF were seeded directly on the glass coverslips. Coverslips were washed with DPBS, fixed with 4% formaldehyde for 10 min, permeabilized with cold (4°C) 0.5% Triton X-100 in DPBS for 10 min, and serially

dehydrated with cold (4°C) 70-100% ethanol for 10 minutes each step. Afterwards coverslips were air dried and hybridized with labeled DNA probes in a humidified chamber at 37°C overnight. Next day, coverslips were washed with 50% formamide in 2x SSC, 2x SSC, then 1x SSC at 42°C for 20 minutes each step. For nuclei staining, coverslips were stained with DAPI and mounted with Vectashield (Vector labs: H-1000). Double-stranded DNA probes were generated from full length cDNA constructs or BACs as described previously (Solovei, 2010). The BACs used include XIST (RP11-13M9), XACT (RP11-35D3), *GPC3* (RP11-678F20), *SMS* (RP11-147O5), *SMARCA1* (RP11-137A15), *THOC2* (RP11-121P4), *UTX* (RP11-256P2) and *SPON1* (RP11-774G22). Every new batch of probes was first tested on normal human dermal fibroblasts. For X chromosome paints combined with RNA FISH, XIST RNA FISH was done as described above and afterwards the coverslips were post-fixed with 4% PFA for 10 minutes and X chromosome painting was done following to manufacturer's instructions (Metasystems Probes, cat:D-0323-100-OR). Images were taken using LSM 880 Confocal Instrument (Zeiss) or Zeiss Axio Imager M1. For image processing and analysis Fiji (ImageJ) was used. For XACT signal quantification, images were converted into 8-bit images using Fiji and afterwards signal intensity was measured by drawing line over signal of interest and extracting pixel intensities with profile plot tool. Intensity values were exported as csv file for plotting.

Signal intensity of the nascent transcription focus of *THOC2* and *UTX* on the X-chromosome was done as described in Ding et al., (Ding and Elowitz, 2019).

**Generation of XIST KO lines**

To generate XIST KO lines, we electroporated primed H9 hESCs with two PX459 plasmids (Addgene #48139) carrying two different gRNAs, one targeting XIST promoter and one XIST exon 1 and both in case of successful transfection would generate 2kb deletion. gRNA sequences are added to STAR methods. gRNAs were cloned into PX459 as described before (Ran et al., 2013). Cell were electroporated using Lonza 4D-Nucleofector and P3 Primary Cell kit (Cat. V4XP-3024) according to manufacturer's instructions. Electroporated cells were seeded on feeder-coated plated in primed hESC media with 10uM ROCK inhibitor Y-27632. After 48 hours, cells were selected with puromycin (Puro cassette was expressed from PX459 plasmid) and surviving colonies were further propagated. To confirm the deletion, genomic DNA was isolated from the bulk population using Zymo Quick-DNA isolation kit (Cat. D4069) for PCR genotyping. PCR primers XP-Frw (CACAAAGATGTCCGGCTTTCA) and XE-Rev (CCTGCTGAATGCAAATGGGG) generate 1.5 kb bend upon deletion of 2kb. After this step, individual colonies were isolated from the targeted hESC population and screened for homozygous and heterozygous XIST deletions. To detect the WT allele, we used the primers XP-Frw and WT-Rev (CTCTGCCAAAGCGGTAGGTAC), which amplify a 942bp region from the WT allele and cannot amplify anything upon deletion as the WT-Rev complementary sequence is not present upon deletion. With this screening strategy, 2 homozygous clones were selected for the experiments.

**siRNA-mediated knockdown of *SPEN***

For siRNA-mediated depletion of *SPEN*, we used a mix of two different siRNAs targeting different exons of human SPEN (Thermo Fisher #4427037, IDs: s22831, s22829). Equal amounts of siRNA were mixed prior to transfection. As a negative, control scramble siRNA was used (Thermo #AM4611). Cells were transfected using Lipofectamine RNAiMAX reagent (Life Technologies Cat #13778150). After 24 hours of transfection, cells were transfected again using the same siRNA mix. After 2 rounds of transfection and in total 48 hours, cells were harvested, and RNA was isolated using Qiagen RNAeasy kit (#74104) according to manufacturer's instructions. Afterwards RNA-seq libraries were constructed as described above.

**Hi-C data analysis**

Xa and Xi Hi-C contacts (.hic files) from GM12878 somatic cells were downloaded from Rao et al. (Rao et al., 2014) (GSE63525). Knight-Ruiz (KR) normalized observed and expected contact matrixes were generated using Juicebox dump command of the Juicebox tool (Durand et al., 2016). Distance normalized interaction signals (observed/expected) were calculated at 1Mb resolution for the X chromosome.

## REFERENCES

Alexa, A., Rahnenführer, J., Lengauer, T., 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22, 1600–1607. https://doi.org/10.1093/bioinformatics/btl140

An, C., Feng, G., Zhang, J., Cao, S., Wang, Y., Wang, N., Lu, F., Zhou, Q., Wang, H., 2020. Overcoming Autocrine FGF Signaling-Induced Heterogeneity in Naive Human ESCs Enables Modeling of Random X Chromosome Inactivation. Cell Stem Cell 27, 482-497.e4. https://doi.org/10.1016/j.stem.2020.06.002

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Augui, S., Nora, E.P., Heard, E., 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. 12, 429–442. https://doi.org/10.1038/nrg2987

Avner, P., Heard, E., 2001. X-chromosome inactivation: counting, choice and initiation. Nat. Rev. Genet. 2, 59–67. https://doi.org/10.1038/35047580

Berletch, J.B., Ma, W., Yang, F., Shendure, J., Noble, W.S., Disteche, C.M., Deng, X., 2015. Escape from X inactivation varies in mouse tissues. PLoS Genet. 11, e1005079. https://doi.org/10.1371/journal.pgen.1005079

Bonora, G., Deng, X., Fang, H., Ramani, V., Qiu, R., Berletch, J.B., Filippova, G.N., Duan, Z., Shendure, J., Noble, W.S., Disteche, C.M., 2018. Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. Nat. Commun. 9, 1445. https://doi.org/10.1038/s41467-018-03694-y

Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.-B., Vassilev, I., Galupa, R., Servant, N., Barillot, E., Surani, A., Chen, C.-J., Heard, E., 2017. Xist -dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. 24, 226–233. https://doi.org/10.1038/nsmb.3365

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

Brockdorff, N., Bowness, J.S., Wei, G., 2020. Progress toward understanding chromosome silencing by Xist RNA. Genes Dev. 34, 733–744. https://doi.org/10.1101/gad.337196.120

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

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. https://doi.org/10.1038/nbt.4096

Cao, Y., Chen, G., Wu, G., Zhang, X., McDermott, J., Chen, X., Xu, C., Jiang, Q., Chen, Z., Zeng, Y., Ai, D., Huang, Y., Han, J.-D.J., 2019. Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions. Genome Res. 29, 40–52. https://doi.org/10.1101/gr.235747.118

Carrel, L., Willard, H.F., 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434, 400–404. https://doi.org/10.1038/nature03479

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., Tartaglia, G.G., 2019. Phase separation drives X-chromosome inactivation: a hypothesis. Nat. Struct. Mol. Biol. 26, 331–334. https://doi.org/10.1038/s41594-019-0223-0

Chitiashvili, T., Dror, I., Kim, R., Hsu, F.-M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K., Clark, A., 2020. Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. Nat. Cell Biol. 22, 1436–1446. https://doi.org/10.1038/s41556-020-00607-4

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., Chang, H.Y., 2015. Systematic discovery of Xist RNA binding proteins. Cell 161, 404–416. https://doi.org/10.1016/j.cell.2015.03.025

Cline, T.W., Meyer, B.J., 1996. Vive la différence: males vs females in flies vs worms. Annu. Rev. Genet. 30, 637–702. https://doi.org/10.1146/annurev.genet.30.1.637

Collier, A.J., Panula, S.P., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., Rugg-Gunn, P.J., 2017. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. Cell Stem Cell 20, 874-890.e7. https://doi.org/10.1016/j.stem.2017.02.014

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., Meyer, B.J., 2015. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature 523, 240–244. https://doi.org/10.1038/nature14450

Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.-C., Sanborn, A.L., Machol, I., Shamim, M., Seberg, A.P., Lander, E.S., Chadwick, B.P., Aiden, E.L., 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. 113, E4504–E4512. https://doi.org/10.1073/pnas.1609643113

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., Onate, K.C., Graham, K., Miyasato, S.R., Dreszer, T.R., Strattan, J.S., Jolanki, O., Tanaka, F.Y., Cherry, J.M., 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 46, D794–D801. https://doi.org/10.1093/nar/gkx1081

Deng, X., Berletch, J.B., Nguyen, D.K., Disteche, C.M., 2014. X chromosome regulation: diverse patterns in development, tissues and disease. Nat. Rev. Genet. 15, 367–378. https://doi.org/10.1038/nrg3687

Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J.B., Blau, C.A., Shendure, J., Duan, Z., Noble, W.S., Disteche, C.M., 2015. Bipartite structure of the inactive mouse X chromosome. Genome Biol. 16, 152. https://doi.org/10.1186/s13059-015-0728-8

Ding, F., Elowitz, M.B., 2019. Constitutive splicing and economies of scale in gene expression. Nat. Struct. Mol. Biol. 26, 424–432. https://doi.org/10.1038/s41594-019-0226-x

Dossin, F., Pinheiro, I., Żylicz, J.J., 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., 2020. SPEN integrates transcriptional and epigenetic control of X-inactivation. Nature 578, 455–460. https://doi.org/10.1038/s41586-020-1974-9

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., Aiden, E.L., 2016. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. 3, 99–101. https://doi.org/10.1016/j.cels.2015.07.012

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., Plath, K., Guttman, M., 2013. The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. Science 341, 1237973. https://doi.org/10.1126/science.1237973

Ercan, S., Dick, L.L., Lieb, J.D., 2009. The C. elegans Dosage Compensation Complex Propagates Dynamically and Independently of X Chromosome Sequence. Curr. Biol. 19, 1777–1787. https://doi.org/10.1016/j.cub.2009.09.047

Ferrari, F., Alekseyenko, A.A., Park, P.J., Kuroda, M.I., 2014. Transcriptional control of a whole chromosome: emerging models for dosage compensation. Nat. Struct. Mol. Biol. 21, 118–125. https://doi.org/10.1038/nsmb.2763

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., Chambers, I., 2012. Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell 11, 477–490. https://doi.org/10.1016/j.stem.2012.08.002

Froberg, J.E., Pinter, S.F., Kriz, A.J., Jégu, T., Lee, J.T., 2018. Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. Nat. Commun. 9, 5004. https://doi.org/10.1038/s41467-018-07446-w

Galupa, R., Heard, E., 2015. X-chromosome inactivation: new insights into cis and trans regulation. Curr. Opin. Genet. Dev., Genome architecture and expression 31, 57–66. https://doi.org/10.1016/j.gde.2015.04.002

Gardner, H.P., Belka, G.K., Wertheim, G.B., Hartman, J.L., Ha, S.I., Gimotty, P.A., Marquis, S.T., Chodosh, L.A., 2000. Developmental role of the SNF1-related kinase Hunk in pregnancy-induced changes in the mammary gland. Dev. Camb. Engl. 127, 4493–4509.

Gendrel, A.-V., Heard, E., 2014. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu. Rev. Cell Dev. Biol. 30, 561–580. https://doi.org/10.1146/annurev-cellbio-101512-122415

Gimferrer, I., Farnós, M., Calvo, M., Mittelbrunn, M., Enrich, C., Sánchez-Madrid, F., Vives, J., Lozano, F., 2003. The accessory molecules CD5 and CD6 associate on the membrane of lymphoid T cells. J. Biol. Chem. 278, 8564–8571. https://doi.org/10.1074/jbc.M209591200

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., Dekker, J., 2016. Structural organization of the inactive X chromosome in the mouse. Nature 535, 575–579. https://doi.org/10.1038/nature18589

Gizard, F., Robillard, R., Gross, B., Barbier, O., Révillion, F., Peyrat, J.-P., Torpier, G., Hum, D.W., Staels, B., 2006. TReP-132 Is a Novel Progesterone Receptor Coactivator

Required for the Inhibition of Breast Cancer Cell Growth and Enhancement of Differentiation by Progesterone. Mol. Cell. Biol. 26, 7632–7644. https://doi.org/10.1128/MCB.00326-06

Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., Plath, K., Smith, A., 2017. Epigenetic resetting of human pluripotency. Dev. Camb. Engl. 144, 2748–2763. https://doi.org/10.1242/dev.146811

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., Nichols, J., 2016. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. Stem Cell Rep. 6, 437–446. https://doi.org/10.1016/j.stemcr.2016.02.005

Gyllborg, D., Ahmed, M., Toledo, E.M., Theofilopoulos, S., Yang, S., ffrench-Constant, C., Arenas, E., 2018. The Matricellular Protein R-Spondin 2 Promotes Midbrain Dopaminergic Neurogenesis and Differentiation. Stem Cell Rep. 11, 651–664. https://doi.org/10.1016/j.stemcr.2018.07.014

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

Hall, L.L., Clemson, C.M., Byron, M., Wydner, K., Lawrence, J.B., 2002b. Unbalanced X;autosome translocations provide evidence for sequence specificity in the association

of XIST RNA with chromatin. Hum. Mol. Genet. 11, 3157–3165. https://doi.org/10.1093/hmg/11.25.3157

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, Kimiko, Tsutui, Ken, Nakagawa, S., 2010. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. Dev. Cell 19, 469–476. https://doi.org/10.1016/j.devcel.2010.08.006

Horakova, A.H., Moseley, S.C., McLaughlin, C.R., Tremblay, D.C., Chadwick, B.P., 2012. The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. Hum. Mol. Genet. 21, 4367–4377. https://doi.org/10.1093/hmg/dds270

Huang, K., Maruyama, T., Fan, G., 2014. The Naive State of Human Pluripotent Stem Cells: A Synthesis of Stem Cell and Preimplantation Embryo Transcriptome Analyses. Cell Stem Cell 15, 410–415. https://doi.org/10.1016/j.stem.2014.09.014

Jiang, J., Jing, Y., Cost, G.J., Chiang, J.-C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., Pearl, J.R., Rebar, E.J., Byron, M., Gregory, P.D., Brown, C.J., Urnov, F.D., Hall, L.L., Lawrence, J.B., 2013. Translating dosage compensation to trisomy 21. Nature 500, 296–300. https://doi.org/10.1038/nature12394

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., Gribnau, J., 2008. Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. 28, 5583–5594. https://doi.org/10.1128/MCB.02269-07

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., Kent, W.J., 2004. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 32, D493-496. https://doi.org/10.1093/nar/gkh103

Kelsey, A.D., Yang, C., Leung, D., Minks, J., Dixon-McDougall, T., Baldry, S.E.L., Bogutz, A.B., Lefebvre, L., Brown, C.J., 2015. Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. Genome Biol. 16, 208. https://doi.org/10.1186/s13059-015-0774-2

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, X., Nefzger, C.M., Rossello, F.J., Chen, J., Knaupp, A.S., Firas, J., Ford, E., Pflueger, J., Paynter, J.M., Chy, H.S., O'Brien, C.M., Huang, C., Mishra, K., Hodgson-Garms, M., Jansz, N., Williams, S.M., Blewitt, M.E., Nilsson, S.K., Schittenhelm, R.B., Laslett, A.L., Lister, R., Polo, J.M., 2017. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. Nat. Methods 14, 1055–1062. https://doi.org/10.1038/nmeth.4436

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., Mesirov, J.P., Cech, T.R., Chang, H.Y., 2016. RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell 165, 1267–1279. https://doi.org/10.1016/j.cell.2016.04.028

Lyon, M.F., 1961. Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190, 372–373. https://doi.org/10.1038/190372a0

Mahadevaiah, S.K., Sangrithi, M.N., Hirota, T., Turner, J.M.A., 2020. A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. Nature 586, 612–617. https://doi.org/10.1038/s41586-020-2629-6

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., Brockdorff, N., 2004. Reactivation of the paternal X chromosome in early mouse embryos. Science 303, 666–669. https://doi.org/10.1126/science.1092674

Marahrens, Y., Loring, J., Jaenisch, R., 1998. Role of the Xist Gene in X Chromosome Choosing. Cell 92, 657–664. https://doi.org/10.1016/S0092-8674(00)81133-2

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., Jaenisch, R., 1997. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 11, 156–166. https://doi.org/10.1101/gad.11.2.156

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., Smith, A., 2012. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell 11, 491–504. https://doi.org/10.1016/j.stem.2012.06.008

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., Wills, Q.F., 2017. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33, 1179–1186. https://doi.org/10.1093/bioinformatics/btw777

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M.J., Shishkin, A.A., Su, J., Lander, E.S., Hess, S., Plath, K., Guttman, M., 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521, 232–236. https://doi.org/10.1038/nature14443

Medstrand, P., van de Lagemaat, L.N., Mager, D.L., 2002. Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes. Genome Res. 12, 1483–1495. https://doi.org/10.1101/gr.388902

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., Lee, J.T., 2015. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science 349. https://doi.org/10.1126/science.aab2276

Minkovsky, A., Patel, S., Plath, K., 2012. Concise Review: Pluripotency and the Transcriptional Inactivation of the Female Mammalian X Chromosome. STEM CELLS 30, 48–54. https://doi.org/10.1002/stem.755

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J.C., Reik, W., 2017. Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. 20, 1215–1228. https://doi.org/10.1016/j.celrep.2017.07.009

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., Wutz, A., 2015. Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. Cell Rep. 12, 554–561. https://doi.org/10.1016/j.celrep.2015.06.067

Navarro-Cobos, M.J., Balaton, B.P., Brown, C.J., 2020. Genes that escape from X-chromosome inactivation: Potential contributors to Klinefelter syndrome. Am. J. Med. Genet. C Semin. Med. Genet. 184, 226–238. https://doi.org/10.1002/ajmg.c.31800

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., Zakian, S.M., 2001. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res. 11, 833–849. https://doi.org/10.1101/gr.174901

Nesterova, T.B., Wei, G., Coker, H., Pintacuda, G., Bowness, J.S., Zhang, T., Almeida, M., Bloechl, B., Moindrot, B., Carter, E.J., Alvarez Rodrigo, I., Pan, Q., Bi, Y., Song, C.-X., Brockdorff, N., 2019. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. Nat. Commun. 10, 3129. https://doi.org/10.1038/s41467-019-11171-3

Nichols, J., Smith, A., 2009. Naive and primed pluripotent states. Cell Stem Cell 4, 487–492. https://doi.org/10.1016/j.stem.2009.05.015

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., Heard, E., 2004. Epigenetic Dynamics of Imprinted X Inactivation During Early Mouse Development. Science 303, 644–649. https://doi.org/10.1126/science.1092727

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

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W.R., Damianov, A., Chronis, C., Papp, B., Chen, C.-K., McKee, R., Wang, X.-J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, D.L., Plath, K., 2020. A protein assembly mediates Xist localization and gene silencing. Nature 587, 145–151. https://doi.org/10.1038/s41586-020-2703-0

Pandya-Jones, A., Plath, K., 2016. The “lnc” between 3D Chromatin Structure and X Chromosome Inactivation. Semin. Cell Dev. Biol. 56, 35–47. https://doi.org/10.1016/j.semcdb.2016.04.002

Payer, B., Lee, J.T., 2008. X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. 42, 733–772. https://doi.org/10.1146/annurev.genet.42.110807.091711

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., Brockdorff, N., 1996. Requirement for Xist in X chromosome inactivation. Nature 379, 131–137. https://doi.org/10.1038/379131a0

Petri, R., Brattås, P.L., Sharma, Y., Jönsson, M.E., Pircs, K., Bengzon, J., Jakobsson, J., 2019. LINE-2 transposable elements are a source of functional human microRNAs and target sites. PLoS Genet. 15. https://doi.org/10.1371/journal.pgen.1008036

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., Lanner, F., 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell 165, 1012–1026. https://doi.org/10.1016/j.cell.2016.03.023

Picard Tools - By Broad Institute [WWW Document], n.d. URL http://broadinstitute.github.io/picard/ (accessed 12.8.20).

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., Panning, B., 2002. Xist RNA and the Mechanism of X Chromosome Inactivation. Annu. Rev. Genet. 36, 233–278. https://doi.org/10.1146/annurev.genet.36.042902.092433

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308. https://doi.org/10.1038/nprot.2013.143

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Ridings-Figueroa, R., Stewart, E.R., Nesterova, T.B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., Bageghni, S.A., Albadrani, G., Mansfield, W., Roulson, J.-A., Brockdorff, N., Ainscough, J.F.X., Coverley, D., 2017. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. Genes Dev. 31, 876–888. https://doi.org/10.1101/gad.295907.117

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. https://doi.org/10.1038/nbt.1754

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., Ali, S., Chin, S.-F., Palmieri, C., Caldas, C., Carroll, J.S., 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393. https://doi.org/10.1038/nature10730

Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., Plath, K., 2017a. Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. Cell Stem Cell 20, 87–101. https://doi.org/10.1016/j.stem.2016.10.006

Sahakyan, A., Plath, K., 2016. Transcriptome Encyclopedia of Early Human Development. Cell 165, 777–779. https://doi.org/10.1016/j.cell.2016.04.042

Sahakyan, A., Plath, K., Rougeulle, C., 2017b. Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372. https://doi.org/10.1098/rstb.2016.0363

Samata, M., Akhtar, A., 2018. Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs. Annu. Rev. Biochem. 87, 323–350. https://doi.org/10.1146/annurev-biochem-062917-011816

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Schulz, E.G., Heard, E., 2013. Role and control of X chromosome dosage in mammalian development. Curr. Opin. Genet. Dev. 23, 109–115. https://doi.org/10.1016/j.gde.2013.01.008

Sheffield, N.C., Bock, C., 2016. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics 32, 587–589. https://doi.org/10.1093/bioinformatics/btv612

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., Lee, J.T., 2013. High-resolution Xist

binding maps reveal two-step spreading during X-chromosome inactivation. Nature 504, 465–469. https://doi.org/10.1038/nature12719

Solovei, I., 2010. Fluorescence in situ Hybridization (FISH) on Tissue Cryosections. Methods Mol. Biol. Clifton NJ 659, 71–82. https://doi.org/10.1007/978-1-60761-789-1_5

Straub, T., Becker, P.B., 2007. Dosage compensation: the beginning and end of generalization. Nat. Rev. Genet. 8, 47–57. https://doi.org/10.1038/nrg2013

Strome, S., Kelly, W.G., Ercan, S., Lieb, J.D., 2014. Regulation of the X Chromosomes in Caenorhabditis elegans. Cold Spring Harb. Perspect. Biol. 6, a018366. https://doi.org/10.1101/cshperspect.a018366

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Sunwoo, H., Colognori, D., Froberg, J.E., Jeon, Y., Lee, J.T., 2017. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). Proc. Natl. Acad. Sci. 114, 10654–10659. https://doi.org/10.1073/pnas.1711206114

Takagi, N., Sasaki, M., 1975. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature 256, 640–642. https://doi.org/10.1038/256640a0

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., Reik, W., Bertone, P., Smith, A., 2014. Resetting

Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell 158, 1254–1269. https://doi.org/10.1016/j.cell.2014.08.029

Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., Duc, J., Cohen, M.A., Wert, K.J., Castanon, R., Zhang, Z., Huang, Y., Nery, J.R., Drotar, J., Lungjangwa, T., Trono, D., Ecker, J.R., Jaenisch, R., 2016. Molecular Criteria for Defining the Naive Human Pluripotent State. Cell Stem Cell 19, 502–515. https://doi.org/10.1016/j.stem.2016.06.011

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., Lungjangwa, T., Imsoonthornruksa, S., Stelzer, Y., Rangarajan, S., D'Alessio, A., Zhang, J., Gao, Q., Dawlaty, M.M., Young, R.A., Gray, N.S., Jaenisch, R., 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15, 471–487. https://doi.org/10.1016/j.stem.2014.07.002

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., Cummings, B.B., Castel, S.E., Karczewski, K.J., Aguet, F., Byrnes, A., GTEx Consortium, Lappalainen, T., Aviv Regev, Ardlie, K.G., Hacohen, N., MacArthur, D.G., 2017. Landscape of X chromosome inactivation across human tissues. Nature 550, 244–248. https://doi.org/10.1038/nature24265

TxDb.Hsapiens.UCSC.hg38.knownGene [WWW Document], n.d. . Bioconductor. URL http://bioconductor.org/packages/TxDb.Hsapiens.UCSC.hg38.knownGene/ (accessed 12.8.20).

Vallot, C., Huret, C., Lesecque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A., Duret, L., Rougeulle, C., 2013. XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. Nat. Genet. 45, 239–241. https://doi.org/10.1038/ng.2530

Vallot, C., Patrat, C., Collier, A.J., Huret, C., Casanova, M., Ali, T.M.L., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J., Rougeulle, C., 2017. XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. Cell Stem Cell 20, 102. https://doi.org/10.1016/j.stem.2016.10.014

Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C., Zhuang, X., 2016. Spatial organization of chromatin domains and compartments in single chromosomes. Science 353, 598–602. https://doi.org/10.1126/science.aaf8084

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

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

Yang, C., Chapman, A.G., Kelsey, A.D., Minks, J., Cotton, A.M., Brown, C.J., 2011. X-chromosome inactivation: molecular mechanisms from the human perspective. Hum. Genet. 130, 175–185. https://doi.org/10.1007/s00439-011-0994-9

Yang, L., Kirby, J.E., Sunwoo, H., Lee, J.T., 2016. Female mice lacking Xist RNA show partial dosage compensation and survive to term. Genes Dev. 30, 1747–1760. https://doi.org/10.1101/gad.281162.116

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., Orkin, S.H., Shen, X., 2015. Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. Cell Stem Cell 16, 504–516. https://doi.org/10.1016/j.stem.2015.03.007

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

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., Wang, L., 2014. CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinforma. Oxf. Engl. 30, 1006–1007. https://doi.org/10.1093/bioinformatics/btt730

Figure 1

Figure 1 Dror et al.

Click here to access/download;Figure;Fig. 1 Dror et al.pdf

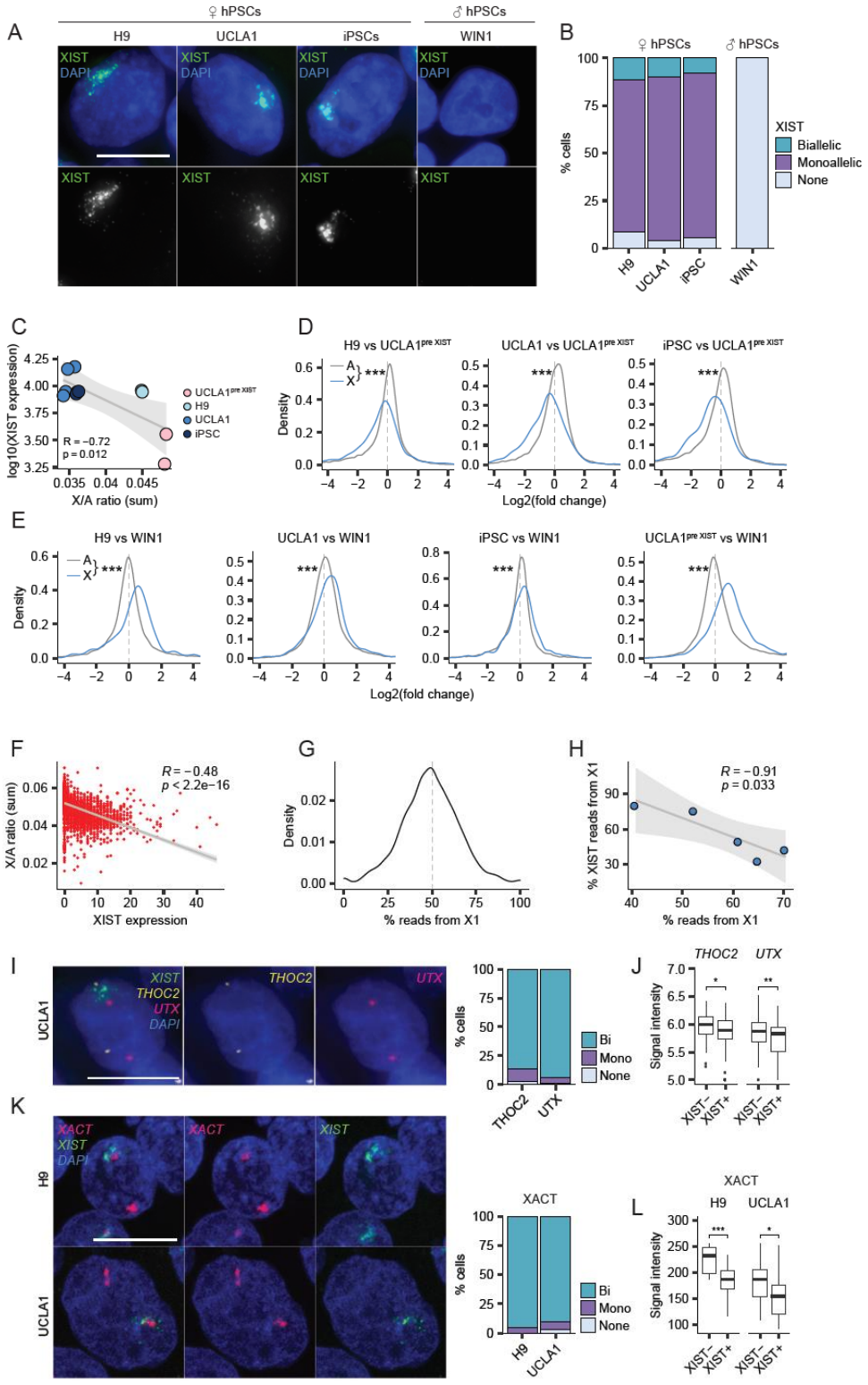


Figure 2

Click here to access/download;Figure;Fig. 2 Dror et al.pdf

Figure 2 Dror et al.

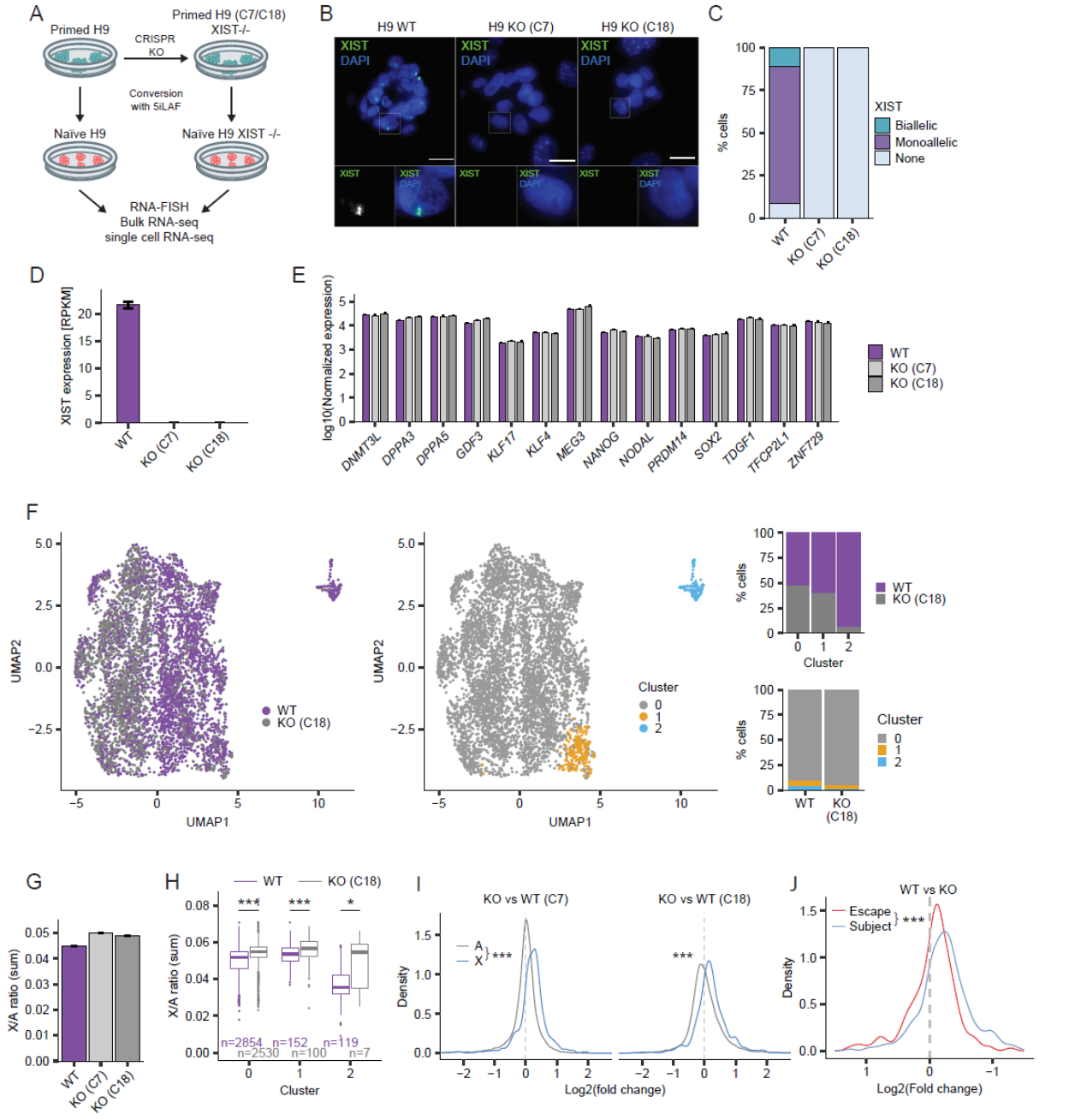


Figure 3

# Figure 3 Dror et al.

Click here to access/download;Figure;Fig. 3 Dror et al.pdf

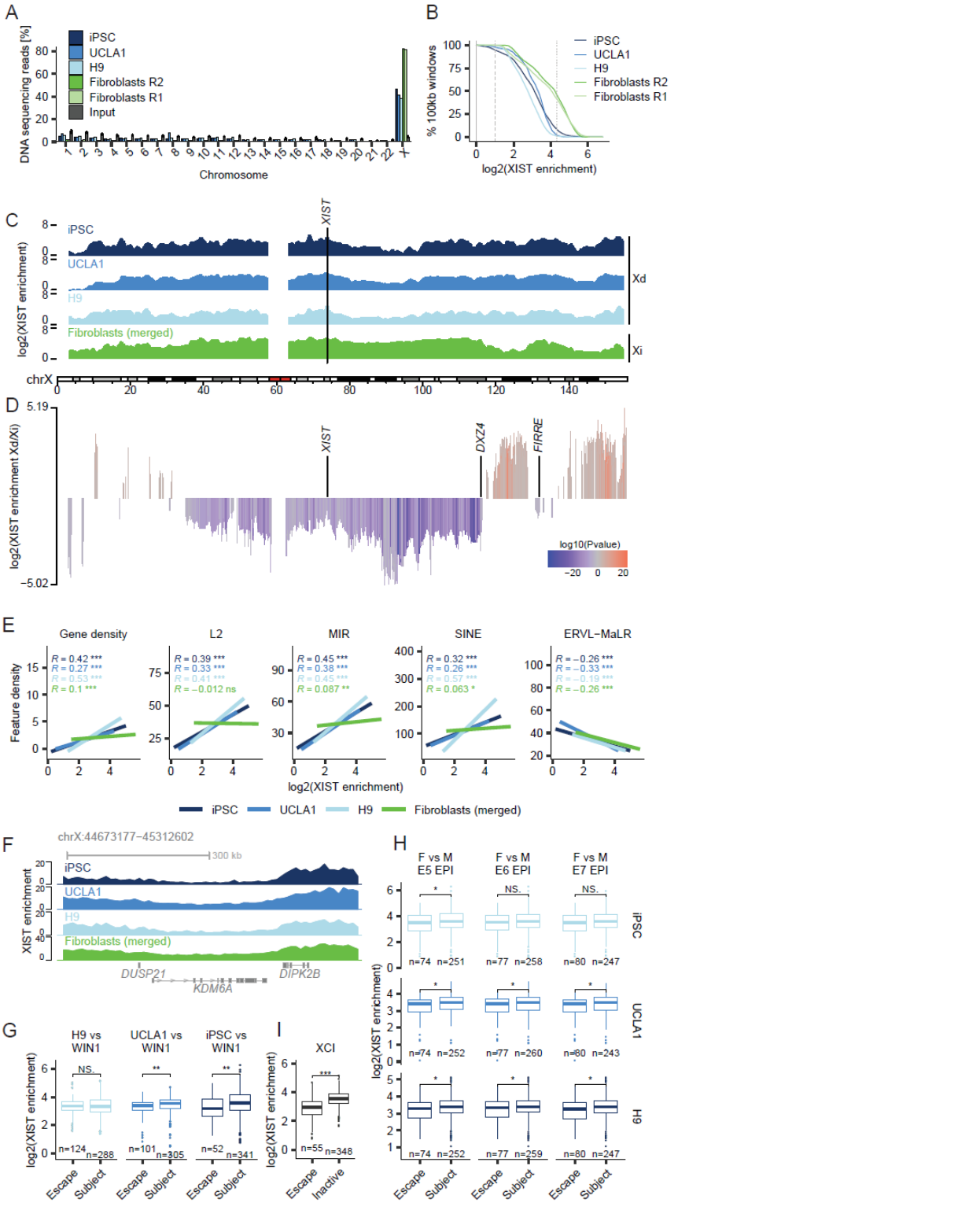


Figure 4

Click here to access/download;Figure;Fig. 4 Dror et al.pdf

# Figure 4 Dror et al.

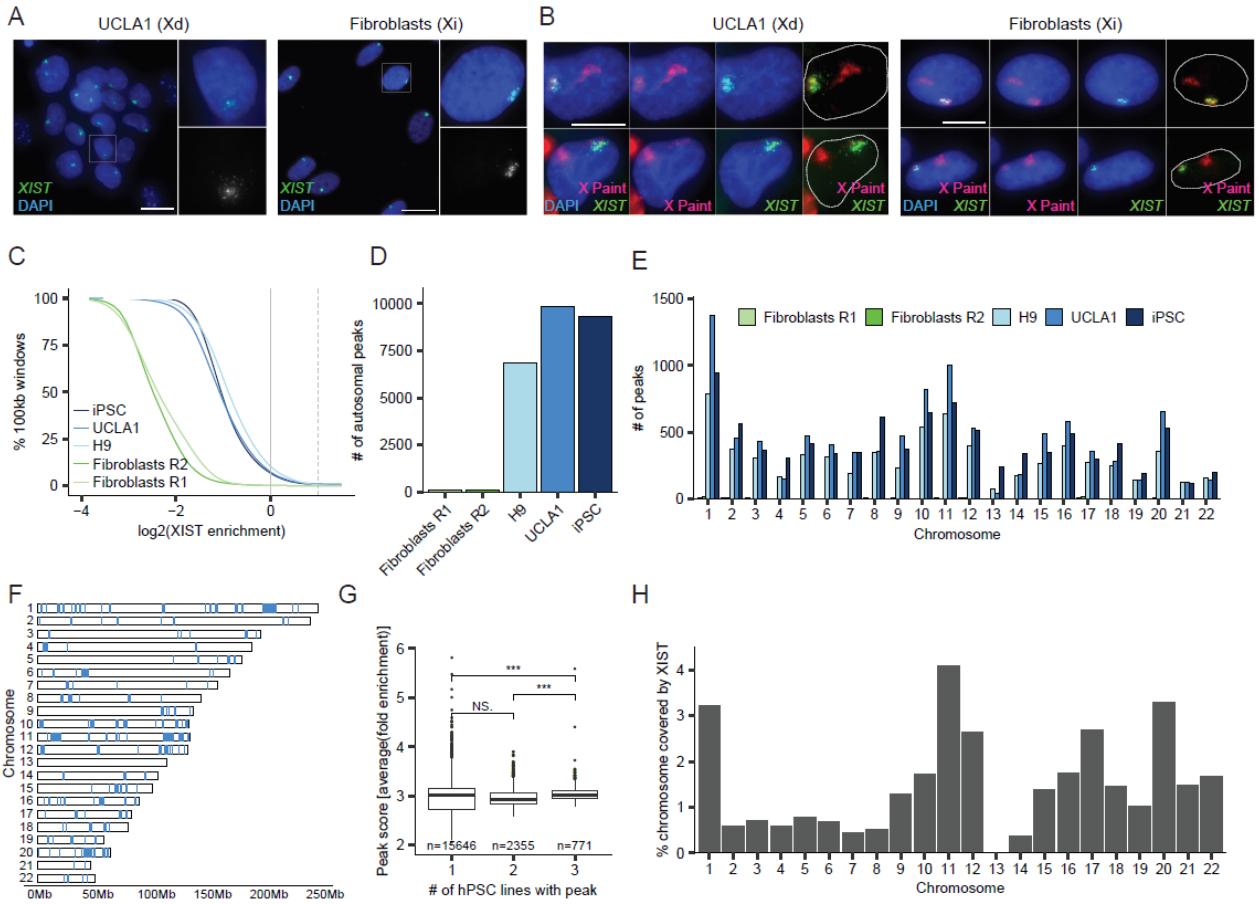


Figure 5

Click here to access/download;Figure;Fig. 5 Dror et al.pdf

# Figure 5 Dror et al.

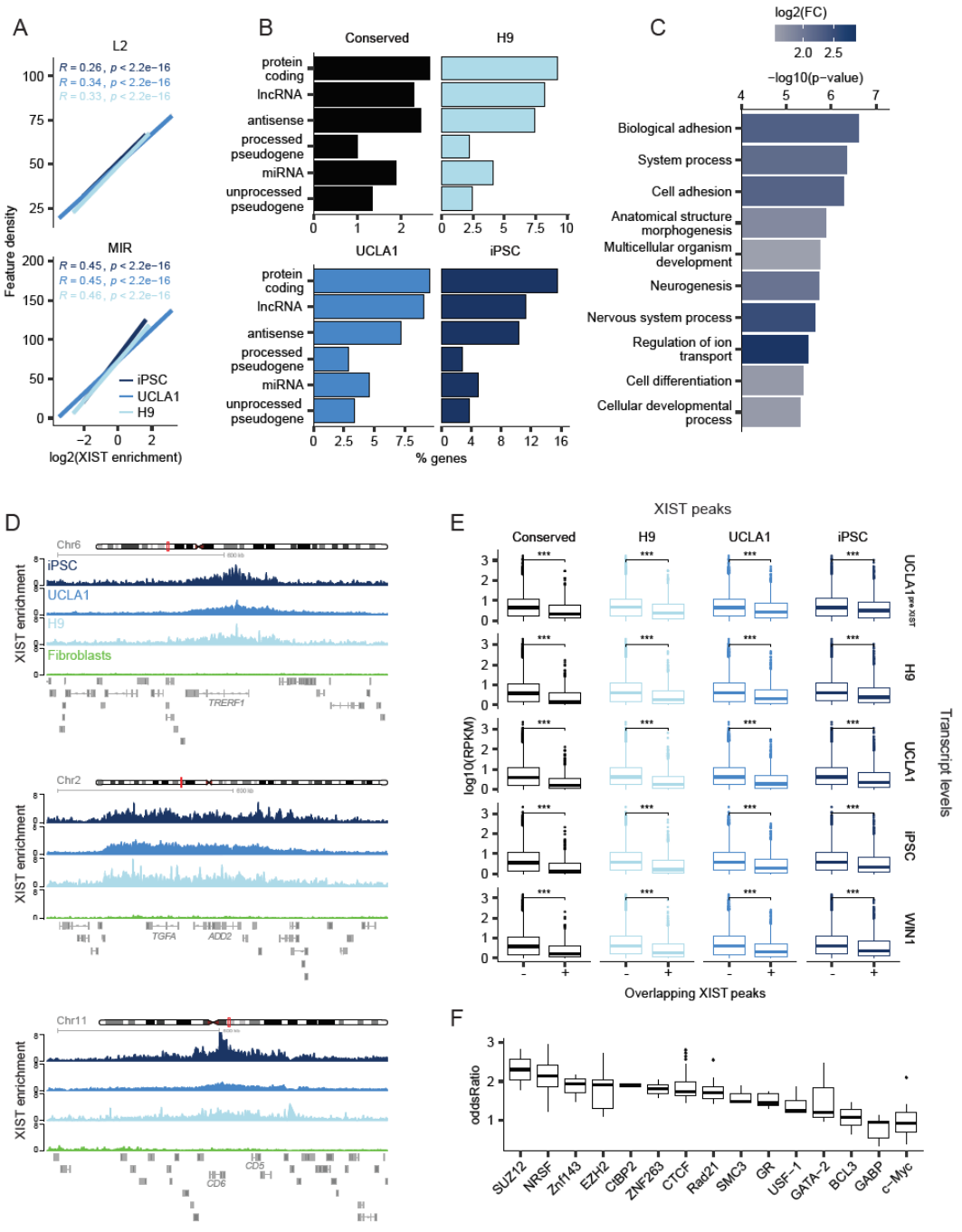


Figure 6

Click here to access/download;Figure;Fig. 6 Dror et al.pdf

# Figure 6 Dror et al.

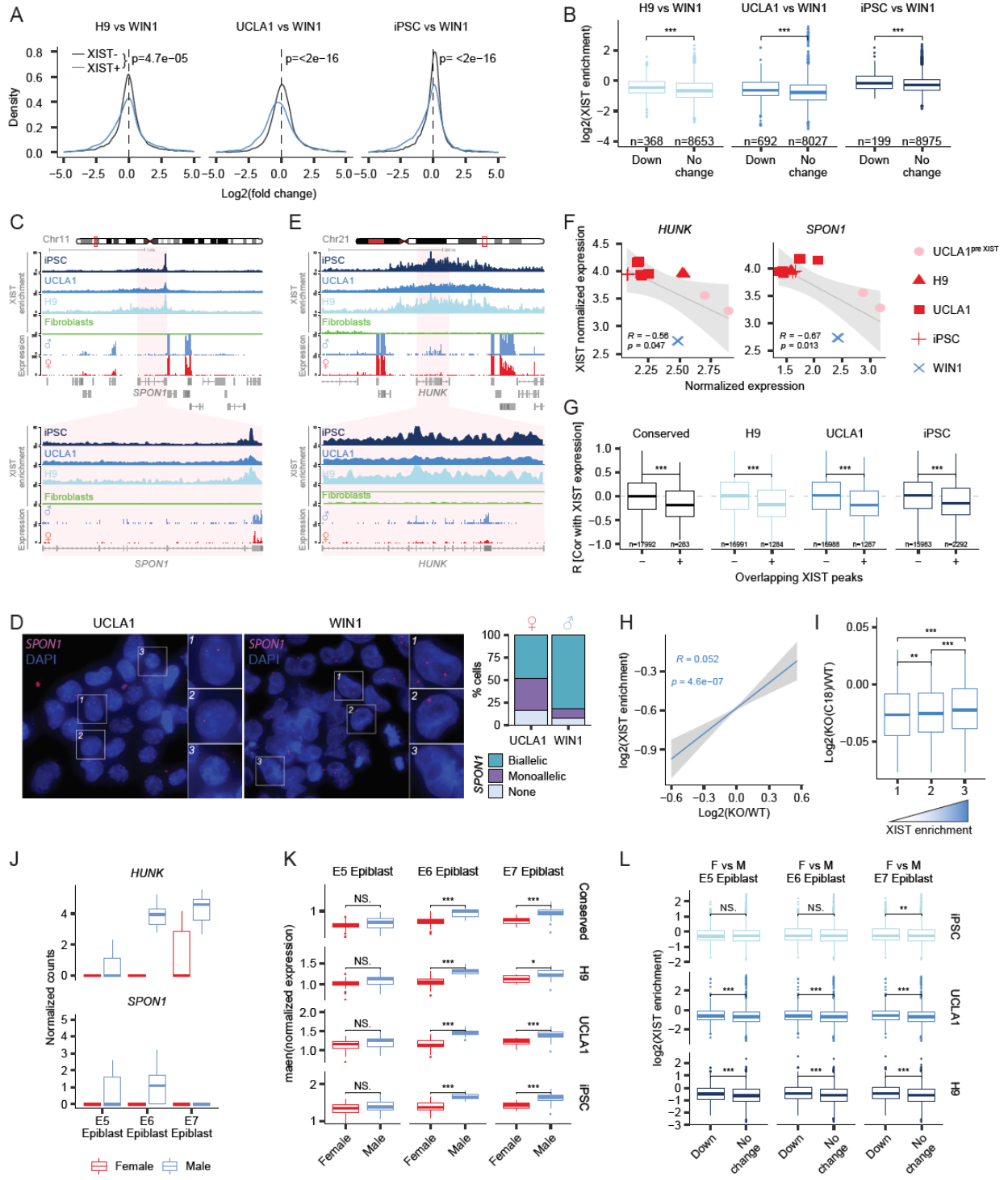
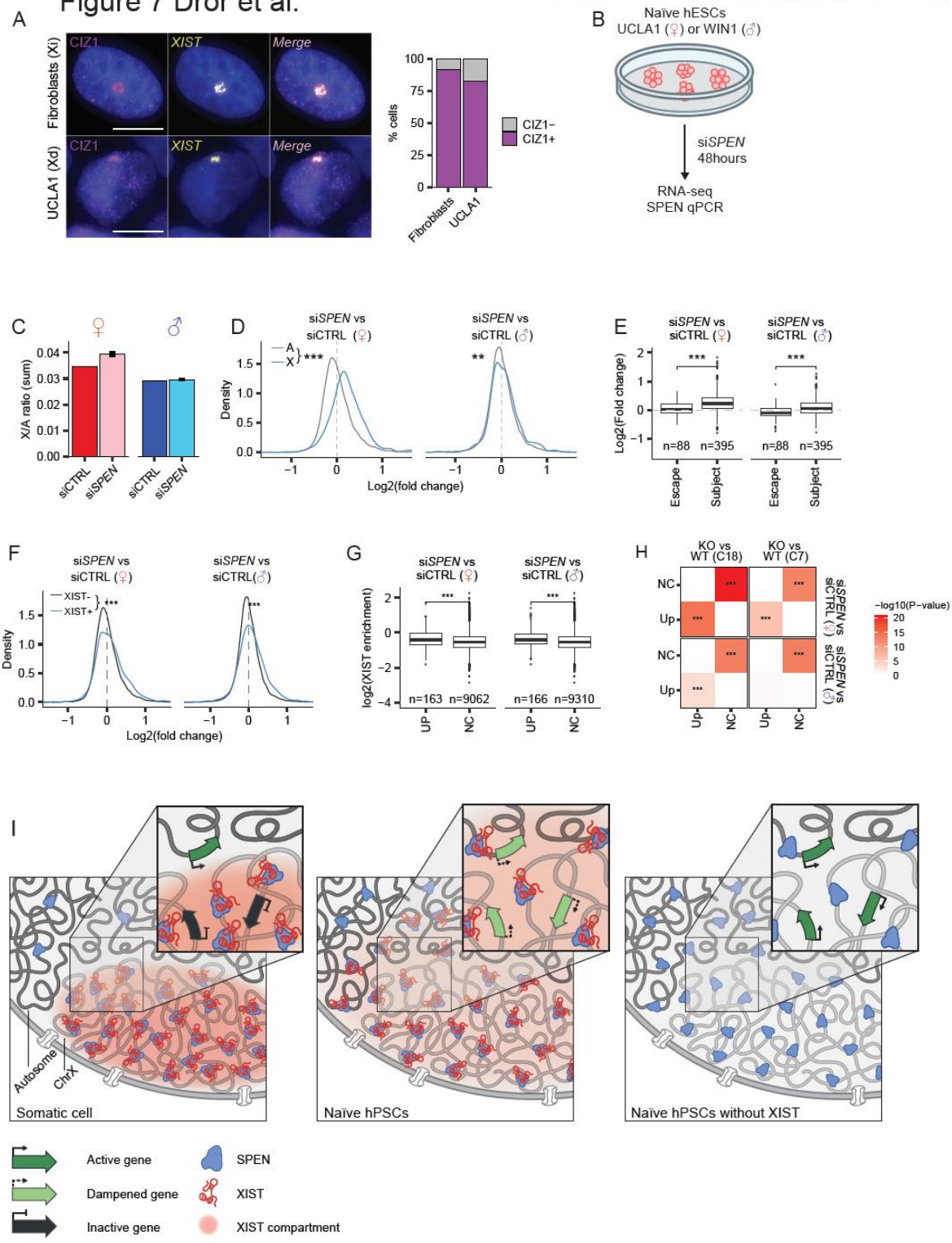


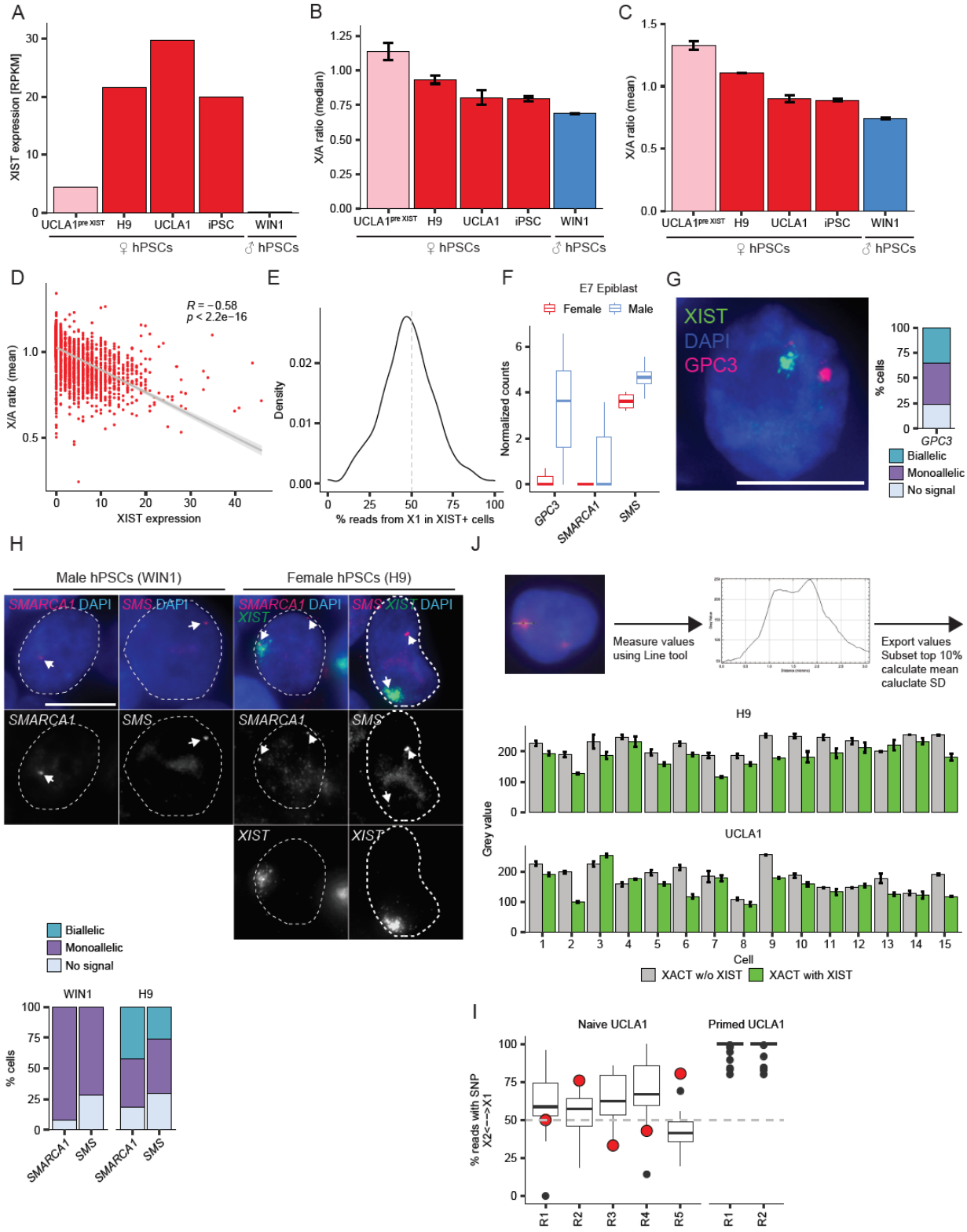
Figure 7

Click here to access/download;Figure;Fig. 7 Dror et al.pdf



Supplemental Figure S1

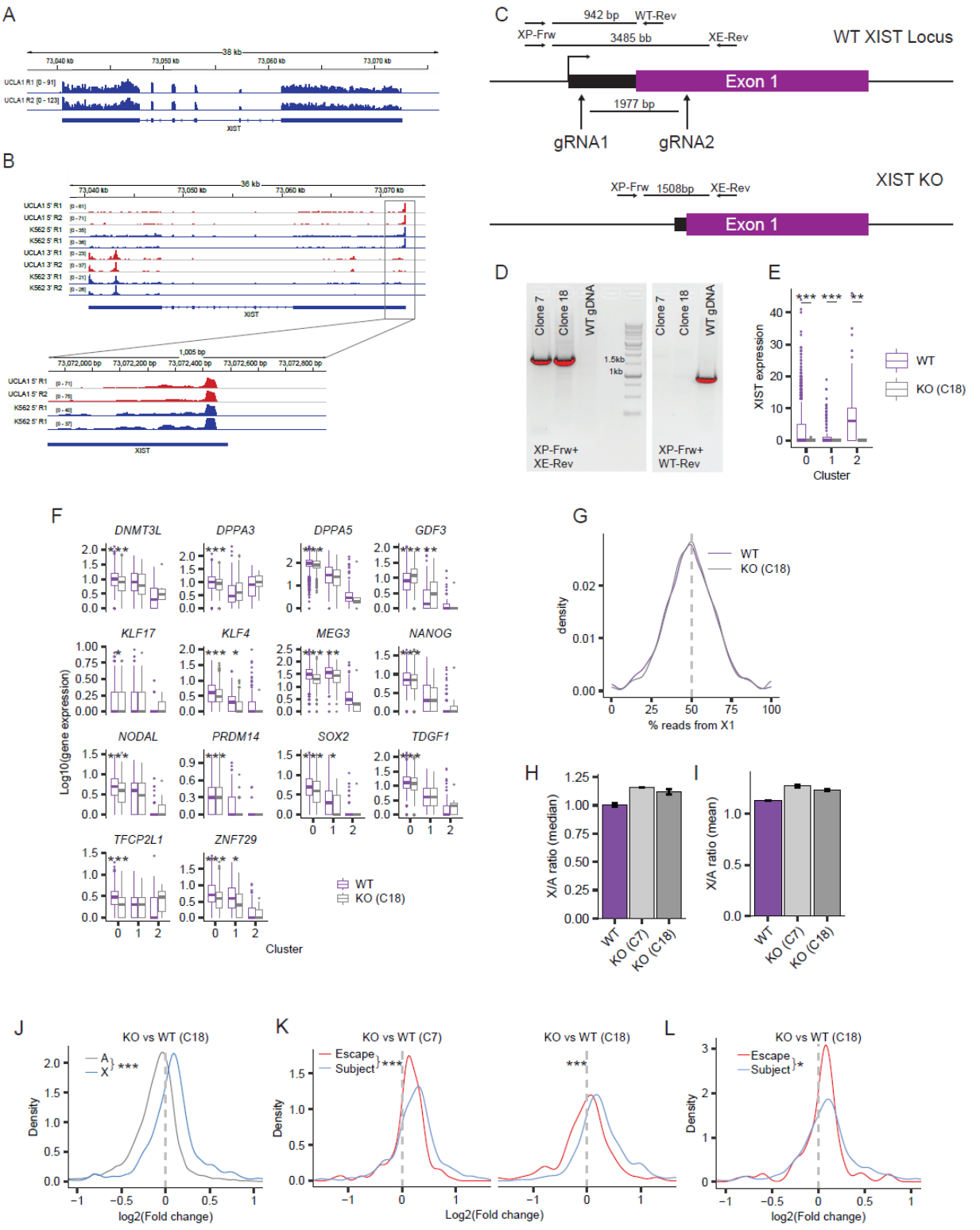
Click here to access/download;Supplemental Figure;Fig. S1 Dror et al.pdf



Supp Figure 2

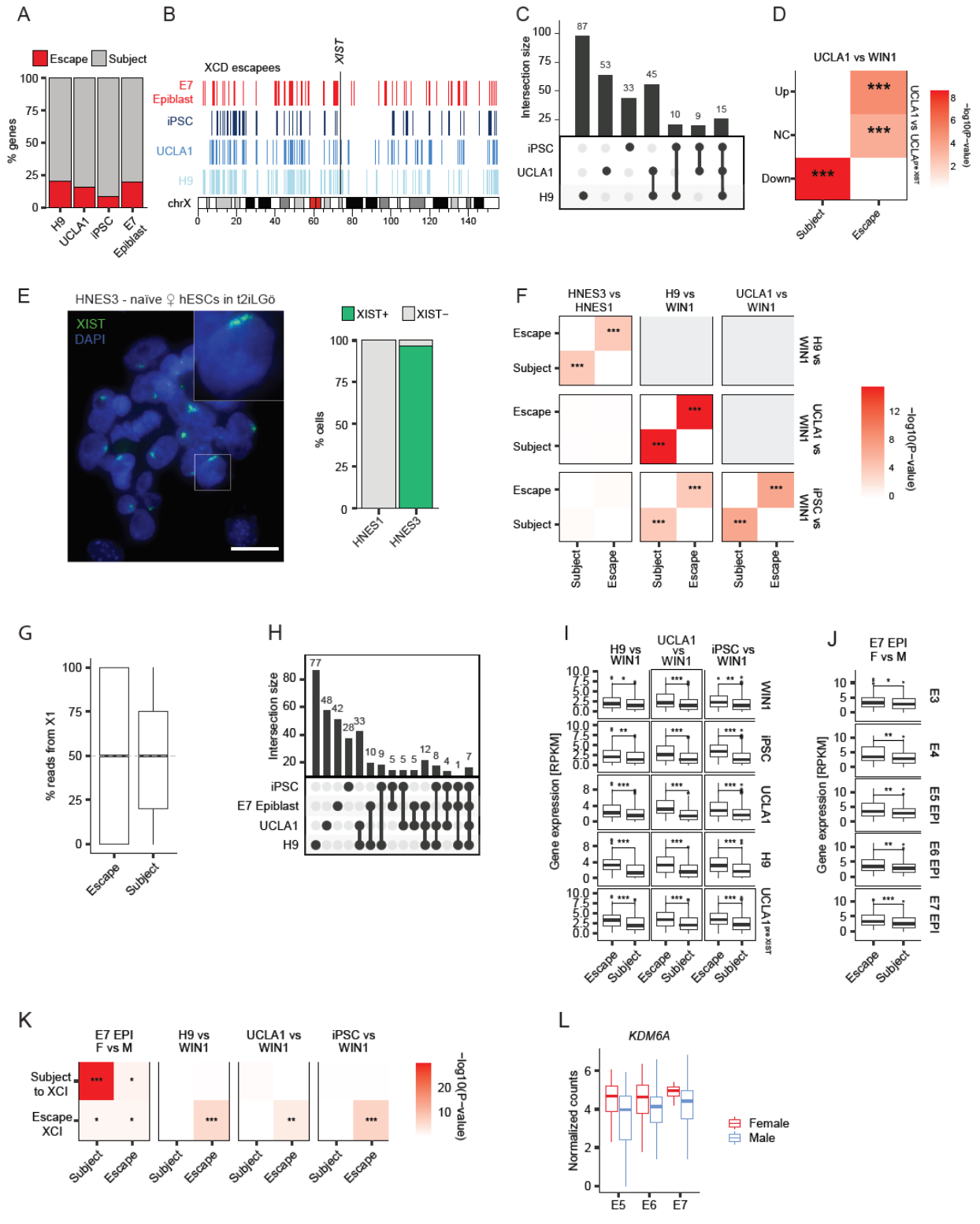
Figure S2 Dror et al.

Click here to access/download;Supplemental Figure;Fig. S2 Dror et al.pdf



Supp Figure 3
Figure S3 Dror et al.

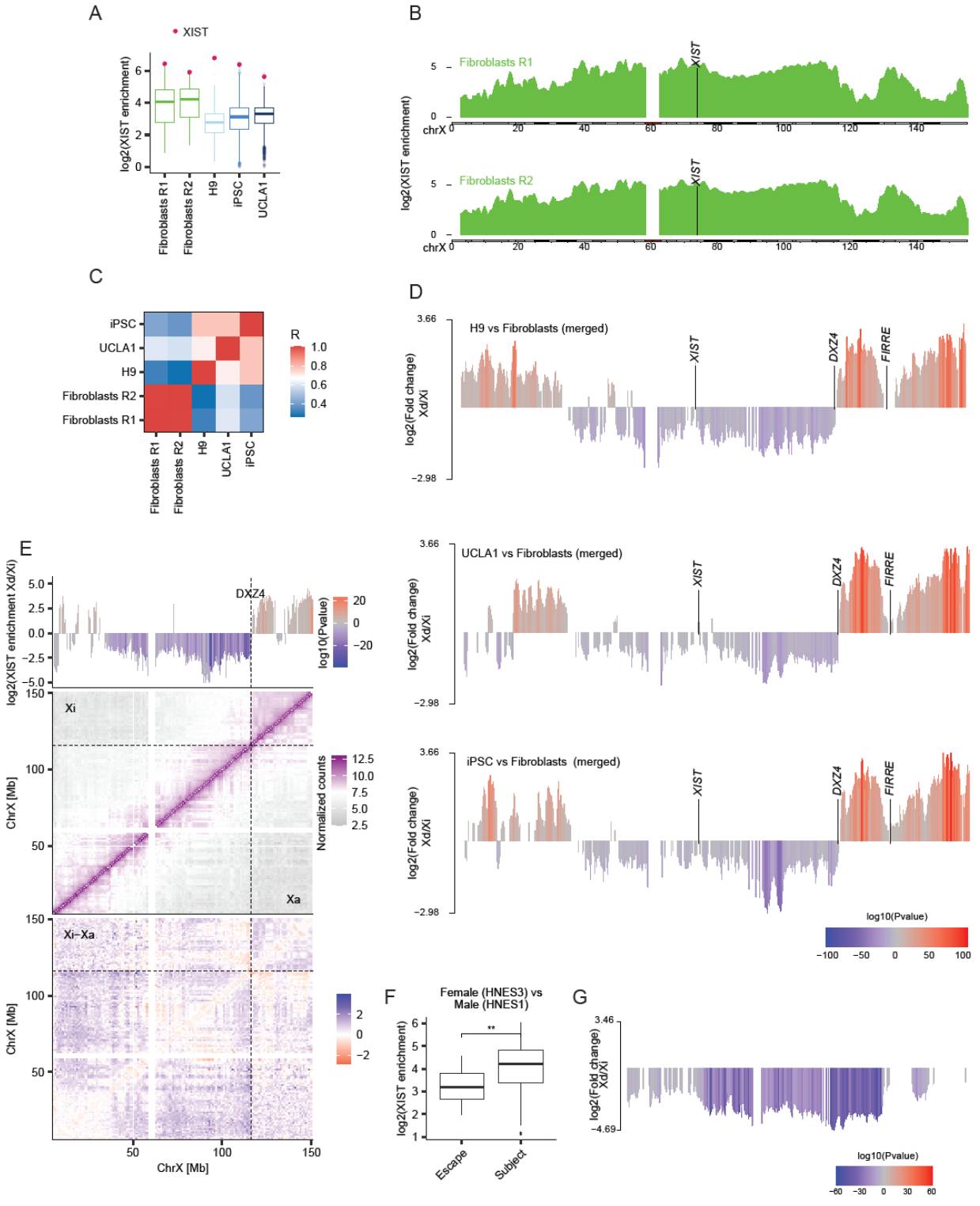
Click here to access/download;Supplemental Figure;Fig. S3 Dror et al.pdf



Supp Figure 4

# Figure S4

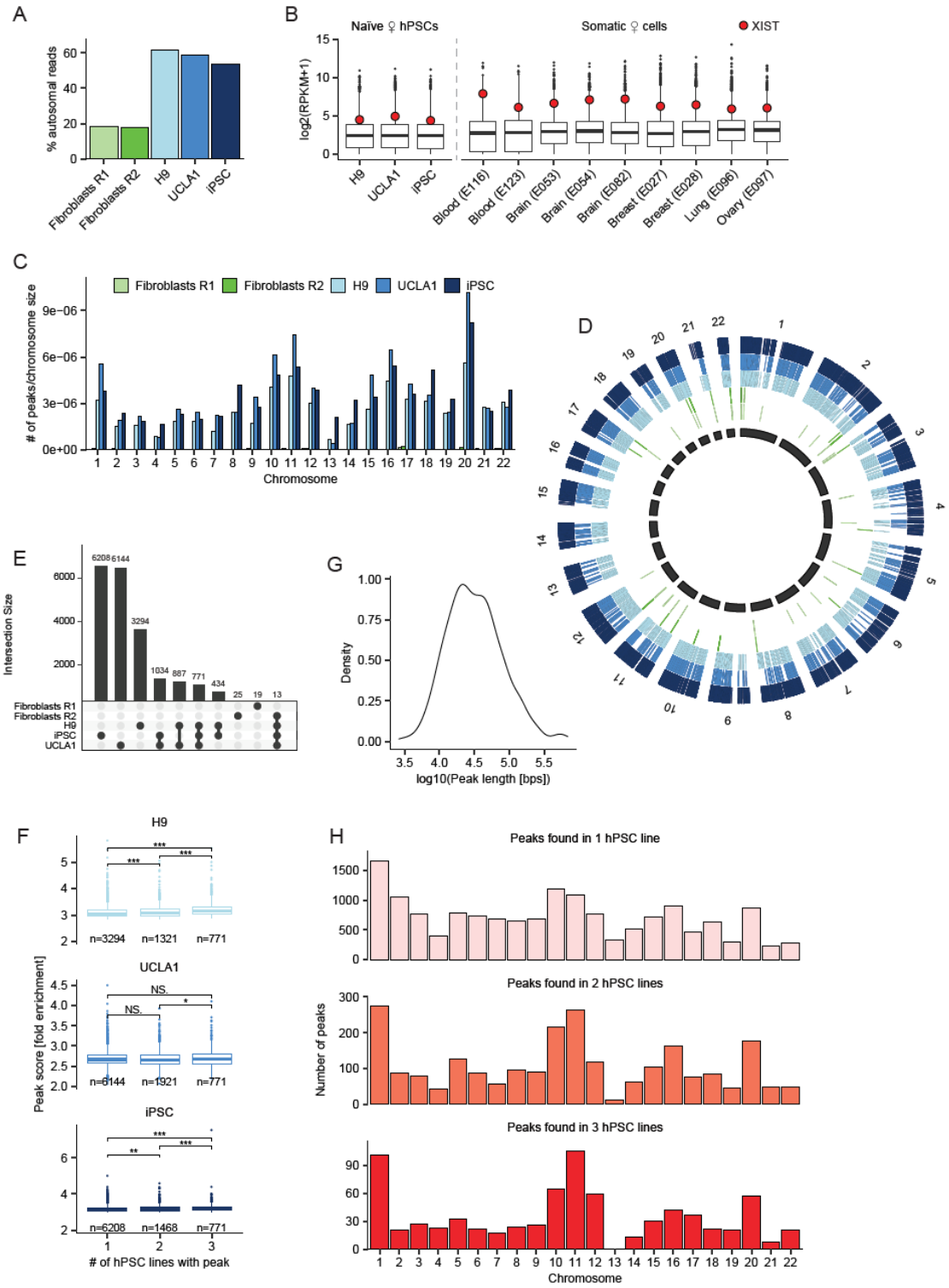
Click here to access/download;Supplemental Figure;Fig. S4 Dror et al.pdf



Supp Figure 5

# Figure S5 Dror et al.

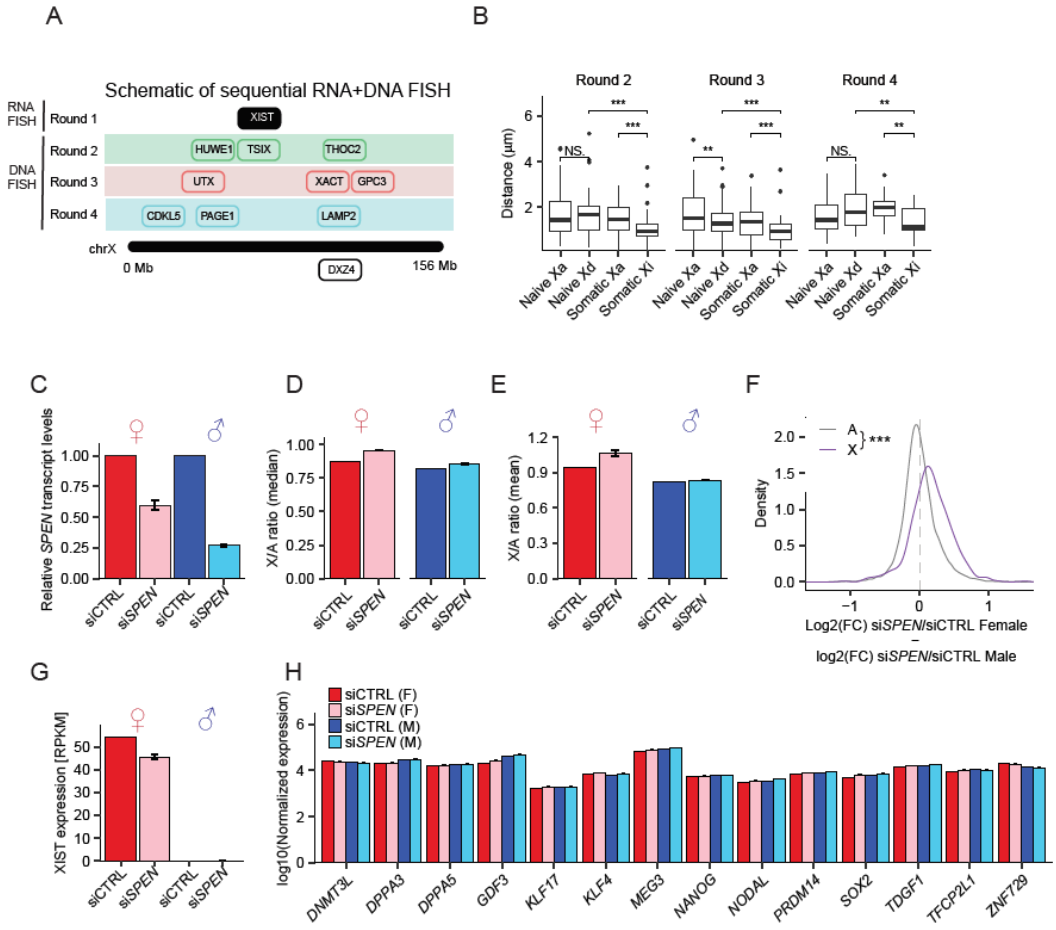
Click here to access/download;Supplemental Figure;Fig. S5 Dror et al.pdf



Supp Figure 6

# Figure S6 Dror et al.

Click here to access/download;Supplemental Figure;Fig. S6 Dror et al.pdf



# Chapter 4

Research Project

## Characterizing the genome-wide localization of the *Xist* lncRNA and its effect on gene silencing

Shawn Y.X. Tan[#1,2], Iris Dror[#1] and Kathrin Plath[*1]

1. Department of Biological Chemistry, Molecular Biology Institute, Jonsson Comprehensive Cancer Center, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA 90095, USA

2. Molecular Biology Interdepartmental Doctoral Program, University of California Los Angeles, Los Angeles, CA 90095, USA

# co-first authors

* correspondence: kplath@mednet.ucla.edu

**ABSTRACT**

Expression of the long non-coding RNA *Xist* and its spreading across the X-chromosome have been closely associated with X-linked dosage compensation, yet some studies have reported the presence of *Xist* expression without gene silencing. Additionally, the silencing ability of *Xist* is thought to be restricted within a narrow developmental window surrounding pluripotency, beyond which *Xist* can no longer induce transcriptional repression, except for a small subset of hematopoietic progenitors that can transiently re-establish X-chromosome inactivation (XCI). However, a recent study showed that *XIST* can trigger chromosome-wide silencing when incorporated into an autosome in differentiated neural cells. To address these conundrums, we employed genomics and imaging approaches, and demonstrated the lower enrichment of *Xist* across the X, which correlates with reduced X-linked gene silencing in early XCI, along with the novel spreading of *Xist* to autosomes and partial repression of autosomal target genes Surprisingly, we found that *Xist* is still able to induce partial silencing of several X-linked genes in differentiated cells, and we speculate that the lack of complete silencing is due to the distribution of *Xist* over the transcriptionally inactive regions of chromatin. Together, our results highlight that differentiated cells still retain epigenetic plasticity to respond to *Xist*, and illustrate how the localization of *Xist* may serve as a good predictor of its consequent gene regulation output on the X-chromosome.

## INTRODUCTION

X-chromosome inactivation (XCI) is a dosage compensation mechanism utilized by female placental mammals early in embryonic development to silence one of the two X-chromosomes, thereby achieving similar X-linked gene expression levels between XX females and XY males (Avner and Heard, 2001; Plath et al., 2002; Payer and Lee, 2008; Augui et al., 2011; Gendrel and Heard, 2014; Galupa and Heard, 2015). The X-inactive specific transcript (*Xist*) long noncoding RNA (lncRNA) is the master regulator of XCI that is transcribed from the X, and directs chromosome-wide gene repression by physically coating the future inactive X-chromosome (Xi) *in cis* (Clemson et al., 1996; Engreitz et al., 2013; Simon et al., 2013) to form an *Xist* cloud, as visualized by fluorescence *in situ* hybridization (FISH). *Xist* also recruits a myriad of protein complexes that contribute to the transcriptional shut-off of X-linked genes (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015).

In mouse, *Xist* first triggers XCI at the four- to eight-cell stage pre-implantation embryo (Wang et al., 2016; Borensztein et al., 2017). In support of the role of *Xist* within a pluripotency context, *Xist* appears to lose its ability to trigger silencing in mouse embryonic stem cells (ESCs) that are differentiated past 48 hours (Wutz and Jaenisch, 2000). On the contrary, one recent study showed that an *XIST* transgene incorporated into one of three copies of chromosome 21 (chr21) in differentiated neural stem cells derived from trisomic Down syndrome patients can induce efficient chromosome-wide gene repression (Czerminski and Lawrence, 2020). Thus, it is of fundamental interest to

decipher the developmental and epigenetic context beyond pluripotency in which *Xist* can still exert its repressive function, so as to better harness its silencing ability for potential translational applications, as well as understanding the mechanisms by which *Xist* execute XCI.

*Xist* upregulation and localization across the X-chromosome had always been associated with X-linked gene silencing (Avner and Heard, 2001; Plath et al., 2002). Yet, in human, both the X-chromosomes in female pre-implantation embryos and in naïve ESCs remain transcriptionally active despite *XIST* coating both Xs (Okamoto et al., 2011; Petropoulos et al., 2016; Dror et al., in review). Similarly, in the mouse, recent imaging data from our lab also highlighted the lack of X-linked gene silencing with the presence of biallelic transcripts for two X-linked genes, *Atrx* and *Mecp2*, which are normally subjected to silencing, in early differentiating cells that displayed *Xist* expression and cloud formation (Markaki et al., in review).

Therefore, in this study we set out to investigate the mechanism underlying the decoupling of *Xist* expression from transcriptional silencing,. By employing genomics approaches, we uncovered the differential *Xist* spreading phenotypes in female differentiating mouse ESCs at different stages of development, with reduced accumulation of *Xist* across the Xi in early differentiating ESCs compared to their late differentiated counterparts. This correlates with the attenuation of X-linked dosage compensation in early differentiating cells, in contrast to complete XCI in late

differentiated cells, suggesting that the failure of *Xist* to implement silencing in early differentiating ESCs may represent a transient state on the way to XCI.

Additionally, we identified a novel localization of *Xist* to autosomes, which we recently observed in human naïve ESCs (Dror et al., in review), but was never previously reported in the mouse system. The propagation of *Xist* to autosomes can also be recapitulated in various mouse cell types, including in fully differentiated male mouse embryonic fibroblasts (MEFs) carrying an inducible *Xist* transgene on the Xa (tetO-*Xist* MEFs). Furthermore, *Xist* is able to transcriptionally downregulate the autosomal genes that it targets, suggesting a potential physiological relevance in the female-specific autosomal gene regulation during early embryonic development.

## RESULTS

### X-linked dosage compensation is reduced in female early differentiated EpiLCs

At the onset of XCI, the upregulation of *Xist* is accompanied by its physical spreading across the entire X-chromosome to form a cloud, a process that can be recapitulated *in vitro* by differentiating female mouse ESCs to epiblast-like cells (EpiLCs) (Guyochin et al., 2014). *Xist* expression and cloud formation had typically been closely tied to X-linked gene silencing (Avner and Heard, 2001; Plath et al., 2002). Surprisingly, we discovered from our bulk RNA-seq data that early Day 2 differentiated EpiLCs exhibited deficient X-linked dosage compensation, as evidenced by the higher ratio of read counts aligned to X-linked versus autosomal genes (X/A ratio), compared to their late Day 4 differentiated counterparts (Figure 1a). This higher X/A ratio closely matches that in undifferentiated ESCs that do not undergo XCI (Figure 1a), despite having similar levels of *Xist* expression and the presence of *Xist* clouds in both differentiation timepoints (Figures 1b,c).

When we defined a strict cutoff of X-linked genes that were significantly downregulated (log2 fold change < -1 and p-value < 0.05), we identified 25 such genes in Day 2 EpiLCs, all of which were further repressed by Day 4 of differentiation except for Gpc4, which is likely an outlier as its expression is higher than that in undifferentiated mESCs with both X-chromosomes still active (Figure 1d). Thus, this implies that *Xist*-mediated silencing is incomplete during early mESC differentiation, but continues progressively over development before full-fledged XCI is achieved.

Closer inspection of the *Xist* cloud morphology revealed a greater extent of its nuclear dispersal in Day 2 EpiLCs, in which *Xist* appeared to spread outside of the inactive Xi territory, as observed by *Xist* RNA FISH, as opposed to the conventional compact *Xist* cloud that is well characterized for complete XCI in late Day 4 differentiated EpiLCs (Figure 1c) and in female somatic cells like MEFs.

To elucidate the genome-wide localization pattern of *Xist*, we performed RNA Antisense Purification, followed by high-throughput DNA sequencing (RAP-seq), a technique that utilizes antisense biotinylated probes targeting *Xist* RNA to map its interaction with chromatin at high resolution (Engreitz et al., 2013), and found that *Xist* distributes broadly across the entire length of the X-chromosome (Figure 1e), with a considerable overlap of *Xist* peaks in both Day 2 and Day 4 EpiLCs (Figure 1f). These *Xist* peaks show a greater amount of enrichment in Day 4 compared to Day 2 EpiLCs (Figure 1g). When we plotted *Xist* enrichment against X-linked genes that were significantly downregulated in Day 2 and Day 4 EpiLCs compared to mESCs, defined as log2 fold change<-0.5 and p-value<0.05, or not significantly downregulated, we found that genes that were downregulated in Day 4, but not in Day 2 EpiLCs, had higher levels of *Xist* binding (Figure 1h), implying that X-linked genes are hindered from silencing despite *Xist* enrichment in Day 2 EpiLCs.

### *Xist* localizes beyond the X-chromosome territory in female EpiLCs

Given the similar amount of *Xist* expression at both timepoints of differentiation (Figure 1b), but less *Xist* binding on the X in Day 2 EpiLCs (Figure 1g), we posited that *Xist* likely

localizes beyond the X-chromosome territory. Indeed, close examination of the RAP-seq data uncovered the unique spreading of *Xist* to autosomes (Figures 2a,b,c), with more abundant autosomal *Xist* binding in Day 2 than in Day 4 EpiLCs (Figure 2c). This concurs with our imaging data that illustrate the dispersed *Xist* cloud phenotype in Day 2 EpiLCs becoming condensed into a compact cloud over the Xi by Day 4 of differentiation (Figure 1c). The spreading of *Xist* to autosomes is an unexpected discovery, as *Xist* localization has always been known to be restricted *in cis* to the X-chromosome territory (Jonkers et al., 2008). Nonetheless, these results closely mirror recent findings from our lab in naïve human ESCs, in which the lower enrichment of *XIST* across the dampened X-chromosome, which displays incomplete XCI with partial repression of X-linked genes, is accompanied by *XIST* spreading to specific autosomal regions (Dror et al., in review).

### *Xist* enriches over the A-compartment and represses autosomal genes in female EpiLCs

To gain further insight into this novel *Xist* localization on chromosomes *in trans*, we called PC1 scores using published mouse ESC Hi-C datasets (Dixon et al., 2012), and found *Xist* accumulation mainly over autosomal domains enriched for the open, transcriptionally active A-compartment (Figure 2d), especially over autosomal regions that harbor active histone modifications, such as H3K4me2 and H3K9Ac, but not repressive chromatin domains marked by H3K9me3 (Figure 2e). Next, we called peaks across the autosomal domains where *Xist* binds, identified genes residing within and outside these peaks and compared their expression levels, and found that autosomal genes under *Xist* peaks tend to be significantly more lowly expressed than those not targeted by *Xist* (Figure 2f).

To interrogate the transcriptional impact of *Xist* spreading on autosomes, we analyzed our bulk RNA-seq data and found that autosomal genes under *Xist* peaks are significantly downregulated in Day 2, but not in Day 4 EpiLCs (Figure 2g), likely because *Xist* binding across all autosomes is reduced from the Day 2 to Day 4 transition (Figure 2c). Furthermore, upon intersecting our data with the *in vivo* single-cell RNA-seq analysis from E5.5 mouse conceptuses (Mohammed et al., 2017), we uncovered a significant correlation between autosomal genes under *Xist* peaks and X-reactivation (Figure 2h), lending support to our *in vitro* data that *Xist* represses the autosomal genes where it spreads to. Finally, Gene Ontology analysis demonstrated that many of the autosomal genes under *Xist* peaks in early EpiLCs are associated with embryonic development, including genes like Bmp4 and Ezh2 involved in epidermis development, and Gsk3b and Notch1 implicated in Wnt signaling (Figure 2i), thereby hinting at a putative biological role for the propagation of *Xist* to autosomes. In a similar vein, *XIST*-targeted autosomal genes in naïve human ESCs are significantly enriched for differentiation and developmental gene ontology terms (Dror et al., in review).

Collectively, these results highlight, for the first time in the mouse system, that *Xist* localization is not confined to the Xi in the early phase of development, as evidenced by the remarkable *trans*-spreading behavior of *Xist* to autosomes, and that majority of these *Xist*-enriched autosomal genes are down-regulated, albeit to a smaller extent compared to *Xist*-mediated transcriptional shut-off of X-linked genes.

### *Xist* localizes to autosomes in male *Xist*-inducible MEFs

The female mouse ESC to EpiLC differentiation scheme offers a unique platform to investigate the link between the differential spreading of *Xist* and its consequent gene regulation output during the course of XCI. However, EpiLCs represent a transient developmental entity that cannot be easily isolated (Hayashi et al., 2011). Thus, we turned to an alternative tractable system that enables stable culture *in vitro*, while also displaying dispersed *Xist* spreading characteristics in mouse differentiated cells, in which *Xist* expression is known to not induce XCI – the male *Xist*-inducible MEFs. These cells may offer insights into how *Xist* localization correlates with gene expression dynamics.

Male cells contain only one X-chromosome and will therefore not have endogenous *Xist* expression. In order to induce *Xist*, we used a MEF line that harbors an inducible *Xist* allele, in which its endogenous promoter is replaced with a tetracycline-responsive promoter, and an M2rtTA trans-activator in its Rosa26 locus, such that doxycycline (Dox) treatment for just 6 hours is sufficient for *Xist* expression and concomitant cloud formation in the majority of the cells, and these clouds exhibit a larger extent of nuclear dispersal with increasing duration of Dox treatment (Figures 3a,b). In fact, the *Xist* cloud was so extensively dispersed, especially for longer periods of *Xist* induction, that it sometimes covered more than half the nuclear volume (Figure 3a), an unusual phenotype that was never observed previously in any *Xist*-expressing cell. This led us to speculate that *Xist* spreads beyond the X and onto autosomes in these MEFs. Indeed, subsequent RAP-seq analysis performed at different timepoints of Dox treatment

(1, 6, 24 and 48 hours) revealed not only the broad coating of *Xist* across the entire X-chromosome, but also on autosomes, by 6 hours of *Xist* expression (Figures 3c,d).

### *Xist* accumulates over the B-compartment and partially represses autosomal genes in male *Xist*-inducible MEFs

Intriguingly, *Xist* enriches over the closed and transcriptionally inert B-compartment of autosomes in these male *Xist*-inducible MEFs (Figure 3e), where compact and gene-poor domains of chromatin preferentially reside, in contrast to its accumulation over the open and transcriptionally permissive A-compartment in female EpiLCs (Figure 2a). However, despite the distinct difference in the autosomal binding patterns of *Xist*, as observed by a minimal overlap of all autosomal *Xist* peaks between both cell types (Figures 3f,g), *Xist* still tends to localize across lowly-expressed autosomal genes in these *Xist*-inducible MEFs (Figures 3c,h), and also mediates the small but significant repression of autosomal genes under *Xist* peaks (Figure 3i), similar to that observed in female EpiLCs (Figure 2g) and in human naïve ESCs (Dror et al., in review).

### SPEN mediates the partial repression of X-linked genes in male *Xist*-inducible MEFs

Even though *Xist* physically coats the entire X-chromosome in male *Xist*-inducible MEFs, X-linked gene silencing appears to be impeded, as demonstrated by the persistent presence of an *Atrx* RNA FISH signal within the *Xist* cloud, regardless of the cloud morphology (Figure 4a). However, when we performed bulk RNA-seq on these MEFs with *Xist* inducibly expressed for up to two weeks, we noticed a partial down-regulation of

X-linked genes by 6 hours of *Xist* expression, compared to control MEFs without *Xist* induction, and the magnitude of this partial repression increases slightly with the duration of *Xist* expression (Figure 4b). Specifically, about 8% of X-linked genes were downregulated within 6-24h of *Xist* induction, which is elevated to roughly 20% of X-linked genes repressed after 1-2 weeks of *Xist* expression (Figure 4c). Importantly, this repression can be abrogated by siRNA-mediated knockdown of SPEN, also known as SMRT- and HDAC-associated repressor protein SHARP (Figure 4d), a key transcriptional repressor that is recruited by and binds directly to *Xist* to initiate X-linked gene silencing (McHugh et al., 2015; Dossin et al., 2020). Thus, the same protein factor SPEN that mediates efficient X-linked gene silencing in differentiating female EpiLCs and male *Xist*-inducible ESCs is also responsible for the partial down-regulation of X-linked genes in these male *Xist*-inducible MEFs.

**Several Xi markers are enriched with *Xist* in male *Xist*-inducible MEFs**

In spite of the mild silencing phenotype in these *Xist*-inducible MEFs, several classic Xi markers that are known to be typical hallmarks of XCI are found to co-localize with the *Xist* cloud within as early as 4 hours of *Xist* expression (Figure 4e,f). These include the protein Cip1-interacting zinc finger protein 1 (CIZ1), which binds directly to *Xist* to enforce its localization on the Xi (Ridings-Figueroa et al., 2017; Sunwoo et al., 2017), the Polycomb repressive complex 1 (PRC1)-dependent histone modification H2AK119ub that is associated with early XCI (Brockdorff, 2017; Zylicz et al., 2019), and RNA Polymerase II (RNAPII) exclusion, which marks the creation of a silent nuclear compartment at the onset of XCI (Chaumeil et al., 2006). However, the canonical heterochromatin mark

H3K27me3 that is known to accumulate on the Xi (Plath et al., 2003; Brockdorff, 2017) did not enrich with the *Xist* cloud in these *Xist*-inducible MEFs, a finding that was nonetheless supported by a previous study demonstrating the failure of H3K27me3 deposition after *Xist* induction beyond a narrow window in ESC differentiation, which is further discussed below (Kohlmaier et al., 2004). Thus, the XCI machinery that permits X-linked gene silencing can be established, but precludes robust and complete chromosome-wide repression in these male *Xist*-inducible MEFs.

### ***Xist*-mediated silencing is gradually diminished in increasingly differentiated male *Xist*-inducible ESCs**

Our discovery that several X-linked genes can be partially repressed in a fully differentiated cell line like MEFs was not expected, as a critical developmental window between pluripotency and 48 hours of differentiation had previously been defined for *Xist* to effectively discharge its gene silencing role, but this was performed in the context of an *Xist* cDNA transgene integrated within an autosome in male mouse ESCs, with silencing assessed using only two genes (Wutz and Jaenisch, 2000). To quantitatively examine the extent of chromosome-wide silencing in its endogenous context, we harnessed the Tet-On *Xist* male pSM33 mouse ESC line carrying a tetracycline-inducible promoter in place of the endogenous *Xist* promoter, with the M2rtTA trans-activator in its Rosa26 locus. *Xist* expression is triggered from the single X-chromosome with 24 hours of Dox treatment at different timepoints of all-*trans* retinoic acid-induced differentiation, and RNA from both Dox-treated and untreated samples for each day of differentiation up to Day 12 is harvested for assembling RNA-seq libraries (Figure 5a).

Bulk RNA-seq analysis revealed that X-linked gene silencing occurred efficiently up to 48 hours of differentiation, but became gradually compromised over time, as seen by the positive correlation of the X/A ratio and duration of differentiation (Figure 5b), despite comparable levels of *Xist* expression across all timepoints (Figure 5c). Silencing was completely abrogated by Day 9 of differentiation (Figure 5b). Hence, *Xist* can only effectively implement silencing within a narrow 48-hour timeframe surrounding pluripotency, beyond which its repressive ability becomes severely attenuated.

### ***Xist*** **localization differs between early and late differentiated male** ***Xist*****-inducible ESCs**

To understand how the differential X-linked gene silencing dynamics may be influenced by changes in the localization of *Xist*, we performed RAP-seq on Day 2 and Day 6-differentiated pSM33 ESCs that inducibly expressed *Xist* for 24 hours, exhibiting efficient and deficient X-linked gene silencing respectively, and found similar amount of *Xist* binding in both conditions (Figure 5d), with *Xist* spreading across the entire length of the X-chromosome (Figure 5e), despite the distinct difference in the output of their gene silencing capacity. Nonetheless, there are some differences in the localization pattern of *Xist* between Day 2 and Day 6-differentiated cells (Figures 5e,f). Notably, *Xist* is more enriched within the vicinity of its transcription locus, as well as the *Dxz4* locus, in Day 2 than in Day 6-differentiated cells (Figure 5f). The *Dxz4* locus is of particular interest in the XCI field, as this conserved macrosatellite repeat locus has been described to separate

the Xi into two distinct mega-domains in both mouse and human female cells (Deng et al., 2015; Giorgetti et al., 2016; Bonora et al., 2018).

Furthermore, we characterized the spreading of *Xist* to autosomes (Figure 6a), with minimal amount of overlap in shared *Xist* peaks between Day 2 and Day 6-differentiated cells (Figure 6b). Indeed, we found that *Xist* spreads preferentially to the transcriptionally permissive A-compartment on the X and autosomes in Day 2-differentiated cells, as opposed to its enrichment over the transcriptionally repressive B-compartment on the X and autosomes in the Day 6-differentiated counterparts, as evidenced by the positive and negative correlations with the PC score, calculated based on Hi-C data in MEFs, respectively (Figures 6c,d). Consistent with this observation, the *Xist* localization patterns across both the X and autosomes were more highly correlated between the Day 6-differentiated ESCs and male *Xist*-inducible MEF datasets, than between the Day 2 and Day 6-differentiated ESCs of the same cell type (Figures 6e,f).

Taken together, these data suggest that the differential genome-wide distribution of *Xist* may account for its inability to mediate X-linked gene silencing in late differentiated ESCs beyond the critical developmental window. Coincidentally, these results also draw important parallels between the ability of *Xist* to execute X-linked gene silencing in female EpiLCs and male *Xist*-inducible Day 2-differentiated ESCs, where *Xist* is found to localize to the transcriptionally active A-compartment, versus impaired chromosome-wide silencing in male *Xist*-inducible Day 6-differentiated ESCs and MEFs, where *Xist* preferentially accumulates over the transcriptionally inactive B-compartment.

## DISCUSSION

*Xist* upregulation and localization across the X-chromosome *in cis* have always been associated with X-linked gene silencing, but in this present study, we demonstrate that high levels of *Xist* expression alone do not automatically confer efficient gene silencing. By exploiting our extensive RAP-seq and RNA-seq datasets gathered from various mouse cell types that model different stages of development, we also made the unprecedented discovery that *Xist* can not only propagate to chromosomes *in trans*, but also mediate the repression of autosomal genes that it targets.

The finding that *Xist* can localize to autosomes was not anticipated, as the current long-held view is that *Xist* spreading is limited to the X-chromosome from which it is transcribed in female mouse cells (Jonkers et al., 2008). Several studies have previously reported that the ectopic expression of *Xist/XIST* from an autosome (Lee and Jaenisch, 1997; Wutz and Jaenisch, 2000; Hall et al., 2002a; Kelsey et al., 2015), or its localization onto autosomal regions in the context of X to autosome (X;A) translocations (Hall et al., 2002b; Popova et al., 2006; Yang et al., 2011; Loda et al., 2017), can enact autosomal gene silencing in both mouse and human cells. In fact, this has heralded the exciting possibility that *Xist*/*XIST* can be exploited for potential therapeutic purposes, as one seminal study demonstrated the remarkable ability of *XIST* to induce chromosome-wide silencing of one of three copies of chr21 where it is integrated in Down syndrome patient-derived induced pluripotent stem cells (iPSCs), thereby aligning overall gene expression levels to that in normal individuals with two copies of chr21 (Jiang et al., 2013). However,

to our knowledge, *Xist* accumulation on autosomes and its subsequent autosomal gene regulation by endogenous *Xist* have never been reported.

Although *Xist* can influence autosomal gene regulation, the physiological relevance of this ability remains enigmatic. Nevertheless, our data raise the interesting possibility that autosomal genes targeted by *Xist* may be involved in the transcriptional regulation of differentiation and embryonic development, based on the enrichment of these gene ontology terms in early female EpiLCs (Figure 2i). Further studies are required to elucidate the biological significance of the female-specific targeting of these autosomal genes, as well as its implication for the possible differential autosomal gene regulation between females and males.

In our female EpiLC differentiation model, lower levels of *Xist* binding on the X-chromosome (Figure 1g), together with higher levels of *Xist* binding on the autosomes (Figure 2c), were detected in Day 2 relative to Day 4 EpiLCs. This is consistent with our observation by FISH of a more dispersed *Xist* cloud, suggesting that *Xist* spreads beyond the X-chromosome, in Day 2 compared to Day 4 EpiLCs (Figure 1c). Combining these RAP-seq and imaging data with our RNA-seq analysis that showed reduced X-linked dosage compensation in Day 2 but not in Day 4 EpiLCs (Figure 1a), we propose a model in which the concentration of *Xist* sequestered on the X-chromosome is critical for determining the extent of X-linked gene silencing, similar to the phenomenon of X-chromosome dampening observed in human naïve ESCs (Dror et al., in review).

The lower amount of *Xist* enrichment on the Xi in Day 2 EpiLCs may lead to a decreased binding and subsequent recruitment of protein partners required for *Xist* to implement gene silencing. This effect is likely enhanced for the subset of proteins involved in condensate formation, whose function is dependent on their higher-order aggregation (Cerase et al., 2019; Pandya-Jones et al., 2020). It is also possible that *Xist* recruits different protein factors, and/or the proteins recruited harbor different post-translational modifications that modify their function, in early versus late differentiation, thereby modulating the ability of *Xist* to mediate the transcriptional repression of the X. Delineating the complete repertoire of protein interactors of *Xist*, along with their modifications across EpiLC differentiation, should define the differential silencing capacity of Xist, and may potentially shed light on the mechanism underlying the compaction of the *Xist* cloud within the X-chromosome territory during the maintenance phase of XCI.

In ascertaining the critical window for X-linked gene silencing (Wutz and Jaenisch, 2000) using male *Xist*-inducible ESCs, we found that *Xist* spreads broadly across the entire length of the X-chromosome in both early and late differentiated ESCs, which showed efficient and deficient gene silencing, respectively (Figure 5e). But upon closer inspection, we noted some differences in their *Xist* enrichment (Figure 5f), one of which is located near the macrosatellite repeat locus *Dxz4* that acts as a hinge region to demarcate the separation of the X into two large “super-domains” (Deng et al., 2015; Giorgetti et al., 2016; Bonora et al., 2018). This unique spatial organization of the X-chromosome is likely to contribute to the differential localization phenotype of *Xist*, as *Xist*

has been known to harness its intra-chromosomal proximity in three-dimensional (3D) space for its initial spread at the onset of XCI (Engreitz et al., 2013).

Beyond this restrictive developmental window surrounding pluripotency that permits *Xist*-mediated silencing, even cancer cells that re-activate their pluripotency program (Sell, 2004; Dreesen and Brivanlou, 2007) can no longer induce silencing, except for a small minority of cancer cells that can partially downregulate a few genes close to the *XIST* locus (Hall et al., 2002a; Chow et al., 2007; Minks et al., 2013). In another study that utilized male *Xist* transgenic mice in which silencing of the single X-chromosome resulted in cell death, the authors noted that only a specific group of hematopoietic progenitor cells could transiently re-establish XCI (Savarese et al., 2006), further supporting the prevailing view that *Xist* is unable to enforce silencing in differentiated/somatic cells.

Intriguingly, we discovered that in fully differentiated male fibroblasts that inducibly expressed *Xist*, *Xist* is able to mediate partial repression of X-linked genes (Figure 4b,c). This surprising result was nonetheless supported by a recent publication showing the remarkable ability of *XIST* to induce substantial dosage compensation of trisomy 21 in Down syndrome patient-derived iPSC-differentiated neural cells that contain a Dox-responsive *XIST* transgene on one of its three chr21 (Czerminski and Lawrence, 2020). However, this study reported that ectopic *XIST* was able to exert efficient silencing within 1-2 weeks, with negligible escape from silencing across the entire chromosome, whereas

transcriptional repression of the X in our *Xist*-inducible MEFs remained incomplete even after 2 weeks of *Xist* expression.

In conclusion, even though differentiated cells extend beyond the supposed strict developmental window for *Xist*-mediated silencing, they still seem to retain considerable epigenetic plasticity to respond to *Xist*, albeit to varying degrees that may be dependent on multiple parameters, such as the specific developmental stage, *Xist* RNA abundance, and differences in RNA sequence between mouse and human *Xist/XIST* (Nesterova et al., 2001) that can affect the type and amount of proteins recruited by *Xist/XIST* to execute silencing. Understanding the mechanistic basis and interplay of factors required by *Xist/XIST* to effectively induce chromosome-wide silencing beyond pluripotency in different cell types will elevate the therapeutic value of this unique lncRNA in ameliorating the adverse phenotypic effects of trisomic disorders, and potentially other chromosomal duplications and abnormalities (Theisen and Shaffer, 2010).

## FIGURE LEGENDS

**Figure 1: Reduced X-linked gene silencing in early female differentiating EpiLCs**

A) Ratio of read counts aligned to X-chromosomal versus autosomal genes (X/A ratio) in mouse embryonic stem cells (mESC), Day 2 (D2) and Day 4 (D4) EpiLCs, based on bulk RNA-seq data. Error bars correspond to the standard deviation of two replicates.

B) Log10 normalized *Xist* expression levels in mESC, D2 and D4 EpiLCs, based on bulk RNA-seq data. Error bars correspond to the standard deviation of two replicates.

C) Representative RNA FISH images of *Xist* (green), which marks the inactive X-chromosome, and *Tsix* (magenta), which marks the active X-chromosome, in Day 2 and Day 4 EpiLCs.

D) Heatmap of the scaled expression levels comparing across mESCs, D2 and D4 EpiLCs of the 25 X-linked genes that were significantly downregulated, defined as log2 fold change<-1 and p-value<0.05, by D2 EpiLCs.

E) *Xist* enrichment, defined as the ratio of *Xist* pulldown over its input, across the X-chromosome for D2 and D4 EpiLCs. Green and blue rectangles beneath the *Xist* enrichment plots represent *Xist* enriched peaks. Unmappable regions are masked in white. The genomic *Xist* locus is marked.

F) Venn diagram depicting the overlap of *Xist* RAP-seq peaks on the X-chromosome in D2 and D4 EpiLCs.

G) Average *Xist* enrichment in D2 and D4 EpiLCs within ±2.5Mb flanking the *Xist* summits, which refer to the highest point called from each *Xist* peak, for D2 EpiLCs.

H) Boxplots of *Xist* enrichment by RAP-seq against X-linked genes that are significantly downregulated (Down), defined as log2 fold change<-0.5 and p-value<0.05, or not significantly downregulated (NS) by bulk RNA-seq, comparing D2 EpiLCs with mESC (left), and D4 EpiLCs with mESC (right). Wilcoxon test p-values: NS≥0.05, ** P<0.01. The number of X-linked genes (n) in each category is provided.

**Figure 2: *Xist* spreads to the A-compartment of autosomes and mediates partial repression of autosomal target genes in female EpiLCs**

A) *Xist* enrichment, defined as the ratio of *Xist* pulldown over its input, across chromosome 16, a representative autosome, for D2 (green) and D4 (blue) EpiLCs.

B) *Xist* enrichment by RAP-seq aligned to each chromosome for D2 (light blue) and D4 (dark blue) EpiLCs.

C) Zoom in of (B), excluding *Xist* enrichment on the X-chromosome.

D) Boxplots of the PC1 scores of regions enriched or not enriched in *Xist* on autosomes in D2 EpiLCs. Red dotted line indicates PC1 score=0. Negative PC1 score represents the A-compartment while positive PC1 score represents the B-compartment. Wilcoxon test p-value: *** P<0.001.

E) Bar graphs showing the mouse ESC histone modifications that correlate with the autosomal regions targeted by *Xist* localization. R represents the Pearson correlation score between *Xist* enrichment in D2 EpiLCs by RAP-seq and the respective histone marks.

F) Boxplots of gene expression levels in ESC, D2 and D4 EpiLCs grouped based on the presence or absence of *Xist* peaks that bind across autosomal genes by RAP-seq.

Wilcoxon test p-values: NS≥0.05, * P<0.05. The number of genes (n) in each group is given.

G) Boxplots of autosomal gene expression fold change comparing D2 EpiLCs with no silencing in ESC (left), and D4 EpiLCs with ESC (right). Autosomal genes are categorized based on the presence or absence of *Xist* peaks in each comparison. Wilcoxon test p-values: NS≥0.05, ** P<0.01. The number of genes (n) in each category is provided.

H) Line graphs indicating the spearman correlation between genes and X-reactivation of autosomal genes in embryonic Day 5.5 female mouse conceptuses when both X-chromosomes are still active, grouped based on whether the gene is under *Xist* peak (blue) or not (grey). Wilcoxon test p-value: *** P<0.001.

I) Gene ontology terms (-log10(p-value)) for autosomal genes targeted by *Xist* in D2 EpiLCs.

**Figure 3: *Xist* spreads to the B-compartment and mediates partial repression of autosomal target genes in male *Xist*-inducible MEFs**

A) Representative RNA FISH images of *Xist* RNA depicting three different classes of *Xist* clouds in male *Xist*-inducible MEFs.

B) Bar graphs showing the proportion of the classes of *Xist* clouds in (A) formed in male *Xist*-inducible MEFs that were treated with Dox for 1h, 6h, 24h and 48h.

C) *Xist* enrichment, defined as the ratio of *Xist* pulldown over its input, across a selected region of chromosome 7 (shown by red box), a representative autosome, for male *Xist*-inducible MEFs treated with Dox for 1h, 6h, 24h and 48h. RNA-seq reads from

control MEFs without Dox treatment for this corresponding region are shown in pink. Green boxes represent regions with low to no gene expression overlapping with higher *Xist* enrichment by RAP-seq.

D) *Xist* enrichment by RAP-seq aligned to each chromosome for male *Xist*-inducible MEFs treated with Dox for 1h, 6h, 24h and 48h (top). Zoom in of *Xist* enrichment for autosomes, excluding *Xist* enrichment on the X-chromosome (bottom).

E) Boxplots of the PC1 scores of regions enriched or not enriched in *Xist* on autosomes in male *Xist*-inducible MEFs treated with Dox for 48h. Red dotted line indicates PC1 score=0. Negative PC1 score represents the A-compartment while positive PC1 score represents the B-compartment. Wilcoxon test p-value: *** P<0.001.

F) Read counts of *Xist* enrichment by RAP-seq across chromosome 6, a representative autosome, for male *Xist*-inducible MEFs treated with Dox for 48h (blue) and female D2 EpiLCs (red).

G) Venn diagram of the overlap of *Xist* RAP-seq peaks on autosomes in male *Xist*-inducible MEFs treated with Dox for 48h and female D2 EpiLCs.

H) Boxplots of gene expression levels grouped based on the presence or absence of *Xist* peaks that bind across autosomal genes by RAP-seq in male *Xist*-inducible MEFs treated with Dox for 48h. Wilcoxon test p-value: *** P<0.001.

I) Boxplots of autosomal gene expression fold change in male *Xist*-inducible MEFs treated with Dox for 48h. Autosomal genes are categorized based on the presence or absence of *Xist* peaks. Wilcoxon test p-value: *** P<0.001. The number of genes (n) in each category is given.

**Figure 4: Partial downregulation of X-linked genes and enrichment of canonical XCI markers in male *Xist*-inducible MEFs**

A) Representative RNA FISH images of *Xist* (red) and an X-linked gene that is normally subjected to silencing, *Atrx* (green), in male *Xist*-inducible MEFs treated with Dox for 48h.

B) Ratio of read counts aligned to X-chromosomal versus autosomal genes (X/A ratio) in male *Xist*-inducible MEFs treated with Dox for 6h, 24h, 1 week and 2 weeks, based on bulk RNA-seq data. Control cells without Dox treatment are shown for comparison. Error bars correspond to the standard deviation of two replicates per time point.

C) Cumulative distribution plots of the percentage of genes downregulated on autosomes (left) and X-chromosome (right) for Dox-induced *Xist* expression for 6h (red), 24h (green), 1 week (blue) and 2 weeks (purple), compared to their respective untreated (no Dox) controls. The solid vertical line indicates a log2 fold change of zero, while the dashed line marks a log2 fold change of 1, which implies a 50% reduction in gene expression.

D) Ratio of read counts aligned to X-chromosomal versus autosomal genes (X/A ratio) in two biological replicates of male *Xist*-inducible MEFs treated with Dox for 48h without siSPEN (red bars on left), and with siSPEN (green bars on right), based on bulk RNA-seq data. Control cells without Dox (blue bars on left) and without siSPEN (green bars on right) treatment are shown for comparison.

E) Representative RNA FISH images of *Xist* (green) combined with immunofluorescence for three canonical XCI markers (magenta) CIZ1, H2AK119ub and exclusion of RNA Polymerase II (RNAPII), in male *Xist*-inducible MEFs treated with Dox for 4h.

F) As in (E), for male *Xist*-inducible MEFs treated with Dox for 48h.

**Figure 5: *Xist*-mediated gene silencing is attenuated with increased differentiation in male *Xist*-inducible ESCs**

A) Schematic showing the timeline of bulk RNA-seq samples harvested for 24h Dox treatment (red arrows) and their respective controls without Dox treatment (blue boxes) across 12 days of differentiation in male *Xist*-inducible ESCs.

B) Ratio of read counts aligned to X-chromosomal versus autosomal genes (X/A ratio) in 24h Dox-treated (blue bars) and untreated (red bars) male *Xist*-inducible ESCs differentiated for up to 12 days, based on bulk RNA-seq data.

C) *Xist* expression level in 24h Dox-treated (blue bars) and untreated (red bars) male *Xist*-inducible ESCs differentiated for up to 12 days, based on bulk RNA-seq data.

D) *Xist* enrichment on the X-chromosome as a percentage of all *Xist* mapped reads by RAP-seq in input, Day 2 and Day 6 differentiated male *Xist*-inducible ESCs treated with Dox for 24h.

E) *Xist* enrichment, defined as the ratio of *Xist* pulldown over its input, across the X-chromosome for Day 2 (green) and Day 6 (blue) differentiated male *Xist*-inducible ESCs treated with Dox for 24h. Unmappable regions are masked in white. The genomic *Xist* locus is marked.

F) Differential enrichment of *Xist* in each 100Kb window (every 25Kb) of the entire X-chromosome between Day 2 and Day 6 differentiated male *Xist*-inducible ESCs treated with Dox for 24h. For this comparison, the *Xist* localization across the X was normalized to only reads that aligned to the X-chromosome, to highlight the differential

distribution of *Xist* on the X, with red color representing genomic regions in which *Xist* was more enriched for Day 2, and blue color representing genomic regions in which *Xist* was more enriched for Day 6 cells. The genomic loci of *Dxz4* and *Xist* are marked.

**Figure 6: *Xist* spreads to the A- and B-compartment for early and late differentiated male *Xist*-inducible ESCs respectively**

A) *Xist* enrichment as a percentage of total DNA sequencing reads aligned to each chromosome for 24h Dox-treated male *Xist*-inducible mESCs differentiated for 2 days (green) or 6 days (blue) (top). Zoom in of *Xist* enrichment for autosomes, excluding *Xist* enrichment on the X-chromosome (bottom).

B) Venn diagram of the overlap of *Xist* RAP-seq peaks on autosomes in 24h Dox-treated male *Xist*-inducible mESCs differentiated for 2 days or 6 days.

C) Scatterplots of the PC score (calculated based on Hi-C data in MEFs) versus *Xist* enrichment on the X-chromosome by RAP-seq for Day 2 (left) and Day 6 (right) differentiated male *Xist*-inducible ESCs treated with Dox for 24h. Pearson correlation (R) and p-value are given. Positive R value denotes higher *Xist* enrichment on the A-compartment while negative R value denotes higher enrichment on the B-compartment. Each data point represents a 500Kb window.

D) As in (C), for *Xist* enrichment on autosomes.

E) Heatmap depicting the Pearson correlation (R) between *Xist* enrichment within 100Kb windows (every 25Kb) across the X-chromosome in indicated cell lines.

F) As in (E), for *Xist* enrichment on autosomes.

## METHODS

### Cell culture

Female wild-type mouse F1 2-1 and male mouse pSM33 *Xist*-inducible embryonic stem cells (ESCs) were cultured on 0.5% gelatin-coated plates (porcine skin gelatin, Sigma) that were pre-seeded with irradiated DR4 feeders, which were obtained from Day 14.5 mouse embryos with appropriate animal protocols in place. These cultures were maintained in mouse ESC media containing knockout Dulbecco's Modified Eagle Medium (DMEM) (Life Technologies), 15% fetal bovine serum (FBS) (Omega), 2mM L-glutamine (Life Technologies), 1x MEM non-essential amino acids (NEAA) (Life Technologies), 0.1mM β-Mercaptoethanol (Sigma), 1x Penicillin/Streptomycin (Life Technologies) and 1000U/mL mouse leukemia inhibitory factor (LIF) (homemade). ESCs were maintained as small colonies, and passaged with trypsin (Life Technologies) and single-cell dissociation at 70-80% confluency. All cell lines used in this study were incubated at 37°C with 5% $CO_2$, and mycoplasma tests (Lonza) were carried out routinely to ensure the absence of such pathogens.

### Female EpiLC differentiation

Female wild-type F1 2-1 ESCs were first adjusted for three passages to feeder-free conditions in the presence of LIF and two inhibitors, 0.4μM PD0325901 and 3μM CHIR99021 (2i+LIF), and then cultured for at least three passages in serum-free 2i+LIF N2B27 media containing 1x N2 supplement (ThermoFisher), 1x B27 supplement (ThermoFisher), 2mM L-glutamine (Life Technologies), 1x NEAA (Life Technologies),

0.1mM β-Mercaptoethanol (Sigma) and 0.5x Penicillin/Streptomycin (Life Technologies), prior to differentiation to epiblast-like cells (EpiLCs). To induce differentiation, cells were dissociated with StemPro Accutase (Life Technologies) and seeded on Geltrex-coated plates (Geltrex LDEV-free Reduced Growth Factor Basement Membrane Matrix, Life Technologies) at a density of $2x10^5$ cells/ml in N2B27 media supplemented with 20ng/ml Activin A (Peprotech) and 12ng/ml basic Fibroblast Growth Factor (bFGF) (R&D Systems). Fresh media was exchanged daily until the cells were harvested for analysis.

**Male ESC differentiation**

Male mouse pSM33 *Xist*-inducible ESCs were dissociated to single cells with trypsin (Life Technologies) and counted. Cells were seeded in 2ml of MEF media (DMEM (Invitrogen) supplemented with 10% FBS (Omega), 2mM L-glutamine (Life Technologies), 1x NEAA (Life Technologies), 0.1mM β-Mercaptoethanol (Sigma) and 1x Penicillin/Streptomycin (Life Technologies)), at a density of $2x10^5$ cells per $4cm^2$ on gelatinized plates (porcine skin gelatin, Sigma). 24 hours after seeding, the culture media was changed and 1μM all-*trans* retinoic acid (Sigma) was added. Fresh media was exchanged daily until the cells were harvested for analysis. To induce *Xist* expression, 2μg/ml doxycycline (Dox) (Sigma) was added to the culture media for 24 hours.

**Male MEF culture**

Male *Xist*-inducible mouse embryonic fibroblasts (MEFs) were cultured in MEF media. MEFs were dissociated to single cells with trypsin (Life Technologies) and seeded in 2ml of MEF media at a density of $2x10^4$ cells per $4cm^2$ on tissue culture plates for RNA

collection or onto sterile 12mm glass coverslips for IF/FISH experiments. *Xist* expression was induced by the addition of 2µg/ml Dox (Sigma) for 6 hours to 2 weeks, depending on the experiment. For knockdown experiments, siRNAs were added upon cell plating (see section 'siRNA-mediated knockdown of SPEN').

**RNA FISH probe synthesis**

RNA FISH probes were labeled using homemade Nick Translation and fluorescent dUTPs as described in Cremer et al., 2008. *Xist* probe was made from a full-length mouse *Xist* cDNA plasmid (p15A-31-17.9kb *Xist*) as template, and *Atrx* probe was created using BAC RP23-265D6, which was purchased from CHORI-BACPAC.

**RNA FISH procedure**

Coverslips were rinsed twice with 1x PBS, and the cells were fixed with 3% paraformaldehyde (PFA) (Electron Microscopy Sciences) in 1x PBS for 10 min at room temperature (RT). After fixation, cells were permeabilized with 0.5% Triton X-100 (Acros) in 1x PBS containing 2mM vanadyl ribonucleoside complex (NEB) for 10-20 min at RT, before rinsing twice with 1x PBS and washing 3 times for 5 min each in 2x SSC (Ambion). Coverslips were then incubated for 1-3 hours at RT with 50% formamide (Fisher) in 2x SSC pH 7.2 solution before storing at 4°C until samples from all time points had been collected. Prior to probe hybridization, probes were denatured for 5 min at 95°C, coverslips were brought back to RT and dried, before incubating with probes for 24 hours at 37°C in a sealed humidified chamber. The next day, coverslips were washed 6 times in 2x SSC, followed by 3 times in 4x SSCT (SSC with 0.05% Tween-20) at 42°C for 15

min per step, protected from light. Finally, cells were post-fixed with 4% PFA in 1x PBS for 10 min at RT, stained with 1:10000 dilution of DAPI (0.5mg/ml) and mounted with Vectashield (Vector Labs) on glass slides, before sealing with Biotium Covergrip coverslip sealant (ThermoFisher). All procedures were performed and used reagents were disposed according to standard laboratory safety practices. Images were acquired using the LSM 880 Confocal Microscope (Zeiss), and Fiji (ImageJ) was used for image processing and analysis.

**Immunofluorescence staining**

Coverslips were rinsed twice with 1x PBS, and the cells were fixed in 2% PFA (Electron Microscopy Sciences) in 1x PBS for 10 min. After fixation, cells were permeabilized with 0.5% Triton X-100 (Acros) in 1x PBS for 10-15 min, before transferring to blocking buffer solution (1x PBST containing 2% bovine serum albumin (BSA) and 0.5% fish skin gelatin) for 1 hour. Cells were then incubated with primary antibody for 45 min, washed 3 times with 1x PBST, followed by incubating with secondary antibody for 30 min and washing 3 times with 1x PBST. Both antibody incubation steps were performed in a dark humidified chamber. Finally, cells were post-fixed with 4% PFA in 1x PBS for 10 min, stained with 1:10000 dilution of DAPI (0.5mg/ml) and mounted with Vectashield (Vector Labs) on glass slides, before sealing with Biotium Covergrip coverslip sealant (ThermoFisher). All procedures were performed at RT, and used reagents were disposed according to standard laboratory safety practices. Images were acquired using the LSM 880 Confocal Microscope (Zeiss), and Fiji (ImageJ) was used for image processing and analysis.

**Immunofluorescence staining combined with RNA FISH**

Immunofluorescence staining preceded RNA FISH whenever these two procedures are performed together.

**Antibodies and dilutions**

Endogenous CIZ1 was detected with a rabbit polyclonal antibody NB100-74624 (1:800 dilution, Novus Biologicals), histone H2AK119Ub with a rabbit monoclonal antibody 8240S (1:1000 dilution, Cell Signaling Technology), and RNA Polymerase II with a mouse monoclonal antibody 05-623 (1:1000 dilution, Millipore). The two secondary antibodies used are donkey anti-rabbit IgG CF568 antibody SAB4600076 (1:400 dilution, Sigma), and goat anti-mouse IgG CF568 antibody SAB4600309 (1:400 dilution, Sigma).

**siRNA-mediated knockdown of SPEN**

Silencer Select siRNAs (Ambion) against SPEN (s72337) were diluted to 20nM in 1x siRNA buffer (60mM KCl, 6mM HEPES pH 7.5 and 0.2mM $MgCl_2$), aliquoted and stored at −80°C until further use. Under sterile conditions at RT, 2.5µl of 20nM siRNAs were added to 80µl of fresh Opti-MEM solution (Gibco). Lipofectamine RNAiMAX transfection reagent (1.6µl, ThermoFisher) was added to 80µl of 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 incubated at RT for 20 min, before adding dropwise to $2x10^4$ cells in 0.8ml of culture media and plated in 1 well of a 12-well plate. The next day, the culture media was exchanged and a second round of siRNA treatment was performed. After both rounds of transfection comprising a total of 48 hours of siRNA treatment, cells

were harvested and their RNA was extracted using the RNeasy kit (Qiagen) according to the manufacturer's protocol. RNA-seq libraries were then assembled as described below.

**Bulk RNA-seq**

Cells were washed with 1x PBS and dissociated to single cells with trypsin. Harvested cells were then lysed with Trizol reagent (Life Technologies) and RNA was extracted with the RNeasy kit (Qiagen) according to the manufacturer's protocol. 4µg of RNA per sample was used for RNA-seq library preparation using the TruSeq Stranded mRNA Library Prep Kit (Illumina), following the manufacturer's protocol. Libraries were sequenced using NovaSeq S2 (Illumina) at 50bp paired-end settings.

**Processing of bulk RNA-seq data**

RNA-seq reads were trimmed using trim_galore with default parameters to remove the standard Illumina adaptor sequences (https://github.com/FelixKrueger/TrimGalore). Reads were then mapped to the mouse genome (mm9 assembly) using TopHat2 (Kim et al., 2013) with default parameters. Reads with mapping quality less than 30 were removed using samtools (Li et al., 2009). Read counts for each gene were calculated using HTSeq using the following parameters: --format=bam --order=pos --stranded=reverse --minaqual=0 --type=exon --mode=union --idattr=gene_name (Anders et al., 2015). Genes with low read count were removed (keeping genes with counts-per-million >/= 0.5 in at least two samples). Regularized log transformation (rlog) of each gene in each sample was calculated using DESeq2 (Love et al., 2014). Values of rlog were used as normalized gene expression data unless noted otherwise. In addition, RPKM (Reads Per Kilobase

per Million mapped reads) values were calculated for each gene. Differential gene expression analysis was performed using DESeq2 (Love et al., 2014).

**RAP-seq**

RNA anti-sense purification followed by DNA sequencing (RAP-seq) was performed as previously described (Engreitz et al., 2013). Briefly, after washing once with 1x PBS, cells were trypsinized and counted, before crosslinking with freshly-made 2mM disuccinimidyl glutarate (DSG) (ThermoFisher) in 1x PBS (used 4ml of DSG crosslinking solution per 10 million cells) for 45 min at RT. Next, cells were further crosslinked with 3% methanol-free formaldehyde (FA) (ThermoFisher) in 1x PBS (used 4ml of 3% FA per 10 million cells) for 10 min at RT, and the reaction was quenched by the addition of 1:5 dilution of freshly-made 2.5M glycine (Promega). Cells were then pelleted at 4°C, flash-frozen in liquid nitrogen and stored in -80°C until samples from all time points had been collected. Fixed cell pellets were lysed in NP-40 containing lysis buffer supplemented with sodium deoxycholate and N-lauroylsarcosine. Chromatin was fragmented by sonication, followed by TURBO DNase (ThermoFisher) digestion. Probe hybridization (mouse *Xist* probes for RAP-seq were generously designed and gifted from the Guttman Lab at Caltech (https://www.guttmanlab.caltech.edu/)) was performed using 50pmol of probe per 5 million cells for 3 hours at 37°C. After multiple stringent washes, DNA was eluted by RNase H (NEB) digestion, and crosslinking was reversed via Proteinase K (Sigma) digestion of eluted DNA at 60˚C overnight. DNA libraries were prepared using NEBNext Ultra End Repair/dATailing Module (NEB), followed by ligation of TruSeq DNA adapters (Illumina) using Quick Ligase (NEB). Finally, libraries were amplified using KAPA HiFi

Polymerase (Roche), pooled, and sequenced on the NovaSeq S2 platform (Illumina) to generate 50bp paired-end reads.

**RAP-seq alignment**

RAP-seq reads were trimmed using trim_galore with default parameters to remove the standard Illumina adaptor sequences (https://github.com/FelixKrueger/TrimGalore). Bowtie2 (Langmead and Salzberg, 2012) was used to align reads to the mouse genome (mm9) using the default parameters. Reads with mapping quality less than 30 were discarded using Samtools (Li et al., 2009), and Picard MarkDuplicates ("Picard Tools - By Broad Institute," n.d.) were used to mark PCR duplicates.

**Calculation of RAP-seq enrichment**

Genomic windows were first defined across the chromosomes using four different approaches: 1) We used bedtools makewindows (Quinlan and Hall, 2010) to create coordinates of genomic intervals in either 100Kb windows every 25Kb (for enrichment analysis), 2) 1Mb every 250Kb (for feature enrichment), 3) or non-overlapping 100kb windows (for differential analysis) along the genome. 4) We also defined genomic regions that encompass 50Kb from transcription start sites (TSSs) and over the gene body to capture *Xist* levels around gene bodies. This was used for comparing gene expression and *Xist* enrichment. We then used bedtools intersect (Quinlan and Hall, 2010) to count RAP-seq reads in each respective genomic region as defined by the above four approaches from each sample. To account for differences in sequencing depth, the read counts in each genomic region were normalized to the sum of all reads in that sample.

To characterize *Xist* enrichment on autosomes, we calculated a scaled enrichment score by normalizing only autosomal windows to the sum of all autosomal reads.

RAP-seq enrichment scores in each genomic region were calculated using the ratio of the normalized read counts in the RAP-seq *Xist* pulldown to the input of each sample. A region was defined as an unmappable region using the inputs of all samples. Specifically, the R function isOutlier (scater package (McCarthy et al., 2017)) was used to identify outliers (having less or more than expected read counts) based on the minimal number of reads across all input samples (nmads>4). Genomic regions identified as outliers were removed from all downstream analysis. The normalized enrichment ratios were used in all subsequent computational analyses. We note that while the read counts of genomic intervals were defined in 100Kb or 1Mb windows every 25Kb and 250Kb, respectively, along the genome, the enrichment scores were assigned to the 25Kb or 250Kb windows in the center of the 100Kb and 1Mb windows. This was done to prevent overlapping windows in the downstream analyses.

**RAP-seq peak calling**

RAP-seq peak calling was performed using MACS2 callpeak (Zhang et al., 2008) with max gap=1000 using the input of each *Xist* pulldown sample as a control sample. For visualization, MACS2 bdgcmp was also used to generate fold-enrichment tracks. The normalized bedgraph files were then converted to tdf format using igvtools toTDF (Robinson et al., 2011). To compare peak scores between samples, we used bedtools merge (Quinlan and Hall, 2010) to compile a list of all detected peaks using either the

narrow or broad peaks of all samples. We then used bedtools intersect (Quinlan and Hall, 2010) to intersect between the merged peaks and the peaks identified in each sample.

**Differential analysis of RAP-seq data**

To examine the differences in *Xist* localization across the X-chromosome we compared Xist enrichment within 500kb windows across the X-chromosome between D2 and D4 of Xist induction in MEFs.

**Analysis of genomic features for *Xist* localization**

Two different approaches were used to identify genomic features associated with *Xist* enrichment. First, we correlated *Xist* enrichment scores in genomic intervals with interval scores of various genomic features, including gene density and repeats. For this, the gene annotation file was obtained from UCSC ("TxDb.Mmusculus.UCSC.mm9.knownGene," n.d.), and DNA repeat annotations were downloaded from the UCSC Genome Browser, track RepeatMasker, table rmsk (Karolchik et al., 2004). The number of annotated genes and repeats in a given genomic interval was calculated, and Pearson correlation was used to calculate the correlation between the *Xist* enrichment score and feature count in each genomic interval (1Mb windows every 250Kb). The *Xist* locus (and an additional 10Mb at either end) was removed for these analyses, similar to Engreitz et al. (Engreitz et al., 2013). Second, we calculated the enrichment of different features with conserved autosomal *Xist* peaks. In addition to the features described above, we took advantage of the LOLA Core database (Sheffield and Bock, 2016), in which the R package LOLA was

used to compute the enrichment of the different features within the conserved autosomal *Xist* peaks.

**Gene Ontology analysis**

Enrichment for gene ontologies was performed for genes overlapping with conserved, EpiLC-specific autosomal *Xist* peaks using the R package topGO (Alexa et al., 2006) with a Fisher test.

## REFERENCES

Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *22*, 1600–1607.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high throughput sequencing data. Bioinformatics *31*, 166–169.

Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X inactivation centre. Nat. Rev. Genet. *12*, 429–442.

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

Bonora, G., Deng, X., Fang, H., Ramani, V., Qiu, R., Berletch, J.B., Filippova, G.N., Duan, Z., Shendure, J., Noble, W.S., and Disteche, C.M. (2018). Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. Nat. Commun. *9*, 1445.

Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.B.,

Vassilev, I., Galupa, R., Servant, N., Barillot, E., Surani, A., Chen, C.J., and Heard, E., (2017). Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. *24*, 226–233.

Brockdorff, N. (2017). Polycomb complexes in X chromosome inactivation. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20170021.

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., and Tartaglia, G.G. (2019). Phase separation drives X-chromosome inactivation: a hypothesis. Nat. Struct. Mol. Biol. *26*, 331–334.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. *20*, 2223–2237.

Chow, J.C., Hall, L.L., Baldry, S.E.L., Thorogood, N.P., Lawrence, J.B., and Brown, C.J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. Proc. Natl. Acad. Sci. *104*, 10104–10109.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

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

Cremer, M., Grasser, F., Lanctot, C., Muller, S., Neusser, M., Zinner, R., Solovei, I., and Cremer, T. (2008). Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. Methods Mol. Biol. *463*, 205–239.

Czerminski, J.T., and Lawrence, J.B. (2020). Silencing trisomy 21 with XIST in neural stem cells promotes neuronal differentiation. Dev. Cell *52*, 294-308.

Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J.B., Blau, C.A., Shendure, J., Duan, Z., Noble, W.S., and Disteche, C.M. (2015). Bipartite structure of the inactive mouse X chromosome. Genome Biol. *16*, 152.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

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

Dreesen, O., and Brivanlou, A.H. (2007). Signaling pathways in cancer and embryonic stem cells. Stem Cell Rev. *3*, 7–17.

Dror, I., Chitiashvili, T., Sahakyan, A., Tan, S.Y.X., Ding, F., Miller, J., and Plath, K. (2021). XIST controls X chromosome dampening and autosomal genes in early human development. In review.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., Plath, K., and Guttman, M. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. Curr. Opin. Genet. Dev. *31*, 57–66.

Gendrel, A.V., and Heard, E. (2014). Noncoding RNAs and epigenetic mechanisms during X chromosome inactivation. Annu. Rev. Cell Dev. Biol. *30*, 561–580.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., and Dekker, J. (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575–579.

Guyochin, A., Maenner, S., Chu, ET-J, Hentati, A., Attia, M., Avner, P., and Clerc, P. (2014). Live cell imaging of the nascent inactive X chromosome during the early differentiation process of naive ES cells towards epiblast stem cells. PLoS ONE *9*.

Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., and Lawrence, J.B. (2002a). An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. *99*, 8677–8682.

Hall, L.L., Clemson, C.M., Byron, M., Wydner, K., and Lawrence, J.B. (2002b). Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum. Mol. Genet. *11*, 3157–3165.

Hall, L.L., and Lawrence, J.B. (2010). XIST RNA and architecture of the inactive X chromosome: implications for the repeat genome. Cold Spring Harb. Symp. Quant. Biol. *75*, 345–356.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519–532.

Heard, E. (2005). Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. Curr. Opin. Genet. Dev. *15*, 482–489.

Jiang, J., Jing, Y., Cost, G.J., Chiang, J.-C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., Pearl, J.R., Rebar, E.J., Byron, M., Gregory, P.D., Brown, C.J., Urnov, F.D., Hall, L.L., and Lawrence, J.B. (2013). Translating dosage compensation to trisomy 21. Nature *500*, 296–300.

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493-496.

Kelsey, A.D., Yang, C., Leung, D., Minks, J., Dixon-McDougall, T., Baldry, S.E.L., Bogutz, A.B., Lefebvre, L., and Brown, C.J. (2015). Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. Genome Biol. *16*, 208.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presences of insertions, deletions and gene fusions. Genome Biol. *14*:R36.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol. *2*:E171.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie2. Nat. Methods *9*, 357–359.

Lee, J.T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275–279.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Loda, A., Brandsma, J.H., Vassilev, I., Servant, N., Loos, F., Amirnasr, A., Splinter, E., Barillot, E., Poot, R.A., Heard, E. and Gribnau, J. (2017). Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. Nat. Commun. *8*, 690.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Markaki, Y., Chong, J.G., Luong, C., Tan, S.Y.X., Wang, Y., Maestrini, D., Dror, I., Mistry, B.A., Schoeneberg, J., Banerjee, A., Guttman, M., Chou, T., and Plath, K. (2021). Xist-seeded supra-molecular complexes dynamically localize transcriptional repressors to silence the X-chromosome. In review.

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186.

McHugh, C.A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M.J., Shishkin, A.A., Su, J., Lander, E.S., Hess, S., Plath, K., and Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., and Lee, J.T. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*:aab2276.

Minks, J., Baldry, S.E., Yang, C., Cotton, A.M., and Brown, C.J. (2013). XIST-induced silencing of flanking genes is achieved by additive action of repeat a monomers in human somatic cells. Epigenet. Chromatin *6*, 23.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J.C., and Reik, W. (2017). Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. Cell Rep. *20*, 1215–1228.

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., and Zakian, S.M. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res. *11*, 833–849.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.P., Renard, J.P., Duranthon, V., and Heard, E. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature *472*, 370–374.

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W.R., Damianov, A., Chronis, C., Papp, B., Chen, C.K., McKee, R., Wang, X.-J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, D.L., and Plath, K. (2020). A protein assembly mediates Xist localization and gene silencing. Nature *587*, 145–151.

Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. *42*, 733–772.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *165*, 1012–1026.

Picard Tools - By Broad Institute [WWW Document], n.d. URL http://broadinstitute.github.io/picard/

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. *36*, 233–278.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131-135.

Popova, B.C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T.B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. Proc. Natl. Acad. Sci. *103*, 7706–7711.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ridings-Figueroa, R., Stewart, E.R., Nesterova, T.B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., Bageghni, S.A., Albadrani, G., Mansfield, W., Roulson, J.-A., Brockdorff, N., Ainscough, J.F.X., and Coverley, D. (2017). The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. Genes Dev. *31*, 876–888.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., Ali, S., Chin, S.-F., Palmieri, C., Caldas, C., and Carroll, J.S. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

Sahakyan, A., Yang, Y., and Plath, K. (2018). The role of Xist in X-chromosome dosage compensation. Trends Cell Biol. *28*, 999–1013.

Savarese, F., Flahndorfer, K., Jaenisch, R., Busslinger, M., and Wutz, A. (2006). Hematopoietic precursor cells transiently reestablish permissiveness for X inactivation. Mol. Cell. Biol. *26*, 7167–7177.

Sell, S. (2004). Stem cell origin of cancer and differentiation therapy. Crit. Rev. Oncol. Hematol. *51*, 1–28.

Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics *32*, 587–589.

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465–469.

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

Theisen, A., and Shaffer, L.G. (2010). Disorders caused by chromosome abnormalities. Appl. Clin. Genet. *3*, 159–174.

Wang, F., Shin, J., Shea, J.M., Yu, J., Boskovic, A., Byron, M., Zhu, X., Shalek, A.K., Regev, A., Lawrence, J.B., Torres, E.M., Zhu, L.J., Rando, O.J., and Bach, I. (2016). Regulation of X-linked gene expression during early mouse development by Rlim. Elife *5*: e19127.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *4*, 695–705.

Yang, C., Chapman, A.G., Kelsey, A.D., Minks, J., Cotton, A.M., and Brown, C.J. (2011). X chromosome inactivation: molecular mechanisms from the human perspective. Hum. Genet. *130*, 175–185.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zylicz, J.J., Bousard, A., Zumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D., Cramer, P., and Heard, E. (2019). The implication of early chromatin changes in X chromosome inactivation. Cell *176*, 182–197.

Figure 1

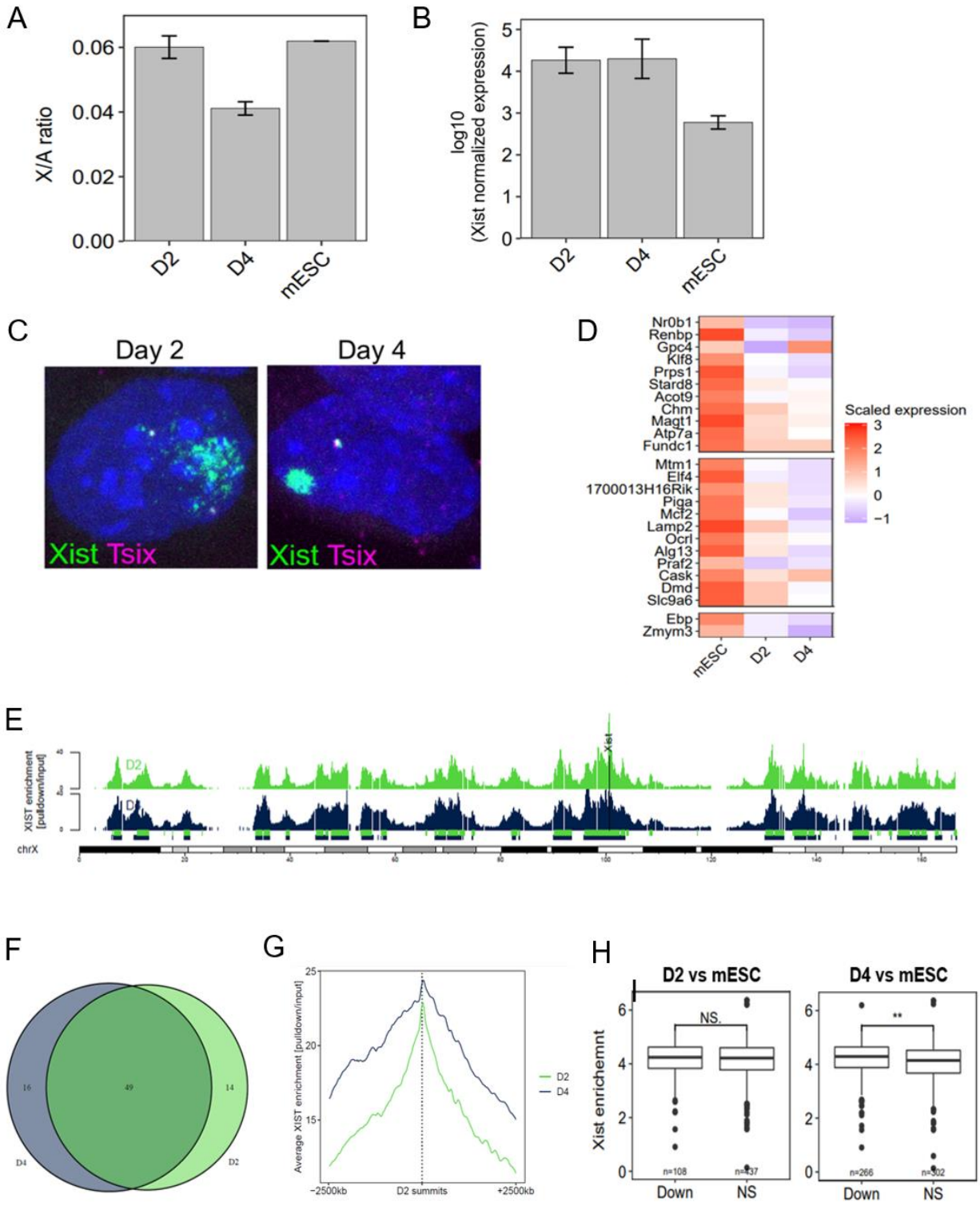

A
X/A ratio
0.06
0.04
0.02
0.00
D2
D4
mESC
B
log10
(Xist normalized expression)
5
4
3
2
1
0
D2
D4
mESC
C
Day 2
Day 4
Xist Tsix
Xist Tsix
D
Nr0b1
Renbp
Gpc4
Klf8
Prps1
Stard8
Acot9
Chm
Magt1
Atp7a
Fundc1
Mtm1
Elf4
1700013H16Rik
Piga
Mcf2
Lamp2
Ocrl
Alg13
Praf2
Cask
Dmd
Slc9a6
Ebp
Zmym3
mESC
D2
D4
Scaled expression
3
2
1
0
−1
E
XIST enrichment [pulldown/input]
D2
D4
Xist
chrX
F
16
49
14
D4
D2
G
Average XIST enrichment [pulldown/input]
25
20
15
−2500kb
D2 summits
+2500kb
D2
D4
H
D2 vs mESC
D4 vs mESC
Xist enrichemnt
NS.
**
6
4
2
0
n=108
n=437
n=266
n=302
Down
NS
Down
NS

Figure 2

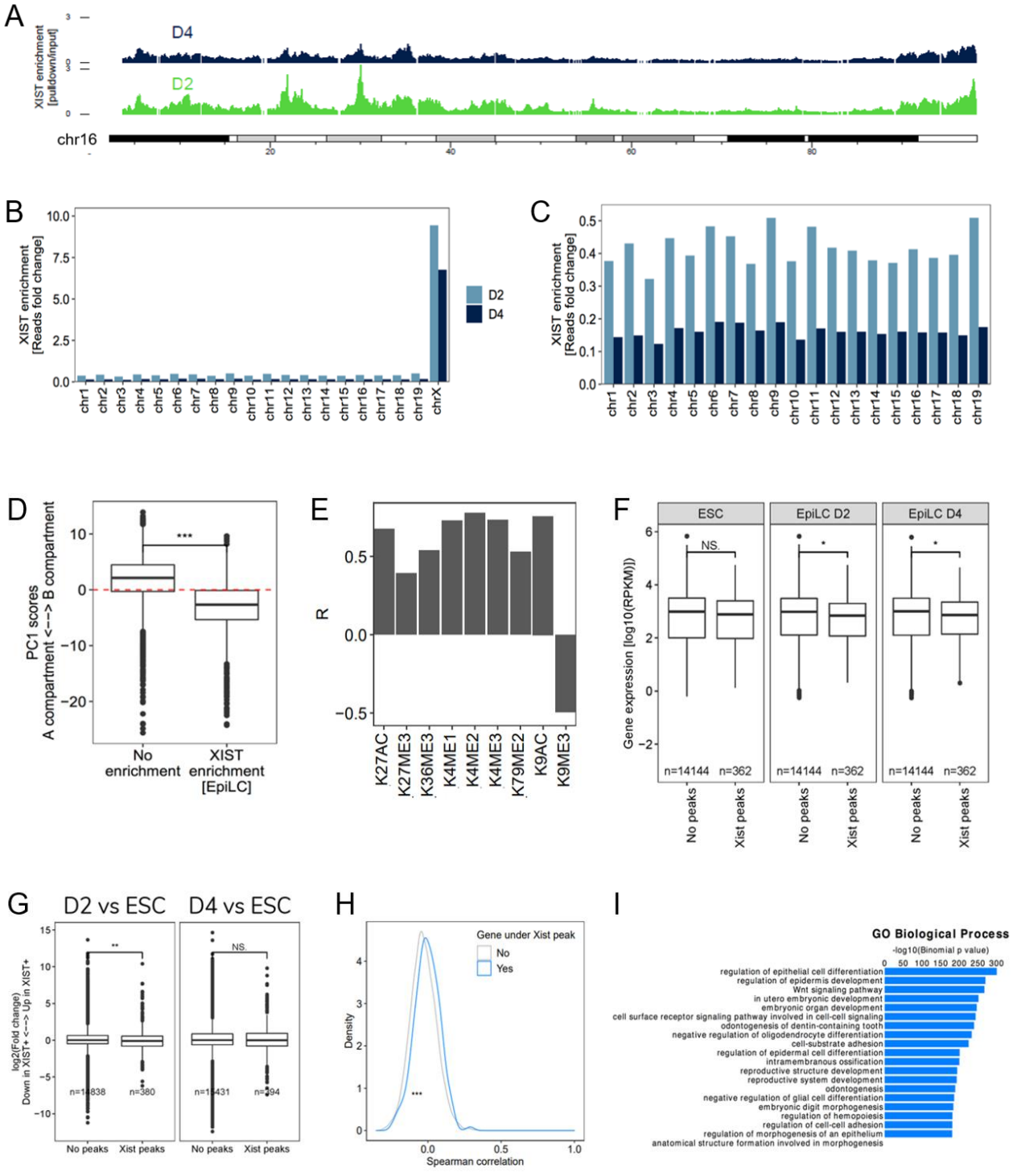

A
D4
D2
XIST enrichment [pulldown/input]
chr16
B
XIST enrichment [Reads fold change]
D2
D4
C
XIST enrichment [Reads fold change]
D
PC1 scores A compartment <---> B compartment
No enrichment
XIST enrichment [EpiLC]
E
R
K27AC
K27ME3
K36ME3
K4ME1
K4ME2
K4ME3
K79ME2
K9AC
K9ME3
F
ESC
EpiLC D2
EpiLC D4
Gene expression [log10(RPKM)]
NS.
n=14144
n=362
No peaks
Xist peaks
G
D2 vs ESC
D4 vs ESC
log2(Fold change) Down in XIST+ <---> Up in XIST+
n=14838
n=380
n=15431
n=394
H
Gene under Xist peak
No
Yes
Density
Spearman correlation
I
GO Biological Process
-log10(Binomial p value)
regulation of epithelial cell differentiation
regulation of epidermis development
Wnt signaling pathway
in utero embryonic development
embryonic organ development
cell surface receptor signaling pathway involved in cell-cell signaling
odontogenesis of dentin-containing tooth
negative regulation of oligodendrocyte differentiation
cell-substrate adhesion
regulation of epidermal cell differentiation
intramembranous ossification
reproductive structure development
reproductive system development
odontogenesis
negative regulation of glial cell differentiation
embryonic digit morphogenesis
regulation of hemopoiesis
regulation of cell-cell adhesion
regulation of morphogenesis of an epithelium
anatomical structure formation involved in morphogenesis

Figure 3

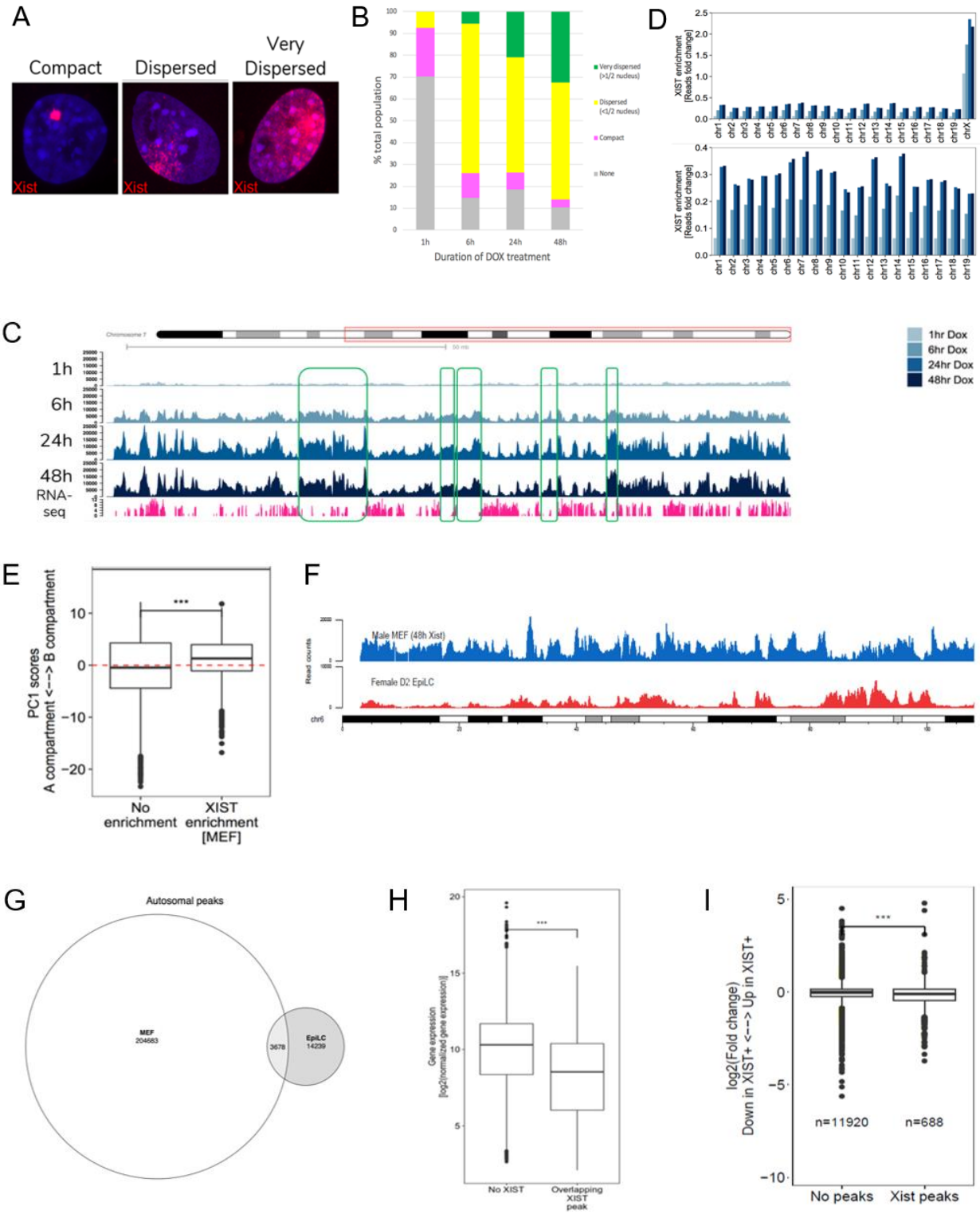

A
Compact
Dispersed
Very Dispersed
Xist
B
% total population
Duration of DOX treatment
1h
6h
24h
48h
Very dispersed (>1/2 nucleus)
Dispersed (<1/2 nucleus)
Compact
None
D
XIST enrichment [Reads fold change]
C
1h
6h
24h
48h
RNA-seq
1hr Dox
6hr Dox
24hr Dox
48hr Dox
E
PC1 scores
A compartment <--> B compartment
No enrichment
XIST enrichment [MEF]
F
Male MEF (48h Xist)
Female D2 EpiLC
Read counts
G
Autosomal peaks
MEF
204683
3678
EpiLC
14239
H
Gene expression [log2(normalized gene expression)]
No XIST
Overlapping XIST peak
I
log2(Fold change) Down in XIST+ <--> Up in XIST+
n=11920
n=688
No peaks
Xist peaks

Figure 4

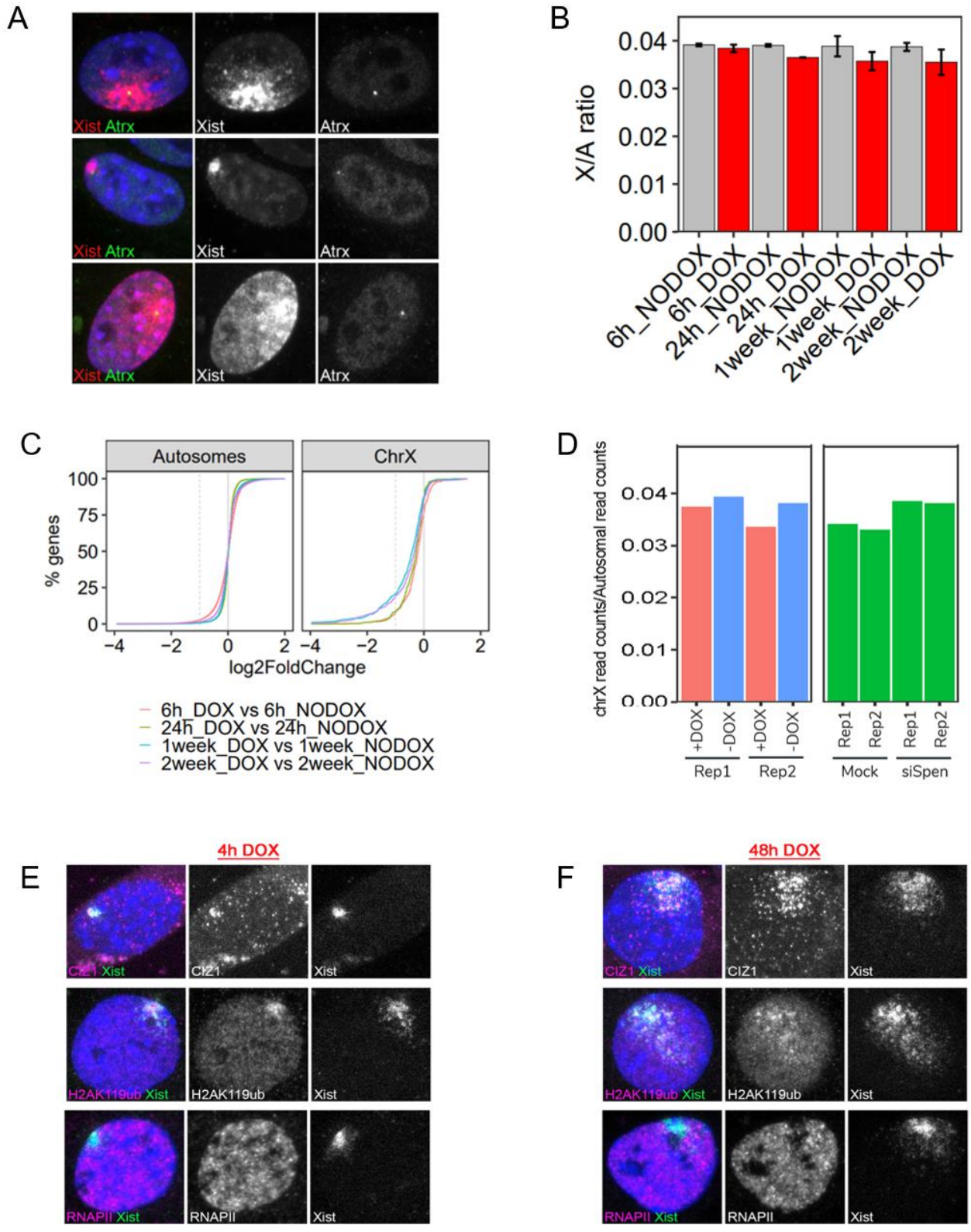

A
Xist Atrx
Xist
Atrx
B
X/A ratio
0.04
0.03
0.02
0.01
0.00
6h_NODOX
6h_DOX
24h_NODOX
24h_DOX
1week_NODOX
1week_DOX
2week_NODOX
2week_DOX
C
Autosomes
ChrX
% genes
100
75
50
25
0
−4
−2
0
2
log2FoldChange
6h_DOX vs 6h_NODOX
24h_DOX vs 24h_NODOX
1week_DOX vs 1week_NODOX
2week_DOX vs 2week_NODOX
D
chrX read counts/Autosomal read counts
0.04
0.03
0.02
0.01
0.00
+DOX
-DOX
Rep1
Rep2
Mock
siSpen
E
4h DOX
CIZ1 Xist
CIZ1
H2AK119ub Xist
H2AK119ub
RNAPII Xist
RNAPII
F
48h DOX

Figure 5

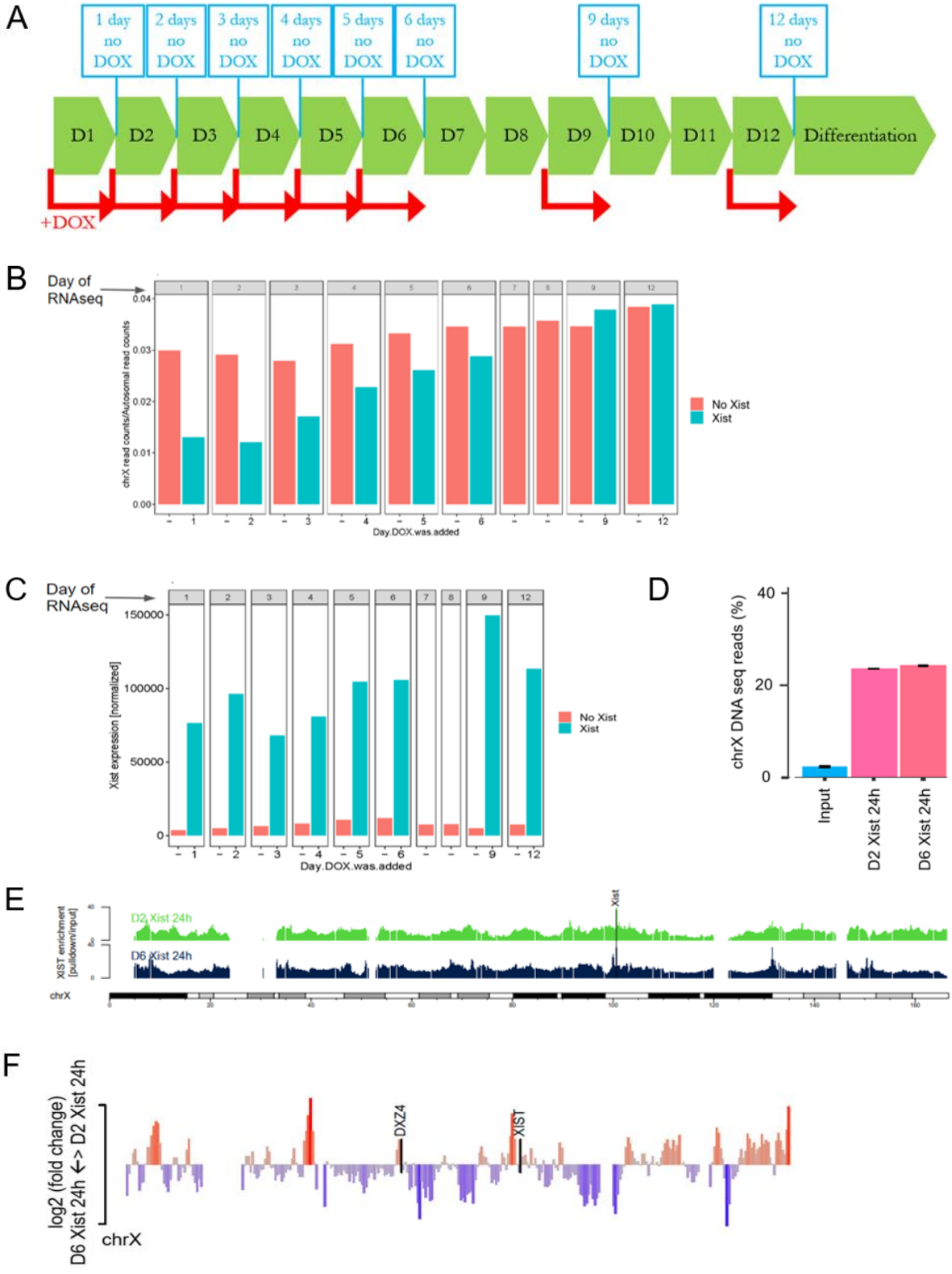
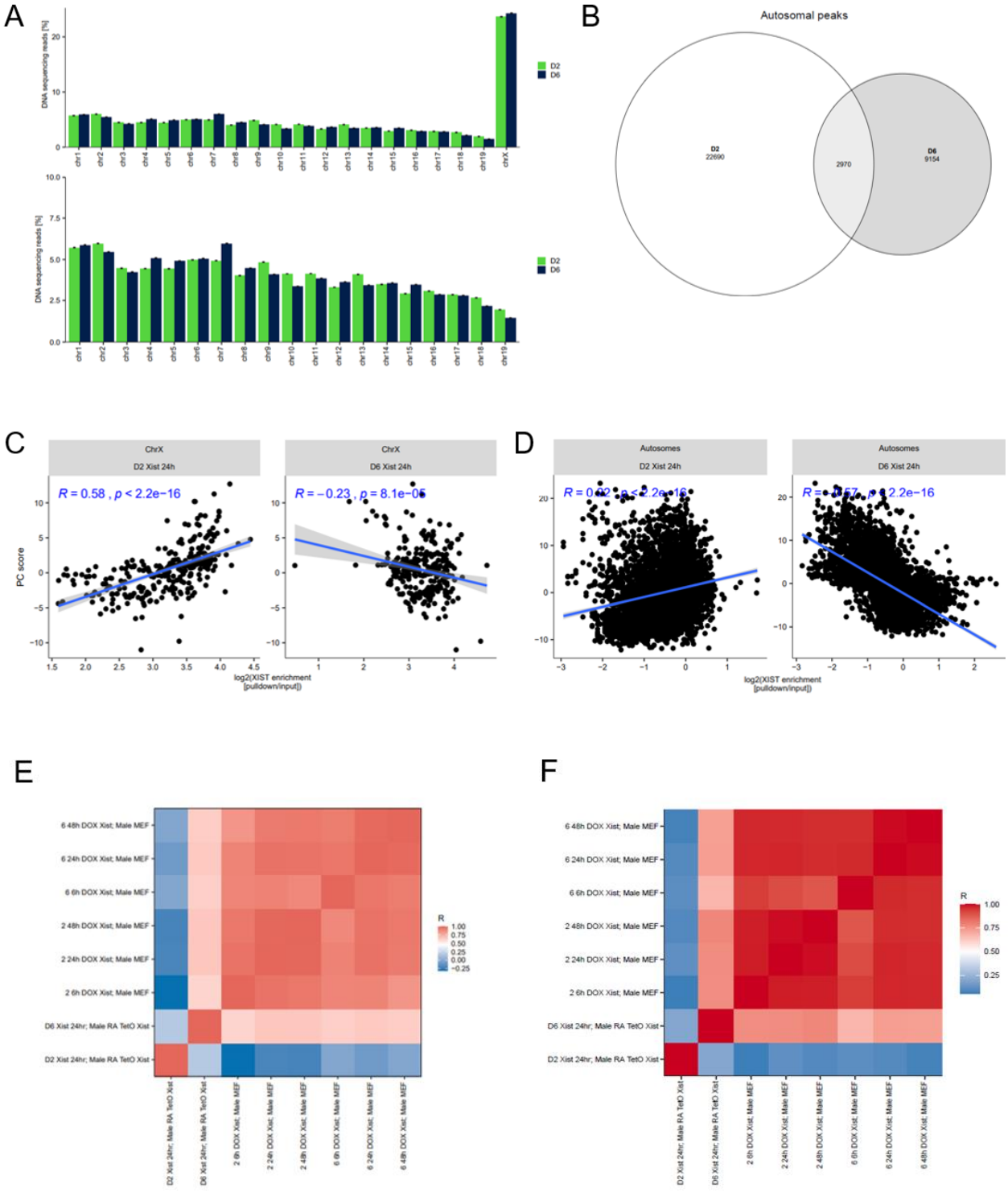

A
1 day no DOX
2 days no DOX
3 days no DOX
4 days no DOX
5 days no DOX
6 days no DOX
9 days no DOX
12 days no DOX
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
D12
Differentiation
+DOX
B
Day of RNAseq
chrX read counts/Autosomal read counts
Day.DOX.was.added
No Xist
Xist
C
Day of RNAseq
Xist expression [normalized]
Day.DOX.was.added
No Xist
Xist
D
chrX DNA seq reads (%)
Input
D2 Xist 24h
D6 Xist 24h
E
XIST enrichment [pulldown/input]
D2 Xist 24h
D6 Xist 24h
Xist
chrX
F
log2 (fold change) D6 Xist 24h ←> D2 Xist 24h
DXZ4
XIST
chrX

Figure 6


A
DNA sequencing reads [%]
D2
D6
B
Autosomal peaks
D2
22690
2970
D6
9154
C
ChrX
D2 Xist 24h
R = 0.58 , p < 2.2e−16
ChrX
D6 Xist 24h
R = −0.23 , p = 8.1e−06
PC score
log2(XIST enrichment
[pulldown/input])
D
Autosomes
D2 Xist 24h
Autosomes
D6 Xist 24h
log2(XIST enrichment
[pulldown/input])
E
6 48h DOX Xist; Male MEF
6 24h DOX Xist; Male MEF
6 6h DOX Xist; Male MEF
2 48h DOX Xist; Male MEF
2 24h DOX Xist; Male MEF
2 6h DOX Xist; Male MEF
D6 Xist 24hr; Male RA TetO Xist
D2 Xist 24hr; Male RA TetO Xist
R
F
6 48h DOX Xist; Male MEF
6 24h DOX Xist; Male MEF
6 6h DOX Xist; Male MEF
2 48h DOX Xist; Male MEF
2 24h DOX Xist; Male MEF
2 6h DOX Xist; Male MEF
D6 Xist 24hr; Male RA TetO Xist
D2 Xist 24hr; Male RA TetO Xist
R

# Chapter 5

Conclusion

By leveraging on multiple diverse approaches ranging from super-resolution live-cell imaging to high throughput genomics, we set out to address three primary objectives: (i) elucidating the molecular mechanism by which the chromosome-wide silencing compartment on the inactive X-chromosome (Xi) is formed and maintained across X-chromosome inactivation (XCI), (ii) determining if *XIST*, the master regulator of XCI, is responsible for X-chromosome dampening (XCD) in early human development, and (iii) characterizing the link between the genome-wide localization patterns of *Xist* and its X-linked gene silencing ability.

### *Xist* induces the spatial aggregation of proteins within locally restricted sites on the X-chromosome

In Chapter 2, we put forth a new model illustrating how *Xist* sets up the Xi compartment that mediates stable, long-term repression of X-linked genes. This model comes from our main observation by super-resolution live-cell imaging that *Xist* forms distinct foci that are confined to only about 50 sites across the entire X-chromosome. Each site contains two *Xist* molecules, which bind and recruit various protein factors early in XCI to form what we call supra-molecular complexes (SMCs), before Xi heterochromatin formation and chromosome-wide silencing are implemented. Over time, these *Xist*-SMCs seed the spatial aggregation of more protein partners, in particular, CIZ1 and CELF1 for localization across the Xi, PCGF5 and other components of Polycomb Repressive Complex 1 (PRC1) for inducing heterochromatinization and Xi compaction, as well as the key transcriptional repressor SPEN, aided by its intrinsically disordered regions (IDRs)

that promote intermolecular protein interactions, thereby gradually expanding this sphere of silencing across the entire X-chromosome.

Our proposed *Xist*-SMC model answers one of the biggest unresolved questions in the XCI field: how does a relatively low stoichiometry of about 100 *Xist* molecules, which is also observed by other groups in female somatic cells (Smeets et al., 2014; Sunwoo et al., 2015), spread across the entire 167Mb length of the X-chromosome to ascertain the robust silencing of more than 1000 genes? The dense nucleation of *Xist*-SMCs within the Xi enables the super-stoichiometric concentration of more *Xist* interactors, and coupled with PRC1-mediated chromatin compaction, facilitates the spatial probing and consequent silencing of a larger number of genomic targets.

Importantly, this model also lends credence to a role for liquid-liquid phase separation (LLPS) in establishing and maintaining the Xi compartment (Cerase et al., 2019; Pandya-Jones et al., 2020), as multi-valent interactions afforded by IDR-containing proteins like SPEN are known to be critical elements for sustaining LLPS (Mittag and Forman-Kay, 2007; Uversky, 2015; Banani et al., 2017; Turoverov et al., 2019), a biological phenomenon that has been increasingly recognized in diverse gene regulatory contexts (Dundr and Misteli, 2010; Shevtsov and Dundr, 2011; Banani et al., 2017; Shin and Brangwynne, 2017). Furthermore, recent work from our lab demonstrated the necessity of *Xist* to scaffold a macromolecular protein assembly comprising four different RNA binding proteins (RBPs), whose concentration increases over time to form a

membrane-free condensate with phase separating properties, for the stable localization of *Xist* and its maintenance of X-linked gene silencing (Pandya-Jones et al., 2020).

Future work is required to define the complete repertoire of protein factors that contribute to *Xist*-SMC formation, as various proteomic and genomic screens for mouse *Xist* interactors have uncovered almost a hundred different protein partners, many of whose functions in XCI remain enigmatic (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015; Moindrot et al., 2015; Monfort et al, 2015). It will also be of interest to determine the similarities and differences by which this *XIST*-SMC model operates in the human system, as mice and humans are known to employ vastly different X-linked dosage compensation strategies, especially during early embryonic development (Sahakyan et al., 2018). This may shed light on the inability of *XIST* to enact complete silencing in human pre-implantation embryos in a process called XCD (Petropoulos et al., 2016; Sahakyan and Plath, 2016; Sahakyan et al., 2017), which is discussed below.

### ***XIST*** **regulates X-chromosome dampening and autosomal genes in early female human development**

In Chapter 3, we showed the requirement of *XIST* in controlling XCD, as knocking out *XIST* in various human pluripotent stem cell (hPSC) lines that model XCD *in vitro* led to the loss of transcriptional downregulation on the dampened X-chromosome (Xd). By applying a genomics technique that examines the chromatin binding of *XIST* with high resolution, we discovered that there is an increased amount of *XIST* accumulation on the Xi than on the Xd, driving our speculation that *XIST* may be unable to exert complete

silencing on the Xd owing to the reduced enrichment of *XIST*. Moreover, we characterized a novel function of *XIST*, that it cannot only propagate to certain autosomal regions outside of the X-chromosome territory, but also repress some of the autosomal genes that it targets, resulting in a sex imbalance of the *XIST*-enriched autosomal gene expression levels that is also observed in human pre-implantation embryos *in vivo*.

Several non mutually-exclusive factors may account for the differential gene silencing capacity in XCD versus XCI: (i) decreased *XIST* levels on the Xd likely result in the concomitant reduction of protein factors bound by *XIST* that contribute to various aspects of XCI, (ii) different effector proteins may be recruited for XCI compared to XCD, and/or (iii) alteration of protein function by post-translational modifications. Therefore, further efforts should aim to delineate and compare the complete *XIST* interactomes, together with their protein states, that are responsible for both XCD and XCI, in order to generate a quantitative framework for determining how *XIST* RNA abundance and their associated protein levels/state collectively influence the X-linked gene regulatory outputs.

Nonetheless, the dual ability of *XIST* to regulate both XCD and XCI implies that this same lncRNA can produce functionally different gene expression outcomes across human development: first dampening of both X-chromosomes in female pre-implantation embryos, and then chromosome-wide silencing of one of the two X-chromosomes. Exactly how and when this developmental switch from XCD to XCI occurs remain open questions. Regardless, this is still a noteworthy discovery, because based on our understanding, no other RNA transcript capable of yielding two unequal gene regulatory

outputs has been reported. The exceptional plasticity of *XIST* to modulate distinct gene expression profiles in different developmental settings offers the potential to create nuclear compartments with specific gene regulatory outcomes, which can be exploited for therapeutic purposes that are discussed below.

Our finding that *XIST* can associate with chromosomes *in trans* is not anticipated, as it directly contradicts the prevailing view that *Xist* spreading is limited to the X-chromosome from which it is transcribed (Jonkers et al., 2008). Past studies have highlighted the ability of *XIST/Xist* to suppress gene expression when it is ectopically induced from an autosome (Lee and Jaenisch, 1997; Wutz and Jaenisch, 2000; Hall et al., 2002a; Jiang et al., 2013; Kelsey et al., 2015), or when it spreads onto certain autosomal regions in X to autosome (X;A) translocations (Hall et al., 2002b; Popova et al., 2006; Yang et al., 2011; Loda et al., 2017), but to our knowledge, this is the first time that autosomal spreading by *XIST/Xist* expressed from its endogenous locus is reported.

Furthermore, the distribution of *XIST* to autosomes appears to have physiological relevance, as gene ontology analysis revealed the enrichment of differentiation and developmental processes for the autosomal genes targeted by *XIST*. *XIST* also mediates the small but significant downregulation of these autosomal target genes, though the ramifications of this female-specific repression merit further research. Nevertheless, this fascinating finding heralds a potential novel function of *XIST* beyond XCD and XCI in early human development that is waiting to be explored.

### *Xist* localization correlates with its gene silencing capacity

In Chapter 4, we sought to dissect the relationship between the genome-wide localization pattern of *Xist* and its differential ability to implement transcriptional silencing, which was first introduced in Chapter 3. By analyzing numerous genomics datasets gathered from mouse cells at different stages of XCI, we found an overall lower enrichment of *Xist* on the Xi in early differentiating cells that is associated with reduced X-linked gene silencing, which is reminiscent of XCD in early human development, relative to the higher levels of *Xist* accumulation in late differentiated cells with complete XCI.

Upon interrogating the genomic features that underpin the differential *Xist* localization patterns, we demonstrate that cells with competent X-linked gene silencing exhibit preferential *Xist* distribution over the transcriptionally active A-compartment, particularly genomic regions that contain active histone marks. In contrast, cells portraying deficient or partial repression of a small fraction of genes tend to correlate with the spreading of *Xist* into the transcriptionally inactive B-compartment, especially chromatin domains where repressive histone modifications reside. Thus, it seems that the genomic locations where *Xist* propagates serve as a good predictor of its ability to trigger gene silencing. Further mining of published genomics datasets and harnessing machine-learning tools should provide deeper insights into numerous other genomic features, including repetitive elements like LINEs and SINEs, DNase hypersensitivity sites and topologically associated domain (TAD) boundaries, which may be important for dictating *Xist* localization.

Last but not least, we showed that *Xist* is still able to induce partial X-linked gene repression in a mouse somatic cell line, which has long exceeded a supposed critical developmental window surrounding pluripotency that is permissive for *Xist*-mediated silencing (Wutz and Jaenisch, 2000). However, the extent of this partial gene suppression cannot be increased to mimic the complete XCI-like silencing with longer duration of *Xist* expression, unlike in human differentiated neural stem cells that have *XIST* inducibly expressed on one of the three copies of chromosome 21 (Czerminski and Lawrence, 2020). Hence, *XIST/Xist* seems to retain a considerable amount of plasticity to execute chromosome-wide gene repression, thereby challenging the long-held view that the function of *XIST/Xist* is restricted to a developmentally regulated narrow timeframe. This remarkable potential of *XIST/Xist* to control gene regulatory outputs in different cells and developmental settings holds great promise in pursuing further research to better understand and capitalize on the exciting therapeutic prospects of this lncRNA for treating disorders caused by chromosomal abnormalities, such as duplications, trisomies and aneuploidies (Theisen and Shaffer, 2010).

## REFERENCES


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

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., and Tartaglia, G.G. (2019). Phase separation drives X-chromosome inactivation: a hypothesis. Nat. Struct. Mol. Biol. *26*, 331–334.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416.

Czerminski, J.T., and Lawrence, J.B. (2020). Silencing trisomy 21 with XIST in neural stem cells promotes neuronal differentiation. Dev. Cell *52*, 294-308.

Dundr, M., and Misteli, T. (2010). Biogenesis of nuclear bodies. Cold Spring Harb. Perspect. Biol. *2*, a000711.

Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., and Lawrence, J.B. (2002a). An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. *99*, 8677–8682.

Hall, L.L., Clemson, C.M., Byron, M., Wydner, K., and Lawrence, J.B. (2002b). Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum. Mol. Genet. *11*, 3157–3165.

Jiang, J., Jing, Y., Cost, G.J., Chiang, J.-C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., Pearl, J.R., Rebar, E.J., Byron, M., Gregory, P.D., Brown, C.J., Urnov, F.D., Hall, L.L., and Lawrence, J.B. (2013). Translating dosage compensation to trisomy 21. Nature *500*, 296–300.

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

Kelsey, A.D., Yang, C., Leung, D., Minks, J., Dixon-McDougall, T., Baldry, S.E.L., Bogutz, A.B., Lefebvre, L., and Brown, C.J. (2015). Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. Genome Biol. *16*, 208.

Lee, J.T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275–279.

Loda, A., Brandsma, J.H., Vassilev, I., Servant, N., Loos, F., Amirnasr, A., Splinter, E., Barillot, E., Poot, R.A., Heard, E. and Gribnau, J. (2017). Genetic and epigenetic

features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. Nat. Commun. *8*, 690.

McHugh, C.A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M.J., Shishkin, A.A., Su, J., Lander, E.S., Hess, S., Plath, K., and Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., and Lee, J.T. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*:aab2276.

Mittag, T., and Forman-Kay, J.D. (2007). Atomic-level characterization of disordered protein ensembles. Curr. Opin. Struct. Biol. *17*, 3–14.

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. Cell Rep. *12*, 562–572.

Monfort, A., Minin, G.D., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. (2015). Cell Rep. *12*, 554–561.

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W.R., Damianov, A., Chronis, C., Papp, B., Chen, C.K., McKee, R., Wang, X.-J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, D.L., and Plath, K. (2020). A protein assembly mediates Xist localization and gene silencing. Nature *587*, 145–151.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *165*, 1012–1026.

Popova, B.C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T.B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. Proc. Natl. Acad. Sci. *103*, 7706–7711.

Sahakyan, A., and Plath, K. (2016). Transcriptome encyclopedia of early human development. Cell *165*, 777–779.

Sahakyan, A., Plath, K., and Rougeulle, C. (2017). Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372.

Sahakyan, A., Yang, Y., and Plath, K. (2018). The role of Xist in X-chromosome dosage compensation. Trends Cell Biol. *28*, 999–1013.

Shevtsov, S.P., and Dundr, M. (2011). Nucleation of nuclear bodies by RNA. Nat. Cell Biol. *13*, 167–173.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*:eaaf4382.

Smeets, D., Markaki, Y., Schmid, V.J., Kraus, F., Tattermusch, A., Cerase, A., Sterr, M., Fiedler, S., Demmerle, J., Popken, J., Leonhardt, H., Brockdorff, N., Cremer, T., Schermelleh, L., and Cremer, M. (2014). Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. Epigenet. Chromatin *7*, 8.

Sunwoo, H., Wu, J.Y., and Lee, J.T. (2015). 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. *112*, E4216–4225.

Theisen, A., and Shaffer, L.G. (2010). Disorders caused by chromosome abnormalities. Appl. Clin. Genet. *3*, 159–174.

Turoverov, K.K., Kuznetsova, I.M., Fonin, A.V., Darling, A.L., Zaslavsky, B.Y., and Uversky, V.N. (2019). Stochasticity of biological soft matter: Emerging concepts in intrinsically disordered proteins and biological phase separation. Trends Biochem. Sci. *44*, 716–728.

Uversky, V.N. (2015). The multifaceted roles of intrinsic disorder in protein complexes. FEBS Lett. *589*, 2498–2506.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *4*, 695–705.

Yang, C., Chapman, A.G., Kelsey, A.D., Minks, J., Cotton, A.M., and Brown, C.J. (2011). X chromosome inactivation: molecular mechanisms from the human perspective. Hum. Genet. *130*, 175–185.