Mechanistic dissection of SPEN function during X chromosome inactivation

François Dossin

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie

Mechanistic dissection of SPEN function during X chromosome inactivation

Fonctions et mécanismes de SPEN pendant l'inactivation du chromosome X

Soutenue par

François Dossin

Le 15 Mars 2021

École doctorale n°577

Structure et Dynamique des Systèmes Vivants

Spécialités

Sciences de la vie et de la santé

Composition du jury :

Pierre CAPY Professeur, Université Paris-Saclay (*Orsay*)	*Président*
Claire ROUGEULLE Directrice de recherche, Epigenetics and Cell Fate Institute (CNRS, *Paris*)	*Rapportrice*
Reiner VEITIA Professeur, Université de Paris (*Paris*); Institut F. Jacob (*Fontenay-aux-Roses*)	*Rapporteur*
Geneviève ALMOUZNI Directrice de recherche, Institut Curie (CNRS, *Paris*)	*Examinatrice*
Dirk SCHÜBELER Professeur, Friedrich Miescher Institute (*Basel*)	*Examinateur*
Edith HEARD Professeure, Collège de France (*Paris*); European Molecular Biology Laboratory (EMBL, *Heidelberg*)	*Directrice de thèse*

institut Curie | PSL UNIVERSITÉ PARIS

Contents

Part I

Introduction

Chapter 1

X Chromosome Inactivation

Disclaimer: The core of my PhD experimental work has been to characterize the function of SPEN in the process of X chromosome inactivation. The following chapter does not aim to provide an exhaustive review on the topic of X chromosome inactivation. Instead, it aims to address certain aspects of this process which are essential to understand the questions I investigated and grasp the relevance of the results I obtained.

1.1 Sex chromosome evolution in mammals

In therian mammals, sex is genetically determined by a pair of sex chromosomes called X and Y : females harbor two X chromosomes within their genome, while males only harbor one X chromosome as well as an additional Y chromosome [Marshall Graves and Shetty, 2001]. Although the X and Y chromosomes both evolved from the same pair of autosomes (proto-sex chromosomes, **Figure 1.1A**), they differ drastically from one another as a consequence of more than 300 million years of evolution (**Figure 1.1**) [Graves, 2006]. Today, while the X chromosome is 165Mb long in humans, encoding over 800 protein-coding genes, the Y chromosome is only 60Mb long and encodes no more than 50 protein-coding genes [Graves et al., 2006].

Approximately 300 million years ago, a mutation of the *SOX3* gene led to the generation of a new variant, called *SRY*, on one of the proto-sex chromosomes: the proto-Y (**Figure 1.1B**) [Graves, 2006]. This variant conferred the ability to its carriers to be male, while individuals carrying the unmodified proto-X remained female. With evolutionary time, genes that were beneficial to males became enriched specifically on the proto-Y (**Figure 1.1C**) [Bachtrog, 2013]. Consequently, meiotic recombination between the proto X and Y would lead to undesired "intersex" states, wherein males would lose

some "beneficial" male-specific genes, while females would inherit "harmful" genes previously found only on the proto-Y (**Figure 1.1D**). Therefore, meiotic recombination around the male-specific genes was suppressed so as to preserve a strict distinction between the proto-X and the proto-Y (**Figure 1.1E**). Unable to recombine with the proto-X, the proto-Y accumulated deleterious mutations which could never be balanced out during meiosis (**Figure 1.1E**) [Muller, 1918; Rice, 1987, 1996]. This mutational burden led to the degeneration of most of the Y chromosome and its encoding genes, thereby providing an explanation for the difference between the modern X and Y chromosomes in mammals (**Figure 1.1E**) [Charlesworth, 1996; Charlesworth et al., 2005; Graves, 2006; Graves et al., 2006; Vicoso and Charlesworth, 2006].

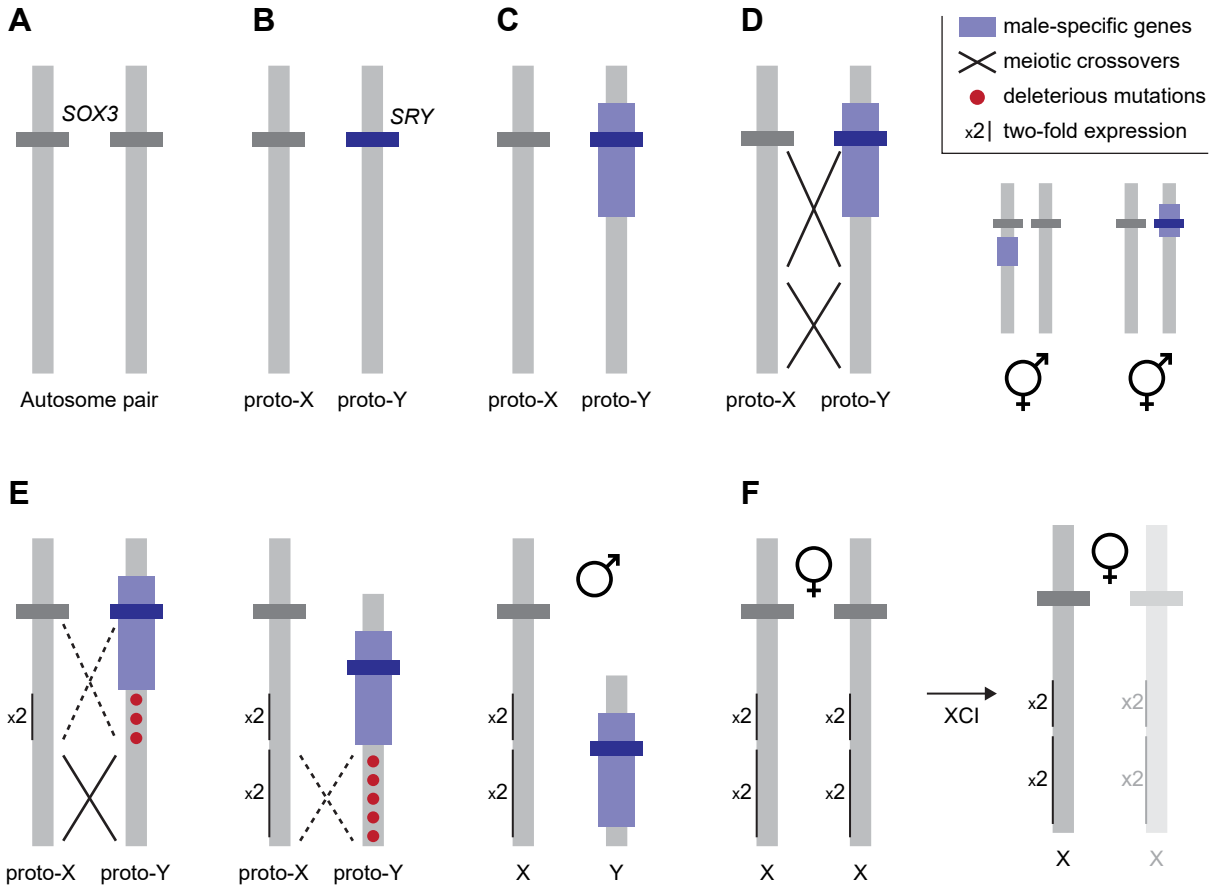


Figure 1.1 – Mammalian sex chromosome evolution and the requirement for dosage compensation. **(A)** Mutations in the *SOX3* gene lead to **(B)** the emergence of a sex-determining gene, *SRY*, on one of the two proto-sex chromosomes. **(C)** Male-specific genes accumulate on the proto-Y, leading to **(D)** deleterious inter-sex arrangements following recombination events in meiosis. **(E)** Consequential suppression of meiotic recombination in the male germline favors fixation of deleterious mutations, resulting in the progressive degeneration of the proto-Y chromosome. As genes are lost on the proto-Y, selective pressures ensure two-fold upregulation of corresponding genes on the proto-X. **(F)** The resulting increase in X-linked gene dosage in females is compensated through X chromosome inactivation.

1.2 Dosage compensation problems and the discovery of X chromosome inactivation

Importantly, as genes were being lost on the Y chromosome, a gene dosage problem arose between XX females and XY males. This problem was first identified by the Japanese geneticist Susumu Ohno, who proposed that as each gene gradually disappeared from the Y chromosome, selective pressures ensured that the expression of its corresponding copy on the X chromosome was increased by two-fold (**Figure 1.1E**) [Ohno, 1967]. Several studies have provided supporting evidence for Ohno's hypothesis that X-linked genes are two-fold upregulated [Deng et al., 2011; **?**; Lin et al., 2007]. However, while a two-fold expression of X-linked genes resolves dosage problems in XY males, it implies that XX females now have to cope with two copies of two-fold expressed X-linked genes (**Figure 1.1F**). Consequently, female mammals evolved a specific mechanism called *X chromosome inactivation* (XCI), during which one of the two X chromosomes is transcriptionally silenced in all cells, ensuring that each female cell expresses one copy of the X chromosome, identically to male cells (**Figure 1.1F**).

This process was first put forth by Mary Lyon in 1961 [Lyon, 1961]. Building up on previous observations that one X chromosome is heteropyknotic in female cells, and that females heterozygous at sex-linked loci affecting coat color show a mosaic phenotype (with hair patches of different color across the body), Lyon proposed that:

- One of the two X chromosomes in female cells is "genetically inactivated" *[sic]*
- This inactivation happens early during embryonic development
- Once inactivated, the same X chromosome remains inactivated in descending cells.

Decades of X chromosome inactivation research have proved that Mary Lyon's hypotheses were correct, and have hitherto provided remarkable insights into this process, which emerged as a paradigm for the study of epigenetics (see **section 1.11**).

1.3 X chromosome inactivation dynamics during early development

Mary Lyon proposed that XCI takes place early during embryonic development. Indeed, the observation of mosaic X-linked hair color phenotypes in females suggests that there should be at least one developmental time-window during which both maternal and paternal X-linked alleles (and, by extension, chromosomes) are expressed, such that each cell "chooses" to inactivate either one or the other X chromosome.

Early research conducted in mice, which emerged as the predominant model to study XCI, and on which we will focus for the rest of this dissertation, revealed that the paternal X chromosome (Xp) is always inactive in extraembryonic tissues, a phenomenon described as *imprinted* XCI (iXCI) [Takagi and Sasaki, 1975; West et al., 1977]. Given that embryonic/somatic tissues instead show a random pattern of XCI (mosaic phenotype), it was initially believed that embryonic and extraembryonic lineages undergo X inactivation independently, and under different modalities.

Later, it was shown that while both maternal and paternal X chromosomes are actively transcribed at zygotic gene activation (2-cell stage, **Figure 1.2A**) [Okamoto et al., 2005; Patrat et al., 2009], all cells initially undergo imprinted XCI (**Figure 1.2B**). Imprinted XCI initiates at the 4-cell stage such that by the early blastocyst stage, 3.5 days post coitum (E3.5), all cells of the embryo have an inactive Xp (**Figure 1.2C**) [Mak et al., 2004; Okamoto et al., 2004].

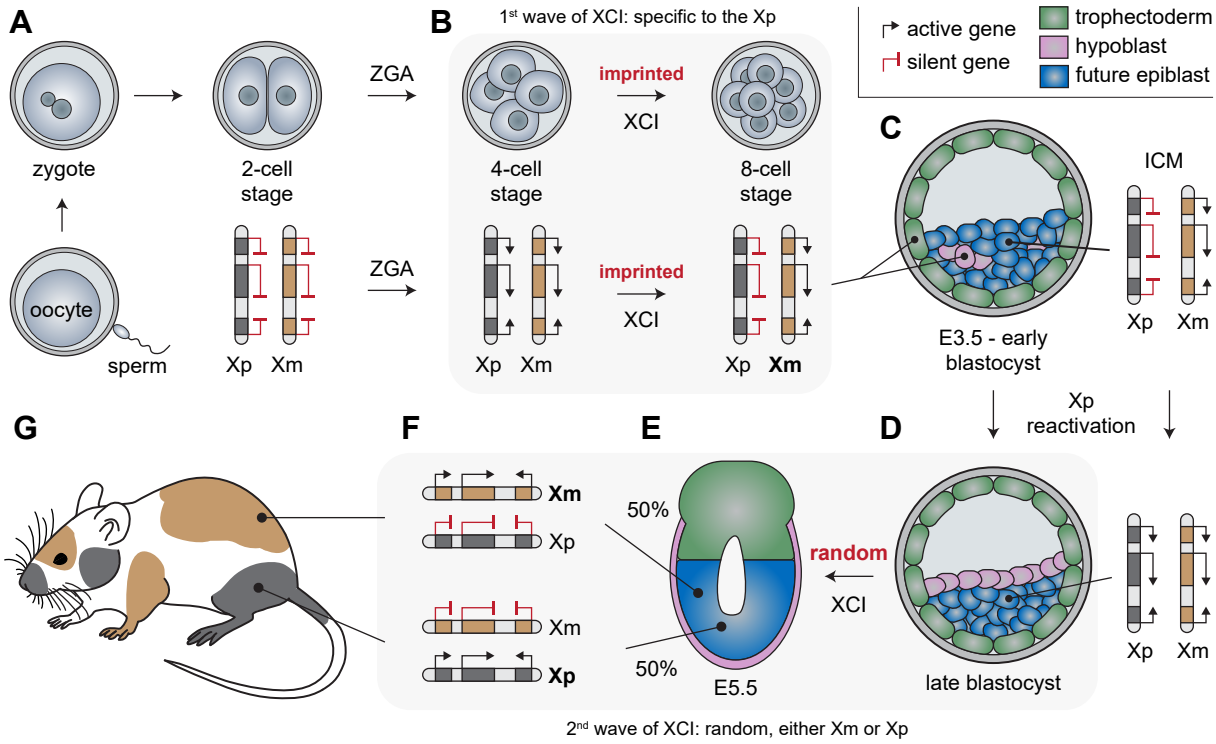


Figure 1.2 – Dynamics of X chromosome inactivation during mouse development. **(A)** Following fertilization and zygotic genome activation (ZGA), both X chromosomes are expressed. **(B)** At the 4-cell stage, imprinted XCI initiates and **(C)** by the early blastocyst stage, all cells have inactivated the Xp. The Xp remains inactivated in extraembryonic cells (trophectoderm and hypoblast), while **(D)** it becomes reactivated only in ICM cells that will give rise to the epiblast. **(E)** In such cells, random XCI occurs such that following implantation and clonal propagation of XCI status, **(F)** half of the cells of the embryo proper have inactivated the Xp, and the other half the Xm. **(G)** Random XCI hence results in cellular mosaicism with respect to X-linked allelic expression, which can for example translate into mosaic hair phenotypes.

The Xp remains stably inactivated in extraembryonic tissues (i.e. trophectoderm and hypoblast, **Figure 1.2C**). Remarkably however, the Xp is *reactivated* specifically in all cells of the *i*nner *c*ell *m*ass (ICM) that will give rise to the epiblast (**Figure 1.2D**) [Mak et al., 2004; Okamoto et al., 2004; Borensztein et al., 2017a]. As the epiblast forms and is further specified, each cell undergoes *random inactivation* of either the paternal or the maternal X chromosome, around 5.5 days post coitum (E5.5, **Figure 1.2E** and **F**) [Mak et al., 2004; Okamoto et al., 2004; Rastan, 1982].

Strikingly, the developmental dynamics and timing of XCI vary extensively among placental mammals [Okamoto et al., 2011; Escamilla-Del-Arenal et al., 2011]. In rabbits and humans for example, where XCI begins at later developmental stages than in mice, *Xist* expression is not imprinted in early embryos and XCI does not occur during pre-implantation development [Okamoto et al., 2011]. Instead, *Xist* RNA is seen accumulating on both human X chromosomes [Okamoto et al., 2011], leading to their transcriptional dampening (rather than complete inactivation of one X chromosome), ensuring dosage compensation by the early blastocyst stage [Vallot et al., 2017; Petropoulos et al., 2016]. Random XCI only proceeds after the late blastocyst stage [Okamoto et al., 2011]. These diverse strategies of XCI are proposed to reflect differences in embryo staging and timing of cellular differentiation between eutherian species [Monk and Harper, 1979; Migeon, 2016].

1.4 X inactivation can be recapitulated *in vitro*

Pluripotent mouse embryonic stem cells (mESCs) derived from the ICM can be cultured *in vitro*, where they can divide for many passages. Similarly to what is observed *in vivo* at the blastocyst stage, cultured female mESCs have two active X chromosomes and undergo random XCI when differentiation is triggered upon changes in culture conditions (**Figure 1.3**).

The ability of female mESCs to recapitulate XCI *in vitro* has made XCI research much more accessible and convenient, as it does not require handling peri-implantation embryos, which are particularly challenging to work with. Furthermore, mESCs in culture can easily be genetically manipulated and yield high cell numbers, allowing for the use of several technical tools and approaches that could not have been employed in early embryos. In fact, over the past 20 years, many important discoveries in the XCI field have been made thanks to the mESC *in vitro* model. The core of my doctoral research was undertaken in such *in vitro* model of XCI, and only a few experiments (see **subsection 6.2.2**) were also performed *in vivo* (i.e. in mouse embryos).

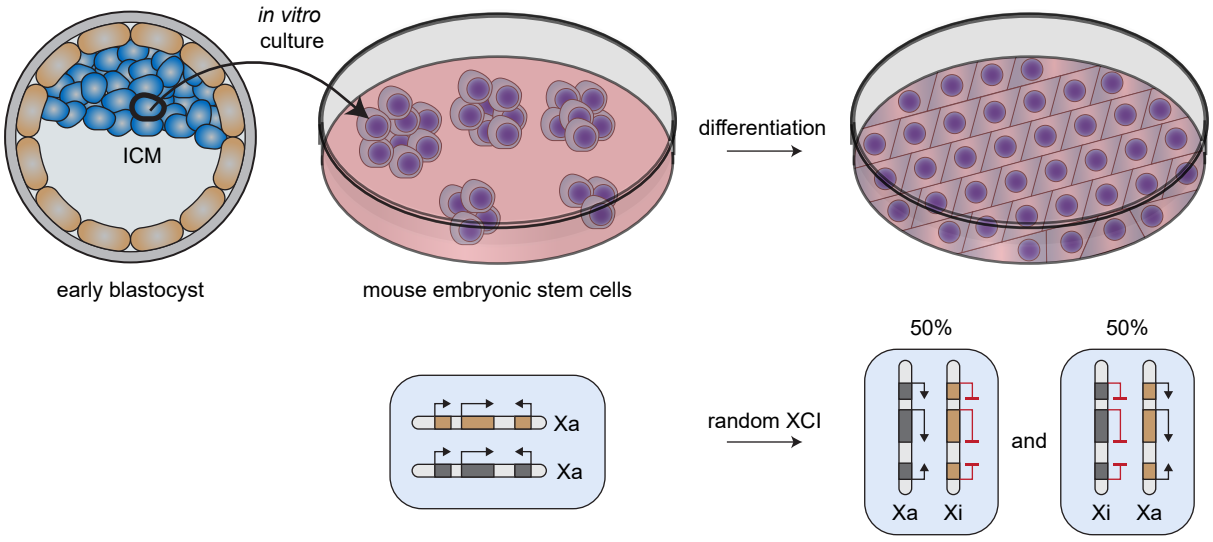


Figure 1.3 – Cultured mouse embryonic stem cells as an *in vitro* model of random XCI. Mouse embryonic stem cells derived from the ICM carry two active X chromosomes. Upon differentiation, these cells undergo random XCI, and one of the two X chromosomes is transcriptionally silenced.

1.5 The *X-inactivation center* and the discovery of *Xist* as an essential mediator of XCI

Twenty years after Lyon's seminal article, through analysis of X-chromosomal deletions and translocations in mouse embryos and embryonic cells, a single X-linked locus, termed the *X-i*nactivation *c*enter (*Xic*, **Figure 1.4A**), was identified to be the region that is necessary and sufficient to trigger XCI [Rastan, 1983; Rastan and Robertson, 1985]. Importantly, this locus has to be physically linked to the X-chromosome for XCI to occur, demonstrating that elements comprised within the *Xic* regulate XCI in *cis*. These studies paved the way for future lines of research aiming to identify the gene(s) responsible for mediating XCI.

In the early 1990s, research led by Carolyn Brown identified a novel human X-linked gene which, contrary to most other X-linked genes, was expressed from the inactive X chromosome but not from the active one [Brown et al., 1991]. This gene, called *XIST* (for *X-i*nactive *s*pecific *t*ranscript), maps within the *Xic* locus (**Figure 1.4B**) in both mouse and humans and shows conservation at both sequence and gene structure levels [Brockdorff et al., 1992; Brown et al., 1992]. The *Xist* transcript (15kb in mouse, 17kb in humans after splicing) is capped and polyadenylated (**Figure 1.4C**), but has no protein-coding potential, as it contains no conserved ORFs and does not associate with the translation machinery.

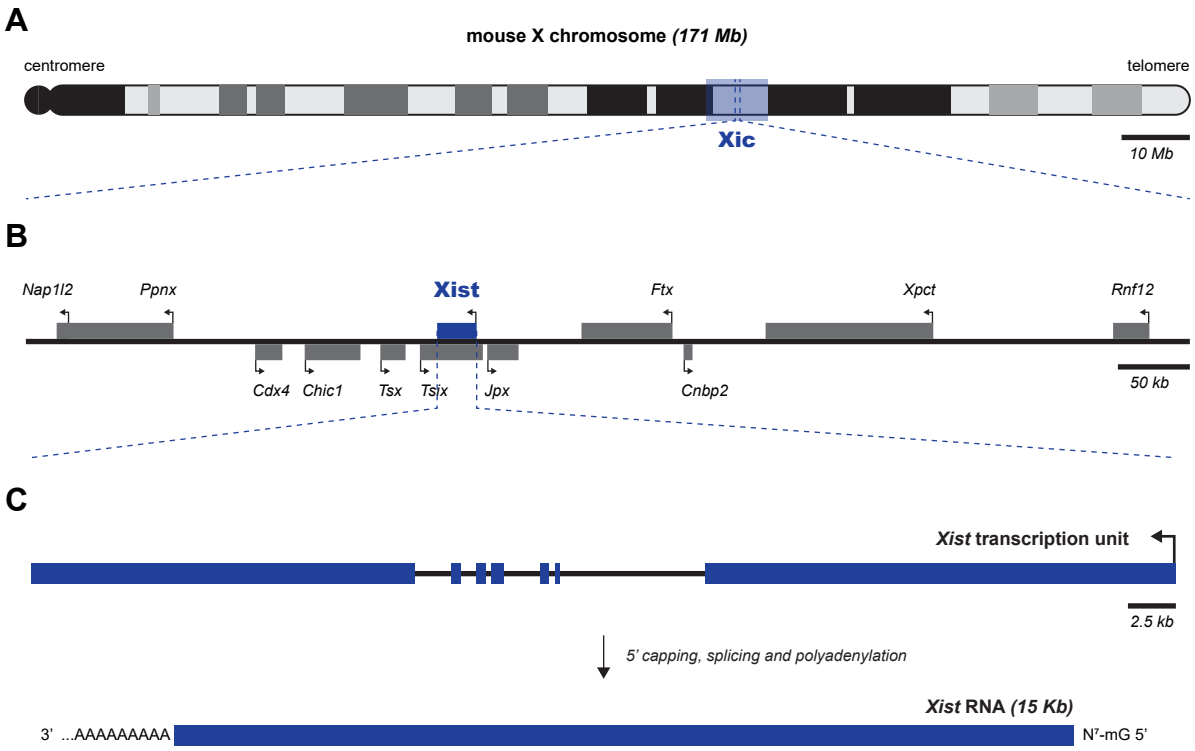


Figure 1.4 – Scheme of the *Xic* region and the *Xist* locus. **(A)** The position of the *Xic* locus (highlighted in blue) is shown relative to the entire mouse X chromosome. **(B)** The *Xist* gene (shown in blue) is located within the *Xic* region and is surrounded by several protein- and non-protein-coding genes (shown in grey). Some of these neighboring genes are implicated in regulating *Xist* expression (see **section 1.6**). **(C)** The predominant *Xist* transcript isoform is 15 kilobases long and comprises 7 exons. Similarly to other RNA polymerase II-transcribed genes, *Xist* RNA is 5'-capped and 3'-polyadenylated.

Instead, *Xist* RNA is retained in the nucleus, where it stably accumulates on the X chromosome from which it is transcribed [Clemson et al., 1996], a phenomenon commonly described as "coating" (**Figure 1.5**). Although several factors – notably the nuclear matrix proteins SAF-A/hnRNPU [Hasegawa et al., 2010] and CIZ1 [Ridings-Figueroa et al., 2017; Sunwoo et al., 2017] – have been shown to play an important role in *Xist* localization, how such coating takes place hitherto remains enigmatic [see Brockdorff, 2019, for a review on the topic]. Remarkably, *Xist* RNA initially does not spread uniformly and linearly along the X chromosome. Instead, *Xist* first accumulates at discrete "entry sites" – which show spatial proximity to the site of *Xist* transcription – and only then spreads to more spatially distant sites [Engreitz et al., 2013].

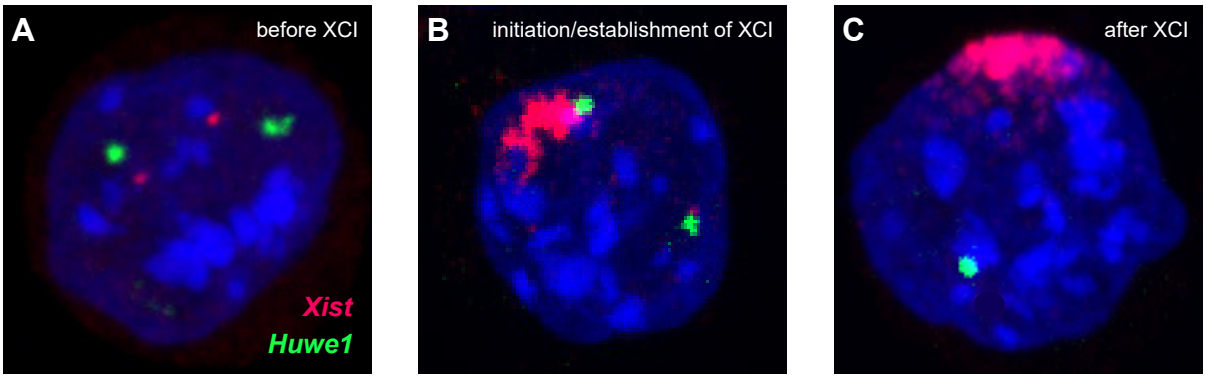


Figure 1.5 – *Xist* coats the X chromosome and mediates X-linked gene silencing. Images from RNA FISH performed on mESCs at different stages of the XCI process show that **(A)** both X chromosomes are initially active in mESCs prior to XCI, with two nascent transcription foci being detected for the X-linked gene *Huwe1* (green signal). **(B)** XCI initiates with the monoallelic upregulation of *Xist* (red signal), whose RNA coats the X chromosome from which it is transcribed, forming a detectable domain called a *Xist*-cloud. **(C)** Eventually, *Xist*-coating leads to chromosome-wide transcriptional silencing of X-linked genes in *cis*, as illustrated with the detection of a single *Huwe1* nascent transcription focus, coming from the active (i.e. non *Xist*-coated) X chromosome.

Deletion of the *Xist* gene in mouse embryonic stem cells [Penny et al., 1996] and embryos [Marahrens et al., 1997] results in failure to undergo XCI, while ectopic expression of *Xist*-encoding transgenes from autosomes leads to *cis*-autosomal gene inactivation [Wutz and Jaenisch, 2000]. Altogether, these studies demonstrate that *Xist* is a long non-conding RNA (lncRNA) whose expression is the earliest event required to mediate X chromosome inactivation in *cis* (**Figure 1.5**). **How *Xist* expression triggers chromosome-wide gene silencing has been the central question of my doctoral work**, and the existing state of knowledge about *Xist*'s silencing function – at the time I started my PhD – will be addressed in **section 1.12** and **section 1.13** of this chapter.

1.6 *Xist* expression is linked to cellular differentiation

Prior to cellular differentiation (i.e. exit from pluripotency), *Xist* is barely expressed in mESCs (**Figure 1.6A**). Upon differentiation however, *Xist* becomes upregulated (**Figure 1.6A**), strictly on one of the two X chromosomes (**section 1.4**). How *Xist* expression is so tightly regulated during differentiation, and how it is restricted to only one of the two X chromosomes has been extensively investigated [see Augui et al., 2011; van Bemmel et al., 2016; Galupa and Heard, 2015; Pontier and Gribnau, 2011, for reviews by experts from the field]. It is now accepted that *Xist* is mainly regulated at the transcriptional level, through action of both *cis* and *trans*-regulators, which will only be briefly covered in the following paragraphs (**Figure 1.6**).

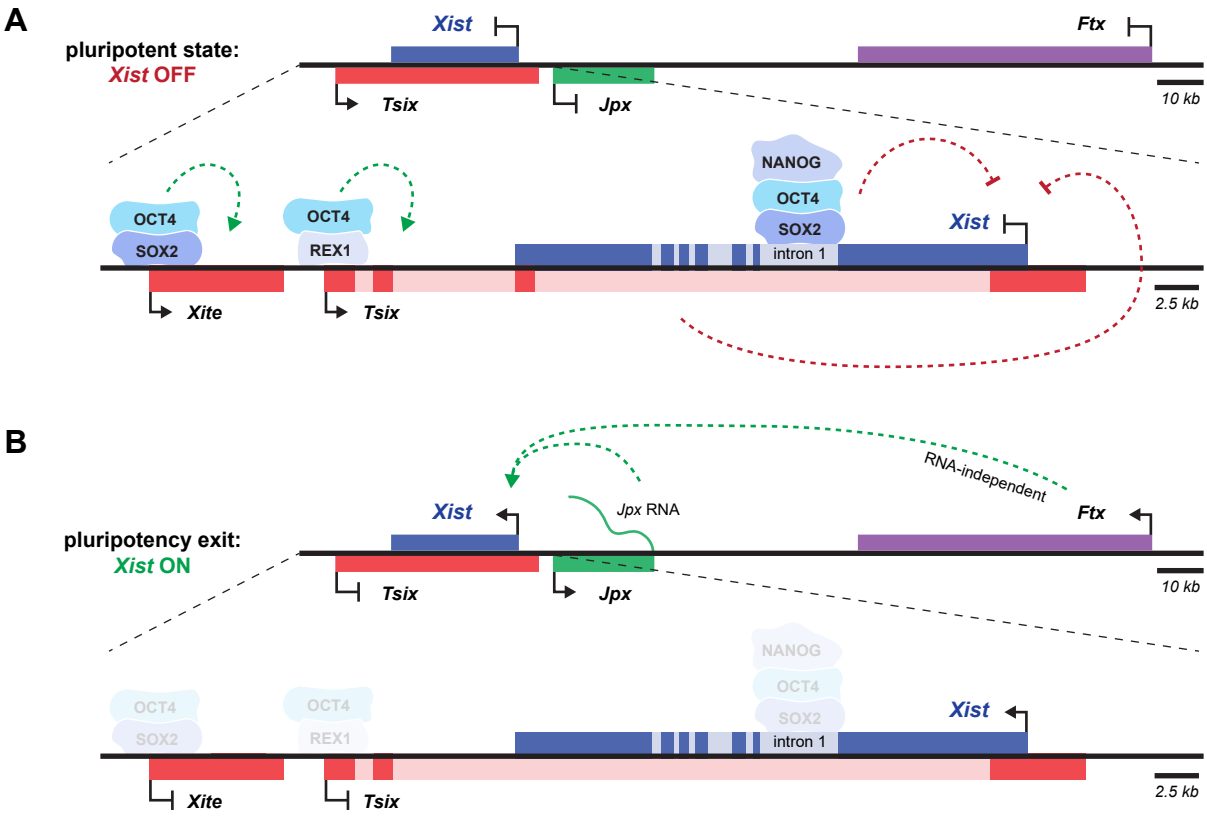


Figure 1.6 – Scheme showing the molecular coupling between *Xist* expression and pluripotency exit. **(A)** At the pluripotent state, *Xist* expression is directly inhibited by OCT4, NANOG and SOX2, which bind at the genomic region coding for *Xist* intron 1. *Xist* is further repressed thanks to the expression of *Tsix*, antisense to *Xist*. *Tsix* transcription is stimulated by binding of OCT4/SOX2 and OCT4/REX1 at its enhancer (*Xite*) and promoter. **(B)** Upon exit from pluripotency, the loss of pluripotency factors results in *Tsix* downregulation and *Xist* upregulation. *Xist* expression is further promoted by the *Jpx* and *Ftx* non-coding RNA, although *Ftx* exerts its *Xist*-promoting function in an RNA-independent manner.

Most of *Xist*'s *cis*-regulators lie within the *Xic* locus and include several non-coding RNAs (e.g. *Tsix* [Lee et al., 1999], *Jpx* [Tian et al., 2010], *Ftx* [Chureau et al., 2011; Furlan et al., 2018]), which have been shown to act either as activators or repressors of *Xist* expression (**Figure 1.6**). The most notable negative *cis*-regulator of *Xist* is the *Tsix* non-coding RNA, which overlaps and is transcribed antisense to the *Xist* gene (**Figure 1.6**) [Lee et al., 1999]. *Tsix* expression prior to differentiation is required to keep *Xist* expression at very low levels in mESCs [Luikenhuis et al., 2001; Stavropoulos et al., 2001], and *Tsix* deletion on one X chromosome results in skewed inactivation of that chromosome [Lee et al., 1999].

Several *trans*-acting factors have also been identified as playing crucial roles in regulating *Xist* expression. Importantly, pluripotency factors including OCT4, NANOG, SOX2 and REX1 have been identified as inhibitors of *Xist* expression, either through activation of *Tsix* transcription or direct repression of *Xist* itself (**Figure 1.6**) [Dono-

hoe et al., 2009; Gontan et al., 2012; Navarro et al., 2008, 2010]. Hence, these studies provided a molecular basis for the temporal link observed between pluripotency exit and upregulation of *Xist* expression (**Figure 1.6B**).

1.7 Silencing kinetics and escape from XCI

Although XCI ultimately affects almost the whole X chromosome, assays to detect allele-specific transcription from the X chromosome in a time-resolved fashion (such as RNA FISH and RNAseq) in differentiating mouse ESCs and pre-implantation embryos have revealed that the kinetics of gene silencing along the X varies in a gene-specific fashion [Borensztein et al., 2017b; Marks et al., 2015; Patrat et al., 2009]. Indeed, some genes are silenced very early during development (e.g. *Rnf12* and *G6pdx*, over the course of a few hours), while others are silenced much later (e.g. *Huwe1*, over the course of several days). How genes across the inactivating X chromosome achieve such differences in gene silencing kinetics remains poorly understood.

Remarkably, some genes, called "escapees", do not undergo silencing during XCI, and remain actively expressed from the Xi in somatic cells. The number of escaping genes varies extensively between tissues, individuals, and species [Berletch et al., 2015]. In humans, more than 15% of X-linked genes escape XCI, while only 3-7% do in mice [Carrel et al., 1999; Disteche and Berletch, 2015; Yang et al., 2010]. Two categories of escapees are to be distinguished. Constitutive escapees evade XCI throughout all stages of development, and in all tissues. On the other hand, facultative escapees show signs of silencing initially during development, but become reactivated in a tissue-specific manner [Berletch et al., 2011, 2015; Giorgetti et al., 2016; Patrat et al., 2009].

How and why some genes manage to escape from XCI is an open question, and whether the mechanisms involved differ between constitutive and facultative escape remain to be answered. Given that biallelic expression of these genes should result in higher protein dosage in females compared to males, it is proposed that escapees have female-specific functions. Another hypothesis is that X-linked genes with functional Y-chromosome homologs (e.g. those lying within the pseudoautosomal region) have evolved to escape so as to ensure their correct dosage. Given the high numbers of genes escaping XCI in humans, escape has been proposed to be implicated in the phenotypes associated with sex chromosome aneuploidies [Carrel and Brown, 2017].

1.8 Chromatin changes associated with XCI and the formation of a repressive compartment

Several studies conducted in both mESCs and developing embryos have unraveled the series of events associated with XCI. Upon *Xist* upregulation and coating of the X chromosome, the earliest event detected is the exclusion of RNA polymerase II (RNAPII) and several of its associated transcription factors (**Figure 1.7**) [Chaumeil et al., 2006]. A repressive nuclear compartment is formed, where silenced and escapee genes respectively locate inside and at the periphery of this compartment (**Figure 1.7**) [Chaumeil et al., 2006].

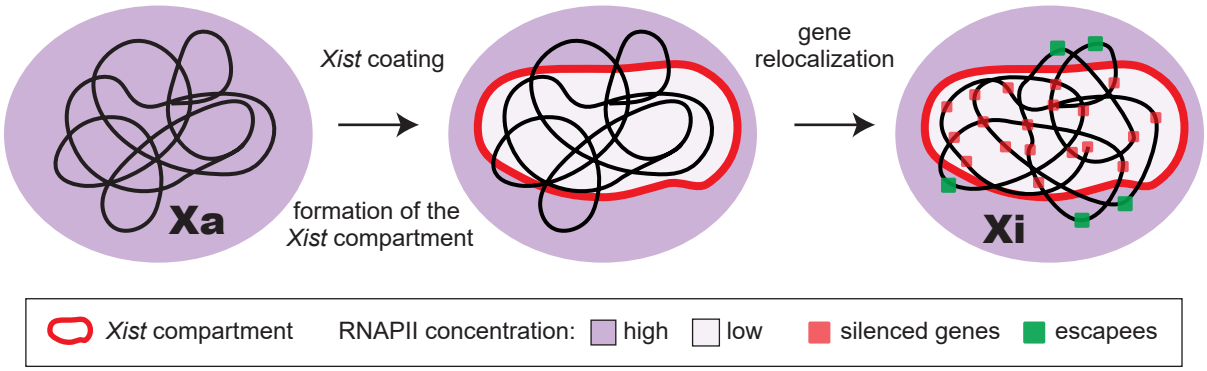


Figure 1.7 – Schematic of *Xist* RNA compartment formation during XCI. Silenced X-linked genes relocate inside the *Xist* compartment, while escaping genes remain at its periphery.

In parallel, the inactivating X chromosome undergoes drastic chromatin changes (**Figure 1.8**) [Lucchesi et al., 2005]:

- active euchromatic histone marks become depleted: H4ac/H3K9ac/H3K27ac [Jeppesen and Turner, 1993] and H3K4me1/3 [Boggs et al., 2002] (**Figure 1.8**),
- repressive heterochromatic histone marks become enriched: H3K9me2 [Heard et al., 2001; Keniry et al., 2016], H4K20me1 [Kohlmaier et al., 2004], and the two Polycomb marks H3K27me3 [Plath et al., 2003; Silva et al., 2003] and H2AK119ub1 [de Napoles et al., 2004] (**Figure 1.8**), catalysed by *P*olycomb *r*epressive *c*omplexes 2 (PRC2) and 1 (PRC1) respectively.

At later stages of the XCI process, nucleosomes on the Xi become specifically enriched for the macroH2A histone variant [Costanzi and Pehrson, 1998], and DNA methylation accumulates at the promoter regions of inactive X-linked genes [Norris et al., 1991] (**Figure 1.8**).

Whether (and how) some of these epigenetic changes contribute to gene silencing during XCI, and what are their relative appearance/disappearance kinetics following *Xist*

expression has been investigated in the context of a project to which I participated at the beginning of my doctoral studies [Żylicz et al., 2019]. This study will be further discussed in **chapter 5** of the results part of my dissertation.

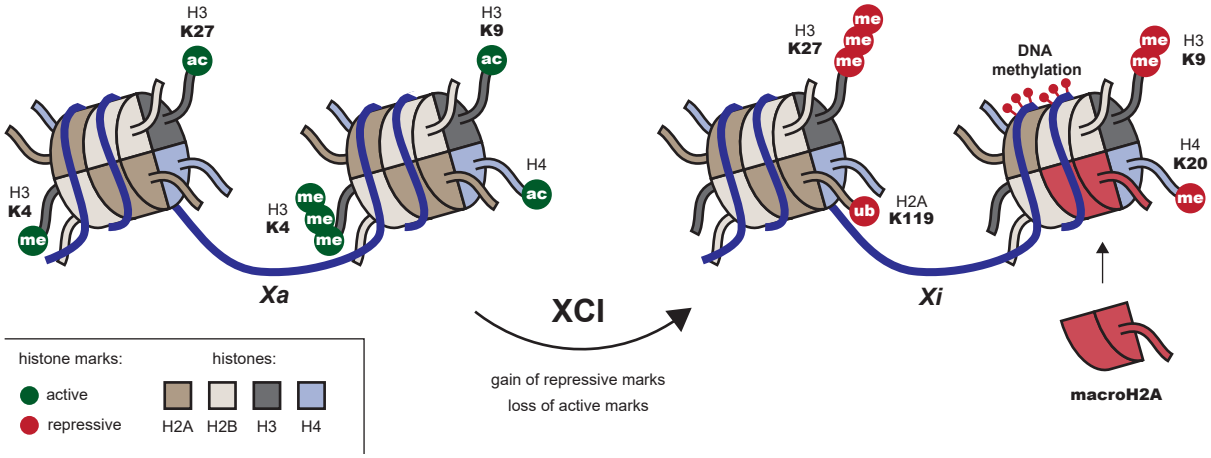


Figure 1.8 – Scheme depicting the chromatin changes occuring during XCI.

1.9 Maintenance of X-chromosome inactivation

During mESC diffentiation, XCI can be separated in two stages: an *initiation* stage, restricted to early differentiation, during which *Xist* expression is essential to establish gene silencing, and a *maintenance* stage, during which *Xist* expression becomes dispensable, and gene silencing is epigenetically maintained [Brown and Willard, 1994; Wutz and Jaenisch, 2000]. Different pathways have been implicated in ensuring that genes remain stably repressed following *Xist* expression and coating.

DNA methylation seems to be important for maintaining *random* XCI specifically, as embryos deficient for *Dnmt1* show reactivation of the Xi in embryonic lineages, while extraembryonic tissues remain unaffected [Sado et al., 2000]. Additionally, an ENU screen in mouse revealed that the *Smchd1* gene is required for maintenance of Xi silencing [Blewitt et al., 2005], with *Smchd1* homozygous-mutant mice showing female specific lethality [Blewitt et al., 2008]. Given that *Smchd1* mutant embryos show hypomethylated promoters on the Xi [Blewitt et al., 2008], and that a plant homolog of SMCHD1 is involved in RNA-dependent DNA methylation [Kanno et al., 2008], these studies further support the idea that DNA methylation plays a crucial role in ensuring maintenance of random XCI.

Intriguingly, maintenance of *imprinted* XCI in extraembryonic tissues is independent of DNA methylation [Sado et al., 2000] and SMCHD1 [Blewitt et al., 2008]. Instead, it seems to rely on PRC2 activity [Wang et al., 2001], suggesting that the repressive

pathways at play during maintenance of XCI differ based on cellular and developmental context.

In line with such context-dependence, gene silencing on the Xi of somatic cells appears to be extremely stable, with simultaneous perturbations of DNA methylation, histone hypoacetylation, and *Xist* expression only leading to partial reactivation of a select few genes [Csankovszki et al., 2001].

1.10 3D structure of the inactive X chromosome

In 1949, Murray Barr and Ewart Bertram reported that upon staining with basic dyes, female cat neurons frequently display a small but densely stained nuclear body (originally termed nucleolar satellite due to its close proximity with the nucleolus), that is however hardly visible in males [Barr and Bertram, 1949]. Barr even proposed that such cytological dichotomy between males and females, in somatic cells, could readily be used to sex animals [Barr and Bertram, 1949]. This nucleolar satellite was later identified to be the inactive X chromosome [Lyon, 1961, 1962; Ohno et al., 1959], and is now commonly referred to as the Barr body.

It became clear from these earlier studies (as well as later ones) that the inactive X chromosome undergoes spatial reorganization to such an extent that its 3-dimensional structure differs significantly from that of its active X counterpart [Eils et al., 1996; Rego et al., 2008; Smeets et al., 2014].

The advent of chromosome conformation capture techniques over the past decade now enables delving into chromosome organization at a hitherto unreached molecular level, allowing to probe spatial positioning of genomic loci relative to one another, chromosome wide [Kempfer and Pombo, 2019]. These methods revealed the existence of topologically associating domains (TADs), corresponding to sub-megabase genomic regions within which loci interact more frequently with one another than they do with loci outside [Dixon et al., 2012; Nora et al., 2012]. Remarkably, TADs are conserved across cell types and species, and are proposed to allow spatial segregation of genes which are co-regulated [Dixon et al., 2012; Nora et al., 2012], although the extent to which they influence transcriptional regulation remains to be addressed.

Importantly, allele-specific conformation capture approaches revealed that contrary to the Xa, the inactive X chromosome in differentiated cells is globally devoid of TADs [Giorgetti et al., 2016; Minajigi et al., 2015; Splinter et al., 2011], except at clusters of escapee genes, where preferential interactions between such genes results in the detection

of TAD-like structures (**Figure 1.9**) [Giorgetti et al., 2016; Splinter et al., 2011].

Remarkably, structural analyses across humans, macaques and mice revealed that the entire Xi folds into two megadomains, which are separated by a boundary element containing the *DXZ4/Dxz4* macrosatellite (**Figure 1.9**), whose deletion results in megadomains fusion [Darrow et al., 2016; Deng et al., 2015; Giorgetti et al., 2016; Minajigi et al., 2015; Rao et al., 2014].

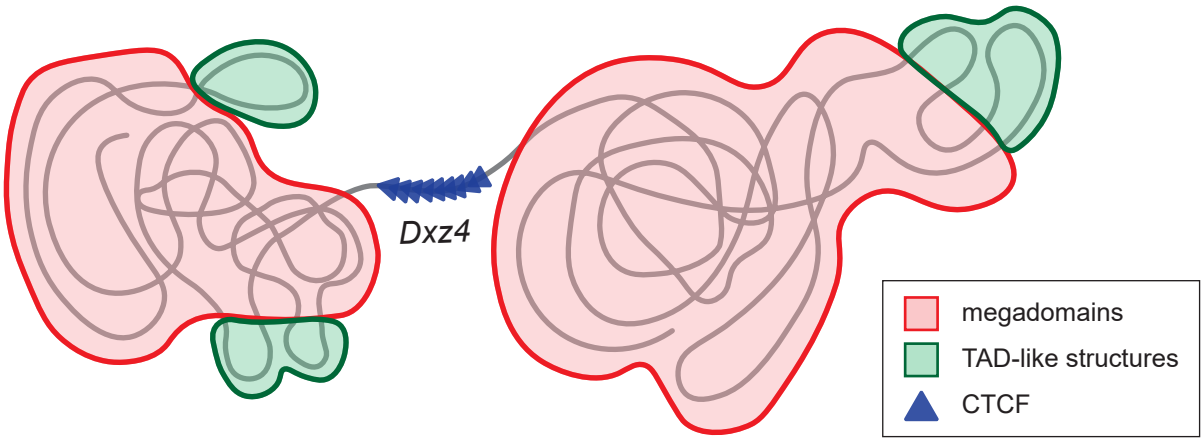


Figure 1.9 – Scheme depicting the 3D structure of the inactive X chromosome. The inactive X chromosome adopts a peculiar three-dimensional structure, characterized by two megadomains separated by the *Dxz4* boundary element, harboring strong CTCF binding. While the megadomains show poor local structure, some level of organization is retained at clusters of escaping genes, which form TAD-like structures.

How the Xi undergoes such peculiar structural reorganization remains poorly understood. The observations that *Xist* deletion results in an Xi conformation that ressembles that of the Xa [Minajigi et al., 2015; Splinter et al., 2011] suggests that *Xist* could play a key role, potentially through the repulsion of architectural factors involved in stabilizing chromosome conformation [Minajigi et al., 2015].

1.11 XCI as a powerful system to study epigenetics

At this stage of the chapter, it appears evident that XCI is a very powerful model for epigenetics:

1. XCI is characterized by chromosome-wide transcriptional changes, chromatin modifications (including DNA methylation, histone marks and histone variants), and spatial reorganization. Given that all these events happen in concert, XCI provides an ideal context to dissect the interplay and interdependencies between these key epigenetic processes.

2. All these events are triggered by *Xist*, the master regulator of XCI. Hence, XCI also constitutes a paradigm for the study of long non-coding RNA-mediated processes.
3. Although XCI is often modeled *in vitro* (i.e. using embryonic stem cells), XCI is a key developmental process, taking place systematically in all female – Therian – mammal embryos. Therefore, dissecting the mechanisms of XCI is relevant to understand development (and diseases to some extent).
4. XCI is particularly amenable to a plethora of quantitative analyses given that all aforementioned molecular events happen chromosome-wide, across thousands of loci, thus providing great power for statistical analyses.

1.12 Early hints into *Xist*-dependent gene silencing mechanisms

Given that it is following coating by *Xist* RNA that the X chromosome undergoes gene silencing and "epigenetic reorganization", most of the research investigating the mechanisms driving XCI have relied on dissecting the function of *Xist* RNA itself.

In a landmark study, Anton Wutz and Rudolf Jaenisch showed that *Xist* RNA is organized in distinct functional "repeat" modules (**Figure 1.10**) [Wutz et al., 2002]. A conserved sequence motif called the *A-repeat*, located towards the 5' end of *Xist* (**Figure 1.10**), is required for its gene-silencing function both *in vitro* and *in vivo* [Wutz et al., 2002; Sakata et al., 2017]. On the other hand, several dispersed sequences, located downstream of the A-repeat, mediate *Xist*'s ability to localize and accumulate (i.e. coat) in *cis* [Wutz et al., 2002]. The regions involved in such localization are functionally redundant and map to domains comprised within the F- [Wutz et al., 2002], C- [Jeon and Lee, 2011] and E-repeats [Sunwoo et al., 2017; Ridings-Figueroa et al., 2017; Yamada et al., 2015] of *Xist* (**Figure 1.10**).

Figure 1.10 – Scheme of the mouse *Xist* spliced transcript highlighting the position of the A-F repeat regions. The regions involved in chromosomal silencing are colored in red while the regions involved in *Xist* localization are colored in blue.

This shed light on the fact that much of the mechanisms at play in inducing gene silencing during XCI occur through the A-repeat. The A-repeat was further shown by

the group of Jeannie Lee to interact directly with EZH2, the catalytic subunit of the PRC2 complex, responsible for histone H3K27 trimethylation [Zhao et al., 2008]. In parallel, other groups reported that two components of the PRC2 complex (EZH2 and EED) are actively "recruited" to the Xi, following *Xist* expression, leading to a strong accumulation of H3K27me3 on the Xi both *in vitro* and *in vivo* [Kohlmaier et al., 2004; Mak et al., 2002, 2004; Plath et al., 2003; Silva et al., 2003].

Considering that polycomb complexes are associated with transcriptional repression [Morey and Helin, 2010], Jeannie Lee proposed that through direct binding of PRC2 to the A-repeat, *Xist* allows chromosome wide recruitment of PRC2 and spreading of H3K27me3, resulting in gene silencing [Lee, 2012; Lee and Bartolomei, 2013; Pinter et al., 2012]. Several lines of evidence however have since seriously called such model into question:

First, a direct interaction between PRC2 and *Xist* RNA is challenged by reports that:

- The PRC2 complex binds RNA promiscuously, with no apparent specificity for *Xist* as compared to other transcripts (including irrelevant transcripts of bacterial origin) [Davidovich et al., 2013, 2015].
- Super-resolution microscopy reveals significant spatial separation between PRC2 proteins and *Xist* RNA [Cerase et al., 2014], to a level that is incompatible with direct PRC2/*Xist* interaction.
- Loss of the A-repeat does not have any dramatic effects on PRC2 recruitment upon $Xist^{\Delta A\text{-}repeat}$ coating [da Rocha et al., 2014; Kohlmaier et al., 2004; Wutz et al., 2002].

Second, initiation of random XCI proceeds normally in embryos and mESCs deficient for the core PRC2 component EED [Kalantry and Magnuson, 2006; Schoeftner et al., 2006].

Hence, although PRC2 had originally appeared as the obvious suspect, these conflicting pieces of evidence left the field with no other tangible factors to explain *Xist*'s gene silencing function.

1.13 *Xist* protein interactome mapping: a revolution

The most intuitive way through which the community envisioned *Xist* RNA coating could trigger chromosome wide gene silencing was through direct recruitment of *bona fide* transcriptional repressors, by *Xist* RNA itself. Identifying the repertoire of *Xist* RNA

binding proteins would therefore be the most direct strategy to decipher the gene silencing mechanisms at play following *Xist* coating. Until recently, this proved to be technically challenging.

In 2015 however, three groups (Howard Chang's, Mitch Guttman's and Jeannie Lee's) independently developed similar approaches allowing for affinity purification of crosslinked *Xist* RNA/protein complexes, followed by mass-spectrometry based identification of *Xist*-bound proteins [Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015]. These three studies will hereafter be referred to as the "proteomic studies". In parallel, two other groups (Neil Brockdorff's and Anton Wutz's) employed high throughput genetic screening in mouse embryonic stem cells to identify protein factors involved in *Xist*-dependent gene silencing [Moindrot et al., 2015; Monfort et al., 2015]. These two studies will hereafter be referred to as the "genetic screens".

Altogether, these five different studies reported several dozens of factors that bind *Xist* RNA and/or are involved in mediating gene silencing during XCI, opening exciting research avenues to further understand the mechanisms of XCI.

Importantly, none of the PRC2 core-components were retrieved, neither in the *Xist*-pulldowns [Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015] nor in the genetic-screen studies [Moindrot et al., 2015; Monfort et al., 2015], further invalidating the previously proposed model of PRC2 recruitment and function in XCI (see **section 1.12**).

Remarkably, the majority of factors identified to interact with *Xist* do not carry chromatin modifying activities and are associated with pathways that were hitherto not suspected to be important for XCI. These discoveries led to a shift in the view *Xist* functions, from directly recruiting chromatin modifying activities, to a model where *Xist* first recruits scaffolding proteins which can then associate with such activities.

In the following paragraphs, I will briefly discuss the three main factors/pathways which were found to interact with *Xist* RNA in these recent proteomic studies, and which were additionally reported to play a role in gene silencing during XCI in more than one report. Then, in **chapter 2** of this introduction, I will delve in greater depths into the SPEN protein, whose function in XCI has been the main focus of my PhD research. Importantly, SPEN is the only factor found to interact with *Xist* consistently across all three proteomic studies [Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015], as well as a top candidate for *Xist*-mediated silencing in both genetic screens [Moindrot et al., 2015; Monfort et al., 2015].

It is worthy to note that several other proteins have been shown to bind *Xist* and influence its localization (i.e. ability to coat), as well as its ability to organize the 3D

structure of the inactive X chromosome. These proteins will not be described here, but have been comprehensively reviewed in [Loda and Heard, 2019] and [Da Rocha and Heard, 2017].

1.13.1 hnRNPK resolves the Polycomb controversy

The *h*eterogeneous *n*uclear *ribo**n*uclear *p*rotein *K* (**hnRNPK**) was identified as an *Xist* interactor in two of the proteomic studies [Chu et al., 2015; Minajigi et al., 2015], and its binding site appeared to be located outside of *Xist*'s A-repeat [Chu et al., 2015]. Consistently, the Brockdorff lab later showed that hnRNPK interacts directly with a region comprised within the B/C repeats of *Xist* RNA (**Figure 1.11**) [Pintacuda et al., 2017].

Deleting this region from *Xist* leads to complete loss of PRC1 and PRC2 recruitment – as well as accumulation of their associated histone marks H2AK119ub1 and H3K27me3 – during XCI [Pintacuda et al., 2017; Bousard et al., 2019; Colognori et al., 2019]. Consequently, this region of *Xist* was named PID (**Figure 1.11**), for *P*olycomb *i*nteraction *d*omain. Importantly, knockdown of *Hnrnpk* phenocopied the Polycomb recruitment defects associated with deletion of *Xist*-PID [Chu et al., 2015; Colognori et al., 2019].

Remarkably, hnRNPK was found to interact directly with PCGF3 and PCGF5, two components of the non-canonical PCGF3/5-PRC1 complex (**Figure 1.11**), and this interaction was found to be necessary for *Xist*-mediated PRC1 recruitment [Pintacuda et al., 2017]. Consistently, RYBP (another component of the non-canonical PRC1 complex) and RING1B (the catalytic subunit of PRC1, responsible for H2AK119ub1 deposition) were retrieved in two of the *Xist* proteomic studies [Chu et al., 2015; Minajigi et al., 2015].

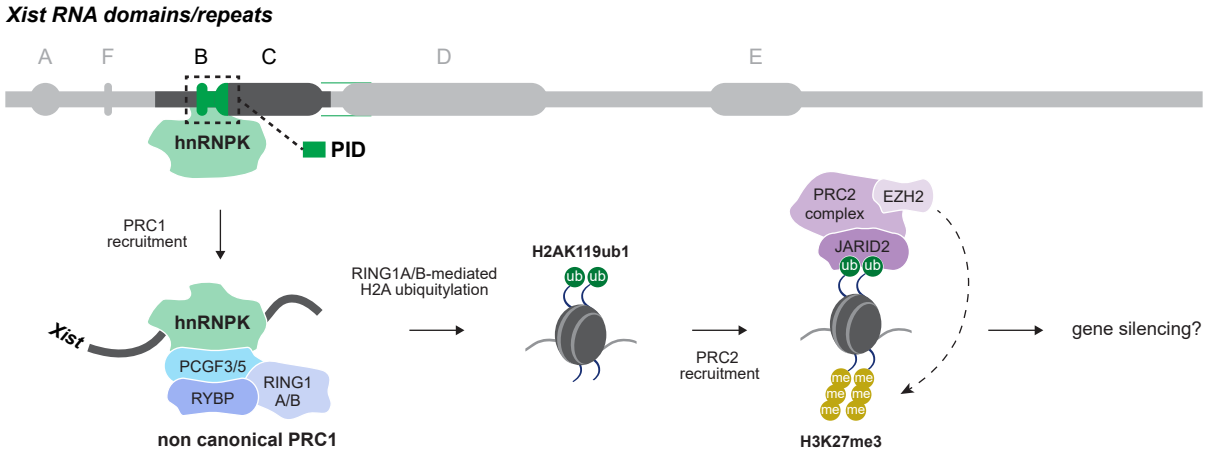


Figure 1.11 – Model for B-repeat/hnRNPK-dependent Polycomb recruitment during XCI. The PID region of *Xist* – which binds hnRNPK – is colored in green.

Furthermore, loss of PCGF3/5-dependent H2AK119ub1 resulted in deficient PRC2 recruitment (and H3K27me3 accumulation) during XCI [Almeida et al., 2017], suggesting that, contrary to what was proposed previously (see **section 1.12**), PRC2 is not directly recruited by *Xist*, and is instead dependent on antecedent *Xist*/hnRNPK-mediated PRC1 recruitment to the X chromosome (**Figure 1.11**).

This model was further supported by evidence that PRC1-mediated H2A ubiquitylation is required for PRC2 recruitment [Blackledge et al., 2014], through direct binding of the PRC2 cofactor JARID2 to H2AK119ub1 (**Figure 1.11**) [Kalb et al., 2014; da Rocha et al., 2014; Cooper et al., 2016].

hnRNPK-mediated recruitment of PRC1 has been proposed to play an important role in *Xist*-mediated gene silencing during XCI, as gene silencing defects are observed upon deletion of *Xist*-PID [Pintacuda et al., 2017; Colognori et al., 2019; Bousard et al., 2019; Nesterova et al., 2019] or loss of PCGF3/5 [Almeida et al., 2017; Nesterova et al., 2019]. In fact, *Pcgf3/5*-KO female mouse embryos show earlier lethality than their male counterparts, and fail to fully undergo dosage compensation [Almeida et al., 2017]. However, the magnitude of the reported gene silencing phenotypes varies greatly between studies, likely reflecting the heterogeneity in cellular systems used to model XCI, as well as the timepoints at which XCI was assayed.

Clearly, all these studies report that silencing still occurs very significantly despite the absence of polycomb recruitment. Furthermore, a gene silencing defect (even if partial) upon loss of PRC1 is hard to reconcile with the fact that deleting *Xist*'s A-repeat (which is physically unlinked from *Xist*'s PID region responsible for PRC1 recruitment) results in fully deficient gene silencing [Wutz et al., 2002; Sakata et al., 2017; Nesterova et al., 2019].

The recent observation that *Xist* RNA fails to properly localize and coat the X chromosome either when deleted of its PID region or when PRC1-components are perturbed [Colognori et al., 2019] could resolve this inconsistency. In such case, the gene silencing defects associated with loss of Polycomb recruitment are most likely an indirect repercussion of impaired *Xist*-coating [Colognori et al., 2019].

1.13.2 LBR tethers the Xi to the nuclear lamina

It has long been described that the condensed Xi is most often localized at the nuclear periphery, in contact with the nuclear envelope [Ohno et al., 1958; Rego et al., 2008; Bourgeois et al., 1985; Dyer et al., 1989; Barton et al., 1964] or around the nucleolus [Barr and Bertram, 1949; Rego et al., 2008; Bourgeois et al., 1985; Zhang et al., 2007]. How

the Xi preferentially becomes localized in the vicinity of the nuclear membrane however was unknown, and whether this peculiar localization is necessary for or consequential to gene silencing remained an open question.

Surprisingly, the *l*amin *B* *r*eceptor (**LBR**) was identified as a *Xist*-binding protein in two of the proteomic studies [McHugh et al., 2015; Minajigi et al., 2015]. Given that LBR is an inner nuclear transmembrane protein, which anchors the lamina and heterochromatin to the nuclear envelope [Nikolakaki et al., 2017], these studies raised the exciting possibility that such *Xist*/LBR interaction could be responsible for localizing the X chromosome to the nuclear periphery. Furthermore, they finally offered the opportunity to address whether such localization is important for gene silencing.

In a follow up study, the Guttman lab showed that through its arginine-serine (RS) rich motif (**Figure 1.12**), LBR binds *Xist* RNA at three distant *L*BR *b*inding *s*ites (LBS), with most prominent binding observed over a region encompassing the 3'-end of the A-repeat as well as the entire F-repeat of *Xist* (**Figure 1.12**) [Chen et al., 2016]. Deletion of the major *Xist*-LBS was shown to abolish the recruitment of the Xi to the lamina (i.e. nuclear periphery) and majorly disrupt *Xist*-mediated gene silencing; and similar phenotypes were consistently reported upon *Lbr* knockdown or knockout [Chen et al., 2016]. Remarkably, "artificial" tethering of *Xist* (deleted of its major LBS) to the lamina rescued gene silencing, suggesting that LBR-mediated recruitment of *Xist* to the nuclear lamina is required for *Xist*-mediated gene silencing [Chen et al., 2016].

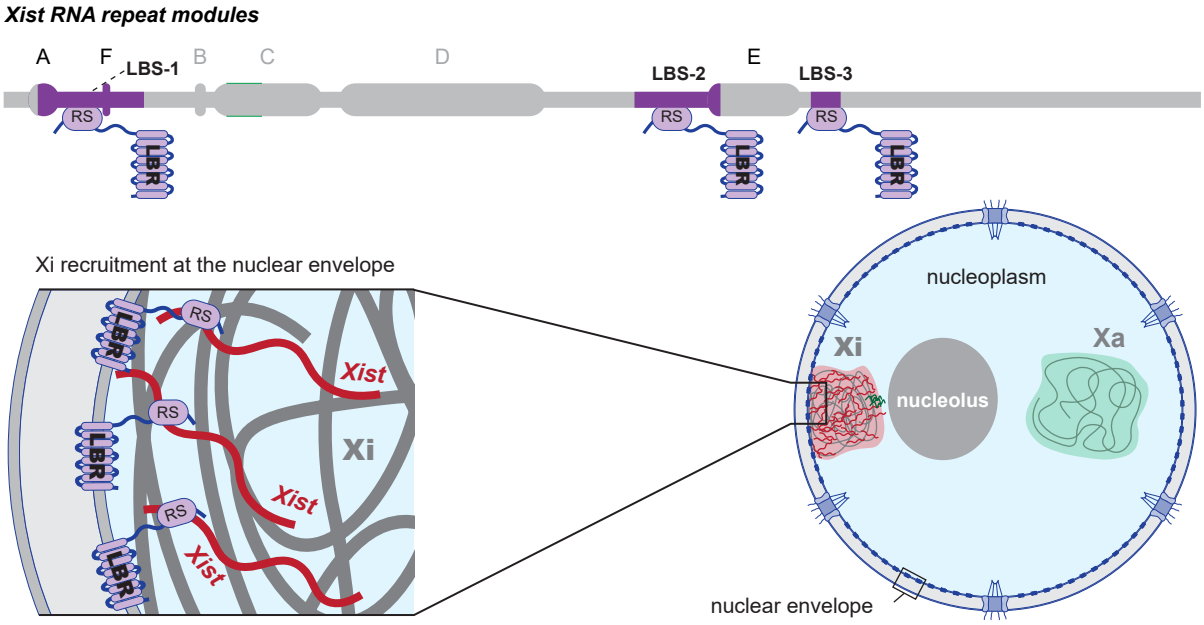


Figure 1.12 – Model for LBR-mediated recruitment of *Xist* and the X chromosome to the nuclear lamina. The LBS regions of *Xist* – which bind the RS motif of LBR – are colored in purple.

Importantly in this study, only 5 X-linked genes were assayed as a proxy for XCI, using RNA FISH [Chen et al., 2016; McHugh et al., 2015]. In a recent report however, in which X-linked transcription was assayed chromosome-wide using RNAseq, the Brockdorff lab showed that *Lbr* knockout – or deletion of the LBS-1 region of *Xist* – resulted in very minor gene silencing defects [Nesterova et al., 2019]. This new study hence suggests that LBR only plays a negligible role in *Xist*-mediated silencing, and that the association of the X chromosome with the nuclear lamina is not crucial for XCI [Nesterova et al., 2019].

1.13.3 RBM15 and WTAP link *Xist* with m^{6}A RNA methylation

Cellular RNAs are subject to a wide array of chemical modifications, among which N^6-methyladenosine (m^{6}A) has recently emerged as an important player in gene expression regulation by affecting mRNA homeostasis at several levels including splicing, stability, localization and translation [Roundtree et al., 2017].

m^{6}A methylation is catalyzed by a nuclear "writer" complex composed of the METTL3/METTL14 methyltransferases and the adaptor proteins WTAP, RBM15, ZC3H13, VIRMA (KIAA1429) and CBLL1 (HAKAI) [Zaccara et al., 2019]. Once deposited, m^{6}A recruits m^{6}A "reader" proteins, among which the YTH domain-containing proteins YTHDC1 and YTHDF1/2/3 – localized in the nucleus and the cytosol respectively – have extensively been shown to affect mRNA fate in their respective subcellular compartments [Zaccara et al., 2019]. In 2015, a role for the m^{6}A RNA methylation pathway during XCI was brought to light (**Figure 1.13**).

Indeed, **RBM15** (*R*NA-*b*inding *m*otif protein 15) and **WTAP** (*W*ilms *t*umor 1 *a*ssociated *p*rotein) were found to interact with *Xist* – seemingly through its A-repeat – in two of the proteomic studies (**Figure 1.13**) [McHugh et al., 2015; Chu et al., 2015]. Furthermore, RBM15 and WTAP were both identified as candidates in one of the two genetic screens, and their knockdown was associated with defects in *Xist*-mediated gene silencing [Moindrot et al., 2015].

Follow up studies showed that *Xist/XIST* is highly m^{6}A-methylated in mouse and humans respectively [Nesterova et al., 2019; Patil et al., 2016], and that binding of RBM15 to *Xist* RNA is required for m^{6}A-methylation of *Xist* (**Figure 1.13**) [Patil et al., 2016]. Patil et al. further showed that RBM15 (and its paralog RBM15B) recruits the m^{6}A-methylation complex to *Xist* by interacting directly with WTAP (**Figure 1.13**).

Importantly, knockdowns of *Mettl3*, *Wtap* or *Rbm15/Rbm15b* were shown to result in deficient *Xist*-mediated gene silencing, suggesting that *Xist* RNA methylation is required for gene silencing during XCI [Patil et al., 2016]. This effect was shown to be

mediated by the binding of the YTHDC1 reader protein to methylated *Xist* [Patil et al., 2016], although the mechanistic basis for YTHDC1-mediated X-linked gene silencing is completely unknown (**Figure 1.13**).

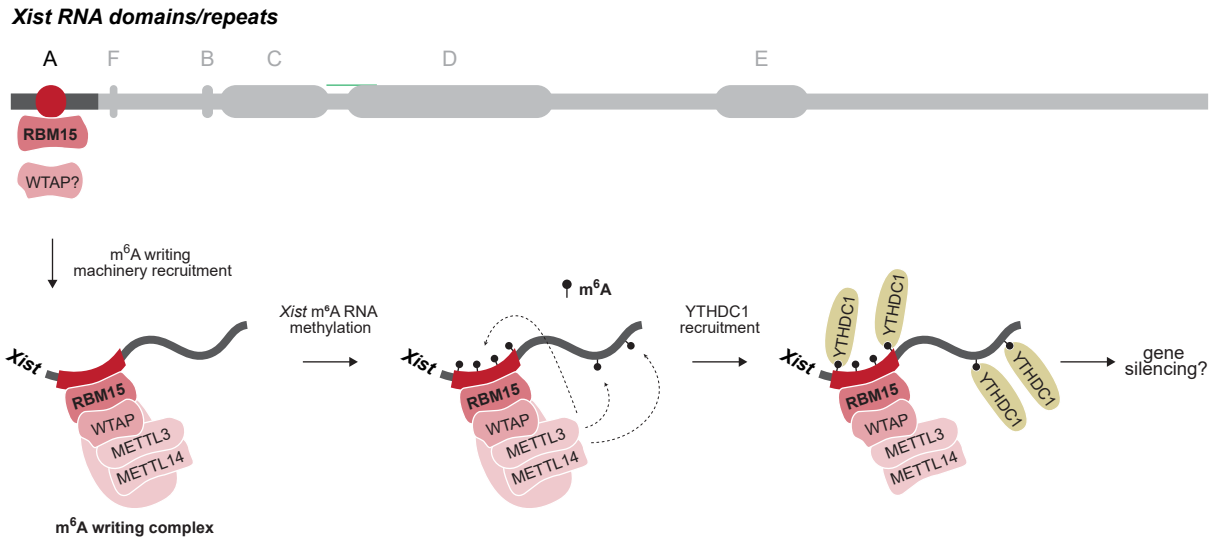


Figure 1.13 – Model for A-repeat dependent recruitment of the m^{6}A RNA methylation machinery. RBM15 binding to *Xist* A-repeat leads to recruitment of the m^{6}A writing machinery and subsequent *Xist* RNA methylation. Binding of the m^{6}A reader YTHDC1 allegedly plays a role in X-linked gene silencing during XCI.

Similar to what was discussed previously in **subsection 1.13.2** about LBR, the evidence for a major role of *Xist* RNA methylation in gene silencing was solely based on RNA FISH assays of 2 X-linked genes. Recently however, the Brockdorff lab used RNAseq to show that gene silencing chromosome wide is only minorly affected upon knockouts of *Rbm15*, *Wtap* or *Mettl3* [Nesterova et al., 2019]. Consistently, they also reported no dramatic silencing defects upon deletion of major m^{6}A sites within *Xist* [Nesterova et al., 2019; Coker et al., 2020]. Taken together, these results strongly suggest that *Xist* RNA methylation is largely dispensable for gene silencing during XCI.

Chapter 2

SPEN: Discovery and function

Although identified more than 20 years ago, the SPEN protein remains very poorly characterized. A few studies, performed across different model organisms, have revealed some of its functions and provided hints into its mechanism of action. In this chapter, I intend to present these studies succinctly and highlight the state of knowledge on SPEN at the time I initiated my PhD work in 2017.

2.1 First discovery of SPEN in *Drosophila*

The gene *Spen* (*Split-ends*) was originally identified in *Drosophila* embryos, where its mutation was associated with sensory axon growth defects [Kolodziej et al., 1995]. *Drosophila Spen* was shown to encode a 600kDa nuclear protein, containing three *R*NA *r*ecognition *m*otifs (RRMs) at its N-terminus and five nuclear localization signals distributed within its center region [Wiellette et al., 1999].

Further genetic studies in *Drosophila* showed that *Spen* regulates signaling by the Wnt/Wingless [Lin et al., 2003], EGFR [Kuang et al., 2000; Rebay et al., 2000] and Notch [Doroquez et al., 2007] pathways, through which it is important for cell positioning [Mace and Tugores, 2004], eye development [Doroquez et al., 2007; Querenet et al., 2015], glial cell development [Chen and Rebay, 2000], lipid metabolism [Hazegh et al., 2017] and intestinal stem cell renewal [Andriatsilavo et al., 2018]. How SPEN mechanistically interacts with such signaling pathways, and how it regulates these cellular processes remains unclear from these studies.

2.2 SPEN is conserved in animals

Conservation analysis reveals *Spen* orthologs only across the animal kingdom, all of which encode very large proteins comprising two to four highly conserved N-terminal RRMs, as well as a highly conserved C-terminal domain (**Figure 2.1**).

This C-terminal domain, called SPOC, for *S*pen *p*aralog and *o*rtholog *C*-terminal domain [Wiellette et al., 1999], is found not only in SPEN, but also in smaller SPEN-related proteins, present across fungi, plantae and animals, suggesting that its function has been conserved throughout evolution [Sánchez-Pulido et al., 2004].

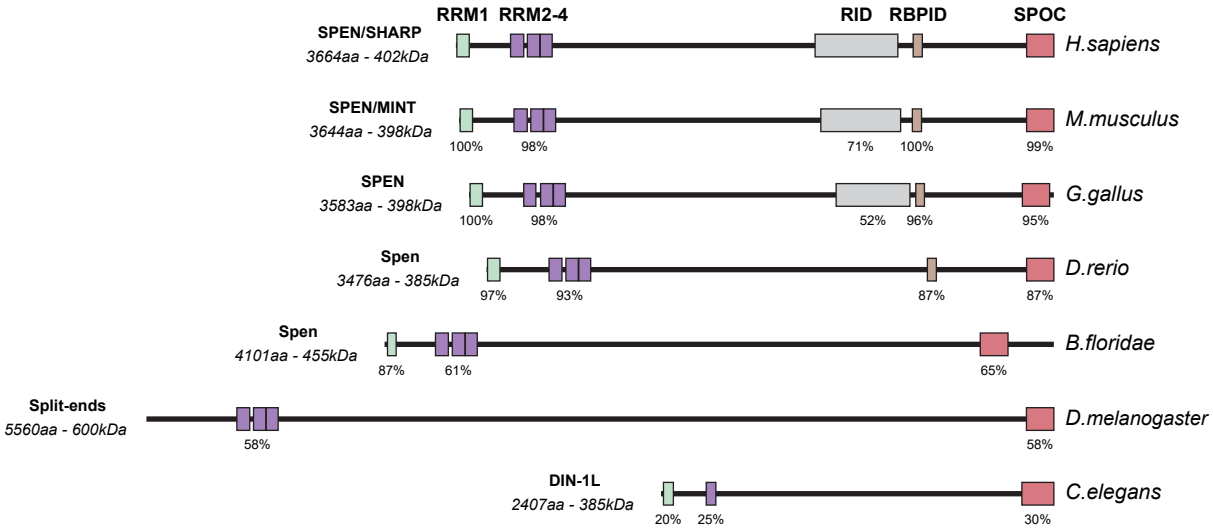


Figure 2.1 – Conservation analysis of SPEN across animals. Each SPEN protein is drawn to scale, and annotated domains are scaled and positioned accordingly. Species names are indicated on the right, and corresponding protein names and sizes on the left. The percent identity of amino acid sequences (compared to *H. sapiens*) is indicated for each domain. RRM: RNA recognition motif, RID: nuclear receptor interacting domain, RBPID: RBPJ-interacting domain, SPOC: Spen paralog and ortholog C-terminal domain. *H. sapiens* (human), *M. musculus* (mouse), *G. gallus* (chicken), *D. rerio* (zebrafish), *B. floridae* (amphioxus), *D. melanogaster* (fruit fly), *C. elegans* (nematode).

2.3 Identification of SPEN/SHARP/MINT in mammals

The first evidence for SPEN function in mammals came from its identification, in rats, as an interactor of MSX2, a transcriptional repressor important for bone development [Newberry et al., 1999]. Through its RRM motifs, SPEN was shown to bind the rat osteocalcin promoter directly, subsequently recruiting MSX2 to repress transcription of the osteocalcin gene [Newberry et al., 1999]. In this early report, SPEN was renamed

MINT, for *M*SX2-*i*nteracting *n*uclear *t*arget.

In 2001, the human homolog of MINT was identified in a yeast two-hybrid screen as an interactor of the nuclear receptor corepressor SMRT, and crucial clues into the function of its RRMs and SPOC domains, discussed below, were provided [Shi et al., 2001]. Human MINT was named SHARP, for *S*MRT/*H*DAC1 *a*ssociated *r*epressor *p*rotein, however, for clarity issues, MINT/SHARP/SPEN will hereafter be referred to as SPEN.

2.4 Functions of SPEN's RRMs

SPEN's RRMs (2, 3 and 4, **Figure 2.1**) were initially reported to interact directly with a non-coding RNA called *SRA* (*s*teroid-receptor *R*NA *a*ctivator, **Figure 2.2**), which plays an important role in the regulation of nuclear-receptor signaling [Shi et al., 2001]. A structural study later revealed that these three RRMs bind both single and double-stranded RNA sequences within *SRA*, and cooperate to ensure binding specificity [Arieti et al., 2014].

The RRM1 domain, located outside of the RRM2-4 block (**Figure 2.1**), seems to be dispensable for binding SRA *in vitro*. Whether RRM1 carries *bona-fide* RNA-binding capacity is unknown, but the fact that RRM1 appeared specifically in chordates, where it is highly conserved (**Figure 2.1**), suggests that it is important for some aspect of SPEN function in mammals.

It is worthy to note that although SPEN's RRMs were originally shown to directly bind DNA at the level of the rat osteocalcin promoter *in vitro* [Newberry et al., 1999], such binding was later found to be fairly weak, and unable to compete with RNA-binding [Arieti et al., 2014]. Hence, it is believed that RNA is the predominant nucleic acid substrate of SPEN's RRMs.

2.5 Functions of SPEN's SPOC domain and links with nuclear receptor signaling

Although previous reports had identified a strong conservation of the SPOC domain across animals [Wiellette et al., 1999; Kuang et al., 2000], its function was brought to light two years later, when the SPOC domain of mammalian SPEN was shown to interact directly with NCoR (NCOR1) and SMRT (NCOR2, **Figure 2.2**) [Shi et al., 2001]. The structural basis of this interaction has been resolved [Ariyoshi and Schwabe, 2003], and

will be further discussed in the final chapter of the results section (**chapter 7**) in the context of a novel interaction between SPOC and RNA polymerase 2 I discovered during my PhD.

NCoR and SMRT are two structurally related founding members of the nuclear receptor corepressor family. These proteins bind unliganded *n*uclear *r*eceptors (NRs) [Lazar, 2003], and mediate transcriptional repression at NR-target genes by recruiting distinct families of histone deacetylases such as HDACs 1/2/3 (class I) and HDACs 4/5/7 (class II) [Zhang et al., 1997; Nagy et al., 1997; Guenther et al., 2001; Huang et al., 2000; Kao et al., 2000].

Remarkably, Shi et al. found that SHARP also binds unliganded nuclear receptors through its *r*eceptor *i*nteraction *d*omain (RID, **Figure 2.2**), and hence proposed that SHARP and NCoR/SMRT could exert coordinated repression through their mutual recruitment at genes bound by unliganded NRs [Shi et al., 2001].

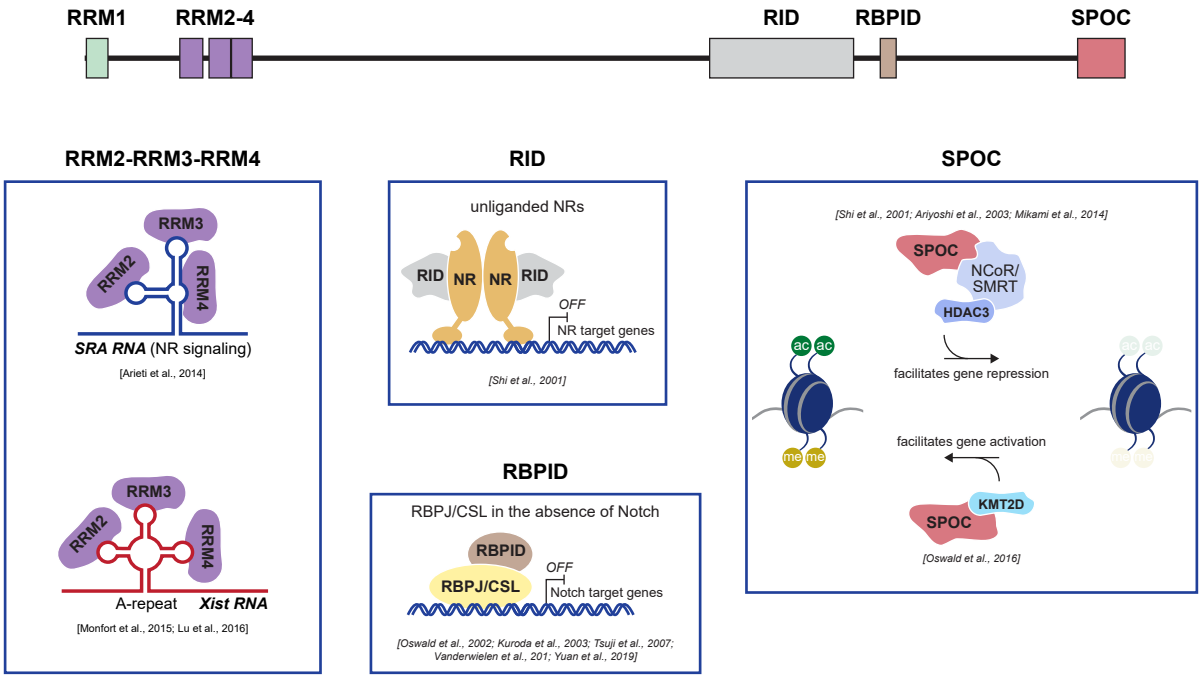


Figure 2.2 – Scheme summarizing the previously reported roles of SPEN's functional domains. SPEN domains are named as in **Figure 2.1**. *NR: nuclear receptor*

2.6 SPEN and Notch signaling

In mice, SPEN was further shown to play a role in thymocyte and splenic B-cell differentiation, through supression of Notch signaling [Kuroda et al., 2003; Tsuji et al., 2007]. Accordingly, *Spen* knock-out mice die during embryonic development between

E13.5 and E14.5, showing morphological abnormalities at the level of the pancreas and the heart, two organs whose genesis and development are heavily regulated by Notch [Kuroda et al., 2003].

SPEN was further found to directly bind RBP-J (the major transcriptional effector of Notch signalling), via its *RBP*J-*i*nteracting *d*omain (RBPID, **Figure 2.2**) [Oswald et al., 2002; Yuan et al., 2019]. Mechanistically, it is proposed that RBPJ tethers SPEN to Notch target genes, where SPEN-mediated recruitment of NCOR1/SMRT leads to transcriptional silencing and suppression of Notch signalling [Oswald et al., 2016; Giaimo et al., 2017; Vanderwielen et al., 2011].

A role for SPEN in Notch signaling suppression has also been reported in *Drosophila* [Doroquez et al., 2007], but whether its mechanistic basis ressembles that observed in mammals is unknown. The fact that *Drosophila* SPEN lacks any sequence homology with the RBPID domain of vertebrate SPEN – despite strong homology between vertebrate RBPJ and its *Drosophila* homolog SU(H) – argues against it.

Intriguingly, one study reported that SPEN can also act as positive regulator of Notch signaling [Oswald et al., 2016]. Indeed, in this report, the authors found that the repressive NCoR/SMRT complex competes with the KMT2D lysine methyltransferase for binding to SPEN's SPOC domain (**Figure 2.2**) [Oswald et al., 2016]. Given that the KMT2D coactivator complex catalyses H3K4 monomethylation (a histone mark associated with active enhancers) and that NCoR/SMRT/HDACs catalyse histone deacetylation (an event associated with transcriptional repression), the authors proposed that SPEN plays a bivalent role in Notch signaling by maintaining a permissive chromatin state around Notch target genes (**Figure 2.2**) [Oswald et al., 2016].

2.7 SPEN and cancer

SPEN is frequently mutated in various types of cancer, including breast cancer [Légaré et al., 2015], ovarian cancer [Légaré et al., 2015], prostate cancer [Armenia et al., 2018], adenoid cystic carcinoma [Stephens et al., 2013; Liu et al., 2017], pancreatic adenosquamous carcinoma [Ma et al., 2020], chronic lymphocytic leukemia [Edelmann et al., 2020], splenic marginal zone lymphoma [Rossi et al., 2012], primary cutaneous marginal zone lymphoma [Maurus et al., 2018], mantle cell lymphoma [Sakhdari et al., 2019] and diffuse large B cell lymphomas [Schmitz et al., 2018; Chapuy et al., 2018].

Studies performed on cancer cell models have highlighted both tumor-suppressing and tumor-promoting effects of SPEN expression depending on cellular context:

In *e*strogen *r*eceptor α (ERα) positive breast cancers (which account for almost 70% of all breast cancers), SPEN loss of function mutations (both homo- and heterozygous) are observed in more than 25% of cases [Légaré et al., 2015]. In such cancers, activation of ERα-signaling is the main oncogenic driver, and SPEN, through its ability to bind unliganded nuclear receptors (see **section 2.5**), seems to play a tumor supressing role by repressing ERα gene targets. Indeed, in the ERα positive T47-D breast cancer cell line, where SPEN is mutated and barely expressed, introduction of a functional SPEN copy suppresses ERα transcription. This SPEN-dependent inhibition of ERα signaling is associated with reduced proliferation and cell survival *in vitro*, and hindered tumorigenic potential in xenograft studies *in vivo* [Légaré et al., 2015].

On the contrary, in ERα negative breast cancers, SPEN seems to play a tumor-promoting role, with high SPEN expression being associated with early metastasis. In such cancers, SPEN promotes primary cilia formation and cellular migration, although the mechanistic basis for such hormone-independent role of SPEN is unclear [Légaré et al., 2017].

A tomor-promoting effect of SPEN was further reported in a human colon cancer model where β-catenin/TCF-mediated transcription is overactivated [Feng et al., 2007]. In such context, stabilized nuclear β-catenin/TCF complexes positively regulate the expression of key genes governing cell fate and proliferation, including the proto-oncogenes *MYC* and *CCND1* (cyclin D1). Through physical interaction with TCF, SPEN was shown to promote such transcription and favor neoplastic transformation in colon cancer [Feng et al., 2007]. Although it is surprising that this study (in addition to [Oswald et al., 2016]) is the only one to show that SPEN can act to promote transcription, one may conceive that its very large size enables interaction with diverse transcription factors involved in both transcriptional activation and repression.

Drosophila and mammalian studies have shown that SPEN interacts genetically and/or physically with the Wnt, Notch, EGFR and NRs signaling pathways, all of which are implicated in cancer [Zhan et al., 2017; Aster et al., 2017; Normanno et al., 2006; Dhiman et al., 2018]. Future studies will most likely reveal that SPEN also plays important roles in Notch and EGFR-dependent cancers.

2.8 SPEN in other animal models

Apart from *Drosophila* and mammalian models, SPEN is very poorly characterized in other animal systems:

A study in *C. elegans* identified the *din-1* gene to encode two protein isoforms – a long DIN-1L and a short DIN-1S – whose functions are separable [Ludewig et al., 2004]. With 2 RRMs and a SPOC domain, DIN-1L is the *C. elegans* homolog of SPEN (**Figure 2.1**), and is required for embryonic and larval development [Ludewig et al., 2004]. DIN-1S on the other hand is important for lipid metabolism (similarly to *Drosophila*, [Hazegh et al., 2017]) and aging in adults. Intriguingly, DIN1-S (but not DIN-1L) interacts with the vitamin D receptor homolog DAF-12 [Ludewig et al., 2004], but its nuclear receptor interacting domain (RID) shows no significant sequence homology with the mammalian SPEN RID.

In zebrafish, *spen* knockdown leads to severe heart failure and arrhythmia [Rattka et al., 2021], a result that is in agreement with the severe cardiac defects observed during embryonic development of SPEN-deficient mice [Kuroda et al., 2003].

2.9 SPEN emerges as a crucial player in XCI

In 2015, SPEN was identified as a *Xist*-binding protein in all the proteomic studies [Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015], and this interaction was shown to take place between SPEN's RRM2-4 domains and the A-repeat of *Xist* [Chu et al., 2015; Minajigi et al., 2015; Lu et al., 2016].

Furthermore, across all three proteomic studies and both genetic screens, *Spen* loss of function was associated with defects in *Xist*-dependent silencing, for several genes tested [Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015; Moindrot et al., 2015; Monfort et al., 2015].

The Guttman lab further reported that knockdowns of *Smrt* or *Hdac3* also result in gene silencing defects [McHugh et al., 2015]. In light of previous works outlining the physical links between SPEN, NCoR/SMRT and HDACs (see **section 2.5**), the Guttman lab proposed a model wherein *Xist*-dependent recruitment of SPEN/SMRT/HDAC3 mediates global histone deacetylation on the X chromosome, ensuring chromosome-wide transcriptional silencing during XCI.

2.10 Goal and output of my PhD research

The evidence to support this model is scarce. Indeed, only one X-linked gene (*Gpc4*) was shown to respond similarly to either *Spen*, *Smrt* or *Hdac3* loss [McHugh et al., 2015]. Hence, key experiments aimed at exploring SPEN's mechanism of action during XCI need

to be conducted.

Most importantly, given that several other *Xist* interactors allegedly mediate gene silencing – to a certain degree – during XCI (see **section 1.13**), the extent to which SPEN (as well as HDAC3) contributes to this process must be determined at the scale of the entire X chromosome.

These studies laid the groundwork for the core of my PhD research, which aimed to dissect **SPEN function and mechanism of action precisely in the context of XCI** (see **chapter 6** and **[Dossin et al., 2020]**).

This project also led me to:

- collaborate with colleagues within the Heard lab to address HDAC3 function in XCI (see **chapter 5** and **[Żylicz et al., 2019]**).
- collaborate with the Müller lab at EMBL (Heidelberg) to explore the molecular/structural basis of the SPOC/RNAPII interaction (see **chapter 7**).
- collaborate with the Bourc'his lab at the Institut Curie (Paris) to establish degrons for factors of the m^6A RNA methylation pathway and study the role of such pathway in regulating expression of endogenous retroviruses (see **section A.1** and **[Chelmicki et al., 2021]**)
- collaborate with the Gribnau lab at the Oncode Institute (Rotterdam) to investigate the function of SPEN in regulating *Xist* expression (see **section A.2** and **[Robert-Finestra et al., 2020]**).

Part II

Material and Methods

Chapter 3

Experimental work

3.1 Plasmid construction for endogenous tagging of *Spen* & *Hdac3* in mESCs

The plasmids to target *OsTIR1* at the *TIGRE* locus and the *TIGRE* specific gRNA encoding plasmid were kindly provided to us by Elphège Nora. The additional *TIGRE* targeting plasmids BglG-mCherry-T2A-OsTir1 (pFD51) and rtTa-VP16-T2A-OsTir1 (pFD68) were cloned using PCR amplification of corresponding gene cassettes followed by traditional cloning into the original *TIGRE* backbone.

Targeting construct (pFD19 and pFD49) to tag endogenous *Spen* at its C-terminus with AID-HaloTag and AID-GFP respectively (and pFD39 to tag endogenous *Hdac3* at its C-terminus with AID-3xFLAG) were generated as follows: 500bp homology arms (flanking both sides of, but excluding the stop codon of *Spen* or *Hdac3*) were PCR amplified from mouse genomic DNA. 1-step Gibson cloning (NEB) was subsequently used to surround the digested AID insert (carrying a puromycin resistance gene under the control of the PGK promoter) in frame with the homology arms and clone the insert into a pBR322 vector. Synonymous mutations in the PAM/SEED target sequence (located on the 5' homology arm) were then introduced using the QuickChange II XL site-directed mutagenesis kit (Agilent) to prevent Cas9 mediated cutting of the targeting vector upon transfection and of the AID tagged allele(s) upon integration.

The pFD90 Targeting construct to replace the endogenous SPOC domain of SPEN by GFP was generated using the same strategy.

For gRNA cloning, the pX459 plasmid (gift from Feng Zhang, addgene #62988) encoding SpCas9 was digested with *BbsI* immediately downstream of the U6 promoter, and

annealed DNA duplex corresponding to the target gRNA sequences were ligated.

3.2 Cell culture

Mouse XX ES cells (TX1072) were grown on 0.1% gelatin-coated flasks in 8% CO2 37°C incubators. For all experiments, cells were cultured in 2i + LIF, and batch-tested fetal calf serum ES cell medium - DMEM (Sigma), 15% FBS (Gibco), 0.1mM β-mercaptoethanol, 1000 U/ml leukemia inhibitory factor (LIF, Chemicon), CHIR99021 (3uM), PD0325901 (1uM).

Neural progenitor cell (NPC) differentiations and subcloning were performed as previously described [Giorgetti et al., 2016]. NPCs were grown in N2B27 medium supplemented with EGF and FGF (10ng/ml each), on 0.1% gelatin-coated flasks.

3.3 Cell transfection and clone isolation

All transgenic insertions were performed using the 4-D nucleofector system from Lonza. For each nucleofection, 5 million cells were electroporated with 2.5ug each of non-linearized targeting vectors and gRNA/Cas9 encoding plasmids (MidiPreps). Nucleofected cells were then serially diluted and plated on 10-cm dish. 48 hours later, selection was added (puromycin: 0.4ug/mL, hygromycin: 250ug/mL, blasticidin: 5ug/mL) except for transfection steps involving flippase-mediated removal of resistance cassettes, during which no selection was applied.

1 week after initial plating, 80 to 96 single colonies were picked from dishes showing ideal clonal density and seeded in 96-well plates. These cells were subsequently split into one high-confluency plate used for PCR genotyping, and one low-confluency plate from which desired clones were further expanded until T25 density was reached. At this stage, some cells were kept to reconfirm correct genotype by PCR, while the remaining cells were frozen.

3.4 Cell treatments

Xist expression in TX1072 mESCs was induced upon administration of doxycycline (1ug/mL). Auxin mediated depletion of target proteins was achieved through supplementing culture media with auxin (IAA, Sigma) at the recommended concentration of

500uM. Auxin-containing medium was renewed every 24 hours. For auxin wash-out, auxin-containing medium was removed, cells were rinsed once with PBS, and exposed to auxin-free medium.

3.5 Protein extraction and western blotting

Cells were trypsinized, washed once in medium, once in PBS and pellets immediately frozen at -80°C. Pellets were then resuspended in RIPA buffer (50mM Tris-HCl pH8.0-8.5, 150mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) containing protease inhibitors (Roche), incubated for 30 minutes on ice and sonicated with a Bioruptor (three 10 second pulses). Lysates were then centrifuged for 20 minutes at 4°C, and supernatants were kept.

Protein concentration was determined using the Bradford (BioRad) assay. Samples were then boiled to 95°C for 10 minutes in LDS buffer (Thermo) containing 200mM DTT. For all Western blots except ones aiming at detecting SPEN, 4-12% Bis-Tris gels were used. For detection of SPEN, a high molecular weight protein (>400 kDa), 3-8% Tris-Acetate poly-acrylamide gels were used. Transfer was performed on a 0.45um nitrocellulose membrane using a wet-transfer system, at 350-400mA for 2h at 4°C.

3.6 RNA extraction, RT, Pyrosequencing and RNAseq

RNA extraction was performed using the RNeasy kit and on-column DNAse digestion (Qiagen). Reverse transcription was performed on 1ug total RNA using Super Script III (Life Technologies). To quantify allelic skewing, cDNA were amplified using biotinylated primers and subsequently sequenced using a Q24 Pyromark (Qiagen). Only samples showing a RNA integrity number (RIN) above 9 were used to prepare RNAseq libraries (TruSeq). Paired-end 100nt sequencing was performed on HiSeq2500 or NovaSeq6000 (Illumina, San Diego, CA).

3.7 RNA FISH

Cells were dissociated using Trypsin (Invitrogen) for ESCs or Accutase (Invitrogen) for NPCs, washed twice in medium, and allowed to attach on Poly-L-Lysine (Sigma) coated coverslips for 10 min. Cells were fixed with 3% paraformaldehyde in PBS for 10

min at room temperature, washed in PBS three times, and permeabilized with ice-cold permeabilization buffer (PBS, 0.5%Triton X-100, 2mM Vanadyl- ribonucleoside complex) for 5 minutes on ice. Coverslips were stored in 70% Ethanol at -20°C.

Samples were dehydrated in 4 baths of increasing Ethanol concentration (80%, 95%, 100% twice) and air-dried quickly. Probes were prepared from minipreps of intron-spanning BACs (clone RP24-157H12 for *Huwe1*, RP23-260I15 for *Atrx*) or plasmid (p510 for *Xist*). Probes were labeled by nick translation (Abbott) using dUTP labeled with spectrum green (Abbott) for *Huwe1*, spectrum red (Abbott) for *Atrx*, and Cy5 (Merck) for *Xist*. Labeled BAC probes were co-precipitated with Cot1 DNA repeats in the presence of ethanol and salt, resuspended in formamide, denatured at 75°C for 10 minutes, and competed at 37°C for 1 hour.

Probes were then co-hybridized in FISH hybridization buffer (50% Formamide, 20% Dextran sulfate, 2x SSC, 1 ug/ul BSA, 10mM Vanadyl-ribonucleoside) at 37°C overnight. The next day, hybridized coverslips were washed three times for 5 minutes with 50% formamide in 2X SSC at 42°C, and three times for 5 minutes with 2X SSC. 0.2mg/ml DAPI was added to the penultimate wash and coverslips were mounted with Vectashield (Vectorlabs).

3.8 HaloTag labeling

HaloTag labeling of SPEN-Halo fusion protein was performed in live TX1072 ESCs and NPCs: Cells were labeled with HaloTag ligand conjugated Janelia Fluor [Grimm et al., 2015] (JF646-HaloTag or JF549-HaloTag, kind gift from Luke Lavis) at a final concentration of 250nM in culture medium. Labeling was performed for 1 hour at 37°C, cells were then washed 4 times with generous volumes of PBS, and incubated with unlabeled medium for 15 minutes before proceeding with downstream experiments. For NPC labeling, cells were washed with unlabeled medium and not PBS, as NPCs detach when exposed to PBS. Auxin and/or doxycycline were kept in the labeling medium when necessary.

3.9 HaloTag labeling followed by RNA FISH

For co-detection of SPEN-Halo and *Xist* RNA, cells were labeled with JF549 as indicated above, and directly processed for fixation and permeabilization as detailed in the RNA FISH section. Importantly, after permeabilization, coverslips were directly washed

twice with PBS, twice with 2xSSC and immediately processed for FISH.

3.10 Immunofluorescence

mESCs were dissociated using Trypsin (Invitrogen), washed extensively in medium, and allowed to attach on Poly-L-Lysine (Sigma) coated coverslips for 10 min. Cells were then fixed with 3% paraformaldehyde in PBS for 10 min at room temperature, washed in PBS three times, and permeabilized with 0.25%Triton X-100 in PBS for 5 minutes at room temperature.

Coverslips were then washed three times in PBS and blocked for 1 hour with blocking buffer (PBS containing 2.5% BSA, 0.1% Tween20 and 10% normal goat serum). Coverslips were then incubated with primary antibodies diluted in blocking buffer at 4°C overnight, washed three times for 5 minutes in PBST (0.1% Tween20) on the next day, incubated with fluorescently labeled secondary antibodies (1/500 in blocking buffer) for 1 hour at room temperature, and washed again three times for 5 minutes in PBST. 0.2mg/ml DAPI was added to the penultimate wash and coverslips were mounted with Vectashield (Vectorlabs).

3.11 Hi-C

Hi-C was performed as previously described [Belaghzal et al., 2017], except that ligated DNA size selection was omitted, and dA-tailing was performed prior to biotin pull-down. Briefly, each Hi-C experiment was performed on 10 million cells (NPCs) per sample. Following fixation, cells were digested with DpnII at 37°C overnight. DNA ends were filled with biotin-14-dATP at 23°C for 4 hours. DNA was then ligated with T4 DNA ligase at 16°C overnight. Binding proteins were removed by treating ligated DNA with proteinase K at 65°C overnight. Purified proximally ligated molecules were fragmented to obtain an average fragment size of 200 bp. After DNA end repair, dA-tailing and biotin enrichment, DNA molecules were ligated to Illumina TruSeq sequencing adapters at room temperature for 2h. Final library PCR productions were carried out following the Illumina TruSeq Nano DNA Sample Prep Kit manual. Paired-end 100nt sequencing was performed on HiSeq4000 (Illumina, San Diego, CA).

3.12 Mouse breeding

Timed natural matings were used for all experiments. Noon of the day when the vaginal plugs of mated females were identified was scored as E0.5. For *Spen* matings a conditional allele was used [Yabe et al., 2007]. For oocyte deletions, a published *Rosa26:Zp3-Cre* allele was used [De Vries et al., 2000].

F1 hybrid $Spen^{+/-}$ males were obtained by crossing $Spen^{+/+}$ CAST/EiJ females with $Spen^{+/-}$ C57BL/6J males. For *Spen* maternally deleted embryos, $Spen^{flox/flox}$ $Zp3\text{-}Cre^{+ve}$ C57BL/6J females were crossed with $Spen^{+/-}$ F1 hybrid males. For *Spen* control embryos, $Spen^{flox/flox}$ $Zp3\text{-}Cre^{-ve}$ C57BL/6J females were crossed with $Spen^{+/-}$ F1 hybrid males.

3.13 Embryo collection and single-embryo RNAseq

Animal care and use for this study were performed in accordance with the recommendations of the European community (2010/63/UE). All experimental protocols were approved by the ethics committee of Institut Curie CEEA-IC118 under the number APAFIS#8812-2017020611033784v2 given by national authority in compliance with the international guidelines. Single-embryo RNA-seq was performed as previously described [Zylicz et al., 2018].

Briefly, E3.5 embryos were collected and morphologically assessed to ensure only viable samples were collected. Zona pellucida was removed by treatment with acidified Tyrode's solution. Single embryos were picked into individual tubes and cDNA was prepared and amplified as previously described [Tang et al., 2010]. Illumina libraries were prepared as published in [Huang et al., 2017]. Paired-end 100nt sequencing was performed with HiSeq2500 (Illumina, San Diego, CA).

3.14 Targeting Bgl stem loops at the *Xist* locus

To tag *Xist* with Bgl stem loops [Chen et al., 2009], we nucleofected cells with pBS-Ptight-Xist-BglSL [Masui et al., 2018] (plasmid harboring 18 repeats of BglSL inserted between homology arms to target *Xist* exon 7, carrying a G418 selection gene, kind gift from Osamu Masui). Following G418/neo selection and FLP-FRT mediated removal of the selection cassette, clones were picked and genotyped to ensure insertion on the *B6* (*Xist* inducible) allele. Positive clones were further tested to ensure that the stem-loop tagged *Xist* could properly be induced and trigger gene silencing upon addition of

doxycycline (data not shown).

3.15 Live-cell imaging and machine learning analysis

Cells were seeded on fibronectin coated 35mm glass bottom dishes (Ibidi) 24 hours before imaging. Doxycycline was added 1 hour before image acquisition. Cells were imaged on the DeltaVision OMX microscope in widefield mode (GE Healthcare) using a 1.4 NA 100x oil immersion objective. Temperature was controlled at 37°C and CO2 at 8% during acquisition. Images were acquired as z-stacks of 40 slices with 400nm steps every 10 minutes for at least 4 hours. Movies were deconvolved using Huygens deconvolution with the following parameters: Iteration 4; S/N 5, 10; Quality Threshold 0.1; and Widefield mode 0.7 was used for background estimation. 2 channels were registered using TetraSpec microspheres 0.1um (Invitrogen) and unwarpJ (Fiji plugin).

For segmentation, z-projected deconvolved registered images were used and pixels were classified as cloud/nuclei using Ilastik. Touching nuclei were seldom manually separated. Cutoffs on resulting probability maps were set to 0.7. We next performed connected component analysis to obtain integer labelled images where each integer label corresponds to a unique nucleus.

In our tailored made Fiji plugin the inputs are the raw max z-projected time-lapse images of the two channels and the integer labelled time-lapse image of the nuclei. The probability maps of the clouds give the region of interest in the time-lapse sequence in which total intensity is calculated. In the plugin, clouds are associated with their corresponding nuclei, they are then linked via Kalman filter tracker over time. These unique links constitute track id's and contain the information about the intensity and area measurements for each cell. For each tracked cell, the first timepoint when a cloud is detected in one channel (*Xist* or SPEN) is labeled as reference timepoint 1.

3.16 Cross-linked CUT&RUN

CUT&RUN against SPEN-GFP was performed during a timecourse of *Xist* induction/SPEN degradation: 0h dox, 4h dox, 8h dox, 24h dox and 8h dox + auxin. Two biological replicates were performed.

The original CUT&RUN protocol [Skene and Henikoff, 2017] was adapted for fixed cells: 10^6 cells in suspension were fixed with 2% formaldehyde diluted in PBS for 10 minutes at room temperature (2mL final volume). Fixation was quenched with 125mM

glycine for 5 minutes and cells were washed twice in 1mL PBS. Fixed cells were then permeabilized with 1mL permeabilization buffer (20mM HEPES pH7.9, 150mM NaCl, 0.5mM Spermidine, 0.25% TritonX-100, c0mplete EDTA free) for 5 minutes and washed twice in 1mL PBS. Cells were then resuspended in 1mL washing buffer (20mM HEPES pH7.9, 150mM NaCl, 0.5mM Spermidine, 0.1% BSA, cOmplete EDTA free), bound to activated concanavalin beads (50uL bead slurry used per 10 million cells) for 10 minutes and blocked in 1mL blocking buffer (wash buffer + 2mM EDTA) for 5 minutes.

At this stage, cells were resuspended in 500uL wash buffer containing target antibodies diluted 1/200, transferred to 0.5mL tubes, and incubated overnight at 4°C on an end-to-end rotator. Cells were washed three times in 500uL washing buffer followed by 1-hour incubation with pA-MNase (500uL of washing buffer containing 700ng/mL pA-MNase, produced by the Protein Expression and Purification Core Facility of Institut Curie) and washed again three times in 500uL washing buffer.

After the last wash, cells were resuspended in 150uL washing buffer, transferred to 1.5mL tubes, and equilibrated to 0°C in a metal block for 10 minutes. To start digestion, CaCl2 was added to 1.5mM final concentration, taking care to return each sample to 0°C immediately after. Digestion was performed at 0°C for 1 hour, before being stopped by adding 150uL of 2X-STOP solution (200mM NaCl, 20mM EDTA, 5mM EGTA, 0.1% NP-40, 40ug/mL glycogen).

RNAse A was added to a final concentration of 50ug/mL and samples were incubated at 37°C for 20 minutes. SDS and proteinase K were then added to final concentrations of 0.1% and 300ug/mL respectively and samples were incubated at 56°C for 2 hours followed with 68°C for 16 hours to reverse crosslinking.

Total DNA was extracted using phenol-chloroform followed by two rounds of ethanol precipitation and DNA size selection (using 0.55x volume of Ampure XP beads relative to DNA sample volume) to remove the large predominating undigested DNA fragments. Each time, beads were discarded, and the supernatant (containing the selected small fragments resulting from MNase digestion) was ethanol precipitated.

Following elution in 50uL TE buffer, samples were quantified and analyzed using Qubit and Tapestation assays. CUT&RUN libraries were prepared from 50ng DNA per samples, using the Accel-NGS 2S Plus DNA Library Kit (Swift) according to manufacturer's protocol. Paired-end 100nt sequencing was performed on HiSeq2500 (Illumina, San Diego, CA).

3.17 SPEN & BglG-SPOC complementation constructs design and analysis

Spen cDNA truncations were generated by splicing out different regions of the *Spen* ORF (Genscript, ORF clone OMu11416C) using overlap extension PCR. Each Spen truncation was cloned downstream of a CAGGS promoter into a vector carrying homology arms for targeted insertion at the *ROSA26* [Barau et al., 2016] locus as well as a SV40-promoter driven hygromycin resistance gene.

The BglG-GFP-SPOC targeting plasmid was designed by inserting a translational fusion between a BglG-GFP cassette and SPEN aa3244-3643 into the same *ROSA26* targeting vector.

Each of these "complementation" constructs were independently targeted at *ROSA26* in SPEN-degron mESCs. Independent clones were picked and protein expression of each SPEN truncation was assessed by western blot. XCI complementation analysis was then performed in 2-3 independent clones for Spen cDNA truncations, and 4 independent clones for BglG-GFP-SPOC expressing clones. The ability of cells to accumulate BglG-mCherry (to track *Xist* RNA), BglG-GFP and BglG-GFP-SPOC upon addition of doxycycline was assessed using microscopy (data not shown).

3.18 SPOC-GFP immunoprecipitation (GFP-Trap) mass-spectrometry

Nuclear extracts were prepared by resuspending 50 million fresh cells (per IP) in 10mL ice-cold buffer A (10mM HEPES pH7.9, 10mM KCl, 1.5mM MgCl2, 0.1% NP-40, c0mplete EDTA free, phosSTOP) and rotating for 10 minutes at 4°C. Nuclei were centrifuged at 2000rpm for 10 minutes at 4°C and resuspended in 1mL IP buffer C150 (20mM HEPES pH7.9, 150mM NaCl, 1.5mM MgCl2, 0.2mM EDTA, 0.25% NP-40, cOmplete EDTA free, phosSTOP).

Lysates were briefly sonicated followed by Benzonase (Merck) digestion for 30 minutes at 4°C. Finally, lysates were cleared through centrifugation at 13000rpm for 20 minutes prior to being incubated with 15uL of GFP-Trap magnetic agarose bead slurry (ChromoTek) overnight at 4°C. Beads were washed 5 times in IP buffer C150.

For Co-IP western blot, washed beads were directly resuspended in LDS buffer (Thermo) containing 200mM DTT, and boiled at 95°C for 10 minutes.

For Co-IP followed by mass-spectrometry, proteins on magnetic beads were washed twice with 100 uL of 25 mM NH4HCO3 on-beads digestion with 0.2 ug of trypsin/LysC (Promega) was performed for 1 hour in 100 uL of 25 mM NH4HCO3. Sample were then loaded onto a homemade C18 StageTips for desalting. Peptides were eluted using 40/60 MeCN/H2O + 0.1% formic acid and vacuum concentrated to dryness.

Online chromatography was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled online to a Q Exactive HF-X with a Nanospay Flex ion source (Thermo Scientific). Peptides were first trapped on a C18 column (75 um inner diameter × 2 cm; nanoViper Acclaim PepMapTM 100, Thermo Scientific) with buffer A (2/98 MeCN/H2O in 0.1% formic acid) at a flow rate of 2.5 uL/min over 4 min. Separation was then performed on a 50 cm x 75 um C18 column (nanoViper Acclaim PepMapTM RSLC, 2 um, 100A, Thermo Scientific) regulated to a temperature of 50°C with a linear gradient of 2% to 30% buffer B (100% MeCN in 0.1% formic acid) at a flow rate of 300 nL/min over 91 min.

MS full scans were performed in the ultrahigh-field Orbitrap mass analyzer in ranges m/z 375-1500 with a resolution of 120 000 at m/z 200. The top 20 intense ions were subjected to Orbitrap for further fragmentation via high energy collision dissociation (HCD) activation and a resolution of 15 000 with the intensity threshold kept at 1.3 × 105. We selected ions with charge state from 2+ to 6+ for screening. Normalized collision energy (NCE) was set at 27 and the dynamic exclusion of 40s.

3.19 Cloning of SPOC constructs for bacterial expression in *Rosetta2*

Fragments of different lengths encompassing SPEN's SPOC domain (see **Table 7.1 Figure 7.2**) were PCR-amplified from *Spen* cDNA (see **section 3.17**) and cloned into the bacterial expression vector pETM11, immediately downstream of either 6xHis-only or 6xHis-SUMO3 tags, both of which can be cleaved with TEV or SUMO proteases respectively. For 6xHis-only tagged SPOC fragments, the SPOC amplicons were cloned between the *NcoI* and *HindIII* restriction sites. For 6xHis-SUMO3 tagged SPOC fragments, the SPOC amplicons were cloned between the *BamHI* and *HindIII* restriction sites.

The resulting plasmids were finally transformed in the *E. coli* BL21(DE3)Rossetta2-pLysS strain (Novagen).

3.20 Bacterial expression and purification of SPOC

50mL bacterial pre-cultures were grown overnight at 37°C. The next morning, 10mL of pre-culture were added to 1 liter of *T*errific *B*roth (TB) containing kanamycin (50ug/mL) and chloramphenicol (33ug/mL) for selection of the SPOC-expressing and pRARE plasmids respectively. Bacterial growth at 37°C was monitored hourly until log-phase was reached (OD_{600}=0.6), at which stage SPOC expression was induced with 0.5mM isopropyl-β-D-thiogalactopyranoside (IPTG). Growth was continued overnight, at 18°C.

The next day, cells were harvested by centrifugation, and the pellet was weighed and resuspended in ice-cold lysis buffer (50mM HEPES pH8.0, 300mM NaCl, 10mM imidazole, 5% glycerol, 1mM DTT, c0mplete EDTA free protease inhibitor cocktail, DNase 1), using 50mL of buffer per 10g of bacterial pellet. Lysis was completed by adding lysozyme and passing the suspension through a French pressure cell press. Lysates were cleared by centrifugation, and the supernatant was applied to a Ni-NTA agarose column (QIAGEN). The column was washed with 20 column-volumes of ice-cold wash buffer (50mM HEPES pH8.0, 500mM NaCl, 10mM imidazole, 5% glycerol, 1mM DTT), followed by 20 column-volumes of ice-cold lysis buffer.

Elution of SPOC was carried out in 5 column-volumes of ice-cold elution buffer (50mM HEPES pH8.0, 300mM NaCl, 300mM imidazole, 5% glycerol, 1mM DTT). Proteolytic cleavage of the 6xHis-only or 6xHis-SUMO3 tag was performed with TEV and SUMO proteases respectively during overnight dialysis against dialysis buffer (50mM HEPES pH8.0, 100mM NaCl, 10mM imidazole, 5% glycerol, 1mM DTT). The dialysed protein solution was loaded a second time onto an Ni-NTA agarose column to remove both the cleaved tag and the proteases (which carry an N-terminal 6x-His tag as well).

SPOC was further purified through size exclusion chromatography (Superdex 75 10/300 GL from GE Healthcare, with a flow of 0.5mL/min) in gel filtration buffer (20mM HEPES pH7.5, 50mM NaCl, 2mM DTT). Fractions showing highest purity (assayed by SDS-PAGE) were pooled and the protein solution was concentrated to 20mg/mL (for ITC experiments) or 80mg/mL (for co-crystallization with RNAPII-CTD peptides), aliquoted and snap-frozen in liquid nitrogen. SPOC aliquots were stored at -80°C for later use.

3.21 Crystallization of SPOC alone

Crystallization was carried out using the sitting drop vapor diffusion technique with a Mosquito nano dispenser (TTPLabtech), at 20°C. Protein and crystallization solutions were mixed in 1:1 volume ratio. A range of crystallization screens were used including the

JCSG-plus, MORPHEUS, PACT premier and LMB screens from Molecular Dimensions, the Classics Suite and PEGsI screens from NeXtal QIAGEN, as well as a homemade EMBL screen.

Crystals appeared within the scope of one day, and grew over a week period. The best SPOC diffracting crystals were grown in 0.05mM $MgNO_3$, 12% PEG 3350. Crystals were fished out with cryo-loops, soaked in cryo-solution containing 20% glycerol, and flash frozen in liquid nitrogen prior to data collection.

3.22 Preparation of RNAPII-CTD phosphopeptides

Peptides matching the consensus heptapeptide repeat sequence of the C-terminal domain (CTD) of RNA polymerase II were chemically synthesized and reverse phase HPLC purified to >95% by Peptide Specialty Laboratories GmbH. These peptides vary in the number of heptapeptide repeats (either 2 or 4) as well as in their Serine phosphorylation status (unphosphorylated, Ser2-, Ser5- and Ser7-phosphorylated).

Upon commercial delivery, each peptide was resuspended in pure water, dialyzed against water overnight and fully dried using a SpeedVac vacuum concentrator. The peptides were then resuspended in gel filtration buffer (see **section 3.20**) and pH was adjusted to 7.5 when necessary. The resuspended peptides were stored at -20°C for later use.

3.23 Isothermal titration calorimetry (ITC) assays of binding between SPOC and RNAPII-CTD peptides

All ITC binding assays were conducted with a MicroCal PEAQ-ITC calorimeter from Malvern Panalytical. ITC experiments were performed at 20°C in gel filtration buffer (20mM HEPES pH7.5, 50mM NaCl, 2mM DTT), with 200uL of 60uM SPOC in the cell and 40uL of 1mM RNAPII-CTD peptide in the syringe. For each peptide, measurements were derived from a minimum of three independent experiments. Data analysis was carried out with the MicroCal PEAQ-ITC software (Malvern Panalytical), using a one-site binding model.

3.24 Co-crystallization of SPOC and RNAPII-CTD Ser5P peptides

Thawed SPOC aliquots (at 80mg/mL) were briefly centrifuged to remove any precipitates formed upon freeze/thawing. The SPOC/Ser5P-CTD peptide (4 repeats of the Ser5-phosphorylated RNAPII-CTD peptide) complex was assembled in gel filtration buffer such that the final SPOC concentration was of 50mg/mL (2.7mM). Complex formation was performed at room temperature for 1 hour, and a brief centrifugation was used to remove any precipitates appearing during that time. Two different peptide:SPOC molar ratios (2:1 and 4:1) were tested for co-crystallization.

Crystallization was carried out using the sitting drop vapor diffusion technique with a Mosquito nano dispenser (TTPLabtech), at 20°C. Protein and crystallization solutions were mixed in 1:1 volume ratio. Various protein crystallization screens were used as mentioned in **Table 7.4**, including the Basic Chemical Space (BCS) and PACT premier from Molecular Dimensions, and the Protein Complexes Suite (PCS), PEGsI and PEGsII from NeXtal QIAGEN.

Crystals appeared very quickly, within the scope of a few hours, and grew over a week period. The best SPOC/Ser5P-CTD diffracting crystals were grown in 0.2M NH_4Cl, 0.1M TRIS pH8, 20% PEG 6000. Crystals were fished out with cryo-loops, soaked in cryo-solution containing 20% glycerol, and flash frozen in liquid nitrogen prior to data collection.

Chapter 4

Bioinformatics analyses

4.1 Genomics analyses

All genomics data were mapped to the mouse genome mm10, using the BL6-EiJ / CAST SNPs from the mouse genome project (v5 SNP142), and the gene annotation from ensembl (v92). Analyses were performed in R (v 3.4.2) and Bioconductor (v3.6). See [Żylicz et al., 2019] for more details.

4.1.1 RNAseq analysis

Reads were trimmed using Trimgalore (v 0.4.4), mapped using STAR (2.5.3a, parameters: –outFilterMultimapNmax 1 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.06 –alignIntronMax 500000 –alignMatesGapMax 500000 –alignEndsType EndToEnd –outSAMattributes NH HI NM MD), and removed when mapping to the mitochondrial genome. Remaining reads were split by allele using SNPsplit (v 0.3.2). Allele specific and the unassigned bam files were sorted, duplicates removed using picard (v2.18.2, parameters: REMOVE_DUPLICATES=true ASSUME_SORTED=true) and pooled as the total reads. Quantification of expression was performed using featureCount (parameters: -p -t exon -g gene_id, -s 1 for stranded RNAseq of in vitro cell, -s 0 for non-stranded RNAseq of single embryo). Data were then analysed in R using DESeq2 (v1.18.1), calculating the sizeFactor on the count of total reads and applying it to the allele specific counts.

For all RNAseq analysis (Spen-degron mESCs, NPCs, and Spen-KO embryos), genes showing less than 10 total allelic reads in at least one sample were discarded from the analysis. Allelic ratios were then computed for genes as follows: $allelic\_ratio =$

$\frac{reads_{B6}}{reads_{B6}+reads_{Cast}}$. Allelic ratios were then averaged between biological replicates.

4.1.1.1 SPEN-degron mESCs RNAseq analysis

Defining differential dependencies on Spen for gene silencing during XCI in mESCs: For this analysis, skewed genes (i.e. genes showing allelic ratios outside of a [0.15;0.85] interval in control conditions) were removed. We then defined a silencing index, translating how much a gene is silenced after 24h of Xist induction with respect to the control condition: $silencing\_index = 1 - \frac{allelic\_ratio_{dox}}{allelic\_ratio_{control}}$. We next filtered out genes showing less than 10% silencing (i.e. silencing_index≤0.1) in Spen non-depleted condition. K-means with 3 clusters was then performed on the raw allelic_ratios across control, 24hdox and 24hdox+aux conditions. Clustering identified 3 groups of genes differing by their response to loss of SPEN during *Xist* induction. To define how dependent on Spen a gene is for silencing, we expressed the silencing defect observed upon loss of SPEN as a fraction of the total silencing that normally occurs in the presence of SPEN. Computationally, this translates in: $Spen\_dependence\_index = 1 - \frac{silencing\_index_{dox+aux}}{silencing\_index_{dox}}$. A Spoc_dependence_index was derived identically.

Integration with HDAC3 KO RNAseq during XCI: Our Spen-degron dataset was integrated with an Hdac3 KO RNAseq dataset [Żylicz et al., 2019] generated from the same mESC background (TX1072) and at the same timepoint of *Xist* induction. The dataset was processed identically, and an Hdac3 dependence index was also computed as follows: $Hdac3\_dependence\_index = 1 - \frac{silencing\_index_{Hdac3KO}}{silencing\_index_{WT}}$.

4.1.1.2 *Spen* KO E3.5 embryo RNAseq analysis

Integration of SPEN and Xist KO embryo datasets: Our Spen KO E3.5 female embryo RNAseq dataset was integrated with a *Xist* KO scRNAseq (processed as pseudo-bulk for our analysis) dataset from E3.5 female embryos [Borensztein et al., 2017b], also generated from a *Mus musculus domesticus* x *Mus musculus castaneus* mouse background.

4.1.1.3 SPEN-degron NPC RNAseq analysis

Defining genes that escape from XCI: In NPCs, X-linked genes were defined as escapees if their transcript allelic ratio was superior to 0.15 in at least one condition (0h, 24h or 48h of SPEN depletion).

4.1.2 CUT&RUN bioinformatics analysis

Reads were trimmed using Trimgalore (v 0.4.4), mapped using STAR (2.5.3a, parameters: –outFilterMultimapNmax 1 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.06 –alignIntronMax 1 –alignMatesGapMax 2000 –alignEndsType EndToEnd –outSAMattributes NH HI NM MD), and removed when mapping to the mitochondrial genome. Remaining reads were split by allele using SNPsplit (v 0.3.2). Allele specific and the unassigned bam files were sorted, duplicates removed using picard (v2.18.2, parameters: REMOVE_DUPLICATES=true ASSUME_SORTED=true) and pooled as the total reads. BigWig of coverage files were done using DeepTools bamCoverage (parameters: –extendReads –binSize 1, with –extendReads 200 for single end data). A scaling factor was calculated as 10^6 / total number of reads, and the same factor was given as the parameter –scaleFactor for both allelic signals. Peak calling was performed using macs2 (v 2-2.1.2.1, parameters for CUT&RUN : –bw 300 -f BAMPE -q 0.01 –keep-dup auto –broad for CUT&RUN and pol2S5 ChIPseq; –bw 300 -f BAMPE -q 0.01 –keep-dup auto –call_summits for other ChIPseq). For quantification of signal in peaks, reads were counted using the featureCounts function from Subread (v1.28.1, parameters: -p -s 0). Data scaling was performed in R using DESeq2 (v1.18.1), calculating the sizeFactor on the counts of total reads in 10kb windows and applying it to the allele specific counts in peaks.

4.1.2.1 Peak filtering

Spen-specific peaks were defined as having a log2FoldChange ≥ 1 compared to auxin treatment (negative control, Spen-degraded), and an adjusted p-value ≤ 0.001.

4.1.2.2 Total Spen enrichment in promoter windows

To compare SPEN accumulation among promoter of all X-linked genes unbiasedly (including genes which fail to have any peak called at their promoters), we performed DEseq analysis on counts spanning total promoter windows.

4.1.2.3 Genomic features and integration with RNAseq

Promoters were defined as ±2kb windows centered around genes TSS. Putative active enhancers and their deacetylation kinetics during XCI were obtained from [Żylicz et al., 2019]. Gene silencing efficiency was determined according to the silencing_index defined in the RNAseq analysis section above. We observed that our silencing_index ranges

between 0 and 0.9. Hence, we split this interval in 3 to define high, medium, and low gene silencing efficiency groups with silencing_index comprised in [0.6,0.9], [0.3,0.6[, and [0,0.3[respectively.

4.1.2.4 Integration with publically available ChIPseq data

SPEN peaks were intersected with other peaks called from publically available ChIPseqs for HDAC3 [Żylicz et al., 2019] (same cellular background, TX1072, 2i+LIF condition), RNAPII-pS5 [Bornelöv et al., 2018], CHD4 [Bornelöv et al., 2018] and MBD3 [Bornelöv et al., 2018] (all in 2i+LIF condition).

4.1.3 Hi-C analysis

Data were processed with HiC-Pro (v2.11.0) in allele specific mode. Only pairs with both reads having MAPQ>30 were kept. Matrices were made using cooler cload (v 0.8.5) at 1kb or 10kb resolution, using HiGlass for visualisation and snapshots. TADs were called using HicExplorer HicFindTads (parameters: –correctForMultipleTesting fdr –minDepth 250000 –maxDepth 4000000 –step 50000 –thresholdComparisons 0.1 –delta 0). Expected value for HiC signal was calculated on the non-allele specific signal using cooltool compute-expected. Average scaled matrices of observed/expected values for allele specific signal were produced with Coolpup.py (parameters: –local –rescale –rescale_size 299), using non-allele specific expected values to normalise both alleles to the same expected values. Average heatmaps were plotted using plotpup.py. For quantification, HiC signal was average over TADs upper-triangle for each allele specific matrices (10kb) using the hicExplorer hicSummarizeScorePerRegion available at https://github.com/heard-lab/HiCExplorer (parameters: –summarizeType mean –rmDiag 1).

4.2 Proteomics data analysis

For identification, the data were searched against the Mus musculus (UP000000589) Uniprot database using Sequest HF through proteome discoverer (version 2.2). Enzyme specificity was set to trypsin and a maximum of two-missed cleavage sites were allowed. Oxidized methionine and N-terminal acetylation were set as variable modifications. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.02 Da for MS/MS peaks.

The resulting files were further processed using myProMS [Poullet et al., 2007] v3.6

(work in progress). FDR calculation used Percolator and was set to 1% at the peptide level for the whole study. The label free quantification was performed by peptide Extracted Ion Chromatograms (XICs) computed with MassChroQ version 2.2 [Valot et al., 2011].

For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN) with two-missed cleavages were used. Median and scale normalization was applied on the total signal to correct the XICs for each biological replicate. To estimate the significance of the change in protein abundance, a linear model (adjusted on peptides and biological replicates) was performed and p-values were adjusted with a Benjamini–Hochberg FDR procedure with a control threshold set to 0.05.

4.3 Protein structure determination and refinement

X-ray diffraction data was collected on beamline P13 (for SPOC) and P14 at the PETRAIII storage ring (DESY, Hamburg, Germany) and processed using the XDS package [Kabsch, 2010b,a].

The structure of murine SPOC was solved by molecular replacement using the PHENIX package [Adams et al., 2010] and the human SPOC structure (PDB 1OW1, [Ariyoshi and Schwabe, 2003]) as a template. Refinement was performed using PHENIX Refine [Adams et al., 2010] and rebuilt using Coot [Emsley and Cowtan, 2004].

The structure of mouse SPOC/Ser5P-CTD was similarly solved by molecular replacement, using our mouse SPOC structure as a template. AutoBuild from PHENIX [Adams et al., 2010] was used to rebuild the model and the same strategy as described in the previous paragraph was applied to refine and complete the model.

4.4 Data availability

Genomic sequencing datasets have been deposited in the Gene Expression Omnibus under accession number GSE131784. The mass-spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD015699.

Part III

Results

Chapter 5

Chromatin changes in XCI *(Zylicz et al., Cell. 2019)*

The overall goal of my PhD was to gain a better understanding of the **mechanisms at play during initiation of gene silencing** during X chromosome inactivation.

When I joined the lab in 2017, a project led by two postdocs (Jan Zylicz and Aurélie Bousard) undertook to gain such insights by profiling the dynamics of histone modifications during XCI and understand their functional link with X-linked gene silencing. Published early 2019 in Cell, my scientific contribution to this project resulted in a co-authorship. The results are to be found in the main text/figures and supplementary figures of the article appended at the end of this chapter (**section 5.4** and **[Żylicz et al., 2019]**).

In this chapter, I will briefly describe the key results derived from this study, outline what my contribution consisted in and highlight which insights were gained thanks to it. Importantly, several results obtained throughout this study are to be integrated with the questions I raised and addressed in the context of my main PhD project centered around SPEN function in XCI (see **chapter 6**).

5.1 General methodology behind *Zylicz et al., 2019*

5.1.1 The TX1072 cellular system

This project made use of an elegant female mESC line called TX1072 [Schulz et al., 2014], in which a doxycycline inducible promoter has been integrated on one allele of the

endogenous *Xist* gene, allowing for conditional monoallelic upregulation of *Xist* RNA and subsequent gene silencing (**Figure 5.1A**).

This offers two benefits: First, XCI is induced synchronously in all cells, overriding the heterogeneity observed with "undinduced" XCI, allowing to follow the process using population-based assays. Second, it enables uncoupling *Xist* upregulation/XCI from cellular differentiation (see **section 1.6**), thereby removing any confounding factors linked to differentiation that would complicate data interpretation. It is also worth noting that although dox-induction of *Xist* results in faster XCI, the order in which genes are silenced is similar to that of endogenous XCI (Heard, unpublished).

Furthermore, TX1072 are derived from a cross between two highly polymorphic mouse strains (*C57BL6/J* x *Cast/EiJ*), hereafter referred to as *B6* and *Cast*, with SNPs occurring every 200-300 base pairs (**Figure 5.1B**). This high SNP density allows discrimination between alleles of the two X chromosomes. Hence, allele-specific information can be retrieved across a diverse range of genomic approaches, including transcriptomics, chromatin-fragment sequencing as well as chromosome conformation capture (**Figure 5.1B**).

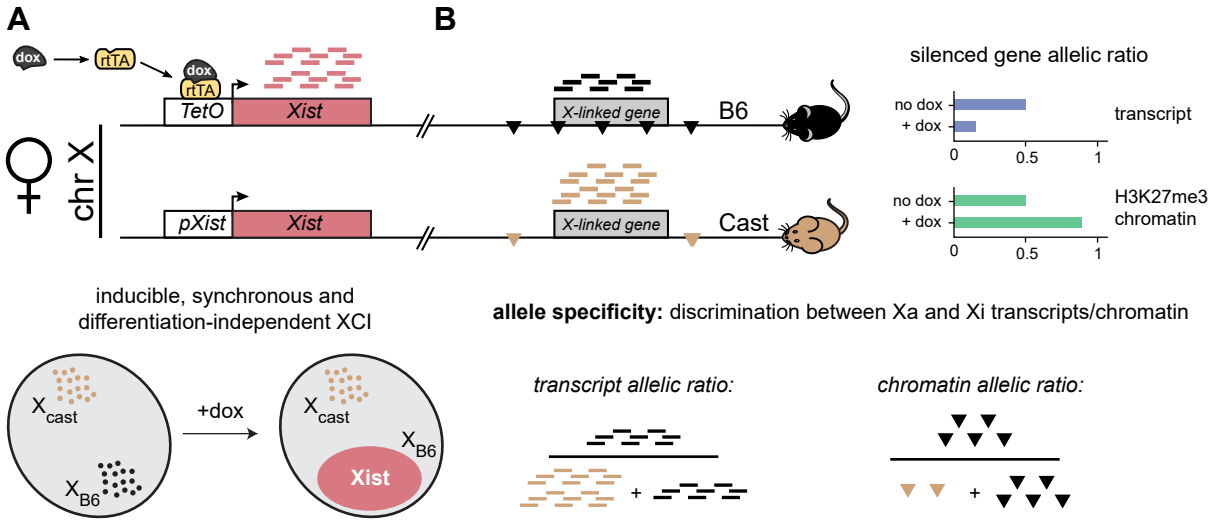


Figure 5.1 – Scheme depicting the TX1072 mouse embryonic stem cell system. (A) A doxycycline-inducible promoter (TetO) upstream of the endogenous *B6 Xist* allele allows inducible, synchronous and differentiation-independent XCI. **(B)** Being derived from a cross between *B6* and *Cast* mouse strains, the TX1072 cell line allows discrimination between the Xa and the Xi using allele-specific sequencing approaches. Allelic ratios can then be derived, directly informing about the transcriptional status (shown as horizontal lines and the blue barplot) as well as the chromatin environment (shown as triangles and the green barplot) of loci on the Xi.

Routinely, in a given experiment involving genomics, each locus will be attributed an *allelic ratio*, which is defined as $\frac{B6_{reads}}{B6_{reads}+Cast_{reads}}$. When looking at a typical X-linked gene (i.e. one that gets silenced), its associated transcript allelic ratio (derived from e.g.

RNAseq) will be ≈0.5 prior to *Xist* induction (**Figure 5.1B**) – as both *B6* and *Cast* alleles are expressed at similar levels. Similarly, the H3K27me3-marked chromatin allelic ratio (derived from e.g. ChIPseq) at that same gene will also be ≈0.5 (**Figure 5.1B**) – as chromatin is equally marked (albeit at low levels) by H3K27me3 on both alleles of that gene.

However, since the inducible *Xist* allele is on the *B6* chromosome (**Figure 5.1A**), the transcript allelic ratio will decrease below 0.5 following *Xist* induction (**Figure 5.1B**) – as the *B6* allele gets transcriptionally silenced, and the fraction of *B6* reads diminishes. Conversely, the H3K27me3-marked chromatin allelic ratio will increase above 0.5 (**Figure 5.1B**) – as PRC2 (and consequent H3K27me3) accumulates on the *B6* X chromosome following *Xist* induction (see **section 1.8**), and the fraction of *B6* reads increases.

5.1.2 Experimental approach

To characterize the dynamics of chromatin modifications during XCI, Jan Zylicz and Aurélie Bousard generated and analyzed allele-specific ChIP-seq datasets for 7 different active and repressive histone marks (**Table 5.1**) during a timecourse (0, 4, 8, 12 and 24 hours) of *Xist* induction in TX1072 mESCs (see Zylicz et al., Figure 1A).

genomic feature	associated histone marks
active promoter	H3K4me3, H3K9ac, H4ac, and H3K27ac
active enhancer	H3K4me1 and H3K27ac
Polycomb-marked	H2AK119ub1 (PRC1), H3K27me3 (PRC2)

Table 5.1 – List of histone modifications profiled in Zylicz et al., Cell. 2019

To further understand the link between the dynamics of these histone modifications and transcriptional silencing of X-linked genes, allele-specific *t*ransient *t*ranscriptome sequencing (TT-seq) [Schwalb et al., 2016] – which sensitively monitors transcriptional activity through specific sequencing of newly synthesized (within less than five minutes) RNAs – was performed at each timepoint of the aforementioned *Xist* induction timecourse.

5.2 Key results independent of my contribution

5.2.1 Spreading of Polycomb marks during XCI requires gene silencing

As expected, it was found that the Polycomb repressive marks (H2AK119ub1 and H3K27me3) rapidly accumulate across the X chromosome upon *Xist* coating (see Zylicz et al., Figure 6). These marks first accumulate intergenically, and only spread within genic regions upon gene silencing (see Zylicz et al., Figure 6 and Figure S6F). Consistently, in A-repeat mutant cells (in which *Xist* fails to silence genes), Polycomb marks still accumulate strongly intergenically, but fail to spread into active genes (see Zylicz et al., Figure 7 and Figure S7).

Collectively, these results are consistent with previous studies conducted in other systems, which found that Polycomb activity can only be marked at transcriptionally inactive sites [Riising et al., 2014; Kaneko et al., 2014; Hosogane et al., 2016]. Furthermore, the observation that Polycomb marks still strongly accumulate in *Xist* A-repeat deletion cells is consistent with the fact that Polycomb recruitment predominantly occurs through the B-repeat of *Xist* (see **subsection 1.13.1**).

5.2.2 PRC1 recruitment precedes PRC2 during XCI

The PRC1-associated H2AK119ub1 mark was found to accumulate before the PRC2-associated H3K27me3 mark (see Zylicz et al., Figure 6), suggesting that PRC1 recruitment during XCI precedes PRC2 recruitment. Such genomics-based evidence confirms a PRC1 first, PRC2 second model which was previously put forth through microscopy-based observations (see **subsection 1.13.1**). Importantly, this is in sharp contrast with what is classically observed with polycomb-marked loci in non-XCI contexts, where PRC2 recruitment and H3K27me3 deposition is required for the subsequent recruitment of PRC1 [Di Croce and Helin, 2013].

5.2.3 HDAC3 is important for *Xist*-dependent gene silencing

Among all active histone marks profiled, loss of histone acetylation at H3K27 occured most rapidly upon *Xist* upregulation (see Zylicz et al., Figure 1B). Notably, loss of H3K27ac at enhancers and promoters occurred with faster kinetics than loss of H3K4me1 and H3K4me3 respectively (see Zylicz et al., Figure 1C-E). Furthermore, when the profiles

of active mark dynamics at X-linked promoters were compared to the transcriptional silencing dynamics of corresponding genes (obtained by TT-seq), a highest correlation was observed with loss of H3K27 acetylation (see Zylicz et al., Figure 2A-C). Collectively, these data suggested that there could be a functional link between histone deacetylation and *Xist*-induced gene silencing during XCI.

To test this hypothesis, the impact perturbing *h*istone *deac*etylases (HDACs) has on X-linked gene silencing during XCI was investigated. Using a combination of HDAC inhibitors, selectively inhibiting HDACs expressed in ES cells (HDAC1/2, HDAC3, HDAC4, HDAC5, HDAC6 and HDAC8), it was found that inhibition of HDAC3 (but not other HDACs) resulted in deficient silencing of several X-linked genes (see Zylicz et al., Figure 3A-B). This observation was further confirmed by monitoring gene silencing chromosome-wide using RNAseq in *Hdac3*-KO TX1072 cells (see Zylicz et al., Figure 3C), which showed significant defects in gene silencing after 24 hours of *Xist* induction (see Zylicz et al., Figure 3D-G).

Furthermore, ChIP-seq revealed that loss of H3K27ac on the X chromosome is severely hindered in *Hdac3* KO cells (see Zylicz et al., Figure 4), and that genes whose silencing is most affected by HDAC3 loss also show most drastic defects in H3K27 deacetylation at their promoters (see Zylicz et al., Figure 4C-D).

Taken together, these results highlight that HDAC3-mediated deacetylation is important for efficient gene silencing during X chromosome inactivation. However, whether and how HDAC3 is recruited to its X-linked targets, and which partners does it interact with in the context of XCI remained unclear at that stage of the project. With no good HDAC3 antibodies to perform IP-based experiments necessary to bring some insights to these questions, I had the opportunity to contribute to this project with experimental tools I had in parallel developed in the context of my main project centered on SPEN (**chapter 6**).

5.3 Key results to which I contributed

5.3.1 Generation and characterization of endogenous FLAG-tagged HDAC3 mESCs

To circumvent the lack of good antibodies against HDAC3, I opted to generate a cell line in which the endogenous HDAC3 protein would be fused to a tag to allow subsequent affinity purification and identify HDAC3 binding dynamics at genomic targets during XCI

(ChIP-seq) as well as its associated protein partners (Co-IP-MS).

Taking advantage of CRISPR/Cas9-mediated homology directed repair, I performed a homozygous knock-in of a 3xFLAG cassette immediately upstream of the STOP codon of *Hdac3* (**Figure 5.2A**). Homozygous knock-in could be confirmed using SNP-genotyping following PCR on genomic DNA (**Figure 5.2A**). The endogenously tagged HDAC3 protein was expressed at levels similar to its untagged counterpart (**Figure 5.2B**), and its function remained preserved given that gene silencing during XCI was not affected (**Figure 5.2C**).

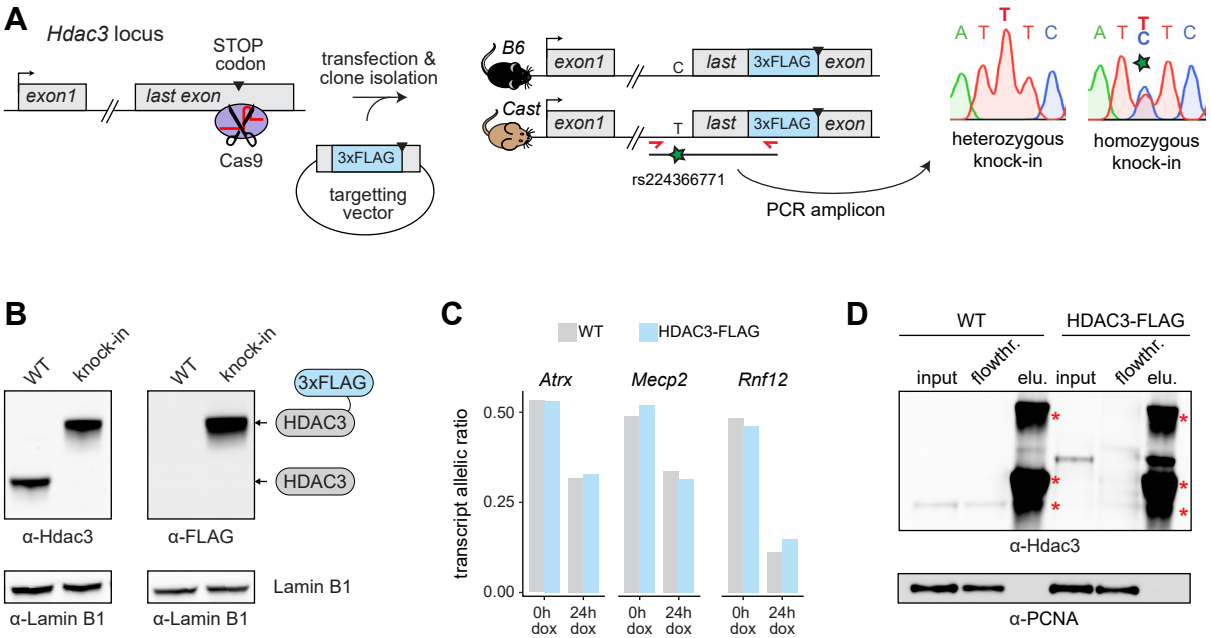


Figure 5.2 – Generation and characterization of HDAC3-FLAG TX1072 mESCs. **(A)** Scheme depicting CRISPR-mediated targeting of a 3xFLAG cassette at both alleles of the endogenous *Hdac3* gene. The insertion was targeted immediately upstream of the stop codon. Thanks to the polymorphic TX1072 background, homozygous knock-ins can be identified through Sanger sequencing of a PCR product spanning across the insertion site, and covering a SNP within the *Hdac3* locus (rs224366771). **(B)** Western blot detection of HDAC3 from wild-type and HDAC3-FLAG knock-in whole cell extracts. Blots were probed with anti-HDAC3 (left) or anti-FLAG (right) antibodies. **(C)** Barplots showing transcript allelic ratio (derived from pyrosequencing assays) for three X-linked genes before and after doxycycline-mediated induction of *Xist* for 24 hours, in WT and HDAC3-FLAG knock-in cells. **(D)** Western blot detection of HDAC3 following anti-FLAG bead-based immunoprecipitation in wild-type and HDAC3-FLAG nuclear extracts. Three fractions (input, flowthrough, and bead eluate) were sampled. * : signal corresponding to heavy and light chains from bead-bound FLAG IgG.

5.3.2 HDAC3 is pre-bound to the X-chromosome and functions strictly through NCoR/SMRT

I next optimized FLAG-based HDAC3 immunoprecipitation and identified conditions wherein the total pool of nuclear HDAC3 was immobilized, with no detectable traces of "free" HDAC3 following immunoprecipitation (**Figure 5.2D**). Under these conditions, we used mass spectrometry (MS) to characterize the protein interactome of HDAC3, in the presence of *Xist* RNA (see Zylicz et al., Figure 5B-C).

Importantly, none of the previously reported *Xist* RNA-binding proteins [Minajigi et al., 2015; Chu et al., 2015; McHugh et al., 2015] were identified in our analysis (see Zylicz et al., Figure 5C). Instead, the significantly enriched peptides retrieved mapped to the NCoR/SMRT complex – whose interaction with HDAC3 is known to be required for HDAC3 enzymatic activity [Guenther et al., 2001] – and the CCT/TRiC chaperonin which is required for proper HDAC3 folding [Guenther et al., 2002] (see Zylicz et al., Figure 5C-E).

Furthermore, anti-FLAG ChIP-seq during a timecourse of *Xist* induction revealed that HDAC3 binding on the X chromosome does not significantly increase during XCI (see Zylicz et al., Figure 5G). Instead, HDAC3 is pre-bound (i.e. prior to *Xist* induction) to X-linked enhancers (and a minute fraction of promoters, see Zylicz et al., Figure 5F and Figure S5C), to promote efficient histone deacetylation following *Xist* coating.

Consistently, regions that show higher pre-binding of HDAC3 show more efficient histone deacetylation upon *Xist* induction (see Zylicz et al., Figure 5H and Figure S5G-H).

Taken together with the IP-MS analysis, these results invalidate the previously proposed hypothesis putting forth that HDAC3 is actively recruited to the X chromosome through *Xist* coating [McHugh et al., 2015]. Instead, they favor a model wherein HDAC3 is "dormant" on the future inactive X, and is activated only through secondary factors brought in by*Xist*.

In the next chapter (**chapter 6**), I will describe my main PhD work aimed at characterizing SPEN's function during XCI. Importantly, the results obtained throughout this previous chapter will help complement the model for SPEN function which emerged during my PhD.

5.4 Article 1: The implication of early chromatin changes in X chromosome inactivation *(Cell. 2020)*

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

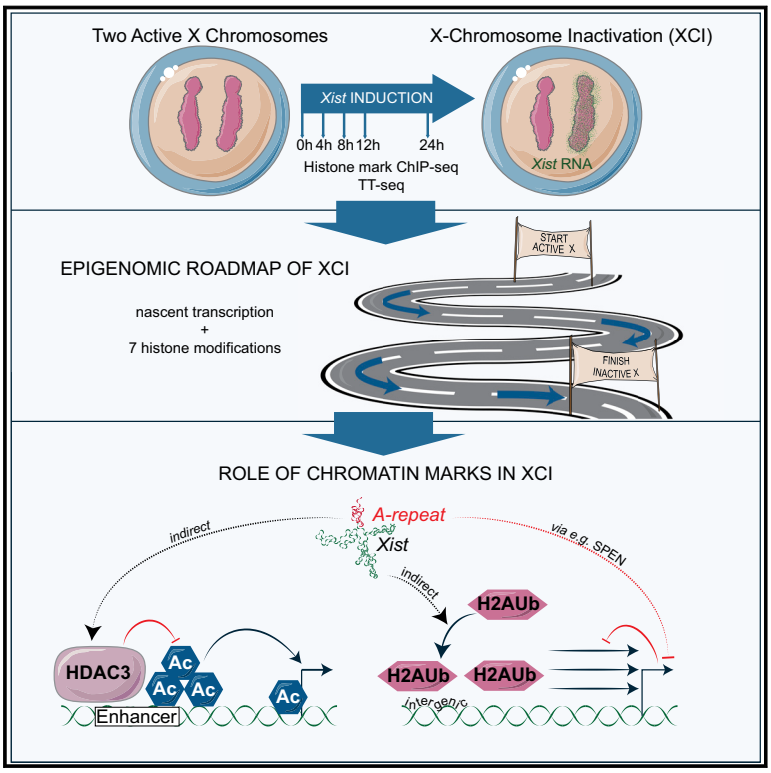
ARTICLE ON NEXT PAGE

Article

Cell

The Implication of Early Chromatin Changes in X Chromosome Inactivation

Graphical Abstract



Highlights

- An epigenomic roadmap for initiation of X chromosome inactivation (XCI)
- Histone deacetylation and H2A ubiquitination are among the earliest XCI events
- HDAC3-mediated histone deacetylation is required for efficient XCI
- PcG marks are first deposited intergenically and spread when gene silencing occurs

Authors

Jan Jakub Żylicz, Aurélie Bousard, Kristina Žumer, ..., Damarys Loew, Patrick Cramer, Edith Heard

Correspondence

edith.heard@curie.fr

In Brief

Żylicz et al. provide a detailed characterization of the earliest stages of X chromosome inactivation, tracing chromatin modification dynamics and uncovering the key role of chromatin changes in initiation of gene silencing.

Żylicz et al., 2019, Cell *176*, 182–197
January 10, 2019
https://doi.org/10.1016/j.cell.2018.11.041

CellPress

Cell

Article

The Implication of Early Chromatin Changes in X Chromosome Inactivation

Jan Jakub Żylicz,[1,2,6] Aurélie Bousard,[1,6] Kristina Žumer,[3] Francois Dossin,[1] Eusra Mohammad,[3] Simão Teixeira da Rocha,[4] Björn Schwalb,[3] Laurène Syx,[1] Florent Dingli,[5] Damarys Loew,[5] Patrick Cramer,[3] and Edith Heard[1,7,*]

[1]Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, UPMC Paris-Sorbonne, 75005 Paris, France
[2]University of Cambridge, Department of Physiology, Development and Neuroscience, Cambridge CB2 3EG, UK
[3]Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, 37077 Göttingen, Germany
[4]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal
[5]Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, Paris 75248 Cedex 05, France
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: edith.heard@curie.fr
https://doi.org/10.1016/j.cell.2018.11.041

SUMMARY

During development, the precise relationships between transcription and chromatin modifications often remain unclear. We use the X chromosome inactivation (XCI) paradigm to explore the implication of chromatin changes in gene silencing. Using female mouse embryonic stem cells, we initiate XCI by inducing *Xist* and then monitor the temporal changes in transcription and chromatin by allele-specific profiling. This reveals histone deacetylation and H2AK119 ubiquitination as the earliest chromatin alterations during XCI. We show that HDAC3 is pre-bound on the X chromosome and that, upon *Xist* coating, its activity is required for efficient gene silencing. We also reveal that first PRC1-associated H2AK119Ub and then PRC2-associated H3K27me3 accumulate initially at large intergenic domains that can then spread into genes only in the context of histone deacetylation and gene silencing. Our results reveal the hierarchy of chromatin events during the initiation of XCI and identify key roles for chromatin in the early steps of transcriptional silencing.

INTRODUCTION

Successful development requires the establishment and maintenance of euchromatin and heterochromatin in different parts of the genome. Progressively stable silencing of some genic regions occurs as epigenetic memory continues to accrue, finally leading to facultative heterochromatin formation (Trojer and Reinberg, 2007). Multiple layers of chromatin modifications are believed to enable stable transcriptional silencing. However, little is known about how facultative heterochromatin is dynamically formed and to what extent chromatin changes are involved in the establishment of gene silencing. A powerful model for developmentally induced gene silencing and formation of facultative heterochromatin is X chromosome inactivation (XCI) in female mammals. Although the role for chromatin changes in maintenance of the inactive state has been extensively studied (Disteche and Berletch, 2015), almost nothing is known about their role in the initiation of gene silencing.

In female mouse embryos, one of the two X chromosomes is randomly chosen for inactivation around the time of implantation (Lyon, 1962). This phenomenon is dependent on the coating of the future inactive X chromosome (Xi) by the long non-coding RNA *Xist* (Penny et al., 1996). The conserved A-repeat region of *Xist* mediates transcriptional silencing (Wutz et al., 2002), whereas other parts of *Xist* ensure chromosome coating (Almeida et al., 2017; Wutz et al., 2002). Assays coupling immunofluorescence with RNA fluorescence *in situ* hybridization (IF/RNA FISH) have revealed that, upon *Xist* RNA coating, a program of striking chromatin rearrangements ensues. These include rapid loss of histone modifications associated with active promoters (H3K4me3, H4ac, H3K9ac, and H3K27ac) and enhancers (H3K27ac and H3K4me1) (Chaumeil et al., 2002; Jeppesen and Turner, 1993). Furthermore, there is accumulation of H2AK119Ub and H3K27me3, two repressive histone marks dependent on the activity of Polycomb repressive complex (PRC) 1 and 2, respectively (for a review, see Brockdorff, 2017). These progressive alterations of chromatin states are associated with stable repression of the majority of genes on the Xi. Not all X-linked loci are affected in the same way, however, with some genes being silenced much faster than others or even entirely resisting repression ("escapees") (Disteche and Berletch, 2015). The reason for this striking diversity in gene inactivation dynamics remains unclear. It is likely that the susceptibility of loci to *Xist* spreading plays a role in this process (Borensztein et al., 2017; Engreitz et al., 2013); however, differences in their chromatin status could also underpin transcriptional silencing dynamics.

Although some chromatin marks have previously been mapped on the X chromosome in female embryonic stem cells (ESCs) and differentiated cells, nothing is known about the dynamics of the XCI process, especially during early stages when gene silencing actually takes place (Marks et al., 2009; Pinter et al., 2012). Specifically, the order of chromatin changes

and their possible role(s) in mediating transcriptional silencing remain largely unknown. Previous studies have relied on ESC differentiation for XCI initiation, which results in highly asynchronous *Xist* induction and elevated levels of heterogeneity in the population. These complications entirely mask primary chromatin events occurring in a small subset of cells embarking on the process of XCI. Recent advances in the identification of *Xist* binding proteins revealed that most histone-modifying activities are not directly recruited by *Xist* RNA but, rather, by its binding partners, like SPEN and HNRNPK (Chu et al., 2015; McHugh et al., 2015; Monfort et al., 2015; Pintacuda et al., 2017). A major question is whether the chromatin changes that occur early on in XCI are actually involved in gene silencing during XCI and, if so, how.

Here we investigate the earliest events accompanying transcriptional silencing of the X chromosome. Using a female ESC line in which one X chromosome can be specifically inactivated via an inducible *Xist* gene, we performed allele-specific native chromatin immunoprecipitation sequencing (ChIP-seq) as well as nascent transcript profiling at the initiation stages of XCI. Using 4-hr time resolution, we were able to reveal the precise order of chromosome-wide epigenetic events that are intricately coupled to gene repression during XCI. We find that loss of histone acetylation and, in particular H3K27ac, is one of the first events following *Xist* RNA accumulation during initiation of XCI. We also show that histone deacetylation, via HDAC3, is vital for efficient silencing of most genes on the *Xist*-coated X chromosome. Contrary to the prevailing hypothesis, HDAC3 is not acutely recruited to the X chromosome by *Xist*; rather, it is pre-loaded, mainly at putative enhancers. Upon *Xist* coating, pre-loaded HDAC3 likely mediates histone deacetylation and facilitates transcriptional silencing. We also uncover a surprisingly rapid accumulation of the PRC1-dependent H2AK119Ub mark on the X chromosome, particularly at intergenic regions lying in proximity to *Xist* RNA entry sites, which are pre-marked by Polycomb (PcG) marks. Accumulation of H3K27me3 appears slightly later and is delayed compared with gene silencing. Using *Hdac3* and *Xist* mutant cells, we also show that spread of Polycomb marks into gene bodies can only occur when transcriptional repression begins. Our study reveals the chromatin choreography across the X chromosome during the early steps of XCI initiation and uncovers a role for histone deacetylation in the establishment of gene silencing.

RESULTS

Allele-Specific Native ChIP-Seq Monitors Chromatin Changes during XCI

A molecular roadmap of the earliest chromatin changes during the initiation of XCI has not yet been established. To reduce heterogeneity and obtain a highly resolved time series of chromatin events upon *Xist* RNA coating, we used the TX1072 ESC line (Figure 1A; Schulz et al., 2014). This hybrid (*Mus musculus castaneus* x C57BL/6) line harbors a doxycycline (DOX) inducible promoter upstream of the endogenous *Xist* at the C57BL/6 (B6) allele (Schulz et al., 2014; Wutz et al., 2002). By adding DOX, we synchronously induce *Xist* expression and can thus decouple the onset of XCI from differentiation-dependent chromatin changes (Figure S1C). We performed allele-specific native ChIP-seq (nChIP-seq) for seven histone modifications across five time points at up to 4-hr resolution on biological duplicates (Figure 1A; Figures S1A and S1B). We focused on histone modifications associated with active promoters (H3K4me3, H3K9ac, H4ac, and H3K27ac) and active enhancers (H3K4me1 and H3K27ac) as well as repressive Polycomb-dependent marks (H3K27me3 and H2AK119Ub). All nChIP-seq datasets were sequenced to a depth of at least 40 million reads and showed a high signal-to-noise ratio (Figures S1B and S1E). The reads were split according to content of allele-specific SNPs (*B6*, mapping to Xi; *Cast*, mapping to active X); this allelic information was then analyzed in detail to define XCI-specific changes.

Histone Deacetylation Is One of the Earliest Events during Initiation of XCI

Upon *Xist* coating, the Xi becomes rapidly depleted of active histone modifications. Although this global loss of euchromatin occurs in a similar time window as the changes in X-linked transcriptional activity (Chaumeil et al., 2002), the precise chronology and possible mechanistic relationship of these two events have remained unclear. To address this, we first analyzed nChIP-seq datasets for a panel of active histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K9ac, and H4ac; Figure S1E). This revealed progressive depletion of B6-specific reads (originating from Xi) upon *Xist* induction, with different marks showing different kinetics of loss (Figure S1D). To analyze this in detail, we performed peak calling on all the reads (STAR Methods). For each peak with at least 50 allele-specific reads, we calculated its d-score ($reads^{B6}/[reads^{Cast} + reads^{B6}]$). This parameter of allelic skewing uses the active X chromosome as a powerful internal control, rendering the analysis less sensitive to variability in nChIP-seq quality and, thus, more stable between replicates.

We first set out to track the relative dynamics of active chromatin marks by comparing d-score dynamics of all peaks. We found that loss of H3K27ac and H4ac is among the earliest and most robust chromatin events on the X chromosome following *Xist* induction (Figure 1B). Furthermore, we found that not only did different histone modifications show distinct behaviors but, also, that specific regions of the X chromosome followed rather different dynamics of inactivation (Figure S1G). Thus, there is a temporal hierarchy of chromatin events both at the level of different histone marks and genomic loci.

To better understand in which context chromatin changes might affect gene silencing, we separately focused on promoters and putative enhancers. For each window, we plotted a d-score (normalized to t = 0 hr) as a function of time and fitted a sigmoidal curve (STAR Methods). We found that the maximal dynamic range of inactivation was from 0.5 to ~0.2 and, thus, set a threshold for when the curve reached 50% of this dynamic range; i.e., d = 0.35 (STAR Methods). This IC35 (inactivation criterium 0.35) parameter indicates how quickly a given peak becomes significantly decreased or lost (Figure S1F). Pairwise comparisons of the IC35 at X-linked gene promoters revealed that H3K27ac becomes efficiently deacetylated prior to demethylation of H3K4me3 (Figure 1C). This trend holds true for ~88% of measured promoters (Figure S1H), including a

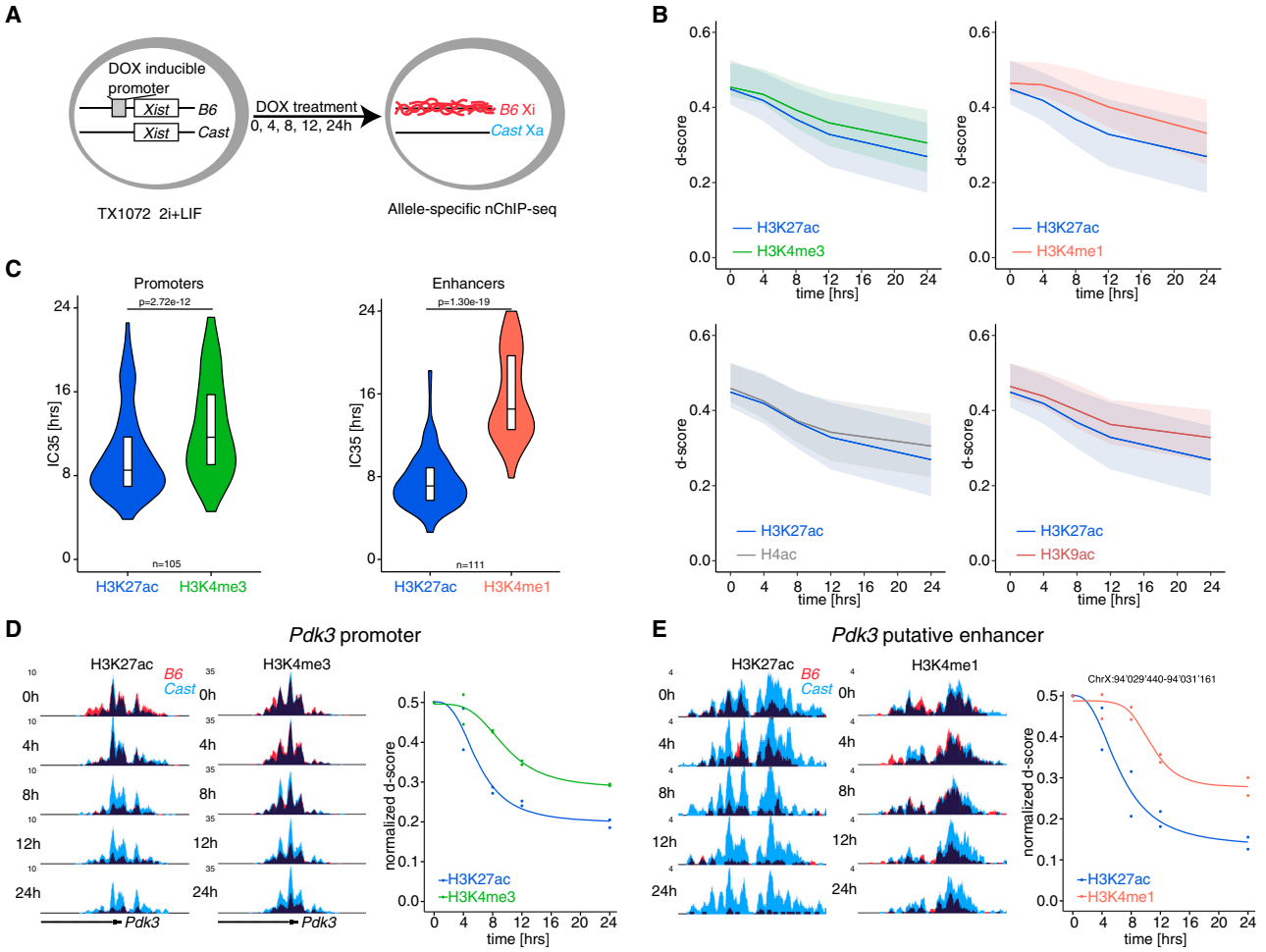


Figure 1. Histone Deacetylation Is among the First Events of XCI

(A) Schematic representation of the experimental design. The hybrid TX1072 mouse ESC line was used, in which *Xist* can be induced from the endogenous B6 allele.

(B) H3K27ac (blue) deacetylation dynamics in comparison with the loss of other active histone marks. Shown are average d-scores (shading is the interquartile range) of all peaks at the X chromosome.

(C) Paired comparison of the dynamics in H3K27ac and H3K4 methylation loss. Plotted are IC35 parameters of peaks at active promoters (left) and putative active enhancers (right) and within the quantitative range of the experiment (IC35 < 24 hr). The p value is from a paired Wilcoxon rank-sum test.

(D and E) Genome browser tracks showing H3K27ac and H3K4me3 at the *Pdk3* promoter (D) or H3K27ac and H3K4me1 at a putative *Pdk3* enhancer (E). Allele-specific tracks are overlaid (Cast reads in blue; B6 reads in red). The plots show the quantification of the inactivation dynamics with fitted sigmoidal curves. See also Figure S1.

representative example of the *Pdk3* promoter (Figure 1D). Similar analysis of the enhancer inactivation dynamics revealed that H3K27ac is efficiently depleted prior to loss of H3K4me1 (Figure 1C). This robust difference is observed in ~95% of the cases; e.g., at a putative *Pdk3* enhancer element (Figure 1E; Figure S1I). Thus, histone deacetylation is an early and efficient event during XCI.

Next, to extract information about relative timing rather than the efficiency threshold, we obtained the time when the curve reaches its maximum slope (effective dose 50%, ED_{50}; Figure S1F). ED_{50} analysis revealed that both H3K27ac and H3K4me3 become depleted concomitantly (Figure S1J). Together with the IC35 data, this indicates that, even though demethylation of promoters and their deacetylation are concurrent, the latter process is significantly more efficient. Furthermore, unlike promoters, enhancers showed significant differences in the ED_{50} parameter between H3K27ac and H3K4me1, indicating that enhancer demethylation is not only less efficient but also significantly delayed (Figure S1J).

To assess the relative deacetylation dynamics of enhancers and promoters, we compared the IC35 parameter for H3K27ac at both types of genomic regulatory elements. Intriguingly, enhancer deacetylation was slightly more efficient than that of promoters (Figure S1K). After pairing active promoters to their putative enhancers (STAR Methods), ~76% (37 of 49) of them showed more efficient deacetylation at enhancers than at

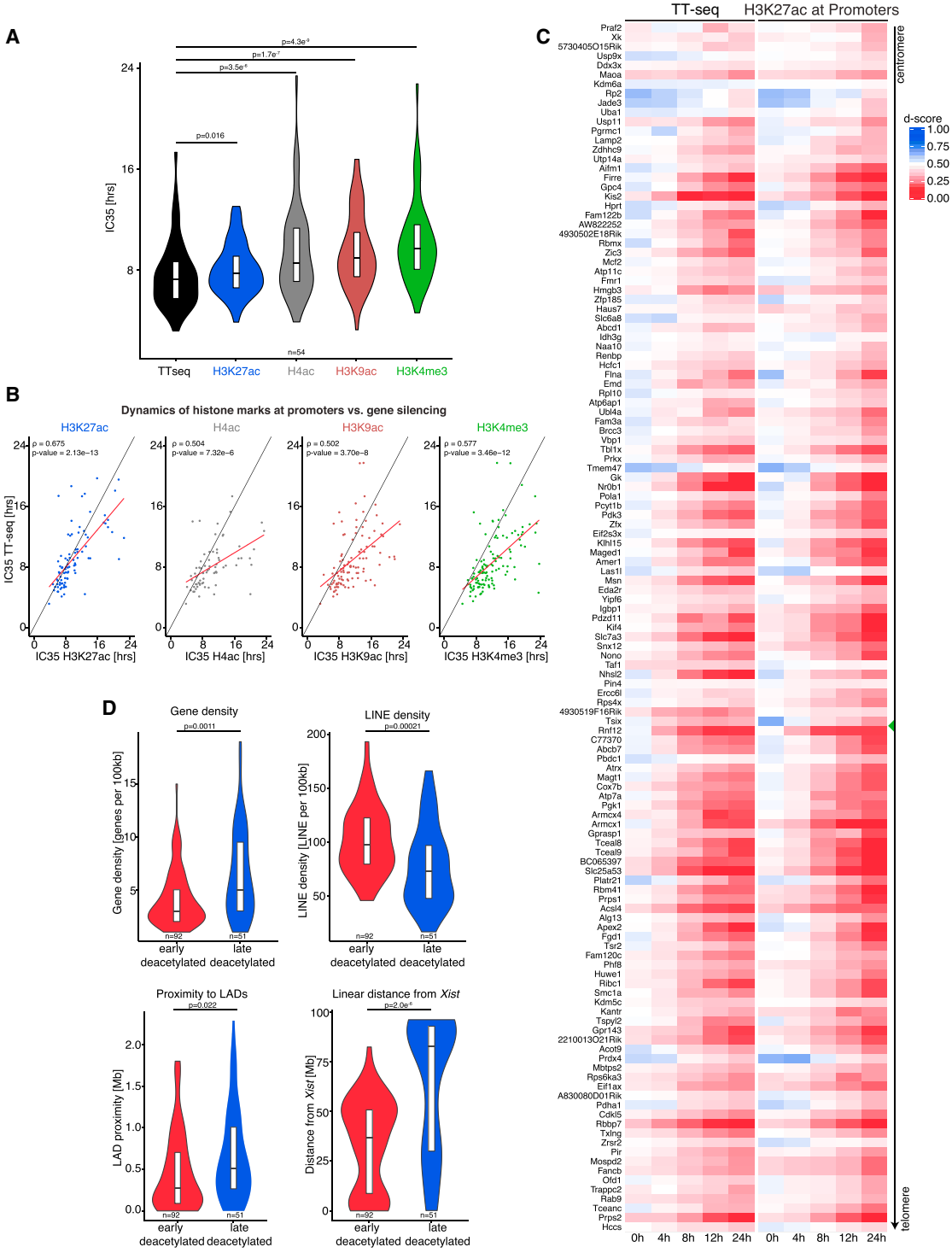


Figure 2. Histone Deacetylation Is Tightly Correlated with Transcriptional Silencing

(A) Comparison of IC35 parameters for gene silencing (TT-seq) and histone modifications at associated promoters. The p values are from a paired Wilcoxon rank-sum test.

(B) Pairwise comparison of IC35 from gene silencing and histone mark dynamics as in (A). Linear regression fitting (red) and perfect correlation with slope = 1 (black) is shown. Pearson's correlation (ρ) and p value are shown for each set.

(legend continued on next page)

promoters (Figure S1M). On the other hand, loss of H3K4me1 on the Xi, which is indicative of enhancer decommissioning, was found to be significantly less efficient compared with the loss of H3K4me3 at promoters (Figure S1L). Thus, inactivation (loss of H3K27ac) of enhancers proceeds at a slightly higher efficiency than of promoters, although stable decommissioning (loss of H3K4me) of promoters precedes that of enhancers. Intriguingly, upon H3K4me3 loss following *Xist* induction, we observe a slight increase in H3K4me1 levels around transcriptional start sites (TSSs), which probably reflects an intermediate of stepwise enzymatic H3K4me3 demethylation (Figures S1N and S1O).

In summary, our data show that efficient histone deacetylation at promoters and enhancers is one of the first events during XCI, indicating that it might contribute to initiation of transcriptional silencing. In comparison, demethylation of H3K4me3 and H3K4me1 is less efficient and delayed, respectively. Finally, deacetylation of promoters and enhancers seems to follow broadly comparable dynamics, whereas their decommissioning differs significantly.

Histone Deacetylation Is Tightly Linked to Transcriptional Silencing

To gain insight into the relationship between loss of active chromatin marks and transcriptional silencing of X-linked genes, we compared our nChIP-seq dataset with transient transcriptome sequencing (TT-seq; Schwalb et al., 2016). This allows a direct measure of newly synthesized RNA and, thus, enabled us to measure transcriptional activity as the X chromosome becomes inactivated. Comparison of the IC35 for TT-seq and individual histone modifications at associated promoters revealed that transcriptional silencing dynamics are most tightly linked to promoter deacetylation at H3K27 and, to a lesser extent, to loss of H4ac, H3K9ac, and H3K4me3 (Figures 2A and 2B; Figure S2A). A similar pattern of gene silencing and promoter deacetylation dynamics further exemplifies this correlation (Figure 2C).

The striking variability in dynamics of promoter chromatin changes begged the question of whether these differences could be due to transcriptional status or genomic context; for example, proximity to *Xist* (the locus and/or *Xist* RNAs entry sites; Engreitz et al., 2013), gene density, or LINE (long interspersed nuclear elements) density. Indeed, the latter feature has been proposed previously to influence the efficiency of XCI (Chow et al., 2010; Loda et al., 2017; Figure S2B and S2C). We found that gene promoters that are rapidly H3K27-deacetylated preferentially reside in gene poor, LINE-dense regions in proximity to lamina-associated domains (LADs) and the *Xist* locus (Figure 2D).

Given that the histone deacetylation is the primary chromatin change observed upon *Xist* induction and that its kinetics are tightly linked to transcriptional silencing, we decided to investigate whether the loss of histone acetylation might play a functional role in mediating the initiation of gene silencing during XCI.

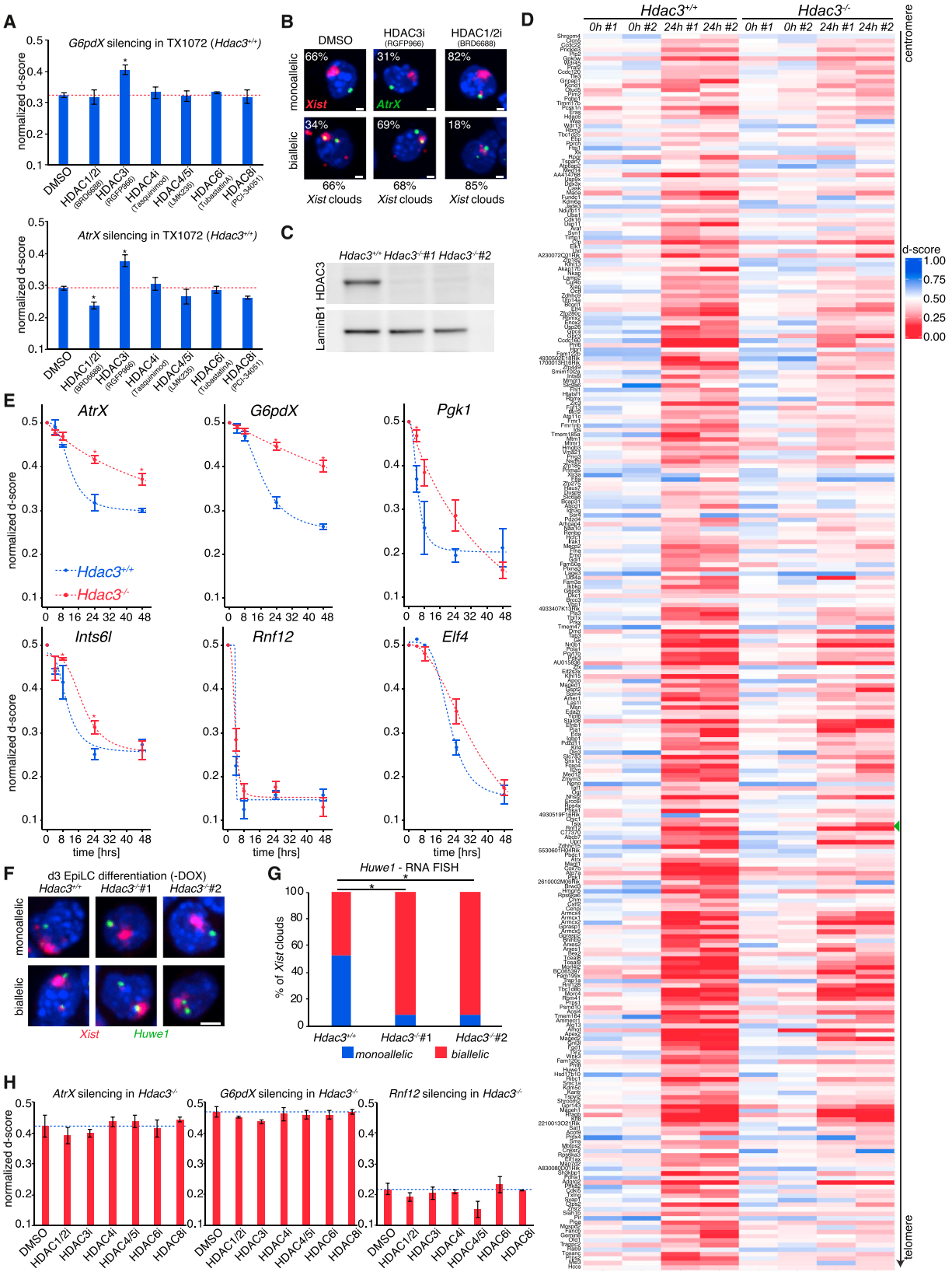
HDAC3 Mediates Efficient Transcriptional Silencing during XCI

To investigate whether this rapid histone deacetylation is mediated by histone deacetylases (HDACs), we first tested selective pharmacological inhibitors against all HDACs expressed in ESCs (Figure S3A). Importantly, after 24 hr of DOX induction, only HDAC3 catalytic inhibition resulted in a significant silencing defect compared with the control (DMSO) (Figure 3A). We further validated these findings using nascent RNA FISH and, unexpectedly, found that HDAC1 and 2 inhibition promoted more efficient *Xist* upregulation, possibly because of rtTA upregulation (Figure 3B). The finding that HDAC3 is involved in XCI is in line with previous reports that it might interact with SPEN, which is one of the key *Xist*-silencing factors (Chu et al., 2015; McHugh et al., 2015; Monfort et al., 2015). Also, *Hdac3* knockdown was recently shown to result in defective silencing of another X-linked gene: *Gpc4* (McHugh et al., 2015).

To better interrogate the specific functions of HDAC3 in XCI, we generated two independent *Hdac3*$^{-/-}$ TX1072 deletion clones using CRISPR/Cas9 technology (STAR Methods; Figure 3C). We performed RNA sequencing (RNA-seq) on these, both prior to and 24 hr following DOX addition. Differential gene expression analysis revealed relatively mild changes apart from some meiosis-associated genes (Figures S3B–S3D). Consistent with our inhibitor treatment results, allele-specific analysis revealed that both *Hdac3*$^{-/-}$ clones are unable to efficiently silence X-linked genes (Figure S3E). Although the majority of genes across the X chromosome were affected, ~24% of genes (43 of 182) were not, showing, instead, almost normal silencing (Figure 3D; Figures S3F and S3G). We further validated our findings in the context of EpiLC (epiblast like-cells) differentiation both with and without DOX treatment (Figures 3E–3G; Figure S3H). In our pyrosequencing experiments, we monitored the allelic bias in expression of genes affected (*AtrX*, *G6pdX*, and *Pgk1*) in their silencing by the HDAC3 loss and three less affected genes (*Ints6l*, *Rnf12*, and *Elf4*) (Figure 3E).

Our results indicate that loss of HDAC3 leads to a delay in transcriptional silencing of most X-linked genes. We noted that, over time, XCI is not fully prevented because gene silencing progressively ensues, although more slowly than in control cells (Figure 3E), indicating that mechanisms parallel to HDAC3 must also contribute to XCI. To address this further, we tested whether other HDACs might compensate for HDAC3 loss. To this end, we treated *Hdac3*$^{-/-}$ ESCs with selective inhibitors against all expressed HDACs (1,2,3,4,5,6, and 8). We found that allelic skewing (i.e., XCI) of *AtrX*, *G6pdX*, and *Rnf12* were not affected in any of the treated cells (Figure 3H), arguing against the involvement of other HDACs in XCI. Thus, our data reveal that HDAC3 is the main histone deacetylase important for the initiation of silencing of most genes on the X chromosome. Because HDAC3 can mediate deacetylation of many substrates, we

(C) Heatmap showing gene silencing and associated promoter H3K27 deacetylation dynamics. Rows were ordered by their genomic position from centromere to telomere (the green triangle indicates the *Xist* location). The averages from biological duplicates are shown.
(D) Quantification of features associated with early and late deacetylated promoters (H3K27ac). Shown are gene density, LINE density, proximity to LADs, and distance from the *Xist* locus. The p values are from a Wilcoxon rank-sum test with Benjamini Hochberg correction.
See also Figure S2.



(legend on next page)

next investigated whether its effect on gene silencing is mediated by changes in chromatin status.

HDAC3 Promotes Silencing by Rapidly Deacetylating Histones

Although our experiments using inhibitors and knockout ESCs indicated the importance of HDAC3 catalytic activity in mediating efficient XCI, it remained unclear whether histone deacetylation dynamics were also affected by loss of HDAC3. To address this question, we first monitored the global distribution of histone modifications by IF/RNA FISH. We observed that the Xi silent compartment formed much less efficiently in *Hdac3*$^{-/-}$ ESCs, with increased levels of H3K27ac and H4ac remaining within the *Xist* cloud (Figures 4A and 4B; Figures S4A–S4D). To further explore the chromatin phenotype of *Hdac3*$^{-/-}$ ESCs, we performed nChIP-seq for H3K27ac and H4ac at 0, 8, and 24 hr of DOX treatment. Differential peak analysis has revealed that less than 9% of peaks significantly change in *Hdac3*$^{-/-}$ compared with WT cells at t = 0 hr (Figure S4E). Consistent with the IF/RNA FISH results, allele-specific analysis revealed delayed histone deacetylation in the *Hdac3*$^{-/-}$ clones compared with the parental line (Figure 4C; Figures S4F and S4G). As with transcriptional silencing, the deacetylation defects affected most of the X chromosome but were not uniform, with some peaks being almost entirely resistant to HDAC3 loss (~29%, 38 of 133). Indeed, protracted loss of H3K27ac at promoters was linked to a severe defect in transcriptional silencing in mutant cells (Figures 4C and 4D). Together, our data show that HDAC3 mediates efficient transcriptional silencing of the majority of genes through deacetylating histones. However, it remains unclear how HDAC3 is targeted to the X chromosome to help silencing.

HDAC3 Is Pre-bound on the X Chromosome, Likely to Promote Deacetylation during XCI

It has been proposed that HDAC3 might be either *de novo* recruited to the Xi by *Xist* through interaction with the SPEN protein (McHugh et al., 2015) or, alternatively, that *Xist*-coating might lead to the activation of already pre-bound HDAC3. To test these two hypotheses, we measured sub-nuclear localization of HDAC3 after 24 hr of DOX treatment. IF analysis revealed no HDAC3 enrichment on the Xi and even a slight depletion (Figure 5A). This surprising result prompted us to further explore the question of HDAC3 recruitment. To circumvent problems with antibody quality, we knocked in a 3xFLAG tag into the 3′ region of the endogenous *Hdac3* gene (Figure 5B). This novel ESC line (*Hdac3*$^{FLAG/FLAG}$) expressed nearly normal levels of HDAC3 (Figure S5A) and robustly silenced the X chromosome, indicating that HDAC3 remained functional (Figure S5B).

We then set out to identify potential common interactors of *Xist* and HDAC3 in DOX-treated cells. To this end, we performed anti-FLAG immunoprecipitation followed by quantitative label-free mass spectrometry analysis (*Hdac3*$^{FLAG/FLAG}$ versus *Hdac3*$^{+/+}$ in five replicates) (Figure 5C). Among significant interactors, we identified nearly all components of the NCOR and SMRT complexes as well as their chaperone complex CCT (Figures 5D and 5E; Guenther et al., 2002). Intriguingly, of all *Xist* binding partners (Chu et al., 2015), only RYBP (a PRC1 factor) showed a weak binding to HDAC3 (Figure 5C). Our results confirm the previously reported interactions between HDAC3 and NCOR and SMRT but argue against strong HDAC3 recruitment by *Xist* RNA binding partners like SPEN.

To directly measure HDAC3 recruitment to the Xi, we performed anti-FLAG ChIP-seq prior to and during XCI. We found robust binding of HDAC3 to multiple H3K27ac and H4ac peaks throughout the genome (~65% of HDAC3 peaks are enriched for H3K27ac [~65% on the X chromosome] and ~61% for H4ac [~58% on the X chromosome]) (Figure 5F). The vast majority of HDAC3 peaks were intriguingly found to reside within putative enhancers rather than at gene promoters (Figure 5F; Figure S5C). Allele-specific analysis revealed a transient and modest enrichment after 4 hr of induction on the Xi (B6 reads) at the *Xist* locus and at some HDAC3 peaks (Figure 5G; Figures S5D and S5E). Together, this indicates that HDAC3 indirectly and unstably interacts with some of the *Xist* binding partners to allow for low-level and transient recruitment to the X chromosome. However, no significant correlation between HDAC3 *de novo* recruitment and timing of deacetylation was found (Figure S5F). On the other hand, the histone acetylation peaks that are lost early during XCI did show higher levels of HDAC3 pre-binding (Figure S5G). In fact, HDAC3-bound regions show more efficient (low IC35) deacetylation of H4ac and a similar but statistically not significant trend for H3K27ac (Figure 5H; Figure S5H). Finally, we observed a mild but statistically insignificant

Figure 3. HDAC3 Mediates Efficient Transcriptional Silencing during XCI

(A) Screening for HDAC activity involved in XCI using selective inhibitors and pyrosequencing for *G6pdX* and *AtrX*. The red line represents silencing in the DMSO control after 24 hr of DOX treatment. Shown are averages (± SD) from 2 experiments. *p < 0.05, Student's t test.

(B) RNA FISH for *Xist* (red) and the X-linked *AtrX* gene (green) on ESCs treated as in (A). Percentages of biallelic and monoallelic *AtrX* expression in *Xist*-expressing cells are shown (>100 nuclei quantified, scale bar, 10 μm). The quantification of *Xist* cloud formation is shown below.

(C) Immunoblot analysis of whole-cell lysates from TX1072 (*Hdac3*$^{+/+}$) and two *Hdac3*$^{-/-}$ clones probed with antibodies specific for the N terminus of HDAC3 and Lamin B1.

(D) Heatmap showing gene silencing after 0 and 24 hr of DOX treatment in TX1072 (*Hdac3*$^{+/+}$) and two *Hdac3*$^{-/-}$ clones. Rows were ordered by their genomic position from centromere to telomere (the green triangle indicates the *Xist* location).

(E) RNA-seq validation using pyrosequencing on ESCs (TX1072: *Hdac3*$^{+/+}$ and *Hdac3*$^{-/-}$) differentiated to EpiLCs with DOX. Shown are averages (± SEM) from 3 independent experiments and two *Hdac3*$^{-/-}$ clones for 3 genes sensitive to HDAC3 loss (top) and 3 genes less dependent on its activity (bottom). *p < 0.05, Student's t test.

(F and G) Validation of *Hdac3*$^{-/-}$ phenotype in day 3 EpiLCs differentiated without DOX. (F) FISH was performed with probes for *Xist* (p510, red) and an X-linked gene, *Huwe1* (green). Scale bar, 5 μm. (G) Also shown is quantification of *Huwe1* silencing status in more than 100 nuclei with *Xist* clouds. χ^2 test, *p < 0.05.

(H) Screening for HDAC activity compensating for HDAC3 loss using pyrosequencing for *AtrX*, *G6pdX*, and *Rnf12*. The experiment was performed like in (A) but in *Hdac3*$^{-/-}$ cells. Shown are averages (± SD) from two independent clones. *p < 0.05, Student's t test.

See also Figure S3 and Table S1.

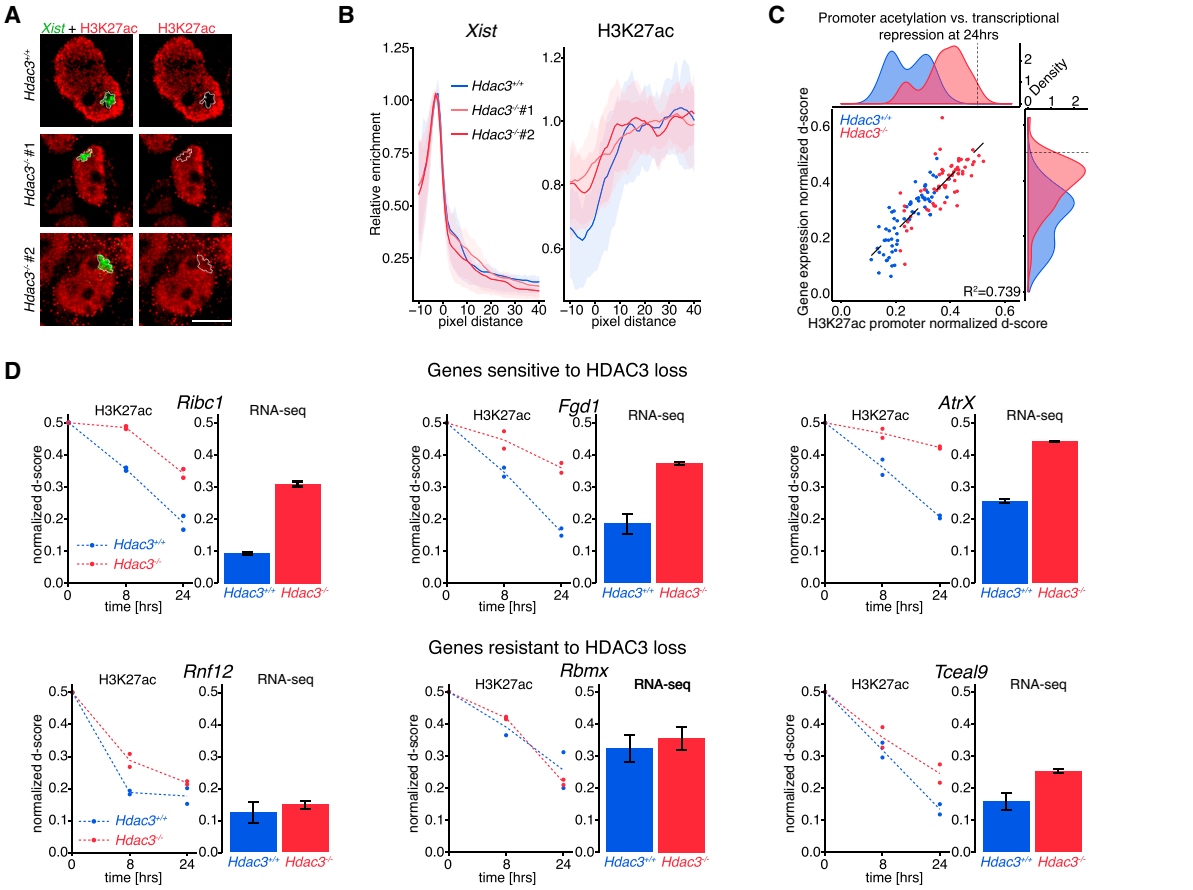


Figure 4. HDAC3 Promotes Silencing by Rapidly Deacetylating Histones

(A) IF/RNA FISH on TX1072 (*Hdac3*$^{+/+}$) and two *Hdac3*$^{-/-}$ ESC clones induced with DOX for 24 hr. Cells were probed using anti-H3K27ac (red) antibodies and a *Xist* intronic probe (green). The white dotted line encircles the *Xist* domain. Because of technical variability, contrast settings were individually adjusted to reflect relative enrichment. Scale bar, 10 μm.

(B) Quantification of (A) over at least 100 *Xist* clouds. Shown is the average signal for *Xist* and H3K27ac when profiles were aligned to the *Xist* cloud boundary (for more details, refer to Figure S4A and D).

(C) Relationship between gene repression (RNA-seq) and promoter deacetylation (H3K27ac) at 24 hr in TX1072 (*Hdac3*$^{+/+}$, blue) and *Hdac3*$^{-/-}$ (red). On the scatterplot, the black dotted line represents fitted linear regression. ρ = 0.860. Density plots show the distribution of the normalized d-score for H3K27ac (horizontal) and RNA-seq (vertical); note that d-scores in mutant cells are shifted toward biallelic levels (0.5, dotted lines on density plots). Each dot represents the average normalized d-score from biological duplicates.

(D) Comparison of promoter deacetylation dynamics and transcriptional silencing after 24 hr of DOX treatment in TX1072 (*Hdac3*$^{+/+}$) and *Hdac3*$^{-/-}$ ESCs. Examples of genes sensitive (top) and more resistant (bottom) to HDAC3 loss are shown. For RNA-seq, shown is the average from biological duplicates (± SD). See also Figure S4.

trend for HDAC3 pre-bound enhancers to surround rapidly silenced genes (Figure S5I). In summary, we show that HDAC3 does not appear to be robustly recruited *de novo* to the X chromosome via *Xist* RNA coating and does not strongly bind known partners of *Xist*. Rather, the majority of HDAC3 is pre-bound at putative enhancers, and, upon *Xist* induction, this results in efficient histone deacetylation, promoting transcriptional silencing.

PRC1-Dependent H2AK119Ub Accumulation Precedes H3K27me3 Deposition

Repressive chromatin modifications are also known to appear shortly after *Xist* RNA coating, in particular H2AK119Ub (PRC1-dependent) and H3K27me3 (PRC2-dependent). However, the early dynamics of this process, as well as the relative timing of H3K27me3 and H2AK119Ub deposition, have remained largely unexplored. To address this question, we have performed nChIP-seq for both PcG-dependent marks. These datasets showed a reproducible and stable pattern of enrichment on autosomes, both between time points and biological replicates (Figure S6A). In contrast, the X chromosome showed a bias in H2AK119Ub enrichment toward the B6 allele (Xi) after 4 hr of DOX induction and after 8 hr in the case of H3K27me3 (Figure S6B). First, we analyzed relative B6-read enrichment within 10-kb windows across the whole X chromosome normalized to t = 0 hr and observed rapid deposition of H2AK119Ub

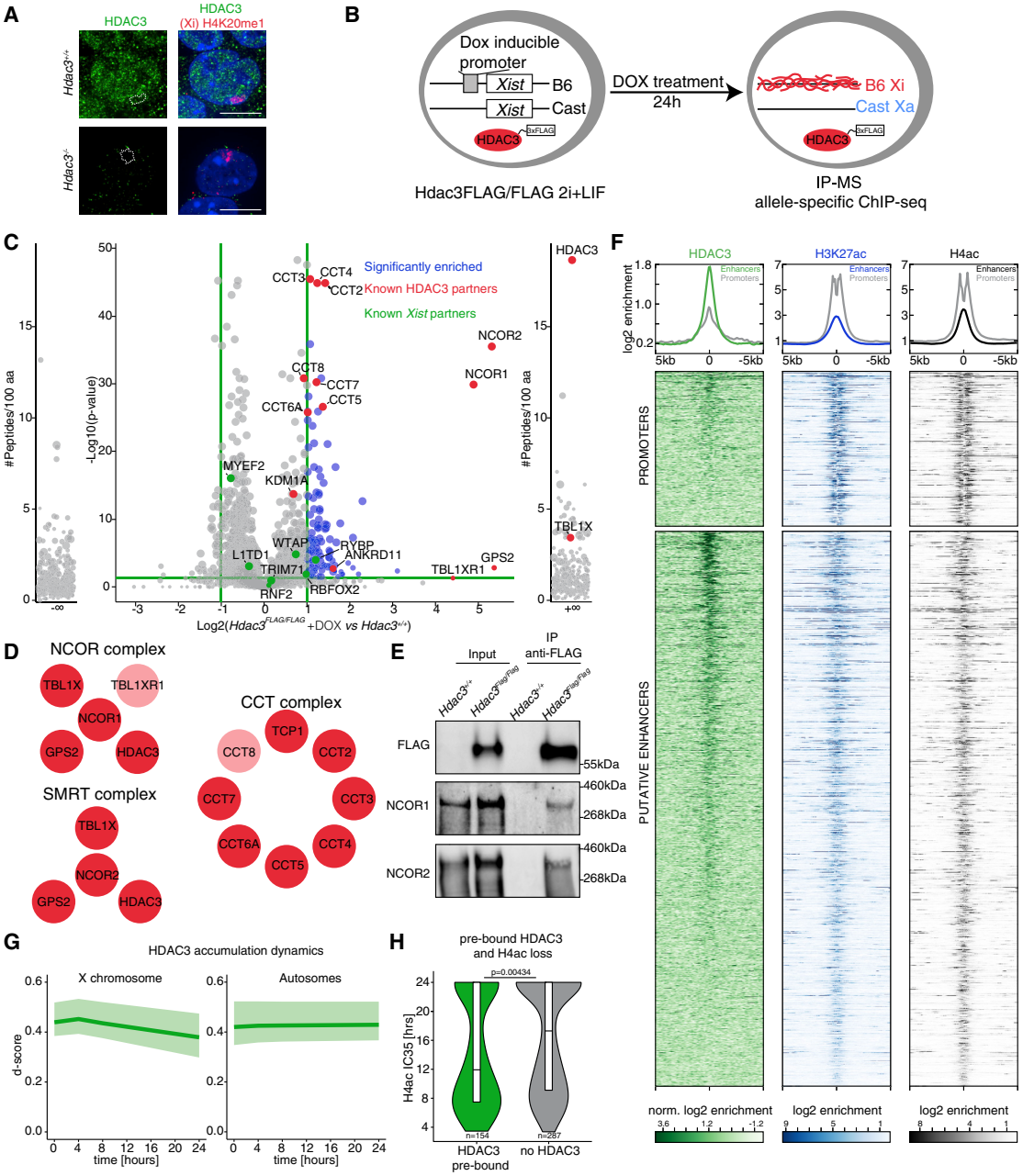


Figure 5. HDAC3 Is Pre-bound on the X Chromosome, Likely to Promote Efficient Histone Deacetylation

(A) HDAC3 sub-nuclear localization in TX1072 (*Hdac3*$^{+/+}$) and *Hdac3*$^{-/-}$ ESCs induced with DOX for 24 hr. Xi was detected by H4K20me1 enrichment (dotted area) and shows no enrichment for HDAC3. Scale bar, 10 μm.

(B) Schematic representation of the *Hdac3*$^{FLAG/FLAG}$ ESC line, which expresses endogenous HDAC3 with a 3xFLAG tag.

(C) Volcano plot analysis identifying interactors of HDAC3 in ESCs. Binding partners were obtained by using quantitative label-free mass spectrometry analysis performed from five replicates. Shown are the fold changes (*Hdac3*$^{FLAG/FLAG}$ versus *Hdac3*$^{+/+}$), quantified with an absolute fold change of 2 or more with an adjusted p value of ratio significance of 0.05 or less and with 3 or more peptides. Known HDAC3 (red) and *Xist* (green) interactors are shown. External plots show proteins with peptides identified only in one sample type (left in *Hdac3*$^{+/+}$ and right in *Hdac3*$^{FLAG/FLAG}$).

(D) Representation of complexes identified as HDAC3 interactors, with pink proteins representing partners, which did not reach significance thresholds.

(E) Immunoprecipitation; western blot validating HDAC3 interaction with both NCOR1 and NCOR2.

(F) Heatmap of HDAC3 enrichment at all H3K27ac and H4ac peaks across the genome. Metaplots show average enrichment of HDAC3, H3K27ac, and H4ac at promoters and putative enhancers. For HDAC3, the signal was normalized to the no-FLAG ChIP-seq control.

(legend continued on next page)

preceding that of H3K27me3 (Figure 6A). To quantify the dynamics at individual windows, we extracted the ED_{50} parameter, as was done for active marks, but using B6-read accumulation (STAR Methods; Figure S1F). Pairwise comparison of ED_{50} at all 10-kb windows across the X chromosome showed that H2AK119Ub accumulates most efficiently prior to H3K27me3 (Figure 6B), with very few exceptions (~1%) (Figure S6C).We also validated this finding using IF/RNA FISH during ESC differentiation toward EpiLCs (Figure S6D). Evidently, the Xi becomes enriched for H2AK119Ub prior to the PRC2-dependent H3K27me3 across the whole chromosome, although there may be local differences in the dynamics of this process. Indeed, certain regions preferentially accumulate Polycomb marks already at early time points (i.e., by 4 hr for H2AK119Ub or 8 hr for H3K27me3; Figure 6D). This early accumulation was found to be prominent around the *Xist* locus (green line) as well as the first regions with which it associates ("entry sites," blue lines) (Engreitz et al., 2013). There is also significant overlap between H2AK119Ub and H3K27me3 patterns of accumulation, and, interestingly, these primary deposition sites are typically marked by PcG prior to *Xist* upregulation (Figure 6C, Figure S6E). Thus, most H2AK119Ub accumulation occurs at sites pre-marked by PcG and rapidly coated by *Xist*. Subsequent H3K27me3 deposition closely follows this early deposition pattern.

We next investigated the accumulation dynamics of both marks at specific types of genomic windows. ED_{50} analysis revealed that H2AK119Ub accumulates first intergenically and is deposited only later at active gene bodies, promoters, and putative enhancers (Figure S6F). This pattern of accumulation is tightly followed by H3K27me3 deposition and spread. What is more, PcG-dependent mark accumulation seems to be closely linked to gene silencing kinetics (Figure 6E). This is exemplified by tracks for genes that are either silenced rapidly (*Rnf12*) or slowly (*Huwe1* and *Alg13*) or escape the process altogether (*Kdm6a*) (Figures 6F and 6G). Direct comparison of gene silencing and promoter deacetylation dynamics with PcG mark accumulation at gene bodies revealed a significantly lower ED_{50} parameter for H2AK119Ub (Figure S6G). This indicates that the H2AK119Ub deposition reaches its maximum rate earlier than gene silencing and promoter deacetylation. Next, we measured the time when the sigmoid's second derivative reaches its minimum or maximum. We use this parameter as a proxy for when the curve starts to increase or decrease (Figure S6I). This revealed that gene silencing, promoter deacetylation (H3K27ac loss), and H2AK119Ub accumulation at gene bodies are all initiated at the same time, whereas H3K27me3 is significantly delayed (Figure S6H). Thus, we see dynamic spreading of H2AK119Ub from intergenic regions to gene bodies coinciding with transcriptional silencing.

We also investigated genes that normally escape from XCI (constitutive) or escape more variably and in specific tissues (facultative escapees) (Disteche and Berletch, 2015). We examined a highly expressed constitutive escapee (*Kdm6a*; Figures 6F and 6G) and found no PcG-dependent mark accumulation at its promoter or gene body in our dataset. In line with this, both constitutive and facultative escapees show no or lower accumulation of H2AK119Ub and H3K27me3 at their gene bodies (Figures S6J and S6K). This is consistent with significantly delayed H3K27ac loss at their promoters (Figure S6L). Thus, genes escaping transcriptional silencing do not undergo promoter deacetylation and do not accumulate PcG-associated marks.

All in all, our analysis reveals the hierarchy of PcG marks during XCI, with H2AK119Ub accumulating significantly prior to H3K27me3. This process is first initiated at large chromatin domains that are pre-marked by PcG (megabase scale) and that correspond to some of the first regions with which *Xist* RNA interacts (Engreitz et al., 2013). We demonstrate that PcG marks accumulate first intergenically and subsequently spread into regulatory elements and gene bodies. This spread into genes coincides with transcriptional silencing, raising the question of whether the two processes are functionally linked.

Transcriptional Silencing Is Necessary for PcG Spreading

To address whether spreading of these repressive histone modifications from intergenic regions into genes is a cause or consequence of transcriptional silencing, we decided to investigate PcG mark distribution in a *Xist* mutant lacking the A-repeat element (*Xist:ΔA*). *Xist:ΔA* RNA is able to coat the X chromosome and lead to PcG mark accumulation (based on IF) but cannot initiate gene silencing (Kohlmaier et al., 2004; Sakata et al., 2017; Schoeftner et al., 2006; Wutz et al., 2002). *Xist:ΔA* thus allows the relationship between repressive chromatin changes and transcriptional repression to be addressed. For this, we used previously published male ESCs harboring wild-type (TXY:*Xist*) or mutant *Xist* (TXY:*XistΔA*) under a DOX-inducible promoter at the endogenous locus (Wutz et al., 2002; Figure 7A). These ESCs do not require allele-specific analysis, as all X-mapping reads will originate from the single X chromosome. After collecting DOX-treated and untreated differentiated cells (48 hr), we performed nChIP-seq for three histone modifications (H3K27ac, H3K27me3, and H2AK119Ub) in biological duplicates. Samples were normalized for ChIP quality prior to further analysis (STAR Methods).

First, we investigated the relationship between PcG-associated mark spread and transcriptional silencing in WT and *Xist:ΔA* cells. We observed that most regions of the X chromosome recruit nearly normal levels of H3K27me3 in the *Xist:ΔA* cells. However, clusters of active genes are a notable exception to this (Figure 7B; Figure S7A). The same regions are also found not to accumulate H2AK119Ub. Intriguingly, reduced recruitment of H2AK119Ub was observed throughout the chromosome in *Xist:ΔA* cells compared with WT *Xist* (Figure 7B). We next quantitated the deposition of both Polycomb marks in relation to regions with actively transcribed genes (Figure 7C). Indeed,

(G) HDAC3 accumulation dynamics (shading is the interquartile range) on peaks at the X chromosome (left) and autosomes (right).
(H) Comparison of peak deacetylation (H4ac) dynamics (IC35) and HDAC3 pre-binding. The p value is from a Wilcoxon rank-sum test.
See also Figure S5 and Table S2.

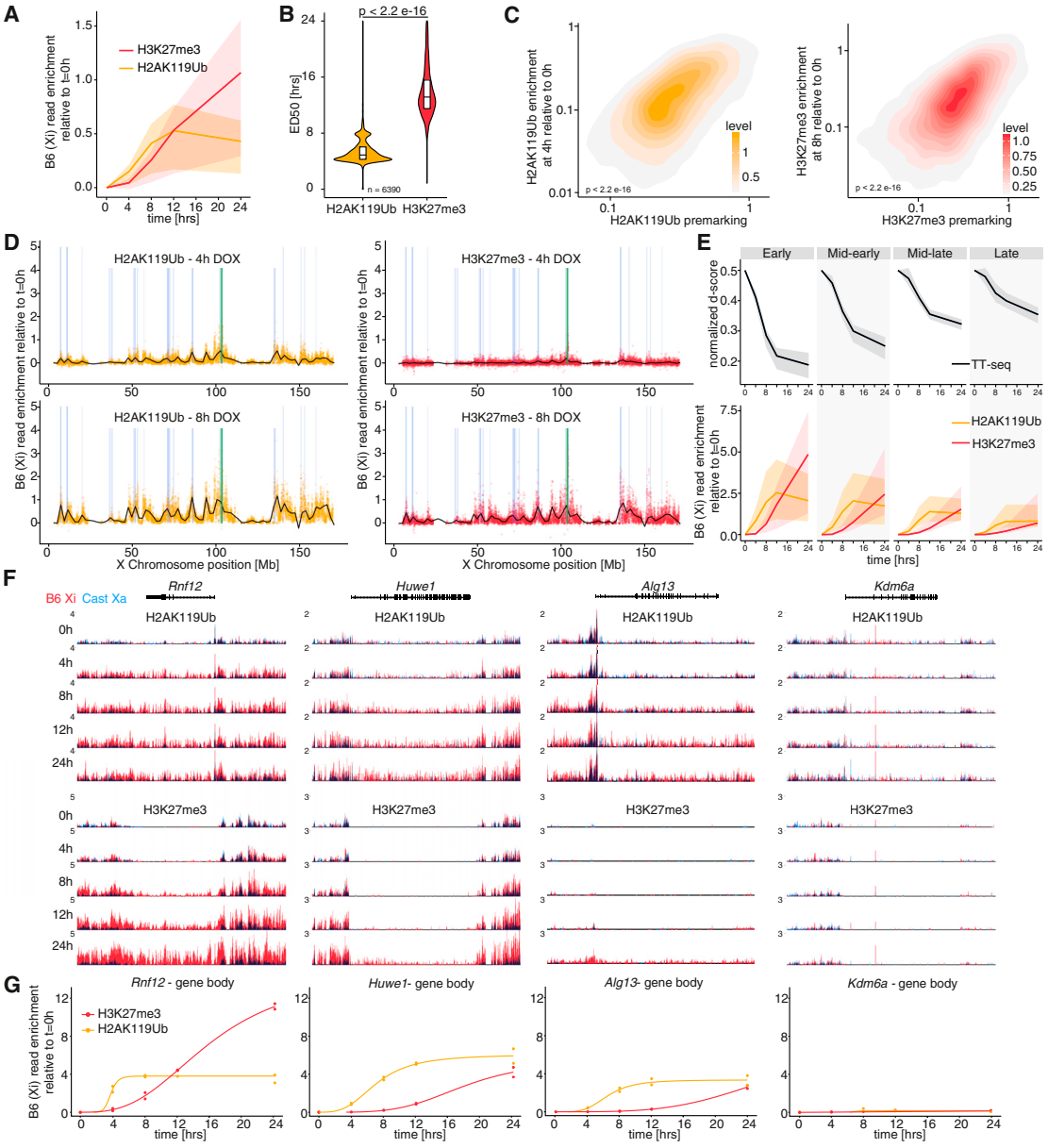


Figure 6. PRC1-Dependent H2AK119Ub Accumulation Precedes H3K27me3 Deposition

(A) Quantification of average H2AK119Ub (yellow) and H3K27me3 (red) enrichment at the Xi (B6 allele) compared with t = 0 hr in 10-kb windows spanning the whole chromosome. Shading is the interquartile range.

(B) Pairwise comparison of H2AK119Ub and H3K27me3 accumulation dynamics (ED_{50}) at the X chromosome. All 10-kb windows with $ED_{50} < 24$ hr are plotted. The p value is from a paired Wilcoxon rank sum test.

(C) Correlation between H2AK119Ub (left) or H3K27me3 (right) pre-marking (t = 0 hr) and *de novo* enrichment of those marks shortly after *Xist* RNA induction (t = 4 hr for H2AK119Ub and t = 8 hr for H3K27me3). Plotted are all 10-kb windows spanning the X chromosome. All scales are logarithmic. The p values are from Pearson's correlation test with $\rho = 0.538$ (H2AK119Ub) and 0.441 (H3K27me3).

(D) H2AK119Ub (yellow) and H3K27me3 (red) accumulation dynamics across the Xi after 4 and 8 hr of DOX treatment. The black line is a locally estimated scatterplot smoothing (LOESS) regression on all 10-kb windows (dots). Shown is the *Xist* locus (green) and *Xist* entry sites (blue).

(E) Dynamics of transcriptional silencing (top) and H2AK119Ub (yellow) and H3K27me3 (red) accumulation at gene bodies. Genes were separated based on transcriptional silencing dynamics. Shading is the interquartile range.

(F) Genome browser tracks showing H2AK119Ub (top) and H3K27me3 (right) accumulation at genes silenced rapidly (*Rnf12*), slowly (*Huwe1, Alg13*), or not at all (*Kdm6a*). Allele-specific tracks were overlaid (B6 in red and Cast in blue).

(G) Quantification of PcG mark accumulation at gene bodies (from F) with sigmoidal curves fitted.

See also Figure S6.

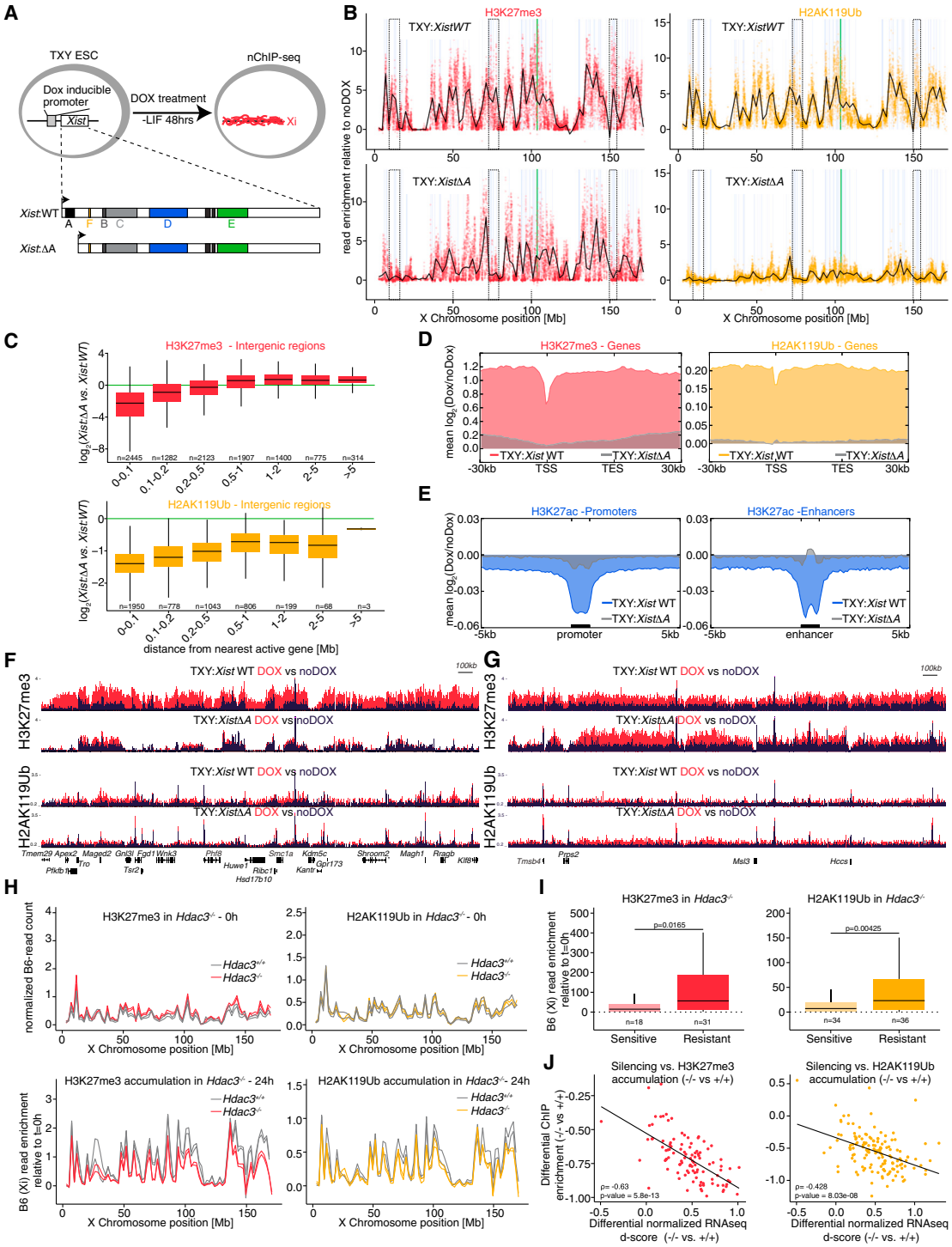


Figure 7. PcG Spreading along the X Chromosome Requires Transcriptional Silencing

(A) Schematic representation of the experimental design. By DOX addition, male TXY mouse ESC lines allow expression of *Xist* or *Xist:ΔA*.

(B) H3K27me3 (red) and H2AK119Ub (yellow) accumulation across the Xi after 48 hr of DOX compared with no DOX in WT (top) and *Xist:ΔA* (bottom) cells. Each dot represents a single 10-kb window. The black line indicates a LOESS regression on all windows. Shown are the *Xist* locus (green line), active genes (light blue), and examples of gene-dense regions (dotted boxes).

(C) Correlation between the distance from an active gene and the H3K27me3 (top) or H2AK119Ub (bottom) accumulation in *Xist:ΔA* cells compared with WT cells.

(legend continued on next page)

both marks show negligible accumulation in the *Xist:ΔA* mutant when within 100 kb from active genes. On the other hand, regions more distal (a minimum of 200 kb away) to transcribed loci gain normal levels of H3K27me3 and intermediate levels of H2AK119Ub, as exemplified in genome browser tracks (Figures 7F and 7G). Thus, *Xist:ΔA* leads to PRC2 and PRC1 recruitment to regions without active transcription; however, both complexes seem impeded from spreading into domains of active genes. In line with this, we observed no H3K27me3 or H2AK119Ub enrichment within promoters and bodies of active genes (Figure 7D; Figures S7B–S7E).

Our finding that H3K27 deacetylation promotes gene silencing during XCI prompted us to investigate H3K27ac loss at gene regulatory elements upon coating by *Xist:ΔA* RNA (Figure 7E; Figures S7C and S7E). Importantly, we did not observe significant deacetylation of either promoters or enhancers in the *Xist:ΔA* mutant. This was in stark contrast to robust inactivation of such regulatory elements when using full-length *Xist*. Thus, promoter and enhancer deacetylation are tightly linked to transcriptional silencing and dependent on the *Xist* A-repeat. This is in line with our HDAC3 experiments indicating a functional role for histone deacetylation in initiating gene repression.

To further assess how PcG accumulation relates to histone deacetylation and gene silencing, we performed nChIP-seq in *Hdac3*$^{-/-}$ cells. We mapped H2AK119Ub and H3K27me3 in two independent *Hdac3*$^{-/-}$ clones at 0 and 24 hr of DOX treatment. Although HDAC3 loss results in inefficient transcriptional silencing, it should not have major direct effects on PcG marks. Indeed, we found that the pattern of both PcG marks at t = 0 hr was very similar along the X chromosome (Figure 7H). After 24 hr of DOX treatment, both marks accumulate globally along the X chromosome in a similar pattern to that observed in *Hdac3*$^{+/+}$ cells (Figure 7H), albeit with slightly lower efficiency. This result differs from the patterns observed in *Xist* A-repeat mutant cells because, in *Hdac3*$^{-/-}$ cells, even gene-rich regions accumulate both PcG-associated marks. We then focused on what happens at genic regions. This revealed that genes that were not silenced in *Hdac3*$^{-/-}$ cells (i.e., sensitive) accumulate neither H2AK119Ub nor H3K27me3 (Figure 7I; Figure S7F). On the other hand, genes efficiently repressed in *Hdac3*$^{-/-}$ (resistant) robustly accumulated both marks. These relationships are exemplified by a *AtrX* (resistant gene) and *Rnf12* (sensitive gene) (Figure S7G). Thus, *Xist* coating and intergenic PcG mark deposition are not sufficient for H2AK119Ub spread into active genes. Only the induction of transcriptional silencing enables H2AK119Ub spread to occur.

In conclusion, we uncoupled the general deposition of the Polycomb marks over the X chromosome from transcriptional silencing. By doing so, we demonstrate that neither PRC1 nor PRC2 dependent marks are able to spread into actively transcribed regions. In the A-repeat mutant, this may be a result of defective *Xist* spreading or the presence of transcription (Engreitz et al., 2013). On the other hand, our experiments in *Hdac3*$^{-/-}$ cells revealed that reduced transcription and histone deacetylation are required to enable PcG spreading into actively transcribed regions.

DISCUSSION

X chromosome inactivation is a powerful model for the study of transcriptional repression and the formation of facultative heterochromatin. The link between chromatin changes and transcriptional silencing during XCI has remained rather elusive. In this study, we provide the detailed choreography of early transcriptional and chromatin changes during the initiation of XCI at high molecular and temporal resolution (Figure S7H). We also explore the chromatin regulatory pathways that help to achieve efficient gene repression. All in all, our findings point to multiple parallel epigenetic mechanisms being at play during the initiation of XCI to ensure rapid and robust transcriptional silencing.

Xist RNA as the Architect for Early Chromatin Changes along the X Chromosome

Our results indicate that the dynamics and efficiency of chromatin alterations along the X chromosome are, in part, shaped by the spreading of *Xist* RNA itself. Indeed, rapid histone deacetylation is most efficient in regions proximal to the *Xist* locus, in line with previous allele-specific RNA-seq experiments (Borensztein et al., 2017; Marks et al., 2015).The PcG-dependent marks accumulate first at regions pre-marked by H3K27me3 and H2AK119Ub, which were previously identified as *Xist* entry sites (XESs), lying in 3D proximity to the *Xist* locus (Engreitz et al., 2013). This striking PcG mark enrichment and 3D proximity are reminiscent of the Polycomb-enriched regions specifically interacting together in *Drosophila melanogaster* (Cheutin and Cavalli, 2014). Thus, it is an exciting possibility that the chromatin landscape of the X chromosome prior to its inactivation instructs the folding of the chromosome, resulting in a specific pattern of *Xist* spreading and subsequent dynamics in gene silencing.

Spreading of PcG Marks, but Not Their Initial Recruitment, Requires Gene Silencing

Our high-resolution XCI analysis also sheds light on the kinetics of PcG mark deposition and spreading along the X chromosome. The very rapid accumulation of PRC1-dependent H2AK119Ub

(D and E) Average plots showing accumulation or depletion of H3K27me3 (D, left), H2AK119Ub (D, right), and H3K27ac (E) over all X-linked active genes (D), promoters (E), or enhancers (E) in *Xist:ΔA*- and *Xist:WT*-expressing cells.
(F and G) Genome browser plots showing H3K27me3 and H2AK119Ub enrichments in regions with a high (F) or low (G) density of active genes. Signals from DOX-treated (red) and untreated (blue) samples were overlaid. Plotted are only expressed genes.
(H) H3K27me3 (left) and H2AK119Ub (right) enrichment across the Xi at 0 hr (top) and their accumulation at 24 hr (bottom) in *Hdac3*$^{+/+}$(TX1072, gray) and *Hdac3*$^{-/-}$ cells. Lines indicate LOESS regression of 10-kb windows spanning the whole X chromosome.
(I) Quantification of H3K27me3 (red) and H2AK119Ub (yellow) accumulation on gene bodies at 24 hr compared to 0 hr in *Hdac3*$^{-/-}$. Genes were separated based on their sensitivity to HDAC3 loss. The p values are from a Wilcoxon rank-sum test.
(J) Correlation between the transcriptional silencing defect and H3K27me3 (red) and H2AK119Ub (yellow) accumulation defect in *Hdac3*$^{-/-}$ compared with *Hdac3*$^{+/+}$(TX1072). The black line represents linear regression fitting. ρ and p values from Pearson's correlation are shown.
See also Figure S7.

that occurs prior to the deposition of H3K27me3 is in contrast with some previous reports suggesting direct PRC2 recruitment by the *Xis*t RNA (Zhao et al., 2008) but is in line with recent reports where a non-canonical PRC1 complex (including PCGF3 and PCGF5) is first recruited, and this then recruits PRC2 (Almeida et al., 2017). Using *Xist* transgenes on autosomes, this was proposed to occur through HNRNPK bridging the B-repeat region of *Xis*t with PRC1 (Pintacuda et al., 2017). This finding appears to be contradicted by the defect in H2AK119Ub deposition in the *Xist* A-repeat mutant. However, we propose that transcriptional silencing may be required to allow for secondary canonical PRC1 recruitment. Indeed, initial H2AK119Ub enrichment is thought to recruit the PRC2 complex via JARID2 or AEBP2, allowing H3K27me3 accumulation, which, in turn, would allow canonical PRC1 binding (da Rocha et al., 2014). Further studies of how transcription affects this feedback and of the timing of PRC1 and PRC2 recruitment are necessary to further validate the revised model.

Here we find that, during the first hours of XCI, PcG marks initially accumulate at intergenic regions and later spread into genes. In cells expressing the *Xist* A-repeat mutant, neither PRC1 nor PRC2-dependent marks are able to spread into genic regions because they are probably impeded by ongoing transcription (Kaneko et al., 2014). Alternatively, the *Xist* A-repeat deletion might lead to local defects in *Xist* spreading, as suggested by Engreitz et al. (2013). Furthermore, our detailed analysis of *Hdac3* mutant cells revealed that H2AK119Ub spreading to genes requires their—at least partial—deacetylation and transcriptional inactivation. This indicates that PRC1 mark spreading is not the trigger for initiating gene repression. On the other hand, because both processes seem to occur around the same time, there might be a potential role of PRC1 in facilitating efficient XCI but not initiating it. Indeed, in an autosomal context, a mutant of *Xist*, which is unable to recruit PRC1, has slightly reduced silencing capacity (Almeida et al., 2017; Pintacuda et al., 2017).

Histone Deacetylation Has an Early Role in Gene Silencing Events during XCI

Here we identify H3K27 deacetylation as one of the earliest chromatin alterations during XCI, tightly linked to transcriptional silencing. This is in line with previous IF/FISH studies in differentiating ESCs and embryos (Chaumeil et al., 2002; Okamoto et al., 2004). Moreover, by using *Hdac3* loss-of-function experiments, we demonstrate that histone deacetylation is not simply due to reduced transcription. Rather, it appears to be an active process that promotes gene silencing. We propose a model whereby histone deacetylation by the pre-bound HDAC3 promotes efficient silencing of most X-linked genes (Figure S7I). A previous study has shown that *Hdac3* knockdown results in defective silencing of a single locus during XCI (McHugh et al., 2015). However, our study reveals that other *Xist* A-repeat-dependent but HDAC3-independent mechanisms are also at play during XCI. This is presumably to safeguard and mediate timely gene repression, especially at genes that need to become rapidly silenced; e.g., *Rnf12*. Although we have excluded the involvement of other histone deacetylases during XCI initiation, future studies will reveal whether these HDAC3-independent mechanisms act via other chromatin-modifying or -remodeling activities or whether they occur through entirely different pathways.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Doxycycline and inhibitor treatment
  - Native ChIP-seq
  - EpiLC induction
  - RNA FISH
  - IF with RNA FISH
  - RNA extraction, reverse transcription, pyrosequencing, qPCR
  - RNA-seq
  - TT-seq
  - CRISPR/Cas9 knockout of *Hdac3*
  - CRISPR/Cas9 FLAG tagging of *Hdac3*
  - HDAC3-FLAG IP-Mass Spectrometry
  - Anti-FLAG ChIP-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Time course analysis on TX1072 cell line
  - Analysis of active histone marks (H3K27ac, H4ac, H3K9ac, H3K4me3, H3K4me1)
  - Analysis of repressive marks (H3K27me3, H2AK119Ub)
  - TT-seq
  - Analysis of WT and Hdac3$^{-/-}$
  - Analysis of FLAG ChIP-seq on Hdac3$^{Flag/Flag}$ ESC cell line
  - Analysis of TXY:Xist and TXY:XistΔA
  - Proteomics data analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and four tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.11.041.

ACKNOWLEDGMENTS

We are grateful to Magdalena Zernicka-Goetz and Elizabeth Blackburn for their continuing support of this project. We are thankful to Samuel Collombet, Tomasz Chelmicki, Pierre Gestraud, Aurelie Teissandier, Nicolas Servant, and Emmanuel Barillot for help with the project and members of the Heard lab for helpful discussions. We are grateful to Anton Wutz for sharing his TXY ESC line. We also thank the tissue-imaging platform—PICT-IBiSA (UMR3215/U934) and the next-generation sequencing platform of Institut Curie. This work was funded by ERC advanced investigator awards (ERC-ADG-2014 671027 to E.H. and ERC-ADG-2015 693023 to P.C.), Labellisation La Ligue, ANR (DoseX 2017: ANR-17-CE12-0029, Labex DEEP: ANR-11- LBX-0044, ABS4NGS: ANR-11-BINF-0001, and part of the IDEX PSL: ANR-10-IDEX-0001-02 PSL to E.H.), a Sir Henry Wellcome Postdoctoral Fellowship (201369/Z/16/Z; to J.J.Z.), Volkswagen Foundation funding (to P.C.), and

"Région Ile-de-France" and Fondation pour la Recherche Médicale grants (to D.L.).

AUTHOR CONTRIBUTIONS

Conceptualization, J.J.Z. and E.H.; Methodology, J.J.Z. and F. Dossin; Software, A.B., B.S., E.M., and L.S.; Validation, J.J.Z.; Formal Analysis, A.B., B.S., E.M., L.S., F. Dingli, and D.L.; Investigation, J.J.Z., K.Z., S.T.R., and F. Dossin.; Data Curation, A.B., B.S., E.M., L.S., F. Dingli, and D.L.; Writing – Original Draft, J.J.Z. and E.H.; Writing – Review & Editing, J.J.Z., E.H., A.B., B.S., E.M., K.Z., S.T.R., and P.C.; Visualization, J.J.Z., A.B., and E.M.; Supervision, E.H. and P.C.; Project Administration, J.J.Z. and E.H.; Funding Acquisition, E.H., J.J.Z., and P.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 2, 2018
Revised: October 20, 2018
Accepted: November 26, 2018
Published: December 27, 2018

REFERENCES

Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. Science *356*, 1081–1084.

Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.B., Vassilev, I., Galupa, R., Servant, N., et al. (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.

Chaumeil, J., Okamoto, I., Guggiari, M., and Heard, E. (2002). Integrated kinetics of X chromosome inactivation in differentiating embryonic stem cells. Cytogenet. Genome Res. *99*, 75–84.

Cheutin, T., and Cavalli, G. (2014). Polycomb silencing: from linear chromatin domains to 3D chromosome folding. Curr. Opin. Genet. Dev. *25*, 30–37.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956–969.

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.

da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Mol. Cell *53*, 301–316.

Disteche, C.M., and Berletch, J.B. (2015). X-chromosome inactivation and escape. J. Genet. *94*, 591–599.

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.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. Curr. Protoc. Bioinformatics *51*, 11.14.1-19.

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

Guenther, M.G., Yu, J., Kao, G.D., Yen, T.J., and Lazar, M.A. (2002). Assembly of the SMRT-histone deacetylase 3 repression complex requires the TCP-1 ring complex. Genes Dev. *16*, 3130–3135.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. *17*, 148.

Jeppesen, P., and Turner, B.M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell *74*, 281–289.

Kaneko, S., Son, J., Bonasio, R., Shen, S.S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev. *28*, 1983–1988.

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.

Krueger, F., and Andrews, S. (2016). SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. F1000Res. *5*, 1479.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

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.

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

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

Marks, H., Kerstens, H.H., Barakat, T.S., Splinter, E., Dirks, R.A., van Mierlo, G., Joshi, O., Wang, S.Y., Babak, T., Albers, C.A., et al. (2015). Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. Genome Biol. *16*, 149.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 1.

McHugh, C.A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and 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.

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.

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.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell *38*, 603–613.

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., et al. (2017).

hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol. Cell *68*, 955–969.e10.

Pinter, S.F., Sadreyev, R.I., Yildirim, E., Jeon, Y., Ohsumi, T.K., Borowsky, M., and Lee, J.T. (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. Genome Res. *22*, 1864–1876.

Poullet, P., Carpentier, S., and Barillot, E. (2007). myProMS, a web server for management and validation of mass spectrometry-based proteomic data. Proteomics *7*, 2553–2556.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Ritz, C., Baty, F., Streibig, J.C., and Gerhard, D. (2015). Dose-Response Analysis Using R. PLoS ONE *10*, e0146021.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Sakata, Y., Nagao, K., Hoki, Y., Sasaki, H., Obuse, C., and Sado, T. (2017). Defects in dosage compensation impact global gene regulation in the mouse trophoblast. Development *144*, 2784–2797.

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.

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

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.

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

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol. Cell *28*, 1–13.

Valot, B., Langella, O., Nano, E., and Zivy, M. (2011). MassChroQ: a versatile tool for mass spectrometry quantification. Proteomics *11*, 3572–3577.

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.

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.

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-LaminB1	Abcam	Cat#ab16048
anti-HDAC3	Santa Cruz	Cat#sc-376957 X
Anti-H3K4me3	Merck-Millipore	Cat#05-1339
anti-H3K4me1	Cell Signaling	Cat#5326S
anti-H3K27ac	Active Motif	Cat#39685
anti-H3K27me3	Cell Signaling	Cat#9733S
anti-H3K9ac	MBL Lifescience	Cat#MABI0305
anti-pentaH4ac	Merck-Millipore	Cat#06-946
anti-H2AK119Ub1	Cell Signaling	Cat#8240S
anti-NCOR1	Abcam	Cat#ab3482
anti-NCOR2	Abcam	Cat#ab5802
anti-FLAG	Sigma-Aldrich	Cat#F1804-50UG
Chemicals, Peptides, and Recombinant Proteins		
RGFP966	Abcam	Cat#ab144819
BRD6688	Bertin Pharma	Cat#19836.1 mg
EZ-Link HPDP-Biotin	ThermoFisher Scientific	Cat#21341
4-Thiouridine	Carbosynth	Cat#NT06186
Tasquinimod	Bertin Pharma	Cat#17692
Tubastatin A	Bertin Pharma	Cat#10559
PCI-34051	Bertin Pharma	Cat#T6325
LMK235	Bertin Pharma	Cat#14969
Trypsin/LysC Mix, Mass Spec Grad	Promega	Cat#v5071
Critical Commercial Assays		
PyroMark Gold Q24	QIAGEN	Cat#970802
Ovation Ultralow System V2 1-16	Nugen Technologies	Cat#0344-32
Ovation Universal RNA-Seq System	Nugen Technologies	Cat#7102-08
Ovation Mouse RNA-Seq System 1-16	Nugen Technologies	Cat#348
Nick Translation Kit	Roche	Cat#10976776001
MODified Histone Peptide Array	Active Motif	Cat#13005
TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat	Illumina	Cat# RS-122-2201
μMACS Streptavidin Kit	Miltenyi Biotec	Cat# 130-074-101
miRNeasy Micro kit	QIAGEN	Cat# 217084
MEGAscript T7 Transcription Kit	ThermoFisher Scientific	Cat# AM1333
Deposited Data		
TT-seq; RNA-seq; ChIP-seq	This study	GSE116480
IP-MS	This study	PXD011344
Experimental Models: Cell Lines		
mouse: ESC TX1072	Heard Lab	(Schulz et al., 2014)
mouse: ESC TX1072-Hdac3−/−#1	Heard Lab	this study
mouse: ESC TX1072-Hdac3−/−#2	Heard Lab	this study
mouse: ESC TX1072-Hdac3Flag/Flag	Heard Lab	this study

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
mouse: ESC TXY:XistWT	Wutz Lab	(Wutz et al., 2002)
mouse: ESC TXY:XistΔA	Wutz Lab	(Wutz et al., 2002)
Oligonucleotides		
Primers	This study	See Table S3
Recombinant DNA		
pSpCas9n(BB)-2A-Puro (PX462)	Feng Zang Lab	Addgene #Cat:48141
p510	Avner Lab	(Chaumeil et al., 2002)
AtrX Fosmid probe	BacPac Consortium at Children's Hospital Oakland Research Institute	Cat#WI1-2039P10
Huwe1 Fosmid probe	BacPac Consortium at Children's Hospital Oakland Research Institute	Cat#RP24-157H12
TT-seq Spike-ins	Cramer Lab	See Table S4
Software and Algorithms		
Bowtie2 (2.2.5)	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
TrimGalore (v0.4.0)	Felix Krueger lab	https://github.com/FelixKrueger/TrimGalore
Cutadapt (1.8.2)	(Martin, 2011)	https://github.com/marcelm/cutadapt
SNPSplit (0.3.2)	(Krueger and Andrews, 2016)	http://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/
Bedtools (2.25.0)	(Quinlan and Hall, 2010)	http://bedtools.readthedocs.io/en/latest
MACS2 (2.0.10)	(Zhang et al., 2008)	https://github.com/taoliu/MACS/
Subread - featureCounts (1.5.1)	(Liao et al., 2014)	http://subread.sourceforge.net/
DeepTools (3.0.2)	(Ramírez et al., 2014)	http://deeptools.ie-freiburg.mpg.de
STAR (2.3.0)	(Dobin and Gingeras, 2015)	https://github.com/alexdobin/STAR
Samtools (1.3)	(Li et al., 2009)	https://github.com/samtools/samtools
Tophat (2.1.0)	(Trapnell et al., 2009)	https://github.com/infphilo/tophat
R (3.4.0)	R Core Team	http://www.R-project.org/
edgeR (R package)	(Robinson et al., 2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Limma (R package)	(Ritchie et al., 2015)	http://bioconductor.org/packages/release/bioc/html/limma.html
drc (R package)	(Ritz et al., 2015)	https://cran.r-project.org/web/packages/drc/index.html
Proteome Discoverer (v 2.2)	ThermoFisher Scientific	N/A
myProMS	(Poullet et al., 2007)	N/A
Bowtie2 (2.2.5)	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
TrimGalore (v0.4.0)	Felix Krueger lab	https://github.com/FelixKrueger/TrimGalore
Other		
Xist oligo probe	Roche	Custom made
S220 Focused-ultrasonicator	Covaris	S220
HiSeq	Illumina	2500
Column LC Acclaim PepMap 100 C18	ThermoFisher Scientific	Cat# 164942 and 164535
Mass Spectrometer	ThermoFisher Scientific	Orbitrap FusionTribrid
Liquid Chromatography System	ThermoFisher Scientific	UltiMate 3000 RSLCnano

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, Edith Heard (edith.heard@curie.fr)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

TX1072 (mouse, female, [*Mus musculus castaneus X* C57BL/6] embryonic stem cells) cells have been previously derived in the lab (Schulz et al., 2014). *Hdac3*$^{-/-}$ and *Hdac3*$^{Flag/Flag}$ clones have been derived from the TX1072 line by CRISPR/Cas9 targeting. Cells were cultured on gelatine-coated plates in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 15% fetal bovine serum, 2-mercaptoethanol (0.1mM), LIF(1000u/mL) and 2i (PD0325901 [0.4 mM], CHIR99021 [3 mM]) . TXY:*Xist* and TXY:*Xist⊿A* (mouse, male embryonic stem cells) were obtained from the Wutz team (Wutz et al., 2002). These lines were cultured on gelatine-coated plates in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 15% fetal bovine serum, 2-mercaptoethanol (50mM) and LIF (1000u/mL). All cells were incubated at 37°C with 8% CO_2.

METHOD DETAILS

Doxycycline and inhibitor treatment

TX1072 based ESC lines were plated at a density of 1.8mln/T75 flask. After 24hrs of culture in DMEM/15%FCS+LIF+2i cell the medium was supplemented with doxycycline (1ug/ml). Cells were then collected at regular intervals. For TXY based ESC lines we have plated them at a density of 1mln/T75. After 24hrs of culture in DMEM/15%FCS+LIF cells were washed twice with PBS and medium was changed to DMEM/10%FCS-LIF with or without doxycycline (1.5ug/ml). Cells were collected after 48hrs of differentiation. For HDAC inhibition cells were pre-treated for 1hr with: 10 μM BRD668; 10 μM RGFP966; 10 μM Tasquinimod; 5 μM Tubastatin A; 5 μM PCI-34051; 30nM LMK235.

Native ChIP-seq

Cells were collected using Accutase (Thermo Fisher Scientific), washed twice in ice-cold PBS and counted. Typically, 3.5mln cells were used per immunoprecipitation (IP). A fraction of cells was always used for RNA/FISH verification of *Xist* induction. Cell pellet was resuspended in 90 μL (per 10 mln cells) of Lysis Buffer (50 mM Tris-HCl, pH7.5; 150mM NaCl; 0.1% sodium deoxycholate; 1% Triton X-100; 5mM $CaCl_2$; Protease Inhibitor Cocktail; 5 mM sodium butyrate). After lysing cells on ice for 10 min we added 62 μL (per 10 mln cells) of Lysis Buffer with MNase (500 μL buffer + 0.5 μL MNase). Chromatin was digested for exactly 10 min at 37°C and reaction was stopped by the addition of 20 mM EGTA. To remove undigested debris the lysates were centrifuged at 13000 rpm for 5 min at 4°C. Supernatant was transferred to a fresh tube, an equal volume of STOP Buffer (50 mM Tris-HCl, pH7.5; 150mM NaCl; 0.1% sodium deoxycholate; 1% Triton X-100; 30 mM EGTA; 30 mM EDTA; Protease Inhibitor Cocktail; 5 mM sodium butyrate) was added, samples were stored on ice.

5 μL of lysate was digested in 45 μL of ProtK Digestion Buffer (20 mM HEPES; 1 mM EDTA; 0.5% SDS) for 30 min at 56°C. 50 μL of AMPure XP beads were added to the digested lysate together with 60 μL of 20% PEG8000 1.25M NaCl. After mixing the samples were incubated for 15 min at RT. Beads were separated on a magnet and washed twice with 80% Ethanol for 30sec. DNA was eluted in 12 μL of Low-EDTA TE and DNA concentration was measured using Qubit DNA High-Sensitivity kit. These measurements were used to normalize lysate concentration between samples. DNA isolated in this step was used as the input sample. The volume of each undigested lysate was adjusted for equal concentration and to obtain 1mL per IP using a 1:1 mix of Lysis Buffer and STOP Buffer.

Anti-mouse Dynabeads (50ul/IP) and Protein-A Dynabeads (10ul/IP) were washed twice in Blocking Buffer (0.5% BSA; 0.5% Tween in PBS). Beads were then resuspended in Blocking buffer and coated with antibodies for 4hrs at 4°C (anti-mouse Dynabeads: H3K9ac[1ug/IP], H3K4me3[2.5ug/IP], H3K27ac[1ug/IP]; Protein-A dynabeads: H3K4me1[0.4ug/IP], H3K27me3 [1ug/IP], H2AK119Ub[0.4ug/IP], H4ac[2ug/IP]). Once coated beads were magnet-separated and resuspended in 1mL of concentration-adjusted lysate. Samples were left rotating overnight at 4°C.

Following day beads were magnet-separated and washed quickly with ice-cold washing buffers. anti-H3K4me3 IP was washed 8 times with Low Salt Buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris-HCl, pH 8.1; 150 mM NaCl; 0.1% sodium deoxycholate). All remaining IPs were washed 4-times with Low Salt Buffer, 2-times with High Salt Buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris-HCl, pH 8.1; 360 mM NaCl; 0.1% sodium deoxycholate) and 2-times with LiCl buffer (0.25 M LiCl; 1% NP40;1.1% sodium deoxycholate; 1 mM EDTA; 10 mM Tris-HCl pH 8.1). Prior to elution all samples were rinsed once in TE. ChIP-DNA was eluted in ProtK-Digestion buffer for 15 min at 56°C. Beads were separated and the supernatant was further digested for another 2 hr at 56°C. DNA was isolated using AMPure XP beads as described for the input sample.

For each nChIP-seq, 0.5 μL of each sample was used for qPCR validation of enrichment at control regions. 0.5 μL of input samples were also used to verify the digestion efficiency using D1000 tapestation. Remaining DNA concentration was adjusted and used for library preparation using Ovation® Ultralow Library System V2 following suppliers protocol. Amplified libraries were size-selected for dinucleotide fraction (350-600bp fragments) using agarose gel-separation and MinElute Gel Extraction Kit (QIAGEN). Sample quality was inspected using D1000 tapestation. Samples were sequenced with HiSeq2500 using PE100 mode for nChIP-seq on TX1072 cell line and SE50 mode for nChIP-seq on TXY cell line.

EpiLC induction

$2.5x10^5$ of ESC were plated onto fibronectin-coated 6-well in N2B27 medium (50%DMEM/F12 and 50% Neurobasal Medium supplemented with: 2-Mercaptoethanol[0.1mM], L-Glutamine[20mM]; B27 supplement[1x], Ndiff Neuro-2 Medium Supplement[1x]) with bFGF (12 ng/ml) and ActivinA (20 ng/ml) and with or without doxycycline (1ug/mL). Medium was changed every day and cells were collected at appropriate intervals.

RNA FISH

ESC were dissociated using Accutase (Invitrogen) and adsorbed onto Poly-L-Lysine (Sigma) coated coverslips #1.5 (1mm) for 5 min. Cells were fixed with 3% paraformaldehyde in PBS for 10 min at room temperature and permeabilised for 5 min on ice in PBS containing 0.5%Triton X-100 and 2mM Vanadyl- ribonucleoside complex (New England Biolabs). Coverslips were preserved in 70% EtOH at −20°C. Prior to FISH, samples were dehydrated through an ethanol series (80%, 95%,100% twice) and air-dried quickly. For detecting *Huwe1* and *Atrx*, a BAC spanning the respective genomic region (RP24-157H12, RP23-160I15) was labeled by nick translation (Roche) using spectrum green (Abbot). Per coverslip, 60ng probe was ethanol precipitated with Cot1 repeats, resuspended in formamide, denatured (10min 75°C) and competed for 45min at 37°C. For *Xist* detection or intron-spanning plasmid probe p510 was labeled by nick translation (Roche) using spectrum red (Abbot). Probe was ethanol precipitated and resuspended in formamide and denatured (10min 75oC). Both *Xist* and AtrX or Huwe1 probes were co- hybridized in FISH hybrization buffer (50% Formamide, 20% Dextran sulfate, 2x SSC, 1 μg/μl BSA, 10mM Vanadyl-ribonucleoside) over-night. Washes were carried out at 42°C three times 5min in 50% formamide in 2X SSC at pH = 7.2 and three times 5min in 2X SSC. 0.2mg/ml DAPI was used for counterstaining and mounting medium consisted of Vectashield (Vectorlabs). Images were acquired using an Inverted Confocal Spinning Disk Roper/Nikon.

IF with RNA FISH

Cells were grown on fibronectin coated coverslips #1.5 (1mm) and treated with doxycycline (1ug/mL) for 24hrs. Cells were fixed with 3% paraformaldehyde in PBS for 10 min at room temperature and permeabilised for 5 min on ice in PBS containing 0.5%Triton X-100 and 2mM Vanadyl- ribonucleoside complex (New England Biolabs). Samples were blocked for 15 min at room temperature with 1% BSA/PBS. Coverslips were incubated with primary antibodies diluted in blocking solution (anti-H4ac: 1/500; anti-H3K27ac:1/100; anti-H2AK119Ub: 1/200) in the presence of a Ribonuclease Inhibitor (0.8u/mL; Euromedex) for 45 min at room temperature. Coverslips were washed thrice with PBS for 5min and subsequently incubated with secondary antibody solution (goat anti-mouse or rabbit conjugated with Alexa fluorophores; 1/500; supplemented with Ribonuclease Inhibitor [0.8u/mL; Euromedex]) for 45 min at room temperature. Coverslips were washed three times with PBS for 5 min at room temperature. Cells were fixed again with 2% paraformaldehyde in PBS for 10 min at room temperature. After rinsing twice in 2xSSC coverslips were hybridized for *Xist* detection. Here we used a custom designed strand-specific probe that covers all exons with ~75 bp long oligo nucleotides end-labeled with the Alexa 488 fluorophore (Roche). Probes was hybridized in FISH hybridization buffer (50% Formamide, 20% Dextran sulfate, 2x SSC, 1 μg/μl BSA, 10mM Vanadyl-ribonucleoside) over-night. Washes were carried out at 42°C three times 5min in 50% formamide in 2X SSC at pH = 7.2 and three times 5min in 2X SSC. 0.2mg/ml DAPI was used for counterstaining and mounting medium consisted of Vectashield (Vectorlabs). Images were acquired using an Inverted Confocal Spinning Disk Roper/Nikon. For signal quantification, signal intensity was measured along an axis across the *Xist* domain (z-plane with *Xist* maximum intensity) and part of the nucleus. Signal was not measured across the nucleoli. Profiles from 70-100 nuclei were centered at the point of *Xist* cloud boundary (maximum inflection point of *Xist* signal profile). Signal was normalized at per cell basis. In Figure 5B and Figure S5B shown are average profiles from all nucleoli with shading representing 25 and 75 quantiles.

RNA extraction, reverse transcription, pyrosequencing, qPCR

RNA was extracted according to the manufacturer's recommendations using RNeasy Mini Kit (QIAGEN) with on-column DNase digestion (QIAGEN). For cDNA synthesis 1.1 μg RNA was reverse transcribed using Superscript III Reverse Transcriptase (Thermo Fisher Scientific). For allelic-skewing analysis the cDNA was PCR-amplified with biotinylated primers and sequenced using the Pyromark Q24 system (QIAGEN). To quantify ChIP DNA enrichments at control regions 2x SybRGreen Master Mix (Applied Biosystems) and a ViiA7 system (Applied biosystems) were used.

RNA-seq

RNA was prepared as described above and quality of samples were verified by Tapestation. Only samples with RIN score above 9 were processed. 100 ng of RNA was used for library preparation using Ovation Mouse RNA-Seq System that uses mouse-specific InDA-C depletion of rRNA. Manufacturers recommendations were followed. Libraries were sequenced using HiSeq2500 at PE100 settings.

TT-seq

The TT-seq protocol was performed as previously described with modifications (Schwalb et al., 2016). Cells were induced with DOX as described above in two T75 flasks per sample. After 0, 4, 8, 12 and 24 hr, RNA was labeled by the addition of 0.5 mM 4-thiouridine (4sU, Carbosynth) for 5 min. Medium was quickly removed and cells were lysed with 10 mL of TRI reagent (Sigma-Aldrich). Lysates were stored at −80°C.

The spike-in mix, containing three 4sU-labeled and three unlabelled RNAs, was synthesized *in vitro* with MEGAscript T7 Transcription Kit (Thermo Fisher), where 1/10 UTP in the transcription reactions was replaced with 4sUTP (Jena BioScience) for the 4sU-labeled spike-ins. RNA lysates were thawed on ice and 24 ng of spike-in mix was added per 10 million cells. The RNA was then purified according to the manufacturer's protocol. 4sU-labeled RNA was isolated from 300 μg total RNA, for this two aliquots of 150 μg were sheared in Covaris MicroTubes with the Covaris S220 System (Covaris) for 10 s with settings 100 W and 1% duty cycle. These were then pooled and 2 μg of sonicated total RNA was taken for total RNA fraction. The RNA was biotinylated with EZ-Link HPDP-Biotin (Thermo Fisher) in two biotinylation reactions per sample (150 μg RNA in 10 mM Tris pH 7.5; 1 mM EDTA, 40% DMSO; and 200 μg/ml HPDP-biotin) that were incubated 1.5 h at 24°C. The biotinylated RNA was extracted with chloroform and precipitated with isopropanol. 4sU-labeled RNA was isolated by incubating the biotinylated RNA with paramagnetic streptavidin μMACS MicroBeads (Miltenyi) for 15 min at room temperature and binding to MACS columns. The columns were washed three times with wash buffer (100 mM Tris pH 7.5; 10 mM EDTA; 100 mM NaCl; and 0.1% Tween 20) at 65°C and three times with wash buffer at room temperature. The RNA was eluted from the columns with 100 mM DTT. The eluted 4sU-labeled RNA and reserved total RNA were purified with the miRNeasy Micro kit (QIAGEN) according to the manufacturer's protocol with on-column DNase I treatment. 150 ng and 500 ng of isolated newly synthesized RNA and total RNA fractions, respectively, were used to prepare NGS libraries with the TruSeq Stranded Total RNA Library Prep Kit (Illumina). The sample libraries were pooled and sequenced as a single run on the Nextseq550 (Illumina) using PE150 mode.

CRISPR/Cas9 knockout of *Hdac3*

Two gRNAs have been designed as previously described (Haeussler et al., 2016) and cloned into pX462 vector allowing for their expression under U6 promoter. Additionally, this vector leads to expression of Cas9. TX1072 ESCs were transfected using AmaxaTM 4D-NucleofectorTM system (Lonza) as per manufacturer's protocol. After a pulse of puromycin selection cells were plated at clonal density. Single clones were picked, expanded and screened for deletion by PCR. Final two knockout clones were further validated by sequencing both alleles. *Hdac3*$^{-/-}$*#1* harbors a deletion of exon 4-7 on the Cast allele and exon 9 on the B6 allele. *Hdac3*$^{-/-}$*#2* harbors a deletion of exon 4-7 on the B6 allele and exon 7 on the B6 allele. Both mutants showed negligible expression on the mRNA level as well as no detectable protein when using a N terminus specific antibody (Figure 4B).

CRISPR/Cas9 FLAG tagging of *Hdac3*

The *Hdac3-Aid-3xFlag* targeting construct was generated as follows: 500bp homology arms (flanking both sides of, but excluding the stop codon of *Hdac3*) were PCR amplified from mouse genomic DNA. 1-step Gibson cloning (NEB) was subsequently used to simultaneously surround the digested AID-3xFLAG insert (carrying a puromycin resistance gene under the control of the PGK promoter) in frame with the homology arms and clone the insert into the pBR322 vector. Synonymous mutations in the PAM/SEED target sequence (located on the 5′ homology arm) were then introduced using the QuickChange II XL site-directed mutagenesis kit (Agilent) to prevent Cas9 mediated cleavage of the targeting vector upon transfection and of the 3xFLAG tagged allele(s) upon integration.

Single gRNAs has been designed as previously described (Haeussler et al., 2016) and cloned into pX462 vector allowing for its expression under U6 promoter.

Hdac3 targeting was performed using the 4-D nucleofector system from Lonza. 5 million cells were electroporated with 2.5 μg each of non-linearized targeting vector and gRNA/Cas9 encoding plasmids. The day after, puromycin was added at a concentration of 0.4ug/mL. Single clones were picked, expanded and screened for insertion by PCR. One clone showing homozygous insertion of the FLAG tagged cassette at the *Hdac3* locus was validated by sanger sequencing and western blotting and kept for subsequent experiments (Figure S5).

HDAC3-FLAG IP-Mass Spectrometry

50 mln (per IP) ESCs (TX1072 –DOX; *Hdac3*$^{Flag/Flag}$ +DOX 24h) were trypsinised and washed in PBS. Pellets were snap frozen and stored at −80°C. Defrosted pellets were resuspended in 10 mL of buffer A (10 mM HEPES pH 7.9; 5 mM $MgCl_2$; 10 mM KCl; 1 mM DTT; 0.1% NP-40, protease inhibitors) and incubated for 10 min at 4°C. Nuclei were pelleted at 2000 rpm at 4°C for 10 minutes. Nuclei were resuspended in 500μL of buffer C (20mM HEPES pH 7.9; 20% (v/v) glycerol; 150 mM KCl; 5 mM MgCl2; 2 mM EDTA; 2 mM EDTA; 1 mM DTT; protease inhibitors) and sonicated 3 times for 10 s (Bioruptor, medium setting). Lysates were cleared by centrifugation 12000 rpm at 4°C for 20 minutes. Lysate concentrations were equalised based on protein concentration. Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) were washed twice in buffer C and then added to each lysate (50μL per IP). Lysates were left rotating over-night at 4°C. Next day beads were washed 5 times in wash buffer (20 mM HEPES pH 7.9; 0.1% NP-40; 150mM KCl; 5mM MgCl2; 2mM EDTA; 2 mM EDTA; 1 mM DTT).

Proteins on magnetic beads were washed twice with 100 μL of 25mM NH4HCO3 and on-beads digestion was performed with 0.2 μg of trypsin/LysC (Promega) for 1 hour in 100μL of 25mM NH4HCO3. Samples were then loaded onto a custom-made C18 StageTips for desalting. Peptides were eluted using 40/60 MeCN/H2O + 0.1% formic acid and vacuum concentrated to dryness.

Online chromatography was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled online to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). Peptides were trapped on a C18 column (75 μm inner diameter × 2 cm; nanoViper Acclaim PepMap 100, Thermo Scientific) with buffer A (2/98 MeCN/H2O in 0.1% formic acid) at a flow rate of 4.0μL/min over 4 min. Separation was performed on a 50cm x 75 μm C18 column (nanoViper Acclaim PepMap RSLC, 2 μm, 100Å, Thermo

Scientific) regulated to a temperature of 55°C with a linear gradient of 5% to 25% buffer B (100% MeCN in 0.1% formic acid) at a flow rate of 300 nL/min over 100 min. Full-scan MS was acquired in the Orbitrap analyzer with a resolution set to 120,000 and ions from each full scan were HCD fragmented and analyzed in the linear ion trap.

Anti-FLAG ChIP-seq

Cells were grown in 145mm plate format. Cells were initially fixed in 1.5 mM ethylene glycolbis [succinimidyl succinate] (EGS) for 30 min, next 2% formaldehyde was added and further incubated for 10 min. After rinsing twice in PBS, cells were scraped and pellets directly used for ChIP.

Fixed pellets were lysed in 6 mL Nuclear Lysis Buffer (0.5% Triton X-100, 0.1 M sucrose, 5 mM MgCl2, 1 mM EDTA, 10 mM Tris-HCl pH 8.0, 1x Protease Inhibitors) for 10min on ice and dounce-homogenized. Nuclei were pelleted and resuspended in 0.9 mL Lysis Buffer (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 8.0, 0.5% N-Lauroylsarcosine, 1x Protease Inhibitors). Chromatin was sheared using a BioruptorPlus (Diagenode) set to high for 30 cycles. Unsonicated chromatin was removed by centrifugation. Supernatant was diluted with 3.6 mL of Dilution Buffer (1.25% Triton, 0.125% sodium deoxycholate, 6 mM EDTA, 10 mM Tris-HCL pH8.0, 1x Protease Inhibitors) and antibody coated beads were added (per IP: 10 μL of anti-FLAG antibody). Samples were left rotating overnight at 4°C.

Following day beads were magnet-separated and washed 5-times with LiCl buffer (0.50 M LiCl; 1% NP40; 1.1% sodium deoxycholate; 1 mM EDTA; 10 mM Tris-HCl pH 8.1) and 5 times with High Salt Buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris-HCl, pH 8.1; 1M NaCl; 0.1% sodium deoxycholate). Each wash was performed for 10 min on a rotating wheel at 4°C. Prior to elution all samples were rinsed once in TE. ChIP-DNA was eluted in ProtK-Digestion Buffer (20mM HEPES; 1mM EDTA; 0.5% SDS; 0.8mg/mL Proteinase K) for 15min at 56°C. Beads were separated and the supernatant was further digested for another 2hrs at 56°C and then decrosslinked for 4hrs at 68°C. DNA was isolated using AMPure XP beads. Libraries were prepared using Ovation Ultralow Library System V2 following suppliers protocol. Samples were sequenced with HiSeq2500 using PE100 mode.

QUANTIFICATION AND STATISTICAL ANALYSIS

Time course analysis on TX1072 cell line

nChIP-seq Data processing

Adapters and low quality bases (< Q20) have been removed with TrimGalore (v0.4.0; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore) and Cutadapt (1.8.2) (Martin, 2011). An 'N-masked'' genome has been generated with SNPSplit (0.3.2) (Krueger and Andrews, 2016) which is a version of the mouse reference genome mm10 where all the polymorphic sites for the hybrid strain *Mus musculus* CAST/EiJ and *Mus musculus* C57BL/6 are masked by ambiguity nucleobase 'N'. For all samples, reads were then mapped to the 'N-masked'' genome with Bowtie2 (2.2.5) with options [--end-to-end -N1 -q] (Langmead and Salzberg, 2012). Duplicates were discarded with Picard MarkDuplicates (1.65) with options [REMOVE_DUPLICATES = true] (https://broadinstitute.github.io/picard/) and reads mapped on blacklisted regions from Encode Consortium were discarded. SNPSplit (0.3.2) (Krueger and Andrews, 2016) was then used to generate allele-specific BAM files by separating the alignment into two distinct alleles (CAST and B6) based on SNPs information downloaded from Sanger. Bigwig files were created with bedtools genomeCoverageBed (2.25.0) (Quinlan and Hall, 2010), using a scale factor calculated on the total library (10.000.000/total reads) for both allele specific bigwigs, and loaded on UCSC genome browser.

Analysis of active histone marks (H3K27ac, H4ac, H3K9ac, H3K4me3, H3K4me1)

Peak identification

Peak calling was done with MACS2 (2.0.10) (Zhang et al., 2008) with options [--broad -B -f BAMPE --broad-cutoff 0.01], on total (non allele-specific) ChIP-seq signal with input as control. Then, peaks with a fold change inferior to 3 were filtered out. For each histone mark, consensus peaks were defined as follow. For each replicate, all peaks coordinates were merged using bedtools merge (2.25.0) (Quinlan and Hall, 2010). Then, common regions between merged peaks coordinates of each replicate were selected using bedtools intersectBed (2.25.0) (Quinlan and Hall, 2010).

Counts and normalization

Total and allelic reads overlapping consensus peaks were counted using featureCounts (1.5.1), with options [-C -p -P] (Liao et al., 2014). Only peaks with more than 50 total allelic reads were selected. Next, ratios of allelic counts, called d-scores [$reads^{B6}$/($reads^{Cast}$+ $reads^{B6}$)], were calculated.

Analysis of the dynamics

The dynamic evolution of d-score with time was analyzed for each peak with a d-score calculated on the 10 samples (5 times, 2 replicates). Moreover, only peaks that had a biallelic d-score at time 0 (comprised between 0.3 and 0.7) were selected. Then normalized d-scores were calculated by dividing each d-score by the corresponding initial d-score (time 0) and by 2, so that initial normalized d-scores start at 0.5. Initially we have fitted either exponential or sigmoidal curve to these values. We compared the quality of this fitting by taking into account the number of parameters used in the models (AIC and BIC criteria) and concluded that the sigmoidal model best reflects the observed dynamics (data not shown). Sigmoidal fitting of normalized d-scores in function of time has been done for each peak with a four-parameter log-logistic function from drc R package (Ritz et al., 2015). Sigmoidal fittings with low residuals (< 0.2; corresponding to more than 98% of the peaks) were selected, and three parameters were obtained from these

fittings: IC35, which is the time where the normalized d-score reaches 0.35 on the sigmoidal fitting; ED_{50}, which corresponds to the maximum slope of the sigmoidal fitting; and the minimum second derivative used as a proxy for when the curve starts to decrease. The minimum of the second derivative has been calculated using optimize function from R in the interval between 0 and the calculated ED_{50} to avoid finding a local minimum not corresponding to the real minimum. IC/ED/minimum second derivative superior to 24h (outside the time range of the experiment) or not calculated were considered windows with a late loss of active histone mark, and replaced by 24h.

Moreover, for all active histone marks, IC25, IC30 and IC40 (thresholds of normalized d-score:0.25; 0.3; 0.4) were also tested and were able to replicate all our major findings (data not shown).

Features comparison between early and late silenced peaks

Early, intermediate and late silenced peaks located on TSS were defined with a 3 clusters k-means based on IC35 values. Early and late silenced peaks were compared for several features using Wilcoxon test: level of expression of associated genes (from TT-seq - see below), distance to *Xist*, distance to the closest TAD boundary (Dixon et al., 2012), distance to the closest LAD (Peric-Hupkes et al., 2010), LINE density in a window of 100kb around the peak (from RepeatMasker database) and gene density in a window of 100kb around the peak. Moreover, IC35 was compared for peaks that are inside or close to *Xist* entry sites (Engreitz et al., 2013)(< 100kb) and those that are distant.

For comparison of dynamics (Figure 6E) of deacetylation and accumulation of repressive marks, early, mid-early, mid-late and late silenced genes were defined with a 4 clusters k-means based on IC35 values.

H3K4me1 accumulation at TSS

Average plots of H3K4me1 signal were created using DeepTools (3.0.2) (Ramírez et al., 2014). Matrix counts were created using DeepTools computeMatrix around active TSS (see above) on chrX and autosomes separately (with option [--binSize 50]), plots were then created using DeepTools PlotProfile.

Analysis of repressive marks (H3K27me3, H2AK119Ub)

Windows definition

For repressive marks, global analysis was first done on fixed windows (10 kb) spanning the whole genome, then on different genomic subcategories: active gene bodies, active promoters, active enhancers and intergenic regions. Active genes were defined as genes with a transcript having its TSS (refFlat annotation) overlapping a consensus peak of H3K9ac and a consensus peak of H3K4me3. For genes having several active transcripts detected, the active gene was defined as starting at the minimum start of transcripts, and ending at the maximum end of transcripts. Then, the active gene bodies were defined as those active genes excluding the 2 first kb downstream of TSS. Active promoters were defined as ± 2kb windows around the TSS of active genes. Putative, active enhancers were defined as intersect between consensus H3K27ac peaks and consensus H3K4me1 peaks, located at a minimal distance of 1kb from a TSS. Intergenic regions were defined as 10 kb windows not overlapping a gene (active or inactive) and its promoter (2kb downstream) or an active enhancer. Enhancer coordinates were also used for the analysis of active histone mark (H3K4me1 and H3K27ac).

Counts and normalization

It should be noted that a d-score analysis, as used for active mark loss, involves the use of the active allele as an internal control, but in the case of H2AK119Ub and H3K27me3, this was hindered by very low enrichment for repressive marks across the Xa. For all defined windows, total and allelic reads overlapping those features were then counted using featureCounts (1.5.1), with options [-C -p -P] (Liao et al., 2014). Then, analysis was done based on normalized reads from B6 allele (allele of the inactive X chromosome). For each sample, a normalization factor was calculated with the trimmed mean of M-values method (TMM) from edgeR package (Robinson et al., 2010), based on B6 reads overlapping consensus peaks located on autosomes (identified and defined as consensus such as for active histone marks analysis). To correct for chromatin accessibility or mappability bias, 10kb windows with outliers counts in the input (counts superior or inferior to mean +-1.5 sd) were discarded from the analysis. Moreover, to represent B6 read accumulation compared to time 0, subtraction of normalized initial counts (time 0) was then applied to all other time points.

Analysis of the dynamics

Sigmoidal fitting of B6 read accumulation in function of time has also been done with the four-parameter log-logistic function from drc R package. Sigmoidal fittings with low residuals (< mean(residuals) + 1.5 sd(residuals)) were selected, representing more than 90% of the analyzed windows. The ED_{50} and the maximum second derivative of the sigmoidal fitting were calculated for each window. Windows with ED_{50}/maximum 2nd derivative superior to 24h, or not calculated, were considered as windows with a late accumulation of repressive marks, those were then replaced by 24h. For pairwise comparison between marks, only windows with ED_{50}/maximum 2nd derivative inferior to 24h for both marks were selected. For comparison of ED_{50} between different genomic subcategories (active promoters, gene bodies, enhancers, intergenic regions) for a same histone mark, no filter was done.

To test the stability of our results, main analysis was also performed with different thresholds of residuals (0.2; 0.15; 0.1; 0.05). Moreover, repressive marks dynamics were also analyzed using allelic ratio (d-scores), such as for analysis of active histone marks, using ED_{50}, IC60 and IC65 as criteria. All those analysis revealed stable and reproducible results (data not shown).

TT-seq

Adapters ("TruSeq Universal Adapter") and low quality bases (< Q20) have been removed with Cutadapt (v1.16) (Martin, 2011). For all samples, reads were then mapped to the 'N-masked' genome (masked for SNPs of the hybrid strain *Mus musculus* CAST/EiJ and

Mus musculus C57BL/6) with STAR 2.3.0 (Dobin and Gingeras, 2015) with maximum 2 percent mismatches. Only unique alignments were retained and soft clipping was disabled [mapping parameters: '--outFilterMismatchNoverReadLmax 0.02 --outFilterMultimap ScoreRange 0 --alignEndsType EndToEnd']. Samtools (Li et al., 2009) was used to quality filter SAM files, where alignments with MAPQ smaller than 7 (-q 7) were skipped and only proper pairs (-f 2) were selected. SNPSplit (v0.3.2) (Krueger and Andrews, 2016) was used to generate allele-specific SAM files. Then, total and allelic reads per gene were counted with featureCounts (1.5.1) [options: -C -p -s 2 -T 8] (Liao et al., 2014) on longest gene coordinates (refFlat annotation). The next steps of analysis (from d-scores to sigmoidal fitting) were done such as for nChIP-seq analysis, except that d-scores were calculated for genes with minimum 10 allelic reads instead of 50. Total normalized expression used for features comparison of early and late silenced peaks (see upper) was calculated based on signal in introns and exons, normalized by the size of the library and the size of the entire gene.

Analysis of WT and Hdac3$^{-/-}$

RNA-seq

Reads were first mapped on rRNA with Tophat (2.1.0) (Trapnell et al., 2009), with options [-g 1 --no-coverage-search --library-type fr-secondstrand]. Paired unmapped reads were then used to reconstruct fastq files with bedTools bamToFastq (2.25.0) (Quinlan and Hall, 2010). Those files were then mapped with Tophat (2.1.0) (Trapnell et al., 2009), with options [-p 8 -g 1 -x 1 -N 3 --read-edit-dist 3 --no-coverage-search --library-type fr-secondstrand], with refFlat annotation. Reads covering exons of each gene were then counted with featureCounts (1.5.1) with options [-C -p -s 1 -T 8] (Liao et al., 2014). D-scores were then calculated for genes with minimum 10 allelic reads.

For differential gene expression analysis prior to DOX treatment, analysis was done on genes for which 2 among the 4 samples have a TPM superior to 1. Normalization was done using TMM (Robinson et al., 2010) method from edgeR and differential analysis using Voom function from Limma R package (Ritchie et al., 2015).

nChIP-seq

The processing from raw data (fastq) to clean mapped file was done such as for time course nChIP-seq. Then, for active histone marks, the reads counts of H3K27ac and H4ac nChIP-seq were done on consensus H3K27ac peaks and consensus H4ac peaks defined in time course analysis, using featureCounts (1.5.1), with options [-C -p -P] (Liao et al., 2014). D-score was calculated with a slight modification for one of the 2 mutant sample. Indeed, this sample showed a non biallelic signal on X chromosome in its input, due to the presence of some XO cells. To take this bias in account, a normalization factor was calculated as follow: normFactor = (reads^{B6}/(reads^{B6} + reads^{CAST})) / (1-(reads^{B6}/(reads^{B6} + reads^{CAST}))).

Then d-score for this mutant was adjusted from classic d-score as follow: d-score = d-score/(d-score + normFactor - normFactor * d-score). After this adjustment, only the peaks that were biallelic at time 0 in WT and mutant samples were selected (d-scores between 0.3 and 0.7) for analysis. For repressive marks, analysis was done such as for time course nChIP-seq, based on signal accumulation of 10 kb windows all along the X chromosome.

For differential peak signal analysis at t = 0hr, normalization was done using TMM (Robinson et al., 2010) method from edgeR and differential analysis using Voom function from Limma R package (Ritchie et al., 2015).

Comparison of sensitive and resistant peaks (nChIP-seq on active histone marks) / genes (RNA-seq, ChIP-seq on repressive histone marks)

Sensitive and resistant peaks/genes were defined as follow. For nChIP-seq on active histone marks and RNA-seq, only peaks/genes with d-scores considered as stable between replicates (diff replicate < median + 2sd of all differences between replicates) for both clones (WT, mutant) at each time point were selected. D-scores were normalized to time 0 (d-score / d-score t0), then only peaks/genes considered as not escaping were selected (normalized d-score at t24 < 0.8). Difference of normalized d-score at 24h between WT and mutant was then calculated, and divided by the difference of normalized WT d-score between 0 and 24h for the WT sample. K-means with 3 clusters was done on this normalized difference, identifying 3 groups: sensitive, intermediate and resistant peaks/genes. For nChIP-seq on repressive histone marks, differential enrichment between WT and mutant was defined as the difference between WT normalized BL6 enrichment relative to time 0 and Mutant normalized BL6 enrichment relative to time 0, divided by WT normalized BL6 enrichment relative to time 0.

Analysis of FLAG ChIP-seq on Hdac3$^{Flag/Flag}$ ESC cell line

The processing from raw data (fastq) to d-score calculation was done such as for time course nChIP-seq, with slight modifications. The peak calling was done with MACS2 (2.0.10) (Zhang et al., 2008) with options [-B -f BAMPE], on total (non allele-specific) ChIP-seq signal against the FLAG ChIP-seq on TX1072 cell line as control. Consensus peaks were defined as peak regions present at least in 2 of the 4 samples, using bedtools multiIntersectBed (2.25.0) (Quinlan and Hall, 2010). Only peaks with more than 20 total allelic reads were selected.

Heatmaps and average plots of HDAC3 signal in comparison with H3K27ac and H4ac signal were created using DeepTools (3.0.2) (Ramírez et al., 2014). Merged peaks of H3K27ac and H4ac at time 0 were split in 2 categories: active promoters for merged peaks overlapping a TSS and putative active enhancer for merged peaks at a distance of minimum 2 kb of a TSS. The merged peaks were ordered by normalized signal of HDAC3 peaks (signal/length of the peak) overlapping those peaks. For HDAC3, signal of the control (FLAG ChIP-seq on TX1072 cell line) was subtracted from the signal of FLAG ChIP-seq on *Hdac3$^{Flag/Flag}$* ESC cell line using DeepTools bamCompare with options [--ratio subtract --binSize 100] before to proceed to the next step. Then, matrix counts were

created using DeepTools computeMatrix around TSS of active promoters (with option [--binSize 100]) or centers of putative enhancers, plots were then created using DeepTools PlotHeatmap.

For H3K27ac and HDAC3 peaks annotation comparison (Figure S5C), peaks were annotated as follow: peaks at a distance < 2000 bp of a TSS were considered in promoters, other peaks were considered as intragenic if overlapping a gene body, intergenic if not.

Category of peaks accumulating signal or not on Xi (Figures S5E and S5F) were created based on clustering with hclust function with parameter [method = 'Ward.D'], using Pearson correlation as distance. 11 clusters were identified using cutree function. Those clusters were grouped in 2 super groups according to the presence or not of an accumulation of signal on Xi at 4hr in each cluster.

Analysis of TXY:Xist and TXY:XistΔA

nChIP-seq

Processing steps to clean bam files were similar to time course nChIP-seq analysis, except that mapping was not allele-specific but done on mm10 genome and that reads mapped with low quality (< q10) were removed with samtools (1.3)(Li et al., 2009). Such as for time course nChIP-seq analysis, ChIP-seq signal was analyzed differently depending on the histone mark: per peak for active mark (H3K27ac) and per window for repressive marks (H3K27me3, H2AK119Ub). Peak calling was done similarly to time course nChIP-seq analysis of active histone marks. Here, consensus peaks were defined as common regions between peaks identified in minimum 2 among the 4 samples using bedtools multiIntersectBed (2.25.0) and bedtools merge (2.25.0).

The windows for repressive mark analysis were defined similarly as for nChIP-seq time course analysis for repressive marks, except that active TSS were defined here based on consensus peaks of H3K27ac overlapping TSS, and putative active enhancers as H3K27ac peaks located at minimum 1kb of TSS.

Reads overlapping defined peaks (for active marks) or windows (for repressive marks) were then counted with featureCounts (1.5.1) (Liao et al., 2014) with default options.

For global analysis on repressive marks, counts normalization was done such as for time course nChIP-seq analysis, based on counts falling in autosomal consensus peaks. For windows and peaks analysis, windows that had less than one read per 50kb for more than 2 among the 8 samples were removed for the analysis. Normalization factors were calculated based on windows located on autosomes (with selection of the 10000 most higher signal for intergenic window analysis), with TMM method using edgeR (Robinson et al., 2010).

Because of the high variability in proportion of cells involved in X chromosome inactivation quantified by the presence of *Xist* Cloud by FISH experiments (TXY:*XistWT*#1 46.64%, TXY:*XistWT*#2 59.44%, TXY:*XistΔA*#1 50.61%, TXY:*XistΔA*#2 48.24%), linear regression including the percentage of induction calculated by FISH was fitted for each window according the following model: ~0 + clones + clones: induction, using Voom function from Limma R package (Ritchie et al., 2015). The slope of this regression represents then the logFC between noDox and Dox conditions if the induction of the cell population was complete (corrected logFC).

Metaplots were created using DeepTools (3.0.2) (Ramírez et al., 2014). Bigwigs of log2(FC) between Dox and noDox samples were created with personalized scaling according to normalized factors calculated above using DeepTools bamCompare, bigwig of mean of log2(FC) between replicates was then created using DeepTools bigwigCompare (binSize: 100 bp for H3K27ac, 1000 bp for H3K27me3 and H2Ub), matrix counts was then created using DeepTools computeMatrix on active genes coordinates (see above) on chrX and autosomes separately and plots were then created using DeepTools PlotProfile.

Proteomics data analysis

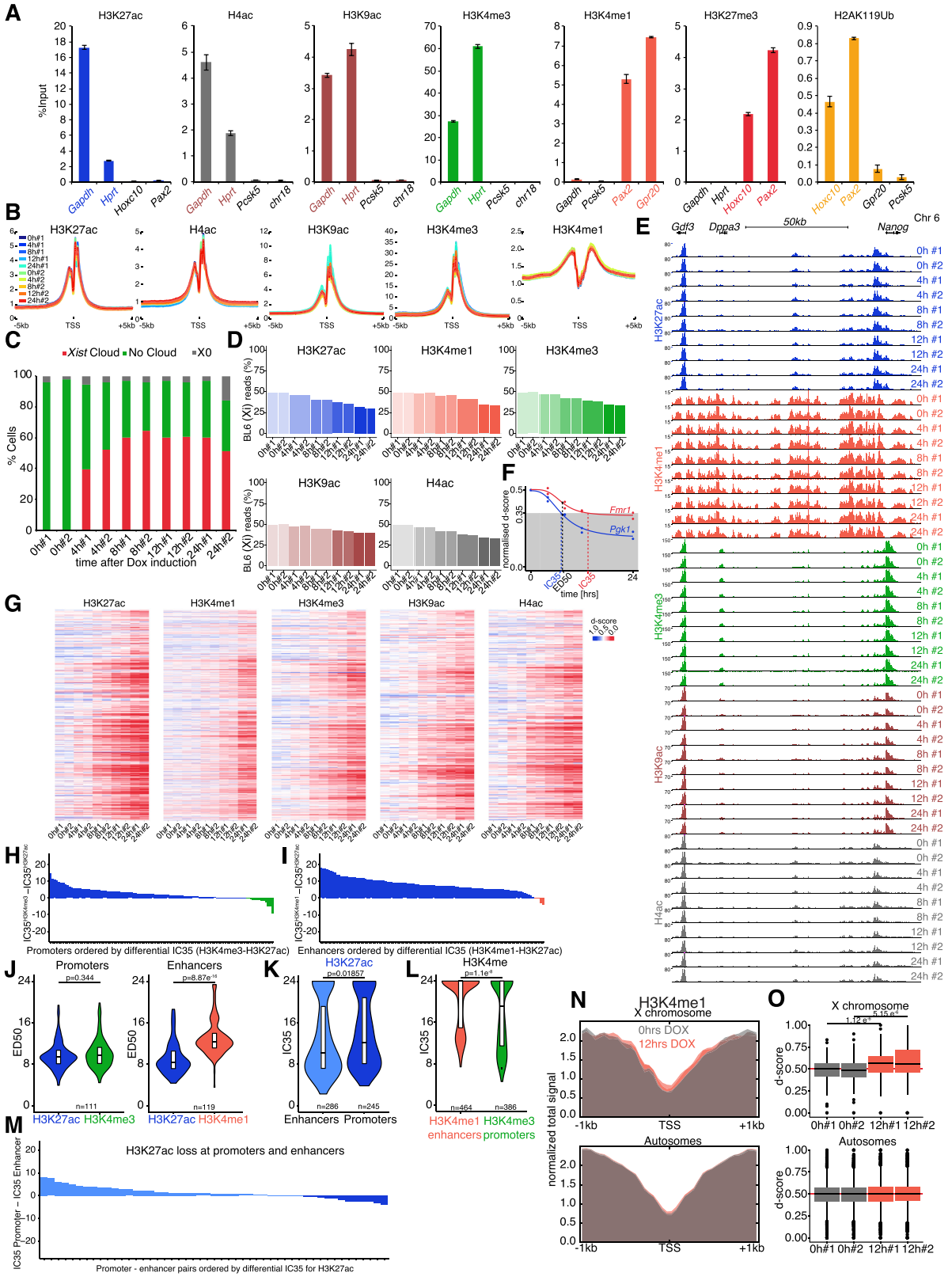
For identification, the data were searched against the *Mus musculus (UP000000589) UniProt* database using Sequest HF through proteome discoverer (version 2.2). Enzyme specificity was set to trypsin and a maximum of two-missed cleavage sites were allowed. Oxidized methionine, N-terminal acetylation, and carbamidomethyl cysteine were set as variable modifications. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.6 Da for MS/MS peaks.

The resulting files were further processed using myProMS v3.6 (Poullet et al., 2007). FDR calculation used Percolator and was set to 1% at the peptide level for the whole study. The label free quantification was performed by peptide Extracted Ion Chromatograms (XICs) computed with MassChroQ version 2.2.2 (Valot et al., 2011). For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN matching) with no missed cleavages were used. Median and scale normalization was applied on the total signal to correct the XICs for each biological replicate. To estimate the significance of the change in protein abundance, a linear model (adjusted on peptides and biological replicates) was performed and *p-value*s were adjusted with a Benjamini–Hochberg FDR procedure with a control threshold set to 0.05.

DATA AND SOFTWARE AVAILABILITY

The accession number for the sequencing datasets reported in this paper is GEO: GSE116480. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011344

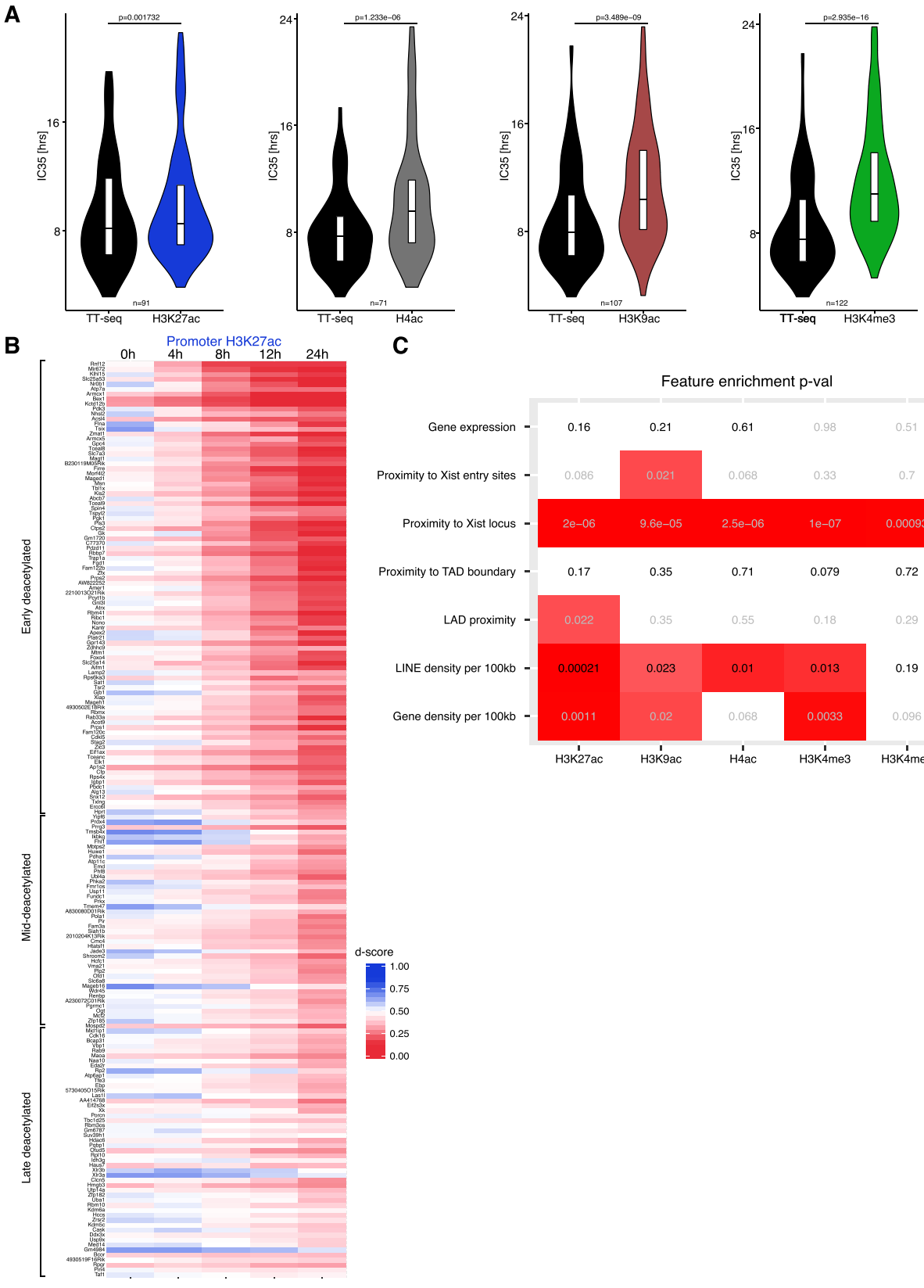
Supplemental Figures



(legend on next page)

Figure S1. Allele-Specific Native ChIP-Seq Monitors Changes in Active Histone Modifications during XCI, Related to Figure 1

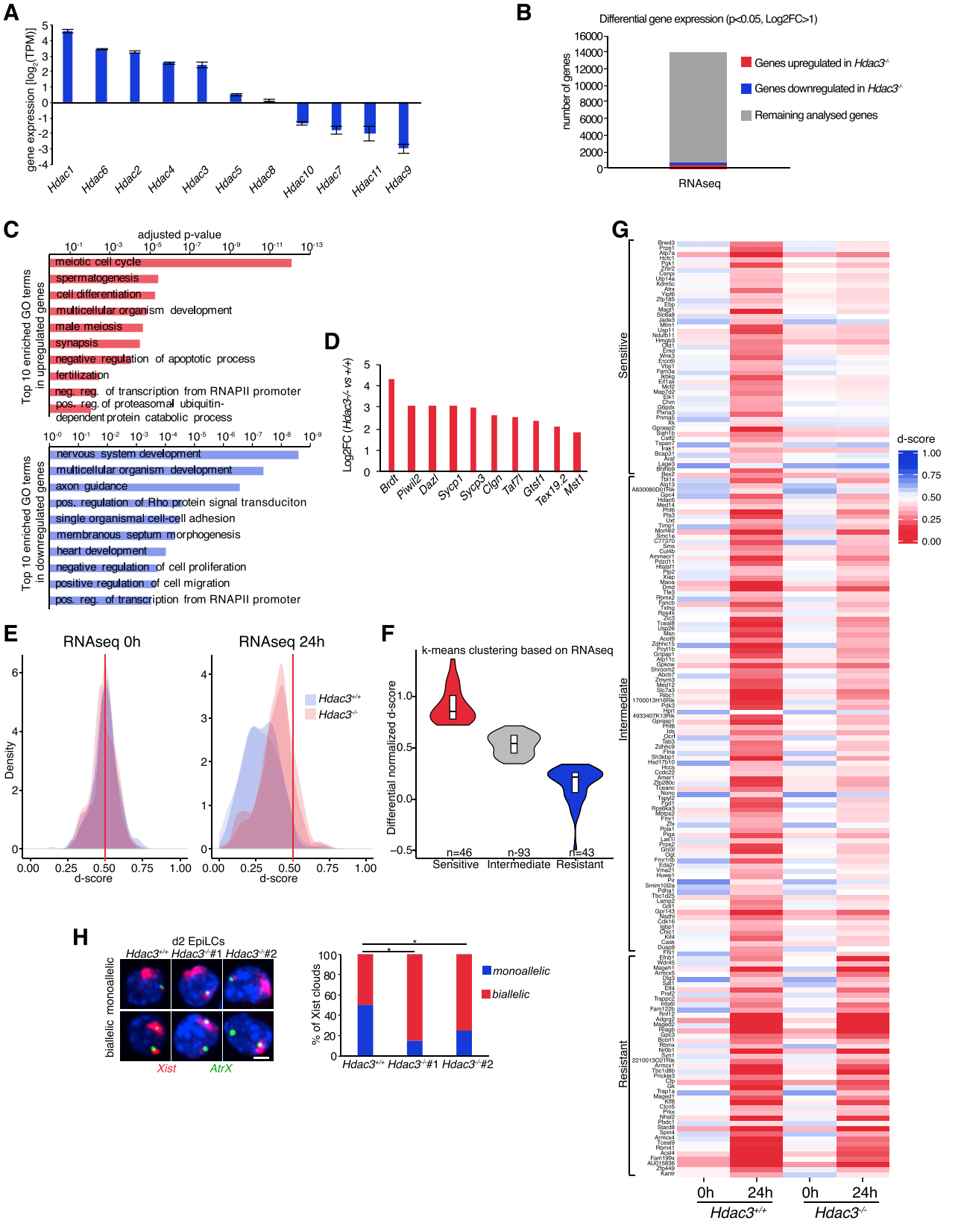
(A) nChIP enrichment validation using qPCR. Shown is the enrichment of nChIP signal at positive (color) and negative (black) control regions for each mark individually. Shown is a representative result for one biological replicate. Values were normalized to the input sample. (B) Average enrichment plots for active histone modifications across all transcriptional start sites. (C) RNA FISH quantification of *Xist* induction (p510 probe) during both replicates of nChIP-seq time course. At least 70 nuclei were quantified for each sample. (D) Bar plots showing the percentage of B6(Xi) reads mapping to the X chromosome in the nChIP-seq time course. (E) Genome browser plots of active histone modification at the *Gda3*/*Dppa3*/*Nanog* cluster. Shown are all mapped reads at all time points in biological duplicates. Note highly reproducible and stable pattern of enrichment. (F) Example showcasing the differences between IC35 and ED_{50} parameters. The IC35 measures when the loss of a histone mark reaches an efficiency threshold and not the timing when this process occurs most rapidly. For example, if two marks show different efficiency of loss from the Xi (i.e., different IC35) the process might still be occurring at the same time. Shown is H3K27 deacetylation dynamics for two promoters (*Fmr1*:red, *Pgk1*:blue). Both promoters have the same ED_{50} (i.e., time when sigmoid reaches its maximum slope), however show very different efficiency of deacetylation. The latter difference is captured by the IC35 parameter (i.e., time when sigmoid reaches the 0.35 threshold). (G) Heatmaps showing the d-scores of all peaks on the X chromosome. Windows were sorted from centromere (top) to telomere (bottom). (H-I) Plots showing the distribution of differential IC35 (H): H3K4me3-H3K27ac; (I): H3K4me1-H3K27ac) for all promoters (H) and putative enhancers (I). Each bar represents a single window, these were ordered accordingly to their differential IC35. (J) Violin plots comparing the dynamics of H3K27ac and H3K4 methylation loss at active promoters (left) and putative active enhancers (right). Plots compare the ED_{50} parameter of H3K27ac and H3K4me peaks within the quantitative range of the experiment (IC35 < 24hrs). p value was calculated using paired Wilcoxon rank sum test.(K-L) Violin plots comparing the distribution of IC35 for H3K27ac (K) or H3K4me1/3 (L) at enhancers and promoters. p values calculated using Wilcoxon rank sum test. (M) Plots showing the distribution of differential IC35 for H3K27ac at promoter-putative enhancer pairs. Windows within one TAD were paired based on proximity. Individual pairs were ordered accordingly to their differential IC35. (N) Average H3K4me1 distribution plots around the transcriptional start site of active genes (TSS ± 1kb). Plots were done using all mapped reads on TSS of active genes on X chromosome (top) and on autosomes (bottom). Average profiles show duplicates at 0hr (gray) and 12hrs (orange) of DOX treatment. (O) Boxplots showing the d-score distribution for H3K4me1 around TSS of active genes (±200bp) after 0 (gray) and 12hrs (orange) of DOX treatment, on X chromosome (top) and on autosomes (bottom). p values calculated using Wilcoxon rank sum test.



(legend on next page)

Figure S2. Histone Deacetylation Correlates with Transcriptional Silencing and Some Genomic Features, Related to Figure 2

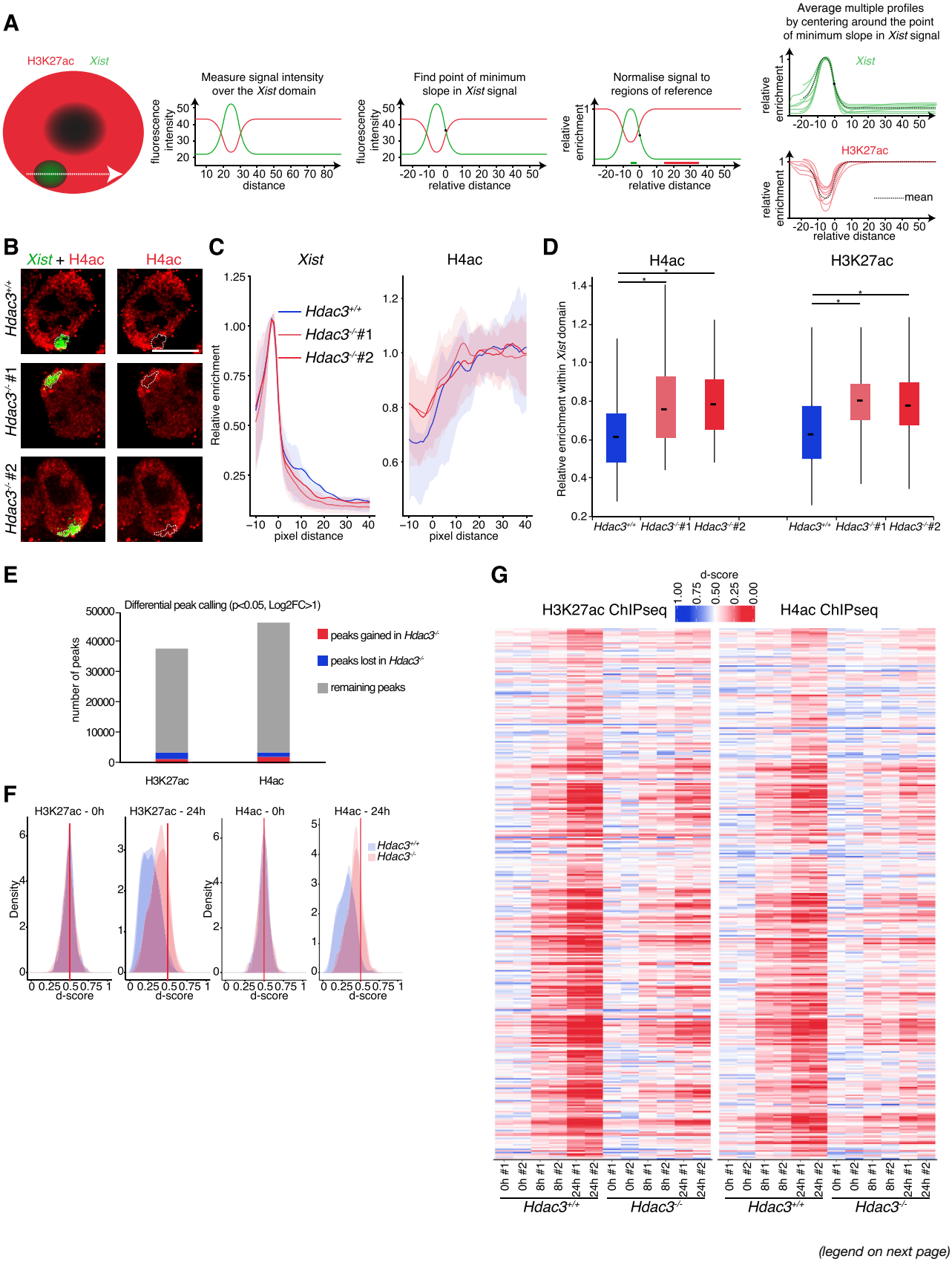
(A) Violin plots comparing the dynamics of transcriptional silencing and the loss of active histone modifications from associated promoters. Plots compare the IC35 parameter calculated from TT-seq nascent transcription profiling to the IC35 of different active histone marks (H3K27ac, H4ac, H3K9ac, H3K4me3). p values were calculated using paired Wilcoxon rank sum test. (B) Heatmap of H3K27ac allelic dynamics at all promoters clustered using k-means. (C) An array showing features enriched (gray font) or depleted (black font) at promoters efficiently inactivated (low IC35) compared to late inactivated (higher IC35) using a panel of histone modifications. Shown are p values calculated using Wilcoxon rank sum test with Benjamini Hochberg correction. Red cells are reaching statistical significance of $p < 0.05$.



(legend on next page)

Figure S3. HDAC3 Mediates Efficient Trascriptional Silencing during XCI; RNA-Seq Analysis and Validation Related to Figure 3

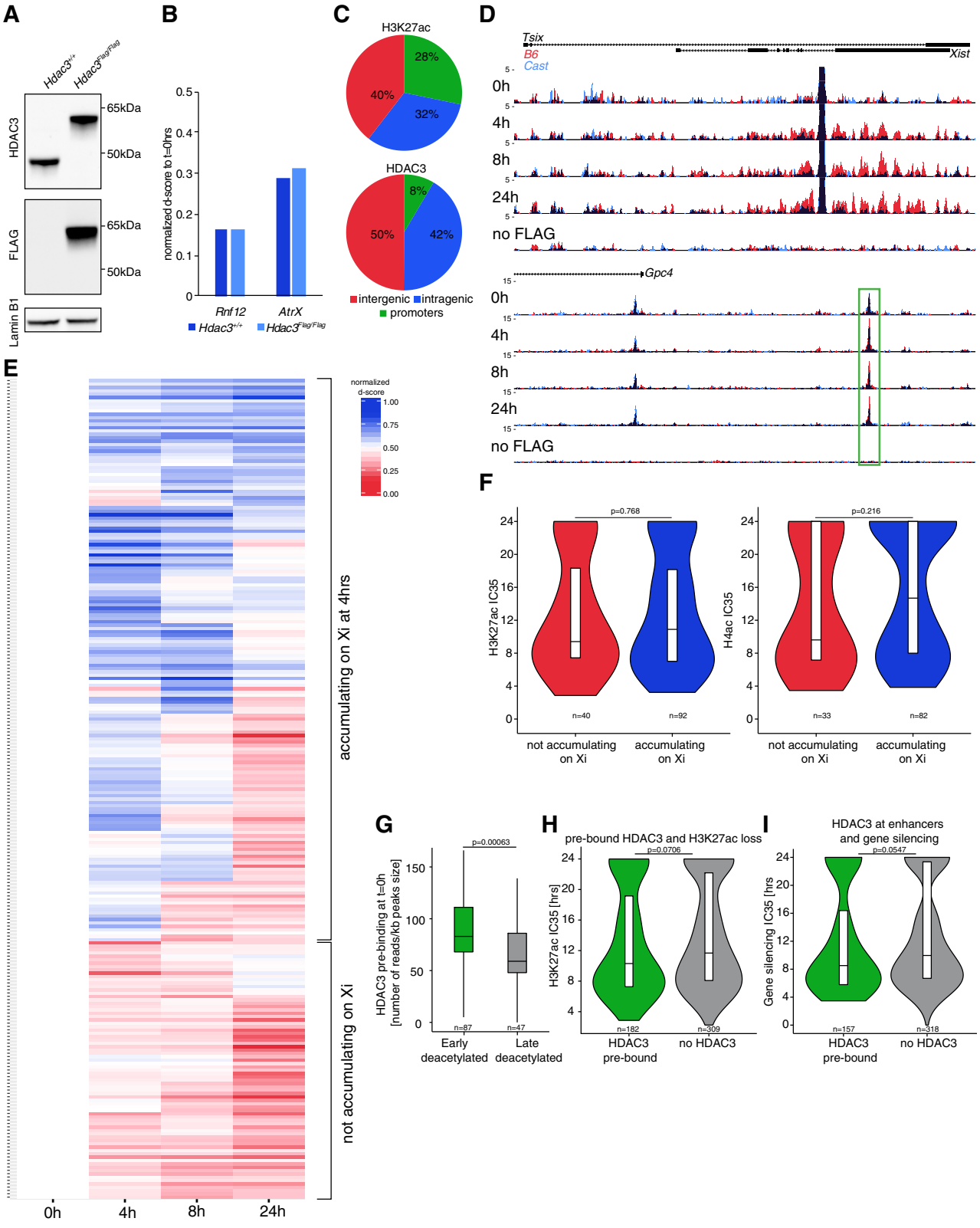
(A) Expression level of all *Hdac* enzymes in TX1072 ESC. Shown is average from two biological replicates (+/− StDev). (B) Plot showing the number of differentially expressed genes in *Hdac3*$^{-/-}$ cells. EdgeR p value < 0.05 and Log2FC > 1 (upregulated) or Log2FC < −1 (downregulated). Related to Table S1. (C) Top ten enriched Gene Ontology terms in upregulated (top) and downregulated genes (bottom) in *Hdac3*$^{-/-}$ cells. (D) Examples of upregulated meiosis genes in *Hdac3*$^{-/-}$ cells. Shown is average Log2FC from two biological replicates. Typically, these genes are stably repressed by DNA methylation, however here in a hypo-methylated context of ESCs grown in 2i/LIF conditions, HDAC3 is responsible for suppressing their transcription. (E) Density plot representing allelic skewing of all X-linked genes after 0 and 24hrs of DOX treatment in *Hdac3*$^{+/+}$ and *Hdac3*$^{-/-}$ lines. Data was extracted from RNA-seq results and presented is the distribution of d-scores, shown are duplicates of the control ESC line and two independent *Hdac3* knockout clones. Red line shows d-score = 0.5 equating to biallelic expression. (F) Violin plot showing the distribution of normalized differential d-scores calculated from RNA-seq experiments in *Hdac3*$^{+/+}$ and *Hdac3*$^{-/-}$ ESCs. Genes were clustered using k-means. The gene clusters are presented in a heatmap of average d-scores (G). (H) Validation of *Hdac3*$^{-/-}$ phenotype in day 2 EpiLCs induced without DOX. FISH was performed with probes for *Xist* (p510-Red) and an X-linked gene *AtrX* (Green). At least 100 nuclei with *Xist* clouds were imaged to assess the *AtrX* silencing status. *Chi2 p value < 0.05. Scale bar = 5 μm.



(legend on next page)

Figure S4. HDAC3 Promotes Silencing by Rapidly Deacetylating Histones; IF/FISH and nChIP-Seq Analysis, Related to Figure 4

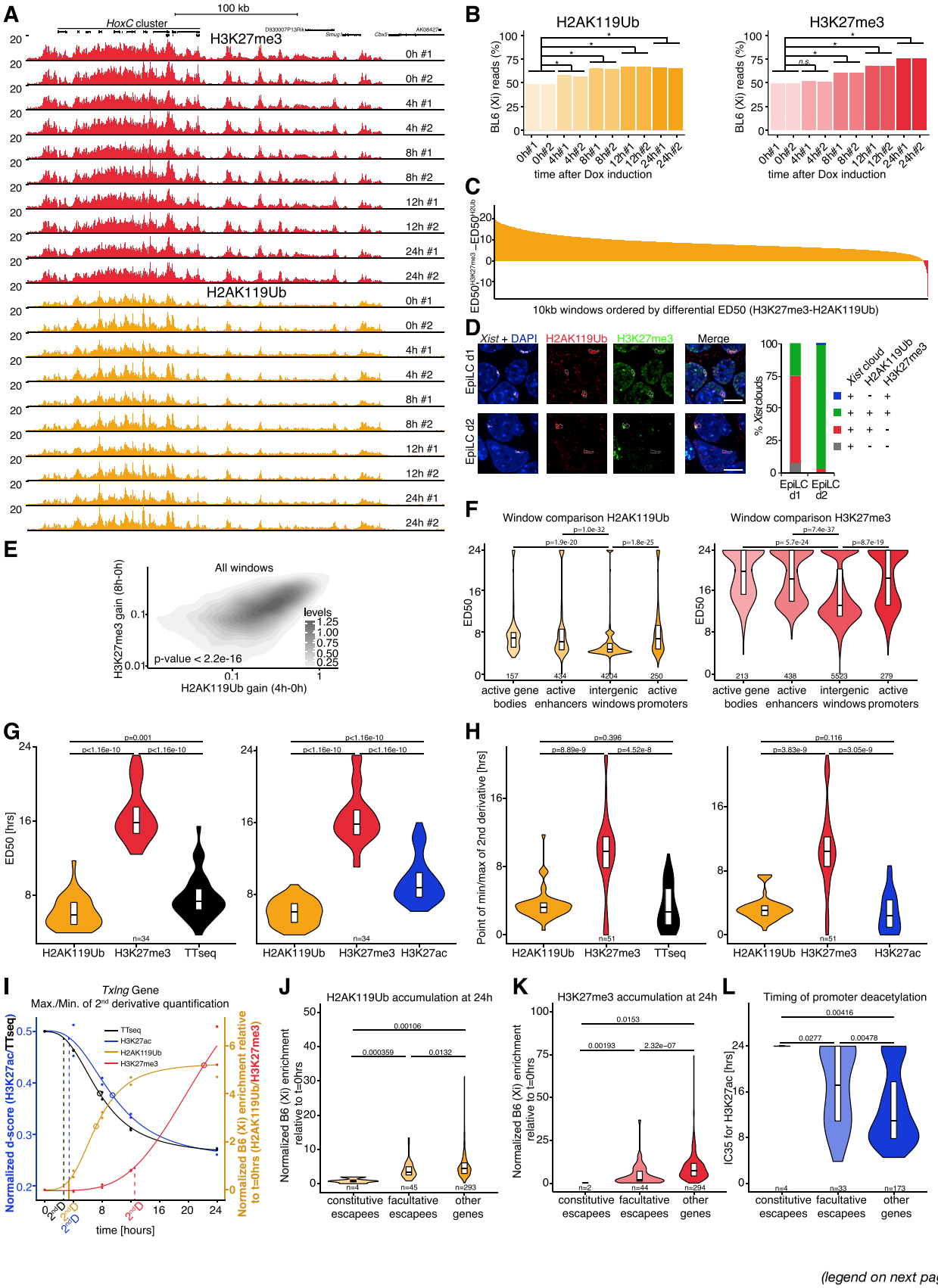
(A) Schematics of IF/FISH quantification. For each nucleus with a *Xist* cloud a z-slice was selected where *Xist* signal reaches its maximum. A line was drawn across the *Xist* domain and into the nucleus avoiding the nucleolus region. Signal intensity for *Xist* (green) and IF (e.g., H3K27ac in red) was recorded along this axis. Next, *Xist* boundary was defined as a point of minimum slope of *Xist* signal. The X coordinates were centered around this boundary point. The fluorescence signal for each cell was normalized to reference regions of *Xist* peak (−4 to −2 pixels from the boundary) or IF signal plateau (15 to 35 pixels from the boundary), these regions are represented as color bars. Finally, profiles from multiple nuclei were averaged for each channel separately. (B) IF/RNA FISH on $Hdac3^{+/+}$ and two $Hdac3^{-/-}$ ESC clones induced with DOX for 24hrs. Cells were probed using anti-H4ac (red) antibodies and a *Xist* intronic probe (green). White dotted line encircles the *Xist* domain. Due to technical variability contrast settings were individually adjusted to reflect relative enrichment. Scale bar = 10 μm. (C) Quantification of IF/RNA FISH signal over at least 70 *Xist* clouds. Shown is the average signal for *Xist* and H4ac when profiles were aligned to the *Xist* cloud boundary. Signal was normalized per cell. For detailed explanation of IF/FISH quantification see point A.(D) RNA/FISH quantification based on point (C) and Figure 4A. Boxplots quantifying the relative enrichment of H4ac and H3K27ac within the *Xist* cloud (−10 to 0 pixel distance) in $Hdac3^{+/+}$ and $Hdac3^{-/-}$. * p value < 0.05 from Wilcoxon rank sum test. (E) Plot showing the number of differentially called peaks in $Hdac3^{-/-}$. P value < 0.05, Log2FC > 1 or Log2FC < −1. (F) Density plots showing the distribution of d-scores at H3K27ac (left) and H4ac (right) peaks on the X chromosome in $Hdac3^{+/+}$ and $Hdac3^{-/-}$. Initially, peaks are on average biallelic (0h) and become skewed toward the Cast allele after 24h of DOX treatment. Shown is an overlay of biological duplicates including two independent $Hdac3^{-/-}$ clones. (G) Heatmaps showing the d-score evolution of all H3K27ac and H4ac peaks in $Hdac3^{+/+}$ and $Hdac3^{-/-}$. X-linked peaks were sorted based on their genomic position from the centromere to telomere.



(legend on next page)

Figure S5. HDAC3 Is Pre-bound on the X Chromosome and Only Weakly Accumulates during XCI, Related to Figure 5

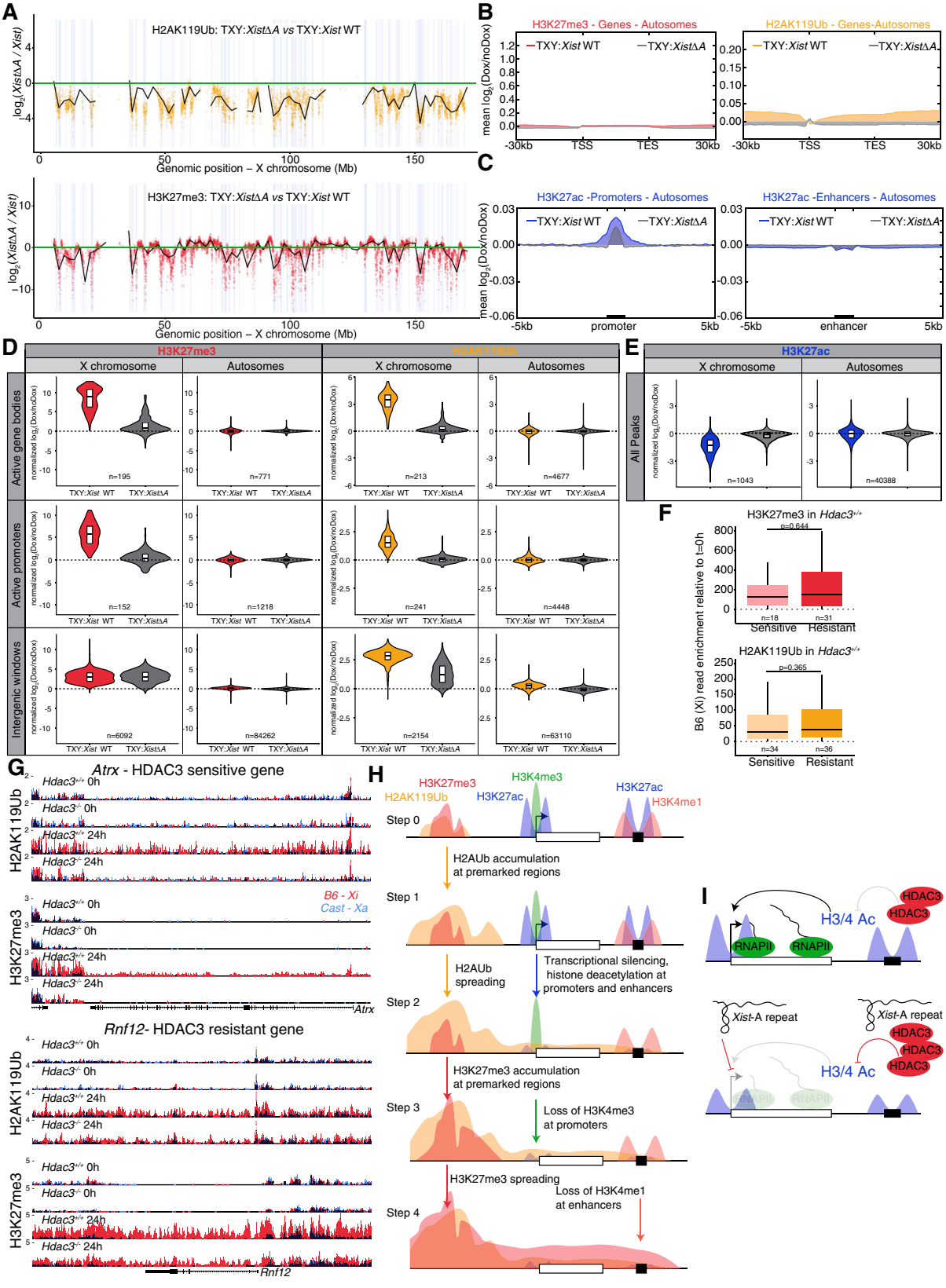
(A) Western blot analysis of *Hdac3*$^{Flag/Flag}$ line in comparison to *Hdac3*$^{+/+}$ ESC line probed with antibodies specific for FLAG, HDAC3 and Lamin B1. (B) Pyrosequencing validation of XCI efficiency in *Hdac3*$^{Flag/Flag}$ compared to *Hdac3*$^{+/+}$ ESC. Samples were collected after 0 and 24hrs of DOX treatment. (C) General annotation of all H3K27ac and HDAC3 peaks. (D) Genome browser track showing HDAC3 binding around *Xist* and *Gpc4*. Shown are allele-specific tracks (Cast-Xa: blue; B6-Xi: red). Green box points to HDAC3 peak becoming enriched for B6-Xi reads. Low-level accumulation of HDAC3 at the *Xist* locus is observed as well as skewing of 69% of HDAC3 peaks along the X chromosome. (E) Heatmap of normalized d-score for all HDAC3 peaks on the X chromosome. Peaks were clustered according to Xi (B6) read accumulation at 4hrs. (F) Violin plots showing the timing of deacetylation (left: H3K27ac; right: H4ac) at peaks accumulating HDAC3 on the Xi (B6) at 4hrs or not. (G) Boxplot showing the level of HDAC3 pre-bound at t = 0hr on the X chromosome at peaks deacetylated early and late. Outliers are not shown. (H) Violin plot showing timing of efficient H3K27ac loss (IC35) at regions pre-bound or not by HDAC3. (I) Violin plot showing timing of efficient transcriptional silencing (IC35) in relation to enhancer pre-binding by HDAC3. Promoter-enhancer pairs were identified as in Figure S1M. All p values were calculated using Wilcoxon rank sum test.



(legend on next page)

Figure S6. Allele-Specific Natic ChIP-Seq Monitors Changes in Representative Histone Modifications during XCI, Related to Figure 6

(A) Genome browser plots of H3K27me3 (red) and H2AK119Ub (yellow) enrichment at the *Hoxc* cluster. Shown are all mapping reads at all time points and biological duplicates. Note highly reproducible and stable pattern of enrichment. (B) Bar plot showing the percentage of B6 reads mapping to the X chromosome in the nChIP-seq time course. *p value < 0.05 calculated using Student's t test. (C) Plot showing the distribution of differential ED_{50} (H3K27me3-H2K119Ub) for all 10kb windows along the X chromosome. Each bar represents a single window, these were ordered accordingly to their differential ED_{50}. (D) IF/RNA FISH for *Xist* (gray), H2AK119Ub (red) and H3K27me3 (green) in day 1 or 2 EpiLCs. White dotted lines outline *Xist* domains. Bar plot quantifies PcG-mark enrichment at 100 *Xist* clouds. Scale bar = 10 μm. (E) Density plots correlating H2AK119Ub gain (4 versus 0hrs) with H3K27me3 gain (8 versus 4hrs). Shown is the behavior of all 10kb windows spanning the X chromosome. All scales are logarithmic. p value was calculated using Pearson correlation test. The correlation coefficient is $\rho = 0.622$ (F) Violin plots showing the distribution of ED_{50} for H2AK119Ub (left) and H3K27me3 (right) at windows mapping to the active gene bodies, active enhancers, intergenic regions and active promoters. p values calculated using Wilcoxon rank sum test. (G) ED_{50} analysis of H2AK119Ub (yellow), H3K27me3 (red) accumulation at gene bodies in comparison to gene silencing (black, left) and promoter H3K27ac loss (blue, right). p values calculated using paired Wilcoxon rank sum test. (H) Analysis of the point when sigmoid's second derivative reaches max/min. This is a proxy for when H2AK119Ub (yellow) and H3K27me3 (red) start to accumulate at gene bodies in comparison to the beginning of gene silencing (black, left) and promoter H3K27ac loss (blue, right). p values calculated using paired Wilcoxon rank sum test. (I) Dynamics of H2AK119Ub (yellow), H3K27me3 (red) accumulation, promoter deacetylation (H3K27ac, blue) and transcriptional silencing (black) of an X-linked gene: *Txlng*. Empty dots present ED_{50} of sigmoid fittings, while perpendicular lines max/min of second derivative (2ndD). Left axis of normalized d-score is for H3K27ac and TT-seq dynamics, right axis of normalized B6-read accumulation is for H2AK119Ub and H3K27me3. (J-K) Violin plots showing H2AK119Ub (J) and H3K27me3 (K) accumulation at gene bodies of constitutive and facultative escapees when compared to remaining genes. Shown is the normalized enrichment of B6-reads after 24hrs of DOX treatment when normalized to t = 0hr. p values calculated using Wilcoxon rank sum test. (L) Violin plot showing H3K27 deacetylation dynamics (IC35) at promoters of constitutive and facultative escapees when compared to remaining genes. p values calculated using Wilcoxon rank sum test.



(legend on next page)

Figure S7. Dependence of PcG Spreading on Gene Silencing; Proposed Models, Related to Figure 7

(A) Plots showing H3K27me3 (red) and H2AK119Ub (yellow) differential accumulation between *Xist:ΔA* and *Xist:WT* samples. Shown is average log2 fold change enrichment. Each dot represents a single 10kb window. Black line is a loess regression on all windows, active genes are in light blue. (B) Average plots showing average accumulation of H3K27me3 (left) and H2AK119Ub (right) over autosomal active genes ± 30kb in *Xist:ΔA* cells compared to *Xist:WT*. Shown is the mean normalized log2 enrichment of doxycycline versus no-doxycycline samples. (C) Average H3K27ac plots over autosomal promoters and enhancers (+/−5kb) in *Xist:ΔA* and *Xist:WT* expressing cells. Shown is the normalized mean log2 enrichment of doxycycline versus no-doxycycline samples. (D) Violin plots quantifying H3K27me3 and H2AK119Ub accumulation over active gene bodies, promoters and intergenic regions. All windows were split into X-linked and autosomal. (E) Violin plots quantifying H3K27ac depletion at its peaks. All peaks were split into X-linked and autosomal. (F) Boxplots quantifying H3K27me3 (red) and H2AK119Ub (yellow) accumulation on gene bodies at 24hrs compared to 0 hr in *Hdac3*$^{+/+}$. Shown are only B6 (Xi) accumulated reads. Genes were separated based on their sensitivity to HDAC3 loss. P values were calculated using Wilcoxon rank sum test. (G) Genome browser tracks showing H2AK119Ub and H3K27me3 accumulation at a gene sensitive (*AtrX;* top) or resistant (*Rnf12;* bottom) to HDAC3 loss. B6- and Cast-specific reads are shown in red and blue. Allele-specific tracks were overlaid. (H) Stepwise model of the epigenomic roadmap for XCI. Step 1: H2AK119Ub accumulates intergenically at sites of *Xist* binding, which are pre-marked by PcG. This stage is complete by 4 hr. Step 2: Concurrently with the initiation of transcriptional silencing, H3K27 starts to be deacetylated at promoters and H2AK119Ub commences to spread into gene bodies. Both transcriptional silencing and histone deacetylation are entirely dependent on the A-repeat of *Xist*. Moreover, HDAC3 is a key player involved in promoting efficient transcriptional silencing by deacetylating promoters and enhancers. This step is complete for early silenced genes by 4 hr, while other genes will typically reach this point by 8hrs. H2AK11Ub spreading into genes necessitates the initiation of gene silencing. Step 3: Decommissioning of promoters by H3K4me3 loss follows a protracted dynamic. H3K27me3 deposition at intergenic regions is also delayed when compared to H2AK119Ub enrichment. These events take place around the 8hrs time point. Step 4: Even more delayed is the decommissioning of enhancers by the loss of H3K4me1, indicating that it is the consequence of gene silencing rather than its cause. Similarly, H3K27me3 spreading into genes occurs after, and is dependent on, gene silencing. This indicates that its role in initiating transcriptional silencing is unlikely but does not preclude its importance in XCI maintenance. Rapidly silenced genes can reach this point already after 8 hr of DOX treatment, while other genes need 24 hr. (I) Model of the involvement of HDAC3 in XCI. Gene silencing is mediated mainly through the HDAC3, which is pre-bound at enhancers already prior to *Xist* upregulation. Upon XCI induction, HDAC3 is only modestly recruited to the Xi, while the pre-bound HDAC3 presumably becomes activated. Full activation of HDAC3 is entirely dependent on the *Xist* A-repeat region as no deacetylation is observed along the Xi when this element is deleted. How *Xist* A-repeat achieves this remains unclear. Most likely HDAC3 activation occurs thanks to *Xist* A-repeat interaction with SPEN, and in turn, this has been proposed to recruit SMRT (Chu et al., 2015; McHugh et al., 2015; Monfort et al., 2015). We observe that HDAC3 in the context of mouse ESC is strongly bound to both SMRT and NCOR components. Such SPEN-SMRT interaction could allow for HDAC3 mediated histone deacetylation and efficient gene silencing (McHugh et al., 2015).

Chapter 6

SPEN function and mechanisms in XCI *(Dossin et al., Nature. 2020)*

6.1 Rationale for investigating SPEN and XCI

Although XCI is associated with a plethora of chromatin changes (see **section 1.8** and **chapter 5**), no chromatin modifying enzymes were found to be recruited directly by *Xist* RNA [Minajigi et al., 2015; Chu et al., 2015; McHugh et al., 2015]. Intriguingly, *Xist* instead interacts with a vast array of proteins, which were hitherto never associated with XCI and whose functions are poorly characterized (see **section 1.13**). Understanding how these *Xist*-binding proteins govern subsequent changes in transcription, chromatin, and spatial organization of the X chromosome is now of paramount importance.

When I joined Edith's lab in 2017, the lab of Neil Brockdorff had extensively characterized how one of these proteins, hnRNPK, contributes to the process of XCI (from its direct association with the B-repeat of *Xist* to its requirement for Polycomb recruitment, see **subsection 1.13.1**). However, such characterization had not been done for **SPEN**.

The originally proposed model for the role of SPEN in X chromosome inactivation suggested that *Xist* recruits SPEN which in turn recruits NCoR/SMRT and HDAC3 to silence transcription [McHugh et al., 2015]. Importantly in this study, there were no direct evidence that SPEN is indeed recruited by *Xist* to the Xi during XCI. Furthermore, its mechanism of action (through HDAC3) only relied on the observation that siRNA-mediated, individual knockdowns of *Hdac3* or *Spen* show similar defects in silencing of one X-linked gene, *Gpc4* [McHugh et al., 2015], with no physical evidence for SPEN-dependent recruitment of HDAC3 during XCI.

Our observation that HDAC3 is not recruited during XCI but is instead pre-bound to active enhancers (see **subsection 5.3.2** and [Żylicz et al., 2019]) suggests that this model needs to be revised. Furthermore, our analysis of *Hdac3* KO cells, which show *delayed* but not fully deficient X-linked gene silencing (see **subsection 5.2.3** and [Żylicz et al., 2019]), demonstrates that HDAC3 is not the only factor at play in initiating XCI.

How about SPEN?

Is SPEN also only partially important for XCI, or is it required to a larger extent than HDAC3? Is SPEN actively recruited to the X chromosome? If so, with which dynamics and where does it preferentially bind? Which protein partners does it engage with? Is SPEN also important for maintenance of XCI, namely to keep the inactive X chromosome silenced and properly structured in somatic cells?

Despite five studies universally pointing out to SPEN as an important player during XCI [McHugh et al., 2015; Chu et al., 2015; Minajigi et al., 2015; Monfort et al., 2015; Moindrot et al., 2015], these fundamental questions remained unanswered.

I hence decided that I would dedicate the core of my PhD work to properly characterize the function and essentiality of SPEN in the process of XCI, and attempt to dissect its mechanism of action.

In the following sections of this chapter, I intend to report the scientific methodology I followed – outlining it chronologically, as it went by throughout my PhD – to uncover different clues about SPEN's function (Goal 1, see **section 6.2**) and mechanism of action (Goal 2, see **section 6.3**) in XCI (**Figure 6.1**).

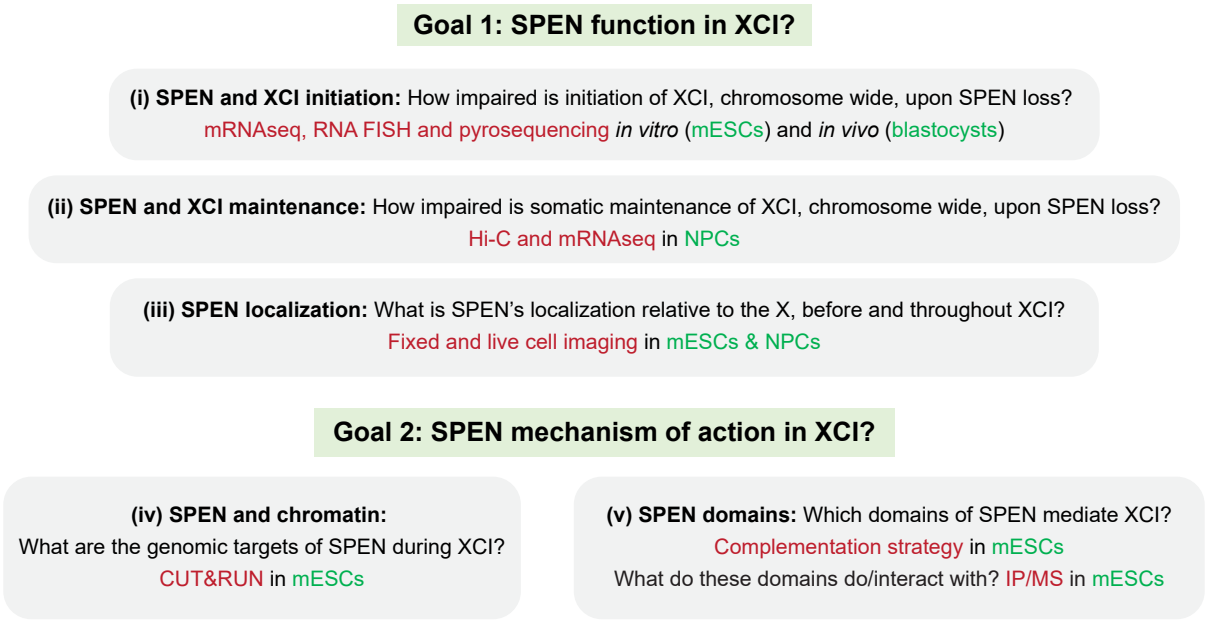


Figure 6.1 – Flowchart depicting the questions addressed and methods employed for each goal of the SPEN project.

The results I obtained will be presented very summarily in the following sections, as they were published early 2020 in Nature, and are to be found in the main text/figures as well as the extended data figures of the article appended at the end of this chapter (**section 6.4** and **[Dossin et al., 2020]**).

6.2 Goal 1: Addressing SPEN function and window of opportunity during XCI

The first line of experiments I set up were aimed to gain insights into (**Figure 6.1**):

(i) The extent to which SPEN regulates gene silencing chromosome wide during initiation of XCI.

(ii) Whether SPEN plays a role for keeping genes silenced (and the Xi structured) during maintenance of XCI in somatic cells.

(iii) Whether SPEN is actively recruited to the Xi throughout the XCI process.

6.2.1 Dissecting SPEN function in initiating XCI *in vitro*

To address points (i) and (ii), a SPEN loss of function approach had to be employed. The observation that *Spen* KO mouse embryos die only at E13.5 [Yabe et al., 2007], long after the blastocyst stage, suggests that SPEN is not essential in embryonic stem cells. Hence, a classical constitutive knock-out of *Spen* could have been carried out in our TX1072 mESC system to interrogate its functional importance during initiation of XCI.

However, such constitutive KO approach would have been incompatible with my aim to address whether SPEN also regulates maintenance of XCI in differentiated cells. Indeed, such aim requires that mESCs are first differentiated in the presence of SPEN, only after which SPEN could be removed and the XCI status interrogated. Furthermore, no previous reports allowed us to infer whether SPEN-deficient mESCs could undergo differentiation properly. Hence, a *conditional* SPEN loss of function approach was preferred.

I made use of the auxin-inducible degron strategy [Nishimura et al., 2009], which allows acute and rapid depletion of any protein of interest, upon treatment with the plant hormone auxin (**Figure 6.2**). This approach relies on the hetereologous expression of TIR1, a rice-derived F-box protein [Gray et al., 2001; Nishimura et al., 2009]. Expressed in mammalian cells, TIR1 associates with endogenous components of the SCF complex, responsible for polyubiquitylation and subsequent proteasomal degradation of protein targets [Cardozo and Pagano, 2004]. Remarkably, through its TIR1 moiety, and only in

the presence auxin, the SCF^{TIR1} complex has the ability to interact with and promote the degradation of any protein carrying an *a*uxin-*i*nducible *d*egron (AID) motif (**Figure 6.2**) [Dharmasiri et al., 2005; Gray et al., 2001; Tan et al., 2007].

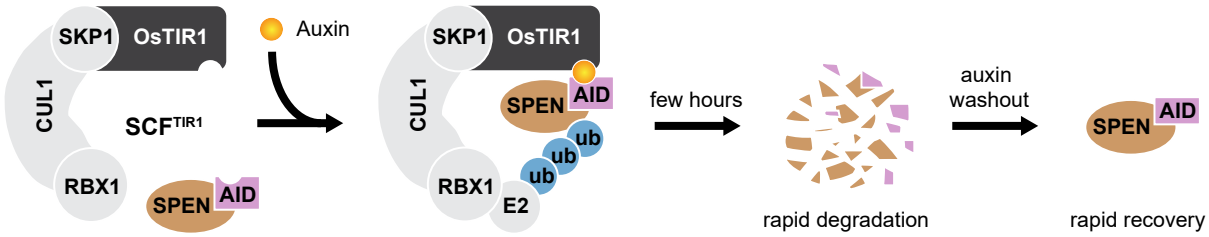


Figure 6.2 – Scheme depicting the auxin inducible degron system applied to SPEN. Auxin leads to TIR1-dependent polyubiquitylation and proteasomal degradation of SPEN tagged with the auxin-inducible degron (AID). Auxin removal leads to protein recovery.

This system has routinely been used across several model organisms, and allows to address protein function directly, with highest specificity, without perturbing neither the gene nor the transcript encoding such protein. Furthermore, the rapidity with which target proteins are depleted in this system minimizes the appearance of compensatory effects that would otherwise complicate data interpretation.

6.2.1.1 Generation of SPEN auxin-inducible degron mESCs

To implement this approach in our TX1072 mESC system (see **subsection 5.1.1**), I used CRISPR/Cas9 to target a constitutively expressed *OsTir1* gene at the *Tigre* locus and, using the same methodology as I did for *Hdac3* (see **subsection 5.3.1** and **Figure 5.2A**), I performed a homozygous knock-in of the AID degron[¶] at the C-terminus of *Spen* (see Dossin et al., External Data Fig. 1a). I also included a HaloTag [Los et al., 2008], additionally to the AID, allowing for visualization of SPEN both by western blot and microscopy, using HaloTag antibodies and fluorescent-ligands respectively. This work was done under the mentorship of Inês Pinheiro, a postdoc who initially trained me upon my arrival in Edith's lab.

With this system, SPEN could be efficiently depleted within 1 hour of auxin treatment (see Dossin et al., Fig. 1b and External Data Fig. 1b). Hence, this cellular system allows us to controllably induce XCI while simultaneously depleting SPEN (**Figure 6.3**).

[¶]Throughout my PhD, I generated several constructs to target AID-fusion cassettes at the N- or C-termini of endogenous genes/proteins. These constructs were used in the context of a collaboration with the Bourc'his lab, to investigate the role of the m^6A RNA methylation pathway in controlling the expression of transposable elements, resulting in a publication in Nature **[Chelmicki et al., 2021]**, which is appended at **section A.1**.

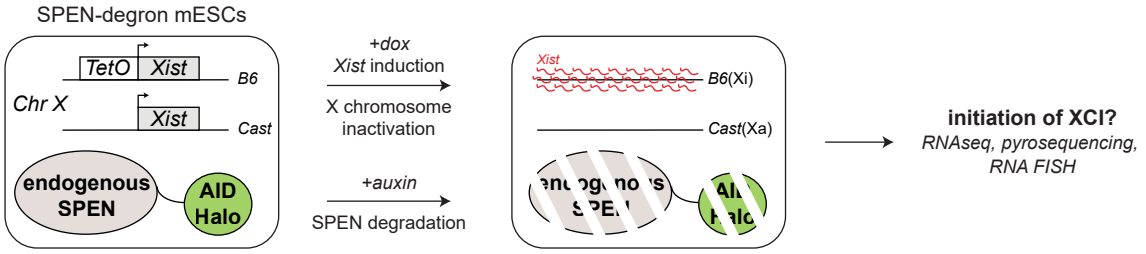


Figure 6.3 – Scheme depicting the strategy to investigate SPEN's importance during *initiation* of XCI *in vitro*. SPEN-degron TX1072 mESCs allow conditional dox-triggered XCI combined with auxin-mediated SPEN depletion.

6.2.1.2 SPEN is essential for initiating XCI *in vitro*

Using allele specific RNA-seq, pyrosequencing and RNA FISH, I revealed that acute depletion of SPEN in mESCs leads to drastic defects in X-linked gene silencing, without affecting *Xist* RNA localization (see Dossin et al., Fig. 1a-f and External Data Fig. 1a-e, and **chapter 8** for further discussion).

6.2.2 Dissecting SPEN function in initiating imprinted XCI *in vivo*

To extend this observation *in vivo*, I adressed the status of XCI in *Spen* KO mouse embryos, collaborating with Jan Zylicz, an embryologist and postdoc from Edith's lab:

As previously mentioned in **section 1.3**, two waves of XCI (imprinted and random) occur during mouse embryonic development (**Figure 1.2**). Monitoring random XCI (which initiates at the late blastocyst stage) using an allele specific readout, chromosome-wide, is challenging. Indeed, it requires that the extraembryonic cells are removed (since they have undergone imprinted XCI), and then ivolves the use of single-cell RNA-seq, as half of the embryonic cells will inactivate paternal X-linked alleles, and the other half will inactivate maternal X-linked alleles.

Instead, monitoring *imprinted* XCI (which has been established in all cells by the early blastocyst stage) is compatible with total-embryo allele specific RNAseq, since all the cells have inactivated strictly the paternal X-linked alleles.

6.2.2.1 Mouse crossing strategy

Hence, for practical reasons, we addressed the importance of SPEN only during imprinted XCI. To do so, we used a published conditional (i.e. floxed) *Spen* KO allele [Yabe

et al., 2007], and generated F_1 hybrid *Spen*$^{KO/WT}$ males carrying a *Cast/EiJ*-derived X chromosome (**Figure 6.4A**). These males were crossed with *C57BL/6J Spen*$^{flox/flox}$ females (**Figure 6.4A**), carrying a *Rosa26:Zp3-Cre* allele [De Vries et al., 2000], allowing for maternal *Spen* deletion in oocytes (**Figure 6.4B**).

Following this cross, both maternal (i.e. inheriting only a functional *Spen* allele of paternal origin) and maternal-zygotic (i.e. inheriting non-functional *Spen* alleles from both parents) *Spen* KO female embryos were obtained (**Figure 6.4C**). By the early blastocyst stage (E3.5), only the paternal X chromosome – of *Cast/EiJ* origin – should be inactivated (**Figure 6.4D**). X-linked transcription in these female embryos was profiled using allele-specific single embryo RNAseq (**Figure 6.4D**).

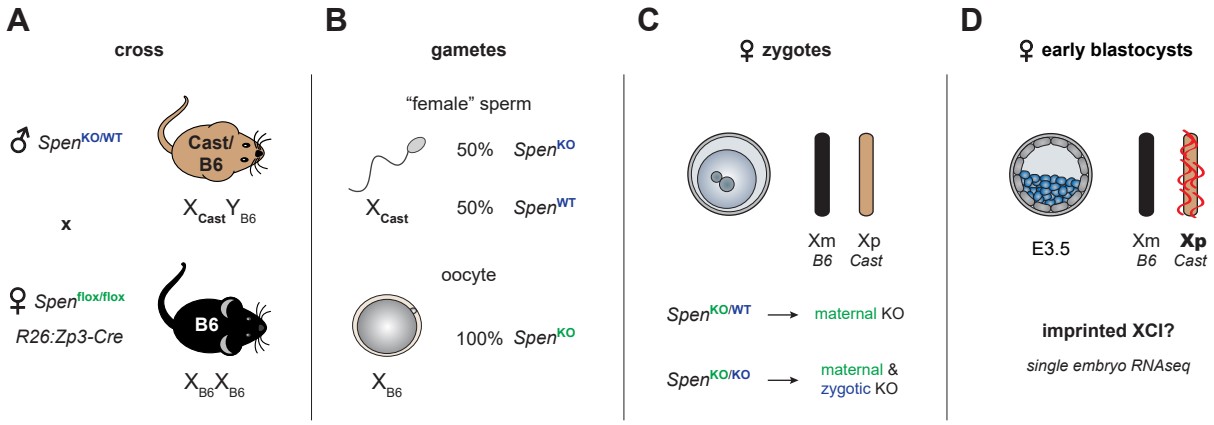


Figure 6.4 – Mouse crossing strategy to investigate SPEN's importance during *imprinted* XCI *in vivo*. Imprinted XCI is assayed using single embryo RNAseq in *Spen* maternal and maternal-zygotic KO early blastocysts.

6.2.2.2 SPEN is essential for imprinted XCI *in vivo*

This experiment revealed that similarly to what is observed in SPEN-degron mESCs or *Xist* KO embryos [Borensztein et al., 2017b], *Spen* KO female embryos are almost fully deficient for X-linked gene silencing during imprinted XCI, despite sustained paternal *Xist* expression (see Dossin et al., Fig. 1g, h and External Data Fig. 1f-h, and **section 8.4** for further discussion).

6.2.3 Dissecting SPEN function during maintenance of XCI in somatic cells

Having established that SPEN is essential for initiating *Xist*-dependent gene silencing during XCI, it became relevant to investigate whether SPEN is also required to *maintain*

genes silenced after XCI has been established in differentiated cells. Hence, I differentiated SPEN-degron mESCs into *n*eural *p*rogenitor *c*ells (NPCs), isolated indendent NPC clones and performed allele-specific RNA-seq following either 24 or 48 hours of SPEN depletion (**Figure 6.5**).

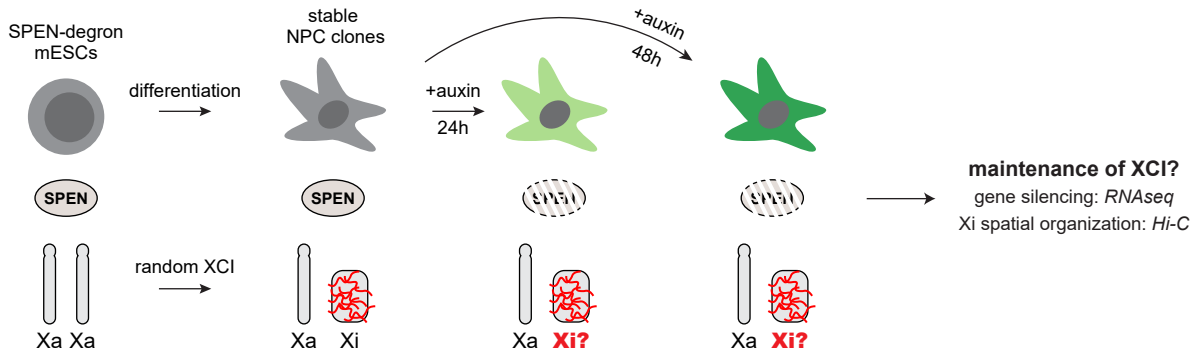


Figure 6.5 – Scheme depicting the strategy to investigate SPEN's importance during *maintenance* of XCI in NPCs. Following differentiation of SPEN-degron mESCs into stable NPC clones, the importance of SPEN in maintaining gene silencing and structure of the Xi is assayed using RNAseq and Hi-C respectively, following 24 or 48 hours of SPEN depletion.

6.2.3.1 SPEN is dispensable for keeping genes silenced, but buffers transcription of escapees

Loss of SPEN did not lead to reactivation of fully silenced genes (see Dossin et al., Fig. 2c, d), demonstrating that maintenance of gene silencing on the Xi is ensured independently of SPEN. However, transcription of escaping genes was moderately, but significantly, upregulated (see Dossin et al., Fig. 2e, f), suggesting that SPEN buffers overexpression of escapees in female somatic cells (discussed in **section 8.5**).

6.2.3.2 The peculiar organization of the Xi is maintained independently of SPEN

Given that *Xist* plays a central role in reorganizing the 3D-structure of the Xi in somatic cells (**see section 1.10**), I further examined whether SPEN – as a *Xist*-binding protein – is also involved in such structural effects. To do so, I depleted SPEN for 48 hours in NPCs and, in collaboration with Job Dekker's lab, performed allele-specific Hi-C (**Figure 6.5**).

Loss of SPEN had no effect on the Xi conformation, neither at the level of megadomains nor at the level of escapee TAD-like structures (see Dossin et al., External Data Fig. 2e-g), hinting that *Xist* RNA structures the Xi independently of SPEN.

6.2.4 Imaging of SPEN and *Xist* RNA dynamics during XCI

The importance of SPEN during initiation of XCI prompted me to explore its cellular localization relative to the X chromosome, throughout the process of XCI. Importantly, I aimed to address whether SPEN is actively recruited by *Xist*, and if so, with which dynamics.

6.2.4.1 Fixed-cell imaging

First, I monitored SPEN localization throughout a timecourse of *Xist* induction, in fixed cells. To do so, dox-induced SPEN-HaloTag degron mESCs were labeled in culture with a fluorescent Halo-ligand [Grimm et al., 2015], fixed, and processed for RNA FISH against *Xist*.

Remarkably, as early as 2 hours following *Xist* expression, SPEN strongly accumulated into domains corresponding to *Xist* RNA clouds (see Dossin et al., Fig. 2a). This accumulation was sustained and amplified throughout later time points; and SPEN also accumulated around the Xi in differentiated NPCs (see Dossin et al., Fig. 2a, b).

While this first set of results suggested that SPEN is recruited to the Xi fairly early (2 hours) upon *Xist* accumulation, the low temporal resolution achieved within the fixed-cell imaging timecourse prevented me from monitoring whether SPEN is in fact recruited *immediately* following *Xist* expression. To circumvent this limitation, I had to track SPEN and *Xist* molecules in live cells, allowing me to record *Xist*/SPEN dynamics with higher temporal resolution, in individual cells.

6.2.4.2 Live-cell imaging

To directly follow SPEN, I performed a homozygous knock-in of GFP at its C-terminus. To visualize *Xist* RNA in live cells, Julia Roensch, a researcher from the lab, inserted an array of Bgl stem loops (BglSL, [Chen et al., 2009]) within *Xist* exon 7 (on the dox-inducible allele), in a background where a BglG-mCherry fusion protein is constitutively expressed (**Figure 6.6A**). Upon *Xist* expression, BglG-mCherry binds the BglSL, allowing direct, live cell visualization of *Xist* RNA molecules (**Figure 6.6B**).

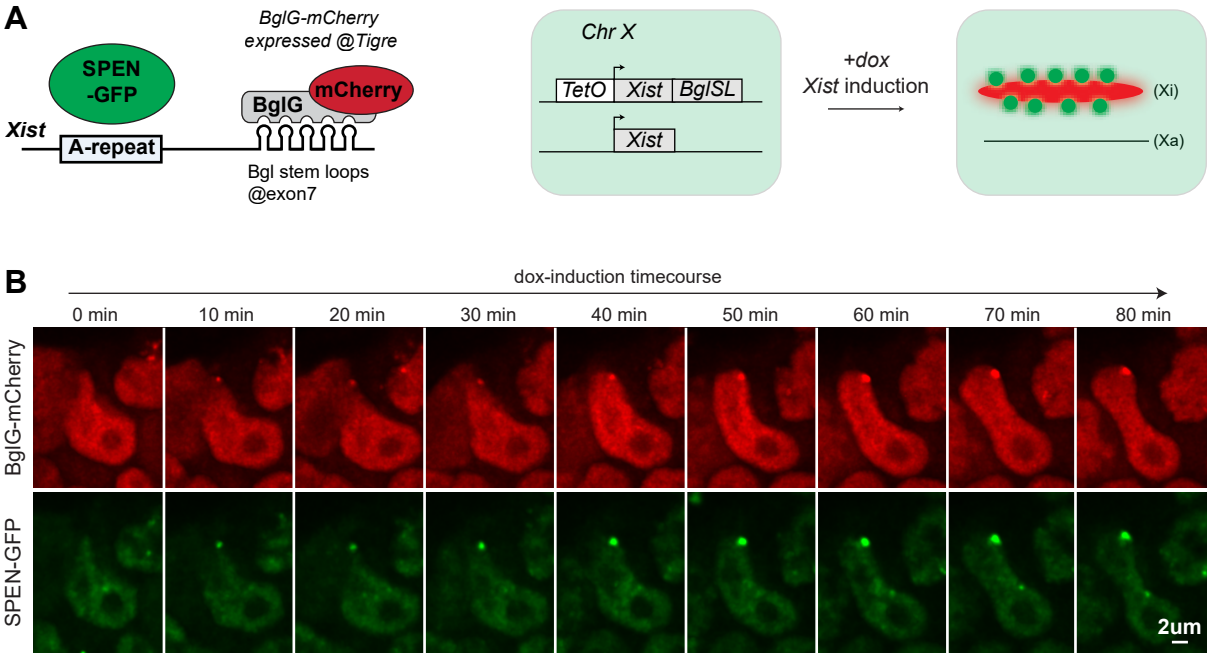


Figure 6.6 – Live cell imaging of SPEN and *Xist* RNA in mESCs. **(A)** BglSL insertions within *Xist* exon 7 allow *Xist* RNA visualization in live cells through binding of a BglG-mCherry fluorescent fusion protein. The BglG-mCherry is constitutively expressed from the *Tigre* locus, while SPEN is endogenously tagged with a GFP. **(B)** Live-cell microscopy snapshots of a representative cell showing *Xist* (red) and SPEN (green) dynamics every 10 minutes following doxycycline-mediated induction of *Xist* RNA.

Using this cell line, Julia and I performed time-lapse microscopy, imaging cells every 10 minutes, for 4 hours following doxycycline induction. We acquired dozens of good quality movies, from which nuclei and *Xist* clouds could be segmented for each timeframe, either manually, or with machine learning using Ilastik [Berg et al., 2019]. For each nucleus and within each segmented cloud surface, pixel intensities for both the SPEN (green) and *Xist* (red) channels were retrieved and tracked through time (see **section 3.15** for further details).

Strikingly, SPEN accumulation was detected as soon as (within ±10 minutes) *Xist* domains appeared (see Dossin et al., External Data Fig. 2a-d). Altogether, these results demonstrate that SPEN is recruited to the X chromosome to initiate gene silencing *immediately* upon *Xist* RNA upregulation. Its sustained accumulation around the Xi in NPCs is consistent with its role in buffering transcription of escaping genes.

6.3 Goal 2: Gaining mechanistic insights into SPEN function during XCI

Having established that SPEN is an essential player for chromosome-wide gene silencing during initiation (but not maintenance) of XCI, I next sought to gain further mechanistic insight into how SPEN carries out such function.

The second line of experiments I set up were aimed to identify (**Figure 6.1**):

(iv) What are the genomic targets – if any – of SPEN during initiation of XCI.
(v) Which domains of SPEN mediate its function, and which protein partners do they interact with.

6.3.1 Profiling SPEN chromatin binding during XCI

The remarkable accumulation of SPEN to the inactivating X chromosome upon *Xist* upregulation observed by microscopy suggested that identifying the genomic regions SPEN targets throughout XCI could help unravel how it contributes to gene silencing.

Classically, one would perform ChIP-seq against SPEN to identify its genomic targets. However, whether SPEN genuinely binds DNA/chromatin directly is unknown, and all my attempts to immunoprecipitate it, either under native or cross-linked conditions, and using a wide array of affinity epitopes (HA, GFP, Flag, HaloTag) were not successful. Furthermore, in a classical ChIP protocol, chromatin fragmentation (typically achieved using sonication) would majorly disrupt the *Xist* RNA domain, and potentially lead to a major loss in the SPEN fraction enriched on the X chromosome.

6.3.1.1 Adapting CUT&RUN for SPEN and XCI

At that time of my PhD, a novel chromatin profiling strategy, which involved neither immunoprecipitation nor chromatin fragmentation, was developed by the lab of Steve Henikoff [Skene and Henikoff, 2017]. This method, called CUT&RUN (*c*leavage *u*nder *t*argets and *r*elease *u*sing *n*uclease), is performed *in situ*, on intact cells (**Figure 6.7A**):

Following gentle membrane permeabilization, cells are incubated with an antibody against any chromatin-associated factor of interest. Then, a pA-MNase fusion protein is tethered to the antibody/chromatin-associated factor complex (through its protein A moiety). Upon activation with divalent cations (e.g. Ca^{2+}), the MNase moiety cleaves DNA around the transcription factor binding site, resulting in the "release" of this specific

chromatin fragment, which allegedly diffuses out of the nucleus. These released fragments are then retrieved from the supernatant, and subsequent DNA sequencing directly reveals the genomic targets of the chromatin-associated factor of interest (**Figure 6.7A**).

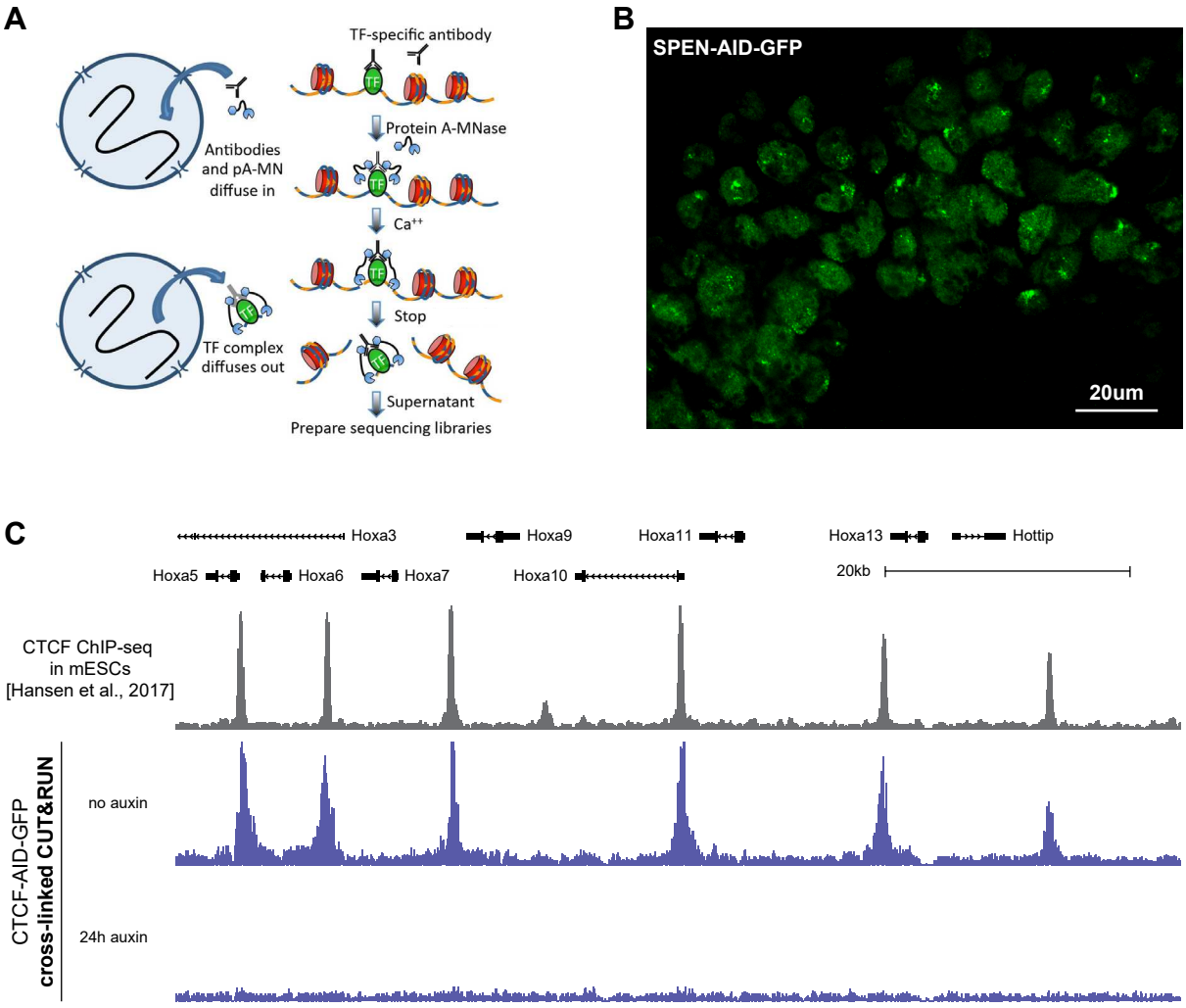


Figure 6.7 – Cross-linked CUT&RUN successfully maps transcription factor binding sites. **(A)** Scheme depicting the CUT&RUN methodology. Image directly reproduced from [Skene and Henikoff, 2017]. **(B)** Microscopy image of dox-induced SPEN-GFP cells processed for optimized cross-linked CUT&RUN shows that SPEN accumulation on the X-chromosome is preserved for 48 hours under such conditions. **(C)** UCSC genome browser tracks show that cross-linked CUT&RUN faithfully maps CTCF binding sites in mESCs. All tracks are scaled identically, and a cluster of *Hox* genes is shown as an example.

Performed *in situ*, this method would in principle be more suitable than ChIP to preserve the *Xist*/SPEN nuclear domain intact. However, domains of *Xist* RNA are readily destabilized following permeabilization, with major RNA degradation occurring. Given that SPEN's chromatin enrichment depends on *Xist* RNA, I opted to perform CUT&RUN under *cross-linked* conditions, which would ensure stabilization of SPEN domains.

With optimization, I identified cross-linking and permeabilization conditions that preserved SPEN accumulation throughout the CUT&RUN procedure, for up to 48 hours following time of fixation (**Figure 6.7B**).

Under these conditions, I first performed CUT&RUN against the transcription factor CTCF, whose genomic binding sites in mESCs have been characterized extensively. Importantly, I made use of a published CTCF-GFP auxin-degron mESC line [Nora et al., 2017], performing CUT&RUN with a GFP antibody (**Figure 6.7C**), as I would then perform CUT&RUN on a SPEN-GFP mESC line with that same antibody.

Remarkably, CTCF binding profiles looked very similar between my cross-linked CUT&RUN dataset and a previously published CTCF ChIP-seq in mESCs [Hansen et al., 2017] (**Figure 6.7C**). Furthermore, the profiles obtained upon auxin-mediated CTCF degradation demonstrated that the signal observed is specific to GFP-tagged CTCF, and that the GFP antibody can be used successfully for cross-linked CUT&RUN.

6.3.1.2 SPEN CUT&RUN sheds light on its mechanism of action during XCI

I hence performed cross-linked CUT&RUN against SPEN, during a time-course of *Xist* induction (0h, 4h, 8h and 24h dox as well as 8h dox+auxin as a negative control) in 2 biological replicates. Several remarkable observations emerged from this experiment:

First, in the absence of *Xist* upregulation, SPEN shows very few binding sites across the genome (see Dossin et al., External Data Fig. 2a), suggesting that SPEN does not readily associate with chromatin in mESCs. This is consistent with our observation that very few genes are differentially expressed following SPEN depletion in SPEN-degron mESCs.

Upon *Xist* RNA coating (as early as 4 hours following dox-mediated induction), several hundreds of SPEN binding sites appear specifically on the X chromosome undergoing inactivation (see Dossin et al., Fig. 4a and External Data Fig. 4a), confirming that SPEN association with X-linked chromatin is dependent on *Xist* RNA. Furthermore, the enrichment of SPEN, early after *Xist* coating, is highest within *Xist* entry sites (see **section 1.5**), hinting that SPEN enrichment on X-linked chromatin follows the spatial dynamics of *Xist* coating (see Dossin et al., Fig. 4e).

A strong enrichment of SPEN is observed across a large fraction of the *Xist* gene body, mostly at the 5' half of the transcription unit (see Dossin et al., Fig. 4b). Given that the A-repeat of *Xist* – located at the very beginning of the transcript – directly binds SPEN *in vitro* [Monfort et al., 2015], I interpret this diffuse CUT&RUN profile as the MNase footprint resulting from SPEN binding to nascent *Xist* RNA transcripts

(**Figure 6.8B**). Such conclusion is consistent with the microscopy-based observation that SPEN is recruited to the X chromosome immediately upon *Xist* upregulation (see Dossin et al., External Data Fig. 2a-d).

Apart from *Xist*, no other X-linked loci show such peculiar diffuse binding pattern. Instead, SPEN shows *focal* and *discrete* enrichment, specifically targeted to gene promoters and enhancers (see Dossin et al., Fig. 4c, d, External Data Fig. 4c, and External Data Fig. 5a-g).

Strikingly, SPEN specifically targets promoters of *initially active* genes (see Dossin et al., Fig. 4f and External Data Fig. 4d). Consistently, X-linked genes which are most dependent on SPEN for silencing (see Dossin et al., Fig. 1e) – expectedly showing highest levels of SPEN recruitment at their promoters – are in fact more highly expressed initially than genes which are less dependent on SPEN for their silencing (see Dossin et al., Fig. 4g, and External Data Fig. 4e). Taken together, these observations strongly suggest that active transcription provides a favorable context for SPEN recruitment to chromatin (**Figure 6.8C**), which is further discussed in **subsection 9.1.2** and **subsubsection 9.2.2.3**.

Another important observation is that the levels of early SPEN recruitment at X-linked promoters (within 4 hours of *Xist* coating) are higher for genes which are efficiently (i.e. early) silenced than they are for genes which are less efficiently (i.e. late) silenced during XCI (see Dossin et al., Fig. 4h). The same trend is observed between SPEN recruitment at X-linked enhancers and the efficiency with which these enhancers are deacetylated during XCI (see Dossin et al., Fig. 4i). Finally, SPEN binding at promoters is lowest – nay absent – at genes that undergo no significant silencing – if not complete escape – in our *Xist*-inducible system (see Dossin et al., External Data Fig. 4f, g). Taken together, these three observations strongly indicate that SPEN recruitment to promoters and enhancers is directly responsible for transcriptional silencing during XCI (**Figure 6.8C** and **D**).

Remarkably, SPEN binding to X-linked chromatin decreases drastically after 24 hours of *Xist* expression (see Dossin et al., Fig. 4a and Extended Data Fig. 4c), despite very strong *Xist*/SPEN clouds being observed by microscopy at that timepoint.

To better understand the biological signification of this drop in SPEN binding, X-linked promoters were clustered according to their dynamics of SPEN binding/"release" within the 24h timecourse of *Xist* coating (see Dossin et al., Extended Data Fig. 4i), and the silencing dynamics of their corresponding genes was compared (see Dossin et al., Extended Data Fig. 4j):

This analysis distinguished three groups of promoters: In the first group, SPEN binding was maximal within 4h of *Xist* coating but decreased thereafter, with major loss of SPEN binding observed by 24 hours (see Dossin et al., Fig. 4j). In the second group however, SPEN binding was maximal only after 8h of *Xist* coating, and loss of SPEN binding was only mildly observed by 24 hours (see Dossin et al., Fig. 4j). Finally, the third group, which comprises significantly fewer promoters than the two previous ones, was characterized by significantly reduced levels of SPEN binding (and loss thereof) throughout the *Xist* coating timecourse (see Dossin et al., Fig. 4j).

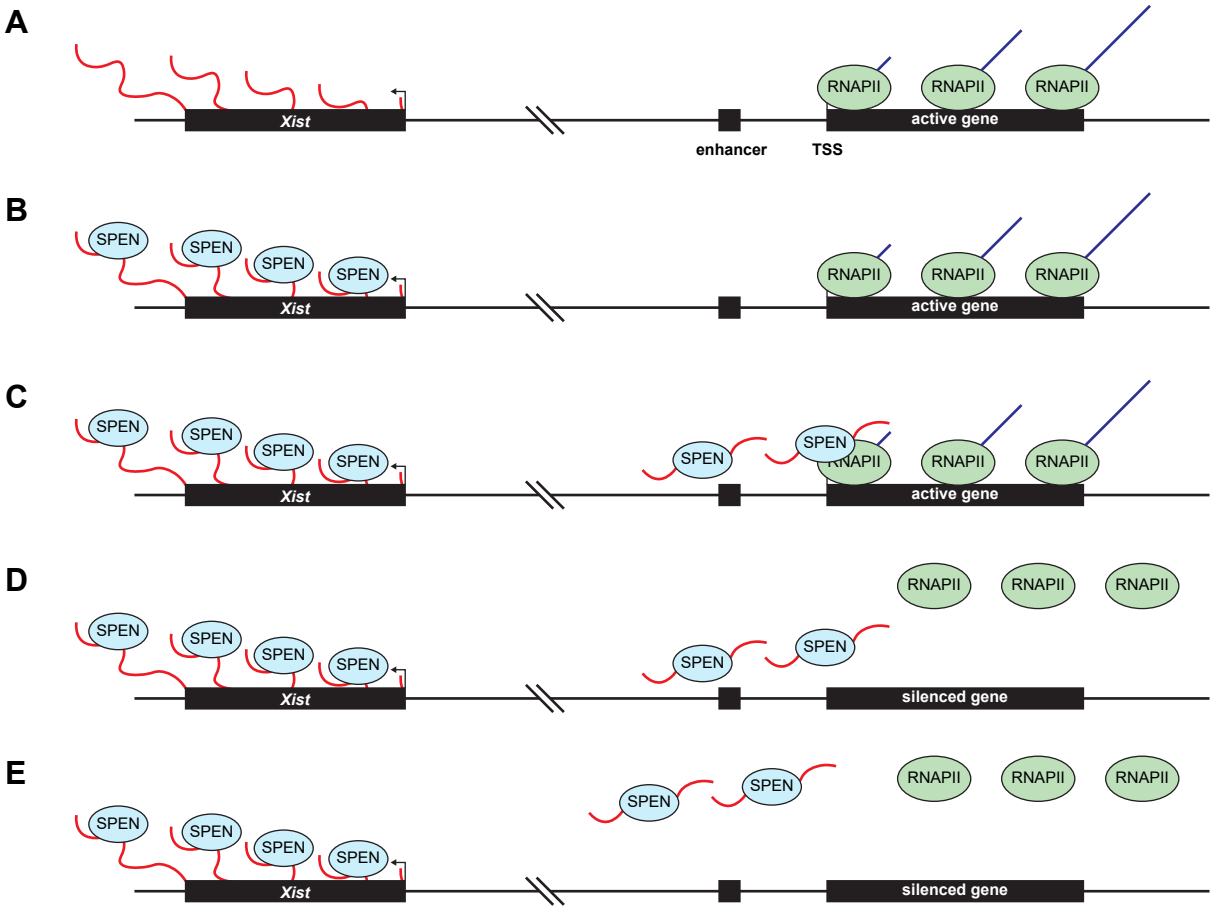


Figure 6.8 – Model for SPEN chromatin recruitment during XCI, derived from CUT&RUN analysis. (A) XCI initiates with the monoallelic upregulation of *Xist*. **(B)** *Xist* recruits SPEN immediately after its A-repeat has been transcribed, leading to CUT&RUN signal across the entire *Xist* transcription unit throughout XCI. **(C)** Once associated with *Xist*, SPEN is guided to enhancers and promoters of active genes specifically, **(D)** where it silences transcription. Gene silencing (i.e. loss of transcribing RNAPII) results in **(E)** dissociation of SPEN from chromatin

When comparing the efficiency of gene silencing (by 24h of *Xist* coating) between each of the three promoter/gene groups aforementioned, the group which lost SPEN binding most efficiently showed strongest gene silencing by 24h compared to the two other groups that showed protracted SPEN binding (see Dossin et al., Fig. 4k). This suggests that

upon gene silencing, SPEN binding to chromatin is no longer possible (**Figure 6.8D** and **E**), further supporting the idea that active transcription is required for SPEN targeting.

Altogether, this CUT&RUN analysis favors a model wherein *Xist* recruits SPEN to gene promoters and enhancers, which serve as the substrates for SPEN's silencing action. SPEN associates with chromatin in a transcription-dependent manner, as it only targets active promoters and enhancers, and disengages from chromatin upon gene silencing (**Figure 6.8**).

6.3.2 Identifying the domains of SPEN required for its function during XCI

Although a lot of insight into how SPEN works to silence genes was gained from CUT&RUN analysis, the protein factors SPEN could engage with to mediate such silencing remained unknown. Originally, I had planned to perform SPEN immunoprecipitation followed by mass spectrometry to unravel the protein partners of SPEN. Unfortunately, and as mentioned in **subsection 6.3.1**, all my attempts at immunoprecipitating SPEN (carrying different affinity tags) were unsuccessful. Importantly, protein extraction under native conditions consistently resulted in significant full-length SPEN degradation (data not shown), further complicating my attempts at characterizing its protein interactome.

6.3.2.1 SPEN functional complementation methodology

Hence, I opted to first determine which regions of the SPEN protein are required for its gene silencing function during XCI, and only then characterize the protein interactome of these specific regions. I hence developed a strategy to express a variety of truncated SPEN proteins (**Figure 6.9A**) and assess whether these truncations could functionally rescue the XCI defects observed upon the loss of endogenous SPEN (**Figure 6.9B**).

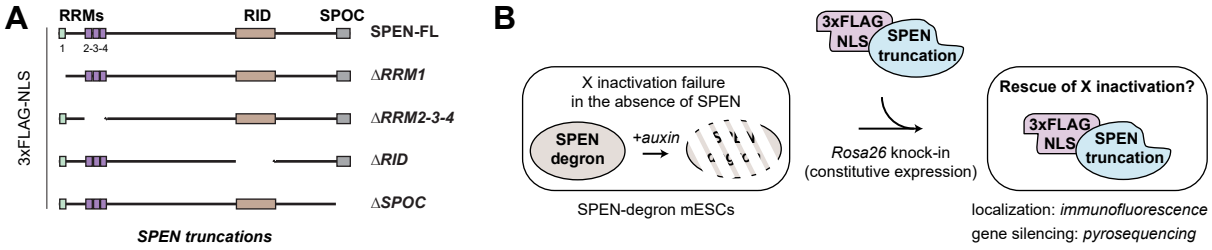


Figure 6.9 – A functional complementation strategy to identify the domains of SPEN involved in XCI. (A) Different SPEN cDNA truncations are **(B)** knocked-in at *Rosa26* and constitutively expressed in SPEN-degron mESCs. Upon *Xist* upregulation and simultaneous depletion of endogenous SPEN, the ability of each of these truncations to rescue defects in X-linked gene silencing is assayed using pyrosequencing. All these truncations are fused to a 3xFLAG epitope at their N-terminus, allowing protein detection and localization relative to the X chromosome by immunofluorescence. An additional N-terminal SV-40 NLS ensures nuclear localization of all truncations.

The full-length SPEN cDNA was purchased from Genscript, and different annotated SPEN domains (RRM1, RRM2-4, RID and SPOC) were individually spliced out using overlap extension PCR in collaboration with Agnès Le Saux, a research scientist from the lab (see Dossin et al., Fig. 3a).

A 3xFLAG epitope was included at the N-terminus of each cDNA to allow protein detection (both by western blot and immunofluorescence). Furthermore, an "artificial" SV-40 *n*uclear *l*ocalization *s*ignal (NLS) was also added to compensate for the removal of any endogenous NLSs in the generated SPEN truncations (**Figure 6.9A**).

I then knocked-in each of these truncations at *Rosa26*[¶¶] in an endogenous SPEN-degron mESC background (**Figure 6.9B**), and confirmed their constitutive expression using western blot (see Dossin et al., Extended Data Fig. 3b).

For each truncation, 2 or 3 independent mESC clones were generated and the ability of each truncation to rescue *Xist*-mediated gene silencing upon depletion of endogenous SPEN was assayed using pyrosequencing of 4 X-linked genes (see Dossin et al., Fig. 3b). Furthermore, the localization of each SPEN truncation relative to the inactive X chromosome was assayed using immunofluorescence (H2AK119ub1 was used as a marker of the Xi, see Dossin et al., Fig. 3c).

[¶¶]Some of these knock-in "rescue" constructs were shared with the Gribnau lab in the context of a project investigating the function of SPEN in regulating the expression of *Xist* RNA. The results from this project have been shared on the bioRxiv preprint server **[Robert-Finestra et al., 2020]**, and are appended at **section A.2**.

6.3.2.2 SPEN's RRM2-4 and SPOC domains appear vital for XCI

This analysis revealed that the RRM1 and RID domains of SPEN are dispensable for SPEN recruitment to the Xi as well as for gene silencing (see Dossin et al., Fig. 3b, c).

Loss of RRM2-4 however resulted in failure to recruit SPEN to the Xi (see Dossin et al., Fig. 3b, c), a finding which is consistent with previous reports that RRM2-4 directly bind *Xist in vitro* [Monfort et al., 2015; Lu et al., 2016]. Furthermore, this truncation failed to rescue gene silencing upon loss of endogenous SPEN (see Dossin et al., Fig. 3b, c), further confirming that active *Xist* mediated recruitment of SPEN (through its RRM2-4 domains) is required for X inactivation.

Importantly, when SPEN was truncated of its SPOC domain, it retained its ability to accumulate around the Xi, but failed to rescue the gene silencing defects associated with loss of endogenous SPEN (see Dossin et al., Fig. 3b, c).

6.3.2.3 Endogenous SPOC-deletion confirms its essentiality in XCI

I further validated this observation by performing a homozygous deletion of the SPOC domain at the endogenous *Spen* locus – replacing it with a GFP – in TX1072 mESCs (**Figure 6.10**). Characterizaion of 3 independent $Spen^{\Delta SPOC}$-*GFP* clones confirmed that SPOC is dispensable for proper *Xist* localization and SPEN recruitment to the Xi (see Dossin et al., Extended Data Fig. 3d-f).

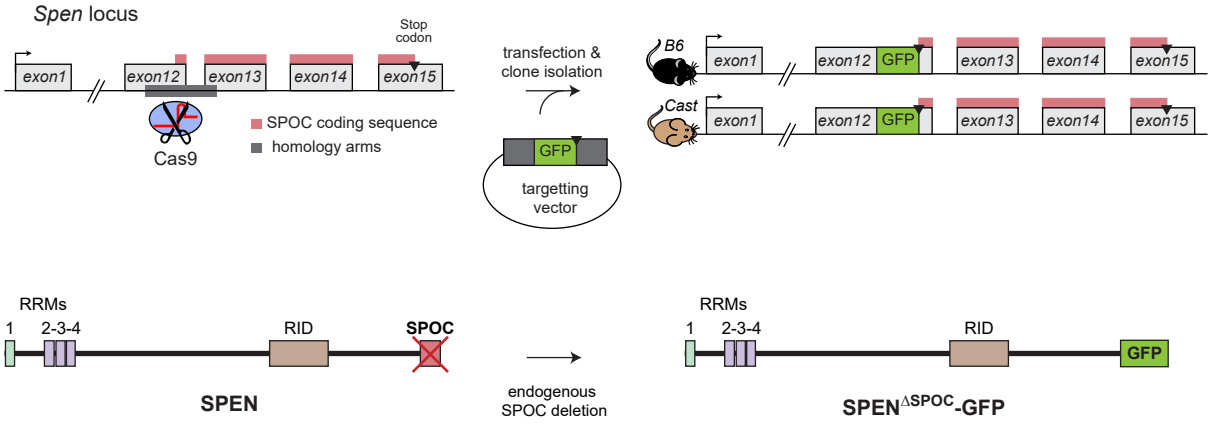


Figure 6.10 – Scheme depicting the genetic strategy to delete SPEN's SPOC domain endogenously and replace it with a GFP. A homozygous knock-in of a GFP cassette (ending with a Stop codon) was performed immediately upstream of the *Spen* sequence encoding its SPOC domain (shown as thin red rectangles on the *Spen* locus) using homology-directed repair (the homology sequence is shown as a thin dark grey rectangle).

Transcriptomic analysis following 24h of *Xist*-induction in $Spen^{\Delta SPOC}$-*GFP* cells re-

vealed dramatic defects in gene silencing, chromosome wide (see Dossin et al., Extended Data Fig. 3g-i), demonstrating that the SPOC domain is an essential effector of gene silencing during XCI, acting downstream of *Xist* localization.

Remarkably, the gene silencing defect was moderately weaker than the one observed upon loss of full-length SPEN (see Dossin et al., Extended Data Fig. 3j), suggesting that other uncharacterized regions of SPEN still contribute – albeit weakly – to *Xist*-mediated gene silencing (see **subsection 9.2.1** for further discussion).

6.3.2.4 SPOC tethering to the X chromosome is sufficient to mediate gene silencing

I next questioned whether the SPOC domain alone could be sufficient to trigger gene silencing along the X-chromosome. To do so, I made use of the Bgl stem-loop system (see **subsubsection 6.2.4.2**) to artificially tether a BglG-GFP-SPOC fusion protein (or BglG-GFP as control) to *Xist* RNA, in a SPEN-degron background (**Figure 6.11**).

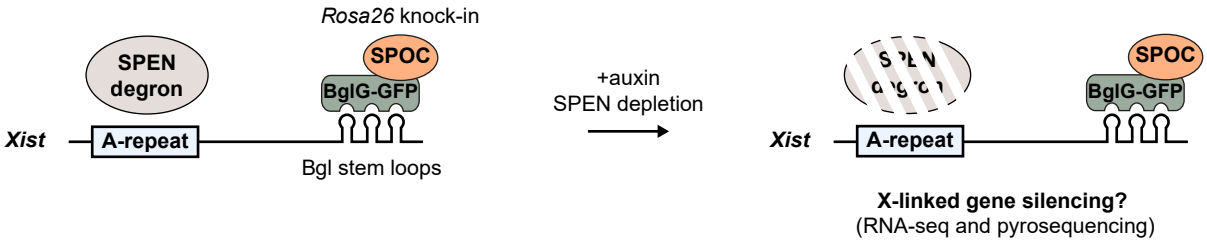


Figure 6.11 – Scheme of the SPOC tethering experiment. A BglG-GFP-SPOC fusion protein is targeted at and constitutively expressed from the *Rosa26* locus, in a SPEN-degron background. This fusion protein is "artificially" recruited to *Xist* RNA (and consequentially, the X-chromosome) through binding to Bgl stem-loops. Upon depletion of endogenous SPEN and upregulation of *Xist* RNA, the ability of SPOC to mediate gene silencing chromosome wide can be monitored using RNA-seq and pyrosequencing.

Remarkably, in the absence of endogenous SPEN, *Xist*-mediated recruitment of SPOC (but not GFP alone) leads to drastic silencing of most genes along the X-chromosome (see Dossin et al., Fig. 3d-g). This result suggests that once recruited to the X chromosome, SPOC is sufficient to instruct a significant fraction of gene silencing (see **subsection 9.2.1** for further discussion).

6.3.3 Identifying the protein partners of SPEN's SPOC domain

To gain further mechanistic insights into how SPOC instructs such gene silencing, I next set out to characterize its protein partners in mESCs.

6.3.3.1 Rationale for characterizing the SPOC-interactome

Knowing that:

1. SPEN's SPOC domain is a very potent interactor of NCoR and SMRT in humans [Shi et al., 2001; Ariyoshi and Schwabe, 2003; Oswald et al., 2016],
2. HDAC3-mediated histone deacetylation plays an important role in promoting efficient gene silencing during XCI in mESCs [Żylicz et al., 2019; McHugh et al., 2015],
3. HDAC3 alone is not functional, and its interaction with SMRT or NCoR – both of which directly interact with SPOC – is necessary for its enzymatic activity [Guenther et al., 2001],

it could at this stage be argued that the function of SPEN's SPOC domain during XCI is *solely* to recruit NCoR/SMRT to HDAC3-bound sites, thereby enabling HDAC3 activity on the X-chromosome, leading to subsequent chromosome-wide gene silencing. In fact, ever since the identification of SPEN as a potential key player in XCI, this model was put forth by the Guttman lab [McHugh et al., 2015].

However, evidence for a physical interaction between SPOC and NCoR/SMRT *in mouse* to support this hypothesis is missing. Furthermore, several lines of evidence demonstrate that invoking only an HDAC3-axis is not sufficient to fully explain how SPEN/SPOC functions during XCI:

1. Using the same cellular background (female TX1072 mESCs), and assaying XCI under identical conditions (24 hours of *Xist* upregulation), I find that the gene silencing defects observed following the endogenous deletion of SPEN's SPOC domain are much more severe than those reported upon the loss of endogenous HDAC3 (see Dossin et al., Extended Data Fig. 3j, k).
2. In fact, the nature of the gene silencing phenotypes associated with HDAC3 and SPOC perturbations appear to be different: In *Hdac3* KO cells, XCI is only delayed: most affected genes are still significantly silenced, only not as quickly as in a WT scenario, and in reality, many genes undergo silencing identically to WT cells [Żylicz et al., 2019], hinting that HDAC3 acts as a catalyst/facilitator rather than as a genuine mediator of *Xist*-dependent gene silencing. In the absence of SPOC on the other hand, virtually completely deficient silencing is observed across all X-linked genes.
3. HDAC3 predominantly acts at X-linked enhancers (see **subsection 5.3.2**), while I found that SPEN is actively recruited not only to enhancers but also to promoters during XCI (see **subsection 6.3.1**).

Taken together, these facts highlight the necessity to map the SPOC protein interactome (notably in mESCs) so as to identify other SPOC-associated cofactors and pathways involved in mediating XCI in parallel to HDAC3.

6.3.3.2 SPOC proteomics methodology

I hence performed *i*mmuno*p*recipitation (IP) of BglG-GFP-tagged SPOC, or BglG-GFP as a control, both stably integrated at (and constitutively expressed from) the *Rosa26* locus. Co-immunoprecipitating proteins were then identified using *m*ass *s*pectrometry (MS, see Dossin et al., Extended Data Fig. 3l). Each IP was performed in biological quadruplicates, using an anti-GFP nanobody-coupled affinity resin (GFP-trap), and key SPOC-partners were orthogonally validated using co-IP western blot (see Dossin et al., Fig. 3h).

Importantly, I performed these IPs in the *absence* of *Xist* induction, so as to warrant that the co-immunoprecipitated proteins are not indirectly bridged to SPOC *via Xist* RNA.

6.3.3.3 SPOC interacts with a variety of complexes involved in direct transcription regulation, RNA metabolism, chromatin modification and nucleosome remodeling

Remarkably, SPOC co-immunoprecipitated with several protein complexes involved in processes pertaining to regulation of gene expression including: transcription, RNA metabolism, chromatin modification and nucleosome remodeling.

SPOC and NCoR/SMRT-HDAC3:

Expectedly, most members of the NCoR/SMRT complex were retrieved (**Figure 6.12**, see Dossin et al., Fig. 3h and Extended Data Fig. 3l). Surprisingly, HDAC3 was found to immunoprecipitate with SPOC (**Figure 6.12**, see Dossin et al., Fig. 3h and Extended Data Fig. 3l), although we previously reported that contrary to SPEN, HDAC3 is not recruited to the X chromosome during XCI but is instead pre-bound to X-linked enhancers (see **chapter 5**). Hence, it is most likely that HDAC3 "saturates" chromatin on the X, leaving little room for it to be recruited. Consequently, SPEN recruitment constitutes the limiting step in the SPEN-HDAC3 axis during XCI.

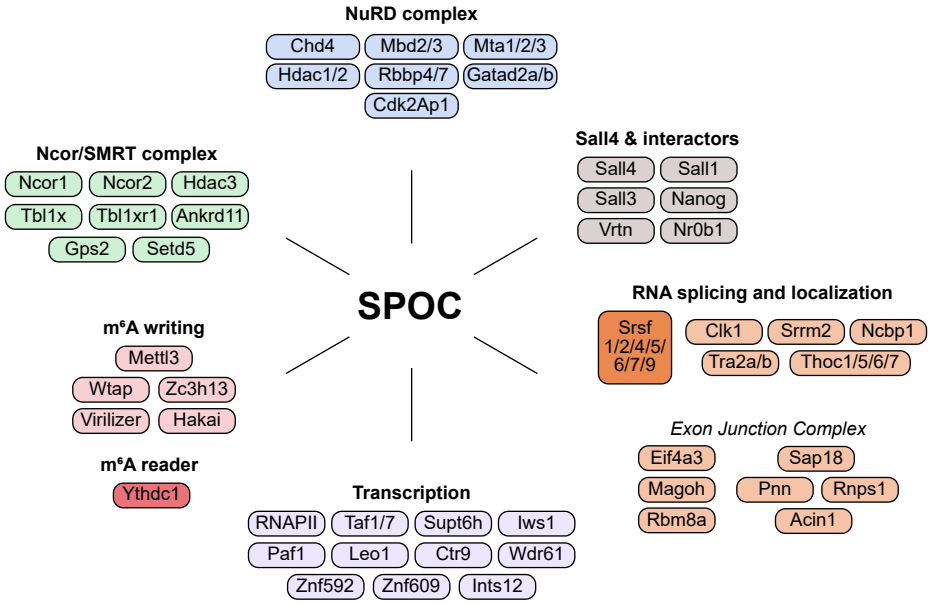


Figure 6.12 – Scheme showing the different proteins (and the complexes/biological pathways they belong to) identified in the SPOC interactome.

SPOC and RNA homeostasis/m^{6}A machinery:

Remarkably, several proteins involved in RNA splicing and localization – notably members of the exon junction complex – were identified in our CoIP-MS analysis (**Figure 6.12**, see Dossin et al., Fig. 3h and Extended Data Fig. 3l), hinting that SPEN could be important for regulating the fate of *Xist* RNA to some extent, as well as that of other X-linked genes.

Furthermore, all members of the m^{6}A RNA methylation machinery as well as the nuclear m^{6}A reader YTHDC1 (see **subsection 1.13.3**) were also identified (**Figure 6.12**, see Dossin et al., Fig. 3h and Extended Data Fig. 3l). These results raise the possibility that similarly to RBM15 (see **subsection 1.13.3**), SPEN constitutes a physical link between *Xist* and the RNA-methylation pathway. Furthermore, this analysis defines SPOC as an m^{6}A-writing machinery interaction domain (see **subsubsection 9.2.2.4** for further discussion of these results).

SPOC and the NuRD complex:

SPOC also co-immunoprecipitated with most members of the *nu*cleosome *r*emodeling *d*eacetylase complex (NuRD), a protein complex carrying both histone deacetylation and chromatin remodeling activities (**Figure 6.12**, see Dossin et al., Fig. 3h and Extended Data Fig. 3l). The SALL4 transcription factor (and many of its interactors) was also highly enriched in the SPOC Co-IP (**Figure 6.12**, see Dossin et al., Fig. 3h and Extended Data Fig. 3l), consistent with previous reports that a large fraction of the NuRD pool is complexed with SALL4 in mESCs [Miller et al., 2016; Kloet et al., 2015; Lu et al., 2009].

SPOC and RNAPII:

The NuRD complex acts as a potent transcriptional repressor, by displacing RNA polymerase II (RNAPII) from transcription start sites, specifically through its nucleosome remodeling activity [Bornelöv et al., 2018]. Therefore, SPEN could mediate gene silencing by affecting RNAPII through the NuRD complex, which will be further discussed in **subsubsection 9.2.2.2**. Remarkably, SPOC also co-immunoprecipitated with several factors directly involved in transcription, including RNAPII itself as well as transcription initiation and elongation factors (**Figure 6.12**, see Dossin et al., Fig. 3h and Extended Data Fig. 3l). Hence, it is possible that SPEN could also repress transcription directly through its interaction with these factors. Such interaction was investigated in the context of **chapter 7** and will be further discussed in **subsubsection 9.2.2.3**.

Integrating SPEN CUT&RUN with SPOC-interactors binding profiles:

The facts that SPOC interacts with other protein complexes than NCoR/SMRT, and that those are also involved in transcriptional regulation (particularly repression), strongly suggest that such other complexes constitute SPEN/SPOC-bridged silencing pathways, acting in parallel to the SPEN-HDAC3 axis during XCI (see **subsubsection 6.3.3.1**).

To confirm this hypothesis, I integrated the SPEN CUT&RUN data (see **subsection 6.3.1**) with ChIP-seq datasets for some of the protein factors identified as SPOC-interactors. These datasets include:

- HDAC3 ChIP-seq, performed in the same TX1072 female mESC background as the SPEN CUT&RUN, with cells cultured under identical 2i+LIF conditions (see **chapter 5** and [Żylicz et al., 2019]).
- CHD4 and MBD3 ChIP-seq (two core components of the NuRD complex), performed in male E14 mESCs, also cultured under 2i+LIF conditions [Bornelöv et al., 2018].
- RNAPII ChIP-seq, performed in male E14 mESCs, also cultured under 2i+LIF conditions [Bornelöv et al., 2018].

This analysis revealed that SPEN binding strongly overlaps that of HDAC3 at enhancers, but not at promoters (see Dossin et al., Extended Data Fig. 4h). Such observation is in line with our previous report that HDAC3 is bound predominantly at enhancers (and not promoters, [Żylicz et al., 2019]), and further supports a model wherein *Xist*-mediated recruitment of SPEN activates HDAC3-dependent deacetylation at X-linked enhancers.

On the contrary, SPEN binding strongly overlapped that of the CHD4/MBD3 NuRD

subunits at gene promoters, but not at enhancers (see Dossin et al., Extended Data Fig. 4h).

This dichotomy is in strong favor of a model wherein SPEN, through its SPOC domain, bridges different pathways (operating distinctively at promoters or enhancers) to robustly silence X-linked transcription during XCI.

Towards a SPOC/RNAPII interaction research axis:

Furthermore, SPEN profiles overlap majorly with RNAPII-bound chromatin sites on the X chromosome (see Dossin et al., Extended Data Fig. 4h). This result is consistent with the observation that SPEN is recruited predominantly to actively transcribed gene promoters and enhancers, and that SPEN-binding to chromatin appears to be transcription-dependent (see **subsection 6.3.1**). Most importantly, they corroborate the preliminary evidence that SPOC interacts with the transcription machinery, through which it could be envisionned that SPEN represses transcription directly during XCI.

In the next chapter (**chapter 7**), I will describe the biochemical and biophysical experiments I undertook in collaboration with Brice Murciano, from Christoph Müller's lab at EMBL Heidelberg, to further explore the molecular basis of a possible interaction between SPOC and the RNAPII transcription machinery.

6.4 Article 2: SPEN integrates transcriptional and epigenetic control of X-inactivation *(Nature. 2020)*

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

ARTICLE ON NEXT PAGE

Article

SPEN integrates transcriptional and epigenetic control of X-inactivation

https://doi.org/10.1038/s41586-020-1974-9

Received: 4 June 2019

Accepted: 10 January 2020

Published online: 5 February 2020

François Dossin[1], Inês Pinheiro[2,8], Jan J. Żylicz[2,3,8], Julia Roensch[2], Samuel Collombet[1], Agnès Le Saux[2], Tomasz Chelmicki[2], Mikaël Attia[2], Varun Kapoor[2], Ye Zhan[4], Florent Dingli[5], Damarys Loew[5], Thomas Mercher[6], Job Dekker[4,7] & Edith Heard[1,2]*

Xist represents a paradigm for the function of long non-coding RNA in epigenetic regulation, although how it mediates X-chromosome inactivation (XCI) remains largely unexplained. Several proteins that bind to *Xist* RNA have recently been identified, including the transcriptional repressor SPEN[1–3], the loss of which has been associated with deficient XCI at multiple loci[2–6]. Here we show in mice that SPEN is a key orchestrator of XCI in vivo and we elucidate its mechanism of action. We show that SPEN is essential for initiating gene silencing on the X chromosome in preimplantation mouse embryos and in embryonic stem cells. SPEN is dispensable for maintenance of XCI in neural progenitors, although it significantly decreases the expression of genes that escape XCI. We show that SPEN is immediately recruited to the X chromosome upon the upregulation of *Xist*, and is targeted to enhancers and promoters of active genes. SPEN rapidly disengages from chromatin upon gene silencing, suggesting that active transcription is required to tether SPEN to chromatin. We define the SPOC domain as a major effector of the gene-silencing function of SPEN, and show that tethering SPOC to *Xist* RNA is sufficient to mediate gene silencing. We identify the protein partners of SPOC, including NCoR/SMRT, the m^{6}A RNA methylation machinery, the NuRD complex, RNA polymerase II and factors involved in the regulation of transcription initiation and elongation. We propose that SPEN acts as a molecular integrator for the initiation of XCI, bridging *Xist* RNA with the transcription machinery—as well as with nucleosome remodellers and histone deacetylases—at active enhancers and promoters.

To assess the importance of SPEN during the initiation of XCI, we used an auxin-inducible degron (AID)[7] that enables controlled and acute depletion of the endogenous SPEN protein. We used our previously described female hybrid (*Mus musculus castaneus* × C57BL/6) TX1072[8] mouse embryonic stem cells (ES cells), in which a doxycycline (DOX)-inducible promoter upstream of the endogenous *Xist* locus enables conditional *Xist* RNA expression and XCI (Fig. 1a). In ES cells expressing the *Oryza sativa* TIR1 (OsTIR1) E3 ligase, we generated a homozygous knock-in that expressed the AID fused to a HaloTag at the C terminus of endogenous SPEN, in order to ensure auxin-dependent SPEN depletion (Extended Data Fig. 1a). Efficient degradation of SPEN occurred within 1 h of auxin treatment (Fig. 1b, Extended Data Fig. 1b, Supplementary Fig. 1) whereas the removal of auxin led to rapid recovery of SPEN (Fig. 1b), demonstrating potent AID-dependent modulation of SPEN levels.

To evaluate the immediate consequences of the loss of SPEN on the initiation of XCI, we acutely depleted SPEN for 4 h before inducing *Xist* expression for 24 h and performing RNA sequencing. Loss of SPEN had no effect on the formation of *Xist* RNA clouds (Extended Data Fig. 1c, e), confirming that SPEN is dispensable for *Xist* localization[2–5]. However, gene silencing was almost completely abolished along the entire X chromosome in the absence of SPEN (Fig. 1c, d, Supplementary Table 1), whereas auxin had no effect on XCI in wild-type cells (Extended Data Fig. 1d). Clustering analysis highlighted three groups of genes that differed in their silencing defects upon the loss of SPEN (Fig. 1e). Most X-linked genes (80% of 382) were found to be entirely dependent on SPEN for silencing, whereas only a small subset (6%) showed unaltered silencing in the absence of SPEN. This notable defect in XCI was confirmed by pyrosequencing (Fig. 1f) and nascent RNA fluorescence in situ hybridization (FISH) (Extended Data Fig. 1e).

We next assessed the requirement for SPEN in XCI in vivo during mouse early embryogenesis, using allele-specific RNA sequencing in embryonic day (E)3.5 *Spen*-knockout female embryos[9] harbouring hybrid X chromosomes (Fig. 1g, Extended Data Fig. 1f, g). At this stage in wild-type embryos, imprinted XCI has taken place[10] and only the paternal X chromosome is inactivated (Fig. 1h, Extended Data Fig. 1h).

[1]European Molecular Biology Laboratory, Director's Unit, Heidelberg, Germany. [2]Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, UPMC Paris-Sorbonne, Paris, France. [3]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. [4]Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA. [5]Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, Paris, France. [6]INSERM U1170, Gustave Roussy Institute, Université Paris-Sud, Villejuif, France. [7]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [8]These authors contributed equally: Inês Pinheiro, Jan J. Żylicz. *e-mail: edith.heard@embl.org

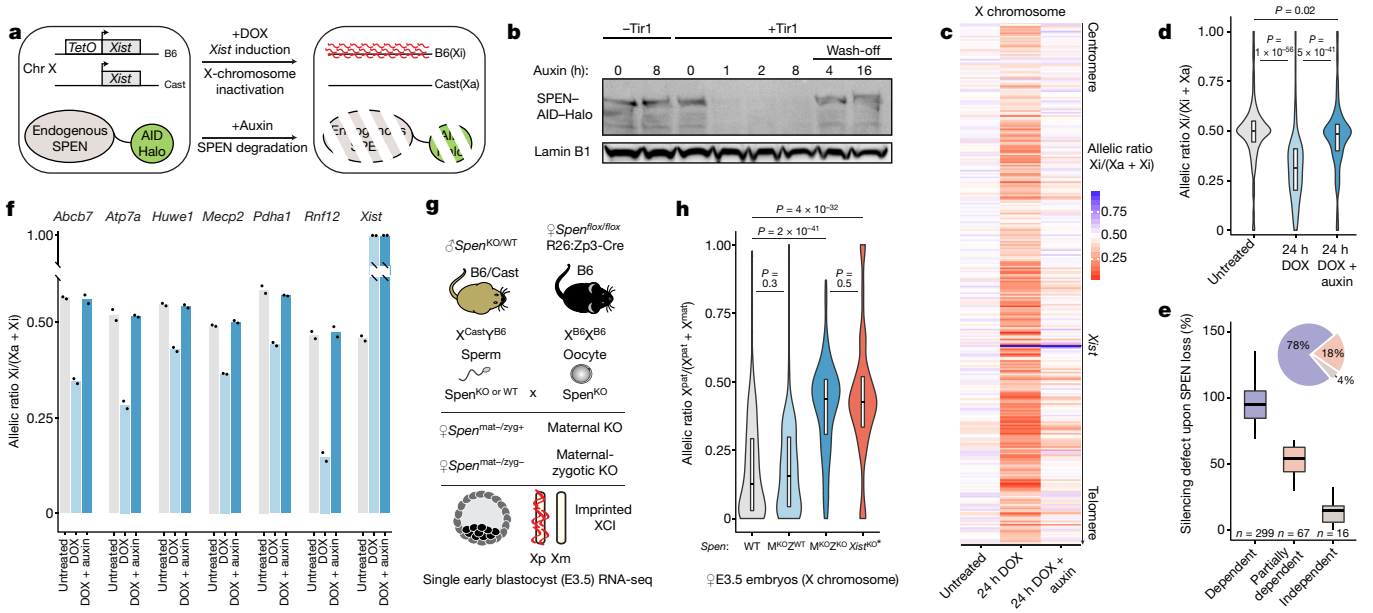


Fig. 1 | SPEN mediates gene silencing across the entire X chromosome in vitro and in vivo. a, Schematic of SPEN-degron *Xist*-inducible mouse ES cells. Xa, active X chromosome; Xi, inactive X chromosome. **b**, Western blot showing auxin-induced degradation of endogenous HaloTagged SPEN. This experiment was repeated at least twice with similar results. **c**, **d**, Heat map (**c**) and violin plots (**d**) showing X-chromosomal transcript allelic ratios after 0 h, 24 h DOX or 24 h DOX + auxin treatment in SPEN-degron mouse ES cells ($n = 434$ genes, two-sided Student's *t*-test). **e**, Box plot representation of gene-silencing defect upon SPEN loss in three groups of genes differing by their level of dependence on SPEN for *Xist*-mediated silencing. The pie chart shows the relative number of genes in each group. **f**, Pyrosequencing assay of seven X-linked transcripts in mouse ES cells after 0 h, 24 h DOX or 24 h DOX + auxin treatment. Data in **c**–**f** are averages of two independent clones; in **f**, individual data points are shown. **g**, The mouse crossbreeding scheme for the *Spen*-knockout experiment. KO, knockout; WT, wild type; mat., maternal; zyg., zygotic; Xm, maternal X chromosome; Xp, paternal X chromosome. **h**, X-chromosomal transcript allelic ratio distribution ($n = 256$ genes) in wild-type ($n = 2$), maternal-only (M) *Spen*-knockout ($n = 3$), maternal-zygotic (Z) *Spen*-knockout ($n = 5$), and *Xist*-knockout E3.5 embryos ($n = 30$ single cells, two-sided Wilcoxon rank-sum test. For * refer to ref. [10]). In **d**, **e**, **h**, horizontal lines denote the median, box limits correspond to the upper and lower quartiles.

In maternal-zygotic *Spen* knockouts, imprinted XCI is severely hindered although paternal *Xist* is expressed. Both maternal and paternal X chromosomes are expressed equally, phenocopying *Xist*-knockout E3.5 embryos[10] (Fig. 1h, Extended Data Fig. 1g, h, Supplementary Table 2). A maternal-only *Spen* knockout has no effect on imprinted XCI (Fig. 1h), suggesting that the zygotic pool of SPEN is necessary and sufficient for this process. Therefore, the early gene-silencing mechanism(s) involved in imprinted and random XCI are dependent on SPEN.

We next assessed precisely when SPEN is recruited during XCI. HaloTag labelling[11] of SPEN combined with *Xist* RNA FISH revealed that SPEN associates with *Xist* RNA rapidly upon *Xist* coating and throughout XCI (Fig. 2a). To capture early *Xist*–SPEN dynamics during the short time window in which *Xist* becomes upregulated, we followed both *Xist* and SPEN in living cells. We tagged endogenous SPEN with GFP in a background in which *Xist* RNA is visualized via a BglG–mCherry fusion protein binding to Bgl stem–loops inserted within *Xist*[12] (Extended Data Fig. 2a, b). Live-cell imaging revealed that SPEN colocalizes with *Xist* from the very onset of *Xist* upregulation (Extended Data Fig. 2c, d, Supplementary Video 1). Therefore, SPEN can initiate gene silencing immediately upon *Xist* coating.

We also found that SPEN robustly accumulated on the inactive X chromosome after differentiation into neural progenitor cells (NPCs, Fig. 2b), in which XCI is epigenetically maintained. The depletion of SPEN for up to two days in independent NPC clones (Fig. 2c) did not lead to reactivation of fully silenced genes (Fig. 2d, Supplementary Table 3); however, we observed moderate but significant upregulation of genes escaping XCI (Fig. 2e, f), which suggests that SPEN buffers the overexpression of X-linked escapee genes in female cells.

Chromosome conformation capture has revealed that, in differentiated cells, the inactive X chromosome is folded into megadomains[13–15] and is globally depleted of topologically associating domains except in regions that contain clusters of escapee genes[13]. *Xist* RNA has been found to have a role in the conformation of the inactive X chromosome[1,14]. To assess whether SPEN is involved, we performed allele-specific Hi-C in NPCs after 48 h of SPEN depletion. No notable conformational changes were observed on the inactive X chromosome (Extended Data Fig. 2e–g); we therefore conclude that the structural effects mediated by *Xist* RNA in differentiated cells occur independently of SPEN.

In summary, our data suggest that SPEN exerts its role by actively promoting gene silencing during the earliest stages of XCI. However, it has no major role in stabilizing the transcriptionally inactive state of the inactive X chromosome, or in ensuring the maintenance of its conformation.

We next sought to identify which parts of SPEN ensure its function during XCI. SPEN is a very large protein (around 400 kDa) that contains four RNA recognition motifs (RRMs), a nuclear receptor interaction domain (RID) and a SPEN paralogue/orthologue C-terminal (SPOC) domain (Fig. 3a). We overexpressed a series of SPEN complementary DNA truncations, stably targeted into the *Rosa26* locus in the SPEN-degron mouse ES cell line (Extended Data Fig. 3a, b, Fig. 3a). We then induced *Xist* expression for 24 h and assessed which SPEN fragments could rescue XCI-initiation function in the context of auxin-mediated depletion of endogenous SPEN. We found that the RRM1 domain and the RID are dispensable for SPEN accumulation on the inactive X chromosome, as well as for X-linked gene silencing (Fig. 3b, c). By contrast, a SPEN truncation lacking the RRM2–4 domains failed to accumulate on the inactive X chromosome and failed to rescue XCI (Fig. 3b, c). SPEN recruitment to the inactive X chromosome is therefore mediated by the RRM2–4 domains and is necessary for gene silencing. This is consistent with studies showing that RRM2–4 directly bind the A-repeat of *Xist*

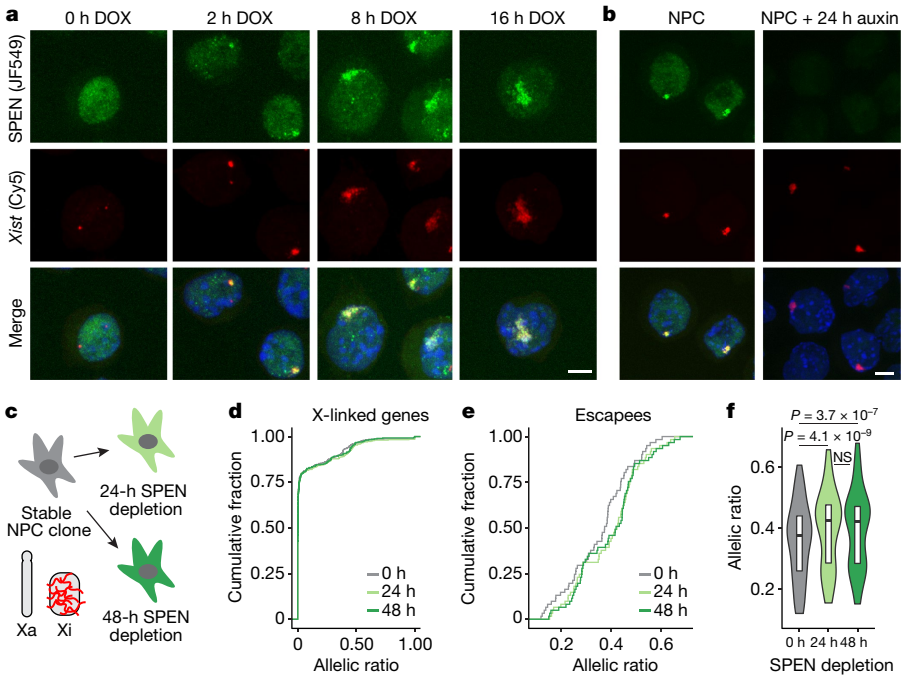


Fig. 2 | SPEN localizes to the X chromosome immediately upon *Xist* upregulation and throughout the stages of XCI, but is dispensable for the maintenance of X-linked gene silencing. a, **b**, Images from combined HaloTag labelling of SPEN (green) and FISH for *Xist* RNA (red) in mouse ES cells during a time course of *Xist* induction (**a**) and in NPCs (**b**). Scale bars, 5 µm. **c**, Schematic of the SPEN-degron experiment in NPCs. **d**, Cumulative distribution of transcript allelic ratios across the X chromosome ($n = 387$ genes) after SPEN depletion in NPCs. **e**, **f**, Cumulative distribution (**e**) and violin plot representation (**f**) of the transcript allelic ratio of escapees after SPEN depletion in NPCs ($n = 65$, two-sided Wilcoxon signed-rank test. NS, not significant. Horizontal lines denote the median, box limits correspond to upper and lower quartiles). Data in **d**–**f** are the average of two independent NPC clones. The experiments in **a**, **b** were repeated at least twice with similar results.

RNA in vitro[4,16]—a region of *Xist* that is necessary for gene silencing[17]. Conversely, a truncation of the SPOC domain enabled efficient SPEN accumulation on the inactive X chromosome, but failed to rescue XCI (Fig. 3b, c). To validate this observation, we performed homozygous deletion of the SPOC domain at the endogenous *Spen* locus in mouse ES cells (Extended Data Fig. 3c). Deletion of the SPOC domain had no effect either on SPEN recruitment to the inactive X chromosome or on *Xist* RNA clouds (Extended Data Fig. 3d–f), but resulted in strongly deficient XCI, albeit milder than that in SPEN-depleted cells (Extended Data Fig. 3g–j). Collectively, these results demonstrate that the SPOC domain is essential for XCI. However, other uncharacterized regions of SPEN contribute—albeit to a lesser extent—to ensure its full silencing potential.

To test whether the SPOC domain alone could mediate X-linked gene silencing, we used SPEN-degron ES cells to introduce an array of BglI stem–loops at the *Xist* locus (identical to the live-imaging strategy). In this background, we generated several independent ES cell lines expressing a BglG–GFP–SPOC protein fusion (or BglG–GFP as a control) targeted into *Rosa26*. These proteins would become tethered to *Xist*–BglI stem–loop RNA via BglG (Fig. 3d). Notably, upon induction of *Xist* RNA in the absence of endogenous SPEN, tethering of BglG–GFP–SPOC (but not of BglG–GFP alone) resulted in substantial gene silencing across the X chromosome, with over half of the genes being silenced by more than 50% (Fig. 3e, f). SPOC-specific rescue was confirmed using pyrosequencing (Fig. 3g). Consistent with previous studies[18–20], our results reveal SPOC as a key domain of SPEN that enables gene silencing once recruited to the X chromosome by *Xist* RNA.

The SPOC domain of SPEN was originally identified as an interactor of the NCoR and SMRT corepressors in human cells[18,21,22]. Given that NCoR and SMRT interact with and activate HDAC3[23], it was proposed that SPEN triggers XCI via HDAC3[2], the activity of which is important for *Xist*-mediated silencing[2,24]. However, XCI is more markedly affected upon the loss of SPEN and SPOC than upon the loss of HDAC3 (Extended Data Fig. 3j, k). These observations suggest that a model involving HDAC3 only partially explains the function of SPEN, and that SPOC must exert its key role in gene silencing also through other, HDAC3-independent pathways. To identify such pathways, we characterized the protein interactome of the SPOC domain by performing GFP pull-downs from mouse ES cells that stably expressed BglG–GFP–SPOC (or BglG–GFP as a control, Fig. 3h, Supplementary Table 4), followed by mass spectrometry analysis.

We identified NCoR and SMRT as expected, but we also found HDAC3 (Fig. 3h, Extended Data Fig. 3l), which further supports the proposed model for the function of SPEN in XCI[2]. Notably, we identified the m^{6}A methyltransferase complex and the m^{6}A reader YTHDC1 (Fig. 3h, Extended Data Fig. 3l), which have been proposed to play a role in XCI[5,20,25]. One of these factors, WTAP, co-purified with *Xist* RNA in an A-repeat-dependent manner[3]—although, contrary to the case of SPEN, a direct interaction between WTAP and *Xist* A-repeat has not been reported. Our results therefore suggest that SPOC may participate in the recruitment of m^{6}A machinery to *Xist* RNA. We also identified the NuRD complex—a potent repressor that displaces RNA polymerase II (RNAPII) from transcription start sites through chromatin remodelling[26]—and RNAPII, together with factors that are involved in the regulation of transcription initiation and elongation (Fig. 3h, Extended Data Fig. 3l). Together these findings show that, through its SPOC domain, SPEN bridges *Xist* to multiple factors that are involved in transcription and chromatin regulation, and together they mediate efficient gene silencing. Given that SPOC immunoprecipitation was performed in the absence of *Xist* induction, the identified interactions are not mediated by *Xist* RNA.

We also investigated where SPEN binds to the X chromosome during XCI, and whether it has distinct binding sites or whether it associates with chromatin diffusely across the entire chromosome, as anticipated

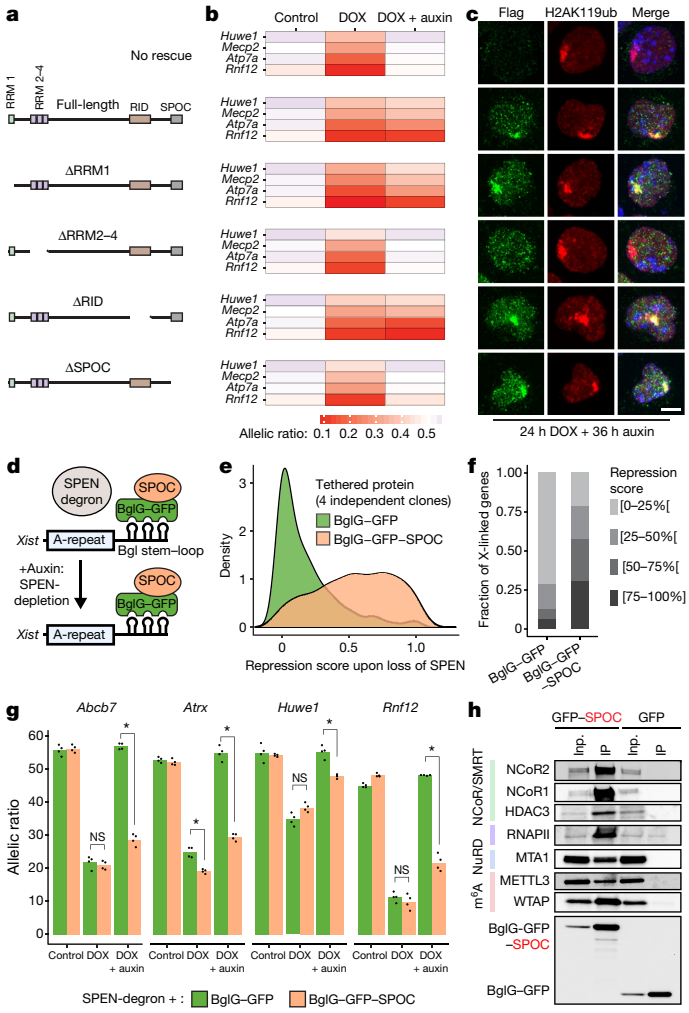


Fig. 3 | The SPOC domain of SPEN mediates gene silencing and interacts with multiple molecular pathways. a, *Spen* cDNA fragments used for the rescue experiment. **b**, Heat map representation of four X-linked transcript allelic ratios (obtained by pyrosequencing) in control, 24 h DOX- and 24 h DOX + 36 h auxin-treated SPEN-degron mouse ES cells overexpressing each cDNA construct. Data represent averages of two to three independent clones. **c**, Immunofluorescence detection of Flag-tagged SPEN truncations (green) and H2AK119ub1 (red), a marker of the inactive X chromosome, in SPEN-degron mouse ES cells treated with DOX and auxin. Scale bar, 5 µm. **d**, Schematic showing the tethering of BglG–SPOC to *Xist*. **e**, Distribution of gene-repression scores observed across the X chromosome upon the depletion of endogenous SPEN and the tethering of BglG–GFP (green) or BglG–GFP–SPOC (orange) to *Xist*. **f**, Bar graphs showing the fraction of X-linked genes within four windows of repression score. **g**, Transcript allelic ratio (obtained by pyrosequencing) for four X-linked genes upon the depletion of endogenous SPEN and the tethering of BglG–GFP or BglG–GFP–SPOC to *Xist* (*$P < 0.01$, two-sided Student's *t*-test). **h**, Western blot showing co-immunoprecipitated proteins in BglG–GFP and BglG–GFP–SPOC immunoprecipitation experiments. One per cent of the input was loaded (0.1% for RNAPII), and 10% of the pull-down. The experiments in **c**, **h** were repeated at least twice with similar results. The data in **e**–**g** are the average of four independent clones.

from our imaging results. We performed allele-specific, cross-linked CUT&RUN[27] experiments on SPEN during a time course of *Xist* induction (0 h, 4 h, 8 h, 24 h DOX, or 8 h DOX + auxin as a negative control).

We found that there are few binding sites for SPEN across the genome of uninduced ES cells (Extended Data Fig. 4a). Conversely, hundreds of SPEN-binding sites appeared specifically on the X chromosome as early as 4 h after *Xist* induction (Fig. 4a, Extended Data Fig. 4a). This is consistent with imaging data (Extended Data Fig. 2). We note that SPEN accumulation is seen across the gene body of *Xist* (Fig. 4b), suggesting that SPEN binds *Xist* RNA while it is transcribed. In sharp contrast to the *Xist* locus, SPEN shows focal binding on the rest of the genome, with peaks falling almost exclusively on promoters and enhancers (Fig. 4c, d, Extended Data Figs. 4b, 5a–g).

After *Xist* induction, recruitment of SPEN to the inactive X chromosome reaches a maximum at 4 h (Fig. 4a, Extended Data Fig. 4c), showing the highest enrichment within regions that were coated earliest by *Xist*[28] (entry sites, Fig. 4e). SPEN accumulation thus follows the spatial dynamics of *Xist* spreading. Among promoter targets on the X chromosome, SPEN preferentially binds those of actively expressed genes (Fig. 4f, Extended Data Fig. 4d), which suggests that the ability of SPEN to target chromatin depends on transcriptional activity. Consistently, genes that are classified as fully dependent on SPEN for silencing (Fig. 1e)—which show a greater degree of SPEN binding at their promoters within 4 h of *Xist* coating than less-dependent genes (Extended Data Fig. 4e)—also show initially higher transcription levels (Fig. 4g).

Furthermore, within 4 h of *Xist* induction, SPEN binding is greater at the promoters of efficiently silenced genes than at the promoters of less-efficiently silenced genes (Fig. 4h). Similarly, upon *Xist* coating, efficiently deacetylated enhancers[24] show a higher enrichment of SPEN than less-efficiently deacetylated enhancers (Fig. 4i). Finally, genes that are subject to very little silencing—or those that completely escape XCI in our *Xist*-inducible system—show a significantly lower SPEN signal at their promoters (Extended Data Figs. 4f, g, 5h–n). This pattern of SPEN recruitment at discrete sites to the X chromosome that is undergoing XCI indicates that transcriptional silencing is caused by the binding of SPEN to active promoters and enhancers.

To understand how SPEN might function at enhancers and promoters, we integrated CUT&RUN profiles with publicly available data from chromatin immunoprecipitation followed by sequencing (ChIP–seq) experiments for transcription and chromatin-associated factors identified in our mass spectrometry analysis. We included HDAC3[24], RNAPII[26] and two members of the NuRD complex (MBD3 and CHD4)[26]. SPEN binding strongly overlaps with HDAC3 at enhancers but not at promoters (Extended Data Fig. 4h). Our recent findings revealed that HDAC3 is pre-bound predominantly at enhancers on the X chromosome[24]. Therefore, *Xist*-mediated recruitment of SPEN to enhancers may activate HDAC3. Conversely, a strong overlap with SPEN binding is observed for the NuRD complex specifically at promoters but not at enhancers (Extended Data Fig. 4h). Furthermore, SPEN peaks extensively overlap with RNAPII phosphorylated on serine 5, which is associated with transcription initiation (Extended Data Fig. 4h). This analysis suggests that SPEN may operate at enhancers and promoters through distinct pathways to promote gene silencing.

Notably, the binding of SPEN to chromatin decreases across the whole X chromosome after 24 h of *Xist* induction (Fig. 4a, Extended Data Fig. 4c). Clustering of CUT&RUN profiles at SPEN-bound promoters (Extended Data Fig. 4i, Supplementary Table 5) revealed distinct groups of promoters, grouped on the basis of how efficiently SPEN was lost within 24 h of XCI (Fig. 4j). In the 'strong SPEN loss' group, binding was maximal by 4 h but decreased after 8 h, and even more markedly after 24 h (Fig. 4j). Conversely, the 'mild SPEN loss' group showed maximal and persistent SPEN binding at 4 h and 8 h of *Xist* induction, respectively, with only a mild reduction of SPEN binding by 24 h (Fig. 4j). Finally, a third group—comprising fewer promoters—showed both mild SPEN enrichment at 4 h and low loss at 24 h (Fig. 4j). The group that lost SPEN most efficiently also showed the most pronounced gene silencing by 24 h when compared with the groups that significantly retained SPEN (Fig. 4k, Extended Data Fig. 4j). Altogether this analysis suggests that, once recruited to the X chromosome by *Xist* RNA, SPEN associates with enhancers and promoters in a transcription-dependent manner. This recruitment leads to gene silencing, after which the favourable transcriptional context for SPEN binding is lost, and SPEN binding to chromatin decreases. Despite loss of the chromatin-bound SPEN

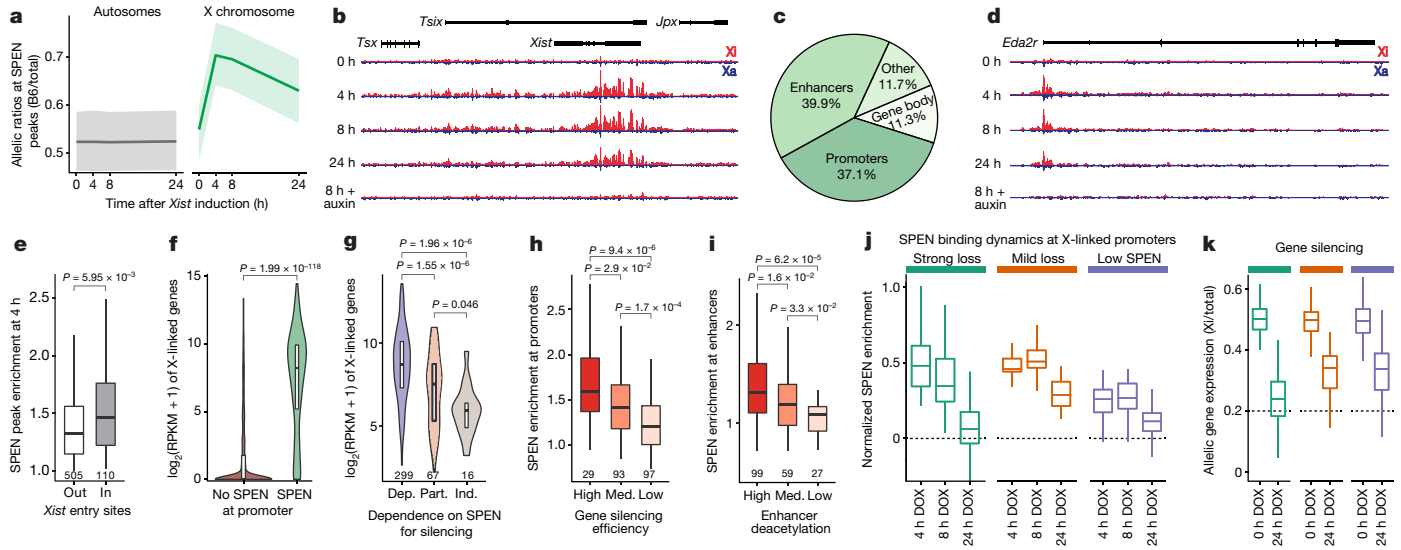


Fig. 4 | SPEN is recruited by *Xist* to active gene promoters and enhancers, where it silences transcription and subsequently disengages from chromatin. a, SPEN allele-specific accumulation (obtained from CUT&RUN experiments) on peaks at autosomes (grey, $n = 948$) and on the X chromosome (green, $n = 635$) after 0 h, 4 h, 8 h and 24 h of *Xist* induction in mouse ES cells. Shown are average allelic-ratios (shading is the interquartile range) of all peaks. **b**, UCSC Genome Browser allele-specific track showing SPEN binding around *Xist*. **c**, Annotation of SPEN peaks on the X chromosome. **d**, UCSC Genome Browser allele-specific track showing SPEN binding around *Eda2r*, an X-linked gene. In **b**, **d**, blue denotes Cast-Xa; red denotes B6-Xi; tracks are scaled identically. **e**, Box plot showing SPEN enrichment at 4 h in peaks outside or within *Xist* entry sites. **f**, Violin plot showing gene expression (reads per kilobase per million reads, RPKM) of genes accumulating SPEN ($n = 289$) or not accumulating SPEN ($n = 2,325$) at their promoters. **g**, Violin plot showing gene expression levels (RPKM in control conditions) of genes grouped on the basis of their level of dependence on SPEN for gene silencing (see Fig. 1e). **h**, **i**, Box plots showing SPEN enrichment after 4 h of *Xist* induction within peaks at promoters grouped on the basis of how efficiently their respective genes are silenced (**h**) or at enhancers grouped on the basis of how efficiently they are deacetylated during XCI (**i**). **j**, **k**, Box plots showing normalized SPEN enrichment at promoters (**j**) and gene silencing (transcript allelic ratio) during XCI (**k**) within 3 groups of X-linked genes showing different dynamics of SPEN accumulation and loss ($n = 86$ strong loss, $n = 92$ mild loss, $n = 39$ low SPEN). In **e**–**i**, the two-sided Wilcoxon rank-sum test was used; in **e**–**k**, horizontal lines denote the median, box limits correspond to upper and lower quartiles.

fraction, persistent *Xist* RNA expression and coating ensure that SPEN remains strongly accumulated around the inactive X chromosome (Fig. 2a, b).

Our study demonstrates that SPEN is a crucial factor that collaborates with *Xist* RNA to initiate gene silencing across the X chromosome, both during XCI in vitro and imprinted XCI in vivo. SPEN becomes dispensable for maintaining gene silencing after XCI has been established, but partially represses escapees, which suggests that *Xist* may have a silencing role even in somatic cells. Although SPEN coats the X chromosome immediately upon *Xist* induction, it contacts chromatin only at active promoters and enhancers, which serve as substrates for SPEN-mediated gene silencing. SPEN association with chromatin is favoured by active transcription, as SPEN disengages from chromatin when X-linked genes become silenced. We identify the SPOC domain of SPEN as a potent transcriptional repressor, which is crucial for SPEN-dependent XCI. On the basis of our mass spectrometry analysis, we propose that the SPOC domain is key for bridging *Xist* with other factors implicated in XCI—such as HDAC3—which we find to be present at most X-linked enhancers to which SPEN is recruited. In particular, the interaction of the SPOC domain with the NuRD complex and the transcription machinery points to a role for SPEN in direct transcriptional repression. We also identify SPOC as an interactor of the m^{6}A methyltransferase complex, which has a role in *Xist* RNA methylation, a modification that is important for *Xist*-dependent silencing[25]. Methylation of *Xist* is mediated by RBM15[25], which interacts with the m^{6}A machinery directly through ZC3H13[29]—the most highly enriched m^{6}A machinery factor identified in our mass spectrometry experiments. Because RBM15 also carries a SPOC domain, our study raises the possibility that the interaction with the RNA methylation machinery is not restricted solely to the SPOC domain of SPEN, but may instead be a feature that is shared across SPOC-containing proteins.

SPEN binds other non-coding RNAs, including *SRA*[18], which is involved in steroid-receptor regulation. Furthermore, another *SRA*-binding protein—SLIRP—has been shown to bind promoters in an *SRA*-dependent manner[30]; this raises the possibility that, similarly to *Xist*, *SRA* could guide SPEN to target gene regulatory elements.

In conclusion, our study suggests that RNA-mediated recruitment of SPEN and other SPOC-containing proteins—which are found across fungi, plants and animals—may be a widespread means by which to acutely repress transcription by co-ordinately engaging several layers of epigenetic and transcriptional control. We propose that SPEN bridges *Xist* to the transcription machinery, histone deacetylases and chromosome remodelling factors to ensure robust and efficient XCI (Extended Data Fig. 4k).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-1974-9.

1. Minajigi, A. et al. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).
2. McHugh, C. A. et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
3. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
4. Monfort, A. et al. Identification of *Spen* as a crucial factor for *Xist* function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554–561 (2015).
5. Moindrot, B. et al. A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.* **12**, 562–572 (2015).
6. Nesterova, T. B. et al. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat. Commun.* **10**, 3129 (2019).

7. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).
8. Schulz, E. G. 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).
9. Yabe, D. et al. Generation of a conditional knockout allele for mammalian Spen protein Mint/SHARP. *Genesis* **45**, 300–306 (2007).
10. Borensztein, M. et al. *Xist*-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nat. Struct. Mol. Biol.* **24**, 226–233 (2017).
11. Grimm, J. B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250 (2015).
12. Masui, O., Heard, E. & Koseki, H. in *X-Chromosome Inactivation* (ed. Sado, T.) *Methods Mol. Biol*. Vol. 1861, 67–72 (Humana, 2018).
13. Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575–579 (2016).
14. Deng, X. et al. Bipartite structure of the inactive mouse X chromosome. *Genome Biol*. **16**, 152 (2015).
15. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
16. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279 (2016).
17. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.* **30**, 167–174 (2002).
18. Shi, Y. et al. Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev*. **15**, 1140–1151 (2001).
19. Oswald, F. et al. RBP-Jκ/SHARP recruits CtIP/CtBP corepressors to silence Notch target genes. *Mol. Cell. Biol*. **25**, 10379–10390 (2005).
20. Ha, N. et al. Live-cell imaging and functional dissection of *Xist* RNA reveal mechanisms of X chromosome inactivation and reactivation. *iScience* **8**, 1–14 (2018).
21. Ariyoshi, M. & Schwabe, J. W. R. A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes Dev*. **17**, 1909–1920 (2003).
22. Oswald, F. et al. A phospho-dependent mechanism involving NCoR and KMT2D controls a permissive chromatin state at Notch target genes. *Nucleic Acids Res*. **44**, 4703–4720 (2016).
23. Guenther, M. G., Barak, O. & Lazar, M. A. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol. Cell. Biol*. **21**, 6091–6101 (2001).
24. Żylicz, J. J. et al. The implication of early chromatin changes in X chromosome inactivation. *Cell* **176**, 182–197 (2019).
25. Patil, D. P. et al. m^{6}A RNA methylation promotes *XIST*-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
26. Bornelöv, S. et al. The nucleosome remodeling and deacetylation complex modulates chromatin structure at sites of active transcription to fine-tune gene expression. *Mol. Cell* **71**, 56–72 (2018).
27. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
28. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
29. Knuckles, P. et al. Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m^{6}A machinery component Wtap/Fl(2)d. *Genes Dev*. **32**, 415–429 (2018).
30. Hatchell, E. C. et al. SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. *Mol. Cell* **22**, 657–668 (2006).

Methods

Data reporting and statistical analysis

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. All statistical tests, resulting *P* values and observation numbers are indicated in the figure panels or in the figure legends.

Data visualization

All heat maps, violin plots, box plots, density plots, bar graphs and pie charts were generated using ggplot2. Unless stated otherwise, box plots always show the median as the centre line, box limits correspond to upper and lower quartiles, and whiskers cover 1.5× the interquartile range.

Plasmid construction

The plasmids to target OsTIR1 at the *TIGRE* locus (Addgene plasmid 92141) and the *TIGRE*-specific guide-RNA-encoding plasmid (Addgene plasmid 92144) were provided by E. Nora. The additional *TIGRE*-targeting plasmids BglG-mCherry-T2A-OsTir1 (pFD51) and rtTa-VP16-T2A-OsTir1 (pFD68) were cloned using PCR amplification of corresponding gene cassettes followed by traditional cloning into the 92141 backbone.

Targeting constructs (pFD19 and pFD49) to tag endogenous SPEN at its C terminus with AID–HaloTag and AID–GFP, respectively, were generated as follows: 500-bp homology arms (flanking both sides of, but excluding the stop codon of *Spen*) were amplified from mouse genomic DNA by PCR. One-step Gibson cloning (New England Biolabs) was subsequently used to simultaneously surround the digested AID insert (carrying a puromycin-resistance gene under the control of the PGK promoter) in frame with the homology arms and clone the insert into the pBR322 vector. Synonymous mutations in the PAM/SEED target sequence (located on the 5′ homology arm) were then introduced using the QuickChange II XL site-directed mutagenesis kit (Agilent) to prevent Cas9-mediated cutting of the targeting vector upon transfection and of the AID-tagged allele(s) upon integration. The targeting construct (pFD90) to replace the endogenous SPOC domain of SPEN by GFP was generated using the same strategy. For guide RNA cloning, the pX459 plasmid (a gift from F. Zhang, Addgene 62988) encoding *Streptococcus pyogenes* Cas9 was digested with BbsI immediately downstream of the U6 promoter, and annealed DNA duplexes corresponding to the target guide RNA sequences were ligated.

Cell culture

Mouse XX ES cells (TX1072) were grown on 0.1% gelatin-coated flasks in 8% CO_2 37 °C incubators. For all experiments, cells were cultured in 2i + LIF, and batch-tested fetal calf serum ES cell medium – DMEM (Sigma), 15% FBS (Gibco), 0.1 mM β-mercaptoethanol, 1,000 U ml^{-1} leukaemia inhibitory factor (LIF, Chemicon), CHIR99021 (3 μM), PD0325901 (1 μM).

NPC differentiations and subcloning were performed as previously described[13]. NPCs were grown in N2B27 medium supplemented with EGF and FGF (10 ng ml^{-1} each), on 0.1% gelatin-coated flasks.

Cell transfection and clone isolation

All transgenic insertions were performed using the 4D nucleofector system from Lonza. For each nucleofection, five million cells were electroporated with 2.5 μg each of non-linearized targeting vectors and guide RNA–Cas9 encoding plasmids (MidiPreps). Nucleofected cells were then serially diluted and plated on 10-cm dishes. Forty-eight hours later, antibiotic selection was performed (puromycin, 0.4 μg ml^{-1}; hygromycin, 250 μg ml^{-1}; blasticidin, 5 μg ml^{-1}), except for transfection steps involving flippase-mediated removal of resistance cassettes, during which no selection was applied. One week after the initial plating, 80 to 96 single colonies were picked from dishes showing ideal clonal density and seeded in 96-well plates. These cells were subsequently split into one high-confluency plate used for PCR genotyping, and one low-confluency plate from which desired clones were further expanded until T25 density was reached. At this stage, some cells were kept to reconfirm the correct genotype by PCR, while the remaining cells were frozen.

Cell treatments

Xist expression in TX1072 mouse ES cells was induced upon administration of doxycycline (1 μg ml^{-1}). Auxin-mediated depletion of target proteins was achieved by supplementing culture media with auxin (Sigma) at the recommended concentration of 500 μM. Auxin-containing medium was renewed every 24 h. For auxin wash-out, auxin-containing medium was removed, cells were rinsed once with PBS, and exposed to auxin-free medium.

Protein extraction and western blotting

Cells were trypsinized, washed once in medium and once in PBS and then pellets were immediately frozen at −80 °C. Pellets were then resuspended in RIPA buffer (50 mM Tris-HCl pH 8.0–8.5, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) containing protease inhibitors (Roche), incubated for 30 min on ice and sonicated with a Bioruptor (three 10-s pulses). Lysates were then centrifuged for 20 min at 4 °C, and supernatants were kept. Protein concentration was determined using the Bradford (BioRad) assay. Samples were then boiled at 95 °C for 10 min in LDS buffer (Thermo) containing 200 mM DTT. For all western blots except those aimed at detecting SPEN, 4–12% Bis-Tris gels were used. For the detection of SPEN, a high-molecular-weight protein (>400 kDa), 3–8% tris-acetate polyacrylamide gels were used. Transfer was performed on a 0.45-μm nitrocellulose membrane using a wet-transfer system, at 350–400 mA for 2 h at 4 °C.

RNA extraction, reverse transcription, pyrosequencing and RNA sequencing

RNA extraction was performed using the RNeasy kit and on-column DNase digestion (Qiagen). Reverse transcription was performed on 1 μg total RNA using SuperScript III (Life Technologies). To quantify allelic skewing, cDNA was amplified using biotinylated primers and subsequently sequenced using Q24 Pyromark (Qiagen). Only samples showing a RNA integrity number greater than 9 were used to prepare RNA sequencing (RNA-seq) libraries (TruSeq). Paired-end 100-nt sequencing was performed on a HiSeq2500 or NovaSeq6000 (Illumina).

RNA FISH

Cells were dissociated using Trypsin (Invitrogen) for ES cells or Accutase (Invitrogen) for NPCs, washed twice in medium, and allowed to attach on poly-L-lysine (Sigma)-coated coverslips for 10 min. Cells were fixed with 3% paraformaldehyde in PBS for 10 min at room temperature, washed in PBS three times, and permeabilized with ice-cold permeabilization buffer (PBS, 0.5% Triton X-100, 2 mM vanadyl–ribonucleoside complex) for 5 min on ice. Coverslips were stored in 70% ethanol at −20 °C. Samples were dehydrated in 4 baths of increasing ethanol concentration (80%, 95%, 100% twice) and air-dried quickly. Probes were prepared from minipreps of intron-spanning bacteria artificial chromosomes (BACs) (clone RP24-157H12 for *Huwe1*, RP23-260I15 for *Atrx*) or plasmid (p510 for *Xist*). Probes were labelled by nick translation (Abbott) using dUTP labelled with spectrum green (Abbott) for *Huwe1*, spectrum red (Abbott) for *Atrx*, and Cy5 (Merck) for *Xist*. Labelled BAC probes were co-precipitated with Cot-1 DNA repeats in the presence of ethanol and salt, resuspended in formamide, denatured at 75 °C for 10 min, and competed at 37 °C for 1 h. Probes were then co-hybridized in FISH hybridization buffer (50% formamide, 20% dextran sulfate, 2X SSC, 1 μg $μl^{-1}$ BSA, 10 mM vanadyl-ribonucleoside) at 37 °C overnight. The next day, hybridized coverslips were washed three times for 5 min

with 50% formamide in 2X SSC at 42 °C, and three times for 5 min with 2X SSC. DAPI (0.2 mg ml^{-1}) was added to the penultimate wash and coverslips were mounted with Vectashield (Vectorlabs).

HaloTag labelling

HaloTag labelling of the SPEN–Halo fusion protein was performed in live TX1072 ES cells and NPCs. Cells were labelled with HaloTag-ligand-conjugated Janelia Fluor[11] (JF646–HaloTag or JF549–HaloTag, a gift from L. Lavis) at a final concentration of 250 nM in culture medium. Labelling was performed for 1 h at 37 °C, cells were then washed 4 times with generous volumes of PBS, and incubated with unlabelled medium for 15 min before proceeding with downstream experiments. For NPC labelling, cells were washed with unlabelled medium and not PBS, because NPCs detach when exposed to PBS. Auxin and/or doxycycline were kept in the labelling medium when necessary.

HaloTag labelling followed by RNA FISH

For co-detection of SPEN–Halo and *Xist* RNA, cells were labelled with JF549 as indicated above, and directly processed for fixation and permeabilization as detailed in the section 'RNA FISH'. Importantly, after permeabilization, coverslips were directly washed twice with PBS, twice with 2X SSC and immediately processed for FISH.

Mouse breeding, embryo collection and single-embryo RNA-seq

Timed natural matings were used for all experiments. Noon of the day when the vaginal plugs of mated females were identified was scored as E0.5. For *Spen* matings a conditional allele was used[9]. For oocyte deletions the published *Rosa26:Zp3-Cre* allele was used[31]. F_1 hybrid *Spen*$^{+/-}$ males were obtained by crossing *Spen*$^{+/+}$ CAST/EiJ females with *Spen*$^{+/-}$ C57BL/6J males. For *Spen* maternally deleted embryos, *Spen*$^{flox/flox}$ *Zp3-Cre*$^{+ve}$ C57BL/6J females were crossed with *Spen*$^{+/-}$ F_1 hybrid males. For *Spen* control embryos, *Spen*$^{flox/flox}$ *Zp3-Cre*$^{-ve}$ C57BL/6J females were crossed with *Spen*$^{+/-}$ F_1 hybrid males. The care and use of animals in this study was performed in accordance with the recommendations of the European Community (2010/63/UE). All experimental protocols were approved by the ethics committee of Institut Curie CEEA-IC118 under the number APAFIS#8812-2017020611033784v2 given by national authority in compliance with the international guidelines. Single-embryo RNA-seq was performed as previously described[32]. In brief, E3.5 embryos were collected and morphologically assessed to ensure that only viable samples were collected. The zona pellucida was removed by treatment with acidified Tyrode's solution. Single embryos were picked into individual tubes and cDNA was prepared and amplified as previously described[33]. Illumina libraries were prepared as published in ref. [34]. Paired-end 100-nt sequencing was performed with HiSeq2500 (Illumina).

Live-cell imaging and analysis

Cells were seeded on fibronectin-coated 35-mm glass-bottom dishes (Ibidi) 24 h before imaging. Doxycycline was added 1 h before image acquisition. Cells were imaged on the DeltaVision OMX microscope in widefield mode (GE Healthcare) using a 1.4 numerical aperture 100× oil immersion objective. The temperature was controlled at 37 °C and CO_2 at 8% during acquisition.

Images were acquired as *z*-stacks of 40 slices with 400-nm steps every 10 min for at least 4 h. Movies were deconvolved using Huygens deconvolution with the following parameters: Iteration 4; *S/N* 5, 10; quality threshold 0.1; and widefield mode 0.7 was used for background estimation. Two channels were registered using TetraSpec microspheres 0.1 µm (Invitrogen) and unwarpJ (Fiji plug-in). For segmentation, *z*-projected deconvolved registered images were used and pixels were classified as cloud or nuclei using Ilastik. Touching nuclei were sometimes manually separated. Cut-offs on resulting probability maps were set to 0.7. We next performed connected component analysis to obtain integer-labelled images in which each integer label corresponds to a unique nucleus. In the tailor-made Fiji plug-in the inputs are the raw max *z*-projected time-lapse images of the two channels and the integer labelled time-lapse image of the nuclei. The probability maps of the clouds give the region of interest in the time-lapse sequence in which total intensity is calculated. In the plug-in, clouds are associated with their corresponding nuclei; they are then linked via Kalman filter tracker over time. These unique links constitute track IDs and contain information about the intensity and area measurements for each cell. For each tracked cell, the first time point when a cloud is detected in one channel (*Xist* or SPEN) is labelled as reference time point 1.

Hi-C

Hi-C was performed as previously described[35], except that ligated DNA size selection was omitted, and dA-tailing was performed before biotin pull-down. In brief, each Hi-C experiment was performed on 10 million cells (NPCs) per sample. Cells were digested with DpnII at 37 °C overnight. DNA ends were filled with biotin-14-dATP at 23 °C for 4 h. DNA was then ligated with T4 DNA ligase at 16 °C overnight. Binding proteins were removed by treating ligated DNA with proteinase K at 65 °C overnight. Purified proximally ligated molecules were fragmented to obtain an average fragment size of 200 bp. After DNA end repair, dA-tailing and biotin enrichment, DNA molecules were ligated to Illumina TruSeq sequencing adapters at room temperature for 2 h. Final library PCR productions were carried out following the Illumina TruSeq Nano DNA Sample Prep Kit manual. Paired-end 100-nt sequencing was performed on a HiSeq4000 (Illumina).

Genetic engineering strategy for *Xist*–Bgl stem–loop tagging and SPEN complementation analysis constructs

To tag *Xist* with Bgl stem–loops[36], we nucleofected cells with pBS-Ptight-Xist-BglSL[12] (plasmid harbouring 18 repeats of Bgl stem–loops inserted between homology arms to target *Xist* exon 7, carrying a G418 selection gene, a gift from O. Masui). After G418 selection and FLP-FRT mediated removal of the selection cassette, clones were picked and genotyped. Positive clones were further tested to ensure that the stem–loop-tagged *Xist* could properly be induced and trigger gene silencing upon addition of doxycycline (data not shown).

Spen cDNA truncations were generated by splicing out different regions of the *Spen* open reading frame (Genscript, ORF clone OMu11416C) using overlap extension PCR. Each *Spen* truncation was cloned downstream of a CAGGS promoter into a vector carrying homology arms for targeted insertion at the *Rosa26*[37] locus as well as a SV40-promoter driven hygromycin-resistance gene. The BglG–GFP–SPOC targeting plasmid was designed by inserting a translational fusion between a BglG–GFP cassette and SPEN amino acids 3244–3643 into the same *Rosa26* targeting vector. Each of these 'complementation' constructs were independently targeted at *Rosa26* in SPEN-degron mouse ES cells. Independent clones were picked and protein expression of each SPEN truncation was assessed by western blot. XCI complementation analysis was then performed in 2–3 independent clones for *Spen* cDNA truncations, and 4 independent clones for BglG–GFP–SPOC expressing clones. The ability of cells to accumulate BglG–mCherry, BglG–GFP and BglG–GFP–SPOC upon the addition of doxycycline was assessed using microscopy (data not shown).

Immunofluorescence

ES cells were dissociated using trypsin (Invitrogen), washed extensively in medium, and allowed to attach on poly-L-lysine (Sigma)-coated coverslips for 10 min. Cells were then fixed with 3% paraformaldehyde in PBS for 10 min at room temperature, washed in PBS three times, and permeabilized with 0.25% Triton X-100 in PBS for 5 min at room temperature. Coverslips were then washed three times in PBS and blocked for 1 h with blocking buffer (PBS containing 2.5% BSA, 0.1% Tween20 and 10% normal goat serum). Coverslips were then incubated with primary antibodies diluted in blocking buffer at 4 °C overnight, washed three

times for 5 min in PBST (0.1% Tween20) the next day, incubated with fluorescently labelled secondary antibodies (1/500 in blocking buffer) for 1 h at room temperature, and washed again three times for 5 min in PBST. DAPI (0.2 mg ml^{-1}) was added to the penultimate wash and coverslips were mounted with Vectashield (Vectorlabs).

Immunoprecipitation

Nuclear extracts were prepared by resuspending 50 million fresh cells in ice-cold 10 ml buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1% NP-40, c0mplete EDTA free, phosSTOP) and rotating for 10 min at 4 °C. Nuclei were centrifuged at 800*g* for 10 min at 4 °C and resuspended in 1 ml IP buffer C150 (20 mM HEPES pH 7.9, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.25% NP-40, cOmplete EDTA free, phosSTOP). Lysates were briefly sonicated followed by Benzonase (Merck) digestion for 30 min at 4 °C. Finally, lysates were cleared through centrifugation at 13,000 rpm for 20 min before being incubated with 15 µl of GFP trap magnetic agarose bead slurry (ChromoTek) overnight at 4 °C. Beads were washed 5 times in IP buffer. For co-immunoprecipitation (Co-IP) western blot, washed beads were directly resuspended in LDS buffer (Thermo) containing 200 mM DTT, and boiled at 95 °C for 10 min.

Proteomics and mass spectrometry analysis

Proteins on magnetic beads were washed twice with 100 µl of 25 mM NH_4HCO_3 and we performed on-bead digestion with 0.2 µg of trypsin/LysC (Promega) for 1 h in 100 µl of 25 mM NH_4HCO_3. Samples were then loaded onto homemade C18 StageTips for desalting. Peptides were eluted using 40/60 MeCN/H_2O + 0.1% formic acid and concentrated to dryness under vacuum. Online chromatography was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled online to a Q Exactive HF-X with a Nanospay Flex ion source (Thermo Scientific). Peptides were first trapped on a C18 column (75 µm inner diameter × 2 cm; nanoViper Acclaim PepMap 100, Thermo Scientific) with buffer A (2/98 MeCN/H_2O in 0.1% formic acid) at a flow rate of 2.5 µl min^{-1} over 4 min. Separation was then performed on a 50 cm × 75 µm C18 column (nanoViper Acclaim PepMap RSLC, 2 µm, 100 Å, Thermo Scientific) regulated to a temperature of 50 °C with a linear gradient of 2% to 30% buffer B (100% MeCN in 0.1% formic acid) at a flow rate of 300 nl min^{-1} over 91 min. Mass spectrometry full scans were performed in the ultra-high-field Orbitrap mass analyser over the range m/z 375–1,500 with a resolution of 120,000 at m/z 200. The top 20 most intense ions were subjected to Orbitrap for further fragmentation via high-energy collision dissociation activation and a resolution of 15,000 with the intensity threshold kept at 1.3×10^5. We selected ions with charge state from 2^+ to 6^+ for screening. Normalized collision energy was set at 27 and the dynamic exclusion at 40 s. For identification, the data were searched against the *M. musculus* (UP000000589) Uniprot database using Sequest HF through Proteome Discoverer (v.2.2). Enzyme specificity was set to trypsin and a maximum of two missed-cleavage sites were allowed. Oxidized methionine and N-terminal acetylation were set as variable modifications. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.02 Da for MS/MS peaks. The resulting files were further processed using myProMS[38] v3.6 (work in progress). Calculation of the false discovery rate used Percolator and was set to 1% at the peptide level for the whole study. The label-free quantification was performed by peptide extracted ion chromatograms (XICs) computed with MassChroQ version 2.2[39]. For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN) with two missed cleavages were used. Median and scale normalization was applied on the total signal to correct the XICs for each biological replicate. To estimate the significance of the change in protein abundance, a linear model (adjusted on peptides and biological replicates) was performed and *P* values were adjusted with a Benjamini–Hochberg false-discovery rate procedure with a control threshold set to 0.05. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD015699.

Cross-linked CUT&RUN

CUT&RUN against SPEN was performed during a timecourse of *Xist* induction/SPEN degradation: 0 h DOX, 4 h DOX, 8 h DOX, 24 h DOX and 8 h DOX + auxin. Two biological replicates were performed. The original CUT&RUN protocol[27] was adapted for fixed cells: 10^6 cells in suspension were fixed with 2% formaldehyde diluted in PBS for 10 min at room temperature (2 ml final volume). Fixation was quenched with 125 mM glycine for 5 min and cells were washed twice in 1 ml PBS. Fixed cells were then permeabilized with 1 ml permeabilization buffer (20 mM HEPES pH 7.9, 150 mM NaCl, 0.5 mM spermidine, 0.25% TritonX-100, c0mplete EDTA free) for 5 min and washed twice in 1 ml PBS. Cells were then resuspended in 1 ml washing buffer (20 mM HEPES pH 7.9, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA, cOmplete EDTA free), bound to activated concanavalin beads (50 µl bead slurry used per 10 million cells) for 10 min, and blocked in 1 ml blocking buffer (wash buffer + 2 mM EDTA) for 5 min. At this stage, cells were resuspended in 500 µl wash buffer containing target antibodies diluted 1/200, transferred to 0.5-ml tubes, and incubated overnight at 4 °C on an end-to-end rotator. Cells were washed three times in 500 µl washing buffer followed by 1-h incubation with pA-MNase (500 µl of washing buffer containing 700 ng ml^{-1} pA-MNase, produced by the Protein Expression and Purification Core Facility of Institut Curie) and washed again three times in 500 µl washing buffer. After the last wash, cells were resuspended in 150 µl washing buffer, transferred to 1.5-ml tubes, and equilibrated to 0 °C in a metal block for 10 min. To start digestion, $CaCl_2$ was added to a final concentration of 1.5 mM, taking care to return each sample to 0 °C immediately afterwards. Digestion was performed at 0 °C for 1 h, before being stopped by adding 150 µl of 2X-STOP solution (200 mM NaCl, 20 mM EDTA, 5 mM EGTA, 0.1% NP-40, 40 µg ml^{-1} glycogen). RNase A was added to a final concentration of 50 µg ml^{-1} and samples were incubated at 37 °C for 20 min. SDS and proteinase K were then added to final concentrations of 0.1% and 300 µg ml^{-1}, respectively, and samples were incubated at 56 °C for 2 h followed by 68 °C for 16 h to reverse cross-linking. Total DNA was extracted using phenol/chloroform followed by two rounds of ethanol precipitation and DNA size selection (using 0.55× volume of Ampure XP beads relative to the DNA sample volume) to remove the large predominating undigested DNA fragments. Each time, beads were discarded and the supernatant (containing the selected small fragments resulting from MNase digestion) was precipitated with ethanol. After elution in 50 µl TE buffer, samples were quantified and analysed using Qubit and Tapestation assays. CUT&RUN libraries were prepared from 50 ng DNA per sample, using the Accel-NGS 2S Plus DNA Library Kit (Swift) according to the manufacturer's protocol. Paired-end 100-nt sequencing was performed on a HiSeq2500 (Illumina).

Bioinformatics analyses

All data were mapped to the mouse genome mm10, using the BL6-EiJ/CAST SNPs from the mouse genome project (v.5 SNP142), and the gene annotation from ensembl (v.92). Analyses were performed in R (v.3.4.2) and Bioconductor (v.3.6). See ref. [24] for more details.

RNA-seq analysis

Reads were trimmed using Trimgalore (v.0.4.4), mapped using STAR (2.5.3a, parameters: –outFilterMultimapNmax 1 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.06 –alignIntronMax 500000 –alignMatesGapMax 500000 –alignEndsType EndToEnd –outSAMattributes NH HI NM MD), and removed when mapping to the mitochondrial genome. Remaining reads were split by allele using SNPsplit (v.0.3.2). Allele-specific and the unassigned bam files were sorted, duplicates removed using picard (v.2.18.2, parameters: REMOVE_DUPLICATES = true ASSUME_SORTED = true) and pooled as the total reads.

Quantification of expression was performed using featureCount (parameters: -p -t exon -g gene_id, -s 1 for stranded RNA-seq of in vitro cell, -s 0 for non-stranded RNA-seq of single embryo). Data were then analysed in R using DESeq2 (v.1.18.1), calculating the sizeFactor on the count of total reads and applying it to the allele-specific counts.

For all RNA-seq analysis (SPEN–degron mouse ES cells, NPCs, and *Spen*-knockout embryos), genes showing less than 10 total allelic reads in at least one sample were discarded from the analysis. Allelic ratios were then computed for genes as follows: allelic_ratio = $reads^{B6}$/($reads^{B6}$+$reads^{Cast}$). Allelic ratios were then averaged between biological replicates.

RNA-seq analysis of SPEN-degron mouse ES cells. In order to define differential dependencies on SPEN for gene silencing during XCI in mouse ES cells, we removed skewed genes (that is, genes showing allelic ratios outside of a [0.15;0.85] interval in control conditions) from the analysis in Fig. 1d. We then defined a silencing index, translating how much a gene is silenced after 24 h of *Xist* induction with respect to the control condition: silencing_index = 1 − (allelic_ratio$_{DOX}$/allelic_ratio$_{control}$). We next filtered out genes showing less than 10% silencing (that is, silencing_index ≤ 0.1) in SPEN non-depleted conditions.

k-means with three clusters was then performed on the raw allelic_ratios across control, 24 h DOX and 24 h DOX + auxin conditions. Clustering identified three groups of genes differing by their response to loss of SPEN during *Xist* induction. To define how dependent on SPEN a gene is for silencing, we expressed the silencing defect observed upon loss of SPEN as a fraction of the total silencing that normally occurs in the presence of SPEN. Computationally, this translates in: Spen_dependence_index = 1 − (silencing_index$_{DOX+aux}$/silencing_index$_{DOX}$). A Spoc_dependence_index was derived identically.

For integration with *Hdac3*-knockout RNA-seq during XCI, we integrated the SPEN-degron dataset with an *Hdac3*-knockout RNA-seq dataset[24] generated from the same mouse ES cell background (TX1072) and at the same time point of *Xist* induction. The dataset was processed identically, and an *Hdac3*-dependence index was also computed as follows: Hdac3_dependence_index = 1 − (silencing_index$_{Hdac3KO}$/silencing_index$_{WT}$).

***Spen*-knockout E3.5 embryo RNA-seq analysis.** Integration of *Spen*-knockout and *Xist*-knockout embryo datasets was performed by integrating our *Spen*-knockout E3.5 female embryo RNA-seq dataset with a *Xist*-knockout single-cell RNA-seq (processed as pseudo-bulk for our analysis) dataset from E3.5 female embryos[10], also generated from a *M. musculus domesticus* × *M. musculus castaneus* mouse background.

SPEN-degron NPC RNA-seq analysis. In NPCs, X-linked genes were defined as escapees if their transcript allelic ratio was greater than 0.15 in at least one condition (0 h, 24 h or 48 h of SPEN depletion).

CUT&RUN bioinformatics analysis

Reads were trimmed using Trimgalore (v.0.4.4), mapped using STAR (2.5.3a, parameters: –outFilterMultimapNmax 1 –outFilterMismatchNmax 999 –outFilterMismatchNoverLmax 0.06 –alignIntronMax 1 –alignMatesGapMax 2000 –alignEndsType EndToEnd –outSAMattributes NH HI NM MD), and removed when mapping to the mitochondrial genome. Remaining reads were split by allele using SNPsplit (v.0.3.2). Allele-specific and the unassigned bam files were sorted, duplicates removed using picard (v.2.18.2, parameters: REMOVE_DUPLICATES = true ASSUME_SORTED = true) and pooled as the total reads. BigWig of coverage files were performed using DeepTools bamCoverage (parameters: –extendReads –binSize 1, with –extendReads 200 for single end data). A scaling factor was calculated as 10^6/total number of reads, and the same factor was given as the parameter –scaleFactor for both allelic signals. Peak calling was performed using macs2 (v.2-2.1.2.1, parameters for CUT&RUN: –bw 300 -f BAMPE -q 0.01 –keep-dup auto –broad for CUT&RUN and pol2S5 ChIPseq; –bw 300 -f BAMPE -q 0.01 –keep-dup auto –call_summits for other ChIPseq). For quantification of signal in peaks, reads were counted using the featureCounts function from Subread (v.1.28.1, parameters: -p -s 0). Data scaling was performed in R using DESeq2 (v.1.18.1), calculating the sizeFactor on the counts of total reads in 10-kb windows and applying it to the allele-specific counts in peaks.

Peak filtering. SPEN-specific peaks were defined as having $\log_2$FoldChange ≥ 1 compared to auxin treatment (negative control, SPEN-degraded), and an adjusted *P* value ≤ 0.001.

Total SPEN enrichment in promoter window. To compare SPEN accumulation among promoters of all X-linked genes in an unbiased manner—including genes that fail to have any peak called at their promoters—we performed DEseq analysis on counts spanning total promoter windows.

Genomic features and integration with RNA-seq. Promoters were defined as ±2-kb windows centred around the transcription start sites of genes. Putative active enhancers and their deacetylation kinetics during XCI were obtained from ref. [24]. Gene-silencing efficiency was determined according to the silencing_index defined in the section 'RNA-seq analysis'. We observed that our silencing_index ranges between 0 and 0.9. Hence, we split this interval in three to define high, medium and low gene-silencing efficiency groups with silencing_index comprising [0.6,0.9], [0.3,0.6[and [0,0.3[respectively.

Integration with publicly available ChIP-seq data. SPEN peaks were intersected with other peaks called from publicly available ChIP-seq data for HDAC3[24] (same cellular background, TX1072, 2i + LIF condition), RNAPII-pS5[26], CHD4[26] and MBD3[26] (all in the 2i + LIF condition).

Hi-C analysis

Data were processed with HiC-Pro (v.2.11.0) in allele-specific mode. Only pairs with both reads having MAPQ > 30 were kept. Matrices were made using cooler cload (v.0.8.5) at 1-kb or 10-kb resolution, using HiGlass for visualization and snapshots. Topologically associating domains were called using HicExplorer HicFindTads (parameters: –correctForMultipleTesting fdr –minDepth 250000 –maxDepth 4000000 –step 50000 –thresholdComparisons 0.1 –delta 0). The expected value for the Hi-C signal was calculated on the non-allele-specific signal using cooltool compute-expected. Average scaled matrices of observed/expected values for allele-specific signal were produced with Coolpup.py (parameters: –local –rescale –rescale_size 299), using non-allele-specific expected values to normalize both alleles to the same expected values. Average heat maps were plotted using plotpup.py. For quantification, the Hi-C signal was averaged over topologically associating domains upper-triangle for each allele-specific matrix (10 kb) using the hicExplorer hicSummarizeScorePerRegion available at https://github.com/heard-lab/HiCExplorer (parameters: –summarizeType mean –rmDiag 1).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

RNA-seq, Hi-C and CUT&RUN data used in this study have been deposited in the Gene Expression Omnibus under accession number GSE131784. Source Data for Figs. 1–3 and Extended Data Figs. 1, 3, 4 are provided with the paper, either in the form of supplementary tables or source data files.

31. de Vries, W. N. et al. Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. *Genesis* **26**, 110–112 (2000).
32. Zylicz, J. J. et al. G9a regulates temporal preimplantation developmental program and lineage segregation in blastocyst. *eLife* **7**, e33361 (2018).
33. Tang, F. et al. RNA-seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5**, 516–535 (2010).
34. Huang, Y. et al. Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *eLife* **6**, e22345 (2017).
35. Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **123**, 56–65 (2017).
36. Chen, J. et al. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl Acad. Sci. USA* **106**, 13535–13540 (2009).
37. Barau, J. et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**, 909–912 (2016).
38. Poullet, P., Carpentier, S. & Barillot, E. myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *Proteomics* **7**, 2553–2556 (2007).
39. Valot, B., Langella, O., Nano, E. & Zivy, M. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* **11**, 3572–3577 (2011).

Acknowledgements We thank K. Ancelin for support throughout the project and help with in vivo experiments; J. Barau, D. Holoch and R. Margueron for support and help with the project; M. Carrara for critical reading of the manuscript; and members of the Heard laboratory for discussions. We thank E. Nora for sharing the OsTIR1 and AID-targeting plasmids; T. Honjo for sharing the *Spen*^{flox} mouse line; O. Masui for sharing the *Xist*–Bgl stem–loop targeting constructs and L. Lavis for sharing Halo-JF646 and JF549 with us. We also thank the imaging platform (A. Dauphin and PICT-IBiSA (UMR3215/U934)) and the protein purification and sequencing platforms of Institut Curie as well as L. Villacorta, J. Provaznik and V. Benes of GeneCore at EMBL. This work was funded by a Boehringer Ingelheim doctoral fellowship (20017-2019 to F. Dossin), an ERC Advanced Investigator award (ERC-ADG-2014 671027 to E.H.), Labellisation La Ligue (to E.H.), ANR (DoseX 2017: ANR-17-CE12-0029, Labex DEEP: ANR-11- LBX-0044, ABS4NGS: ANR-11-BINF-0001, and part of the IDEX PSL: ANR-10-IDEX-0001-02 PSL to E.H.), a Sir Henry Wellcome Postdoctoral Fellowship (201369/Z/16/Z to J.J.Z.), 'Région Ile-de-France' and Fondation pour la Recherche Médicale grants (to D.L.), EMBO long-term fellowships (ALTF 549-2014 to I.P., ALTF 301-2015 to T.C.), and Fondation pour la Recherche Medicale (SPF 20140129387 to I.P.) and NIH (HG003143 to J.D.) grants. J.D. is an investigator of the Howard Hughes Medical Institute.

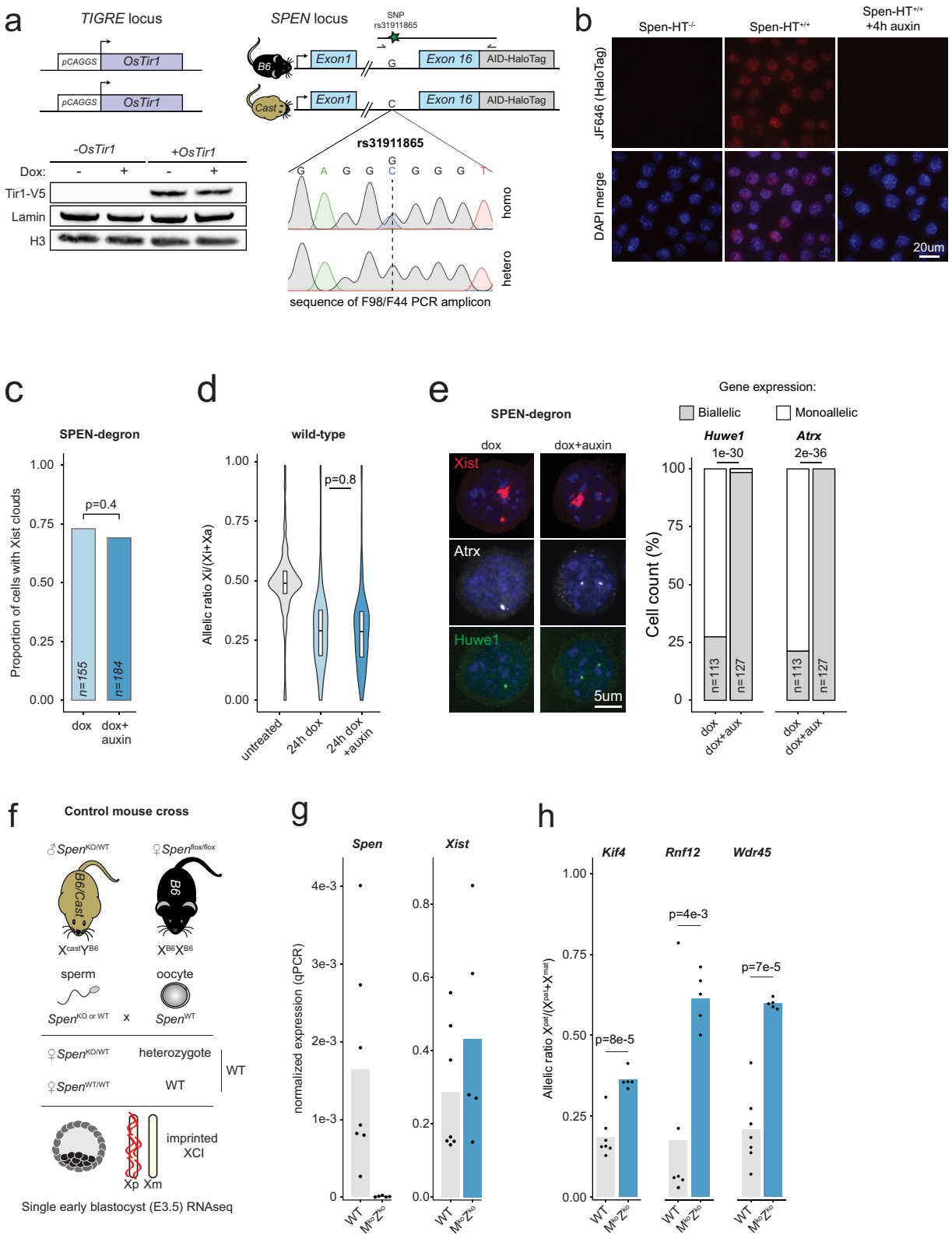
Author contributions F. Dossin and E.H. conceived the experiments and F. Dossin performed them unless stated otherwise. T.M. provided the *Spen*-knockout mice and J.J.Z. performed the embryo experiments. E.H. and I.P. supervised the work. J.R. established the *Xist* stem–loop cell line and acquired live-cell images. F. Dossin, J.R. and V.K. analysed live-cell imaging experiments. V.K. wrote the image analysis software. F. Dossin and S.C. processed and analysed the data. A.L.S. cloned the SPEN cDNA truncation constructs. T.C. helped with some experiments and M.A. derived NPC clones. Y.Z. prepared Hi-C libraries in the laboratory of J.D. F. Dingli carried out the mass spectrometry experimental work and D.L. supervised mass spectrometry experiments and data analysis. F. Dossin and E.H. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information

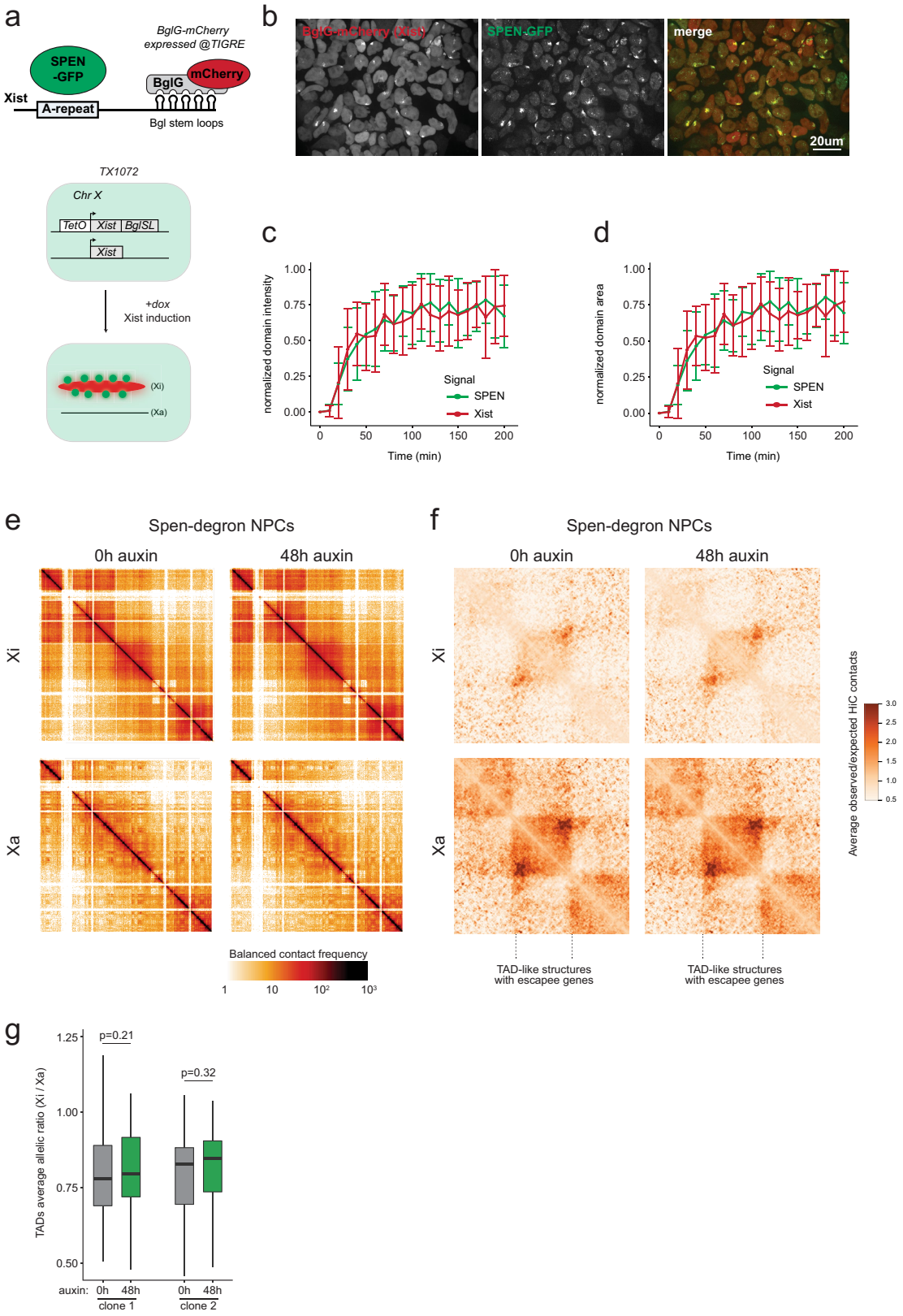
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-1974-9.

Correspondence and requests for materials should be addressed to E.H.



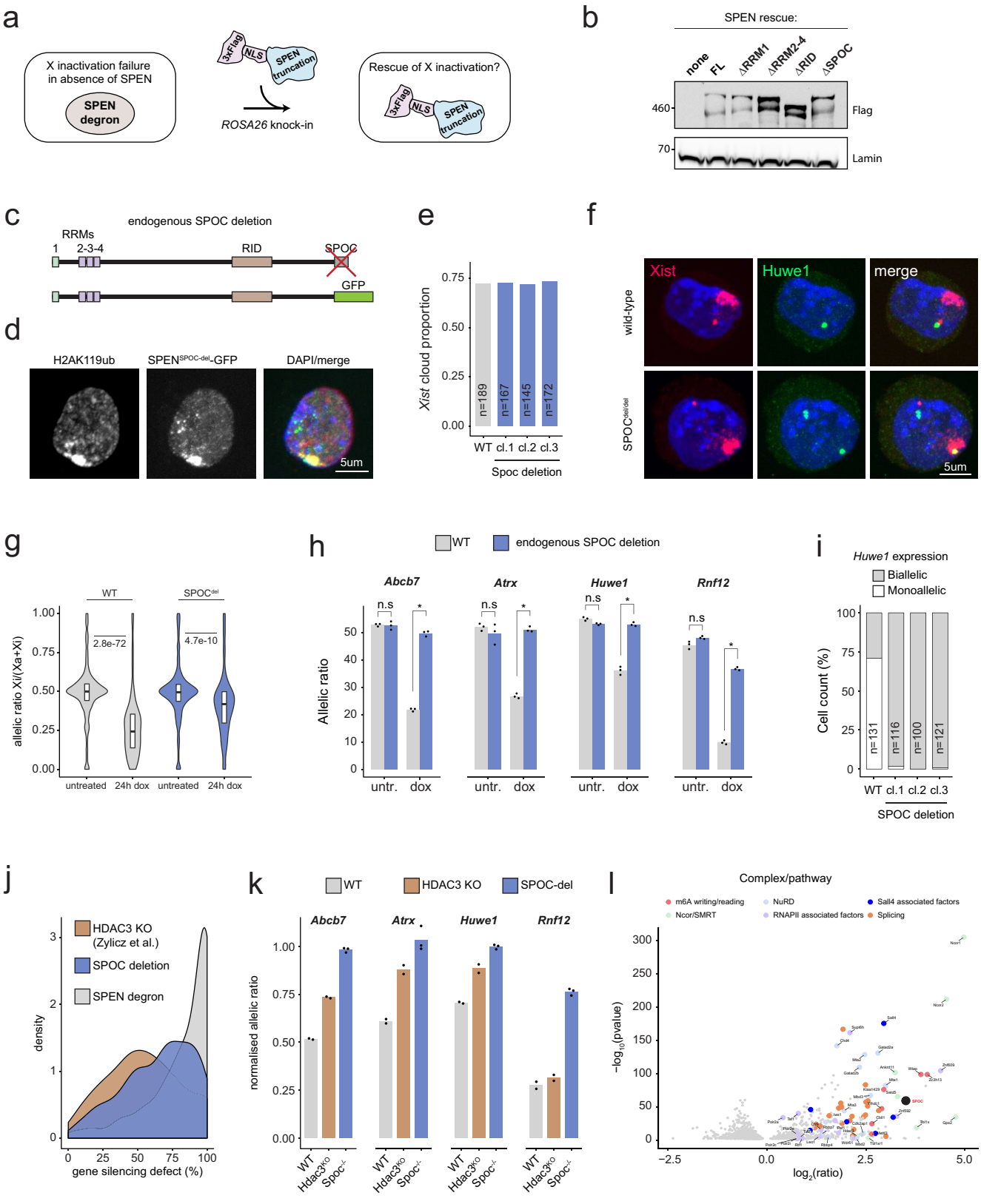
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | SPEN mediates gene silencing across the entire X chromosome in vitro and in vivo. a, Schematic representation of the SPEN-degron genotype with AID-HaloTag insertions in frame with the C terminus of endogenous SPEN. Targeted homozygous insertion of V5-tagged OsTIR1 at the *TIGRE* locus (top left) results in its constitutive protein expression as assessed by western blot (bottom left). Right, Sanger sequencing results for a PCR amplicon specific to AID-HaloTag insertions and covering a SNP outside of the recombined left homology arm. Detection of both alleles in the amplicon confirms homozygous AID knock-in. **b**, Fixed-cell imaging of HaloTag in wild-type cells (left), in SPEN-degron mouse ES cells (middle) and in SPEN-degron mouse ES cells exposed to auxin for 4 h (right). Cells were labelled with Halo-JF646 before fixation. SPEN–Halo is properly localized to the nucleus, and is depleted upon auxin treatment. This experiment was repeated at least twice with similar results. **c**, Bar graph showing the proportion of cells displaying *Xist* RNA clouds (quantified using RNA FISH) before and after degradation of SPEN (*n*, number of cells counted; χ^2 test). **d**, Violin plot showing the distribution of X-chromosomal transcript allelic ratios (obtained by RNA-seq) after 0 h DOX, 24 h DOX or 24 h DOX + auxin treatment in wild-type SPEN-degron mouse ES cells. Horizontal lines denote the median, box limits correspond to upper and lower quartiles, averages of two independent clones shown, $n = 434$ genes, two-sided Student's *t*-test. **e**, RNA FISH experiments for *Xist* (red) and two X-linked genes: *Atrx* (grey) and *Huwe1* (green), in SPEN-degron mouse ES cells treated with DOX only, or DOX in combination with auxin for 24 h. The proportion of *Atrx*/*Huwe1* monoallelic and biallelic expression among *Xist*-expressing cells is shown (*n*, number of cells counted; χ^2 test). **f**, Illustration of the control hybrid mouse crossbreeding scheme for the experiment shown in Fig. 1g, h. **g**, Quantitative PCR (qPCR) analysis of *Spen* and *Xist* transcripts in wild-type ($n = 7$) and maternal-zygotic *Spen*-knockout ($n = 5$) E3.5 embryos. **h**, Pyrosequencing assay of three X-linked transcripts in maternal-zygotic *Spen*-knockout ($n = 5$) and wild-type ($n = 7$) E3.5 embryos (two-sided Student's *t*-test). In **g**, **h**, bars show the mean value and individual data points are shown as dots.



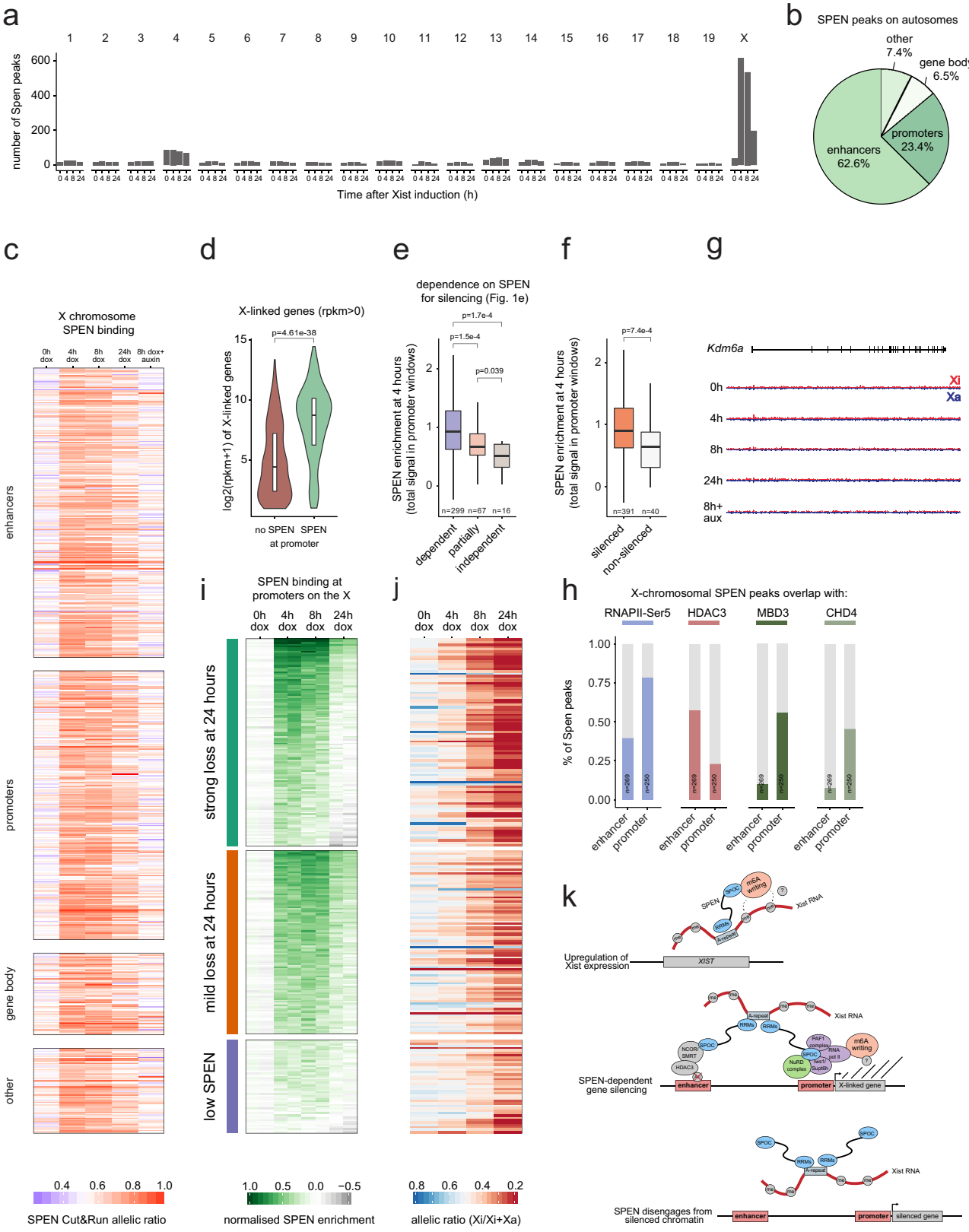
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | SPEN localizes to the X chromosome immediately upon *Xist* upregulation and throughout the stages of XCI, but is dispensable for maintenance of X-linked gene silencing. a, Scheme of the strategy for live-cell imaging of SPEN protein and *Xist* RNA. **b**, Live-cell snapshot after 16 h of *Xist* induction in the cell line shown in **a**. This experiment was repeated at least twice with similar results. **c**, **d**, Kinetics of total intensity (**c**) and area (**d**) of *Xist* (red) and SPEN (green) domains over time during *Xist* induction. The data in **c**, **d** are the averages of 27 tracked cells. Error bars indicate standard deviation. Images were acquired every 10 min. Time point 1 is defined as the earliest time at which a SPEN or *Xist* domain is detected in each cell. Intensity and area values were respectively normalized to the maximum value reached for each signal (SPEN and *Xist*). **e**, Hi-C map of the inactive (top) and active (bottom) X chromosomes (resolution, 1.024 Mb) in NPCs after 0 h or 48 h of auxin-mediated SPEN depletion. **f**, Heat map of the average contact enrichment on scaled topologically associating domains containing escapees in NPCs after 0 h or 48 h of auxin-mediated SPEN depletion. **g**, Quantification of the allelic ratio (inactive/active X chromosome) of the Hi-C signal within topologically associating domains ($n = 37$) shown in **f**, after 0 h or 48 h of auxin-mediated SPEN depletion. Horizontal lines denote the median, box limits correspond to upper and lower quartiles, two-sided Wilcoxon rank-sum test. In **e**, **f**, averages of two independent clones are shown.



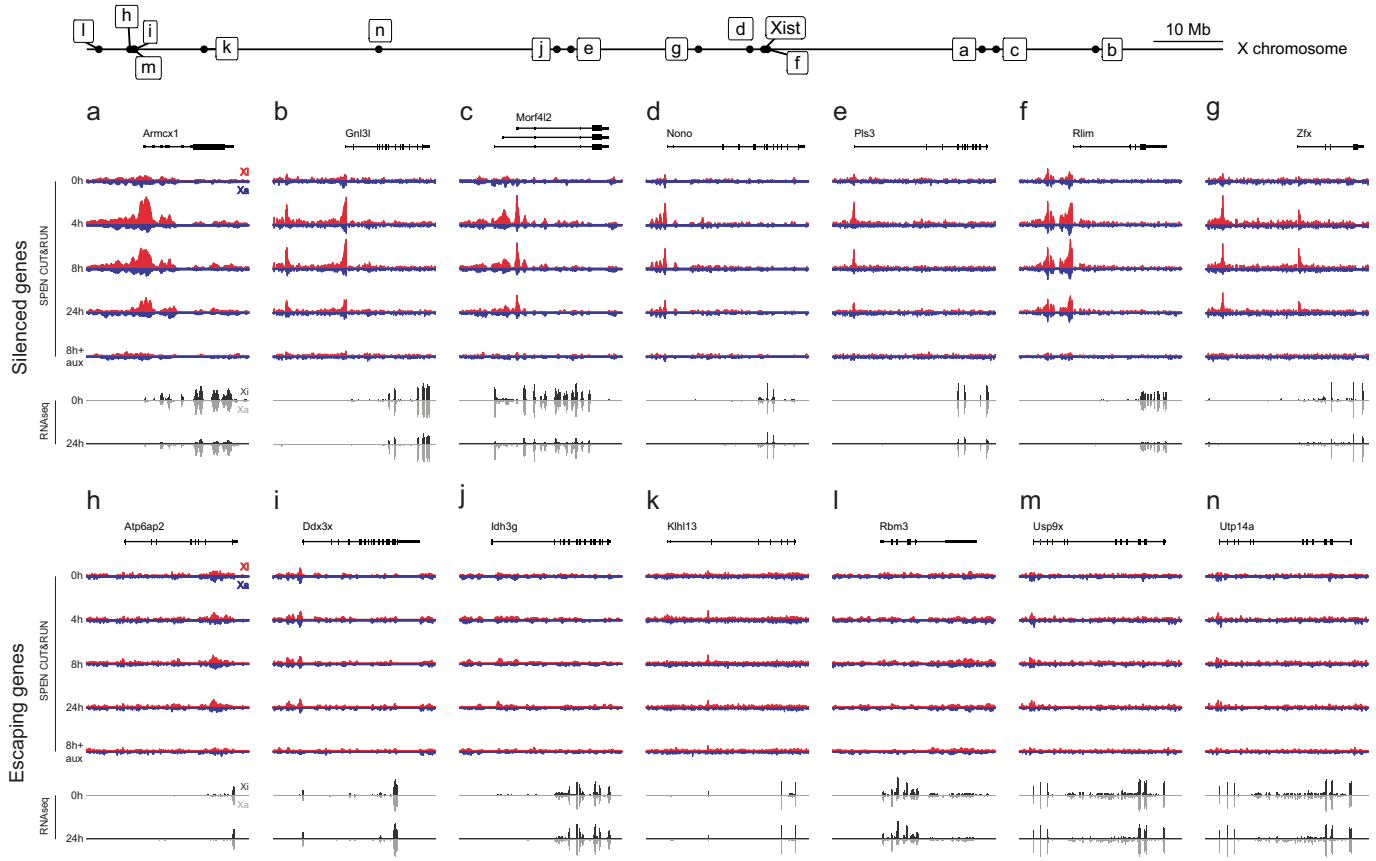
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | The SPOC domain of SPEN mediates gene silencing and interacts with multiple molecular pathways. a, Scheme of complementation strategy. **b**, Western blot detection of overexpressed 3×Flag-tagged SPEN protein rescue fragments. **c**, Scheme showing endogenous deletion of SPOC. **d**, Sub-nuclear localization of endogenous SPEN lacking its SPOC domain upon *Xist* RNA induction. The inactive X chromosome is identified using immunofluorescence detection of H2AK119ub1. **e**, Bar graph showing the proportion of cells with *Xist* RNA clouds (assayed by RNA FISH) in wild-type cells and three independent SPOC-deletion clones after induction of *Xist* for 24 h (*n*, number of counted cells). **f**, RNA FISH for *Xist* (red) and *Huwe1* (green) in SPOC-deletion and wild-type cells treated with DOX for 24 h. **g**, Violin plot showing the distribution of X-chromosomal transcript allelic ratios (measured by RNA-seq) after 0 h or 24 h DOX treatment in wild-type and SPOC-deletion mouse ES cells. Horizontal lines denote the median, box limits correspond to upper and lower quartiles, averages of three independent clones shown, $n = 469$ genes, two-sided Student's *t*-test. **h**, Bar graph of transcript allelic ratios (obtained from pyrosequencing) for four X-linked genes in SPOC-deletion (blue) or wild-type (grey) cells. Bars show mean values for three independent SPOC-deletion clones ($^*P < 10^{-4}$, two-sided Student's *t*-test). **i**, Bar graph showing the proportion of cells expressing *Huwe1* monoallelically (white) or biallelically (grey), assayed by RNA FISH, in wild-type cells and in three independent SPOC-deletion clones after induction of *Xist* for 24 h (*n*, number of counted cells). **j**, Density plot showing the distribution of gene silencing defects (see Methods) observed across the X chromosome in RNA-seq data from HDAC3-knockout[24] SPEN-degron and SPOC-deletion (this study) ES cells after 24 h of *Xist* induction. **k**, Bar graph of normalized allelic ratios (obtained from pyrosequencing) for four X-linked genes in HDAC3-knockout (brown), SPOC-deletion (blue) and wild-type (grey) cells after 24 h of *Xist* induction. Bars show mean values for two independent HDAC3 clones and three independent SPOC deletion clones; individual data points are shown. **l**, Volcano plot of fold changes in GFP-pull-down (BglG–GFP–SPOC compared with BglG–GFP) and their adjusted *P* values (Benjamini–Hochberg procedure, see Methods for statistical analysis). Quantitative label-free mass spectrometric analysis was performed on four independent biological replicates. In **b**, **d**, **f**, experiments were repeated at least twice with similar results.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | SPEN is recruited by *Xist* to active gene promoters and enhancers where it silences transcription and subsequently disengages from chromatin. **a**, Bar graph showing the number of SPEN peaks on each chromosome after 0 h, 4 h, 8 h and 24 h of *Xist* induction in mouse ES cells. **b**, Annotation of SPEN peaks on autosomes. **c**, Heat map showing allelic ratios at SPEN peaks during XCI among different X-linked genomic features. **d**, Violin plot showing expression (RPKM) of genes accumulating SPEN ($n = 259$) or not accumulating SPEN ($n = 689$) at their promoters. Genes showing 0 RPKM were excluded from this plot. **e**, Box plots showing SPEN enrichment after 4 h of *Xist* induction within promoter windows of genes grouped on the basis of their level of dependency on SPEN for gene silencing (see Fig. 1e). **f**, Box plots showing SPEN enrichment after 4 h of *Xist* induction within promoter windows of genes grouped on the basis of whether or not they are silenced at 24 h of *Xist* induction (see Methods). In **d**–**f**, data were analysed using the two-sided Wilcoxon rank-sum test, horizontal lines denote the median, box limits correspond to upper and lower quartiles. **g**, UCSC Genome Browser allele-specific track showing SPEN binding around *Kdm6a*, an escaping gene (blue, Cast-Xa; red, B6-Xi; all tracks are scaled identically). **h**, Bar graphs showing overlap between SPEN-binding sites and the binding sites of four different factors at X-linked enhancers and promoters. **i**, **j**, Heat maps showing normalized SPEN enrichment ($\log_2$) at promoters (both replicates are shown) (**i**) and gene silencing kinetics (allelic ratio) during XCI (**j**) within three groups of X-linked genes showing different dynamics of SPEN accumulation and loss. **k**, Schematic of the function of SPEN in XCI. In **a**–**f**, **h**–**j**, data are from two biological replicates.



Extended Data Fig. 5 | UCSC Genome Browser allelic tracks of SPEN binding and transcript expression at X-linked genes. a–**n**, Top, Genome Browser allelic tracks of SPEN binding (from CUT&RUN) at silenced genes (**a**–**g**) and non-silenced genes (**h**–**n**) during a time course of *Xist* induction in mouse ES cells (blue, Cast-Xa; red, B6-Xi; scaled identically within each panel). Bottom, allelic tracks of transcript expression (from RNA-seq) at 0 h and 24 h of *Xist* induction in mouse ES cells (light grey, Cast-Xa; black, B6-Xi; scaled identically within each panel). The relative position of each gene along the X chromosome is shown at the top of the figure.

nature research

Corresponding author(s): Edith Heard

Last updated by author(s): Dec 10, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All microscopy images were acquired either with an Inverted Confocal Spinning Disk Roper/Nikon for fixed cells or a Super-resolution microscope OMX (Applied Precision Incorporation, DeltaVision) for live cells. Sequencing data was collected using the Illumina platform. Pyrosequencing data was collected using the PyroMark Q24 System from Qiagen. QPCR was performed on a ViiA 7 Real-Time PCR System from Thermo Fisher Scientific. Mass Spectrometry was performed on Thermo Scientific Orbitrap Fusion Tribrid MS from Thermo Fisher Scientific. Western blot (chemiluminescent) images were collected using a ChemiDoc MP from BioRad.
Data analysis	Trimgalore (v 0.4.4), STAR (2.5.3a), SNPsplit (v 0.3.2), picard (v2.18.2), DESeq2 (v1.18.1), R (3.4.1), featureCount (1.6.3), Fiji (2.0.0), dplyr (0.7.4), readr (1.1.1), tidyr (0.8.0), ggplot2 (2.2.1), macs2 (v 2-2.1.2.1), Subread (v1.28.1), HiC-Pro (v2.11.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated in this study are available on GEO database under the number GSE131784.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was not predetermined. For mouse experiments, we used a sample size commonly used and accepted for basic statistical inference while using an justifiable number of mice. For tissue culture based experiments, at least two independent clones were systematically assessed for each genotype. Typically, for the SPOC tethering experiment, 4 independent clones were characterized with very little variation being observed between them.
Data exclusions	No data were excluded from the analysis, with the exception of live imaging analysis, from which dying cells, as well as cells which moved outside from the field of view during movie acquisition were removed.
Replication	All attempts at replication were successful and noted in the relevant figure legend.
Randomization	Samples were not randomized, given that samples were grouped according to their respective genotypes.
Blinding	For most experiments, no blinding was performed as most measurements were derived from third party machines/softwares, and hence not affacted by subjective interpretation. For RNA FISH image analysis (counting of Xist RNA clouds and X-linked gene pinpoints), each experimental condition was blinded from the first author F. Dossin during counting.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Epitope, Antibody, reference, Application, Dilution HaloTag, Promega cat. #G9211, WB, 1/1000 V5, Sigma cat. #V8012, WB, 1/2500, Flag, Sigma cat. #F1804, WB and IF, 1/1000 and 1/200 GFP, Abcam cat. #ab290 LotGR3222604-1, Cut&Run and IF, 1/200 GFP, Roche cat. #11814460001, WB, 1/1000 H2Ak119ub1, Cell Signaling cat. #8240, IF, 1/500 Lamin B1, Abcam cat. #ab16048, WB, 1/3000 PCNA, DAKO cat. #M0879, WB, 1/3000 Ncor1, Abcam cat. #ab2482 LotGR320472-6, WB, 1/1000 Ncor2, Abcam cat. #ab5802 LotGR259451-18, WB, 1/1000 Mta1, Cell Signaling cat. #5647, WB, 1/1000 Wtap, Proteintech cat. #10200-1-AP, WB, 1/1000 Mettl3, Abcam cat. #ab195352, WB, 1/1000 Hdac3, SantaCruz cat. #sc-376957, WB, 1/1000 Rpb1, Abcam cat. #ab817, WB, 1/2000
Validation	The HaloTag antibody is validated for western blot in Fig. 1b, with specific signal disappearing upon auxin treatment, and reappearing upon removal of auxin from the culture medium. The V5 antibody is validated for Western blot in Extended Data Fig. 1a, with specific signal being observed only in cells expressing

V5-tagged Tir1.
The Flag antibody is validated for immunofluorescence in Fig. 3c, with specific nuclear signal being observed only in cells expressing Flag-tagged SPEN truncations.
The GFP antibody is validated for CUT&RUN by showing specific genomic signal for GFP-tagged SPEN, which disappears upon treatment of cells with auxin.
The H2AK119ub1 has been validated for immunofluorescence in Zylicz et al., 2019
The Lamin B1 antibody have been KO validated by abcam.
The PCNA antibody is cited 361 times according to CiteAb.
The Ncor1/Ncor2 antibodies are validated for western blot in Fig3. h, as showing increased signal upon immunoprecipitation of SPOC.
The Mta1 antibody has been cited 12 times previously according to CiteAb.
The Wtap antibody have been KO validated by ProteinTech.
The Mettl3 antibody have been KO validated by Abcam.
The Hdac3 antibody has been cited 9 times according to CiteAb.
The Rpb1 antibody has been cited 280 times previously according to CiteAb.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	All cell lines were derived from the TX1072 female mouse embryonic stem cell line (Schulz et al., 2014)
Authentication	None of the cell lines were authenticated
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	cells used are not in the ICLAC database

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	We used adult animals (from 6 weeks to 3 months for females, and 7 weeks to 1 year for males) for producing preimplantation embryos (first 4 days post mating). For Spen matings a published conditional allele was used (Yabe at al., 2007). For oocyte deletions published Rosa26:Zp3-Cre allele was used (DeVries et al., 2000). F1 hybrid Spen+/- males were obtained by crossing Spen+/+ CAST/EiJ females with Spen+/- C57BL/6J males. For Spen maternally deleted embryos, Spenflox/flox Zp3-Cre+ve C57BL/6J females were crossed with Spen+/- F1 hybrid males. For Spen control embryos, Spenflox/flox Zp3-Cre-ve C57BL/6J females were crossed with Spen+/- F1 hybrid males.
Wild animals	this study does not contain any wild animals
Field-collected samples	this study does not contain animals collected from the fields.
Ethics oversight	Animal care and use for this study were performed in accordance with the recommendations of the European community (2010/63/UE). All experimental protocols were approved by the ethics committee of Institut Curie CEEA-IC118 under the number APAFIS#8812-2017020611033784v2 given by national authority in compliance with the international guidelines.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Chapter 7

Structural and biophysical analysis of the SPOC/Ser5P-CTD interaction

7.1 Rationale for investigating a potential phosphodependent interaction between SPOC and the CTD of RNAPII

Among all protein interactions detected in our SPOC-immunoprecipitation (see **subsection 6.3.3**), the potential interaction with components of the general transcription machinery raises the exciting possibility that SPEN could directly regulate transcription. If confirmed, this could explain how genes are turned off so efficiently upon *Xist* coating of the X chromosome.

While powerful for detecting protein partners of SPOC, the IP-MS framework does not allow to conclude whether any of the detected partners interact *directly* with SPOC, and hence precludes drawing further mechanistic conclusions.

In favor of a direct interaction between SPOC and RNA polymerase II, our CUT&RUN experiment revealed that:

- SPEN recruitment to X-linked chromatin predominantly overlaps with RNAPII-bound sites, at transcriptionally active genes (see Dossin et al., External Data Fig. 4h).
- SPEN binding to chromatin is lost following X-linked gene silencing (i.e. upon decreasing levels of chromatin-bound RNAPII, see Dossin et al., Fig. 4j, k).

However, SPEN binding is restricted to enhancer and promoter features, with virtually

no "diffuse" binding detected within gene bodies, contrary to what would have been expected should SPEN/SPOC interact with RNAPII generally.

SPEN binding does however strongly ressemble that of a peculiar fraction of RNA polymerase II: that which is phosphorylated on Serine 5 of the C-terminal domain of POLR2A, the enzymatic core (and largest subunit) of RNAPII.

7.1.1 The CTD of RNAPII and function of the CTD code

The *C-t*erminal *d*omain (CTD) of POLR2A is unique in that it is strikingly composed of several tandem repeats (52 in mouse and humans [Chapman et al., 2008]) of the $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$ (YSPTSPS) consensus heptapeptide sequence (**Figure 7.1A**). *In vivo*, these repeats are subject to extensive *p*ost-*t*ranslational *m*odifications (PTMs), which constitute a *bona fide* "CTD code" [Komarnitsky et al., 2000; Buratowski, 2003].

The CTD code is intricately linked to different stages of the transcription cycle, as well as several co-transcriptional events including mRNA capping, polyadenylation and splicing, but also histone modifications [Hsin and Manley, 2012]. Indeed, it has been shown that the CTD of RNAPII acts as a docking platform for several transcription and RNA processing factors, as well as chromatin modifiers, whose binding depends on the PTM-status of the CTD [Jeronimo et al., 2013; Jasnovidova and Stefl, 2013].

Among all CTD-PTMs, phosphorylations at Serine 2 (Ser2P), Serine 5 (Ser5P) and Serine 7 (Ser7P) of the heptapeptide (**Figure 7.1A**) have been most extensively studied [Heidemann et al., 2013]. Ser2P and Ser5P represent 75% of the total fraction RNAPII phosphorylated at its CTD, illustrating their essential roles during the transcription process [Schüller et al., 2016].

ChIP analyses [reviewed in Heidemann et al., 2013] revealed that Ser5P-marked RNAPII is most highly enriched at gene promoters, with levels sharply decreasing further into the 3' end of genes (**Figure 7.1B**).

Conversely, Ser2P-marked RNAPII signal is typically low around the promoter region, but increases continuously into the 3'-end of genes, being most highly enriched around the site of transcription termination (**Figure 7.1B**).

Given their peculiar distributions within the transcription unit, it is not surprising that Ser5P has been associated with transcription initiation and the transition into early elongation, while Ser2P has been associated with events promoting the productive phase of transcription elongation as well as termination.

7.1.1.1 Function of Ser5 phosphorylation

Ser5 phosphorylation [by kinases reviewed in Heidemann et al., 2013] results in RNAPII dissociation from the Mediator complex [Søgaard and Svejstrup, 2007], promoting promoter escape. Furthermore, Ser5P is required for the recruitment of mRNA 5'-capping enzymes [Cho et al., 1997; McCracken et al., 1997], as well as the SET1 complex [Ng et al., 2003; Krogan et al., 2003] – which catalyses H3K4 mono-, di- and trimethylation, a set of marks which recruit histone acetylases, nucleosome remodelers and transcription elongation factors [Kouzarides, 2007] – further supporting entry into transcription elongation.

7.1.1.2 Function of Ser2 phosphorylation

Ser2 phosphorylation [by kinases reviewed in Heidemann et al., 2013] requires antecedent Ser5 phosphorylation. Increasing Ser2P levels favor productive transcription elongation through the direct recruitment of elongation factors such as SPT6 [Yoh et al., 2007]. Similarly, Ser2P promotes co-transcriptional mRNA processing by recruiting splicing factors [Morris and Greenleaf, 2000; David et al., 2011]. Finally, high Ser2P levels around the transcription termination site enable efficient and successful transcription termination, by recruiting several transcription termination and mRNA 3'-end processing (e.g. polyadenylation) factors [Meinhart and Cramer, 2004; Lunde et al., 2010].

7.1.1.3 Function of Ser7 phosphorylation

Unlike Ser2 and Ser5, Ser7 phosphorylation [by kinases reviewed in Heidemann et al., 2013] accounts for less than 5% of the total fraction of RNAPII phosphorylated at its CTD. The distribution of Ser7P-marked RNAPII initially ressembles that of Ser5P, peaking at the transcriptional start site (TSS); but unlike Ser5P, Ser7P levels remain high until the polyadenylation site (**Figure 7.1B**). While Ser7P plays a crucial role in the transcription and processing of RNAPII-transcribed snRNAs [Egloff et al., 2007], its function at protein-coding genes is still very unclear.

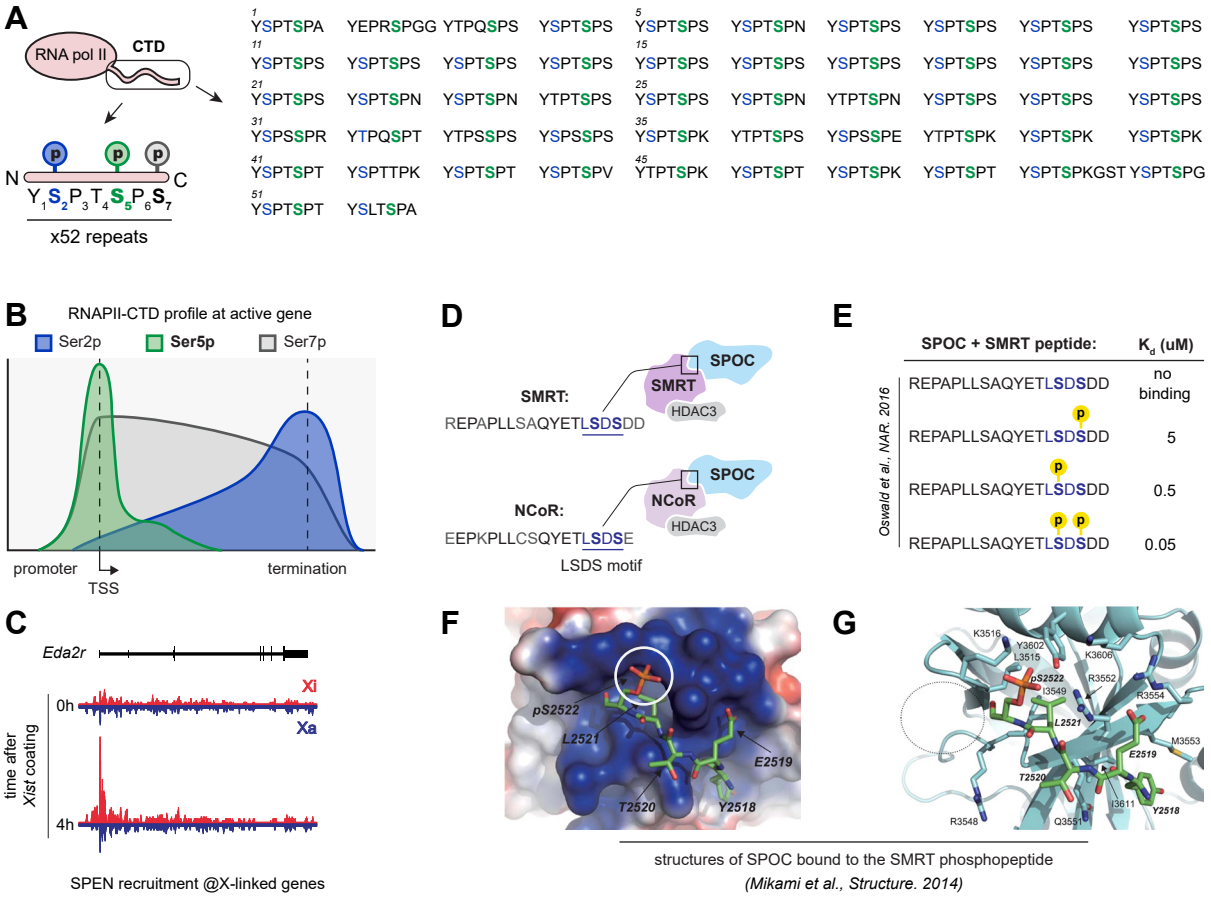


Figure 7.1 – Evidence towards a possible interaction between SPOC and the phosphorylated C-terminal domain of RNA polymerase II. (A) Scheme showing the amino acid sequence of the RNAPII-CTD, comprised of 52 heptapeptide repeats. The consensus sequence of the repeated heptapeptide is shown at the bottom left of the panel. The Serines at position 2 (bold blue), 5 (bold green) and 7 (bold black) are readily phosphorylated *in vivo*. **(B)** Typical average ChIP profiles of RNAPII phosphorylated at Ser2 (blue), Ser5 (green) and Ser7 (black) of its CTD shows the intricate link between the transcription cycle and CTD phosphorylation. *TSS: transcription start site.* **(C)** CUT&RUN profiling of *Xist*-mediated SPEN recruitment to X-linked chromatin (the *Eda2r* gene is shown as an example). **(D)** Scheme showing the amino-acid sequence of the SPOC-interacting regions of SMRT (top) and NCoR (bottom). Direct binding with SPOC occurs through the conserved LSDS motif (shown in purple). The two serines within this motif (shown in bold) are readily phosphorylated *in vivo*. **(E)** Table showing the binding affinities measured between SPOC and several SMRT peptides differing by their level of Serine phosphorylation (values taken directly from [Oswald et al., 2016]). **(F)** Electrostatic potential and **(G)** ribbon representations of the SMRT binding site of the SPOC domain highlight how the phosphorylated SMRT peptide (green licorice) fits into a basic patch of SPOC, where the SMRT phosphoserine (white circle) interacts with SPOC lysine and arginine side chains. *Images shown in **(F)** and **(G)** were taken directly from [Mikami et al., 2014]*

7.1.2 SPOC interaction with NCoR/SMRT is phospho-dependent

The direct interaction between SPOC and NCoR/SMRT occurs at a highly conserved LSDS motif found within the C-terminus of NCoR and SMRT (**Figure 7.1D**) [Ariyoshi and Schwabe, 2003]. Remarkably, this interaction can *only* occur when at least one of the two Serine residues within the LSDS motif of NCoR/SMRT are phosphorylated (**Figure 7.1E**) [Oswald et al., 2016].

The structural basis for this phospho-dependent interaction has been resolved using NMR [Mikami et al., 2014], revealing that the first phosphoserine of the LSDS motif fits tightly into a basic patch exposed at the surface of SPOC (**Figure 7.1F**), engaging with key arginine and lysine side chains of SPOC (**Figure 7.1G**). Altogether, these biochemical and structural studies have highlighted that SPEN's SPOC domain functions as a phosphoserine-binding unit.

Taken together with the observation that SPEN binding on the X chromosome closely ressembles that of Ser5P-RNAPII (**Figure 7.1B** and **C**), I hypothesized that SPEN could similarly interact directly with RNA-polymerase II specifically when its CTD is Ser5 phosphorylated.

To test this hypothesis, I started a collaboration with Christoph Müller's lab at EMBL Heidelberg (where the Heard lab had moved by then). Christoph's lab has world-renowned expertise in using structural biology coupled to biophysical and biochemical tools to dissect the molecular mechanisms of transcription.

All the experimental work presented in this chapter has only been made possible through teaming up with Brice Murciano from Christoph's lab, who also operates the crystallization facility at EMBL Heidelberg.

7.2 Bacterial expression and purification of the SPOC domain

To test a potential direct interaction between SPOC and RNAPII phosphorylated at Ser5 of its CTD, we first attempted to purify murine SPEN's SPOC domain using a bacterial expression system. I designed 5 different constructs (listed in **Table 7.1**), all of which encode protein fragments fully encompassing the SPOC domain (**Figure 7.2A**) as well as a varying number of SPEN amino acids upstream of SPOC (**Figure 7.2B**).

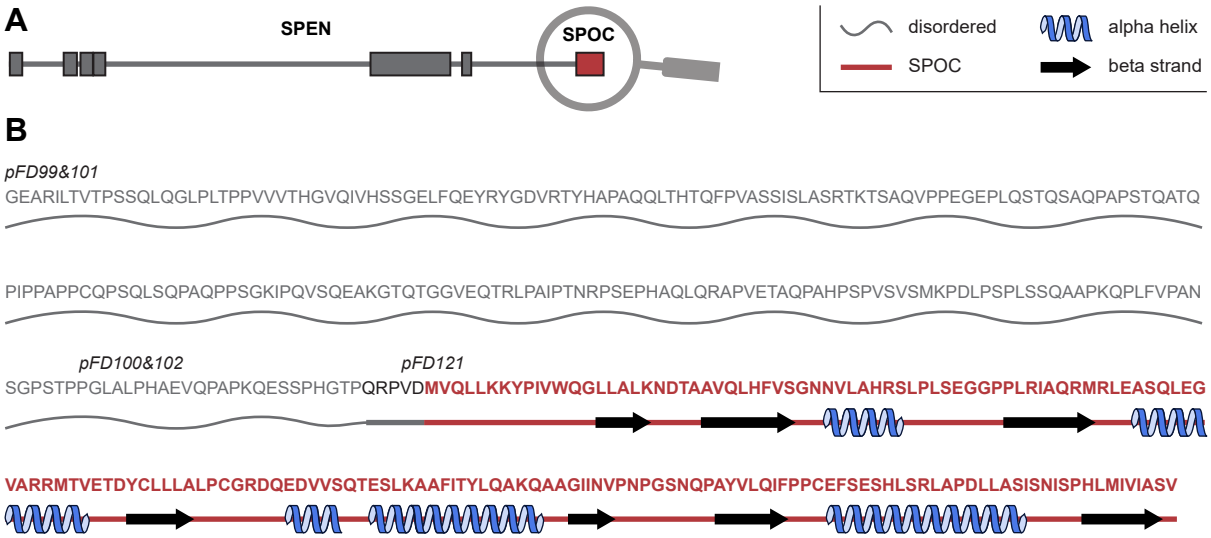


Figure 7.2 – Scheme depicting the different SPOC fragments tested for crystallization. **(A)** Schematic showing the whole SPEN protein, with a focus on the last 407 amino acids, which encompass the SPOC domain. **(B)** Amino acid sequence corresponding to the last 407 amino acids of SPEN. The names of each SPOC expression plasmids (shown above the sequence, see **Table 7.1**) are positioned relative to the start of the protein fragments they encode. Secondary structures – predicted with *Phyre2* [Kelley et al., 2015]– are represented below the sequence. Amino acids shown in grey map to predicted disordered regions, while amino acids shown in bold red map to the SPOC domain.

Each fragment was tagged with 6xHis-only or dual 6xHis-SUMO3 tags, with the 6xHis moiety being used for affinity purification, and the SUMO3 moiety allegedly enhancing protein expression and solubility. Given that our SPOC-encoding DNA fragments were not codon-optimized for bacteria, all these constructs were expressed from the Rosetta 2(DE3)pLysS bacterial strain, which harbors a pRARE plasmid expressing several tRNAs rarely found in *E. coli*.

construct	backbone	affinity tag	a.a.	size (kDa)	pI	yielded crystals
pFD99	pETM11	6xHis-SUMO3	407	43.3	7.3	no
pFD100	pETM11	6xHis-SUMO3	196	21.2	6.7	no
pFD101	pETM11	6xHis	407	43.3	7.3	no
pFD102	pETM11	6xHis	196	21.2	6.7	no
pFD121	pETM11	6xHis-SUMO3	169	18.4	6.5	**yes**

Table 7.1 – List of SPOC constructs tested for purification and crystallization. All constructs were cloned into a pETM11 backbone, and SPOC expression was induced at 18°C in *Rosetta 2(DE3)pLysS* bacterial cells (see **section 3.20**). *a.a: amino acids, pI: isoelectric point*

Adapting the conditions originally used to purify human SPOC [Ariyoshi and Schwabe, 2003], our optimized protocol (see **section 3.20**) resulted in great yields of highly purified SPOC (**Figure 7.3**).

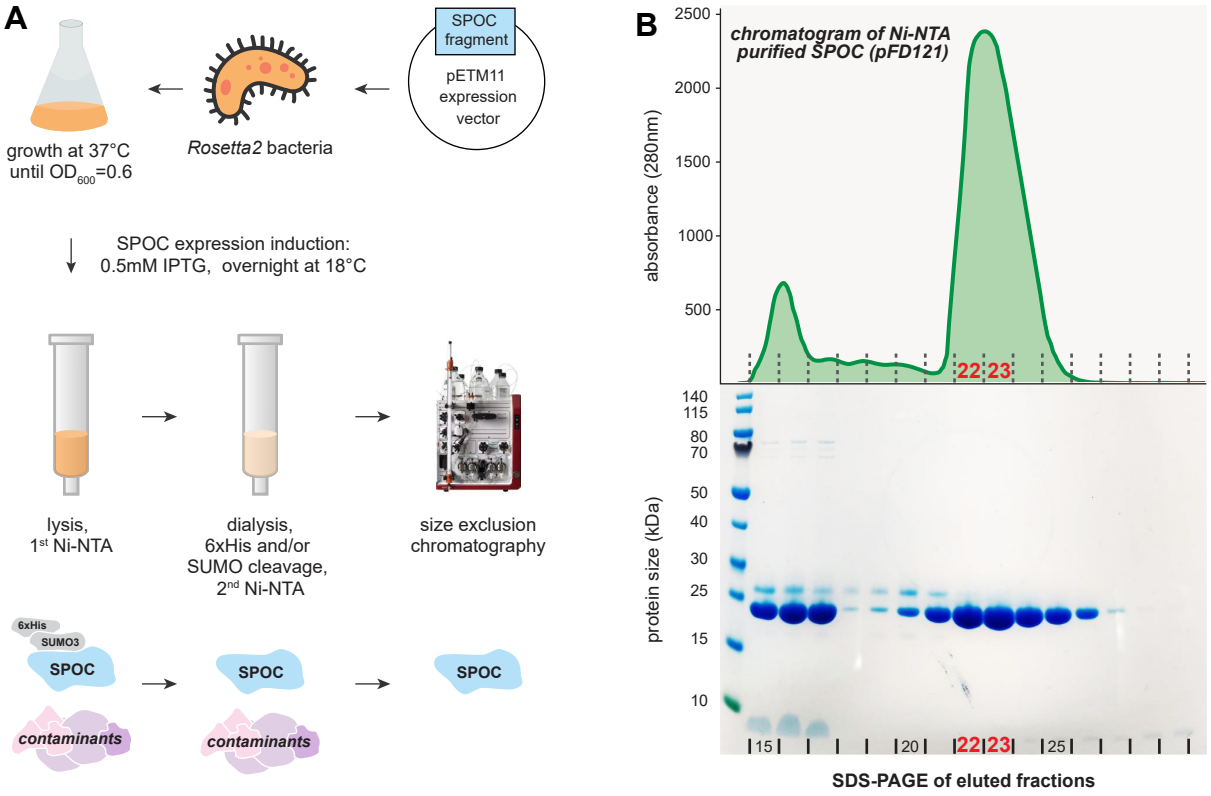


Figure 7.3 – Purification of mouse SPEN's SPOC domain expressed in bacterial cells. **(A)** The different SPOC constructs (see **Table 7.1**) were transformed in the *Rosetta2* bacterial strain, enabling enhanced expression of eukaryotic proteins which have not been codon-optimized for bacteria. Liquid bacterial cultures were grown at 37°C, in terrific broth, until log-phase was reached (OD_{600}=0.6), after which SPOC expression was induced overnight at 18°C with 0.5mM IPTG. SPOC was purified from bacterial lysates using Ni-NTA affinity-resin. SPOC was then cleaved-off of its 6xHis-SUMO3 tag with a SUMO protease. One step of size exclusion chromatography was sufficient to collect great yields of highly-purified SPOC. **(B)** Chromatrogram (top panel) obtained during size exclusion chromatography of Ni-NTA purified SPOC (from the pFD121 construct, see **Table 7.1**) and the resulting SDS-PAGE analysis of eluted fractions (bottom panel). Fractions 22/23 were used for ITC and crystallization experiments.

7.3 Structure of mouse SPOC at 1.2Å resolution

We next attempted to crystallize and resolve the structure of our purified SPOC protein. Mouse and human SPEN's SPOC domains share 99% sequence homology (**Figure 7.4A**), and thus, we aimed to ensure that the 3D-structure of our purified *mouse* SPOC protein was similar to the published structure of *human* SPOC [Ariyoshi and Schwabe, 2003].

Following several rounds of crystallization screening, we were successful at obtaining diffracting SPOC crystals (**Figure 7.4B**). Importantly, only the minimal SPOC fragment (pFD121 construct, see **Table 7.1**) yielded good diffracting crystals.

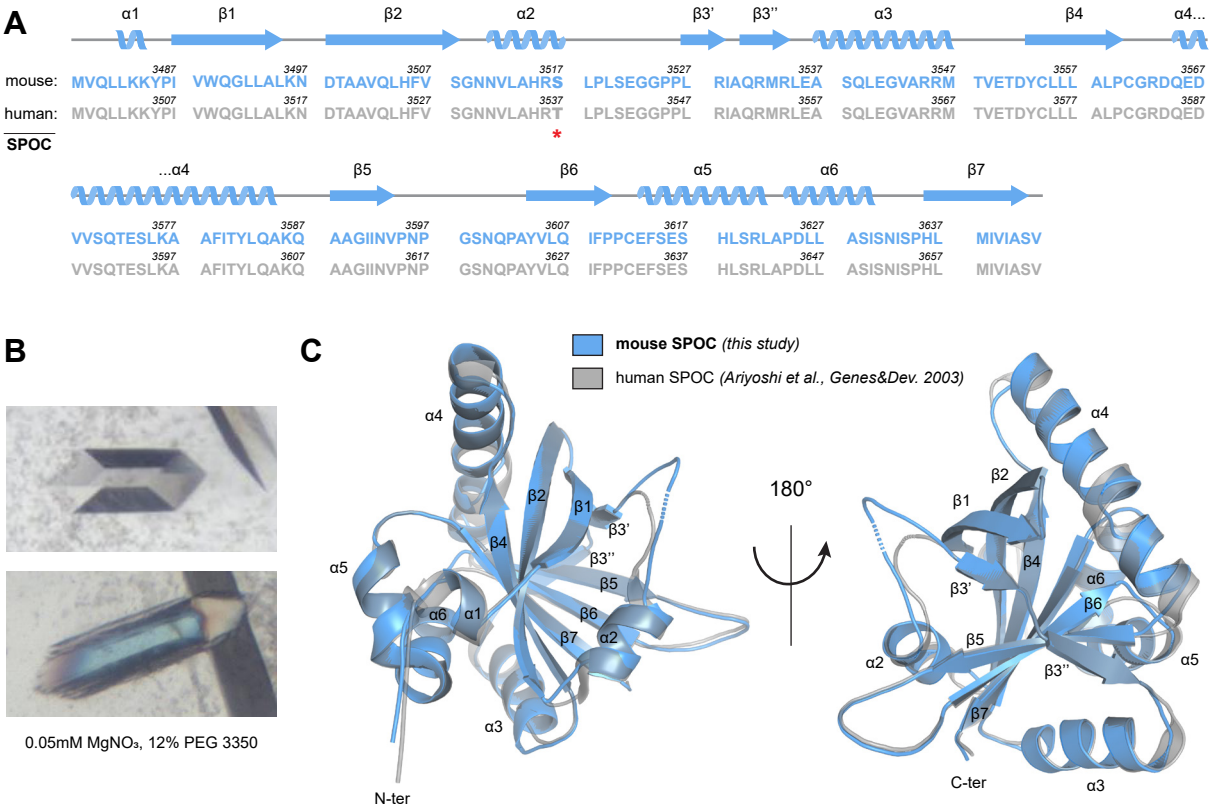


Figure 7.4 – Crystallization and structural determination of mouse SPOC. (A) Amino acid sequences of mouse (blue) and human (grey) SPOC domains. Amino acids are numbered relative to the entire SPEN protein sequence. Both SPOC sequences share 99% homology, with a Serine at position 3517 in mouse being substituted by a Threonine at position 3537 in human. The position of α-helices and β-sheets identified in our SPOC structure is indicated above the sequence. **(B)** Pictures showing typical SPOC crystals. **(C)** Cartoon representation of a structural alignment between mouse SPOC (blue, resolved in this study), and human SPOC (grey, resolved in [Ariyoshi and Schwabe, 2003]).

We were able to solve the structure of mouse SPOC (**Table 7.2** and **Figure 7.4C**), using molecular replacement with human SPOC (PDB 1OW1, [Ariyoshi and Schwabe, 2003]) as a template. We determined SPOC structure at a resolution of 1.2Å, improving it by 0.6Å compared to [Ariyoshi and Schwabe, 2003].

Importantly, our mouse SPOC structure aligned very well with the human SPOC structure, with a *r*oot-*m*ean-*s*quare *d*eviation (RMSD) of atomic positions equal to 0.119Å (**Figure 7.4C**). Similarly to human SPOC [Ariyoshi and Schwabe, 2003], mouse SPOC is characterized by a seven-stranded β-barrel (identified as β1-β7 on **Figure 7.4C**) surrounded by six α-helices (identified as α1-α6 on **Figure 7.4C**).

Parameter	
Resolution range (Å)	30.78 - 1.2 (1.243 - 1.2)
Space group	P $2_12_12_1$
Unit cell (Å)	a=41.604, b=61.502, c=68.437 (α,β,γ=90°)

Table continued on next page

Table 7.2 – *Continued from previous page*

Parameter	
Molecules (a.u)	1
Total reflections	678039 (52895)
Unique reflections	52459 (4357)
Multiplicity	12.9 (12.1)
Completeness (%)	94.28 (79.66)
Mean I/σ(I)	34.04 (13.16)
Wilson B-factor	12.48
R_{merge}	0.04853 (0.1764)
R_{meas}	0.05086 (0.1842)
R_{pim}	0.01475 (0.0517)
$CC_{1/2}$	0.998 (0.994)
Reflections used in refinement	52454 (4355)
Reflections used for R_{free}	1993 (167)
R_{work}/R_{free}	0.1439 (0.1231) / 0.1635 (0.1728)
CC_{work}	0.952 (0.977)
CC_{free}	0.953 (0.954)
Ramachandran favored (%)	97.56
Ramachandran allowed (%)	2.44
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	4.22

Table 7.2 – Crystallization data and refinement statistics for SPOC. Statistics for the highest resolution shell are shown in parentheses.

7.4 SPOC interacts with the CTD of RNAPII specifically when Ser5 is phosphorylated

Altogether, these previous results demonstrate that our conditions for SPOC purification faithfully preserve its 3D-structure, and hence can be subsequently used to investigate a potential interaction with the CTD of RNAPII. To assay such interaction, we first used *i*sothermal *t*itration *c*alorimetry (ITC, **Figure 7.5**).

ITC is one of the most sensitive and quantitative analytical methods to characterize protein-protein interactions, notably to identify the dissociation constant (K_d) and the stoichiometry. In our case, ITC measures (under constant pressure and temperature) the heat exhanges resulting from exposing a SPOC solution (of known concentration) to increasing quantities of a synthetic CTD peptide.

ITC revealed no significant binding between SPOC and synthetic unphosphorylated CTD peptides, containing either two (**Figure 7.5A1**) or four (**Figure 7.5A2**) repeats

of the consensus CTD heptapeptide. Similarly, no binding was detected between SPOC and Ser2- (**Figure 7.5B1**) or Ser7- (**Figure 7.5D1**) phosphorylated CTD peptides , regardless of the number of heptapeptide repeats (**Figure 7.5B2** and **D2**).

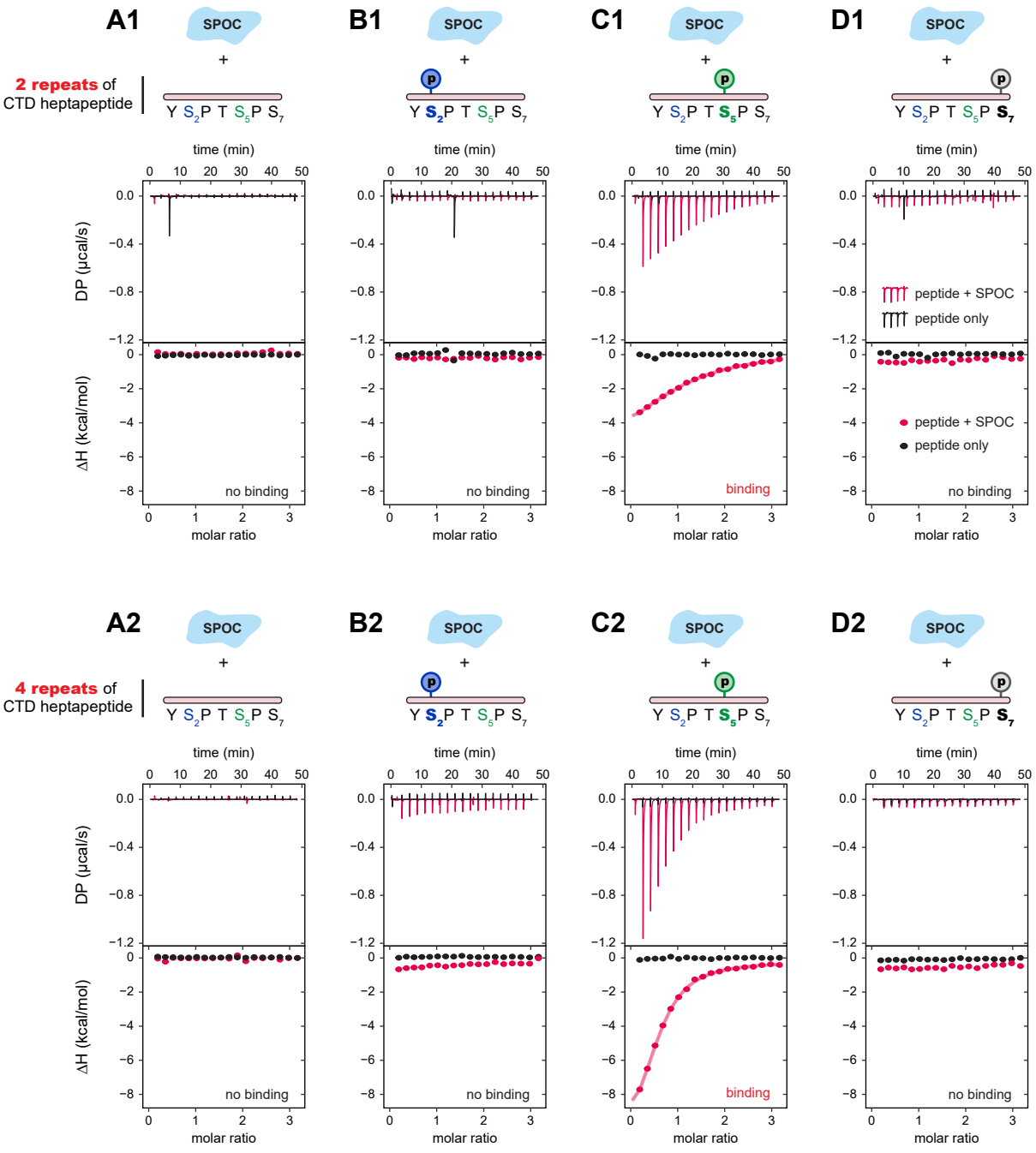


Figure 7.5 – Isothermal titration calorimetry reveals a direct and specific interaction between SPOC and the Ser5-phosphorylated CTD of RNAPII. Representative thermograms (all scalred identically) resulting from ITC measurements of SPOC binding to **(A)** unphosphorylated, **(B)** Ser2-, **(C)** Ser5- or **(D)** Ser7-phosphorylated peptides containing **(1)** two or **(2)** four repeats of the RNAPII-CTD heptapeptidic sequence. Each peptide was titrated against SPOC (data shown in red) or buffer as a control (data shown in black). When binding is observed, the statistical fitting – resulting from the one-site binding model used to derive the thermodynamic parameters reported in **Table 7.3** – is shown as a sigmoid curve. All ITC experiments were repeated at least three times.

Remarkably however, direct binding (albeit moderate, K_d≈10uM, **Table 7.3**) was observed between SPOC and *Ser5 phosphorylated* CTD peptides (**Figure 7.5C1**). Although the measured affinity did not differ whether we used two or four repeats of the Ser5 phosphorylated consensus heptapeptide (**Table 7.3**), we noticed however that greater heat (≈ twice as much) was released when using the 4-repeat peptide in lieu of the 2-repeat peptide (**Figure 7.5C1** and **C2**). This result suggests that there are twice as many SPOC molecules bound on the 4-repeat peptide as there are on the 2-repeat pepide.

Altogether, these results demonstrate that SPOC interacts directly with the CTD domain of RNAPII, only when it is phosphorylated specifically on Serine 5. Furthermore, they indicate that several SPOC molecules can bind simultaneously to the same CTD molecule so long as it contains a sufficient number of heptapeptide repeats. This interaction has important implications for how SPEN could silence transcription during XCI, and will be further discussed in **subsubsection 9.2.2.3**.

phosphorylation	**repeats**	**K_d (uM)**	**N (sites)**	**ΔH (kcal/mol)**	**ΔG (kcal/mol)**
unphosphorylated	2	n.b.d.			
unphosphorylated	4	n.b.d.			
Ser2	2	n.b.d.			
Ser2	4	n.b.d.			
Ser5	2	10.2±3.3	1.01±0.06	-3.2±0.3	-6.7
Ser5	4	13.4±0.8	0.58±0.01	-11.5±0.4	-6.54
Ser7	2	n.b.d.			
Ser7	4	n.b.d.			
Ser2 and Ser7	2	n.b.d.			
Ser5 and Ser2	2	4.0±0.8	1.33±0.03	-4.0±0.2	-7.24
Ser5 and Ser7	2	22.2±4.2	1.30±0.04	-3.7±0.3	-6.24

Table 7.3 – Summary of ITC data resulting from titrating different RNAPII-CTD phosphopeptides against SPOC. *n.b.d.: no binding detected*

7.5 Ser2/7 phosphorylations do not alter SPOC binding to the Ser5P-CTD of RNAPII

Throughout the transcription process, the CTD of RNA polymerase II is not always marked by only one type of phosphorylation (i.e. Ser2-only, Ser5-only, Ser7-only...). Indeed, several CTD phosphorylation-states are found to co-occur within the same RNAPII molecule [Schüller et al., 2016]. For example, simultaneous phosphorylation at Ser2 and Ser5 of the RNAPII-CTD is observed at the beginning of the elongation phase. This doubly-phosphorylated state directly recruits SET2 (which only binds to Ser2P/Ser5P

CTD), a histone lysine methyltransferase which catalyses H3K36me3 and plays a crucial role in transcription elongation [Kizer et al., 2005; Vojnic et al., 2006].

Hence, I questioned whether the additional presence of Ser2- or Ser7-phosphorylation alters (positively or negatively) the ability of SPOC to interact with Ser5P-CTD, so as to shed light on which RNAPII molecules SPOC has the ability to engage with. To address this point, binding between doubly phosphorylated CTD peptides (Ser2P/Ser7P, Ser2P/Ser5P, Ser5P/Ser7P) and SPOC was assayed using ITC.

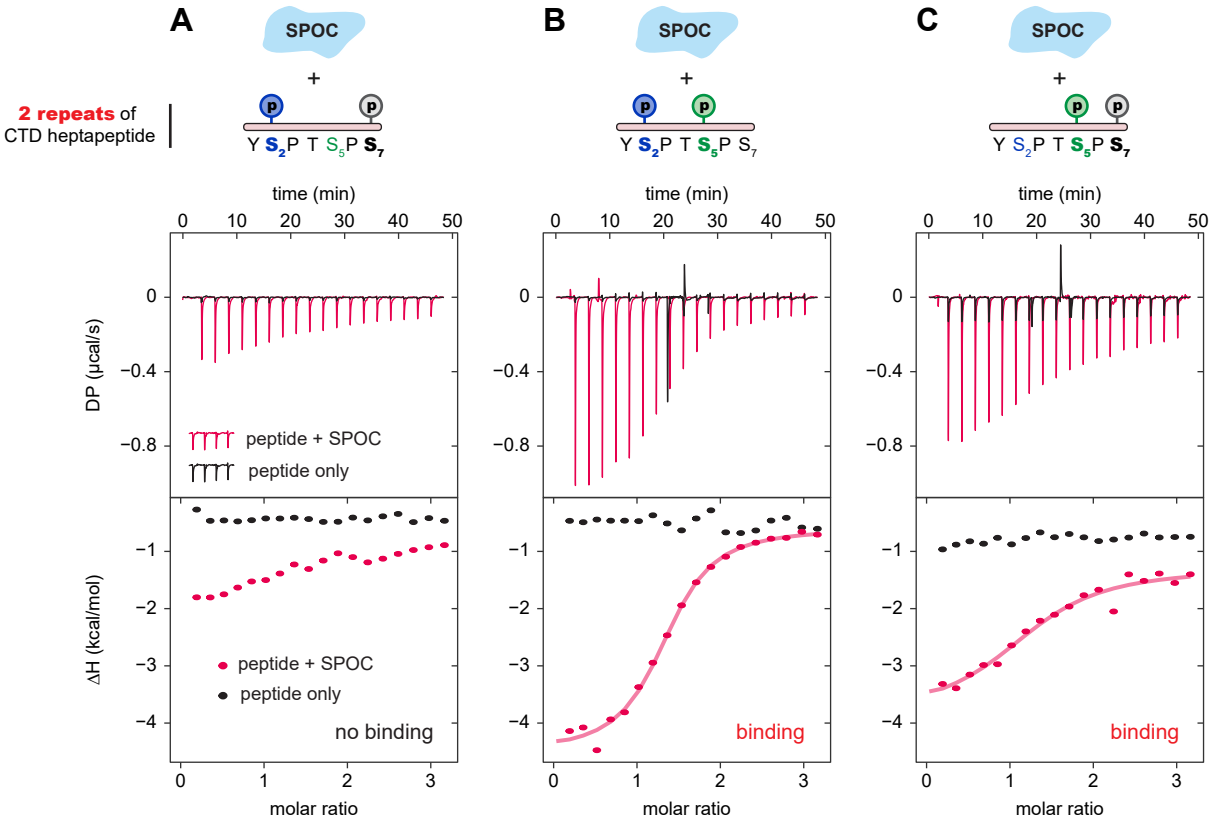


Figure 7.6 – Isothermal titration calorimetry reveals SPOC is able to bind the Ser5-phosphorylated CTD of RNAPII even when other Serines are phosphorylated. Representative thermograms (all scaled identically) resulting from ITC measurements of SPOC binding to **(A)** Ser2Ser7-, **(B)** Ser2Ser5- or **(C)** Ser5Ser7-doubly phosphorylated peptides containing two repeats of the RNAPII-CTD heptapeptidic sequence. Each peptide was titrated against SPOC (data shown in red) or buffer as a control (data shown in black). When binding is observed, the statistical fitting – resulting from the one-site binding model used to derive the thermodynamic parameters reported in **Table 7.3** – is shown as a sigmoid curve. All ITC experiments were repeated at least three times.

Direct binding was observed between SPOC and either Ser2Ser5- or Ser5Ser7-doubly phosphorylated peptides (**Figure 7.6B**, **C** and **Table 7.3**). Importantly, the binding affinities for both peptides did not differ significantly from that observed with CTD peptides phosphorylated only at Ser5 (**Table 7.3**). Taken together, these results show

that SPOC binds Ser5P-CTD irrespectively of the phosphorylation environment present at neighboring residues, suggesting that SPOC interacts with a larger fraction of RNAPII than that which is strictly phosphorylated at Serine 5.

Surprisingly, although no significant binding was detected between SPOC and the Ser2Ser7-doubly phosphorylated peptide (as would be expected from the results obtained in **section 7.4**), a heat response profile differing from that of the control one was observed (**Figure 7.6A**). This could be interpreted as a very weak and transient interaction between SPOC and this peptide, which can occur only when several residues within the heptapeptide sequence are phosphorylated. This interaction is most likely not biologically relevant as no K_d could be computed following ITC analysis.

7.6 The SPOC/Ser5P-CTD complex at 1.8Å resolution

To further capture the molecular basis of such specific SPOC/Ser5P-CTD interaction, we next resolved the structure of the SPOC/Ser5P-CTD complex using X-ray crystallography. For this purpose, we incubated purified SPOC with either a 2- or 4-fold molar excess of the Ser5P-CTD peptide (containing four heptapeptide repeats), and subjected the complex to high-throughput crystallization screening (see **section 3.24**).

id	screen	ratio	salt	buffer	precipitant
1	BCS	4		0.1M PO_4 pH5.5	20% PEG Smear High
2	BCS	2		0.1M PO_4 pH5.5	22% PEG Smear High
3	BCS	4	0.15M NaCl		28% PEG Smear Medium
4	BCS	4	0.1M $NaKHPO_4$ pH5.5, 0.1M RbCl	0.1M NaCitrate pH5.5	25% PEG Smear Medium
5	BCS	2	0.15M $Mg(CH_3COO)_2$*4H20	0.1M NaCitrate pH5.6	20% PEG Smear Broad
6	BCS	4	0.2M $NaKHPO_4$ pH7.5	0.1M HEPES pH7.5	22.5% PEG Smear Medium, 10% Glycerol
7	PCS	4		0.1M NaCacodylate pH6	15% PEG 4000
8	PCS	4	0.2M KI	0.1M MES pH6.5	25% PEG 4000
9	PCS	4		0.1M NaCitrate pH5.5	15% PEG 6000
10	PCS	2	0.2M $(NH_4)_2SO_4$	0.1M MES pH6.5	20% PEG 8000
11	*EMBL*	4		0.1M NaCitrate pH5.6	18% MPEG 5000
12	*EMBL*	2		0.1M HEPES pH7	18% MPEG 5000
13	*EMBL*	2		0.1M MOPS pH7.2	18% MPEG 5000
14	*EMBL*	4		0.1M TRIS pH8	18% MPEG 5000
15	*EMBL*	2		0.1M $NaKHPO_4$ pH6.3	15% PEG 6000
16	*EMBL*	2		0.1M MES pH6.5	15% PEG 6000
17	*EMBL*	4		0.1M BIS-TRIS pH6.2	15% PEG 8000
18	*EMBL*	4		0.1M NaCacodylate pH6	15% PEG 8000
19	PEG1	2		0.1M MES pH6.5	25% PEG 1000
20	PEG1	4		0.1M MES pH6.5	25% PEG 4000
21	PEG1	4		0.1M MES pH6.5	20% PEG 10000
22	PEG1	4		0.1M HEPES pH7.5	25% PEG 3000
23	PEG1	4		0.1M HEPES pH7.5	20% PEG 10000
24	PEG1	2	0.2M NaF		20% PEG 3350
25	PEG1	4	0.2M NaF		20% PEG 3350
26	PEG1	4	0.2M NaCl		20% PEG 3350

Table continued on next page

Table 7.4 – *Continued from previous page*

id	screen	ratio	salt	buffer	precipitant
27	PEG1	2	0.2M $Mg(CH_3COO)_2$		20% PEG 3350
28	PEG1	4	0.2M Na_2CH_3COO		20% PEG 3350
29	PEG1	2	0.2M $Ca(CH_3COO)_2$		20% PEG 3350
30	PEG1	2	0.2M Na_2Tartrate		20% PEG 3350
31	PEG2	4	0.1M Na_2CH_3COO	0.1M MES pH6.5	30% PEG 2000 MME
32	PACT	4		0.1M SPG buffer pH4	25% PEG 1500
33	PACT	4		0.1M SPG buffer pH5	25% PEG 1500
34	PACT	2		0.1M MIB buffer pH6	25% PEG 1500
35	PACT	4	0.2M $MgCl_2$	0.1M MES pH6	20% PEG 6000
36	PACT	2	0.2M NaCl	0.1M HEPES pH7	20% PEG 6000
37	PACT	4	0.2M LiCl	0.1M HEPES pH7	20% PEG 6000
38	PACT	4		0.1M MMT buffer pH4	25% PEG 1500
39	PACT	4		0.1M MMT buffer pH6	25% PEG 1500
40	**PACT**	**2**	**0.2M NH_4Cl**	**0.1M TRIS pH8**	**20% PEG 6000**
41	PACT	2	0.2M NaF		20% PEG 3350
42	PACT	4	0.2M NaBr		20% PEG 3350

Table 7.4 – Crystallization conditions for crystals shown in Figure 7.7. Each crystal is designated with its associated number as in **Figure 7.7**. The ratio column indicates the Ser5P-CTD peptide to SPOC molar ratio. Data collected from X-ray diffraction of crystal 40 (shown in bold) was used to resolve the structure of the SPOC/Ser5P-CTD complex.

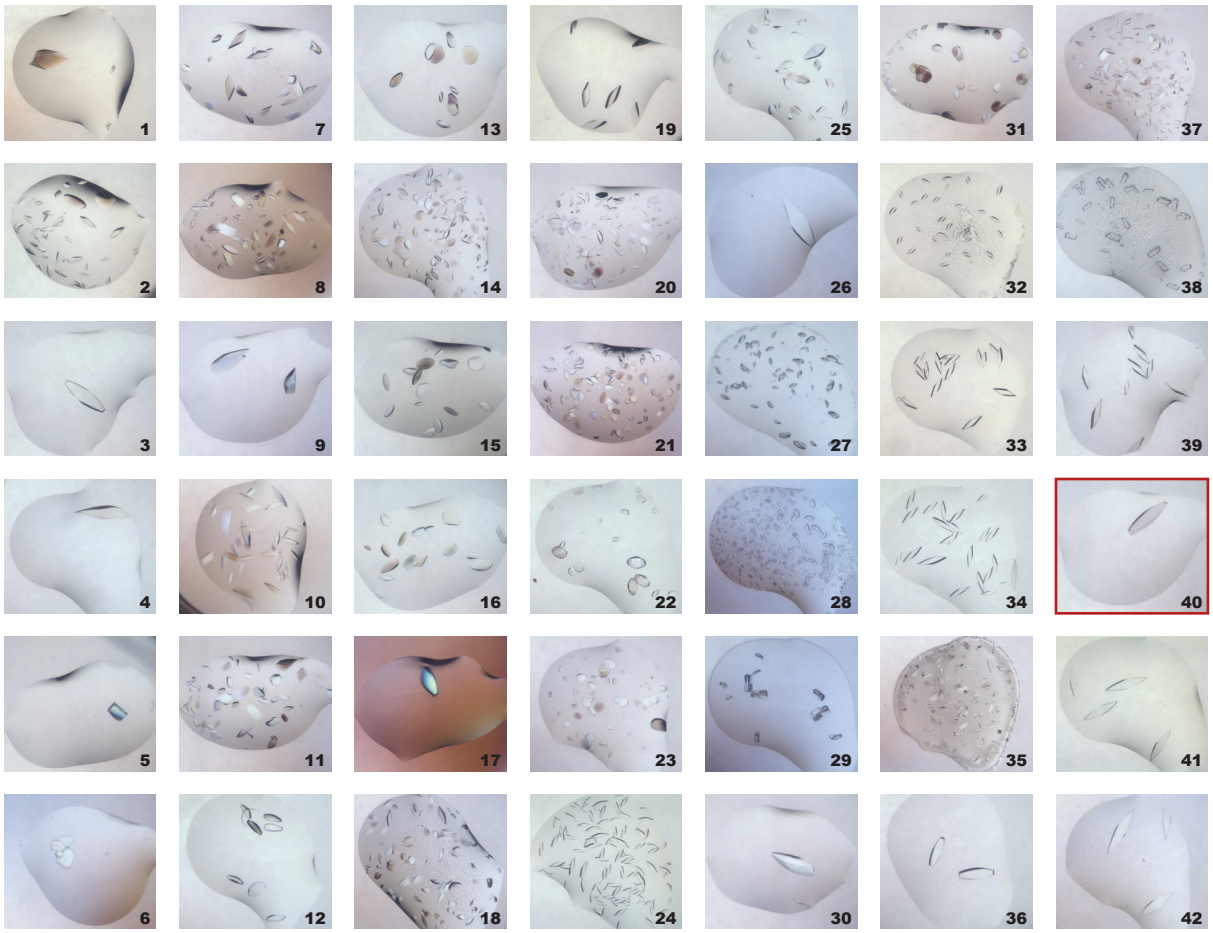


Figure 7.7 – Pictures of SPOC/Ser5P-CTD peptide cocrystals grown across several conditions. Magnification is identical across all pictures. Conditions associated with each crystal are to be found in **Table 7.4**. Data collected from X-ray diffraction of crystal 40 (red-bordered image) was used to resolve the structure of the SPOC/Ser5P-CTD complex.

Remarkably, crystals grew within a few hours (up to within one day), under several tested conditions (see **Figure 7.7** for examples of crystals grown and **Table 7.4** for their associated conditions). We were able to collect several datasets from these diffracting crystals, which allowed us to solve the molecular structure of the SPOC/Ser5P-CTD complex down to 1.8Å resolution (**Table 7.5**).

Parameter	
Resolution range (Å)	41.75 - 1.8 (1.864 - 1.8)
Space group	P $2_12_12_1$
Unit cell (Å)	a=41.302, b=119.788, c=133.646 (α,β,γ=90°)
Molecules (a.u)	1
Total reflections	805364 (78455)
Unique reflections	62477 (6141)
Multiplicity	12.9 (12.8)
Completeness (%)	98.42 (93.96)
Mean I/σ(I)	22.52 (1.16)
Wilson B-factor	34.60
R_{merge}	0.04475 (2.385)
R_{meas}	0.0467 (2.483)
R_{pim}	0.01318 (0.6868))
$CC_{1/2}$	1 (0.667)
CC*	1 (0.895)
Reflections used in refinement	61514 (5771)
Reflections used for R_{free}	1418 (130)
R_{work}/R_{free}	0.2051 (0.2882) / 0.2333 (0.3332)
CC_{work}	0.927 (0.669)
CC_{free}	0.988 (0.627)
Ramachandran favored (%)	97.40
Ramachandran allowed (%)	2.60
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	4.60

Table 7.5 – Crystallization data and refinement statistics for the SPOC/Ser5P-CTD complex. Statistics for the highest resolution shell are shown in parentheses.

7.7 The basic patch of SPOC mediates direct interaction with the Ser5P-CTD of RNAPII

The 3D-structure of the complex shows that binding of the Ser5P-CTD peptide to SPOC occurs at a surface surrounding the β3 strand (**Figure 7.8**), which corresponds to the only basic patch of SPOC (**Figure 7.9**).

Structural alignment between the SPOC domain alone and the SPOC domain in

complex with Ser5P-CTD yields an RMSD value of 0.676Å, indicating that SPOC does not undergo significant structural changes upon binding the CTD molecule. Similarly, no structural changes in SPOC were reported upon interacting with pSMRT [Mikami et al., 2014].

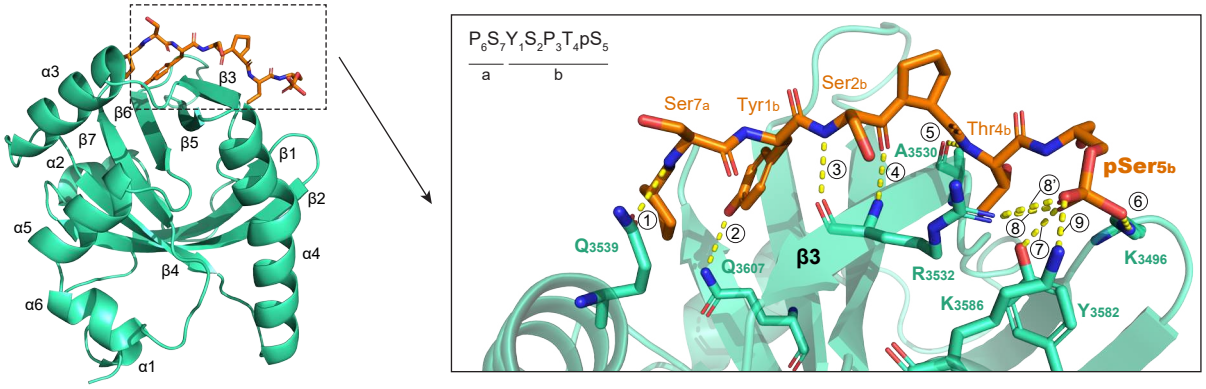


Figure 7.8 – Cartoon representation of the SPOC/Ser5P-CTD complex. Intermolecular hydrogen bonds appear as dotted yellow lines, and are numbered (black circles) as in **Table 7.6**. *green cartoon representation: SPOC, orange licorice representation: Ser5P-CTD peptide*

Although SPOC binds the CTD peptide only when the latter is phosphorylated on Ser5 (see **section 7.4**), structural analysis reveals that SPOC forms several intermolecular interactions with CTD residues located upstream (i.e. N-terminally) of the Ser5-phosphoserine including Tyr_1, Ser_2 and Thr_4, as well as Ser_7 of the preceding heptapeptide (**Figure 7.8** and **Table 7.6**). This observation is reminiscent of the SPOC/pSMRT interaction in which several residues of the pSMRT phosphopeptide, located upstream of the phosphoserine, are required for binding of pSMRT to SPOC [Mikami et al., 2014].

id	SPOC residue	functional group	CTD residue	functional group
1	Q_{3539} ($\alpha 3$)	α-amide (C=O dipole, backbone)	Ser_{7a}	α-amide (N-H dipole, backbone)
2	Q_{3607} ($\beta 6$)	γ-amide (N-H dipole)	Tyr_{1b}	β-hydroxyphenyl
3	R_{3532} ($\beta 3$")	α-amide (C=O dipole, backbone)	Ser_{2b}	α-amide (N-H dipole, backbone)
4	R_{3532} ($\beta 3$")	α-amide (N-H dipole, backbone)	Ser_{2b}	α-amide (C=O dipole, backbone)
5	A_{3530} ($loop_{\beta 3'\text{-}\beta 3"}$)	α-amide (C=O dipole, backbone)	Thr_{4b}	α-amide (N-H dipole, backbone)
6	**K_{3496}** ($loop_{\beta 1\text{-}\beta 2}$)	ϵ-amino	$pSer_{5b}$	β-phosphonooxy (O^{1A})
7	**Y_{3582}** ($\alpha 4$)	β-hydroxyphenyl	$pSer_{5b}$	β-phosphonooxy (O^{2A})
8	**R_{3532}** ($\beta 3$")	δ-guanidino	$pSer_{5b}$	β-phosphonooxy (O^{2A})
8'	**R_{3532}** ($\beta 3$")	δ-guanidino	$pSer_{5b}$	β-phosphonooxy (O^{3A})
9	**K_{3586}** ($loop_{\alpha 4\text{-}\beta 5}$)	ϵ-amino	$pSer_{5b}$	β-phosphonooxy (O^{3A})

Table 7.6 – Summary of intermolecular hydrogen bonds between SPOC and the Ser5P-CTD peptide. Hydrogen bonds are labeled according to **Figure 7.8**, and oxygen atoms within the phosphate group are named following IUPAC recommendations [Blackburn et al., 2017].

The Ser_5 phosphoserine is docked deeply into the basic patch of SPOC, with the phosphate group oriented inwards, toward the depth of a cavity formed by the K_{3496}, R_{3532} and K_{3586} SPOC residues (**Figure 7.8** and **Figure 7.9**). These residues are most likely critical for the SPOC specificity towards the Ser5P-CTD, as they are all involved in hydrogen bonds with the phosphonooxy group of the Ser_5 phosphoserine (**Figure 7.8** and **Figure 7.9**).

Remarkably, binding of phosphorylated SMRT (pSMRT) occurs within the same region of SPOC and involves the same SPOC residues (see **subsection 7.1.2** and [Mikami et al., 2014]), suggesting that NCoR/SMRT and the Ser5P-CTD of RNAPII may compete for interacting with SPOC *in vivo*.

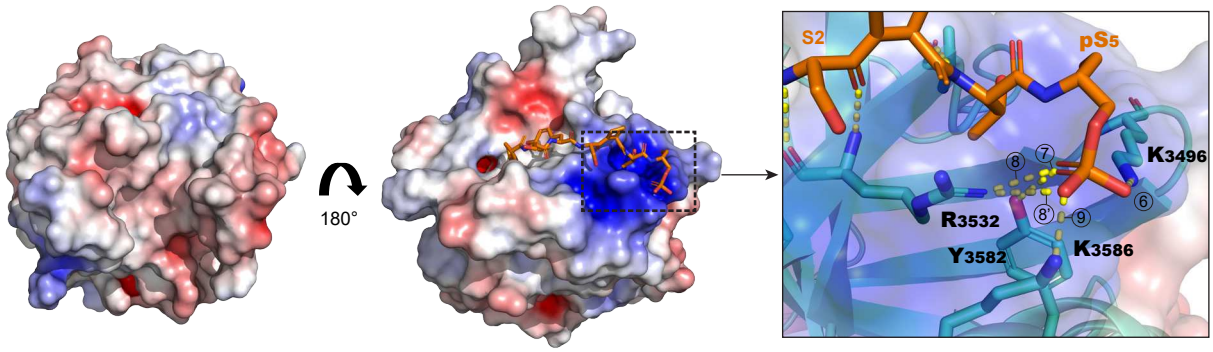


Figure 7.9 – Electrostatic map of the Ser5P-CTD binding site of the SPOC domain. Intermolecular hydrogen bonds appear as dotted yellow lines, and are numbered (black circles) as in **Table 7.6**. *electrostatic potential map/green cartoon representation: SPOC, orange licorice representation: Ser5P-CTD peptide*

Part IV

Discussion

Chapter 8

SPEN is the master effector of gene silencing during XCI

While several studies had shown that *Spen* knockdown leads to silencing defects at a few X-linked genes [Moindrot et al., 2015; Monfort et al., 2015; Chu et al., 2015; McHugh et al., 2015], none addressed the effect of SPEN loss on chromosome wide gene silencing. In this work, I show that SPEN is the master regulator of gene silencing initiation during XCI, both *in vitro* and *in vivo*. Importantly, *Xist* recruits SPEN immediately upon its upregulation, and SPEN remains associated with the *Xist* domain throughout the XCI process.

8.1 Most X-linked genes are silenced entirely through SPEN-dependent mechanisms.

Using SPEN-degron mESCs and RNAseq, I found that depleting SPEN throughout 24 hours of *Xist* upregulation results in dramatic defects in X-linked gene silencing, chromosome wide. Among the 382 X-linked genes monitored by RNAseq, 78% of them show no significant signs of *Xist*-mediated gene silencing upon SPEN depletion, demonstrating that the majority of genes on the X are fully dependent of SPEN for silencing.

Importantly, 18% of the 382 genes only show partial defects in *Xist*-mediated gene silencing upon SPEN depletion. On average, these genes are silenced by 50% (compared to a wild-type scenario), suggesting that their silencing partially relies on SPEN-dependent mechanisms, and that other mechanisms ensure that these genes are silenced to their full potential. Future studies could identify such mechanisms by setting up genetic screens

for XCI using some of these genes as their phenotypic reporters.

8.2 A few X-linked genes are silenced through SPEN-independent mechanisms.

While most genes (96%) show drastic gene silencing defects upon loss of SPEN, I observed that a few genes (16 in total) were fully silenced, independently of SPEN. Among these 16 SPEN-independent genes (**Table 8.1**):

- 5 were also silenced upon deletion of *Xist* A-repeat [Żylicz et al., 2019], suggesting that their silencing depends entirely on mechanisms independent of the SPEN/A-repeat axis. One such mechanism could be *Xist* B-repeat mediated recruitment of PRC1, which has recently been shown to play a role in gene silencing during XCI [Pintacuda et al., 2017; Nesterova et al., 2019].
- 11 failed to be silenced upon deletion of *Xist* A-repeat [Żylicz et al., 2019], suggesting that the A-repeat recruits SPEN-independent mechanisms to silence these few genes. Apart from SPEN, two other proteins, WTAP and RNF20, have been shown to co-precipitate with *Xist* RNA in an A-repeat dependent manner [Chu et al., 2015]. WTAP is a member of the RNA m^6A RNA methylation machinery, and has been suggested to play a role in *Xist*-mediated gene silencing (see **subsection 1.13.3** and [Moindrot et al., 2015; Patil et al., 2016]). RNF20 is an E3 ligase which mediates mono-ubiquitylation of histone H2B at Lysine 120 (H2B120ub1) [Zhu et al., 2005; Kim et al., 2005]. Given that RNF20 is involved in direct transcriptional activation [Zhu et al., 2005; Kim et al., 2005], it is unlikely that this factor is involved in mediating gene silencing during XCI.

Remarkably however, 3 genes were dependent on SPEN, but *not* on the A-repeat for their silencing (**Table 8.1**). This intriguing result suggests that some residual SPEN-recruitment by *Xist* may still happen independently of the A-repeat, supporting silencing of a very minor fraction of genes. Given that the number of such genes is very low, this observation needs to be further reproduced in several other SPEN-degron clones and under non-artificially induced *Xist* expression (i.e. in the context of cellular differentiation).

category	**gene names**
SPEN-independent only	*3830417A13Rik, Cdkl5, Cnksr2, Hmgn5, Mdm4-ps, Nap1l2, Pcsk1n, Rpgr, S100g, Sh3bgrl, Tmem29*
A-repeat independent only	*Fhl1, Phka1, Prrg3*
SPEN/A-repeat independent	***Amot, Col4a5, Fgf13, Fndc3c1, Rnf128***

Table 8.1 – List of SPEN and/or A-repeat-independent genes.

8.3 SPEN is critical for silencing highly expressed genes.

Nevertheless, our analysis reveals that the vast majority of genes are silenced through the SPEN/A-repeat axis. Importantly, we found that the expression level of these genes (prior to XCI) is highly correlated with how much SPEN is required for their silencing. Nesterova et al. also reported that highly expressed genes seem to rely more on SPEN for silencing than less expressed genes, and further proposed a model wherein Polycomb-dependent mechanisms would be involved in silencing these less expressed genes [Nesterova et al., 2019].

These observations are consistent with our CUT&RUN data, which demonstrates that SPEN is preferentially recruited to sites of active transcription. Given that SPEN is quickly recruited to the inactivating X (see **subsection 8.6.3**), this may be a way to guarantee efficient silencing of highly expressed genes, whose dosage might be more important to compensate for. It also highlights SPEN as a potent silencer, capable to repress highly expressed genes.

8.4 SPEN is essential for imprinted XCI *in vivo*

Remarkably, we also found SPEN to be essential for mediating gene silencing *in vivo*, during imprinted XCI. Indeed, although *Xist* was still highly expressed, gene silencing across the paternal X chromosome was completely hindered in *Spen* KO embryos. This finding is important, as it demonstrates that the mechanisms responsible for initiating gene silencing are similar during both waves of XCI (imprinted and random, the latter being modeled in our mESC system), and largely dependent on SPEN. Conversely, several studies have shown that the mechanisms involved in *maintaining* gene silencing after XCI is established differ between imprinted and random XCI (see **section 1.9**).

We found that the iXCI transcriptional phenotype of *Spen* KO embryos is identical to that of *Xist* KO embryos [Borensztein et al., 2017b]. Female *Xist* KO embryos show severe growth defects starting at E6.5, and lethality by E10.5 as a consequence of iXCI failure in extraembryonic tissues [Marahrens et al., 1997]. Kuroda et al. however reported that *Spen* KO embryos appear similar to wild-type embryos – in size and morphology – at E10.5; and lethality is observed only later, at E13.5 [Kuroda et al., 2003]. This difference in timing of lethality is surprising considering that iXCI is fully deficient in *Spen* KO embryos. Kuroda et al. however only examined a few *Spen* KO embryos around E10.5 (n=7) and did not follow sex ratios in their experiments. Hence, it is most likely that *Spen* KO *female* embryos die at E10.5 due to iXCI failure, while *Spen* KO *male* embryos die

at E13.5, due to other SPEN-dependent phenotypes. Future studies should investigate such likely female-specific lethality of *Spen* KO embryos.

8.5 SPEN is dispensable to maintain gene silencing but buffers expression of escapees in somatic cells

No prior studies had explored whether SPEN is important to maintain X-linked gene silencing after it has been established. The observation that SPEN is enriched on the Xi in differentiated cells suggested it could play a role in such process. Thanks to the degron approach, I was able to deplete SPEN in differentiated cells (NPCs), after XCI had been established and maintained. Transcriptomic analysis revealed no sign of reactivation of fully inactivated genes, demonstrating that SPEN becomes dispensable during maintenance of XCI. This observation is not unexpected, given that previous studies highlighted that other pathways (namely Polycomb and DNA methylation) are involved in maintaining gene silencing on the Xi (see **section 1.9**).

The impact of SPEN loss on maintenance of XCI should be further studied in other somatic cell types. Indeed, it is tempting to hypothesize that some cell types may be more epigenetically labile than others, and hence more susceptible to X-linked gene reactivation upon loss of SPEN. For example, it is known that male cells of the hematopoietic compartment can undergo *Xist*-mediated gene silencing [Savarese et al., 2006] and that *Xist*-deficient female hematopoietic cells show X-reactivation, ultimately leading to cancer [Yildirim et al., 2013]. In these cells, regulation of SPEN expression could be a means to affect X-linked gene dosage. Very recently, a study reported that *Spen* loss of function in B-cells resulted in significant reactivation of the Xi [Yu et al., 2021], thereby confirming that SPEN's importance during XCI maintenance depends on cellular context.

Nonetheless, our results in ESCs and NPCs hint towards a model in which SPEN is important for gene silencing only when genes are transcriptionally ON. Consistent with such model, I find that SPEN significantly dampens expression of escapees (which are transcriptionally expressed from the Xi) in NPCs. This transcriptional effect occurs globally, throughout almost all escapees, suggesting that dosage of escapee expression could be buffered directly through SPEN. However, it is important to note that the magnitude of escapee transcriptional changes upon SPEN-loss is fairly low (less than two-fold), and we did not assess whether their respective protein levels were affected.

These findings could hold clinical relevance in the context of certain diseases in women. For example, it may help understand the link between *Spen* loss of function mutations

and breast cancer (see **section 2.7**). It is tempting to hypothesize that in *Spen*-mutated contexts, tumorogenesis may be promoted by overall increases in expression of dosage-sensitive escapees, or even partial reactivation of several X-linked genes.

8.6 *Xist* recruits SPEN to the X chromosome

8.6.1 SPEN acts downstream of *Xist* coating

In ES cells, loss of SPEN does not affect *Xist* RNA coating (i.e. morphology and frequency of appearance of *Xist* clouds) when assayed using RNA FISH, suggesting that SPEN plays no major role in organizing the *Xist* domain. This observation is made using our dox-inducible system, in which forced *Xist* expression may hide some effects SPEN could have on *Xist* coating. Furthermore, *Xist* coating cannot be monitored with great resolution using FISH. Performing *Xist* *R*NA *a*ntisense *p*urification followed by genomic sequencing (RAP-seq) [Engreitz et al., 2013] under SPEN depleted conditions would help unravel whether SPEN affects *Xist* spreading at genomic resolution.

Nonetheless, I propose that the SPEN degron can be a nice model to uncouple gene silencing from other functions mediated by *Xist* RNA, without having to interfere with *Xist*'s sequence itself.

8.6.2 SPEN loss affects *Xist* expression differently *in vitro* and *in vivo*

Intriguingly, I noticed using RNA-seq and qPCR that *Xist* levels following dox-mediated induction in SPEN-depleted ES cells are twice as low as in wild-type cells (data not shown). The gene silencing defects observed upon SPEN loss are unlikely to be linked to such lower *Xist* expression. Indeed, *Xist* still remains very highly expressed, forming strong domains in SPEN-depleted cells, and a previous study showed that *Xist* expression levels poorly correlate with chromosome-wide silencing outcomes [Loda et al., 2017].

The lower *Xist* levels observed in the absence of SPEN *in vitro* can be explained by several hypotheses:

First, it has been reported that when *Xist* is deleted of its A-repeat (which is required for gene silencing), *Xist* cannot spread over active regions of the X-chromosome [Engreitz et al., 2013]. Given that SPEN binds the A-repeat and is essential for mediating gene

silencing, it is most likely that SPEN loss also hinders *Xist* spreading into active regions. Such hypothesis could be tested by performing RAP-seq [Engreitz et al., 2013] in SPEN-depleted cells. Lower *Xist* levels would then be a consequence of its destabilization by nuclear RNAses upon failure to spread properly across the X chromosome.

Second, I observed that *Tsix* is not as silenced in SPEN-depleted cells as it is in wild-type cells. While this suggests that *Tsix* silencing is at least partially dependent on SPEN, it also hints that this higher *Tsix* expression could lead to *Xist* repression (see **section 1.6**), thus explaining the lower *Xist* levels observed in SPEN-depleted cells. In fact, a recent study to which I contributed (see **section 2.10** and **section A.2**) reports that in the context of *differentiation-coupled* XCI, SPEN-mediated silencing of *Tsix* is absolutely required for *Xist* upregulation [Robert-Finestra et al., 2020].

Conversely, *in vivo* during imprinted XCI, we found that *Spen* KO embryos show much higher *Xist* expression than wild-type embryos. This could be a compensatory response to failure to initiate XCI, since imprinted XCI is essential for development. On the other hand, mESCs perfectly tolerate to have two active X chromosomes, which could explain why there is no selective pressure to increase *Xist* expression upon SPEN loss in our *in vitro* system.

8.6.3 *Xist* upregulation immediately triggers SPEN recruitment

While it was reported that SPEN directly interacts with *Xist* RNA [Chu et al., 2015; Monfort et al., 2015; McHugh et al., 2015], whether SPEN is actively recruited to the X chromosome during XCI remained unexplored. Using fixed and live cell imaging during a timecourse of *Xist* expression, I found that immediately upon *Xist* upregulation, SPEN forms "cloud-like" structures identical to and overlapping *Xist* clouds. This finding demonstrates that SPEN can trigger silencing very early on upon *Xist* upregulation, further highlighting its essentiality during XCI.

Consistent with a very early recruitment of SPEN by *Xist*, CUT&RUN revealed SPEN footprints across the *Xist* gene body, suggesting that SPEN binds *Xist* RNA while it is nascently transcribed. In fact, the SPEN footprint is markedly increased at the 5' region of *Xist*, where the A-repeat – to which SPEN binds – is encoded. It is hard to reconcile then, how *Xist* evades transcriptional silencing despite such strong SPEN recruitment at its locus. One possible explanation could be that *Xist* RNA traps SPEN and "carries" it away from its promoter as the RNA is transcribed, thereby preventing SPEN from exerting its silencing function.

8.6.4 SPEN recruitment is required for gene silencing and depends on RRM2-4

Using a complementation strategy, I identified that three of the four RRM domains of SPEN (RRM2, RRM3 and RRM4) are required for its recruitment to the X chromosome in live cells. This finding is consistent with previous reports which demonstrated that the RRM2-4 domains interact directly with *Xist in vitro*.

Importantly, I found that when SPEN was deleted of its RRM2-4 domains, it failed to silence the X chromosome, altogether demonstrating that SPEN recruitment (by *Xist*) is an essential step in the gene silencing process during XCI.

Chapter 9

Mechanisms of SPEN-mediated transcriptional silencing

9.1 SPEN interacts with chromatin to promote transcriptional silencing

9.1.1 SPEN targets promoters and enhancers

CUT&RUN profiling of SPEN revealed virtually no chromatin-binding sites in mESCs, a finding that is consistent with the observation that virtually no genes are differentially expressed upon SPEN depletion in mESCs (data not shown). However, hundreds of binding sites appeared specifically on the X chromosome following *Xist* expression. Taken together, these discoveries strongly indicate that SPEN does not target DNA/chromatin on its own, contrary to what had been proposed in an old report [Newberry et al., 1999]. Instead, SPEN requires an RNA-intermediate to be guided to its target loci.

Following *Xist* expression, SPEN-binding on the X is restricted to promoters and enhancers of X-linked genes, with binding being highest at rapidly silenced genes, and lowest (nay undetectable) at escaping genes. This discovery demonstrates that enhancers and promoters constitute the substrates of SPEN for its function in transcriptional silencing.

It remains to be determined how SPEN is physically directed to these sites. Most probably, where *Xist* accumulates constitutes the first limiting step in governing where SPEN binds, and consistently, it is observed that SPEN spatial dynamics of recruitment closely follow those of *Xist* coating. SPEN-binding to chromatin is then secondarily dictated through interactions with its chromatin-bound protein partners (see **subsection 9.2.2**).

By developing an optimized CUT&RUN methodology, which supports profiling of *Xist* RNA-dependent chromatin recruitment, I report the first genome-binding map of a *Xist*-recruited protein. This approach should be applied to other *Xist*-binding proteins (see **section 1.13**) and could also be extended to study chromatin associated RNAs in general [Li and Fu, 2019].

Remarkably, the discrete profile of SPEN binding to X-linked chromatin is unexpected given that imaging reveals SPEN accumulates globally and massively around the X chromosome. This indicates that CUT&RUN is sufficiently resolutive to only capture the fraction of SPEN that is in the very close vicinity of chromatin, discriminating it from the remaining SPEN fraction which is accumulated around the X (through *Xist*) but not in contact with chromatin.

9.1.2 Active transcription is required for SPEN binding

Although SPEN binds to promoters and enhancers, it does so only at those of *active genes*. Consistently, highly expressed genes show higher SPEN binding and are consequently more dependent on SPEN for silencing than less expressed ones (**section 8.3**). Furthermore, upon gene silencing, SPEN "disengages" from chromatin (as witnessed by a sharp decrease in CUT&RUN signal). High resolution imaging approaches to directly visualize molecules of SPEN, RNAPII and X-linked chromatin, in single cells, should now be employed to further explore the dynamics of such SPEN "disengagement".

Nevertheless, these results indicate that active transcription is required for SPEN binding. Biologically, this guarantees that all ressources are targeted only to genes that need to be dosage compensated (i.e. transcribed ones). Mechanistically, the basis for such a link between SPEN and active transcription could be explained by the direct interaction between SPEN's SPOC domain and the transcription machinery, as is discussed later in **subsubsection 9.2.2.3**.

9.2 SPOC-centered gene silencing mechanisms during XCI

9.2.1 SPEN's SPOC domain is necessary and sufficient for gene silencing

Previously involved in gene repression in other systems (see **section 2.5**), SPEN's SPOC domain also appeared crucial for gene silencing during XCI in my work. Indeed, a SPEN truncation lacking the SPOC domain failed to rescue X-linked gene silencing upon loss of endogenous SPEN and consistently, endogenous deletion of the SPOC domain resulted in chromosome-wide defects in *Xist*-mediated silencing.

Interestingly, knocking-out SPOC endogenously resulted in a slightly less severe phenotype than the one observed upon auxin-mediated depletion of endogenous full-length SPEN. This small phenotypic difference could possibly be explained by compensatory mechanisms arising in the constitutive $Spen^{\Delta SPOC}$ scenario, which cannot take place in the case of the *acute and conditional* SPEN loss achieved in the degron context. However, it is more likely that the small difference between the $Spen^{\Delta SPOC}$ and SPEN-degron phenotypes is explained by the fact that regions distinct from SPOC also play a role in mediating gene silencing during XCI.

Given the very strong $Spen^{\Delta SPOC}$ phenotype, the role of these regions is only minor. The regions involved are neither the RRM1 (see **section 2.4**) nor the RID (nuclear receptor interaction, see **section 2.5**) domains, as highlighted by the complementation analysis. It is also highly unlikely that the RBPID domain – which mediates interaction with RBP-J (and hence Notch signaling, see **section 2.6**) – is involved in XCI. Indeed, Notch signaling is dispensable for early embryonic development, with maternal-zygotic *Rbpj*-KO mouse embryos giving rise to all three germ layers [Souilhol et al., 2006], contrary to what would be observed should XCI be compromised [Marahrens et al., 1997].

The SPEN regions whose function I did not investigate in the complementation assay correspond to very long stretches of amino acids mapping to predicted *i*ntrinsically *d*isordered *r*egions (IDRs). Given that IDRs are known to enable low affinity protein-protein interactions, which favor increases in protein concentration locally [Banani et al., 2017; Chong et al., 2018], it is conceivable that these regions could be involved in higher-order assembly with *Xist* and other protein partners, leading to the molecular compartmentalization of the Xi, which has been proposed to regulate gene silencing during XCI [Cerase et al., 2019]. In fact, a recent report from the Plath lab shows that *Xist* nucleates the formation of SPEN clusters, which dynamically increase in concentration through

protein-protein interactions [Markaki et al., 2020].

Remarkably, the "artificial" recruitment of SPOC alone to *Xist* and the X chromosome, through the Bgl stem-loop tethering system, is sufficient to carry out a major fraction of SPEN-dependent gene silencing, chromosome wide. Importantly, the magnitude of this *Xist*+SPOC rescue effect is not homogeneous throughout the X chromosome, decreasing as the genomic distance to the *Xist* locus increases. This highlights the limitation of the Bgl-tethering approach, in which SPOC is being artificially anchored to *Xist* exon 7 and presumably not binding with the same stoichiometry as it would in the context of SPEN on *Xist* A-repeat. Nevertheless, SPOC clearly emerges as SPEN's essential effector of gene silencing during XCI, gathering all elements required to instruct gene silencing autonomously.

9.2.2 SPOC is a molecular integrator of epigenetic and transcriptional control

The characterization of SPOC's protein interactome revealed that its partners include NCoR/SMRT (expectedly), the NuRD complex, the m^6A RNA methylation machinery as well as the RNAPII transcription machinery. Hence, SPOC acts as a hub during XCI, allowing SPEN to bridge *Xist* with several parallel factors involved in transcription and chromatin regulation.

Identifying protein partners of SPEN allows to go a layer beyond the *Xist*-interactome originally revealed in the proteomic studies (see **section 1.13**), bringing new mechanistic insights into the XCI process. It could be informative to repeat the *Xist*-proteomic experiments [Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015] under SPEN-depleted conditions, so as to reveal whether any of the originally identified *Xist*-interacting proteins are in fact brought in via SPEN.

9.2.2.1 Re-evaluating the essentiality of the SPEN-NCoR/SMRT-HDAC3 axis

The NCoR/SMRT complex was expectedly retrieved upon SPOC immunoprecipitation, and so was HDAC3. Furthermore, HDAC3 was found to be pre-bound to X-linked enhancers (instead of being actively recruited during XCI), where its occupancy strongly overlaps that of SPEN-targeted enhancers. Hence, this work refines the model for SPEN function during XCI previously proposed by McHugh et al., pinpointing that following *Xist* coating, SPEN accumulates at enhancers where it activates pre-bound HDAC3 by

recruiting NCoR/SMRT through its SPOC domain.

Furthermore, HDAC3-dependent deacetylation plays a role in *Xist*-mediated gene silencing, but this role is not as essential as McHugh et al. orignially proposed. Indeed, we report that HDAC3 loss only delays XCI in mESCs [Żylicz et al., 2019], and unpublished data from our lab show that imprinted XCI is not altered in *Hdac3* KO female embryos, indicating that although HDAC3 may facilitate efficient gene silencing during XCI, it does not trigger it.

Conversely, loss of SPEN – or SPOC only – is associated with severely hindered silencing of virtually all X-linked genes, demonstrating that SPEN/SPOC mediates gene silencing through additional mechanisms which do not involve HDAC3. Consistently, HDAC3 is not readily bound at X-linked promoters – contrary to SPEN – suggesting that the function of SPEN at promoters has to be explained independently of HDAC3.

9.2.2.2 A possible role for nucleosome remodeling during XCI

SPOC co-purifies with all components of the NuRD complex [reviewed in Lai and Wade, 2011], which carries nucleosome remodeling (through ATP-dependent histone sliding by CHD3/4/5 helicases) and histone deacetylation (through HDAC1/2) activities, both of which are involved in silencing transcription [Wade et al., 1998; Zhang et al., 1998; Xue et al., 1998; Tong et al., 1998; Bornelöv et al., 2018].

Remarkably, binding at X-linked promoters by two NuRD subunits (CHD4 and MBD3) strongly overlaps that of SPEN. Furthermore, auxin-mediated depletion of CHD4 during XCI in mESCs results in moderate defects in *Xist*-mediated gene silencing [Dossin, unpublished]. Given that neither HDAC1 nor HDAC2 (the two NuRD-associated HDACs) are involved in XCI [Żylicz et al., 2019], these results altogether indicate that the NuRD complex plays a role in *Xist*-mediated gene silencing, specifically through its CHD4-mediated nucleosome remodeling function. Supporting this conclusion, CHD4-dependent nucleosome remodeling was recently shown to displace RNAPII from TSSs, thus participating to transcriptional silencing [Bornelöv et al., 2018]. Importantly, the SPOC/NuRD interaction suggests that SPEN may regulate NuRD's function during XCI, although additional work is now required to determine precisely how, when and where this happens.

9.2.2.3 SPOC as a direct repressor of RNAPII function

SPOC co-immunoprecipitates with several components of the RNAPII transcription machinery, as well as with multiple general transcription initiation and elongation factors.

What is more, SPOC interacts directly with the CTD of the POLR2A (RPB1) RNAPII subunit, specifically when the latter is phosphorylated on Serine 5. Structural analysis reveals that the Ser5 phosphoserine comprised within the CTD YSPTSPS heptapeptide repeat engages in several interactions with at least 3 residues located in the basic patch of SPOC.

These molecular insights provide a mechanistic basis for the observations that:

(i) active transcription is required for SPEN binding to chromatin
(ii) SPEN chromatin recruitment during XCI majorly correlates with Ser5-phosphorylated RNAPII-bound sites
(iii) gene silencing (i.e. loss of transcribing RNAPII) results in SPEN dissociation from chromatin

Biologically, such a direct SPEN/RNAPII interaction strongly hints that SPEN can repress transcription directly. The specificity of SPOC for Ser5-phosphorylated RNAPII (regardless of whether or not Ser2 and Ser7 are also phosphorylated) shows that SPEN has the potential to target any RNA polymerase II molecule engaged in the early stages of the transcription process (i.e. initiation, promoter escape and the switch into productive elongation). Rationally, this seems to be the most straightforward way to perturb ongoing transcription activity, thus explaining how gene silencing can occur so efficiently and robustly during XCI.

Biophysical analysis revealed that the affinity of the SPOC/Ser5P-CTD interaction is relatively weak ($K_d \approx 10$uM). In comparison, the affinity between SPOC and phosphorylated NCoR/SMRT is reportedly at least 10 times greater ($K_d < 1$uM) [Oswald et al., 2016]. Such modest binding between SPEN and RNAPII is to be expected. Indeed, given the potent repressive activity of SPEN, a much stronger affinity would suggest that SPEN is "constitutively" associated with RNAPII, genome wide, hence hindering transcription dramatically. Furthermore, interaction affinities between CTD-phosphopeptides and other CTD-interacting proteins [Eick and Geyer, 2013; Vasiljeva et al., 2008] are within the same order of magnitude as the Ser5P-CTD/SPOC interaction affinity, confirming the relevance of our measurement.

The weaker SPEN/RNAPII interaction (compared to SPOC/NCoR/SMRT) is to be contrasted with the fact that the RNAPII-CTD motif (YSPTSPS) recognized by SPOC is repeated more than 50 times in the RNAPII molecule (see **subsection 7.1.1**), while the NCoR/SMRT motif (LSDS) recognized by SPOC is present in only one copy per NCoR/SMRT molecule (see **subsection 7.1.2**). This would in theory allow multiple SPOC molecules to bind to a single RNAPII molecule *in vivo* (**Figure 9.1A**), and we show that this is indeed the case *in vitro*.

In light of the observation that SPEN accumulates massively around the X chromosome following *Xist* coating, it follows that the local concentration of SPEN in the vicinity of X-linked chromatin increases drastically during XCI. Such crowding of SPEN likely overcomes the weak binding affinity of SPEN for RNAPII, shifting the molecular balance in favor of the SPEN/RNAPII interaction, with many SPEN/SPOC molecules in the vicinity of transcribing RNA polymerases. This makes SPEN a unique transcriptional regulator in the sense that despite having a strong repressive potential, this potential is initially "dormant", and can only be mobilized through its conditional accumulation at target loci, through an RNA intermediate (i.e. *Xist*).

Whether the SPEN/RNAPII interaction is of functional relevance for gene silencing during XCI needs to be investigated. Structural comparison of the SPOC/Ser5P-CTD and SPOC/phospho-SMRT complexes reveals that RNAPII and NCoR/SMRT bind to the same region of SPOC (i.e. the basic patch), engaging in intermolecular interactions with the same SPOC residues. Given that the SPOC/NCoR-SMRT/HDAC3 axis plays a role in *Xist*-mediated gene silencing, the function of the SPOC/RNAPII interaction during XCI has to be addressed in a context where it has been genetically uncoupled from the SPOC-NCoR/SMRT interaction. The experiment I propose is to test the ability of different SPOC mutants (mutated at key residues implicated in the interaction with the Ser5P-CTD) to rescue gene silencing defects associated with auxin-mediated depletion of endogenous SPEN (**Figure 9.1A**, an experiment principally similar to the one described in **subsubsection 6.3.2.4**), in a cellular background in which the LSDS-motif of endogenous NCoR and SMRT – which mediates their interaction with SPOC, and is located at their very C-terminus (see **subsection 7.1.2**) – has been deleted (**Figure 9.1B**).

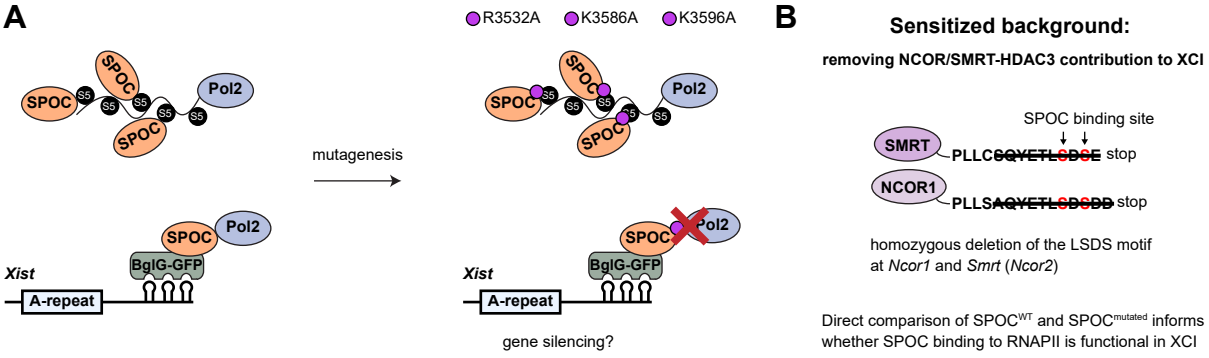


Figure 9.1 – Scheme of the proposed experiment to test the functionality of the SPOC/RNAPII interaction during XCI. (A) Mutating SPOC at key residues (shown as purple circles) allows the disruption of its interaction with RNAPII. Bgl-tethering of the mutated SPOC to *Xist* RNA enables testing whether the absence of the SPOC/RNAPII interaction has an impact on gene silencing during XCI. **(B)** Uncoupling of the SPOC/RNAPII axis from the NCoR/SMRT-HDAC3 axis is achieved by performing the proposed experiment in a background where the LSDS motifs of NCoR and SMRT have been deleted homozygously, thereby ensuring that NCoR/SMRT cannot engage with SPOC anymore.

Assuming that SPEN's interaction with RNAPII is important for gene silencing during XCI, one could then propose that *Xist*-mediated transcriptional silencing takes place through the following steps:

(i) *Xist* extensively recruits SPEN (through the A-repeat) to the X chromosome, leading to an increase in SPEN concentration locally around X-linked chromatin.
(ii) SPEN's interaction with RNAPII inhibits the transcription process and triggers gene silencing.
(iii) SPEN-mediated activation of HDAC3 at enhancers accentuates/facilitates gene silencing.
(iv) *Xist*-dependent recruitment of Polycomb (through the B-repeat) further locks gene silencing, in synergy with DNA methylation.

9.2.2.4 SPOC is a novel interaction domain of the m^{6}A writing machinery

Xist m^{6}A RNA methylation has previously been shown to be important for XCI, although the extent to which this mark plays a role in *Xist*-mediated gene silencing is controversial (see **subsection 1.13.3**). So far, Patil et al. proposed that the m^{6}A machinery is recruited to *Xist* via the RBM15 cofactor – which directly binds *Xist* A-repeat – with subsequent YTHDC1 binding to methylated *Xist* being obscurely implicated in X-linked gene silencing (see **subsection 1.13.3**). Surprisingly however, a recent study reports that RBM15 is dispensable for gene silencing during XCI [Nesterova et al., 2019].

Previous work conducted in *Drosophila* and mouse showed that RBM15 interacts directly with ZC3H13, which bridges RBM15 with the rest of the m^{6}A-writing complex, notably WTAP and the METTL3/METTL14 methyltransferases [Knuckles et al., 2018]. The regions of RBM15 implicated in this interaction were however not explored. Remarkably, SPEN and RBM15 share two striking features. First, they both interact directly with the A-repeat of *Xist* through their RNA recognition motifs. Second, they both carry a SPOC domain at their very C-terminal end, although these two SPOCs only share 35% homology.

In this work, I show that SPEN's SPOC domain co-purifies with all subunits of the m^{6}A-writing machinery, as well as the nuclear m^{6}A reader YTHDC1. Hence, I propose that interaction with the m^{6}A RNA methyltransferase complex is a general feature of SPOC domains, shared across other SPOC-containing proteins. Given the results from Knuckles et al., this interaction is most likely mediated through ZC3H13. Consistently, ZC3H13 is the most highly enriched m^{6}A-associated factor in my SPOC IP-mass spectrometry analysis.

The discovery that SPOC interacts with the m^{6}A machinery suggests that SPEN can functionally replace RBM15, thereby explaining why defects in XCI are inconsistently observed upon *Rbm15* knock-out or knockdown [Nesterova et al., 2019; Patil et al., 2016]. Taken together, my work highlights SPEN as a very probable candidate for linking *Xist* with m^{6}A RNA methylation. It is now important to profile *Xist* m^{6}A methylation upon SPEN loss, so as to confirm that SPEN plays a role in *Xist* methylation.

Nevertheless, it seems unlikely that such SPEN-dependent RNA methylation is important for *Xist*-mediated gene silencing, given that mESCs show unperturbed XCI upon auxin-mediated depletion of ZC3H13 [Dossin, unpublished]. However, SPEN-dependent RNA methylation could regulate *Xist* post-transcriptionally (e.g. promoting its stability), and could thus explain the reduced *Xist* levels observed under SPEN-depleted conditions (see **subsection 8.6.2**).

In one of the *Xist* proteomic studies, Howard Chang's lab reported that the WTAP m^{6}A cofactor co-purifies with full-length *Xist* RNA – but not *Xist* ΔA-repeat – and hence concluded that WTAP interacts with the A-repeat of *Xist* [Chu et al., 2015]. However, although several studies have now shown that SPEN directly binds *Xist* A-repeat [Monfort et al., 2015; Lu et al., 2016; Carter et al., 2020], no such direct interaction has been reported for WTAP. In light of my work revealing that SPEN is a *bona fide* interactor of the m^{6}A RNA methylation machinery (with WTAP scoring as a highly enriched candidate in the SPOC protein interactome), I propose that WTAP is in fact bridged to *Xist* indirectly via SPEN, hence co-purifying with *Xist* in an A-repeat dependent manner.

9.3 SPEN-family proteins as RNA-guided transcriptional repressors

SPEN is conserved across the animal kingdom (see **section 2.2**), and its RRM2-4 and SPOC domains are remarkably well conserved in vertebrates. Furthermore, several proteins containing both RRM and SPOC domains are found across fungi, plant and animal kingdoms. Therefore, it is tempting to speculate that the protein interactome of SPEN's SPOC domain is very similar across vertebrates, and that some of the mechanisms identified in this work may also be bridged by SPOC in other SPOC-containing proteins.

If that were the case, SPEN and other SPEN-related proteins would emerge as RNA-guided transcriptional repressors, bridging several layers of chromatin and transcriptional control to ensure potent gene repression. It is now crucial to characterize the repertoire of RNAs that interact specifically with SPEN, across tissues and animals. A recent study

found that in addition to *Xist*, SPEN's RRM2-4 domains bind to *e*ndogenous *r*etrovirus *K* elements (ERVKs), a family of integrated retroviruses sequences [Carter et al., 2020]. Through binding actively transcribed ERVKs, SPEN represses ERVK transcription and ensures genomic surveilance [Carter et al., 2020]. The silencing mechanisms involved were not addressed, but these mechanisms are likely to be similar to those involved in XCI.

Highlighting that SPOC domains generally may be involved in similar mechanisms, a recent study conducted in plants (*Arabidopsis*) found that BORDER proteins – which carry a SPOC domain – act as transcriptional insulators by promoting 3' RNAPII-pausing [Yu et al., 2019]. The precise mechanism by which this happens was not addressed, but it is tempting to hypothesize that the SPOC domain of BORDER proteins may be interacting with the CTD domain of *At*RNAPII according to CTD phosphorylation status, with a specificity towards phosphorylation at Serine 2, which is most enriched at the 3' end of genes (see **subsection 7.1.1**).

Furthermore, the plant protein FPA – which regulates flowering time and carries RRM and SPOC domains – has been shown to silence chromatin in an RNA-dependent manner [Bäurle et al., 2007] and regulate 3'-end processing [Hornyik et al., 2010; Sonmez et al., 2011]. Importantly, FPA's SPOC domain shares structural similarity with SPEN's SPOC domain, notably at the level of the basic patch [Zhang et al., 2016].

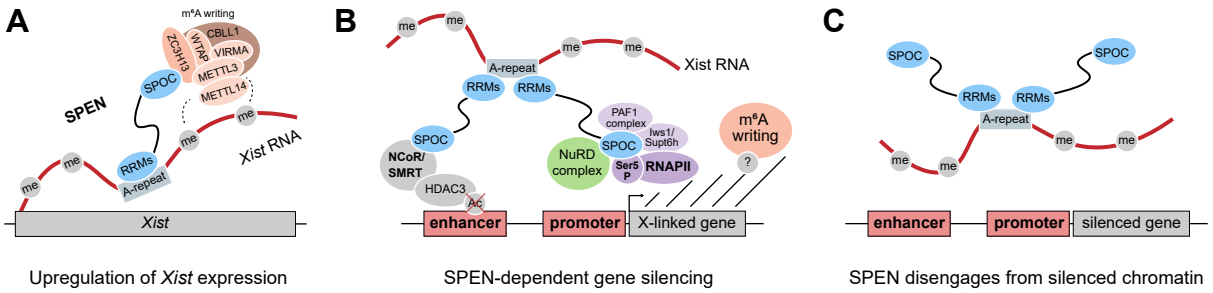


Figure 9.2 – Working model for SPEN function during XCI. (A) At the very beginning of XCI, SPEN is recruited by *Xist* RNA while it is being transcribed. The interaction between the SPOC domain and the m^6A writing machinery may be involved in *Xist* RNA methylation. **(B)** *Xist* then guides SPEN to X-linked enhancers and promoters, where the SPOC domain triggers gene silencing by interacting with the NCoR/SMRT complex, the NuRD complex as well as the transcription machinery. **(C)** Following gene silencing, the "active chromatin" context favorable for SPEN binding is lost, and SPEN hence disengages from X-linked chromatin. *Direct SPOC-interactors are labeled in bold.*

Taken together, my work and the existing literature suggest that RNA-mediated recruitment of SPEN-family proteins is a widespread means to repress transcription acutely at target loci, through simultaneously engaging with multiple layers of epigenetic and transcriptional regulation. In the context of XCI, I propose that SPEN is a molecular integrator which links *Xist* RNA with the transcription machinery, histone deacetylases,

nucleosome remodellers and RNA m^6A-transferases to promote efficient and robust gene silencing (**Figure 9.2**). As a whole, this work demonstrates that the study of X chromosome inactivation provides a powerful model to discover fundamental aspects of protein function and gene regulation.

Part V

References

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):213–221.

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(6342):1081–1084.

Andriatsilavo, M., Stefanutti, M., Siudeja, K., Perdigoto, C. N., Boumard, B., Gervais, L., Gillet-Markowska, A., Al Zouabi, L., Schweisguth, F., and Bardin, A. J. (2018). Spen limits intestinal stem cell self-renewal. *PLoS Genetics*, 14(11):e1007773.

Arieti, F., Gabus, C., Tambalo, M., Huet, T., Round, A., and Thore, S. (2014). The crystal structure of the split end protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. *Nucleic Acids Research*, 42(10):6742–6752.

Ariyoshi, M. and Schwabe, J. W. R. (2003). A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes & development*, 17(15):1909–20.

Armenia, J., Wankowicz, S. A., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W. K., Chakravarty, D., Han, G. C., Coleman, I., Montgomery, B., Pritchard, C., Morrissey, C., Barbieri, C. E., Beltran, H., Sboner, A., Zafeiriou, Z., Miranda, S., Bielski, C. M., Penson, A. V., Tolonen, C., Huang, F. W., Robinson, D., Wu, Y. M., Lonigro, R., Garraway, L. A., Demichelis, F., Kantoff, P. W., Taplin, M. E., Abida, W., Taylor, B. S., Scher, H. I., Nelson, P. S., De Bono, J. S., Rubin, M. A., Sawyers, C. L., Chinnaiyan, A. M., Schultz, N., and Van Allen, E. M. (2018). The long tail of oncogenic drivers in prostate cancer. *Nature Genetics*, 50(5):645–651.

Aster, J. C., Pear, W. S., and Blacklow, S. C. (2017). The Varied Roles of Notch in Cancer. *Annual Review of Pathology: Mechanisms of Disease*, 12(1):245–275.

Augui, S., Nora, E. P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Reviews Genetics*, 12(6):429–442.

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, 14(2):113–124.

Banani, S. F., Lee, H. O., Hyman, A. A., and Rosen, M. K. (2017). Biomolecular condensates: Organizers of cellular biochemistry.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science*, 354(6314):909–912.

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 [2]. *Nature*, 163(4148):676–677.

Barton, D. E., David, F. N., and Merrington, M. (1964). The positions of the sex chromosomes in the human cell in mitosis. *Annals of Human Genetics*, 28(1-3):123–128.

Bäurle, I., Smith, L., Baulcombe, D. C., and Dean, C. (2007). Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science*, 318(5847):109–112.

Belaghzal, H., Dekker, J., and Gibcus, J. H. (2017). Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods*, 123:56–65.

Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J. I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F. A., and Kreshuk, A. (2019). ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, 16(12):1226–1232.

Berletch, J. B., Ma, W., Yang, F., Shendure, J., Noble, W. S., Disteche, C. M., and Deng, X. (2015). Escape from X Inactivation Varies in Mouse Tissues. *PLoS Genetics*, 11(3):e1005079.

Berletch, J. B., Yang, F., Xu, J., Carrel, L., and Disteche, C. M. (2011). Genes that escape from X inactivation. *Human genetics*, 130(2):237–45.

Blackburn, G. M., Cherfils, J., Moss, G. P., Richards, N. G. J., Waltho, J. P., Williams, N. H., and Wittinghofer, A. (2017). IUPAC Recommendations How to name atoms in phosphates, polyphosphates, their derivatives and mimics, and transition state analogues for enzyme-catalysed phosphoryl transfer reactions (IUPAC Recommendations 2016). *Pure Appl. Chem*, 89(5):653–675.

Blackledge, N. P., Farcas, A. M., Kondo, T., King, H. W., McGouran, J. F., Hanssen, L. L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., Ishikura, T., Long, H. K., Sheahan, T. W., Brockdorff, N., Kessler, B. M., Koseki, H., and Klose, R. J. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell*, 157(6):1445–1459.

Blewitt, M. E., Gendrel, A. V., Pang, Z., Sparrow, D. B., Whitelaw, N., Craig, J. M., Apedaile, A., Hilton, D. J., Dunwoodie, S. L., Brockdorff, N., Kay, G. F., and Whitelaw, E. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature Genetics*, 40(5):663–669.

Blewitt, M. E., Vickaryous, N. K., Hemley, S. J., Ashe, A., Bruxner, T. J., Preis, J. I., Arkell, R., and Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21):7629–7634.

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. *Nature Genetics*, 30(1):73–76.

Borensztein, M., Okamoto, I., Syx, L., Guilbaud, G., Picard, C., Ancelin, K., Galupa, R., Diabangouaya, P., Servant, N., Barillot, E., Surani, A., Saitou, M., Chen, C. J., Anastassiadis, K., and Heard, E. (2017a). Contribution of epigenetic landscapes and transcription factors to X-chromosome reactivation in the inner cell mass. *Nature Communications*, 8(1):1–14.

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. (2017b). Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nature Structural & Molecular Biology*.

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. *Molecular Cell*, 71(1):56–72.e4.

Bourgeois, C. A., Laquerriere, F., Hemon, D., Hubert, J., and Bouteille, M. (1985). New data on the in situ position of the inactive X chromosome in the interphase nucleus of human fibroblasts. *Human Genetics*, 69(2):122–129.

Bousard, A., Raposo, A. C., Żylicz, J. J., Picard, C., Pires, V. B., Qi, Y., Gil, C., Syx, L., Chang, H. Y., Heard, E., and da Rocha, S. T. (2019). The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO reports*, 20(10).

Brockdorff, N. (2019). Localized accumulation of Xist RNA in X chromosome inactivation. *Open Biology*, 9(12).

Brockdorff, N., Ashworth, A., Kay, G. F., Cooper, P., Smith, S., McCabe, V. M., Norris, D. P., Penny, G. D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, 351(6324):329–331.

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(3):515–26.

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(6304):38–44.

Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafrenière, 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(3):527–542.

Brown, C. J. and Willard, H. F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature*, 368(6467):154–156.

Buratowski, S. (2003). The CTD code.

Cardozo, T. and Pagano, M. (2004). The SCF ubiquitin ligase: Insights into a molecular machine.

Carrel, L. and Brown, C. J. (2017). When the lyon(Ized chromosome) roars: Ongoing expression from an inactive X chromosome. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1733).

Carrel, L., Cottle, A. A., Goglin, K. C., and Willard, H. F. (1999). A first-generation X-inactivation profile of the human X chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 96(25):14440–14444.

Carter, A. C., Xu, J., Nakamoto, M. Y., Wei, Y., Zarnegar, B. J., Shi, Q., Broughton, J. P., Ransom, R. C., Salhotra, A., Nagaraja, S. D., Li, R., Dou, D. R., Yost, K. E., Cho, S. W., Mistry, A., Longaker, M. T., Khavari, P. A., Batey, R. T., Wuttke, D. S., and Chang, H. Y. (2020). Spen links rna-mediated endogenous retrovirus silencing and x chromosome inactivation. *eLife*, 9:1–58.

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., and Tartaglia, G. G. (2019). Phase separation drives X-chromosome inactivation: a hypothesis.

Cerase, A., Smeets, D., Tang, Y. A., Gdula, M., Kraus, F., Spivakov, M., Moindrot, B., Leleu, M., Tattermusch, A., Demmerle, J., Nesterova, T. B., Green, C., Otte, A. P., Schermelleh, L., and Brockdorff, N. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6):2235–2240.

Chapman, R. D., Heidemann, M., Hintermair, C., and Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD.

Chapuy, B., Stewart, C., Dunford, A. J., Kim, J., Kamburov, A., Redd, R. A., Lawrence, M. S., Roemer, M. G., Li, A. J., Ziepert, M., Staiger, A. M., Wala, J. A., Ducar, M. D., Leshchiner, I., Rheinbay, E., Taylor-Weiner, A., Coughlin, C. A., Hess, J. M., Pedamallu, C. S., Livitz, D., Rosebrock, D., Rosenberg, M., Tracy, A. A., Horn, H., Van Hummelen, P., Feldman, A. L., Link, B. K., Novak, A. J., Cerhan, J. R., Habermann, T. M., Siebert, R., Rosenwald, A., Thorner, A. R., Meyerson, M. L., Golub, T. R., Beroukhim, R., Wulf, G. G., Ott, G., Rodig, S. J., Monti, S., Neuberg, D. S., Loeffler, M., Pfreundschuh, M., Trümper, L., Getz, G., and Shipp, M. A. (2018). Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nature Medicine*, 24(5):679–690.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current biology : CB*, 6(2):149–62.

Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, 95(2):118–128.

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 & development*, 20(16):2223–37.

Chelmicki, T., Roger, E., Teissandier, A., Dura, M., Bonneville, L., Rucli, S., Dossin, F., Fouassier, C., Lameiras, S., and Bourc'his, D. (2021). m6A RNA methylation regulates the fate of endogenous retroviruses. *Nature*, 591(7849):312–316.

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(6311):468–472.

Chen, F. and Rebay, I. (2000). split ends, a new component of the Drosophila EGF receptor pathway, regulates development of midline glial cells. *Current Biology*, 10(15):943–946.

Chen, J., Nikolaitchik, O., Singh, J., Wright, A., Bencsics, C. E., Coffin, J. M., Ni, N., Lockett, S., Pathak, V. K., and Hu, W. S. (2009). High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 106(32):13535–13540.

Cho, E. J., Takagi, T., Moore, C. R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes and Development*, 11(24):3319–3326.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., and Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, 361(6400).

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(2):404–416.

Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and Rougeulle, C. (2011). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Human Molecular Genetics*, 20(4):705–718.

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. *Journal of Cell Biology*, 132(3):259–275.

Coker, H., Wei, G., Moindrot, B., Mohammed, S., Nesterova, T., and Brockdorff, N. (2020). The role of the Xist 5' m6A region and RBM15 in X chromosome inactivation. *Wellcome Open Research*, 5.

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. *Molecular Cell*, 74(1):101–117.e10.

Cooper, S., Grijzenhout, A., Underwood, E., Ancelin, K., Zhang, T., Nesterova, T. B., Anil-Kirmizitas, B., Bassett, A., Kooistra, S. M., Agger, K., Helin, K., Heard, E., and Brockdorff, N. (2016). Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nature Communications*, 7(1):1–8.

Costanzi, C. and Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*, 393(6685):599–601.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *Journal of Cell Biology*, 153(4):773–783.

Da Rocha, S. T. and Heard, E. (2017). Novel players in X inactivation: Insights into Xist-mediated gene silencing and chromosome conformation.

da Rocha, S., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N., Sanulli, S., Chow, J., Schulz, E., Picard, C., Kaneko, S., Helin, K., Reinberg, D., Stewart, A., Wutz, A., Margueron, R., and Heard, E. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Molecular Cell*, 53(2):301–316.

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., and Aiden, E. L. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proceedings of the National Academy of Sciences of the United States of America*, 113(31):E4504–E4512.

David, C. J., Boyne, A. R., Millhouse, S. R., and Manley, J. L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes and Development*, 25(9):972–982.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K. J., Gooding, A. R., Lee, J. T., and Cech, T. R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Molecular Cell*, 57(3):552–558.

Davidovich, C., Zheng, L., Goodrich, K. J., and Cech, T. R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nature Structural and Molecular Biology*, 20(11):1250–1257.

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., Koseki, H., and Brockdorff, N. (2004). Polycomb group proteins ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental Cell*, 7(5):663–676.

De Vries, W. N., Binns, L. T., Fancher, K. S., Dean, J., Moore, R., Kemler, R., and Knowles, B. B. (2000). Expression of Cre recombinase in mouse oocytes: A means to study maternal effect genes. *Genesis*, 26(2):110–112.

Deng, X., Hiatt, J. B., Nguyen, D. K., Ercan, S., Sturgill, D., Hillier, L. W., Schlesinger, F., Davis, C. A., Reinke, V. J., Gingeras, T. R., Shendure, J., Waterston, R. H., Oliver, B., Lieb, J. D., and Disteche, C. M. (2011). Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nature Genetics*, 43(12):1179–1185.

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 Biology*, 16(1):152.

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. *Nature*, 435(7041):441–445.

Dhiman, V. K., Bolt, M. J., and White, K. P. (2018). Nuclear receptors in cancer - Uncovering new and evolving roles through genomic analysis. *Nature Reviews Genetics*, 19(3):160–174.

Di Croce, L. and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins.

Disteche, C. M. and Berletch, J. B. (2015). X-chromosome inactivation and escape. *Journal of Genetics*, 94(4):591–599.

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(7398):376–380.

Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N., and Lee, J. T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature*, 460(7251):128–132.

Doroquez, D. B., Orr-Weaver, T. L., and Rebay, I. (2007). Split ends antagonizes the Notch and potentiates the EGFR signaling pathways during Drosophila eye development. *Mechanisms of Development*, 124(9-10):792–806.

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(7795):455–460.

Dyer, K. A., Canlfield, T. K., and Gartler, S. M. (1989). Molecular cytological differentiation of active from inactive x domains in interphase: Implications for x chromosome inactivation. *Cytogenetic and Genome Research*, 50(2-3):116–120.

Edelmann, J., Holzmann, K., Tausch, E., Saunderson, E. A., Jebaraj, B. M., Steinbrecher, D., Dolnik, A., Blätte, T. J., Landau, D. A., Saub, J., Estenfelder, S., Ibach, S., Cymbalista, F., Leblond, V., Delmer, A., Bahlo, J., Robrecht, S., Fischer, K., Goede, V., Bullinger, L., Wu, C. J., Mertens, D., Ficz, G., Gribben, J. G., Hallek, M., Döhner, H., and Stilgenbauer, S. (2020). Genomic alterations in high-risk chronic lymphocytic leukemia frequently affect cell cycle key regulators and NOTCH1-regulated transcription. *Haematologica*, 105(5):1379–1390.

Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, 318(5857):1777–1779.

Eick, D. and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code.

Eils, R., Dietzel, S., Bertin, E., Schröck, E., Speicher, M. R., Ried, T., Robert-Nicoud, M., Cremer, C., and Cremer, T. (1996). Three-dimensional reconstruction of painted human interphase chromosomes: Active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *Journal of Cell Biology*, 135(6):1427–1440.

Emsley, P. and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I):2126–2132.

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 (New York, N.Y.)*, 341(6147):1237973.

Escamilla-Del-Arenal, M., Da Rocha, S. T., and Heard, E. (2011). Evolutionary diversity and developmental regulation of X-chromosome inactivation. *Human Genetics*, 130(2):307–327.

Feng, Y., Bommer, G. T., Zhai, Y., Akyol, A., Hinoi, T., Winer, I., Lin, H. V., Cadigan, K. M., Cho, K. R., and Fearon, E. R. (2007). Drosophila split ends homologue SHARP functions as a positive regulator of Wnt/β-catenin/T-cell factor signaling in neoplastic transformation. *Cancer Research*, 67(2):482–491.

Furlan, G., Gutierrez Hernandez, N., Huret, C., Galupa, R., van Bemmel, J. G., Romito, A., Heard, E., Morey, C., and Rougeulle, C. (2018). The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products. *Molecular Cell*, 70(3):462–472.e8.

Galupa, R. and Heard, E. (2015). X-chromosome inactivation: New insights into cis and trans regulation. *Current Opinion in Genetics and Development*, 31:57–66.

Giaimo, B. D., Oswald, F., and Borggrefe, T. (2017). Dynamic chromatin regulation at Notch target genes. *Transcription*, 8(1):61–66.

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(7613):575–579.

Gontan, C., Achame, E. M., Demmers, J., Barakat, T. S., Rentmeester, E., Van Ijcken, W., Grootegoed, J. A., and Gribnau, J. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature*, 485(7398):386–390.

Graves, J. A., Koina, E., and Sankovic, N. (2006). How the gene content of human sex chromosomes evolved. *Current Opinion in Genetics and Development*, 16(3):219–224.

Graves, J. A. M. (2006). Sex chromosome specialization and degeneration in mammals. *Cell*, 124(5):901–914.

Gray, W. M., Kepinski, S., Rouse, D., Leyser, O., and Estelle, M. (2001). Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature*, 414(6861):271–276.

Grimm, J. B., English, B. P., Chen, J., Slaughter, J. P., Zhang, Z., Revyakin, A., Patel, R., Macklin, J. J., Normanno, D., Singer, R. H., Lionnet, T., and Lavis, L. D. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nature Methods*, 12(3):244–250.

Guenther, M. G., Barak, O., and Lazar, M. A. (2001). The SMRT and N-CoR Corepressors Are Activating Cofactors for Histone Deacetylase 3. *Molecular and Cellular Biology*, 21(18):6091–6101.

Guenther, M. G., Yu, J., Kao, G. D., Yen, T. J., and Lazar, M. A. (2002). Assembly of the SMRT-histone deacetylase 3 repression complex requires the TCP-1 ring complex. *Genes and Development*, 16(24):3130–3135.

Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife*, 6.

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. *Developmental Cell*, 19(3):469–476.

Hazegh, K. E., Nemkov, T., D'Alessandro, A., Diller, J. D., Monks, J., McManaman, J. L., Jones, K. L., Hansen, K. C., and Reis, T. (2017). An autonomous metabolic role for Spen. *PLoS Genetics*, 13(6):e1006859.

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C. D., and Spector, D. L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell*, 107(6):727–38.

Heidemann, M., Hintermair, C., Voß, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription.

Hornyik, C., Terzi, L. C., and Simpson, G. G. (2010). The Spen Family Protein FPA Controls Alternative Cleavage and Polyadenylation of RNA. *Developmental Cell*, 18(2):203–213.

Hosogane, M., Funayama, R., Shirota, M., and Nakayama, K. (2016). Lack of Transcription Triggers H3K27me3 Accumulation in the Gene Body. *Cell Reports*, 16(3):696–706.

Hsin, J. P. and Manley, J. L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing.

Huang, E. Y., Zhang, J., Miska, E. A., Guenther, M. G., Kouzarides, T., and Lazar, M. A. (2000). Nuclear receptor corepressors partner with class II histone deacetylases in a Sin3-independent repression pathway. *Genes and Development*, 14(1):45–54.

Huang, Y., Kim, J. K., Do, D. V., Lee, C., Penfold, C. A., Zylicz, J. J., Marioni, J. C., Hackett, J. A., and Surani, M. A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *eLife*, 6.

Jasnovidova, O. and Stefl, R. (2013). The CTD code of RNA polymerase II: A structural view.

Jeon, Y. and Lee, J. T. (2011). YY1 Tethers Xist RNA to the inactive X nucleation center. *Cell*, 146(1):119–133.

Jeppesen, P. and Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell*, 74(2):281–9.

Jeronimo, C., Bataille, A. R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code.

Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):133–144.

Kabsch, W. (2010b). XDS. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2):125–32.

Kalantry, S. and Magnuson, T. (2006). The Polycomb Group Protein EED Is Dispensable for the Initiation of Random X-Chromosome Inactivation. *PLoS Genetics*, 2(5):e66.

Kalb, R., Latwiel, S., Baymaz, H. I., Jansen, P. W., Müller, C. W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature Structural and Molecular Biology*, 21(6):569–571.

Kaneko, S., Son, J., Bonasio, R., Shen, S. S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes & Development*, 28(18):1983–1988.

Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Böhmdorfer, G., Gregor, W., Kreil, D. P., Matzke, M., and Matzke, A. J. (2008). A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nature Genetics*, 40(5):670–675.

Kao, H. Y., Downes, M., Ordentlich, P., and Evans, R. M. (2000). Isolation of a novel histone deacetylase reveals that class I and class II deacetylases promote SMRT-mediated repression. *Genes and Development*, 14(1):55–66.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6):845–858.

Kempfer, R. and Pombo, A. (2019). Methods for mapping 3D chromosome architecture. *Nature Reviews Genetics*, 21(4):207–226.

Keniry, A., Gearing, L. J., Jansz, N., Liu, J., Holik, A. Z., Hickey, P. F., Kinkel, S. A., Moore, D. L., Breslin, K., Chen, K., Liu, R., Phillips, C., Pakusch, M., Biben, C., Sheridan, J. M., Kile, B. T., Carmichael, C., Ritchie, M. E., Hilton, D. J., and Blewitt, M. E. (2016). Setdb1-mediated H3K9 methylation is enriched on the inactive X and plays a role in its epigenetic silencing. *Epigenetics and Chromatin*, 9(1):16.

Kim, J., Hake, S. B., and Roeder, R. G. (2005). The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Molecular Cell*, 20(5):759–770.

Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., and Strahl, B. D. (2005). A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. *Molecular and Cellular Biology*, 25(8):3305–3316.

Kloet, S. L., Baymaz, H. I., Makowski, M., Groenewold, V., Jansen, P. W., Berendsen, M., Niazi, H., Kops, G. J., and Vermeulen, M. (2015). Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS Journal*, 282(9):1774–1785.

Knuckles, P., Lence, T., Haussmann, I. U., Jacob, D., Kreim, N., Carl, S. H., Masiello, I., Hares, T., Villaseñor, R., Hess, D., Andrade-Navarro, M. A., Biggiogera, M., Helm, M., Soller, M., Bühler, M., and Roignant, J. Y. (2018). Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor RbM15/spenito to the m6 a machinery component Wtap/Fl(2)d. *Genes and Development*, 32(5-6):415–429.

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 Biology*, 2(7):e171.

Kolodziej, P. A., Yeh Jan, L., and Nung Jan, Y. (1995). Mutations that affect the length, fasciculation, or ventral orientation of specific sensory axons in the Drosophila embryo. *Neuron*, 15(2):273–286.

Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes and Development*, 14(19):2452–2460.

Kouzarides, T. (2007). Chromatin Modifications and Their Function.

Krogan, N. J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Ryan, O. W., Golshani, A., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: Linking transcriptional elongation to histone methylation. *Molecular Cell*, 11(3):721–729.

Kuang, B., Wu, S. C., Shin, Y., Luo, L., and Kolodziej, P. (2000). split ends encodes large nuclear proteins that regulate neuronal cell fate and axon extension in the Drosophila embryo. *Development (Cambridge, England)*, 127(7):1517–29.

Kuroda, K., Han, H., Tani, S., Tanigaki, K., Tun, T., Furukawa, T., Taniguchi, Y., Kurooka, H., Hamada, Y., Toyokuni, S., and Honjo, T. (2003). Regulation of marginal zone B cell development by MINT, a suppressor of Notch/RBP-J signaling pathway. *Immunity*, 18(2):301–312.

Lai, A. Y. and Wade, P. A. (2011). Cancer biology and NuRD: A multifaceted chromatin remodelling complex.

Lazar, M. A. (2003). Nuclear receptor corepressors. *Nuclear Receptor Signaling*, 1(1):nrs.01001.

Lee, J. T. (2012). Epigenetic regulation by long noncoding RNAs. *Science*, 338(6113):1435–1439.

Lee, J. T. and Bartolomei, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*, 152(6):1308–1323.

Lee, J. T., Davidow, L. S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics*, 21(4):400–404.

Légaré, S., Cavallone, L., Mamo, A., Chabot, C., Sirois, I., Magliocco, A., Klimowicz, A., Tonin, P. N., Buchanan, M., Keilty, D., Hassan, S., Laperrière, D., Mader, S., Aleynikova, O., and Basik, M. (2015). The estrogen receptor cofactor SPEN functions as a tumor suppressor and candidate biomarker of drug responsiveness in hormone-dependent breast cancers. *Cancer Research*, 75(20):4351–4363.

Légaré, S., Chabot, C., and Basik, M. (2017). SPEN, a new player in primary cilia formation and cell migration in breast cancer. *Breast Cancer Research*, 19(1):104.

Li, X. and Fu, X. D. (2019). Chromatin-associated RNAs as facilitators of functional genomic interactions.

Lin, H., Gupta, V., Vermilyea, M. D., Falciani, F., Lee, J. T., O'Neill, L. P., and Turner, B. M. (2007). Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. *PLoS Biology*, 5(12):2809–2820.

Lin, H. V., Doroquez, D. B., Cho, S., Chen, F., Rebay, I., and Cadigan, K. M. (2003). Splits ends is a tissue/promoter specific regulator of Wingless signaling. *Development*, 130(14):3125–3135.

Liu, B., Mitani, Y., Rao, X., Zafereo, M., Zhang, J., Zhang, J., Futreal, P. A., Lozano, G., and El-Naggar, A. K. (2017). Spatio-Temporal Genomic Heterogeneity, Phylogeny, and Metastatic Evolution in Salivary Adenoid Cystic Carcinoma. *Journal of the National Cancer Institute*, 109(10).

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. *Nature Communications*, 8(1).

Loda, A. and Heard, E. (2019). Xist RNA in action: Past, present, and future.

Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D. H., Bulleit, R. F., and Wood, K. V. (2008). HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chemical Biology*, 3(6):373–382.

Lu, J., Jeong, H.-W., Kong, N., Yang, Y., Carroll, J., Luo, H. R., Silberstein, L. E., YupoMa, and Chai, L. (2009). Correction: Stem Cell Factor SALL4 Represses the Transcriptions of PTEN and SALL1 through an Epigenetic Repressor Complex. *PLoS ONE*, 4(7).

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(5):1267–1279.

Lucchesi, J. C., Kelly, W. G., and Panning, B. (2005). Chromatin Remodeling in Dosage Compensation. *Annual Review of Genetics*, 39(1):615–651.

Ludewig, A. H., Kober-Eisermann, C., Weitzel, C., Bethke, A., Neubert, K., Gerisch, B., Hutter, H., and Antebi, A. (2004). A novel nuclear receptor/coregulator complex controls C. elegans lipid metabolism, larval development, and aging. *Genes and Development*, 18(17):2120–2133.

Luikenhuis, S., Wutz, A., and Jaenisch, R. (2001). Antisense Transcription through the Xist Locus Mediates Tsix Function in Embryonic Stem Cells. *Molecular and Cellular Biology*, 21(24):8512–8520.

Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nature Structural and Molecular Biology*, 17(10):1195–1201.

Lyon, M. F. (1961). Gene action in the X-chromosome of the mouse (mus musculus L.). *Nature*, 190(4773):372–373.

Lyon, M. F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. *American journal of human genetics*, 14:135–148.

Ma, H., Song, B., Guo, S., Li, G., and Jin, G. (2020). Identification of germline and somatic mutations in pancreatic adenosquamous carcinoma using whole exome sequencing. *Cancer Biomarkers*, 27(3):389–397.

Mace, K. and Tugores, A. (2004). The product of the split ends gene is required for the maintenance of positional information during Drosophila development. *BMC Developmental Biology*, 4:15.

Mak, W., Baxter, J., Silva, J., Newall, A. E., and Otte, A. P. (2002). Mitotically stable association of polycomb group proteins Eed and Enx1 with the inactive X chromosome in trophoblast stem cells. *Current Biology*, 12(12):1016–1020.

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(5658):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 and Development*, 11(2):156–166.

Markaki, Y., Gan Chong, J., Luong, C., Tan, S. Y., Wang, Y., Jacobson, E. C., Maestrini, D., Dror, I., Mistry, B. A., Schöneberg, J., Banerjee, A., Guttman, M., Chou, T., and Plath, K. (2020). Xist-seeded nucleation sites form local concentration gradients of silencing proteins to inactivate the X-chromosome Equal contribution. *bioRxiv*, page 2020.11.22.393546.

Marks, H., Kerstens, H. H. D., Barakat, T. S., Splinter, E., Dirks, R. A. M., van Mierlo, G., Joshi, O., Wang, S.-Y., Babak, T., Albers, C. A., Kalkan, T., Smith, A., Jouneau, A., de Laat, W., Gribnau, J., and Stunnenberg, H. G. (2015). Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. *Genome Biology*, 16(1):149.

Marshall Graves, J. A. and Shetty, S. (2001). Sex from W to Z: Evolution of vertebrate sex chromosomes and sex determining genes. In *Journal of Experimental Zoology*, volume 290, pages 449–462.

Masui, O., Heard, E., and Koseki, H. (2018). Live Imaging of Xist RNA. pages 67–72.

Maurus, K., Appenzeller, S., Roth, S., Kuper, J., Rost, S., Meierjohann, S., Arampatzi, P., Goebeler, M., Rosenwald, A., Geissinger, E., and Wobser, M. (2018). Panel Sequencing Shows Recurrent Genetic FAS Alterations in Primary Cutaneous Marginal Zone Lymphoma. *Journal of Investigative Dermatology*, 138(7):1573–1581.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D. L. (1997). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes and Development*, 11(24):3306–3318.

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(7551):232–236.

Meinhart, A. and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature*, 430(6996):223–226.

Migeon, B. R. (2016). An overview of X inactivation based on species differences. *Seminars in Cell and Developmental Biology*, 56:111–116.

Mikami, S., Kanaba, T., Takizawa, N., Kobayashi, A., Maesaki, R., Fujiwara, T., Ito, Y., and Mishima, M. (2014). Structural insights into the recruitment of SMRT by the corepressor SHARP under phosphorylative regulation. *Structure*, 22(1):35–46.

Miller, A., Ralser, M., Kloet, S. L., Loos, R., Nishinakamura, R., Bertone, P., Vermeulen, M., and Hendrich, B. (2016). Sall4 controls differentiation of pluripotent cells independently of the nucleosome remodelling and deacetylation (NuRD) complex. *Development (Cambridge)*, 143(17):3074–3084.

Minajigi, A., Froberg, J. E., 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(6245):aab2276–aab2276.

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 reports*, 12(4):562–72.

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell Reports*, 12(4):554–561.

Monk, M. and Harper, M. I. (1979). Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos [18]. *Nature*, 281(5729):311–313.

Morey, L. and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends in Biochemical Sciences*, 35(6):323–332.

Morris, D. P. and Greenleaf, A. L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA Polymerase II. *Journal of Biological Chemistry*, 275(51):39935–39943.

Muller, H. J. (1918). Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of Balanced Lethal Factors. *Genetics*, 3(5):422–99.

Nagy, L., Kao, H. Y., Chakravarti, D., Lin, R. J., Hassig, C. A., Ayer, D. E., Schreiber, S. L., and Evans, R. M. (1997). Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell*, 89(3):373–380.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. *Science*, 321(5896):1693–1695.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A. S., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2010). Molecular coupling of Tsix regulation and pluripotency. *Nature*, 468(7322):457–460.

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., and Brockdorff, N. (2019). Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nature Communications*, 10(1):3129.

Newberry, E. P., Latifi, T., and Towler, D. A. (1999). The RRM domain of MINT, a novel Msx2 binding protein, recognizes and regulates the rat osteocalcin promoter. *Biochemistry*, 38(33):10678–10690.

Ng, H. H., Robert, F., Young, R. A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular Cell*, 11(3):709–719.

Nikolakaki, E., Mylonis, I., and Giannakouros, T. (2017). Lamin B Receptor: Interplay between Structure, Function and Localization. *Cells*, 6(3):28.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods*, 6(12):917–922.

Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., and Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, 169(5):930–944.e22.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., and Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–385.

Normanno, N., De Luca, A., Bianco, C., Strizzi, L., Mancino, M., Maiello, M. R., Carotenuto, A., De Feo, G., Caponigro, F., and Salomon, D. S. (2006). Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*, 366(1):2–16.

Norris, D. P., Brockdorff, N., and Rastan, S. (1991). Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mammalian Genome*, 1(2):78–83.

Ohno, S. (1967). Sex chromosomes and sex linked genes. *BerlinSpringer*.

Ohno, S., Kaplan, W. D., and Kinosita, R. (1958). Somatic association of the positively heteropycnotic X-chromosomes in female mice (Mus musculus). *Experimental Cell Research*, 15(3):616–618.

Ohno, S., Kaplan, W. D., and Kinosita, R. (1959). Formation of the sex chromatin by a single X-chromosome in liver cells of Rattus norvegicus. *Experimental Cell Research*, 18(2):415–418.

Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A. P., Disteche, C. M., Avner, P., and Heard, E. (2005). Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature*, 438(7066):369–373.

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(5658):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(7343):370–374.

Oswald, F., Kostezka, U., Astrahantseff, K., Bourteele, S., Dillinger, K., Zechner, U., Ludwig, L., Wilda, M., Hameister, H., Knöchel, W., Liptay, S., and Schmid, R. M. (2002). SHARP is a novel component of the Notch/RBP-Jκ signalling pathway. *EMBO Journal*, 21(20):5417–5426.

Oswald, F., Rodriguez, P., Giaimo, B. D., Antonello, Z. A., Mira, L., Mittler, G., Thiel, V. N., Collins, K. J., Tabaja, N., Cizelsky, W., Rothe, M., Kühl, S. J., Kühl, M., Ferrante, F., Hein, K., Kovall, R. A., Dominguez, M., and Borggrefe, T. (2016). A phospho-dependent mechanism involving NCoR and KMT2D controls a permissive chromatin state at Notch target genes. *Nucleic Acids Research*, 44(10):4703–4720.

Patil, D. P., Chen, C.-K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S. R. (2016). m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*, 537(7620):369–373.

Patrat, C., Okamoto, I., Diabangouaya, P., Vialon, V., Le Baccon, P., Chow, J., and Heard, E. (2009). Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proceedings of the National Academy of Sciences*, 106(13):5198–5203.

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561):131–137.

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(4):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. (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Molecular Cell*, 68(5):955–969.e10.

Pinter, S. F., Sadreyev, R. I., Yildirim, E., Jeon, Y., Ohsumi, T. K., Borowsky, M., and Lee, J. T. (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Research*, 22(10):1864–1876.

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(5616):131–135.

Pontier, D. B. and Gribnau, J. (2011). Xist regulation and function eXplored. *Human Genetics*, 130(2):223–236.

Poullet, P., Carpentier, S., and Barillot, E. (2007). myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *Proteomics*, 7(15):2553–2556.

Querenet, M., Goubard, V., Chatelain, G., Davoust, N., and Mollereau, B. (2015). Spen is required for pigment cell survival during pupal development in Drosophila. *Developmental Biology*, 402(2):208–215.

Rao, S. S., 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., and Aiden, E. L. (2014).

A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7):1665–1680.

Rastan, S. (1982). Timing of X-chromosome inactivation in postimplantation mouse embryos. *Journal of embryology and experimental morphology*, 71:11–24.

Rastan, S. (1983). Non-random X-chromosome inactivation in mouse X-autosome translocation embryos - location of the inactivation centre. *Journal of Embryology and Experimental Morphology*, VOL. 78:1–22.

Rastan, S. and Robertson, E. J. (1985). X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *Journal of embryology and experimental morphology*, 90:379–88.

Rattka, M., Westphal, S., Gahr, B. M., Just, S., and Rottbauer, W. (2021). Spen deficiency interferes with Connexin 43 expression and leads to heart failure in zebrafish. *Journal of molecular and cellular cardiology*.

Rebay, I., Chen, F., Hsiao, F., Kolodziej, P. A., Kuang, B. H., Laverty, T., Suh, C., Voas, M., Williams, A., and Rubin, G. M. (2000). A genetic screen for novel components of the Ras/Mitogen-activated protein kinase signaling pathway that interact with the yan gene of Drosophila identifies split ends, a new RNA recognition motif-containing protein. *Genetics*, 154(2):695–712.

Rego, A., Sinclair, P. B., Tao, W., Kireev, I., and Belmont, A. S. (2008). The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure. *Journal of Cell Science*, 121(7):1119–1127.

Rice, W. R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics*, 116(1):161–167.

Rice, W. R. (1996). Evolution of the Y Sex Chromosome in Animals. *BioScience*, 46(5):331–343.

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., and Coverley, D. (2017). The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes and Development*, 31(9):876–888.

Riising, E. M., Comet, I., Leblanc, B., Wu, X., Johansen, J. V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. *Molecular Cell*, 55(3):347–360.

Robert-Finestra, T., Tan, B. F., Mira-Bontenbal, H., Timmers, E., Gontan-Pardo, C., Merzouk, S., Giaimo, B. D., Dossin, F., van IJcken, W. F. J., Martens, J. W. M., Borggrefe, T., Heard, E., and Gribnau, J. (2020). SPEN is Required for Xist Upregulation during Initiation of X Chromosome Inactivation. *bioRxiv*.

Rossi, D., Trifonov, V., Fangazio, M., Bruscaggin, A., Rasi, S., Spina, V., Monti, S., Vaisitti, T., Arruga, F., Famà, R., Ciardullo, C., Greco, M., Cresta, S., Piranda, D., Holmes, A., Fabbri, G., Messina, M., Rinaldi, A., Wang, J., Agostinelli, C., Piccaluga, P. P., Lucioni, M., Tabbò, F., Serra, R., Franceschetti, S., Deambrogi, C., Daniele, G., Gattei, V., Marasca, R., Facchetti, F., Arcaini, L., Inghirami, G., Bertoni, F., Pileri, S. A., Deaglio, S., Foà, R., Dalla-Favera, R., Pasqualucci, L., Rabadan, R., and Gaidano, G. (2012). The coding genome of splenic marginal zone lymphoma: Activation of NOTCH2 and other pathways regulating marginal zone development. *Journal of Experimental Medicine*, 209(9):1537–1551.

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation.

Sado, T., Fenner, M. H., Tan, S. S., Tam, P., Shioda, T., and Li, E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: Distinct effect of hypomethylation on imprinted and random X inactivation. *Developmental Biology*, 225(2):294–303.

Sakata, Y., Nagao, K., Hoki, Y., Sasaki, H., Obuse, C., and Sado, T. (2017). Defects in dosage compensation impact global gene regulation in the mouse trophoblast. *Development (Cambridge)*, 144(15):2784–2797.

Sakhdari, A., Ok, C. Y., Patel, K. P., Kanagal-Shamanna, R., Yin, C. C., Zuo, Z., Hu, S., Routbort, M. J., Luthra, R., Medeiros, L. J., Khoury, J. D., and Loghavi, S. (2019). TP53 mutations are common in mantle cell lymphoma, including the indolent leukemic non-nodal variant. *Annals of Diagnostic Pathology*, 41:38–42.

Sánchez-Pulido, L., Rojas, A. M., van Wely, K. H., Martinez-A, C., and Valencia, A. (2004). SPOC: A widely distributed domain associated with cancer, apoptosis and transcription. *BMC Bioinformatics*, 5:91.

Savarese, F., Flahndorfer, K., Jaenisch, R., Busslinger, M., and Wutz, A. (2006). Hematopoietic Precursor Cells Transiently Reestablish Permissiveness for XInactivation. *Molecular and Cellular Biology*, 26(19):7167–7177.

Schmitz, R., Wright, G. W., Huang, D. W., Johnson, C. A., Phelan, J. D., Wang, J. Q., Roulland, S., Kasbekar, M., Young, R. M., Shaffer, A. L., Hodson, D. J., Xiao, W., Yu, X., Yang, Y., Zhao, H., Xu, W., Liu, X., Zhou, B., Du, W., Chan, W. C., Jaffe, E. S., Gascoyne, R. D., Connors, J. M., Campo, E., Lopez-Guillermo, A., Rosenwald,

A., Ott, G., Delabie, J., Rimsza, L. M., Tay Kuang Wei, K., Zelenetz, A. D., Leonard, J. P., Bartlett, N. L., Tran, B., Shetty, J., Zhao, Y., Soppet, D. R., Pittaluga, S., Wilson, W. H., and Staudt, L. M. (2018). Genetics and pathogenesis of diffuse large B-Cell lymphoma. *New England Journal of Medicine*, 378(15):1396–1407.

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. *The EMBO Journal*, 25(13):3110–3122.

Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T. M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Molecular Cell*, 61(2):305–314.

Schulz, E. G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell*, 14(2):203–216.

Schwalb, B., Michel, M., Zacher, B., Hauf, K. F., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*, 352(6290):1225–1228.

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 & development*, 15(9):1140–51.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H., 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. *Developmental Cell*, 4(4):481–495.

Skene, P. J. and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, 6.

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. *Epigenetics and Chromatin*, 7(1):8.

Søgaard, T. M. M. and Svejstrup, J. Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *Journal of Biological Chemistry*, 282(19):14113–14120.

Sonmez, C., Bäurle, I., Magusin, A., Dreos, R., Laubinger, S., Weigel, D., and Dean, C. (2011). RNA 3′ processing functions of Arabidopsis FCA and FPA limit intergenic transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 108(20):8508–8513.

Souilhol, C., Cormier, S., Tanigaki, K., Babinet, C., and Cohen-Tannoudji, M. (2006). RBP-Jκ-Dependent Notch Signaling Is Dispensable for Mouse Early Embryonic Development. *Molecular and Cellular Biology*, 26(13):4769–4774.

Splinter, E., de Wit, E., Nora, E. P., Klous, P., van de Werken, H. J., Zhu, Y., Kaaij, L. J., van Ijcken, W., Gribnau, J., Heard, E., and de Laat, W. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes and Development*, 25(13):1371–1383.

Stavropoulos, N., Lu, N., and Lee, J. T. (2001). A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18):10232–10237.

Stephens, P. J., Davies, H. R., Mitani, Y., Van Loo, P., Shlien, A., Tarpey, P. S., Papaemmanuil, E., Cheverton, A., Bignell, G. R., Butler, A. P., Gamble, J., Gamble, S., Hardy, C., Hinton, J., Jia, M., Jayakumar, A., Jones, D., Latimer, C., McLaren, S., McBride, D. J., Menzies, A., Mudie, L., Maddison, M., Raine, K., Nik-Zainal, S., O'Meara, S., Teague, J. W., Varela, I., Wedge, D. C., Whitmore, I., Lippman, S. M., McDermott, U., Stratton, M. R., Campbell, P. J., El-Naggar, A. K., and Futreal, P. A. (2013). Whole exome sequencing of adenoid cystic carcinoma. *Journal of Clinical Investigation*, 123(7):2965–2968.

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). *Proceedings of the National Academy of Sciences of the United States of America*, 114(40):10654–10659.

Takagi, N. and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, 256(5519):640–642.

Tan, X., Calderon-Villalobos, L. I. A., Sharon, M., Zheng, C., Robinson, C. V., Estelle, M., and Zheng, N. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*, 446(7136):640–645.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V. I., Lao, K., and Surani, M. A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols*, 5(3):516–535.

Tian, D., Sun, S., and Lee, J. T. (2010). The long noncoding RNA, Jpx, Is a molecular switch for X chromosome inactivation. *Cell*, 143(3):390–403.

Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E., and Schreiber, S. L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*, 395(6705):917–921.

Tsuji, M., Shinkura, R., Kuroda, K., Yabe, D., and Honjo, T. (2007). Msx2-interacting nuclear target protein (Mint) deficiency reveals negative regulation of early thymocyte differentiation by Notch/RBP-J signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 104(5):1610–1615.

Vallot, C., Patrat, C., Collier, A. J., Huret, C., Casanova, M., Liyakat Ali, T. M., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P. J., and Rougeulle, C. (2017). XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell*, 20(1):102–111.

Valot, B., Langella, O., Nano, E., and Zivy, M. (2011). MassChroQ: A versatile tool for mass spectrometry quantification. *Proteomics*, 11(17):3572–3577.

van Bemmel, J. G., Mira-Bontenbal, H., and Gribnau, J. (2016). Cis- and trans-regulation in X inactivation. *Chromosoma*, 125(1):41–50.

Vanderwielen, B. D., Yuan, Z., Friedmann, D. R., and Kovall, R. A. (2011). Transcriptional repression in the Notch pathway: Thermodynamic characterization of CSL-MINT (Msx2-interacting nuclear target protein) complexes. *Journal of Biological Chemistry*, 286(17):14892–14902.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature Structural and Molecular Biology*, 15(8):795–804.

Vicoso, B. and Charlesworth, B. (2006). Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews Genetics*, 7(8):645–653.

Vojnic, E., Simon, B., Strahl, B. D., Sattler, M., and Cramer, P. (2006). Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription. *Journal of Biological Chemistry*, 281(1):13–15.

Wade, P. A., Jones, P. L., Vermaak, D., and Wolffe, A. P. (1998). A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. *Current Biology*, 8(14):843–848.

Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J. C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nature Genetics*, 28(4):371–375.

West, J. D., Frels, W. I., Chapman, V. M., and Papaioannou, V. E. (1977). Preferential expression of the maternally derived X chromosome in the mouse yolk sac. *Cell*, 12(4):873–882.

Wiellette, E. L., Harding, K. W., Mace, K. A., Ronshaugen, M. R., Wang, F. Y., and McGinnis, W. (1999). spen encodes an RNP motif protein that interacts with Hox pathways to repress the development of head-like sclerites in the Drosophila trunk. *Development*, 126(23):5373–5385.

Wutz, A. and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell*, 5(4):695–705.

Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature Genetics*, 30(2):167–174.

Xue, Y., Wong, J., Moreno, G. T., Young, M. K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Molecular Cell*, 2(6):851–861.

Yabe, D., Fukuda, H., Aoki, M., Yamada, S., Takebayashi, S., Shinkura, R., Yamamoto, N., and Honjo, T. (2007). Generation of a conditional knockout allele for mammalian spen protein Mint/SHARP. *Genesis*, 45(5):300–306.

Yamada, N., Hasegawa, Y., Yue, M., Hamada, T., Nakagawa, S., and Ogawa, Y. (2015). Xist Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-Chromosome. *PLoS Genetics*, 11(8):e1005430.

Yang, F., Babak, T., Shendure, J., and Disteche, C. M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Research*, 20(5):614–622.

Yildirim, E., Kirby, J. E., Brown, D. E., Mercier, F. E., Sadreyev, R. I., Scadden, D. T., and Lee, J. T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell*, 152(4):727–742.

Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., and Jones, K. A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes and Development*, 21(2):160–174.

Yu, B., Qi, Y., Li, R., Shi, Q., Satpathy, A. T., and Chang, H. Y. (2021). B cell-specific XIST complex enforces X-inactivation and restrains atypical B cells. *Cell*, 0(0).

Yu, X., Martin, P. G., and Michaels, S. D. (2019). BORDER proteins protect expression of neighboring genes by promoting 3′ Pol II pausing in plants. *Nature Communications*, 10(1):1–15.

Yuan, Z., VanderWielen, B. D., Giaimo, B. D., Pan, L., Collins, C. E., Turkiewicz, A., Hein, K., Oswald, F., Borggrefe, T., and Kovall, R. A. (2019). Structural and Functional Studies of the RBPJ-SHARP Complex Reveal a Conserved Corepressor Binding Site. *Cell Reports*, 26(4):845–854.e6.

Zaccara, S., Ries, R. J., and Jaffrey, S. R. (2019). Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology*, 20(10):608–624.

Zhan, T., Rindtorff, N., and Boutros, M. (2017). Wnt signaling in cancer. *Oncogene*, 36(11):1461–1473.

Zhang, L. F., Huynh, K. D., and Lee, J. T. (2007). Perinucleolar Targeting of the Inactive X during S Phase: Evidence for a Role in the Maintenance of Silencing. *Cell*, 129(4):693–706.

Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (1997). Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell*, 89(3):357–364.

Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell*, 95(2):279–289.

Zhang, Y., Rataj, K., Simpson, G. G., and Tong, L. (2016). Crystal structure of the SPOC domain of the arabidopsis flowering regulator FPA. *PLoS ONE*, 11(8).

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, 322(5902):750–756.

Zhu, B., Zheng, Y., Pham, A. D., Mandal, S. S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: The factors involved and their roles in HOX gene regulation. *Molecular Cell*, 20(4):601–611.

Zylicz, J. J., Borensztein, M., Wong, F. C., Huang, Y., Lee, C., Dietmann, S., and Surani, M. A. (2018). G9a regulates temporal preimplantation developmental program and lineage segregation in blastocyst. *eLife*, 7.

Żylicz, J. J., Bousard, A., Žumer, 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(1-2):182–197.e23.

Part VI

Appendix

Appendix A

Additional publications

A.1 Article 3: m^6A RNA methylation regulates the fate of endogenous retroviruses *(Nature. 2021)*

Chelmicki T, Roger E, Teissandier A, Dura M, Bonneville LA, Rucli S, **Dossin F**, Fouassier C, Lameiras S, Bourc'his D. *m^6A RNA methylation regulates the fate of endogenous retroviruses*. **Nature**. 2021; 591(7859):312–316.

ARTICLE ON NEXT PAGE

Article

m^{6}A RNA methylation regulates the fate of endogenous retroviruses

https://doi.org/10.1038/s41586-020-03135-1

Received: 4 March 2020

Accepted: 30 November 2020

Published online: 13 January 2021

Check for updates

Tomasz Chelmicki[1]✉, Emeline Roger[1], Aurélie Teissandier[1], Mathilde Dura[1], Lorraine Bonneville[1], Sofia Rucli[1], François Dossin[2], Camille Fouassier[3], Sonia Lameiras[4] & Deborah Bourc'his[1]✉

Endogenous retroviruses (ERVs) are abundant and heterogenous groups of integrated retroviral sequences that affect genome regulation and cell physiology throughout their RNA-centred life cycle[1]. Failure to repress ERVs is associated with cancer, infertility, senescence and neurodegenerative diseases[2,3]. Here, using an unbiased genome-scale CRISPR knockout screen in mouse embryonic stem cells, we identify m^{6}A RNA methylation as a way to restrict ERVs. Methylation of ERV mRNAs is catalysed by the complex of methyltransferase-like METTL3–METTL14[4] proteins, and we found that depletion of METTL3–METTL14, along with their accessory subunits WTAP and ZC3H13, led to increased mRNA abundance of intracisternal A-particles (IAPs) and related ERVK elements specifically, by targeting their 5′ untranslated region. Using controlled auxin-dependent degradation of the METTL3–METTL14 enzymatic complex, we showed that IAP mRNA and protein abundance is dynamically and inversely correlated with m^{6}A catalysis. By monitoring chromatin states and mRNA stability upon METTL3–METTL14 double depletion, we found that m^{6}A methylation mainly acts by reducing the half-life of IAP mRNA, and this occurs by the recruitment of the YTHDF family of m^{6}A reader proteins[5]. Together, our results indicate that RNA methylation provides a protective effect in maintaining cellular integrity by clearing reactive ERV-derived RNA species, which may be especially important when transcriptional silencing is less stringent.

Mammalian genomes host millions of retrotransposons, including ERVs that derive from past retroviral infections and have integrated as permanent residents. Over the course of evolution, successive waves of ERVs have multiplied and diversified, providing a fertile ground for genomic innovations. However, ERVs potentially compromise genomic integrity by disrupting genome structure and expression[6]. In laboratory mice, roughly 12% of pathological mutations result from ERV integrations, half of which emanate from a single family of the ERVK class, the IAPs, that comprise approximately 2,800 full-length copies[7]. By contrast, human ERVs are mostly transposition-defective[8]. However, by providing *cis*-regulatory modules, ERVs can also divert regulatory networks and alter cellular states. Moreover, ERVs generate RNA, cDNA, RNA–DNA hybrid species and proteins, the accumulation of which is associated with and may contribute to senescence, cancer and neurodegenerative diseases[3].

Homeostatic regulation of ERVs is achieved by surveillance at different steps of the ERV life cycle. Notably, chromatin-based silencing by DNA methylation and histone modifications and post-transcriptional control through RNA editing and RNA interference have extensively been characterized[9]. However, these control mechanisms are not active in all cell types or developmental periods. To identify unknown ERV-limiting factors, we performed a CRISPR–Cas9 loss-of-function screen for IAPEz control, a highly active mouse ERV. We engineered mouse embryonic stem (ES) cells to carry constitutively expressed Cas9 and a reporter cassette with IAPEz regulatory elements: IAPEz(5′LTR-UTR+3-60n *gag*)-GFP-Blast^R (Supplementary Table 1), in which 'LTR' denotes long terminal repeat and 'UTR' denotes the untranslated region (Fig. 1a, Extended Data Fig. 1a). Placing a doxycycline (dox)-responsive promoter upstream of the LTR sequence allowed us to test reactivation of the reporter after dox induction, to choose blasticidin-resistance over GFP as a more sensitive marker, and to adjust the blasticidin concentration for selection (Extended Data Fig. 1b, c). We also showed that the IAPEz reporter responded to known IAP repressors, by transducing cells with single guide RNAs (sgRNAs) against KAP1[10] (Extended Data Fig. 1d, e, Supplementary Fig. 1, Supplementary Tables 2, 3).

For the screen, we transduced IAPEz-reporter cells with a lentiviral genome-wide sgRNA library at multiplicity of infection (MOI) of 0.2–0.3 (Fig. 1b). The frequencies of sgRNAs after the application of blasticidin (5, 7 and 9 days) versus non-selected conditions were assessed via sequencing, and candidate genes were identified using MAGeCK (model-based analysis of genome-wide CRISPR–Cas9 knockout)[11]. Selection efficiency was verified by dropout of control intergenic sgRNAs, and genes were ranked on the basis of sgRNA *P* values (Fig. 1c, Supplementary Tables 4, 5). Several known IAP-repressing genes were among the top 100 hits: *Resf1*, *Trp53*, *Daxx*, *Atrx*, *Uhrf1*, *Cbx1* and *Dnmt1*[12–15]. Moreover, we identified several previously unknown candidates for IAP control (Supplementary Table 4). Notably, among the top hits were regulators of the *N*^{6}-methyladenosine (m^{6}A) mRNA methylation pathway, such as

[1]Institut Curie, PSL Research University, INSERM U934, CNRS UMR3215, Paris, France. [2]European Molecular Biology Laboratory, Heidelberg, Germany. [3]CRISPR'IT Genetic Platform Screening, Institut Curie, Paris, France. [4]ICGex Next-Generation Sequencing Platform, Institut Curie, PSL Research University, Paris, France. ✉e-mail: tomasz.chelmicki@curie.fr; deborah.bourchis@curie.fr

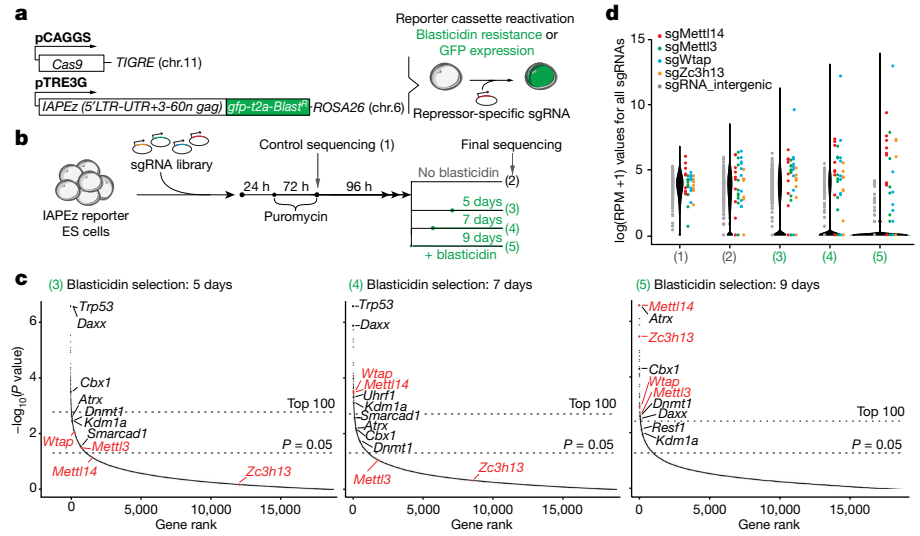


Fig. 1 | CRISPR screen for IAP suppressors in mouse ES cells. a, Schematic of the IAPEz reporter and experimental rationale. **b**, Workflow and timelines for the CRISPR–Cas9 knockout screen (screen I). **c**, Ranked *P* values (permutation test by MAGeCK[11]; Methods) for enriched genes after 5 (left), 7 (middle) and 9 (right) days of blasticidin treatment. Discontinued lines indicate genes ranked in top 100 according to *P* value (top) and $P = 0.05$ (bottom). Known IAP regulators (black, when $P < 0.05$) and *Mettl3*, *Mettl14*, *Wtap* and *Zc3h13* (red) are reported. **d**, Violin plots for all sgRNA read counts represented as the logarithm of reads per million (RPM) after library introduction (1), without selection (2) and after blasticidin selection (3–5); sgRNAs targeting *Mettl14* (sgMettl14; red, $n = 10$), *Mettl3* (sgMettl4; green, $n = 10$), *Wtap* (sgWtap; blue, $n = 10$), *Zc3h13* (shZc3h13; orange, $n = 9$) and intergenic sequences (sgRNA_intergenic; grey, $n = 199$).

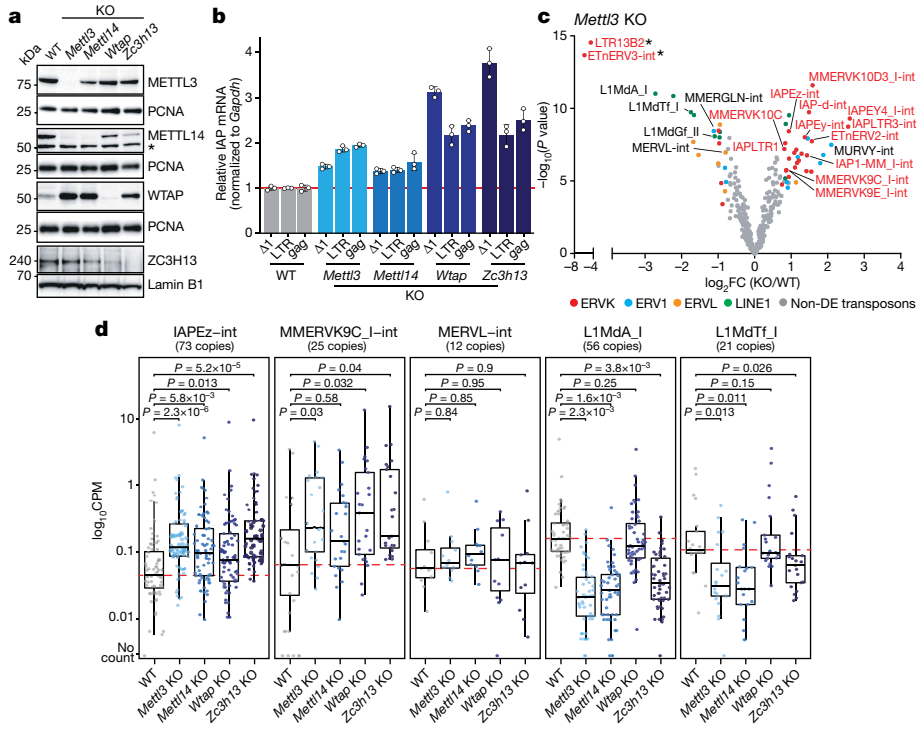


Fig. 2 | Depletion of the m^{6}A methyltransferase complex increases IAP mRNA levels. a, Immunoblot showing *Mettl3*, *Mettl14*, *Wtap* and *Zc3h13* knockout (KO) in ES cells. PCNA and lamin B1 served as loading controls; asterisk indicates unspecific band. Immunoblotting was repeated at least twice with similar results. **b**, RT–qPCR showing normalized IAP mRNA levels using primers against Δ1, LTR and *gag* sequences in the indicated knockouts relative to wild-type (WT) control (set to 1). Data are mean ± s.d. of three independent experiments. **c**, Volcano plot showing $\log_2$-transformed fold change ($\log_2$FC) in retrotransposon expression in *Mettl3* knockout versus wild-type control using random assignment of multi-mapped reads. Red, blue, orange and green denote significantly deregulated RepeatMasker annotations belonging to ERVK, ERV1, ERVL and L1 families, respectively. Grey denotes non-differentially expressed (non-DE) retrotransposons. *P* values were computed using limma and adjusted with the Benjamini–Hochberg correction (Methods). $\log_2$FC > 0.75; false discovery rate (FDR) < 0.05. **d**, Box plots showing expression ($\log_{10}$CPM; in which CPM denotes counts per million) of individual retrotransposon copies in wild-type and knockout ES cells. RNA-seq mapping allowed only unique hits in the reference genome and only copies with at least ten reads in at least one sample were conserved. Horizontal lines denote median, box limits correspond to upper and lower quartiles. *P* values determined by two-sided Student's *t*-test.

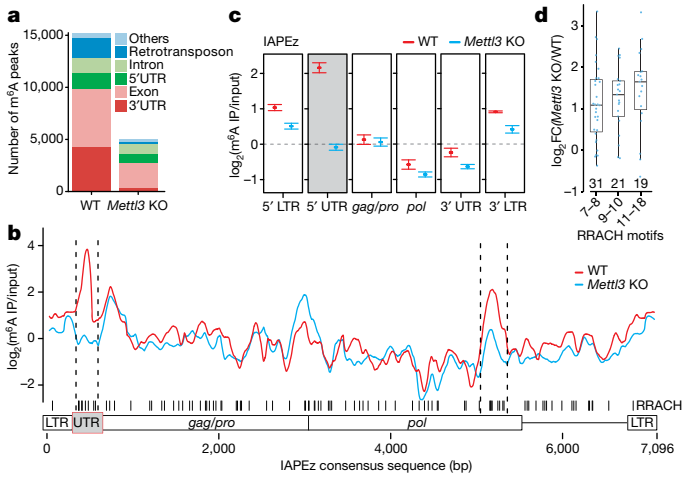


Fig. 3 | IAP mRNAs undergo m^{6}A methylation. a, Bar plot showing m^{6}A peaks within indicated genomic features. **b**, Average of input-normalized m^{6}A signal intensities along the IAPEz consensus sequence in wild-type (red) and *Mettl3* knockout (blue) cells. Vertical black lines denote RRACH motif positions. Discontinued vertical black lines delineate regions of m^{6}A enrichment present in wild-type cells and lost in *Mettl3*-knockout cells. **c**, Average of m^{6}A signal intensities for indicated IAPEz segments in wild-type (red) and *Mettl3*-knockout (blue) ES cells. Error bars indicate mean ± s.d. of three independent MeRIP-seq. **d**, Box plot showing $\log_2$FC in expression of uniquely mapped IAPEz copies grouped according to numbers of 5′ UTR-associated RRACH motifs in *Mettl3*-knockout versus wild-type cells. Box plots as in Fig. 2d. Only copies with a minimum of ten reads in at least one sample were conserved.

Mettl3, *Mettl14*, *Wtap* and *Zc3h13*, the enrichment of which gradually increased with extended blasticidin-mediated pressure (Fig. 1c, d). By repeating the screen under the most stringent selection, we confirmed significant enrichment for *Mettl3*, *Mettl14* and *Wtap* sgRNAs (Extended Data Fig. 2a, b). m^{6}A is the most abundant internal mark on mRNAs, and is crucial for organizing their fate—including export, decay and translation—in an array of biological processes such as development, cell differentiation, stress response and cancer[4]. The deposition of m^{6}A is exerted by a nuclear complex with an enzymatically active core formed by methyltransferase-like METTL3 and METTL14 proteins and other calibrating subunits, including WTAP and ZC3H13. METLL3 and METTL14 form a heterodimer, in which METTL3 is the catalytic component and METTL14 facilitates binding to the RNA substrate. WTAP and ZC3H13 are essential for assembling the complex into the nucleus[16,17].

To confirm that methylation of m^{6}A RNA regulates endogenous IAP copies, we generated individual ES cell lines that contained gene knockouts of *Mettl3*, *Mettl14*, *Wtap* and *Zc3h13* (Fig. 2a, Extended Data Fig. 3a–d). To avoid differentiation and lethality effects previously reported when m^{6}A-depleted ES cells are cultured in metastable conditions (serum plus leukaemia inhibitory factor (LIF)), we derived and cultured ES cells in a medium that stabilizes a ground state of pluripotency ('2i + LIF' medium)[18]. In these conditions, the identity of ES cells was globally preserved in the mutant lines (Extended Data Fig. 3e, f), with a mild reduction in proliferation (Extended Data Fig. 3g). We first confirmed reduced m^{6}A RNA methylation levels in the four knockout lines by enzyme-linked immunosorbent assay (ELISA) (Extended Data Fig. 4a). We used quantitative PCR with reverse transcription (qRT–PCR) to confirm that knockout of each of the four m^{6}A factors increased the abundance of endogenous IAPEz mRNA by two- to three-fold compared to wild-type cells (Fig. 2b). RNA sequencing (RNA-seq) analysis confirmed significant upregulation of IAPEz in m^{6}A mutants, as well as close relatives within the ERVK family—such as MMERVK10C, MMERVK10D3, ETnERV2 and Y chromosome-specific IAPEy elements—that share more than 65% sequence identity with IAPEz (https://www.dfam.org/) (Fig. 2c, Extended Data Fig. 4b–f, Supplementary Table 6). Specific Y-linked elements of the ERV1 family were also more abundant in the knockout lines (MuRVY). Upregulation of ERVK was observed at the subfamily level after random assignment of multi-mapped reads, and confirmed at the level of individual ERVK copies, when considering uniquely assigned reads only (Fig. 2d, Extended Data Fig. 5a). By contrast, MERVL remained globally unaffected and evolutionarily young long interspersed nuclear elements (LINEs or L1s) transcripts showed opposite downregulation in m^{6}A-knockout ES cells (Fig. 2c, d, Extended Data Figs. 4e, f, 5a). These different responses to the loss of m^{6}A mRNA methylation highlight the divergent effects that this pathway may exert depending on the retrotransposon type, with a negative effect on IAP-related ERVK elements, specifically. ERVK-specific upregulation was confirmed in published nuclear RNA-seq datasets from independent *Mettl3*-knockout ES cells[19] (Extended Data Fig. 5b).

We next evaluated the potential effect of increased IAP transcripts on gene regulation. As previously shown, hundreds of gene transcripts were upregulated in m^{6}A mutant ES cells[20], among which 941 were common between the 4 knockout lines (Extended Data Fig. 6a). However, these upregulated genes did not show correlation with proximity of ERVK annotations (−5 kb to +1 kb from the transcription start site (TSS)) (Extended Data Fig. 6b). Moreover, we did not score increased splicing between exonic sequences and IAP fragments in m^{6}A-knockout compared to wild-type ES cells (Extended Data Fig. 6c). As a whole, we conclude that the increased abundance of IAP transcripts after loss of m^{6}A loss minimal *cis*-effects on gene expression. Notably, depletion of m^{6}A did not result in downregulation of known retrotransposon repressors (Extended Data Fig. 6d) and did not alter ES cell identity, as demonstrated by expression levels of pluripotency and early differentiation genes (Extended Data Fig. 3e). This provides strong indication that m^{6}A RNA methylation directly represses IAP elements.

To verify this, we mapped the abundance and topology of m^{6}A methylation on IAPEz transcripts, by m^{6}A immunoprecipitation (methylated RNA immunoprecipitation sequencing (MeRIP–seq)) of total RNA from wild-type and *Mettl3*-knockout ES cells. We scored 15,216 and 4,864 m^{6}A peaks in wild-type and *Mettl3*-knockout ES cells, respectively, with substantially higher m^{6}A signal intensities in wild-type cells (Extended Data Fig. 7a). Notably, in addition to the well-characterized enrichment of m^{6}A methylation at the 3′ UTR and exons of genic mRNAs[20–23] (Fig. 3a, Extended Data Fig. 7b, c), we found that a considerable number of METTL3-dependent m^{6}A events mapped to retrotransposon annotations, comprising 13% of all peaks, including L1s—as recently reported[19,24]—and ERVK elements (Fig. 3a). When we plotted m^{6}A distribution along the IAPEz consensus sequence, we found two distinct regions of METTL3-dependent m^{6}A enrichment, predominantly at the 5′UTR—present in the IAPEz reporter—and to a lesser extent on the *Pol* sequence (Fig. 3b, c, Extended Data Fig. 7d). Enrichment in m^{6}A also coincided with the 5′ UTR region of MMERVK10C (Extended Data Fig. 7e, f) and was rather spread across L1MdA_I (Extended Data Fig. 7g, h). The m^{6}A RNA methylation mostly occurs on conserved RRACH sequence motifs (in which R denotes A or G, and H denotes A, C or U)[21,23]. Accordingly, we found several RRACH motifs on the 5′ UTR of IAPEz and MMERVK10C consensus sequences (Fig. 3b, Extended Data Fig. 7e, i). By focusing on uniquely mapped copies, we found that m^{6}A-dependent repression was proportional to the RRACH content: the more RRACH motifs an IAPEz copy contains in the 5′ UTR, the more upregulated it is in m^{6}A-knockout cells (Fig. 3d, Extended Data Fig. 7j). These data demonstrate for the first time, to our knowledge, that IAPs and their ERVK relatives undergo m^{6}A RNA methylation, and reveal a new pathway of retrotransposon suppression.

Functions of the m^{6}A RNA methylation complex have so far been investigated by conventional gene perturbation techniques—such as knockout or short hairpin RNA (shRNA)-mediated knockdown—which precludes examining the early consequences of m^{6}A loss and can also lead to secondary effects after prolonged selection. To address the early and direct IAP responses to m^{6}A depletion, we used auxin-inducible

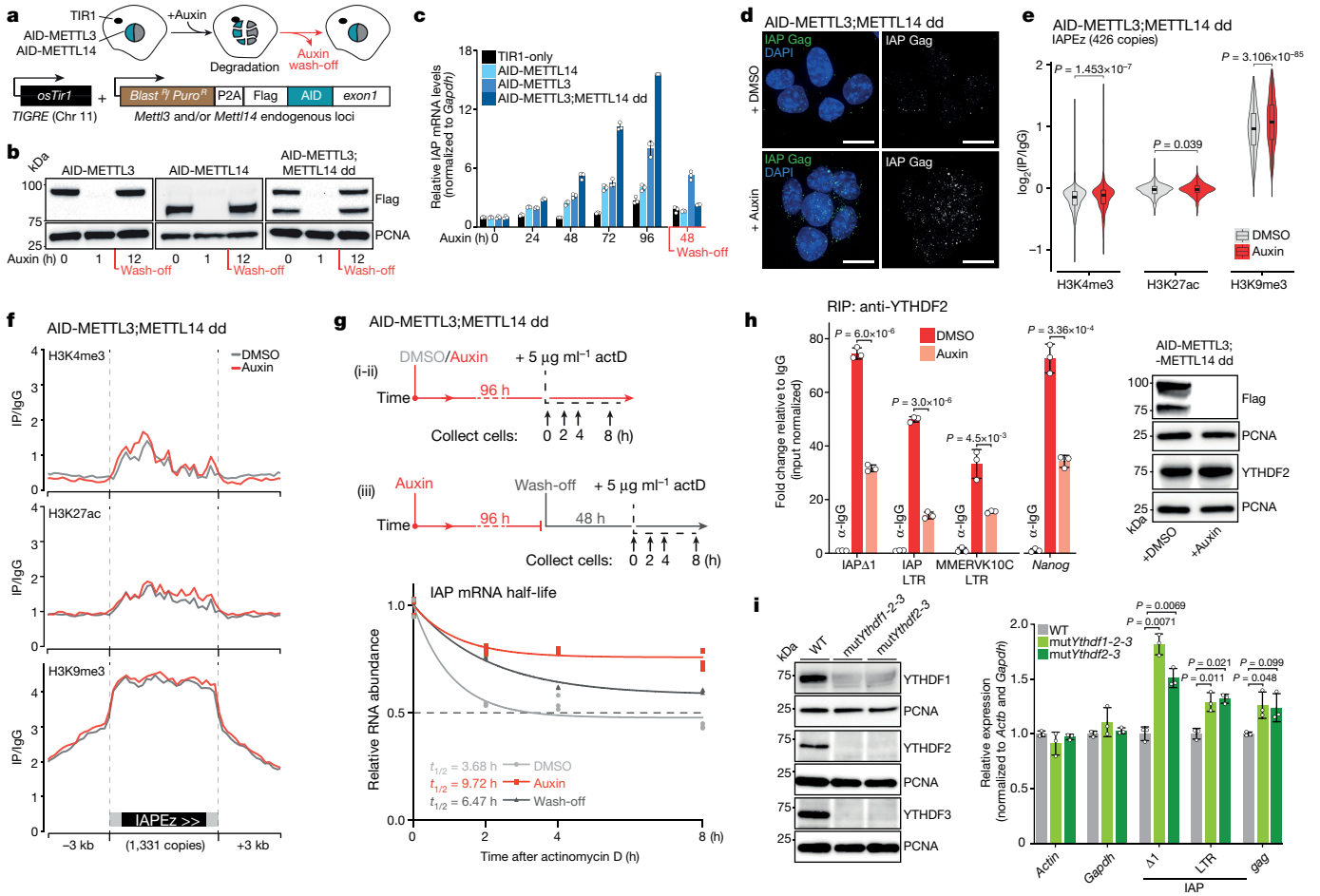


Fig. 4 | IAP chromatin status and RNA stability upon inducible degron of the m^{6}A methyltransferase complex. a, Schematic of METTL3 and METTL14 degron engineering. **b**, Immunoblot showing auxin-induced degradation of endogenous 3×Flag-AID-METTL3, -METTL14 and -METTL3;METTL4 double degron (dd). PCNA served as a loading control. **c**, RT–qPCR showing normalized IAP mRNA levels relative to 0h using LTR-specific primers after degron of METTL14 (light blue), METTL3 (blue) and METTL3;METTL14 dd (dark blue). TIR1-only (black) ES cells served as control. Error bars indicate mean ± s.d. of three independent auxin inductions. **d**, Immunofluorescence staining for IAP Gag in METTL3;METTL14 double degron after 96 h of treatment with DMSO (top) or auxin (bottom). Scale bar, 10 μm. **e**, Violin plots and box plots (inside) showing H3K4me3, H3K27ac and H3K9me3 levels at IAPEz copies (n = 426) determined by CUT&RUN in METTL3;METTL14 dd after 96 h of DMSO (grey) or auxin (red). Only copies with at least 10 reads in at least one sample were conserved. Box plots as in Fig. 2d. P values were determined by two-sided Student's t-test. **f**, Composite profiles showing H3K4me3 (top), H3K27ac (middle) and H3K9me3 (bottom) coverage along full-length (greater than 5 kb) IAPEz copies (n = 1,331) in METTL3;METTL14 double degron after 96 h of DMSO (grey) or auxin (red). **g**, Top, scheme for evaluating IAP mRNA degradation rates in METTL3;METTL14 double degron ES cells exposed to DMSO or auxin and actinomycin D (actD) for 2, 4 and 8 h before (i–ii) and after (iii) auxin wash-off. Bottom, relative IAP mRNA levels and half-life measurements in ES cells exposed to DMSO (light grey), auxin (red) and after auxin wash-off (dark grey) using Δ1-specific primers (n = 3 independent experiments). **h**, Left, YTHDF2 RIP-qPCR for ERVKs and *Nanog* in METTL3;METTL14 double degron cells after treatment for 96 h with DMSO (red) or auxin (pink). Rabbit IgG served as control. Data are mean ± s.d. of three independent experiments. P values were determined by two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli. Right, immunoblot showing METTL3 and METTL14 (Flag), and YTHDF2 in METTL3;METTL14 double degron cells after treatment with DMSO or auxin for 96 h. PCNA served as loading control. **i**, Left, immunoblots showing YTHDF1–YTHDF3 protein levels in mut*Ythdf1-2-3* and mut*Ythfd2-3* ES cells. PCNA served as loading control. Right, RT–qPCR showing normalized IAP mRNA levels (geometric mean between *Actb* and *Gapdh*) in mutant ES cells relative to wild-type cells using Δ1-, LTR- or *gag*-specific primers. Data are mean ± s.d. from three independent experiments. Immunoblots in **b**, **h** and **i** and the experiment in **d** were repeated at least twice with similar results. Results in **e** and **f** represent averages of two independent CUT&RUN.

degron (AID)[25] to control the depletion of METTL3 and METTL14, individually and in combination (Fig. 4a, Extended Data Fig. 8a, b). The addition of auxin (indole-3-acetic acid (IAA)) resulted in efficient and near-total degradation within 1 h, which persisted over prolonged treatment and was reversible after wash-off (Fig. 4b, Extended Data Fig. 8c). Depletion of METTLs was rapidly followed by substantial and sustainable decrease of m^{6}A RNA methylation levels (Extended Data Fig. 8d). Notably, we observed a progressive, time-dependent accumulation of IAP transcripts after the removal of m^{6}A (Fig. 4c, Extended Data Fig. 8e), whereas levels of cellular proliferation were unaffected (Extended Data Fig. 8f). A similar trend was observed after degron-mediated depletion of ZC3H13 (Extended Data Fig. 8g–j). After 96 h of m^{6}A depletion, levels of IAP mRNA were increased by 7-fold, 5-fold and 15-fold in single METTL3, single METTL14 or METTL3;METTL14 double degron, respectively (Fig. 4c), and this translated into accumulation of IAP-encoded Gag proteins in cytoplasmic speckles (Fig. 4d). Notably, the abundance of IAP mRNA was higher after degron depletion than in established knockouts for individual *Mettl* genes (only twofold increase in *Mettl3* or *Mettl14* knockouts) (Fig. 2b), which suggests the implementation of adaptive mechanisms after prolonged m^{6}A loss. Moreover, the relative upregulation after simultaneous METTL3 and METTL14 degron-mediated depletion compared to single degrons

of METTL3 and METTL14 highlight their functional synergy in reducing IAP mRNA levels. Finally, re-stabilizing the m^{6}A methyltransferase complex after auxin removal resulted in a rapid decline in IAP mRNAs (Fig. 4c, Extended Data Fig. 8e, j). Together, these results suggest that m^{6}A RNA methylation dynamically restrains the cellular availability of IAP mRNAs.

Increased abundance of IAP mRNA may result from transcriptional and/or post-transcriptional effects of the m^{6}A loss. Notably, m^{6}A RNA methylation has been shown to affect chromatin-based regulation of transcription[19,26]. We performed CUT&RUN (cleavage under targets and release using nuclease) experiments to profile permissive (H3K4me3 and H3K27ac) and repressive (H3K9me3) histone marks after 96 h of auxin treatment or dimethylsulfoxide (DMSO) control in METTL3;METTL14 double degron ES cells (AID-METTL3;METTL14). On a global scale, the rapid m^{6}A removal did not alter the coverage of these histone marks on m^{6}A target genes or on different retrotransposon families (Extended Data Fig. 9a–d). When focusing on full-length IAPEz annotations, we observed moderate H3K4me3 gain (Fig. 4e, f), which may contribute to the mRNA upregulation observed after AID-METTL3;METTL14 double degron depletion, although this was not accompanied by a decrease in H3K9me3 or an increase in H3K27ac (Fig. 4e, f). We therefore concluded that acute withdrawal of m^{6}A does not strongly modify chromatin-based regulation of ERVKs in the short term.

Owing to its major effect on genic mRNA lifetime[20], we sought to determine whether m^{6}A could destabilize IAP-derived mRNAs. We used METTL3;METTL14 double degron ES cells to monitor IAP RNA levels after the inhibition of transcription with actinomycin D, and then calculated the half-life rate. IAP mRNA levels and half-life rates were substantially higher after degradation of the m^{6}A methyltransferase complex than in cells treated with DMSO control (Fig. 4g). Importantly, the rapid reintroduction of m^{6}A methylation (by auxin wash-off) restored IAP mRNA degradation (Fig. 4g). The fate of m^{6}A-modified mRNAs is determined by a set of readers, among which the YTH-domain containing proteins YTHDF1, YTHDF2 and YTHDF3 have key roles by directing them to cytosolic compartments where they undergo destabilization and decay[5]. Using RNA immunoprecipitation (RIP), we found that YTHDF2 binds to IAP and MMERVK10C mRNAs (Extended Data Fig. 10a–c), and notably, this occurred in an m^{6}A-dependent manner (Fig. 4h). Finally, we reasoned that, similarly to depleting the m^{6}A writing complex, depleting YTHDF m^{6}A readers should increase IAP mRNA abundance. To account for potential redundancy[27,28], we derived two mutant ES cell lines that exhibited loss of all three YTHDF proteins (Fig. 4i, left, Extended Data Fig. 10d). Depletion of YTHDFs resulted in increased IAP mRNA levels (Fig. 4i, right), which supports the idea that the m^{6}A methylation pathway regulates IAP mRNAs in a post-transcriptional and YTHDF-dependent manner, although other mechanisms may also apply.

Our results provide evidence that m^{6}A RNA methylation directly affects ERV mRNA abundance by accelerating their clearance. This mechanism appears specific to IAP-related ERVK elements, whereas other retrotransposon types may undergo alternative directions and modes of m^{6}A-dependent regulation. Considering the abundance and deleterious effect of ERVKs, this pathway may be particularly relevant in situations of relaxation of chromatin-based control, such as in early mammalian embryos—from which ES cells are derived—or during ageing. The link with YTHDF readers indicates that ERV mRNA decay may occur through phase-partitioning into cytoplasmic processing bodies (P-bodies)[29], consistent with previous observations that IAP mRNAs localize to these structures[30]. In addition, this pathway may cooperatively prevent m^{6}A-modified ERV mRNAs from being reverse transcribed, translated or assembled into protective retrotransposition complexes, providing several layers of control. Notably, m^{6}A mostly occurs on the 5′ UTR of ERVKs, a region that contains the tRNA primer-binding site that is essential for reverse transcription. Finally, m^{6}A-dependent regulation could also dampen the immunogenic potential of ERV-derived RNA species and their ability to trigger inflammatory responses, as seen in human neurodegenerative diseases[3].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-03135-1.

1. Johnson, W. E. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat. Rev. Microbiol.* **17**, 355–370 (2019).
2. Barau, J. et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354**, 909–912 (2016).
3. Tam, O. H., Ostrow, L. W. & Gale Hammell, M. Diseases of the nERVous system: retrotransposon activity in neurodegenerative disease. *Mob. DNA* **10**, 32 (2019).
4. Shi, H., Wei, J. & He, C. Where, when, and how: context-dependent functions of RNA methylation writers, readers, and erasers. *Mol. Cell* **74**, 640–650 (2019).
5. Patil, D. P., Pickering, B. F. & Jaffrey, S. R. Reading m^{6}A in the transcriptome: m^{6}A-binding proteins. *Trends Cell Biol.* **28**, 113–127 (2018).
6. Goodier, J. L. & Kazazian, H. H., Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135**, 23–35 (2008).
7. Gagnier, L., Belancio, V. P. & Mager, D. L. Mouse germ line mutations due to retrotransposon insertions. *Mob. DNA* **10**, 15 (2019).
8. Hancks, D. C. & Kazazian, H. H. Jr. Roles for retrotransposon insertions in human disease. *Mob. DNA* **7**, 9 (2016).
9. Zamudio, N. & Bourc'his, D. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? *Heredity* **105**, 92–104 (2010).
10. Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
11. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
12. Fukuda, K., Okuda, A., Yusa, K. & Shinkai, Y. A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Res.* **28**, 846–858 (2018).
13. Liu, X. et al. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat. Commun.* **4**, 1563 (2013).
14. Sadic, D. et al. Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep.* **16**, 836–850 (2015).
15. Maksakova, I. A. et al. H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. *Epigenetics Chromatin* **4**, 12 (2011).
16. Wen, J. et al. Zc3h13 regulates nuclear RNA m6A methylation and mouse embryonic stem cell self-renewal. *Mol. Cell* **69**, 1028–1038.e6 (2018).
17. Ping, X. L. et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).
18. Greenberg, M. V. C. & Bourc'his, D. Cultural relativism: maintenance of genomic imprints in pluripotent stem cell culture systems. *Curr. Opin. Genet. Dev.* **31**, 42–49 (2015).
19. Liu, J. et al. N^6-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription. *Science* **367**, 580–586 (2020).
20. Geula, S. et al. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).
21. Dominissini, D. et al. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature* **485**, 201–206 (2012).
22. Batista, P. J. et al. m^{6}A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15**, 707–719 (2014).
23. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
24. Abakir, A. et al. N^6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nat. Genet.* **52**, 48–55 (2020).
25. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).
26. Li, Y. et al. N^6-Methyladenosine co-transcriptionally directs the demethylation of histone H3K9me2. *Nat. Genet.* **52**, 870–877 (2020).
27. Zaccara, S. & Jaffrey, S. R. A unified model for the function of YTHDF proteins in regulating m^{6}A-modified mRNA. *Cell* **181**, 1582–1595.e18 (2020).
28. Lasman, L. et al. Context-dependent functional compensation between Ythdf m^{6}A reader proteins. *Genes Dev.* **34**, 1373–1391 (2020).
29. Ries, R. J. et al. m^{6}A enhances the phase separation potential of mRNA. *Nature* **571**, 424–428 (2019).
30. Lu, C., Contreras, X. & Peterlin, B. M. P. P bodies inhibit retrotransposition of endogenous intracisternal A particles. *J. Virol.* **85**, 6244–6251 (2011).

Methods

Data reporting and statistical analysis

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. All statistical tests, resulting *P* values and observation numbers are indicated in the figure panels or legends.

Data visualization

Unless stated otherwise, heat maps, violin plots, box plots, Venn diagrams, density plots, dot plots and bar plots that visualize deep-sequencing analyses were generated using ggplot2. Box plots always show the median as the centre line, box limits correspond to upper and lower quartiles, and whiskers cover 1.5× the interquartile range. Volcano plots visualizing retrotransposon expression changes from RNA-seq and bar graphs visualizing ELISA and RT–qPCR experiments were generated using Prism 8.3.0. Dot plots and histogram visualizing FACS analyses were generated using NovoExpress software (v.1.2.1).

Cell culture

Mouse embryonic day (E) 14 ES cells were grown in two different media: serum + LIF consisted in Glasgow medium (Sigma), 15% FBS (Gibco), 2 mM L-glutamine, 0.1 mM MEM non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 0.1 mM β-mercaptoethanol, 1,000 U ml^{-1} leukaemia inhibitory factor (LIF, Miltenyi Biotec); 2i + LIF was made of 50% Neurobasal medium (Gibco), 50% DMEM/F12 (Gibco), 2 mM L-glutamine (Gibco), 0.1 mM β-mercaptoethanol, Ndiff Neuro-2 medium supplement (Millipore), B-27 medium supplement (Gibco), 1,000 U ml^{-1} LIF, 3 μM GSK3 inhibitor (CT-99021), 1 μM MEK inhibitor (PD0325901). Cells were cultured in 0.2% gelatin-coated flasks at 37 °C with 5% CO_2. Except for the CRISPR–Cas9 loss-of-function screens that were performed in serum + LIF medium, all experiments were performed in 2i + LIF medium. Mycoplasma-free status of the cell cultures was verified.

Plasmid construction

IAPEz reporter. Plasmids used to target *ROSA26* (pEN111) and *TIGRE* (also known as *Igs7*) (Addgene 92141) loci, and the *ROSA26*- and *TIGRE*-specific sgRNA-encoding plasmids (Addgene 86234 and 92144, respectively) were provided by E. Nora (UCSF). The IAPEz-5′LTR-5′UTR-gag(3-60nt) consensus sequence was obtained from RepeatMasker and/or from repbase (http://www.repeatmasker.org/) synthetized and cloned into pUC57 by GenScript (Supplementary Table 1). To make the IAPEz reporter (pTCH1), IAPEz-5′LTR-5′UTR-gag(3-60nt) sequence and the GFP-T2A-blasticidin resistance cassette (hereafter denoted as BlastR) were combined using extension PCR and inserted it into the pEN111 backbone using the ClaI site. To insert the *Cas9* gene at the *TIGRE* locus, 3× Flag-NLS-Cas9 was PCR amplified from pX459 expression vector (Addgene 62988) and inserted into plasmid 92141 backbone using BamHI and XhoI sites (pTCH2).

Plasmids for N terminus tagging with AID domain for auxin-inducible degron. The plasmids to target OsTIR1 at the *TIGRE* locus (Addgene 92141) and the *TIGRE*-specific sgRNA-encoding plasmid (see above) were provided by E. Nora (UCSF). The plasmids to target AID inserts into the gene-of-interest N terminus were generated as follows: either puromycin resistance (PuroR)-P2A-3×Flag-AID, or BlastR-P2A-3×Flag-AID inserts were cloned into pUC19 backbone (pFD71 with puromycin resistance gene and pFD75 with blasticidin S-resistance gene). Next, homology arms ranging from 320 to 530 bp depending on the gene (flanking both sides, but excluding ATG start codon) for *Mettl3*, *Mettl14* and *Zc3h13* were PCR amplified from mouse genomic DNA and inserted into pFD71 or pFD75 surrounding and in frame with AID insert using EcoRI/NcoI sites for upstream homology arms and AgeI/HindII for downstream homology arms. Final expression vectors were used as follows: pTCH3 (BlastR-P2A-3×Flag-AID-METTL3_Nter), pTCH4 (BlastR-P2A-3×Flag-AID-METTL4_Nter) and pFD119 (BlastR-P2A-3×Flag-AID-ZC3H13_Nter) to generate individual endogenous degron ES cell lines for aforementioned genes. The pTCH4 and pTCH5 (PuroR-P2A-3×Flag-AID-Mettl3) were used sequentially to generate the AID-METTL3;-METTL14 double degron ES cell line. For the sgRNA cloning, the pX459 plasmid (Addgene 62988) was digested with BbsI immediately downstream of the U6 promoter and annealed DNA duplex corresponding to the target sgRNA sequences were ligated. sgRNA sequences were chosen to overlap with the gene TSS, so that after introduction of the AID the sgRNA-specific sequences were disrupted. sgRNA sequences used for degron targeting are listed in Supplementary Table 2.

Cell transfection and clone isolation

All transgenic insertions and mutations were performed using Amaxa 4D nucleofector (Lonza). For each nucleofection, 3×10^6–5×10^6 cells were electroporated with 1–3 μg of nonlinearized targeting vector and/or sgRNA/Cas9-encoding plasmids and plated at a low density. Two days later, cells were selected with puromycin (1 μg ml^{-1}, Life Technologies) or blasticidin S (5 μg m^{-1}) for 2 and 5 days, respectively, and individual clones were picked and screened by PCR. Flippase-mediated removal of puromycin-resistance cassettes were performed for the IAPEz reporter cell line (from both *ROSA26* and *TIGRE* loci) and for puromycin resistance cassette for AID-METTL3;METTL14 double degron from the *TIGRE* locus. For the IAPEz reporter cell line, functionality of the reporter cassette was confirmed by doxycycline (dox) induction followed by FACS and fluorescence microscopy analyses while Cas9 expression and activity was confirmed by *Kap1*-specific sgRNA introduction (see below) and western blot analysis. To generate *Mettl3*, *Mettl14*, *Wtap*, *Zc3h13*, *Ythdf2* and combined *Ythdf* mutant ES cells, two sgRNAs for each gene were designed using the online CRISPOR Design Tool[31] to introduce indels and/or deletions. For sgRNA cloning, the pX459 plasmid (Addgene 62988) was digested with BbsI immediately downstream of the U6 promoter and annealed DNA duplex corresponding to the target sgRNA sequences were ligated. The *Mettl3*-knockout cells were created by deleting part of exon 4; *Mettl14*-knockout by deleting part of exon 1; *Zc3h13*-knockout by deleting part of exon 9; and sgRNA targeting of *Wtap* gene resulted in single nucleotide insertion and premature stop codon in exon 4. *Ythdf2*-knockout cells were generated by deleting part of exon 4. The mut*Ythdf1-2-3* and mut*Ythd2-3* ES cells were obtained by simultaneous introduction of six sgRNAs targeting the *Ythdf1*, *Ythdf2* and *Ythdf3* genes. Protein loss was confirmed by western blot for all mutant cell lines; in addition, m^6A ELISA assays was carried out for *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout lines. For degron lines, proper insertion and AID-fusion protein activities were confirmed by genotyping, western blot analysis and m^6A ELISA. For sgRNA sequences used for generation of knockout, mutant and degron lines see Supplementary Table 2.

Cell treatments

IAPEz reporter expression was induced after the administration of doxycycline (1 μg ml^{-1}). Auxin-mediated depletion of target proteins was achieved by supplementing culture medium with auxin (Sigma) at the recommended concentration of 500 μM. Auxin-containing medium was renewed every 24 h. For auxin wash-off, auxin-containing medium was removed, cells were rinsed twice with PBS, and exposed to auxin-free medium. Blockade of transcription with actinomycin D was achieved by supplementing culture medium with 5 μg ml^{-1} actinomycin D (Sigma).

Lentivirus production and lentiviral-based *Kap1*-specific sgRNA knockout

Two previously described sgRNAs specific to the *Kap1* gene[32] (Supplementary Table 2) were incorporated into plentiGuide-puro vector

(Addgene 52963). For production of lentiviral particles, HEK293FT cells were co-transfected with 3.33 μg of either of the Kap1 lentiGuide-puro constructs, 2.5 μg psPAX2 packaging plasmid and 1 μg pMD2.G envelope plasmid using Lipofectamine 2000 (Invitrogen). Lentiviral supernatant was collected, filtered with 0.45-μm filter, concentrated using Amicon Ultra centrifugal filter (Millipore, 100-kDa cut-off) and added to pre-plated IAPEz reporter cells supplemented with 8 μg ml^{-1} polybrene (Millipore). Twelve hours after infection, the medium was replaced and supplemented with puromycin (1 μg ml^{-1}). After 48 h of puromycin selection the medium was replaced and supplemented with blasticidin S (5 μg ml^{-1}) for additional 72 h.

Protein extraction and western blotting

Cells were trypsinized, washed once in medium and once in PBS and pelleted for 5 min at 1,000 rpm. Cell pellets were resuspended in RIPA buffer (1× PBS, 0.5% sodium deoxycholate, 0.1% SDS, 1% Igepal CA-630) containing protease inhibitors (Roche) and incubated on ice for 20 min. Lysates were then centrifuged for 20 min at 16,400 rpm, at 4 °C and supernatants were kept. Protein concentration was determined using Bradford assay and protein extracts were boiled for 10 min in LDS buffer (Life Technologies) containing 200 mM DTT. Equal amounts of protein were loaded on 4–12% Bis-Tris gel (NuPAGE), or 3–8% Tris-Acetate gel (NuPAGE) for ZC3H13 detection. Transfer was performed on a 0.45-μm nitrocellulose membrane (GE Healthcare) using wet-transfer system, blocked with 5% milk in PBS (+Igepal CA-630 to a final concentration of 0.3%) for 1 h at room temperature. Membranes were incubated with primary antibodies (Supplementary Table 3) at 4 °C overnight in 1% milk in PBS (supplemented with 0.3% Igepal CA-630), washed five times with PBS supplemented with 0.3% Igepal CA-630 and incubated with HRP-conjugated secondary antibodies for 1 h at room temperature and washed again five times. Signal was detected using LumiLight Plus Kit (Roche) on the Chemidoc MP imaging system (BioRad). For uncropped images of western blot membranes, see Supplementary Fig. 1.

FACS analysis

Cells were collected, washed with PBS to remove residual medium and proceeded to analyse GFP expression using NovoCyte 2000R (ACEA Biosciences) flow cytometer and NovoExpress software (v1.2.1). The percentage of GFP-positive cells was determined upon definition of three gates: (i) FSC-H *vs* SSC-H to isolate cells from debris, (ii) SSC-H versus SSC-A to isolate single cells and (iii) SSC-H versus FITC-H for detection of GFP-positive population. For pseudocolour plots and gating strategy see Supplementary Fig. 2.

Genome-wide screen in IAPEz reporter mouse ES cells (screens I and II)

Approximately 300 × 10^6 IAPEz reporter ES cells expressing Cas9 were lentivirally infected with a genome-wide Mouse Two Plasmid Activity-Optimized CRISPR Knockout Library (Addgene 1000000096) as described above, containing 188,509 sgRNAs targeting 18,986 genes and 199 intergenic sgRNAs at a multiplicity of infection of 0.2–0.3 (measured by puromycin-resistance gene co-delivered with the lentiviral vector) and selected for lentiviral integration using puromycin (1 μg ml^{-1}) for 3 days. In screen I, the culture was expanded for another 4–8 days. On days 4, 6 and 8 of expansion, 200 × 10^6 cells were split into blasticidin S-selecting conditions (for 9, 7 and 5 days, respectively) and non-selection conditions (9 days). Cells in non-selection conditions were maintained at minimum level of 100 × 10^6 cells and logarithmic growth. After 9 days, 3 × 10^6–5 × 10^6 cells from selection conditions and 100 × 10^6 non-selection conditions were washed three times with PBS and pelleted by centrifugation for genomic DNA extraction using GeneElute Mammalian Genomic DNA Miniprep kit (Sigma) and Quick-DNA Midiprep Plus kit (Zymo Research), respectively, following the manufacturers guidelines. The sgRNA-encoding insertions were PCR-amplified using Agilent Herculase II Fusion DNA Polymerase (600675). These libraries were then sequenced using Illumina HiSeq 2500 (approximately 5 million–10 million reads with sgRNA sequence per condition; around 40× coverage per library element in non-selection conditions, screen I). As screen I demonstrated that longer blasticidin S treatment resulted in better intergenic sgRNA depletion, we performed screen II in two biological replicates with 9-day-long blasticidin S selection after either 8-day-long or 17-day-long cell culture (early and late selection, respectively). After genomic DNA extraction and library amplification, libraries were sequenced using Illumina HiSeq 2500 (SE65) (approximately 30 million–35 million reads per condition; around 170× coverage per library element in early and late non-selection conditions, screen II. See Supplementary Table 1 for the primer sequences used to amplify the libraries.

Immunofluorescence

Cells were plated on fibronectin-coated (Sigma) glass coverslips. For IAPEz-GFP reporter reactivation control, doxycycline (1 μg ml^{-1}) was added for 24 h. The next day, cells were fixed with 3% paraformaldehyde for 10 min at room temperature, rinsed three times with PBS, incubated 3 min in 0.3 μg ml^{-1} DAPI and rinsed again with PBS. For detection of IAP-GAG, after fixation, cells were permeabilized for 4 min with PBS/0.5× Triton X-100 on ice, blocked with 1% BSA/PBS for 15 min, incubated for 40 min with rabbit anti-mouse IAP-GAG antibody (gift from B. Cullen), 40 min with secondary antibodies and 3 min in 0.3 μg ml^{-1} DAPI at room temperature. Slides were mounted with VECTASHIELD medium (Vector Laboratories). Images were obtained with an Upright Spinning disk Confocal Microscope (Roper/Zeiss) and processed with Image J.

RT–qPCR analysis

Total RNA was extracted using Trizol (Life Technologies). Genomic DNA was removed by DNase I treatment (Qiagen), precipitated and resuspended in DNase/RNase-free water. Next, 10 μg of RNA was used for a second round of purification using RNeazy Mini columns (Qiagen) and 500 ng RNA was reverse-transcribed using random priming with Superscript III (Life Technologies). Quantitative PCR was performed using the SYBR Green Master Mix on the Viia7 thermal cycling system (Applied Biosystem). Relative expression levels were normalized to *Gapdh* or *Actb* (indicated in respective figures) using the $\Delta\Delta C_t$ method. For primer sequences, see Supplementary Table 1.

RNA stability assay

For RNA stability assay, 0.5 × 10^6 AID-METTL3;METTL14 dd ES cells treated with either 500 μM auxin or DMSO for 96 h or after 48-h auxin wash-off were re-plated on fibronectin-coated 6-cm plates 24 h before addition of actinomycin D. Next, medium were renewed and supplemented with 5 μg ml^{-1} actinomycin D (Sigma) to inhibit transcription. Total RNA was extracted at indicated time points and used for RT–qPCR. The half-life of IAPΔ1 was calculated according to the following equation: $\ln(C_i/C_0) = -kt_i$, in which k is the degradation rate, C_i is the mRNA value at time i, and t_i is the time interval in hours[33]. First, we calculated degradation rate k_i from each time point. The half-life $t_{1/2}$ is $\ln(2)/k_a$, in which k_a is the average degradation rate measured across the different time points.

m^{6}A ELISA

m^{6}A ELISA was performed using an m^{6}A RNA methylation colorimetric assay kit (Abcam, ab185912) according to manufacturer's protocol, and using 200 ng of RNA. After incubation in the last developer solution, the reaction was stopped by adding 100 μl of STOP solution at the moment positive control wells turned medium blue. It is crucial to continuously control the progressive colour change for the positive control; prolonged incubation will result in signal saturation in the experimental wells and potential masking of differences between tested conditions. Absorbance was measured at 450 nm within 5 min using absorbance microplate reader.

RNA immunoprecipitation

For each replicate, 1.5×10^7 cells were collected, washed with ice-cold PBS, resuspended in 1 ml of RIP lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.5% Igepal CA-630, containing 40 U ml^{-1} of RNasin and protease inhibitors (Roche)). Lysates were placed on ice for 20 min and centrifuged at 4 °C to remove cell debris. Supernatants containing RNA–protein complexes were collected and 50 µl was kept for input control. Remaining lysates were precleared with M-280 sheep anti-rabbit IgG magnetic beads (Thermo Fisher Scientific) and incubated with either 5 µg of YTHDF2 antibody or rabbit IgG rotating overnight at 4 °C. Next, 25 µl of beads was added and samples were rotated for additional 2 h at 4 °C. Beads were washed with RIP lysis buffer five times for 10 min and RNA was isolated from the beads as well as input samples using Trizol. Genomic DNA was removed by DNase I treatment (Qiagen), precipitated and resuspended in DNase/RNase-free water. Next, purified RNA underwent a second round of purification using RNeazy Mini columns (Qiagen). Immunoprecipitated and input RNAs were reverse transcribed using random priming with Superscript III (Life Technologies). Reverse transcription and real-time quantitative PCR were used to measure abundance.

Poly-A RNA sequencing

Total RNA was extracted using Trizol (Life Technologies). Genomic DNA was removed by in solution DNaseI treatment (Qiagen), RNA was precipitated and resuspended in DNase/RNase-free water. Next, 10 µg of RNA was used for a second round of purification using RNeazy Mini columns (Qiagen). RNA integrity was evaluated on TapeStation 4200 (Agilent) using RNA ScreenTape (5067-5576), requiring a minimal integrity number (RIN) of 9. Libraries were prepared according to Illumina's instructions accompanying the TruSeq Stranded mRNA Library Prep Kit. Approximately 800 ng of RNA per replicate was used for library preparation. After library preparation, the length profiles were assessed with the LabChip GX Touch HT system (Perkin Elmer) and equimolar pool from all samples was prepared. Molarity of the pool was quantified by qPCR using KAPA Library Quantification Kit and the CFX96 qPCR system (Biorad) before sequencing. Samples were sequenced using Novaseq 6000 (PE100, approximately 90 million clusters per replicate).

MeRIP-seq

MeRIP-seq was carried out using Magna MeRIP m^6A Kit (Millipore) according to the manufacturer's instructions. In short, total RNA was extracted using Trizol (Life Technologies). DNaseI-treated RNA samples were chemically fragmented into 100-nucleotide-long fragments and 350 µg of total RNA were subjected to each immunoprecipitation with affinity purified anti-m^6A antibody in presence of RNase inhibitor. Bound m^6A-methylated RNA fragments were eluted with free N6-methyladenosine, purified using RNeazy Kit (Qiagen) and processed for library generation using SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (TaKaRa) following the manufacturer's recommendations, but without fragmentation step (9 ng of RNA per replicate). Sequencing was performed using Illumina Novaseq 6000 (PE100, approximately 50 million to 90 million clusters per replicate). The m^6A IP for wild-type and *Mettl3*-knockout cells was performed independently three times. Input for each cell line was sequenced as a control.

CUT&RUN

For chromatin profiling, we performed CUT&RUN[34] on AID-METTL3;METTL14 double degron cells treated with either DMSO (control) or auxin for 96 h in two biological replicates. In brief, 2×10^5 cells were washed with PBS, three times with wash buffer (20 mM HEPES-KOH pH 7.9, 150 mM NaCl, 0.5 mM spermidine and Protease Inhibitors (Roche)) at room temperature, then resuspended in 1 ml of washing buffer. Next, 10 µl of concanavalin A-coated magnetic beads (Bangs Laboratories, BP531), pre-washed and resuspended in binding buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 1mM $CaCl_2$, 1 mM $MnCl_2$), were added to the cells. After 10 min incubation at room temperature under rotation, bead-bound cells were isolated on a magnetic stand and resuspended in 400 µl of antibody buffer (wash buffer supplemented with 0.02% digitonin (Millipore, 300410) and 2 mM EDTA) containing 2 µg of anti-H3K4me3, -H3K27ac or -H3K9me3 antibody or 2 µg of rabbit IgG (Supplementary Table 3). After 15 min incubation at room temperature under rotation, cells were washed three times in 1 ml of digitonin(dig)-wash buffer (wash buffer supplemented 0.02% digitonin), incubated with pA-MNase (400 µl of dig-wash buffer containing 700 ng ml^{-1} pA-MNase, produced by the Protein Core Facility of Institut Curie) and washed again three times in 1 ml of dig-wash buffer. Cells were then resuspended in 150 µl of dig-wash buffer, transferred to 1.5 ml tubes, and equilibrated to 0 °C in a metal block for 10 min on ice. To initiate pA-MNase-dependent digestion, $CaCl_2$ was added to a final 2 mM concentration, incubation was carried out at 0 °C for 30 min, and stopped by adding 150 µl of 2 × STOP solution (200 mM NaCl, 20 mM EDTA, 5 mM EGTA, 0.1% Igepal CA-630, 40 µg ml^{-1} glycogen). RNase A was added to a final concentration of 50 µg ml^{-1} and samples were incubated at 37 °C for 20 min. Samples were placed on a magnetic stand, supernatant was transferred to low-binding tubes and SDS and proteinase K were then added to final concentrations of 0.1% and 300 µg ml^{-1}, respectively, and samples were incubated at 70 °C for 30 min. Total DNA was extracted by phenol/chloroform followed by two rounds of ethanol precipitation, eluted in 40 µl of 1 mM Tris-HCl pH 8.0 and 0.1 mM EDTA, and quantified and analysed using Qubit and TapeStation assays. CUT&RUN libraries were prepared using the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences) according to the manufacturer's protocol. Samples were sequenced using NovaSeq 6000 (PE50, approximately 60 million clusters per replicate).

Genome-wide CRISPR–Cas9 screen analysis

The sequenced reads were mapped to the sgRNA library. Only reads that contained one sgRNA sequence without mismatch were counted. The MAGeCK[11] test command line (version 0.5.8) was used to rank sgRNAs and genes with following parameters: –norm-method total–adjust-method fdr–remove-zero-threshold 10–gene-lfc-method alphamean–remove-zero both. For screen II, the sequencing primer was oriented in the opposite direction to the sgRNA, therefore the CRISPR mouse pooled library was reverse-complemented before counting.

RNA-seq analysis

Adapters were trimmed using Atropos v.1.1.16[35]. Paired-end read alignment was performed onto the Mouse reference genome (mm10) with STAR v.2.7.0a[36] reporting randomly one position, allowing 6% of mismatches (–outFilterMultimapNmax 5000–outSAMmultNmax 1–outFilterMismatchNmax 999–outFilterMismatchNoverLmax 0.06). Repeat annotation was downloaded from RepeatMasker (http://www.repeatmasker.org/). To reconstruct full-length LTR copies, we used the same strategy as done previously[10] using the perl tool 'one code to find them all'[37]. Reconstructed transposons annotation and basic genes annotation from GENCODE v.18 were merged and used as input for quantification with FeatureCounts v1.5.1[38]. Differential expression analysis was performed using edgeR's normalization combined with voom transformation from limma R package[39,40]. *P* values were computed using limma and adjusted with the Benjamini–Hochberg correction. Genes and transposon families were declared as differentially expressed if FDR < 5% and $\log_2FC > 0.75$. Upregulated genes in all four knockout lines were annotated with proximal retrotransposon elements (overlap with promoter regions defined as −5 kb to +1 kb from the TSS). Randomized gene sets were created 100 times and were annotated to proximal retrotransposon elements to compute permutation test using regioneR[41] R package.

Transposon element-based analysis was performed using only uniquely mapped reads by retrieving reads with NH tag equal to 1. To avoid confounding effects between expression from retrotransposon copies and from genes, elements overlapping genes in the same direction were removed from RepeatMasker annotations. FeatureCounts v.1.5.1 was used for the quantification and only copies with at least 10 reads in at least one sample were conserved. RRACH motifs were searched into the 5′UTR sequences of individual IAP copies using RSAT dna pattern. Division into three categories of RRACH motif number (7–8, 9–10 and 11–18) was performed to homogenize the number of IAP copies per category.

To estimate intron retention between genes and single IAP copies, reads alignment was performed using specific parameters to report only uniquely mapped reads with STAR v.2.7.0a[36] (–outFilterMultimapNmax 1–outSAMmultNmax 1). Unannotated splice junctions detected by STAR was annotated with GENCODE v.18 and IAP LTR elements from RepeatMasker annotation to retrieve splicing events between a gene and an IAP element. The number of uniquely mapped reads crossing the splicing events was calculated for each sample and normalized by the library size.

MeRIP-seq analysis

Due to the addition of 3 nucleotides on 5′-end of the second sequencing read (R2) from the Pico v.2 SMART Adaptor, paired-end reads were trimmed using Trim Galore v.0.4.4 with the options: –three_prime_clip_R1 3–clip_R2 3 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were aligned onto the mouse ribosomal sequence (GenBank: BK000964.3) using Bowtie v.1.2 allowing at most three mismatches[42]. Previously unmapped reads were aligned onto the mouse reference genome (mm10) using STAR v.2.6.0c reporting randomly one position, allowing 4% of mismatches (–outFilterMultimapNmax 5000–outSAMmultNmax 1–outFilterMismatchNmax 999–outFilterMismatchNoverLmax 0.04). PCR duplicates were removed using STAR with the option –bamRemoveDuplicatesType UniqueIdenticalNotMulti. Bigwig files were produced with deepTools v.2.5.3[43] using the option –normalizeUsingRPKM. Peaks enriched in the MeRIP sample over the input control were defined using MACS2 peak-caller[44] with a genome size of 994,080,837 bp[20] and the FDR threshold of 5%. Reads were extended to 200-bp-long fragments. Only peaks called in at least two replicates were used for downstream analysis. Peaks intensity was calculated using FeatureCounts v.1.5.1[38] and normalized to background (reads not falling into peaks) and to peak length. GENCODE v.18 was used to define 5′UTR, 3′UTR, intronic and exonic regions. Retrotransposon annotations (RepeatMasker) were downloaded from UCSC table browser. Genes overlapping with at least one peak were used to calculate coverage along the genic region (5′UTR, coding sequence and 3′UTR) with trumpet R package[45]. Mapped reads onto the Mouse reference genome overlapping with IAP and L1MdA_I elements were extracted as single-end reads and mapped to the full-length IAP (GenBank: M17551.1) and L1MdA_I consensus sequences with Bowtie2 v.2.2.9[46] with these parameters: –local -N 1. Coverage along the consensus sequence was normalized to background (reads not falling into peaks) as was done previously for peak intensities. Rolling mean was calculated for a window of 50 bp to smooth the signal. RRACH motif was searched into the IAP consensus sequence using RSAT dna pattern. Intragenic peaks in the wild-type condition were used to define m^{6}A-bound genes. In addition, a threshold of 1 RPKM in the input wild-type samples was put to be sure that the gene is expressed. Genes with RPKM higher than 1 and not overlapping with a m^{6}A peak were defined as non-m^{6}A bound genes.

CUT&RUN analysis

Paired-end reads were trimmed using Trim Galore v0.4.4. The alignment was performed onto a concatenated genome using the mouse reference genome (mm10) and the *Escherichia coli* genome (str. K-12 substr. MG1655, GenBank: NC_000913) with STAR v.2.7.0a[36] reporting randomly one position, allowing 6% of mismatches (–outFilterMultimapNmax 5000–outSAMmultNmax 1–outFilterMismatchNmax 999–outFilterMismatchNoverLmax 0.06). PCR duplicates were removed using Picard v.2.6.0 (http://broadinstitute.github.io/picard/). Reconstructed transposon annotation was used as input for quantification with FeatureCounts v.1.5.1[38] using reads mapped onto the mouse genome. Normalization of counts per million was performed using as library size the number of mapped reads onto the mouse genome.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The raw imaging data that support the findings of this study are available from the corresponding authors on request owing to size considerations. All sequencing data derived from CRISPR–Cas9 screens I and II, RNA-seq, MeRIP-seq and CUT&RUN have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE145616. Previously published data were downloaded from GEO: *Mettl3* control and *Mettl3* knockout nuclear RNA-seq: (GSE133585)[19]. Source data are provided with this paper.

31. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148 (2016).
32. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5**, 1–30 (2016).
33. Chen, C. Y. A., Ezzeddine, N. & Shyu, A. B. Messenger RNA half-life measurements in mammalian cells. *Methods Enzymol.* **448**, 335–357 (2008).
34. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, 1–35 (2017).
35. Didion, J. P., Martin, M. & Collins, F. S. Atropos: specific, sensitive, and speedy trimming of sequencing reads. *PeerJ* **5**, e3720 (2017).
36. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
37. Bailly-Bechet, M., Haudry, A. & Lerat, E. 'One code to find them all': a perl tool to conveniently parse RepeatMasker output files. *Mob. DNA* **5**, 1–15 (2014).
38. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
39. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
40. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
41. Gel, B. et al. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).
42. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
43. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165 (2016).
44. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
45. Zhang, T., Zhang, S. W., Zhang, L. & Meng, J. trumpet: transcriptome-guided quality assessment of m^{6}A-seq data. *BMC Bioinformatics* **19**, 260 (2018).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

Acknowledgements We thank the members of the Bourc'his laboratory for their support; M. Greenberg and A. Shkumatava for critical reading of the manuscript; G. Cristofari and J. J. Zylicz for suggestions; I. Pinheiro for help with FACS analysis; E. Nora for targeting plasmids; and M. Schultz and M. Borensztein for assistance with cell culture and immunofluorescence experiments. We acknowledge the ICGex NGS platform of the Institut Curie -supported by grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique) from the Agence Nationale de la Recherche- and the Cell and Tissue Imaging Platform (PICT-IBiSA) of Institut Curie -member of the French National Research Infrastructure France-BioImaging (ANR-10-INBS-04). The laboratory of D.B. is part of the LABEX DEEP (ANR-11-LABX-0044, ANR-10-IDEX-0001-02). This work was supported by the Fondation Bettencourt Schueller, the Association Robert Debré pour la Recherche Médicale (ARDRM), the Fondation pour la Recherche Médicale (FRM) and the Association de Recherche contre le Cancer (ARC-PJA-20191209637). T.C. was a recipient of an EMBO postdoctoral fellowship; E.R. and M.D were supported by PhD fellowships from la Ligue contre le Cancer and from Région Ile-de-France, respectively.

Author contributions D.B. and T.C. conceived and designed the study. T.C. performed the genetic screen, genetic engineering of the different ES cell lines (reporter, genetic knockouts and auxin-degron), RNA-seq, MeRIP-seq, degron experiments, actinomycin D assays, m^{6}A quantification by ELISA and immunoblots. E.R. contributed to generating and characterizing ES cell knockouts of *Mettl3*, *Mettl14*, *Wtap*, *Zc3h13*, *Ythdf1*, *Ythdf2* and *Ythdf3* and aided degron

experiments. T.C., M.D. and L.B. performed CUT&RUN experiments. S.R. and C.F. aided the genetic screen, and M.D. the MeRIP-seq. F.D. provided degron-targeting vectors and helped with degron design. A.T. performed the bioinformatic analyses. S.L. aided sequencing library preparations. D.B. and T.C. interpreted the data and wrote the manuscript. All authors read and approved the final manuscript.

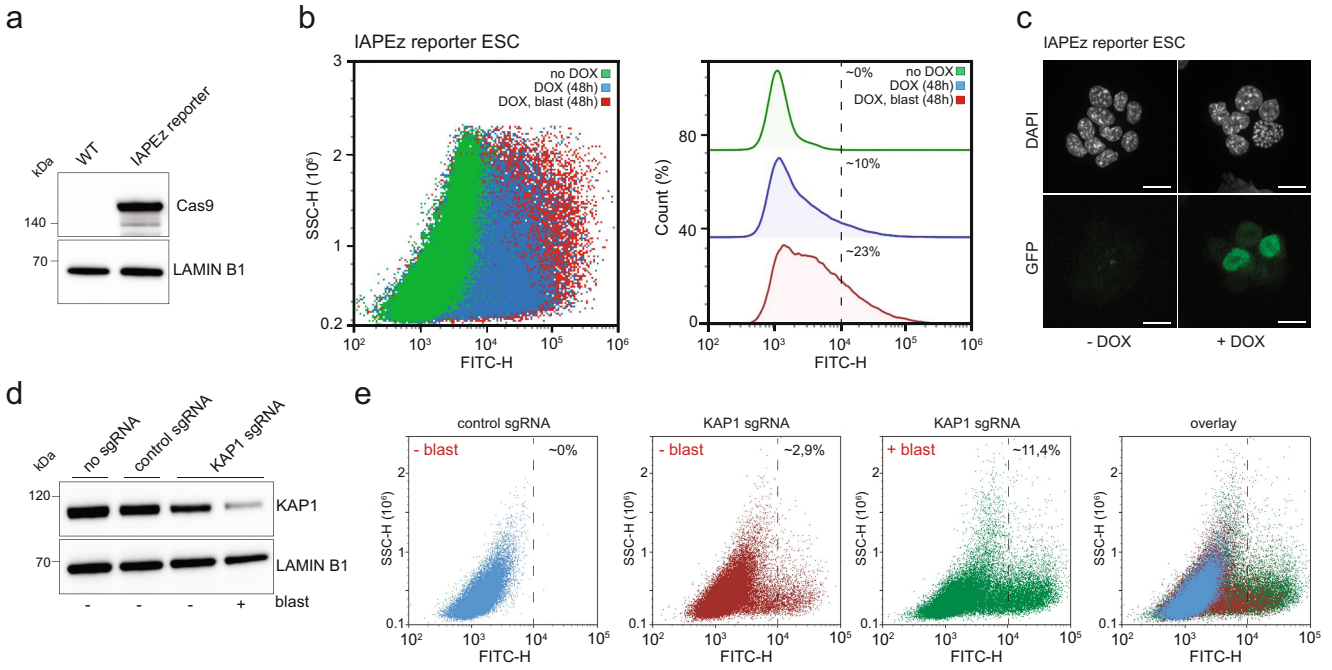
Competing interests The authors declare no competing interests.

Additional information

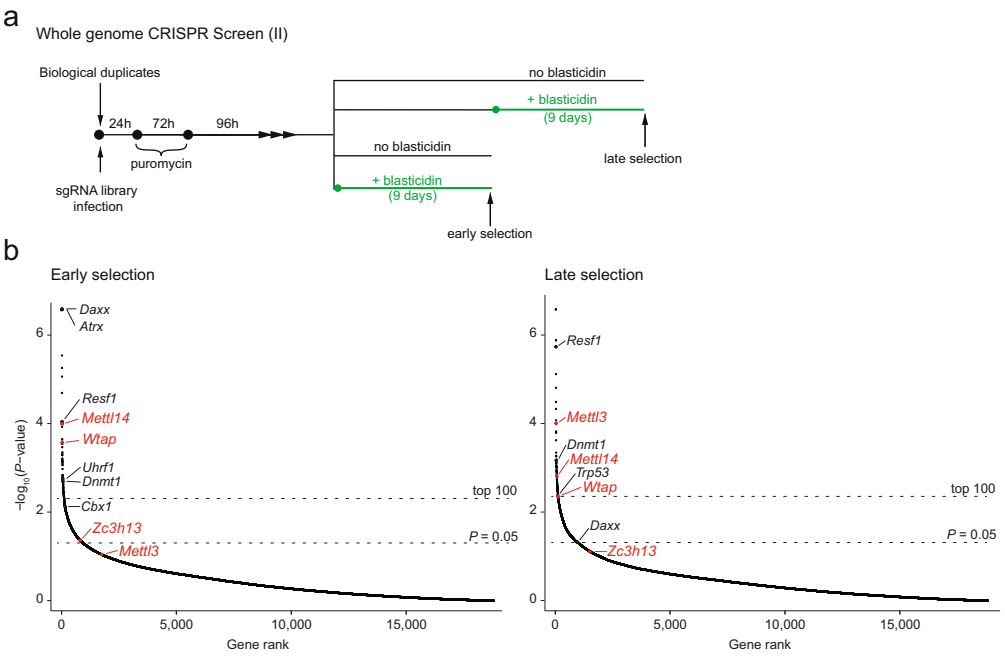
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-020-03135-1.

Correspondence and requests for materials should be addressed to T.C. or D.B.

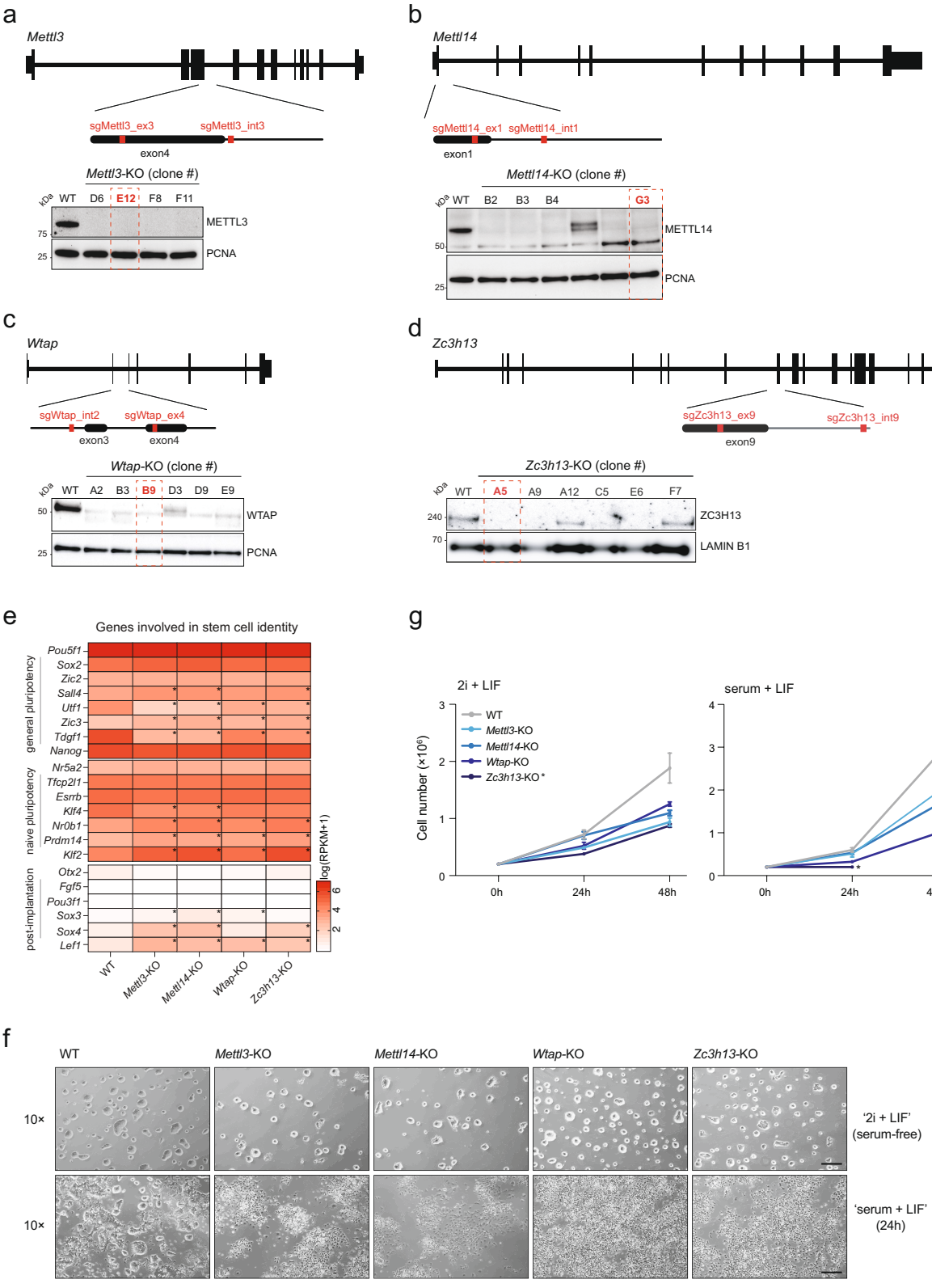
Peer review information *Nature* thanks Miguel Branco and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.



Extended Data Fig. 1 | Validation of the IAPEz-reporter ES cell line. **a**, Immunoblot showing Cas9 protein levels in parental (E14) and doxycycline (dox)-inducible IAPEz-reporter ES cells. Lamin B1 served as loading control. **b**, Flow cytometry analysis of GFP expression in IAPEz-reporter ES cells before and upon 48 h of dox (1 μM) induction and blasticidin (blast, 5 μg ml^{-1}) selection. Left, overlaid dot blots (last 130,000 events) and histograms normalized to 100% scale (right). Analysis was performed using NovoExpress software (Acea Biosciences). For pseudocolour plots and gating strategy, see Supplementary Fig. 2. Note that blasticidin selection further increases the fraction of GFP-positive cells, and not all blasticidin-resistant cells accumulate detectable GFP levels. Therefore, blasticidin selection is more sensitive than the GFP signal to subtle changes in the expression of IAPEz reporter. **c**, Immunofluorescence detection of GFP in dox-induced IAPEz-reporter ES cells. Scale bar, 10 μm. **d**, Immunoblot showing KAP1 protein levels in IAPEz-reporter ES cells after KAP1-specific sgRNA introduction and blasticidin selection. Scrambled sgRNA served as control. Blasticidin treatment shows that only cells with successful KAP1 depletion become antibiotic-resistant. Lamin B1 served as loading control. **e**, Flow cytometry dot blot analysis (last 50,000 events) of GFP expression in cells from **d**. Discontinued lines denote GFP-positive threshold. For pseudocolour plots and gating strategy, see Supplementary Fig. 2. KAP1 depletion combined with blasticidin selection leads to IAPEz-reporter reactivation and GFP expression. Experiments in **a**–**e** were replicated at least twice with similar results.

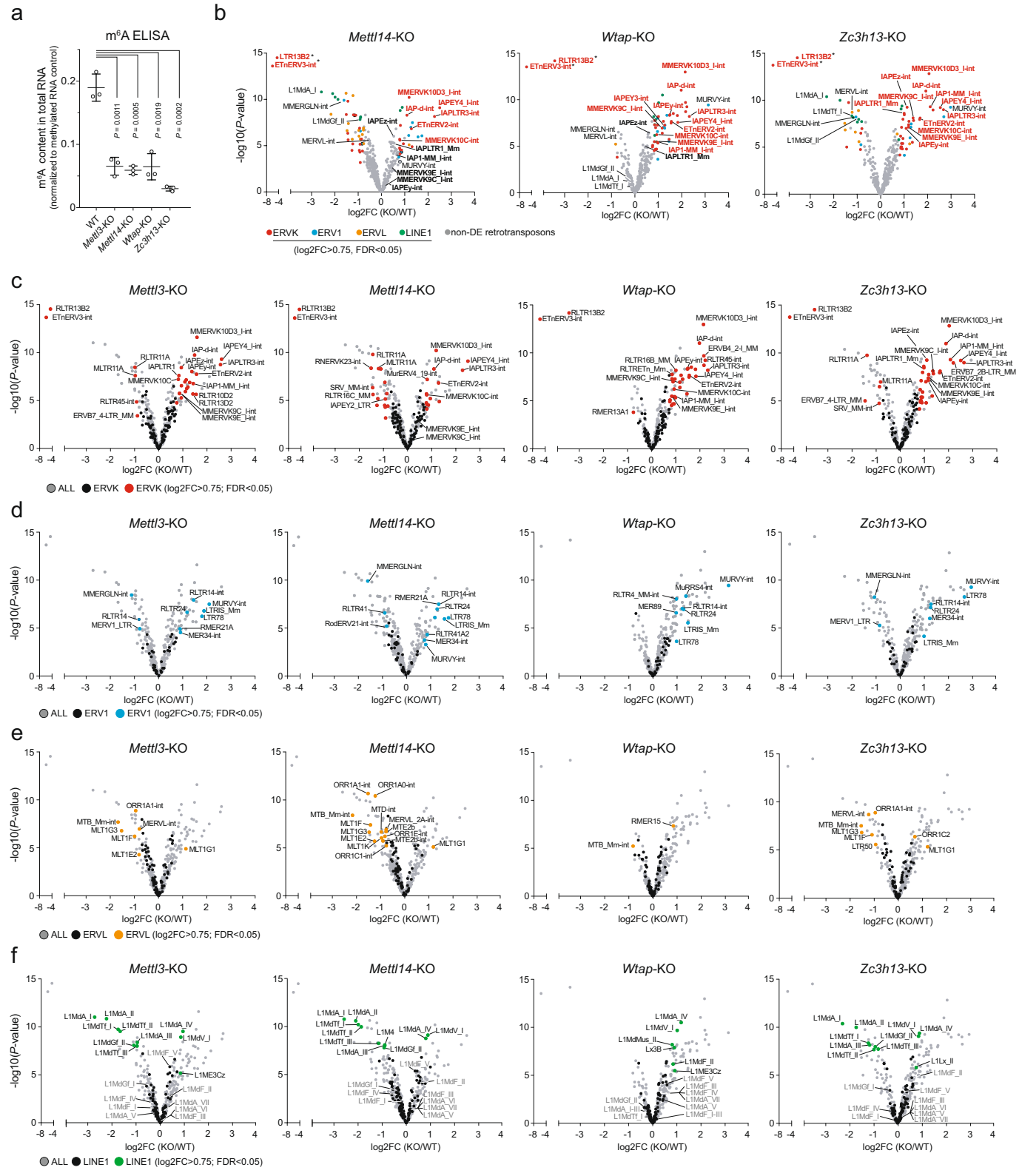


Extended Data Fig. 2 | Early versus late genome-wide CRISPR–Cas9 screen for IAPEz suppressors. a, Schematic of the screening process. **b**, Ranked *P* values (permutation test) for enriched genes in early (left) and late (right) blasticidin treatment, as in **a**. Discontinued lines indicate genes ranked in top 100 according to *P* value (top) and $P = 0.05$ (bottom). Known IAP regulators (black, when $P < 0.05$) and *Mettl3*, *Mettl14*, *Wtap* and *Zc3h13* (red) are reported. Presented data were derived from biological duplicates.

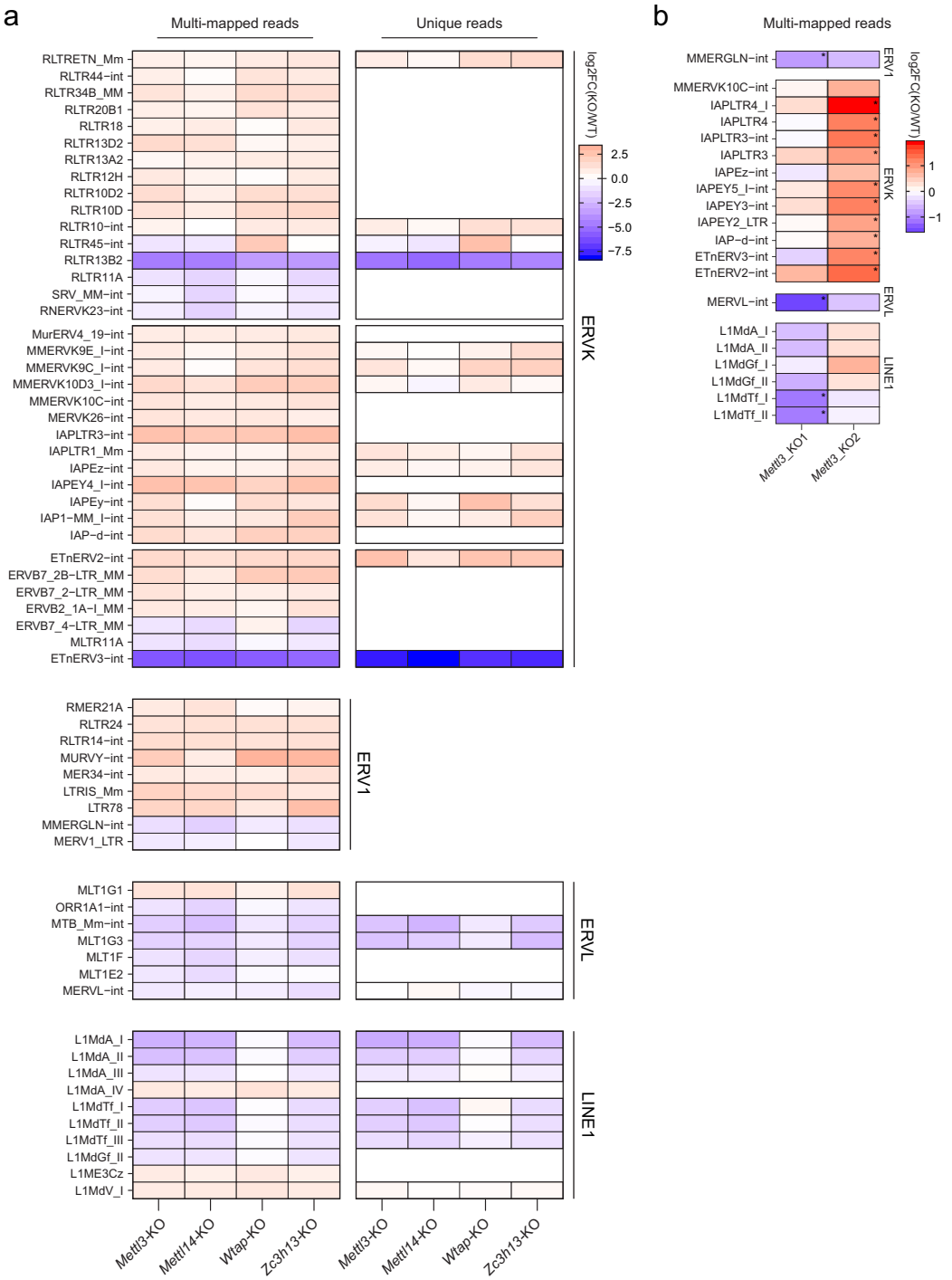


Extended Data Fig. 3 | See next page for caption.

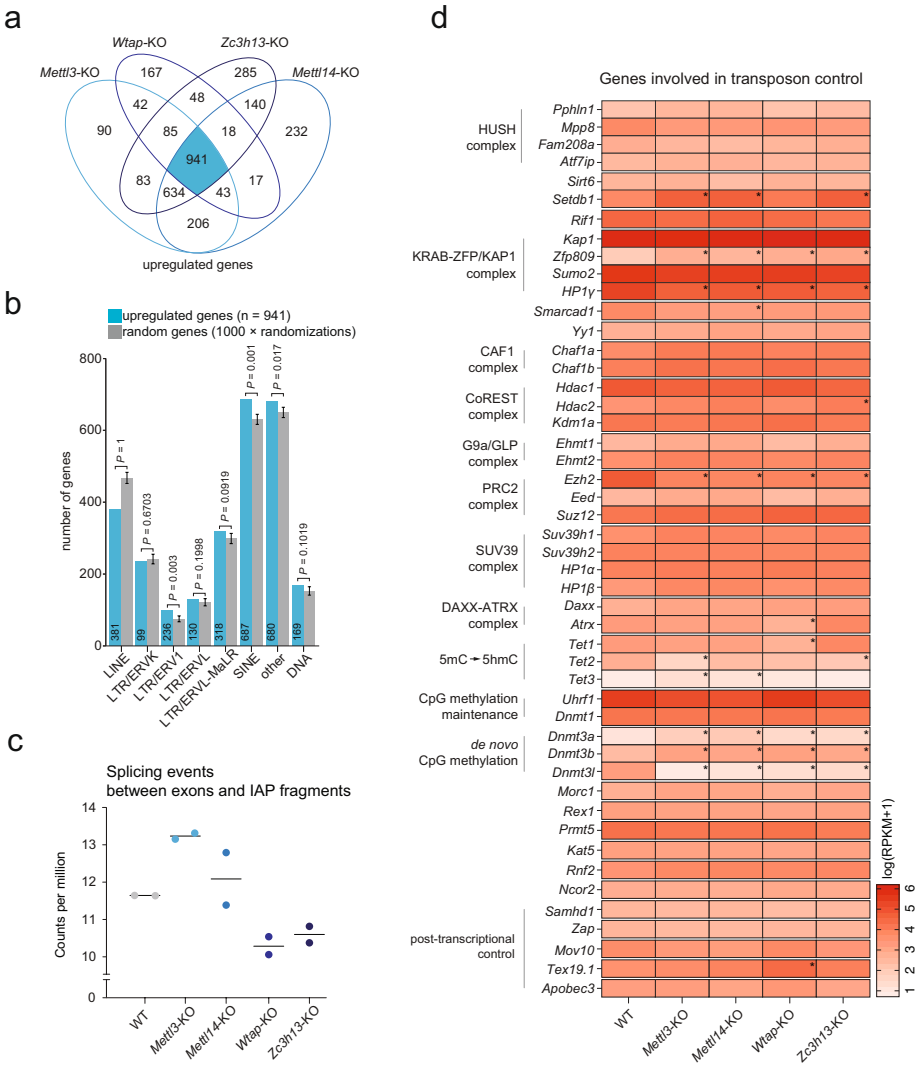
Extended Data Fig. 3 | Generation and validation of *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout mouse ES cells. a–**d**, *Mettl3*-knockout (**a**), *Mettl14*-knockout (**b**), *Wtap*-knockout (**c**), and *Zc3h13*-knockout (**d**), mediated by CRISPR–Cas9. Schematic representation showing sgRNAs targeting of indicated loci (top) and immunoblots (bottom) confirming depletion of indicated proteins. PCNA or lamin B1 served as loading controls. In red, clones selected for downstream analyses presented in Figs. 2, 3, Extended Data Figs. 3e–g, 4–7. Western blots for clones indicated in red were repeated at least twice with similar results. **e**, RNA-seq heat maps showing expression of selected pluripotency and post-implantation markers in wild-type, *Mettl3*-knockout, *Mettl14*-knockout, *Wtap*-knockout and *Zc3h13*-knockout ES cells. Asterisks indicate genes with $\log_2$FC > 0.75 and FDR < 0.05. Data represent averages of two independent RNA-seq. **f**, Representative images of wild-type, *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout ES cells grown in serum-free 2i + LIF medium, or after 24 h conversion to serum + LIF medium. Scale bar, 0.5 mm. When replacing 2i + LIF with serum + LIF medium, marked morphological changes were observed in all knockout lines. Experiments were repeated twice with similar results. **g**, Growth curves of wild-type, *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout ES cells cultured in serum-free 2i + LIF medium, or after conversion to serum + LIF. Data are mean ± s.d. from three independent experiments. When replacing 2i + LIF with serum + LIF medium, the self-renewal ability of *Wtap*-knockout cells was severely impaired and *Zc3h13*-knockout cells disappeared (highlighted by an asterisk). This could explain the lower sgRNA ranking for these two genes upon extended selection time in the screen, which was performed in serum + LIF medium (Extended Data Fig. 2b).



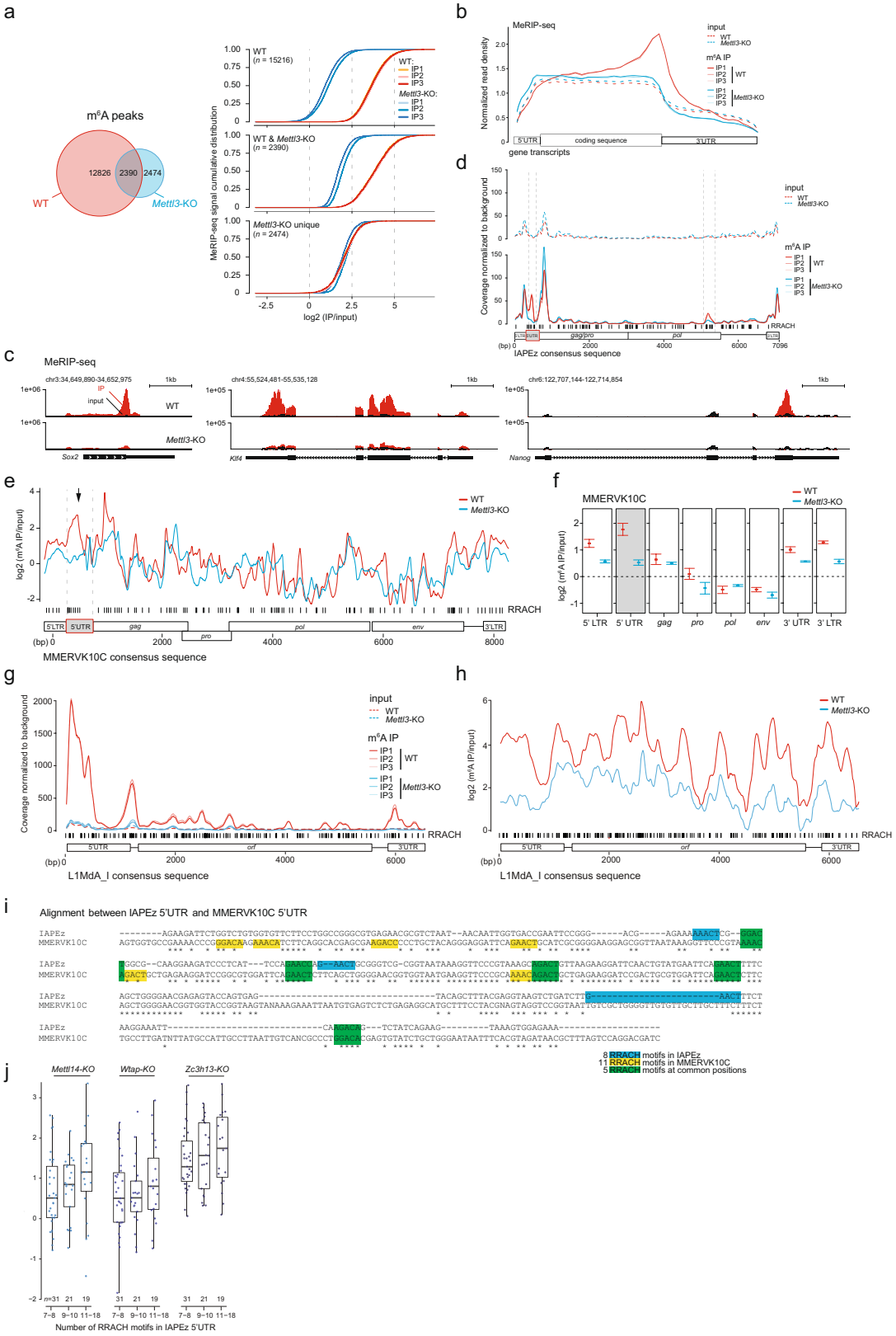
Extended Data Fig. 4 | Depletion of m^{6}A methyltransferase complex results in deregulation of different retrotransposon families. a, ELISA showing normalized m^{6}A levels in total RNA in WT and KO ES cells. Data are mean ± s.d. of three independent RNA samples. *P* values were determined by two-sided Student's *t*-test. **b**, Volcano plot showing $\log_2$FC in retrotransposon expression in *Mettl14*, *Wtap* and *Zc3h13*-knockout versus wild-type cells using a random assignment of multi-mapped reads. In red, blue, orange and green are significantly deregulated RepeatMasker annotations belonging to ERVK, ERV1, ERVL and L1 families, respectively. In grey, non-differentially expressed (non-DE) retrotransposons. *P* values were computed using limma and adjusted with the Benjamini–Hochberg correction (Methods). **c**–**f**, Volcano plots showing $\log_2$FC of retrotransposon expression in *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout versus wild-type ES cells using a random assignment of multi-mapped reads. In red, blue, orange and green are significantly deregulated RepeatMasker annotations belonging to ERVK (**c**), ERV1 (**d**), ERVL (**e**) and L1 (**f**) families, respectively ($\log_2$FC > 0.75 and FDR < 0.05). In grey, non-differentially expressed retrotransposons. The *P* values were computed using limma and adjusted with the Benjamini–Hochberg correction.



Extended Data Fig. 5 | Family-based analysis of retrotransposon expression using random versus unique mapping. a, RNA-seq heat maps showing $\log_2$FC in expression of indicated retrotransposon families in *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout versus wild-type ES cells allowing either multiple mapping with random allocation (left) or unique mapping only (right). In case of unique mappers with intragenic position, only elements transcribed in reverse orientation to the gene were included in the analysis. Note that although ERVK annotations show general upregulation in all knockout ES cells, L1s tended to be downregulated, with the exception of oldest L1MdA_IV, L1ME3Cz and L1MdV_I subfamilies. **b**, Heat maps of published nuclear RNA-seq data[19] showing $\log_2$FC in expression for indicated retrotransposon families in two *Mettl3*-knockout versus control ES cells, allowing multiple mapping with random allocation. The two knockout lines show divergent levels of retrotransposon reactivation but consistently show greater reactivation of ERVK compared to ERVL and LINE1 families. Asterisks indicate retrotransposon families with $\log_2$FC > 0.75 and FDR < 0.05. Data were taken from accession GSE133585.

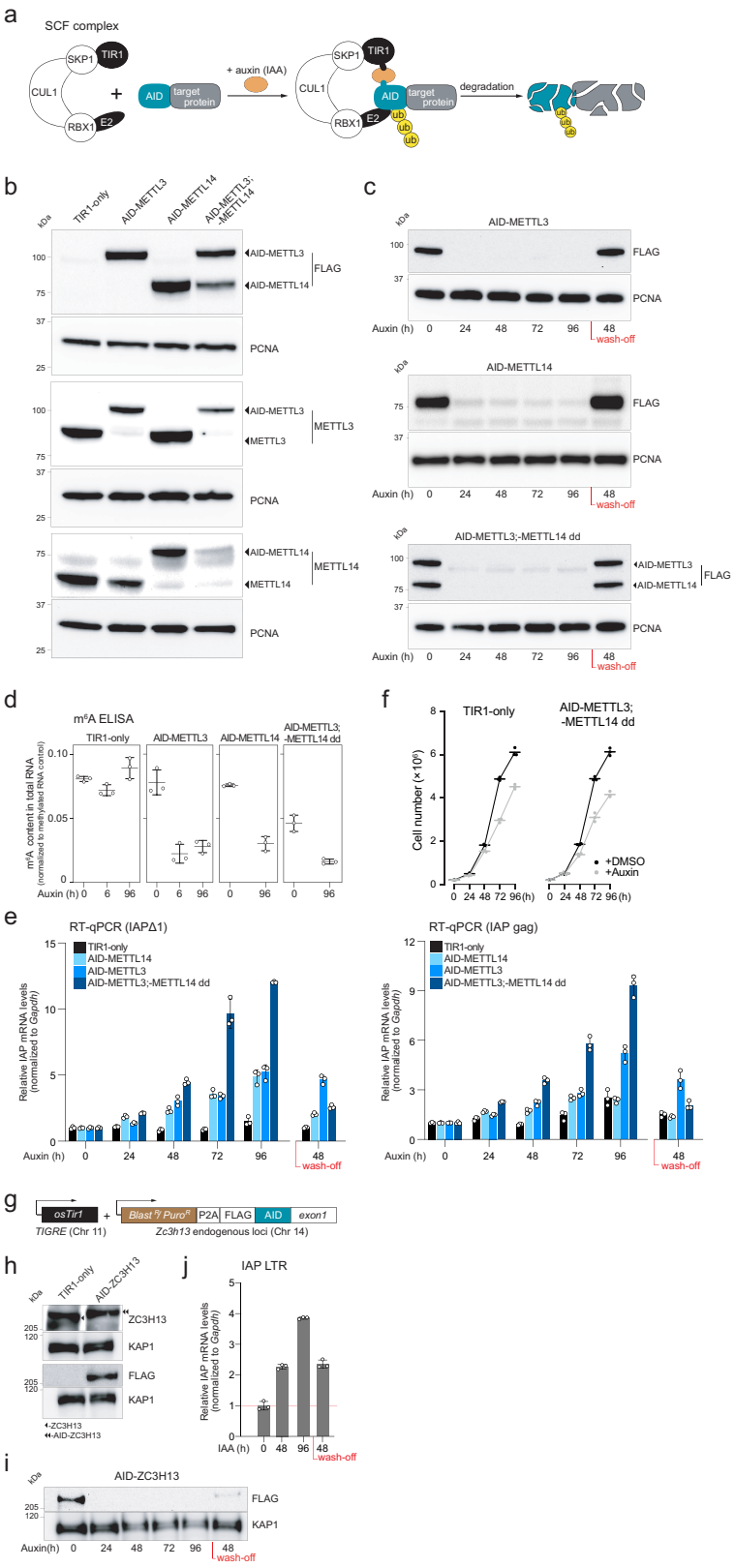


Extended Data Fig. 6 | Analysis of gene expression in mutant ES cells of the m^{6}A methyltransferase complex. a, Venn diagram showing overlap of the upregulated genes (FDR < 0.05, $\log_2$FC > 0.75) in *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout ES cells as identified by RNA-seq. In blue, genes (n = 941) commonly upregulated in all four knockout ES cells. **b**, Correlation between gene upregulation and proximity to retrotransposon annotations (−5 kb to +1 kb from the TSS). In blue, genes (n = 941) commonly upregulated in all four knockout ES cell lines. In grey, random genes plotted as mean + s.d. of n = 1,000 randomizations. *P* values were determined by permutation test. **c**, Dot plot showing splicing events (normalized counts per million) occurring between exons and RepeatMasker-annotated IAPs in wild-type, *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout ES cells. Horizontal lines denote mean (n = 2 independent RNA-seq). **d**, Heat map showing average expression of selected retrotransposon regulators in wild-type, *Mettl3*-, *Mettl14*-, *Wtap*- and *Zc3h13*-knockout cells. Genes with $\log_2$FC > 0.75 and FDR < 0.05 are marked by an asterisk.



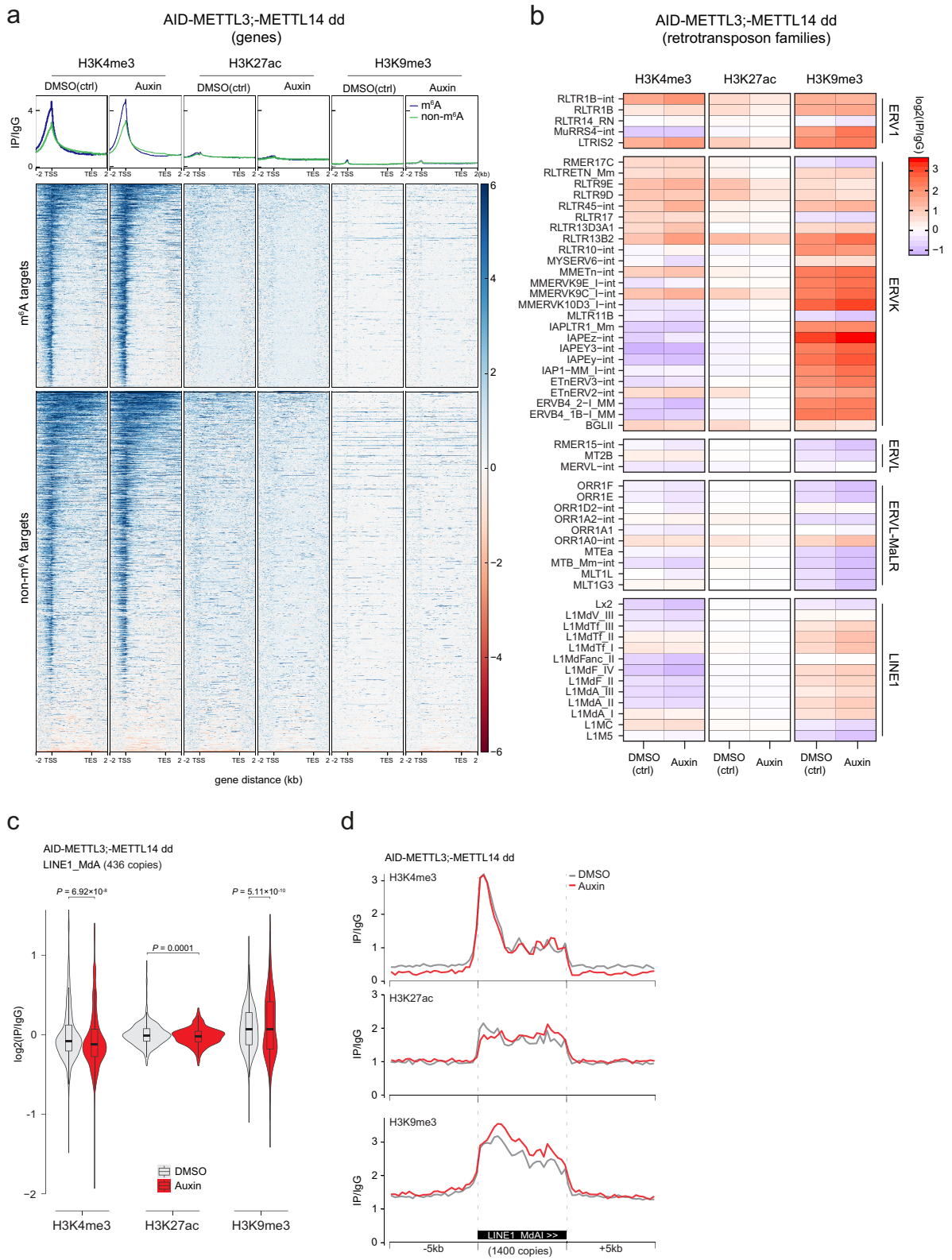
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | IAP mRNAs undergo METTL3-dependent m^{6}A methylation. a, Left, Venn diagram showing overlap between m^{6}A peaks identified in wild-type and *Mettl3*-knockout MeRIP-seq. Right, input-normalized cumulative distribution of signal intensity for m^{6}A peaks in wild-type (top), overlapping (middle) and *Mettl3*-knockout specific (bottom) cells. A set of novel m^{6}A peaks was detected in *Mettl3*-knockout cells; however, they are probably false positives (weak signal intensity compared to canonical wild-type peaks). **b**, Normalized MeRIP-seq read density in wild-type (red) and *Mettl3*-knockout (blue) cells across the 5′ UTR, coding sequence and 3′ UTR of mRNA for the genes with at least one m^{6}A peak. Discontinued lines represent respective inputs. **c**, UCSC genome browser tracks showing m^{6}A distribution (red, normalized read density, RPM) at indicated genes in wild-type and *Mettl3*-knockout ES cells. In black, input RNA. Results represent average signal of three independent MeRIP-seq experiments. **d**, Inputs (top) and background-normalized m^{6}A signal distributions (bottom) across the IAPEz consensus sequence in wild-type (red) and *Mettl3*-knockout (blue) cells. Vertical black lines denote positions of RRACH motifs. Discontinued vertical black lines denote regions of m^{6}A enrichment present in wild-type and lost in *Mettl3*-knockout cells. **e**, Average of input-normalized m^{6}A signal intensities along the MMERVK10C consensus sequence in wild-type (red) and *Mettl3*-knockout (blue) cells. Vertical black lines denote RRACH motif positions. Discontinued vertical black lines denote region of m^{6}A enrichment present in wild-type and lost in *Mettl3*-knockout cells. **f**, Average of m^{6}A signal intensities for indicated MMERVK10C sequence segments in wild-type (red) and *Mettl3*-knockout (red) ES cells. Data are mean ± s.d. of three independent MeRIP-seq experiments. **g**, **h**, Background-normalized m^{6}A signal distribution (**g**) and average of input-normalized m^{6}A signal intensities across the L1MdA_I consensus sequence in wild-type (red) and *Mettl3*-knockout (blue) cells (**h**). Vertical black lines denote RRACH motif positions. **i**, Alignment between 5′ UTRs of IAPEz and MMERVK10C consensus sequences. IAP-specific, MMERVK10C-specific and common RRACH motifs are indicated in blue, yellow and green, respectively. **j**, Box plot showing $\log_2$FC in expression of uniquely mapped IAPEz copies grouped according to number of 5′ UTR-associated RRACH motifs in *Mettl14*-, *Wtap*- and *Zc3h13*-knockout versus wild-type cells. Box plots are as in Fig. 2d. Only copies with a minimum of 10 reads in at least one sample were conserved.



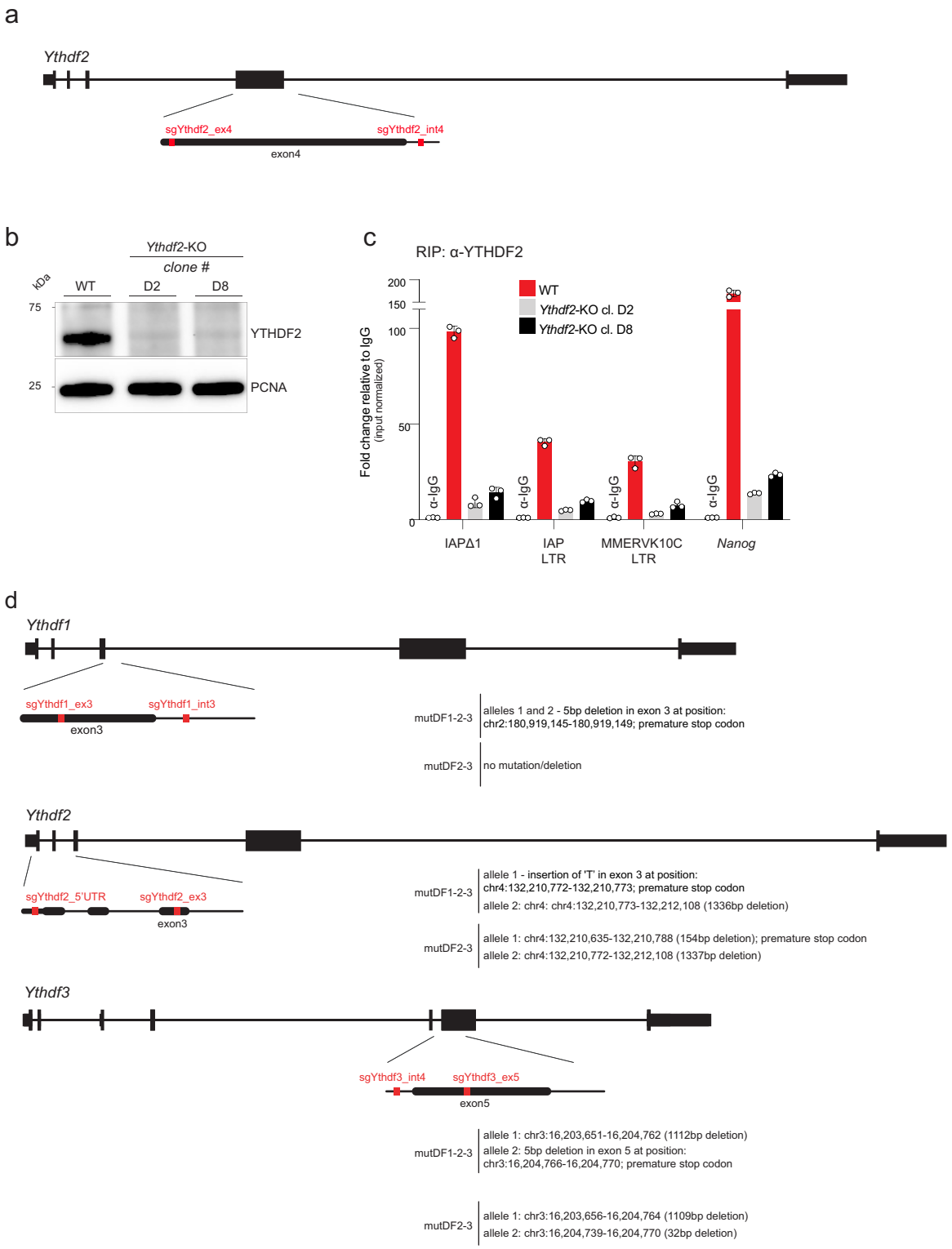
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Auxin-inducible degron of endogenous METTL3, METTL14 and ZC3H13. a, Schematic of TIR1 and SCF1 complex-dependent degradation of endogenously AID-tagged proteins in presence of auxin. **b**, Immunoblot showing protein levels of endogenously 3×Flag-AID-tagged METTL3 and METTL14 in single and double degron ES cell lines. TIR1-only ES cells were used as control for protein levels. PCNA served as loading control. **c**, Immunoblot showing efficiency and reversibility of METTL3 and METTL14 depletion after 0–96 h auxin treatment followed by 48 h auxin wash-off. PCNA served as loading control (related to Fig. 4b). **d**, ELISA showing normalized m^6A levels in total RNA after 0, 6 and 96 h of auxin-induced degradation of METTL3, METTL14 and METTL3;METTL14 double degron. Data are mean ± s.d. of three technical replicates. TIR1-only ES cells served as control. Experiment was repeated twice with similar results. **e**, RT–qPCR showing normalized IAP mRNA levels relative to 0 h using Δ1- or *gag*-specific primers after auxin-induced degradation of METTL14 (light blue), METTL3 (blue) and METTL3;METTL14 double degron (dark blue). TIR1-only (black) ES cells served as control. Data are mean ± s.d. of three independent auxin inductions). **f**, Growth curves of TIR1-only and AID-METTL3;METTL14 double degron ES cells treated with either DMSO (grey) or auxin (black) for 0–96 h. Data are mean ± s.d. of three independent auxin inductions. Note that contrary to constitutive m^6A-knockout ES cells, proliferation rate is not altered by acute m^6A loss (similar rate between AID-METTL3;-METTL14 double degron and TIR1-only ES cells). However, prolonged auxin treatment may have negative effect on the proliferation rate. **g**–**i**, ZC3H13 auxin-dependent degron. Schematic of ZC3H13 degron engineering in mouse ES cells (**g**), immunoblot showing protein levels of endogenously 3×Flag-AID-tagged ZC3H13 (**h**) and degron efficiency in presence of auxin (**i**). TIR1-only ES cells were used as control for protein levels. KAP1 served as loading control. **j**, RT–qPCR showing normalized IAP mRNA levels relative to 0 h using LTR-specific primers after auxin-induced degradation of ZC3H13. Data are mean ± s.d. from three independent auxin inductions. Immunoblots presented in **b**, **c**, **h**, **i** were repeated at least twice with similar results.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Retrotransposon chromatin profiling in AID-METTL3;METTL14 double degron ES cells using random assignment of multi-mapped reads on full-length elements. a, Composite profiles (top) and density plots (bottom) showing enrichment of H3K4me3, H3K27ac and H3K9me3 as determined by CUT&RUN on m^{6}A-bound and non-m^{6}A bound genes (±2 kb) (gene lists were obtained based on MeRIP-seq results) in AID-METTL3;METTL14 double degron ES cells treated with DMSO or auxin for 96 h. **b**, Heat map showing average H3K27ac, H3K4me3 and H3K9me3 levels (log_2(immunoprecipitate/IgG)) for 56 retrotransposon families in AID-METTL3;METTL14 double degron ES cells treated with DMSO or auxin for 96 h. **c**, Violin plots and box plots (inside) showing H3K4me3, H3K27ac and H3K9me3 levels at L1MdA_I-IV copies (n = 436 copies) in AID-METTL3;METTL14 double degron ES cells treated with DMSO (grey) or auxin (red) for 96 h. Only copies with a minimum of 10 reads in at least one sample were conserved. Box plots are as in Fig. 2d. P values were determined by two-sided Student's t-test. **d**, Composite profiles showing H3K4me3 (top), H3K27ac (middle) and H3K9me3 (bottom) coverage along full-length (>5 kb) L1MdA_I copies (n = 1,400) in AID-METTL3;METTL14 double degron ES cells treated with DMSO (control, grey) or auxin (red) for 96 h. Results in **a**–**d** represent averages of two independent CUT&RUN.



Extended Data Fig. 10 | YTHDF2 binds to ERV mRNAs. a, Schematic representation showing sgRNA targeting of the *Ythdf2* gene. **b**, Immunoblot showing YTHDF2 protein levels in wild-type and YTHDF2-depleted cells. PCNA served as loading control. Immunoblotting was repeated twice with similar results **c**, RIP–qPCR showing YTHDF2 enrichment on indicated ERVKs and *Nanog* in wild-type and *Ythdf2*-knockout ES cells (two independent clones). Rabbit IgG served as control. Data are mean ± s.d. from three independent experiments. **d**, Generation of mutant mut*Ythdf2-3* and mut*Ythdf1-2-3* ES cells using CRISPR–Cas9. Schematic representation showing sgRNA sequences; mutation/deletion information based on Sanger sequencing is provided.

nature research

Corresponding author(s): Deborah Bourc'his
Tomasz Chelmicki

Last updated by author(s): Sep 30, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No costum software was used in this study. Microscopy images were acquired using Upright Spinning disk Confocal Microscope (Roper/ Zeiss). Sequencing data was collected using Illumina HiSeq 2500 and Illumina NovaSeq 6000 Platforms. The qPCR was performed on a ViiA 7 Real-Time QPCR System (Thermo Fisher Scientific). For Western blot, membranes were scanned using a ChemiDoc MP (v.) (BioRad). The FACS analysis was performed using NovoCyte 2000R flow cytometer and NovoExpress Software (v1.2.1) (ACEA Biosciences).
Data analysis	MAGeCK (v0.5.8), Atropos (v1.1.16), STAR (v2.7.0a), One code to find them all (v1.0), featureCounts (v1.5.1), R (v3.5.0), edgeR (v3.22.3), limma (v3.38.3), regioneR (v1.14.0), TrimGalore (v0.4.4), Bowtie (v1.2), deepTools (v2.5.3), MACS2 (v2.1.1), Bowtie2 (v2.2.9), STAR (v2.6.0c), RSAT (v), Picard (v.2.6.0), Trumpet (v0.3.6), TapeStation Controller Software (v3.2), NovoExpress (v1.2.1), Prism (v8.3.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated in this study are available on GEO database under the number GSE145616

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples size was chosen to ensure reproducibility of the results at affordable costs. Sample size was determined based on previously published studies and equivalent to what is routinely used for respective assay. Sample sizes are indicated for all experiments. At least 2 and typically 3 independent experiments were carried out for most of the assays. For CRISPR-Cas9 knock-out screening, the sample size was determined based on the design as a positive selection strategy and library recommendations. For Screen_I, one pre-selection time point, three selection time points (5-day, 7-day and 9-day long blasticidin S selection) and one non-selection time point (9-day, negative control) were sequenced. Based on the results from Screen I, showcasing increasing selection stringency with extended blasticidin S treatment, independent Screen_II was performed in 2 biological replicates with 9-day long blasticidin S selection following either 8-day long, or 17-day long cell culture (early and late selection respectively). RNA-seq were obtained from biological duplicates. MeRIP-seq was performed in independent triplicates from wildtype and Mettl3-KO cells. Input for each of the cell line was sequenced as a control. CUT&RUN-seq were performed in biological duplicates consistent with comparable studies.
Data exclusions	When experimental mistakes occurred or in case of low sample quality or occasional loss of samples/reagents, the experiment was discarded and repeated. Otherwise, no data were excluded from the analysis.
Replication	All attempts at replication were successful and noted in the relevant figure legends.
Randomization	No randomization techniques were required, as the study was based on molecular and cellular biology techniques, which did not involve live organisms or required allocation of experimental units.
Blinding	The investigators were not blinded to sample allocation during experiments and outcome assessment, because results were obtained using objective quantitative methods.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*

Randomization

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample

Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection

Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions

Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access and import/export

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance

Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Primary antibodes
Antibody, Supplier, Catalog number, Host&Class, Application, Dilution
anti-METTL14, Abcam, ab220030, Mouse monoclonal [CL4252], Western blot, 1:500
anti-METTL3, Abcam, ab195352, Rabbit monoclonal [EPR18810], Western blot, 1:1000
anti-ZC3H13, Invitrogen, PA5-36515, Rabbit polyclonal, Western blot, 1:500
anti-WTAP, Proteintech, 10200-1-AP, Rabbit polyclonal, Western blot, 1:2000
anti-YTHDF1, Proteintech, 17479-1-AP, Rabbit polyclonal, Western blot, 1:1000
anti-YTHDF2, MBL, RN123PW, Rabbit polyclonal, Western blot, 1:1000
anti-YTHDF2, Proteintech, 24744-1-AP, Rabbit polyclonal, RNA Immunoprecipitation (RIP), (4µg/IP)
anti-YTHDF3, Abcam, ab220161, Rabbit monoclonal [EPR21912-3], Western blot, 1:500
anti-PCNA, DAKO, M0879 Mouse monoclonal [PC10] , Western blot, 1:3000
anti-FLAG, Sigma, F1804, Mouse monoclonal [M2], Western blot, 1:1000
anti-KAP1, Abcam, ab10483, Rabbit polyclonal, Western blot, 1:2000
anti-LAMIN B1, Abcam, ab16048, Rabbit polyclonal, Western blot, 1:3000
anti-Cas9, Active motif, 61757, Mouse monoclonal [8C1-F10], Western blot, 1:1000
anti-H3K4me3, Millipore, 07-473, Rabbit polyclonal, Cut@Run, 2µg/experiment
anti-H3K9me3, Abcam, ab176916, Rabbit monoclonal [EPR16601], Cut@Run, 2µg/experiment
anti-H3K27ac, Abcam, ab4729, Rabbit polyclonal, Cut@Run, 2µg/experiment
IgG from Rabbit serum, Sigma, PP54, Cut&Run (2µg/experiment), RIP, (4µg/IP)
IAP-GAG, gift from Dr. Bryan R. Cullen (Duke University School of Medicine, MI, USA), Immunofluorescence (IF), 1:200

Secondary Antibodies
Antibody, Supplier, Catalog Number, Class, Application, Dilution
Goat-anti Mouse IgG (H+L) HRP, Thermo Fisher Scientific, G-21040 polyclonal, western blot 1:10000
Goat-anti Rabbit IgG (H+L) HRP, Thermo Fisher Scientific, G-21234 polyclonal, western blot 1:10000
Goat-anti Rabbit IgG (H+L) Alexa Fluor 488, Thermo Fisher Scientific, A-11008, polyclonal immunofluorescence (IF) 1:500

Validation

Accept anti-IAP-GAG antibody all used antibodies are commercially available and have been validated by manufacturer.
Antibody validations and validation criteria are available on the following websites:
anti-METTL14 https://www.abcam.com/mettl14-antibody-cl4252-ab220030.html
anti-METTL3 https://www.abcam.com/mettl3-antibody-epr18810-ab195352.html
anti-ZC3H13 https://www.thermofisher.com/antibody/product/ZC3H13-Antibody-Polyclonal/PA5-36515
anti-WTAP https://www.ptglab.com/products/WTAP-Antibody-10200-1-AP.htm
anti-YTHDF1 https://www.ptglab.com/products/YTHDF1-Antibody-17479-1-AP.htm
anti-YTHDF2 https://www.mblbio.com/bio/g/dtl/A/index.html?pcd=RN123PW
anti-YTHDF2 https://www.ptglab.com/products/YTHDF2-Antibody-24744-1-AP.htm
anti-YTHDF3 https://www.abcam.com/ythdf3-antibody-epr21912-3-ab220161.html
anti-PCNA https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/proliferating-cell-nuclear-antigen-(concentrate)-76551
anti-FLAG https://www.sigmaaldrich.com/catalog/product/sigma/f1804?lang=fr&region=FR
anti-KAP1 https://www.abcam.com/kap1-antibody-ab10483.html
anti-LAMIN B1 https://www.abcam.com/lamin-b1-antibody-nuclear-envelope-marker-ab16048.html
anti-Cas9 https://www.activemotif.com/catalog/details/61757/cas9-antibody-mab-clone-8c1-f10
anti-H3K4me3 https://www.merckmillipore.com/FR/fr/product/Anti-trimethyl-Histone-H3-Lys4-Antibody,MM_NF-07-473
anti-H3K9me3 https://www.abcam.com/histone-h3-tri-methyl-k9-antibody-epr16601-chip-grade-ab176916.html
anti-H3K27ac https://www.abcam.com/histone-h3-acetyl-k27-antibody-chip-grade-ab4729.html
IgG from Rabbit serum https://www.sigmaaldrich.com/catalog/product/sigma/i5006?lang=fr&region=FR
Goat-anti Mouse IgG (H+L) HRP https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/G-21040
Goat-anti Rabbit IgG (H+L) HRP https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/G-21234
Goat-anti Rabbit IgG (H+L) Alexa Fluor 488 https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11008

The anti-IAP-GAG antibody was a kind gift from Dr. Bryan R. Cullen (Duke University School of Medicine, MI, USA) and was described in Dewannieux, M., Dupressoir, A., Harper, F. et al. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nat Genet 36, 534–539 (2004). https://doi.org/10.1038/ng1353

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	All embryonic stem cell lines generated in this study were derived from the a well-characterized E14 mouse embryonic stem cell line. HEK293FT cell line was obtained from Institut Curie CRISPRit Screening Platform.
Authentication	Non of the cell lines were authenticated.
Mycoplasma contamination	ES Cell lines were routinely tested for mycoplasma contamination and tested negative. HEK293FT cell line was not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	Cells used are not in the ICLAC database

Palaeontology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJEguidelines for publication of clinical research and a completedCONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE156007

Files in database submission

SuperSeries: GSE145616 m6A methylation regulates the fate of endogenous retrovirus transcripts.
SubSeries:
GSE145309 [RNA-seq]:

GSM4314580 WT ESC_Rep 1
GSM4314581 WT ESC_Rep 2
GSM4314582 Mettl3 KO ESC_Rep 1
GSM4314583 Mettl3 KO ESC_Rep 2
GSM4314584 Mettl14 KO ESC_Rep 1
GSM4314585 Mettl14 KO ESC_Rep 2
GSM4314586 WTAP KO ESC_Rep 1
GSM4314587 WTAP KO ESC_Rep 2
GSM4314588 Zc3h13 KO ESC_Rep 1
GSM4314589 Zc3h13 KO ESC_Rep 2

GSE145315 [RIP-seq]:

GSM4314656 Input_WT
GSM4314657 Input_KO
GSM4314658 m6A_WT_rep1
GSM4314659 m6A_WT_rep2
GSM4314660 m6A_WT_rep3
GSM4314661 m6A_M3KO_rep1
GSM4314662 m6A_M3KO_rep2
GSM4314663 m6A_M3KO_rep3

GSE145615 [CRISPR/Cas9]:

GSM4322248 Screen I_seq1_+puro
GSM4322249 Screen I_seq2_-blast
GSM4322250 Screen I_seq3_blast5D
GSM4322251 Screen I_seq4_blast7D
GSM4322252 Screen I_seq5_blast9D
GSM4322253 Screen II_early_-blast_rep1
GSM4322254 Screen II_early_-blast_rep2
GSM4322255 Screen II_early_+blast_rep1
GSM4322256 Screen II_early_+blast_rep2
GSM4322257 Screen II_late_-blast_rep1
GSM4322258 Screen II_late_-blast_rep2
GSM4322259 Screen II_late_+blast_rep1
GSM4322260 Screen II_late_+blast_rep2

GSE156007 [CUT & RUN]

GSM4718889 Input AID-METTL3:-METTL14 dd (noIAA)
GSM4718890 Input AID-METTL3:-METTL14 dd (+IAA)
GSM4718891 IP:H3K4me3 AID-METTL3:-METTL14 dd (noIAA) Rep1
GSM4718892 IP:H3K4me3 AID-METTL3:-METTL14 dd (noIAA) Rep2
GSM4718893 IP:H3K4me3 AID-METTL3:-METTL14 dd (+IAA) Rep1
GSM4718894 IP:H3K4me3 AID-METTL3:-METTL14 dd (+IAA) Rep2
GSM4718895 IP:H3K27ac AID-METTL3:-METTL14 dd (noIAA) Rep1
GSM4718896 IP:H3K27ac AID-METTL3:-METTL14 dd (noIAA) Rep2
GSM4718897 IP:H3K27ac AID-METTL3:-METTL14 dd (+IAA) Rep1
GSM4718898 IP:H3K27ac AID-METTL3:-METTL14 dd (+IAA) Rep2
GSM4718899 IP:H3K9me3 AID-METTL3:-METTL14 dd (noIAA) Rep1
GSM4718900 IP:H3K9me3 AID-METTL3:-METTL14 dd (noIAA) Rep2
GSM4718901 IP:H3K9me3 AID-METTL3:-METTL14 dd (+IAA) Rep1
GSM4718902 IP:H3K9me3 AID-METTL3:-METTL14 dd (+IAA) Rep2

Genome browser session
(e.g. UCSC)

No web genome browser is available, but an IGV session can be made available upon request.

Methodology

Replicates

CUT@RUN-seq for H3K4me3, H3K27ac and H3K9me3 were performed in biological duplicates for each condition. Single IgG was used as negative control for each condition.

Sequencing depth	For CUT@RUN, on average: 3x in depth, 124 million of sequenced reads (PE50) and 108 million of mapped reads per sample. For precise information on number of sequenced and mapped reads for CUT&RUN-seq, as well as CRISPR-Cas9 screen, RNA-seq and meRIP-seq please see Supplementary Table 5.
Antibodies	Antibody, Supplier, Catalog Number, Host, Class anti-H3K4me3, Millipore, 07-473, Rabbit, polyclonal, anti-H3K9me3, Abcam, ab176916, Rabbit, monoclonal [EPR16601] anti-H3K27ac, Abcam, ab4729, Rabbit, polyclonal IgG from Rabbit serum, Sigma, PP54
Peak calling parameters	Paired-end reads were trimmed using Trim Galore v0.4.4. The alignment was performed onto a concatenated genome using the Mouse reference genome (mm10) and the Escherichia coli genome (str. K-12 substr. MG1655, Genbank: NC_000913) with STAR v2.7.0a35 reporting randomly one position, allowing 6% of mismatches (--outFilterMultimapNmax 5000 --outSAMmultNmax 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.06). PCR duplicates were removed using Picard v2.6.0. Peaks were not called.
Data quality	Quality control (adapters and low-quality bases trimming) was performed using TrimGalore v0.4.4
Software	Trim Galore v0.4.4. STAR v2.7.0a. Picard v2.6.0

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Live cells were collected, washed and resuspended in PBS buffer
Instrument	NovoCyte 2000R flow cytometer (ACEA Biosciences).
Software	NovoExpress v1.2.1 (ACEA Biosciences)
Cell population abundance	Minimum of 50,000 cells were counted for each analysis
Gating strategy	The percentage of GFP-positive cells was determined upon definition of three gates: i) FSC-H vs SSC-H to isolate cells from debris,ii) SSC-H vs SSC-A to isolate single cells and iii) SSC-H vs FITC-H for detection of GFP-positive population.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) *Specify: functional, structural, diffusion, perfusion.*

Field strength *Specify in Tesla*

Sequence & imaging parameters *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*

A.2 Article 4: SPEN is required for *Xist* upregulation during initiation of X chromosome inactivation *(BioRxiv. 2020)*

Robert-Finestra T, Tan BF, Mira-Bontenbal H, Timmers E, Gontan-Pardo C, Merzouk S, Giaimo BD, **Dossin F**, van Ijcken WFJ, Martens JWM, Borgreffe T, Heard E, Gribnau J. *SPEN is Required for Xist Upregulation during Initiation of X Chromosome Inactivation.* **bioRxiv**. 2020.

ARTICLE ON NEXT PAGE

SPEN is Required for *Xist* Upregulation during Initiation of X Chromosome Inactivation

Authors

Teresa Robert-Finestra[1], Beatrice F. Tan[1,6], Hegias Mira-Bontenbal[1,6], Erika Timmers[1,6], Cristina Gontan-Pardo[1], Sarra Merzouk[1], Benedetto Daniele Giaimo[2], François Dossin[3], Wilfred F. J. van IJcken[4], John W. M. Martens[5], Tilman Borggrefe[2], Edith Heard[3] & Joost Gribnau[1,*]

Affiliations

[1] Department of Developmental Biology, Erasmus University Medical Center, Oncode Institute, 3015CN Rotterdam, The Netherlands.

[2] Institute of Biochemistry, University of Giessen, 35392 Giessen, Germany.

[3] European Molecular Biology Laboratory, Director's Research, 69117 Heidelberg, Germany.

[4] Center for Biomics, Erasmus University Medical Center, 3015CN Rotterdam, The Netherlands.

[5] Department of Medical Oncology, Erasmus MC Cancer Institute and Cancer Genomics Netherlands, Erasmus University Medical Center, 3015CN Rotterdam, The Netherlands.

[6] These authors contributed equally.

*Correspondence: j.gribnau@erasmusmc.nl

Abstract

At initiation of X chromosome inactivation (XCI), *Xist* is monoallelically upregulated from the future inactive X (Xi) chromosome, overcoming repression by its antisense transcript *Tsix*. *Xist* recruits various chromatin remodelers, amongst them SPEN, which are involved in silencing of X-linked genes in *cis* and establishment of the Xi. Here, we show that SPEN plays an important role in the initiation of XCI. *Spen* null female mouse embryonic stem cells (ESCs) are defective in *Xist* upregulation upon differentiation. We find that *Xist*-mediated SPEN recruitment to the Xi chromosome happens very early in XCI, and that SPEN-mediated silencing of the *Tsix* promoter is required for *Xist* upregulation. Accordingly, failed *Xist* upregulation in *Spen*$^{-/-}$ ESCs can be rescued by concomitant removal of *Tsix*. These findings indicate that SPEN is not only required for the establishment of the Xi, but is also crucial in the initiation of the XCI process.

Introduction

To compensate for gene dosage imbalance between females (XX) and males (XY), female placental mammals randomly inactivate one X chromosome early during embryonic development[1]. In mice, random X chromosome inactivation (XCI) takes place in the epiblast in three phases: initiation, establishment and maintenance. During the initiation phase, the long non-coding RNA (lncRNA) *Xist* is upregulated from the future inactive X (Xi) chromosome[2–4]. *Xist* is located within the X Chromosome Inactivation Centre (*Xic*), an X-linked region required for XCI that contains different *cis*-regulatory elements, including *Tsix*, another lncRNA gene that is transcribed antisense to and completely overlaps *Xist*[5]. *Tsix* negatively regulates *Xist* expression via antisense transcription and chromatin remodelling[6–12]. Together, *Xist* and *Tsix* form a master switch that controls initiation of XCI: in the pluripotent state, *Tsix* is biallelically expressed, repressing both alleles of *Xist*, while upon initiation of XCI the balance changes, resulting in downregulation of *Tsix* and upregulation of *Xist* on the future Xi. *Trans*-regulators including pluripotency factors OCT4, NANOG, REX1 as well as XCI activators are key in controlling this switch by regulating *Xist* and *Tsix* transcription (reviewed in[13]). Later, during the establishment phase of XCI, the 17 kb lncRNA *Xist* spreads in *cis* along the future Xi and recruits different proteins involved in gene silencing that render the X transcriptionally inactive (reviewed in[14]). Active histone marks (H3K4me2/me3, H3 and H4 acetylation) are removed and repressive histone marks are instigated, catalysed by Polycomb group complexes and

48 other protein complexes (reviewed in[15]). Finally, in the maintenance phase, the inactive state of the Xi is
49 epigenetically propagated across cell divisions.

50 Different studies consisting of *Xist* RNA immunoprecipitations coupled to mass spectrometry[16–18] and
51 genetic screens[19,20] identified SPEN (also known as SHARP in human and MINT in mouse) as a crucial factor
52 in the establishment phase of XCI. SPEN is a large protein with four N-terminal RNA Recognition Motifs
53 (RRM) and a highly conserved C-terminal SPOC domain able to recruit different proteins involved in
54 transcriptional silencing[21,22]. SPEN is also involved in the Notch signalling pathway and nuclear receptor
55 signalling, where it acts as a transcriptional corepressor[23,24].

56 SPEN is crucial for X-linked gene silencing[16,17,19,20] by binding the *Xist* Repeat A (RepA) via its RRM
57 domains[17,19] and interacting via its SPOC domain with the corepressors NCoR/SMRT to recruit/activate
58 histone deacetylase 3 (HDAC3), which is responsible for the removal of histone H3 and H4 acetylation at
59 promoters and enhancers of genes located on the future Xi[16,25]. Despite this crucial role for SPEN in
60 establishment of the Xi, these studies did not report defects in *Xist* upregulation and coating[16,17,19,20,25]
61 (Supplementary Table 1).

62 Here, we show that SPEN accumulates on the Xi very early during differentiation and is required for *Xist*
63 upregulation. We show that SPEN has a dual function, required to silence *Tsix* and facilitate *Xist*
64 upregulation, while also stabilizing *Xist* RNA. Together, our results indicate that SPEN is not only necessary
65 for X-linked gene silencing but also plays a crucial earlier role in the regulation of initiation of XCI.

66

Results

SPEN is required for *Xist* upregulation

Previous work has shown how SPEN is crucial for silencing of X-linked genes, but these studies did not investigate the role of SPEN in initiation of *Xist* expression. Therefore, we generated *Spen* homozygous (*Spen*$^{-/-}$) and heterozygous (*Spen*$^{+/-}$) knockout mouse embryonic stem cells (ESCs) by deleting the complete open reading frame (ORF) using the CRISPR/Cas9 technology (Supplementary Fig. 1a). These lines were generated in a hybrid F1 129/Sv:Cast/EiJ (129/Cast) genetic background with a doxycycline-responsive endogenous *Xist* promoter located on the Cast X chromosome[26] (Fig. 1a). The S*pen* ORF deletion was verified by PCR on genomic DNA (gDNA) and Western blot analysis (Supplementary Fig. 1b-c). Allele-specific RNA-seq analysis (Supplementary Fig. 1d) of wild type (Wt) undifferentiated ESCs (day 0) containing the doxycycline-responsive *Xist* promoter treated with and without doxycycline for 4 days (Fig. 1b) showed skewed X-linked gene silencing towards the Cast allele (Fig. 1c top-left, Supplementary Fig. 1e). On the other hand, the same analysis in *Spen*$^{-/-}$ ESCs showed impaired X-linked gene silencing (Fig. 1c bottom-left), as described before[16,17,19,20] (Supplementary Table 1).

To trigger XCI in the context of differentiation, we forced *Xist* upregulation by doxycycline treatment followed by monolayer differentiation (Fig. 1b). Allele-specific RNA-seq analysis of Wt and *Spen*$^{-/-}$ ESCs treated with doxycycline at day 7 of monolayer differentiation (Supplementary Fig. 1d) also revealed a lack of X-linked gene silencing along the entire X-chromosome (Fig. 1c bottom-right, Supplementary Fig. 1e). Similarly, allele-specific RT-qPCR analysis of the X-linked gene *Rnf12* revealed impaired silencing (Fig. 1d). Although previous work indicated that in *Spen*$^{-/-}$ ESCs a group of lowly expressed X-linked genes are susceptible to SPEN-independent gene silencing[27], our RNA-seq analysis shows no silencing of this specific group of genes (Supplementary Fig. 1f,g). We observed *Xist* upregulation in Wt and *Spen*$^{-/-}$ cells, however doxycycline induction resulted in lower *Xist* expression levels in *Spen*$^{-/-}$ compared to Wt ESCs (Fig. 1e), an effect that was also reported in a recent study[27].

Next, we recapitulated physiological XCI by monolayer differentiation in the absence of doxycycline to allow normal *Xist* upregulation and X-linked gene silencing (Fig. 1b). Allele-specific RNA-seq analysis of Wt and *Spen*$^{-/-}$ ESCs at day 3 without doxycycline (Supplementary Fig. 1d) shows no significant differences in silencing (Fig. 1c top-right, Supplementary Fig. 1e), making evident the time scale differences between forced *Xist* upregulation in the undifferentiated ESC stage and physiological XCI. Even so, at the latest time points of differentiation (day 5 and 7) *Spen*$^{-/-}$ cells clearly lost the capacity to induce *Rnf12* silencing (Fig. 1f). Remarkably, in *Spen*$^{-/-}$ cells, *Xist* upregulation from the 129 allele was completely abrogated (Fig. 1g),

contrasting earlier evidence that suggested that SPEN is not required for *Xist* upregulation and coating [16,17,19,20,25] (Supplementary Table 1). In addition, differentiating *Spen*$^{-/-}$ ESCs lack *Xist* clouds, determined by RNA-FISH (Fig. 1h,i), while *Tsix* was significantly more expressed from the Wt 129 allele in *Spen*$^{-/-}$ cells compared to Wt cells at day 5 and 7 of monolayer differentiation (Fig. 1j), suggesting that SPEN might be necessary for *Tsix* silencing. Importantly, RT-qPCR analysis confirmed proper silencing of pluripotency genes *Rex1* and *Nanog*, and upregulation of endoderm marker *Gata6* (Supplementary Fig. 1h) in *Spen*$^{-/-}$ ESCs upon monolayer differentiation, indicating that loss of *Xist* expression is not related to defective ESCs differentiation.

Furthermore, while *Spen*$^{+/-}$ cells are able to upregulate *Xist* and silence *Rnf12* upon doxycycline treatment followed by monolayer differentiation (Supplementary Fig. 2a,b), they show reduced *Rnf12* silencing upon physiological differentiation without a defect in *Xist* upregulation (Supplementary Fig. 2c,d). Given that SPEN levels are reduced in Spen$^{+/-}$ cells (Supplementary Fig. 1c), these results demonstrate that SPEN dosage is important in XCI.

Spen rescue leads to normal *Xist* expression levels

To confirm previous results, we performed a rescue experiment by stably re-expressing *Spen* through introduction of the full-length *Spen* cDNA in the *ROSA26* locus[25] of *Spen*$^{-/-}$ ESCs (Supplementary Fig. 3a). Successful integration and expression of *Spen* was verified by PCR on gDNA and RT-qPCR, respectively (Supplementary Fig. 3b,c). The generated rescue clones (Clone A, B and C) display two- to three-fold overexpression of *Spen* RNA at the ESC stage and during monolayer differentiation (Supplementary Fig. 3c). In contrast to the *Spen*$^{-/-}$ lines, all *Spen*$^{-/-(cDNA)}$ rescue clones express *Xist* at similar levels to Wt clones during monolayer differentiation (Fig. 2a). Allele-specific expression analysis of *Rnf12* and *Tsix* (129 allele) indicated that the silencing defect is partially rescued in *Spen*$^{-/-(cDNA)}$ clones (Supplementary Fig. 3d,e). Interestingly, in undifferentiated *Spen*$^{-/-(cDNA)}$ ESCs clones, *Xist* levels were higher than in Wt controls (Fig. 2a), and *Xist* RNA-FISH analysis revealed a significant percentage of *Spen*$^{-/-(cDNA)}$ ESCs having *Xist* clouds (Fig. 2b,c). This abnormal cloud formation could be related to higher SPEN abundance due to overexpression in the undifferentiated state, possibly stabilizing *Xist* or silencing *Tsix*. In addition, these results indicate that the observed defect in *Xist* expression is SPEN mediated and takes place at the very early initiation steps of XCI.

SPEN accumulation on the Xi upon ESCs differentiation shows two distinguishable states: early and late SPEN accumulation

In light of the previous results, we expect SPEN to accumulate on the Xi at the very early steps of physiological XCI. To investigate this hypothesis and since commercial SPEN antibodies suitable for immunofluorescence (IF) are lacking, we generated ESCs endogenously expressing a C-terminally tagged SPEN-GFP (Supplementary Fig. 4a). Correct GFP integration was confirmed by PCR on gDNA and by FACS analysis (Supplementary Fig. 4b,c). IF detection of GFP at different time points of monolayer differentiation revealed SPEN accumulation as early as day 1 in about 30% of the nuclei (Fig. 3a,b). Interestingly, we could distinguish two different states: early and late SPEN accumulation on the Xi (Fig. 3a-c). Early accumulation was dispersed and had a lower IF intensity, while SPEN accumulation at later stages was more compact and with a higher IF signal. Early accumulation signals were only detected at the initial days of monolayer differentiation (day 1 to 5), while late accumulation signals appeared at day 2, progressively increased over time and plateaued at day 7 (Fig. 3b).

Previous studies have shown that SPEN and *Xist* co-localize in doxycycline-inducible *Xist* undifferentiated ESC lines[20,25], although the normal timing and relation between SPEN and *Xist* in differentiating cells remained unexplored. Therefore, we investigated SPEN accumulation in relation to *Xist*, by performing GFP IF combined with *Xist* RNA FISH in differentiating cells (Fig. 3d, Supplementary Fig. 4d). This analysis indicated that at day 2 of differentiation, about 55% of the cells with *Xist* clouds showed SPEN co-localization (Fig. 3e). We also studied the relation of SPEN with other key players in XCI, including the PRC2 catalytic subunit EZH2 and its catalytic product H3K27me3. As found for *Xist*, we could detect co-localization of EZH2 and its associated histone modification with SPEN (Supplementary Fig. 4e-g). At day 2 about 60% of the nuclei with early SPEN accumulation displayed EZH2 accumulation in the absence of H3K27me3 deposition, whereas at day 5 about 90% of the cells showed H3K27me3 enrichment at SPEN positive Xi (Fig. 3f). Our results show that SPEN accumulation on the Xi can be distinguished in an early state and late state of SPEN accumulation, where the early state is marked by small dispersed accumulation of SPEN co-localizing with EZH2 without detectable H3K27me3 at day 2, whereas the late state SPEN accumulation is compact and co-localizes with EZH2 and H3K27me3.

Xist RNA stability is compromised in *Spen*$^{-/-}$ ESCs

We have shown that SPEN is required for endogenous *Xist* upregulation and XCI upon ESCs differentiation. In addition, we noticed that forced *Xist* upregulation resulted in reduced *Xist* RNA levels in *Spen*$^{-/-}$ cells compared to Wt cells (Fig. 1e, Fig. 4a). Accordingly, the percentage of nuclei with *Xist* clouds was lower in *Spen*$^{-/-}$ cells with doxycycline induced *Xist* versus Wt cells and *Xist* clouds were in general smaller (Fig. 4b,c). This finding could be explained by a role for SPEN in stabilizing *Xist* RNA by complex formation. To study the role of SPEN in *Xist* stability, we determined the half-life of doxycycline-induced *Xist* RNA in *Spen*$^{-/-}$ and Wt ESCs treated with actinomycin D to block its transcription. The remaining levels of *Xist* RNA at different time-points was assessed by RT-qPCR, and the *Xist* RNA decay rate and half-life were then calculated (Fig. 4d). In Wt ESCs, the *Xist* RNA half-life was 6h 37min, similar to what was described in a previous study[28], whereas the *Xist* half-life was reduced to 3h 52min in *Spen*$^{-/-}$ cells. These results indicate that SPEN plays a role in promoting *Xist* RNA stability. However, these results cannot explain why physiological *Xist* upregulation is lost in *Spen*$^{-/-}$ ESCs upon differentiation.

SPEN, HDAC3 and H3K27ac are enriched at the *Tsix* regulatory region

Our results showed that at late days of monolayer differentiation, *Spen*$^{-/-}$ cells display higher *Tsix* levels than control cells (Fig. 1j), suggesting that SPEN might be recruited by *Xist* to silence *Tsix*. Hence, we explored SPEN genomic binding at the *Xist-Tsix* locus using published SPEN CUT&RUN data in undifferentiated ESCs with a doxycycline-responsive *Xist* promoter[25]. SPEN accumulates on the *Xist* gene body as well as on the *Tsix* regulatory region, SPEN accumulation is evident at 24 hours of doxycycline induction, but more prominent after 4 and 8 hours of induction (Fig. 5a,b). This *Tsix* regulatory region comprises the minor and major *Tsix* promoters and *Xite*, an enhancer of *Tsix*[29]. *Xist*-mediated recruitment of SPEN is important for recruitment and/or activation of HDAC3, responsible for the removal of H3K27ac from the future Xi[16,30]. Analysis of published HDAC3 and H3K27ac ChIP-seq data from undifferentiated female ESCs upon 24h *Xist* induction[30] reveals HDAC3 binding and H3K27ac loss at the *Tsix* regulatory region, where SPEN is recruited (Fig. 5a). Moreover, X-linked *Rnf12* (Supplementary Fig. 5a) and *Pgk1* (Supplementary Fig. 5b) display SPEN and HDAC3 enrichment at their promoters and H3K27ac loss upon *Xist* induction. Interestingly, *Rnf12*, an early silenced gene[31], shows SPEN and HDAC3 promoter binding in the undifferentiated state without *Xist* induction, suggesting that in the undifferentiated state *Xist* might be sufficiently expressed at very low levels to partly silence *Rnf12*, explaining the *Rnf12* allelic ratio difference between Wt and *Spen*$^{-/-}$ ESCs at day 0 (Fig. 1f, Supplementary Fig. 3d).

To test whether SPEN recruitment leads to H3K27me3 accumulation, we performed H3K27me3 ChIP-seq on *Spen*$^{-/-}$ and Wt ESCs prior to and after differentiation (day 3) (Fig. 5a). This analysis revealed H3K27me3 enrichment in the *Tsix* regulatory region in Wt cells upon differentiation, while *Spen*$^{-/-}$ cells do not show H3K27me3 enrichment in accordance with the lack of *Xist* upregulation. Interestingly, the H3K27me3 hotspot, located at the 3' end of *Tsix*[32,33] is clearly reduced upon differentiation both in Wt and *Spen*$^{-/-}$ cells, while Wt and *Spen*$^{-/-}$ cells show no difference in H3K27me3 levels, indicating that SPEN does not play a role in H3K27me3 enrichment at the hotspot (Fig. 5a). Altogether, these data support a model where *Xist*-mediated SPEN recruitment leads to *Tsix* promoter silencing.

SPEN is required to silence *Tsix* to allow *Xist* upregulation

To further investigate the role of SPEN in silencing *Tsix*, we generated compound *Spen*$^{-/-}$:*Tsix* defective hybrid ESC lines. If *Xist*-mediated recruitment of SPEN to the *Tsix* regulatory region is crucial for *Tsix* silencing and *Xist* upregulation, we expect *Xist* upregulation upon differentiation to be rescued in these double knockout cell lines. We made use of a *Tsix*-Stop line containing a triple poly(A) signal downstream of the major *Tsix* promoter that blocks its transcription on the 129 allele[7], and a *Tsix*-Cherry line with a mCherry coding sequence introduced downstream of the major *Tsix* promoter on the Cast allele[34] (Fig. 5c). As controls, we generated *Spen*$^{-/-}$ ESCs in the same hybrid background (F1:129/Cast) where both *Xist* alleles are intact[35], in contrast to the previously studied ESCs that contained one doxycycline-inducible *Xist* allele. Deletion of the *Spen* ORF was confirmed in all cell lines by PCR on gDNA and Western blot analysis (Supplementary Fig. 6a,b). For each line we generated two independent knockout clones. All the ESC lines were differentiated in parallel and in biological duplicates, followed by allele-specific *Xist* RNA expression analysis by RT-qPCR. As expected, the F1:129/Cast control *Spen*$^{-/-}$ line showed no *Xist* upregulation (Fig. 5d), similar to the phenotype observed in the *Spen* knockout clones generated in the heterozygous doxycycline-inducible *Xist* cell line (Fig. 1g). Remarkably, the introduction of a poly(A) signal in *Spen*$^{-/-}$:*Tsix*-Stop lines fully rescued *Xist* upregulation upon differentiation (Fig. 5e), indicating that SPEN is required for *Xist* upregulation via *Tsix* repression. However, the *Spen*$^{-/-}$*:Tsix*-Cherry lines displayed only mild upregulation of *Xist* upon differentiation (Fig. 5e). These differences in the rescue phenotype between *Spen*$^{-/-}$*:Tsix*-Cherry and *Spen*$^{-/-}$:*Tsix*-Stop lines can be explained by remaining *Tsix* transcription in the *Tsix*-Cherry line, which is fully ablated in the *Tsix*-Stop lines (Supplementary Fig. 6c). As expected, the *Spen*$^{-/-}$: *Tsix*-Stop lines are not able to silence X-linked *Rnf12* despite normal *Xist* levels in the *Tsix*-Stop line

(Supplementary Fig. 6d). Moreover, the Spen$^{-/-}$:*Tsix*-Stop lines display *Xist* clouds that were similar in morphology (Fig. 5f) and number compared to Wt F1:129/Cast cells (Fig. 5g).

Taken together, our results indicate a novel and essential role for SPEN in *Xist* upregulation, mainly via silencing of *Tsix*, and to a lesser extent by stabilizing *Xist* RNA. SPEN is therefore not only crucial in X-linked gene silencing but also in the early initiation steps of XCI, playing a role in the feedforward loop leading to *Xist* activation (Fig. 6).

Discussion

The role of the silencing factor SPEN in the establishment phase of XCI has been addressed in various studies, but whether SPEN is relevant for the initiation phase, involving *Xist* upregulation, remains unknown. Here, we show that SPEN-defective ESCs do not upregulate *Xist* upon differentiation and study the molecular mechanism behind this observation. To explore the role of SPEN in XCI, various studies used *Xist*-inducible ESC lines to generate SPEN knockdown or knockout cell lines (Supplementary Table 1). Forcing *Xist* expression in ESCs is a powerful way to understand X-linked gene silencing, but is not suitable to investigate the initiation phase of XCI. Studies exploring the role of SPEN in cells undergoing physiological XCI involved *Spen* knockdown strategies[17,20], therefore, the levels of SPEN might have been sufficient to allow normal *Xist* upregulation, while showing a defect in X-linked gene silencing. Likewise, our *Spen*$^{+/-}$ ESC lines show *Xist* upregulation, but reduced X-linked silencing compared to Wt lines. In agreement with our results, one study reports lower *Xist* abundance and cloud formation upon forced *Xist* induction from the endogenous locus in *Spen*$^{-/-}$ ESCs[27]. Another recent study suggested that *Spen*$^{-/-}$ ESCs are not able to differentiate upon Leukemia Inhibitory Factor (LIF) removal and differentiation towards neural progenitor cells[36], while we observe that *Spen*$^{-/-}$ ESCs display a normal morphology in the undifferentiated state and undergo normal differentiation upon monolayer differentiation. Nevertheless, we observe more cell death of *Spen*$^{-/-}$ cells compared to Wt cells upon monolayer differentiation, although we consider this might not be XCI-related since X0 *Spen*$^{-/-}$ cells also die upon differentiation (data not shown). This observation is probably not surprising since SPEN plays a role in various biological processes as a transcriptional repressor[23,24,37].

Previous studies performed with doxycycline-inducible *Xist* lines show that SPEN and *Xist* co-localize soon after *Xist* induction[20,25]. However, *Xist* upregulation from a doxycycline-inducible promoter happens at a

very different timescale than during physiological XCI. In the present study, we observe that upon *Spen*-GFP ESCs monolayer differentiation we can distinguish early and late SPEN accumulation on the Xi. Cells with SPEN early accumulation are detectable from day 1 of differentiation with their number decreasing overtime, while the number of cells with SPEN late accumulations progressively increases. Other XCI key players, such EZH2 and H3K27me3, also sequentially accumulate on the Xi.

In addition, we provide evidence that SPEN is required to silence *Tsix* to allow *Xist* upregulation. (I) In *Spen*$^{-/-}$ cells, we detect higher levels of *Tsix* at the latest time points of monolayer differentiation, compared to Wt cells. In line with this, paternal *Tsix* levels in female E3.5 blastocysts with a mutated *Xist* RepA, necessary for SPEN recruitment, are higher than in Wt blastocysts[38], suggesting that higher *Tsix* levels in *Spen*$^{-/-}$ cells are due to defective *Tsix* silencing, rather than a lack of *Xist* antisense transcription. (II) SPEN overexpression in ESCs leads to higher *Xist* levels and ectopic *Xist* cloud formation in the undifferentiated state, suggesting that SPEN overexpression might lead to partial silencing of the *Tsix* promoter, facilitating *Xist* expression. (III) SPEN binds to active promoters and enhancers to silence the X chromosome[25]. Spatial proximity to the *Xist* locus is a strong predictor of X-linked gene silencing efficiency[26,39]. One of the closest actively transcribed promoters to *Xist* is *Tsix* and we indeed observe SPEN binding at the *Tsix* regulatory region. (IV) *Spen*$^{-/-}$ cell lines cannot upregulate *Xist*, while compound *Spen*:*Tsix* defective cell lines (*Spen*$^{-/-}$:*Tsix*-Stop) display normal *Xist* levels upon differentiation, indicating that in the absence of SPEN, *Tsix,* which acts a as brake on *Xist* transcription, cannot be silenced. Furthermore, our results show that SPEN plays a role in *Xist* stability, however, we also observe that *Xist* levels in differentiating *Spen*$^{-/-}$:*Tsix*-Stop cells are comparable to Wt cells, suggesting that *Xist* stability might not be compromised. This difference may be attributed to the massive *Xist* overexpression in *Xist*-inducible systems, where the excess of *Xist* RNA molecules, in the absence of SPEN, may affect overall *Xist* RNA stability readings.

During initiation of XCI, SPEN helps remodel the chromatin environment of the *Xist*-*Tsix* locus. SPEN and HDAC3 are present at the *Tsix* regulatory region and H3K27ac levels decrease upon *Xist* induction. Accordingly, we propose that during XCI initiation, SPEN binds *Xist* nascent transcripts[25], recruits and/or activates HDAC3[16,30], which removes histone acetylation marks, weakening *Tsix* promoter activity and facilitating *Xist* expression (Fig. 6).

The mechanism driving the symmetry breaking event leading to monoallelic upregulation of *Xist* has been a focus of many studies. Several studies showed that the *Xic* and more specifically the *Xist*-*Tsix* master switch are tightly regulated in a deterministic process, involving monoallelic down-regulation of *Tsix* or X-pairing mechanisms[40–42]. Our studies, and studies of others, indicate that a gene regulatory network

276 composed of trans-acting activators and inhibitors of XCI, involving positive and negative feedback loops,
277 instruct the *cis*-regulatory landscape of the *Xic* to direct monoallelic upregulation of *Xist*[12,43–45]. Our
278 present work reveals that *Tsix* silencing not only involves downregulation of inhibitors of XCI, including
279 pluripotency factors OCT4, SOX2, KLF4, c-MYC and REX1[46–48], but also involves *Xist*-mediated recruitment
280 of SPEN. This can happen on either allele, but asynchronous *Xist* transcription bursts will facilitate *Xist*-
281 mediated monoallelic silencing of *Tsix* through SPEN, resulting in further upregulation of *Xist* and
282 concomitant silencing of the XCI activator *Rnf12*, providing a negative feedback loop to prevent
283 upregulation of *Xist* on the future active X chromosome.

284

285 ## Acknowledgements

286 We thank Anniek Meesters and Esther Sleddens-Linkels for technical help; Kristian Helin for providing the
287 EZH2 antibody; Martine M. Jaegle for sharing with us a floxed-puromycin resistance cassette and all the
288 members of the Erasmus MC Developmental Biology department for useful discussions.

289 T.R.F. is supported by an Erasmus MC grant (Mrace). B.F.T., H.M.B., E.T. and J.G. are supported by the
290 Oncode Institute. T.B. is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research
291 Foundation) - TRR81- A12 and BO 1639/9-1, the Behring-Röntgen foundation and Excellence Cluster for
292 Cardio Pulmonary System (ECCPS) in Giessen. B.D.G. is supported by a research grant of the University
293 Medical Center Giessen and Marburg (UKGM) and by a Prize of the Justus Liebig University Giessen".

294

295 ## Author contributions

296 T.R.F. and J.G. conceived the project and designed the experiments. T.R.F. and E.T. performed most of the
297 experimental work and data analysis. H.M.B. performed the ChIP-seq experiments. B.F.T. performed all
298 the bioinformatic analysis. C.G.P. helped setup the high molecular weight western blot analysis. S.M.
299 setup the allele-specific RT-qPCR analysis. B.D.G., F.D., T.B. and E.H. provided valuable resources. H.M.B.
300 and J.G. supervised the work. T.R.F. and J.G. wrote the manuscript with input from all the authors. H.M.B.
301 helped review and edit the manuscript. J.W.M.M. and J.G. were responsible for the founding acquisition.
302 RNA-seq and ChIP-seq datasets were generated in the Erasmus MC Center for Biomics led by W.v.IJ.

303

References

1. Lyon, M. F. Gene Action in the X-chromosome of the Mouse (Mus musculus L.). *Nature* **190**, 372–373 (1961).

2. Borsani, G. *et al.* Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **354**, 56–58 (1991).

3. Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329–31 (1991).

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

5. Lee, J. T., Davidow, L. S. & Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat. Genet.* **21**, 400–404 (1999).

6. Stavropoulos, N., Lu, N. & Lee, J. T. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10232–10237 (2001).

7. Luikenhuis, S., Wutz, A. & Jaenisch, R. Antisense Transcription through the Xist Locus Mediates Tsix Function in Embryonic Stem Cells. *Mol. Cell. Biol.* **21**, 8512–8520 (2001).

8. Shibata, S. & Lee, J. T. Tsix Transcription- versus RNA-Based Mechanisms in Xist Repression and Epigenetic Choice. *Curr. Biol.* **14**, 1747–1754 (2004).

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

10. Navarro, P., Pichard, S., Ciaudo, C., Avner, P. & Rougeulle, C. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: Implications for X-chromosome inactivation. *Genes Dev.* **19**, 1474–1484 (2005).

11. Sun, B. K., Deaton, A. M. & Lee, J. T. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol. Cell* **21**, 617–628 (2006).

12. Mutzel, V. *et al.* A symmetric toggle switch explains the onset of random X inactivation in different mammals. *Nat. Struct. Mol. Biol.* (2019). doi:10.1038/s41594-019-0214-1

13. Mutzel, V. & Schulz, E. G. Dosage Sensing, Threshold Responses, and Epigenetic Memory: A Systems Biology Perspective on Random X-Chromosome Inactivation. *BioEssays* **42**, 1–14 (2020).

14. Loda, A. & Heard, E. Xist RNA in action: Past, present, and future. *PLoS Genet.* **15**, 1–17 (2019).

15. Brockdorff, N. & Turner, B. M. Dosage compensation in mammals. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).

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

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

18. Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science (80-.).* **316**, (2015).

19. Monfort, A. *et al.* Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554–561 (2015).

20. Moindrot, B. *et al.* A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Rep.* **12**, 562–572 (2015).

21. Ariyoshi, M. & Schwabe, J. W. R. A conserved structural motif reveals the essential transcriptional repression function of spen proteins and their role in developmental signaling. *Genes Dev.* **17**, 1909–1920 (2003).

22. Oswald, F. *et al.* A phospho-dependent mechanism involving NCoR and KMT2D controls a permissive chromatin state at Notch target genes. *Nucleic Acids Res.* **44**, 4703–4720 (2016).

23. Shi, Y. *et al.* Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.* **15**, 1140–1151 (2001).

24. Oswald, F. *et al.* SHARP is a novel component of the Notch/RBP-Jκ signalling pathway. *EMBO J.* **21**, 5417–5426 (2002).

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

26. Loda, A. *et al.* Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. *Nat. Commun.* **8**, (2017).

27. Nesterova, T. B. *et al.* Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat. Commun.* **10**, 1–15 (2019).

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

29. Ogawa, Y. & Lee, J. T. Xite , X-Inactivation Intergenic Transcription Elements that Regulate the Probability of Choice. **11**, 731–743 (2003).

30. Żylicz, J. J. *et al.* The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* **176**, 182-197.e23 (2019).

31. Jonkers, I. *et al.* RNF12 Is an X-Encoded Dose-Dependent Activator of X Chromosome Inactivation. *Cell* **139**, 999–1011 (2009).

32. Heard, E. *et al.* Methylation of histone H3 at Lys-9 Is an early mark on the X chromosome during X inactivation. *Cell* **107**, 727–738 (2001).

33. Rougeulle, C. *et al.* Differential Histone H3 Lys-9 and Lys-27 Methylation Profiles on the X Chromosome. *Mol. Cell. Biol.* **24**, 5475–5484 (2004).

34. Loos, F. *et al.* Xist and Tsix Transcription Dynamics Is Regulated by the X-to-Autosome Ratio and Semistable Transcriptional States. *Mol. Cell. Biol.* **36**, 2656–2667 (2016).

35. Barakat, T. S., Rentmeester, E., Sleutels, F., Grootegoed, J. A. & Gribnau, J. Precise BAC targeting of genetically polymorphic mouse ES cells. *Nucleic Acids Res.* **39**, 6–13 (2011).

36. Carter, A. C. *et al.* Spen links rna-mediated endogenous retrovirus silencing and x chromosome inactivation. *Elife* **9**, 1–58 (2020).

37. Kuroda, K. *et al.* Regulation of marginal zone B cell development by MINT, a suppressor of Notch/RBP-J signaling pathway. *Immunity* **18**, 301–312 (2003).

38. Hoki, Y. *et al.* A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* **136**, 139–146 (2009).

39. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science (80-.).* **341**, 1237973 (2013).

40. Lee, J. T. & Lu, N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* **99**, 47–57 (1999).

41. Xu, N., Tsai, C. L. & Lee, J. T. Transient homologous chromosome pairing marks the onset of X inactivation. *Science (80-.).* **311**, 1149–1152 (2006).

42. Aeby, E. *et al.* Decapping enzyme 1A breaks X-chromosome symmetry by controlling Tsix elongation and RNA turnover. *Nat. Cell Biol.* **22**, 1116–1129 (2020).

43. Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X Inactivation Counting and Choice Is a Stochastic Process: Evidence for Involvement of an X-Linked Activator. *Cell* **132**, 410–421 (2008).

44. Gayen, S. *et al.* Article A Primary Role for the Tsix lncRNA in Maintaining Article A Primary Role for the Tsix lncRNA in Maintaining Random X-Chromosome Inactivation. *CellReports* **11**, 1251–1265 (2015).

45. Galupa, R. *et al.* A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary. *Mol. Cell* **77**, 352-367.e8 (2020).

46. Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. & Lee, J. T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128–132 (2009).

47. Navarro, P. *et al.* Molecular coupling of Tsix regulation and pluripotency. *Nature* **468**, 457–460 (2010).

48. Gontan, C. *et al.* RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**, 386–390 (2012).

49. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1100 (2013).

50. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).

51. Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Research* **5**, 1479 (2016).

52. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

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

54. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

55. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

56. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).

57. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic data sets. *Bioinformatics* 1–2 (2020). doi:10.1093/bioinformatics/btaa692

58. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

Main figure legends

Fig. 1 Impaired *Xist* upregulation in *Spen*$^{-/-}$ ESCs upon monolayer differentiation.

a Overview of the endogenous doxycycline-inducible *Xist* hybrid system used in this study. Addition of doxycycline leads to inactivation of the Cast X chromosome, while no addition of the drug leads to inactivation of the 129 X chromosome.

b Experimental design to recapitulate XCI. Cells were treated or untreated with doxycycline for 4 days previous to monolayer differentiation.

c Allelic ratio ((Xi)/(Xi+Xa)) of individual genes along the X chromosome of Wt **(top-left)** and *Spen*$^{-/-}$ **(bottom-left)** undifferentiated (day 0) ESCs with and without doxycycline. Allelic ratio of individual genes along the X chromosome of Wt and *Spen*$^{-/-}$ lines at day 3 of differentiation without doxycycline **(top-right)** and day 7 with doxycycline **(bottom-right)**. Only the genes (n) with sufficient reads in both conditions are shown. Xa = active X chromosome.

d Percentage of *Rnf12* allelic expression at different time points of monolayer differentiation of two independent Wt and *Spen*$^{-/-}$ ESC lines treated with doxycycline, determined by RT-qPCR. Relative *Rnf12*

allelic (129 and Cast) expression was normalized to *Rnf12* total expression and averaged ± SD, n=2 biological replicates.

e Relative allele-specific *Xist* expression of two independent Wt and *Spen*$^{-/-}$ ESC lines at different time points of monolayer differentiation treated with doxycycline. Average expression ± Standard Deviation (SD), n=2 biological replicates.

f Same as displayed in **(d)** without doxycycline.

g Same as displayed in **(e)** without doxycycline.

h *Xist* RNA FISH (red) of Wt and *Spen*$^{-/-}$ ESC lines at day 3 and 5 of differentiation. Both *Tsix* pinpoints and *Xist* clouds are visible. DNA is stained with DAPI (blue). Scale bar: 10 µm.

i Quantification of **(h)**, displaying the average percentage of nuclei with *Xist* clouds. Average percentage ± SD, n=2 biological replicates, 150-500 nuclei quantified per replicate.

j Relative allele-specific *Tsix* expression of Wt and Spen$^{-/-}$ ESCs at different time points of monolayer differentiation, determined by RT-qPCR. Average expression ± SD, n=2 biological replicates. Statistical analysis was done using a two-tailed Student's t-test comparing four independent replicates per condition (two biological replicates of two different clones). (*) p-value <0.05, (**) p-value <0.01.

Fig. 2 *Spen* cDNA expression in *Spen*$^{-/-}$ ESCs rescues *Xist* expression.

a Relative total *Xist* expression determined by RT-qPCR at day 0, 3 and 5 of monolayer differentiation of Wt, *Spen*$^{-/-}$ and three independent rescue clones (Clone A, B and C). Each individual clone was differentiated twice as biological duplicates. Average expression ± SD, n=2 biological replicates.

b RNA FISH of *Xist* (red) at day 0 and 3 of monolayer differentiation of Wt, *Spen*$^{-/-}$ and *Spen* rescue clones. DNA stained with DAPI (blue). Scale bar: 10 µm.

c Quantification of **(b)**, showing the percentage of nuclei with *Xist* clouds in each condition. Average percentage ± SD, n=2 biological replicates, 150-400 nuclei quantified per replicate.

Fig. 3 SPEN co-localizes with the Xi chromosome very early during monolayer differentiation. Two different states are detectable: early and late SPEN accumulation on the Xi.

a SPEN-GFP IF staining (αGFP, green) at different time points of monolayer differentiation (day 0, 1, 2, 3, 5, 7, 10 and 12). White arrowheads mark early SPEN accumulation to the Xi and asterisks late SPEN accumulation. DNA is stained with DAPI (blue). Scale bar: 10 µm.

b Quantification of **(a)**, depicting the percentage of early and late SPEN accumulation per nuclei at different time points of monolayer differentiation. Average percentage ± SD, n=2 independent clones, 200-500 nuclei quantified per replicate.

c Examples of early **(left)** and late **(right)** SPEN accumulation (αGFP, green) in differentiating cells. White arrowheads mark early SPEN accumulation to the Xi and asterisks late SPEN accumulation. DNA is stained with DAPI (blue). Scale bar: 10 µm.

d IF (αGFP, green) combined with *Xist* RNA FISH (red) of SPEN-GFP lines at day 0, 2 and 5 of monolayer differentiation. DNA is stained with DAPI (blue). Scale bar: 10 µm.

e Percentage of nuclei with a *Xist* cloud that also present SPEN accumulation (total). Average percentage ± SD, n=2 independent clones, 200-500 nuclei quantified per replicate. Calculated from the same data as for **Supplementary Fig. 4d**.

f Percentage of nuclei with SPEN accumulation (total, early and late) that also present EZH2 accumulation or H3K27me3 pinpoints. Average percentage ± SD, n=2 independent clones, 200-500 nuclei quantified per replicate. Calculated from the same data as for **Supplementary Fig. 4g**.

Fig. 4 SPEN plays a role in *Xist* RNA stability.

a Relative total *Xist* expression in undifferentiated Wt and *Spen*$^{-/-}$ ESCs upon *Xist* induction with doxycycline for 4 days. Average expression ± SD, n=2 biological replicates.

b *Xist* RNA FISH (red) of undifferentiated Wt and *Spen*$^{-/-}$ ESCs treated with doxycycline to induce *Xist* expression. DNA is stained with DAPI (blue). Scale bar: 10 µm.

c Quantification of **(b)**, displaying the percentage of nuclei with *Xist* clouds. Average percentage ± SD, n=2 biological replicates, 100-150 nuclei quantified per replicate.

d Percentage of *Xist* RNA remaining at different time points of actinomycin D treatment (t = 0, 2, 4, 6 and 10 hours) in undifferentiated Wt and *Spen*$^{-/-}$ ESCs treated with doxycycline. Normalized relative expression to t=0 ± SD, n=2 biological replicates.

Fig. 5 *Tsix* regulatory region shows SPEN, HDAC3 and H3K27ac enrichment. SPEN is required to silence *Tsix*, to allow *Xist* upregulation upon XCI initiation.

a Genome browser tracks showing allele-specific SPEN, HDAC3, H3K27ac and H3K27me3 binding at the *Xist*-*Tsix* locus. The top part (dark) of each track represents the Xi and the bottom (light) the Xa. SPEN CUT&RUN **(blue, top)** profile in *Xist* inducible undifferentiated ESCs (day 0) untreated or treated with doxycycline (24h)[25]. HDAC3 **(green, middle-top)** and H3K27ac ChIP-seq **(yellow, middle-bottom)** in *Xist* inducible undifferentiated ESC (day 0) untreated or treated with doxycycline (24h)[30]. H3K27me3 ChIP-seq **(red, bottom)** in Wt and *Spen*$^{-/-}$ ESCs at day 0 and 3 of monolayer differentiation. The light-blue rectangle highlights the *Tsix* regulatory region, including the *Tsix* minor and major promoters, and *Xite*. The light-gray rectangle highlights the H3K27me3 hotspot region. The dashed rectangle indicates the genomic region in **(b)**.

b Zoom-in view of the *Tsix* promoter region, showing the SPEN CUT&RUN profile in undifferentiated ESCs (day 0) untreated or treated (4h, 8h and 24h) with doxycycline.

c Schematic overview of the *Tsix*-defective ESC lines used to study the role of SPEN in *Tsix* silencing and *Xist* upregulation, namely, Wt (F1:129/Cast) **(top)**, *Tsix*-Stop **(middle)** and *Tsix*-Cherry **(bottom)**. The *Tsix*-Stop line is defective for *Tsix* in the 129 allele; the *Tsix*-Cherry line is defective in the Cast allele. Homozygous deletion of *Spen* was performed in the three lines.

d Relative allele-specific *Xist* expression in Wt and *Spen*$^{-/-}$ F1:129/Cast ESC lines at different time points of monolayer differentiation. Each individual clone was differentiated twice in biological duplicates. Average expression ± SD, n=2 biological replicates.

e Same as in **(d)** for Wt and *Spen*$^{-/-}$ *Tsix*-Stop and *Tsix*-Cherry ESC lines.

f *Xist* RNA FISH (red) at day 3 of monolayer differentiation in F1:129/Cast and *Tsix*-Stop Wt and *Spen*$^{-/-}$ ESC lines. DNA is stained with DAPI (blue). Scale bar: 10 µm.

g Quantification of **(f)**, showing the percentage of nuclei with *Xist* clouds. Average percentage ± SD, n=2 biological replicates, 150-500 nuclei quantified per replicate.

Fig. 6 Model of the role of SPEN in XCI.

In the undifferentiated state, *Tsix* transcription represses *Xist* and *Tsix* RNA levels are high in ESCs. Upon differentiation, at the start of XCI, *Xist* transcription increases, the nascent *Xist* transcripts recruit SPEN that is necessary to silence the *Tsix* promoter via removal of active H3K27ac marks from its promoter. Consequently, the silencing of *Tsix* allows *Xist* upregulation, accumulation and spreading. The spreading of SPEN-*Xist* along the X chromosome allows the silencing of X-linked genes. SPEN also plays a role in *Xist* stability. This model, in addition to previous evidence, proposes that SPEN is not only key in X-linked gene silencing, but also in initiation of XCI. In the figure, the blue circles depict SPEN and the grey circles RNA Polymerase II.

Supplementary figure legends

Supplementary Fig. 1 Generation and characterization of *Spen*$^{+/-}$ and *Spen*$^{-/-}$ ESC lines. RNA-seq analysis of Wt and *Spen*$^{-/-}$ ESCs. | Related to **Fig. 1.**

a Targeting strategy used to generate *Spen* knockout ESC lines using the CRISPR/Cas9 system. Two single guide RNAs (sgRNA) targeting the 5' and 3' region of the *Spen* ORF were used to integrate a Puromycin resistance (PuroR) cassette.

b PCR genotyping to identify optimal heterozygote ($^{+/-}$) and homozygote ($^{-/-}$) *Spen* knockout ESC clones. Specific 5' and 3' integration of the PuroR cassette in the *Spen* locus. Restriction Fragment Length Polymorphism (RFLP) analysis on *Spen* exon 13 to determine the absence of *Spen* ORF in one or both alleles. Verification of the presence of two X chromosomes per ESC line. Genotyping of the *Xist* endogenous doxycycline-inducible promoter. M = DNA ladder. (*) = unspecific band.

c SPEN western blot of two independent Wt, *Spen*$^{+/-}$ and *Spen*$^{-/-}$ ESC clones. VCP was used as a loading control. M = High molecular weight protein ladder.

d Overview of the RNA-seq libraries generated in this study, summarizing the experimental conditions performed in Wt and *Spen*$^{-/-}$ ESCs at different time points of monolayer differentiation (day 0, 3 and 7),

546 treated with or without doxycycline (X in the panel). Each condition includes two biological replicates,
547 adding up to a total of 16 RNA-seq libraries.

548 **e** Violin plots depicting the distribution of the allelic ratios ((Xi)/(Xi+Xa)) of X-linked genes in Wt and Spen^{-}
549 $^{/-}$ ESCs, untreated or treated with doxycycline at days 0, 3 and 7 of differentiation. This figure summarizes
550 data in **Fig. 1c**. The box plots inside the violin plots show the median and interquartile range.

551 **f** Scatter plot showing the allelic ratio (Xi/(Xi+Xa)) of X-linked genes in Wt ESCs treated with doxycycline
552 at day 0 (x-axis) and day 7 (y-axis) of differentiation. Highlighted in orange are the lowly silenced genes in
553 *Spen*$^{-/-}$ ESC previously identified[27]. Only the genes (n) with sufficient reads in both conditions are shown.
554 Dashed lines = allelic ratio of 0.5; Diagonal solid line = equal ratio in both conditions.

555 **g** Same as in **(f)**, for *Spen*$^{-/-}$ ESCs.

556 **h** *Rex1*, *Nanog* and *Gata6* relative expression of Wt and *Spen*$^{-/-}$ ESC lines upon monolayer differentiation
557 not treated with doxycycline. Average expression ± SD, n=2 biological replicates.

558

559 **Supplementary Fig. 2** *Spen*$^{+/-}$ ESCs lines are able to upregulate *Xist*, but show less X-linked gene silencing,
560 compared to Wt ESCs upon monolayer differentiation. | Related to **Fig. 1.**

561 **a** *Rnf12* percentage of allelic expression of Wt and *Spen*$^{+/-}$ lines treated with doxycycline at day 0 and 7 of
562 differentiation. Wt samples are the same as in **Fig. 1d**. Relative *Rnf12* allelic (129 and Cast) expression was
563 normalized to *Rnf12* total expression and averaged ± SD, n=2 biological replicates.

564 **b** Relative allele-specific *Xist* expression of two Wt and *Spen*$^{+/-}$ ESC clones treated with doxycycline at day
565 0 and 7 of monolayer differentiation, determined by RT-qPCR. Wt samples are the same as in **Fig. 1e**.
566 Average expression ± SD, n=2 biological replicates.

567 **c** Same as displayed in **(a)** without doxycycline. Wt samples are the same as in **Fig. 1f**.

568 **d** Same as displayed in **(b)** without doxycycline. Wt samples are the same as in **Fig. 1g**.

569

Supplementary Fig. 3 *Spen* cDNA rescue ESC lines characterization. |Related to **Fig. 2.**

a Schematic overview of the generation of *Spen*$^{-/-}$ ESC lines able to stably express the *Spen* cDNA from the *ROSA26* locus. These lines were made using a sgRNA targeting the *ROSA26* locus and a vector coding for the *Spen* cDNA and a hygromycin resistance cassette (HygroR)[25].

b Strategy to identify correct *Spen* cDNA rescue clones by PCR on gDNA. Specific 3' *Spen* cDNA integration in the *ROSA26* locus. Primer on the Flag-tag present in the *5' Spen* cDNA end to identify those clones containing the *Spen* cDNA vector. Specific 5' and 3' integration of the PuroR cassette in the *Spen* locus to identify the *Spen*$^{-/-}$ line. RFLP analysis on *Spen* exon 13 to determine the presence of *Spen* in the *ROSA26* locus and its absence in the *Spen*$^{-/-}$ cells; the *Spen* rescue cDNA sequence is C57BL/6 (B6). Verification of the presence of two X chromosomes per ESC line, making use of a length polymorphism on the X chromosome. M = DNA ladder.

c Relative *Spen* expression in Wt, *Spen*$^{-/-}$ and *Spen* cDNA rescue ESC lines (Clone A, B and C) at day 0, 3 and 5 of monolayer differentiation. Each individual clone was differentiated twice in biological duplicates. Average expression ± SD, n=2 biological replicates.

d Percentage of *Rnf12* allelic expression at day 0, 3 and 5 of monolayer differentiation of Wt, *Spen*$^{-/-}$ and three *Spen* cDNA rescue clones (Clone A, B and C), determined by RT-qPCR. Relative *Rnf12* allelic (129 and Cast) expression was normalized to *Rnf12* total expression and averaged ± SD, n=2 biological replicates.

e Relative allele-specific *Tsix* expression at day 0 and 5 of monolayer differentiation of Wt, *Spen*$^{-/-}$ and three *Spen* cDNA rescue clones (Clone A, B and C), determined by RT-qPCR. Average expression ± SD, n=2 biological replicates.

Supplementary Fig. 4 Generation and characterization of a SPEN-GFP C-terminal tag line. |Related to **Fig. 3.**

a eGFP knock-in strategy in *Spen* exon 16 using the CRISPR/Cas9 system.

b Genotyping of the *Spen*-GFP knock-in clones by PCR on gDNA to determine the specific integration of the 5' and 3' HA and the presence of two different X chromosomes making use of a length polymorphism. M = DNA ladder.

c Flow cytometry histograms comparing the GFP fluorescence in the parental line (gray) with a negative clone **(left)** and two independent SPEN-GFP-tagged ESC clones **(middle and right)**.

d Percentage of nuclei with *Xist* clouds and SPEN accumulation in Wt (untagged) and *Spen*-GFP ESCs. Quantification of **Fig. 3d.** Average percentage ± SD, n=2 independent clones, 200-500 nuclei quantified per replicate.

e Double IF staining of SPEN-GFP (αGFP, green) and EZH2 (red) at day 0, 2 and 5 of differentiation of a *Spen*-GFP ESC line. Early accumulation is indicated with a white arrowhead, late accumulations with an asterisk. DNA is stained with DAPI (blue). Scale bar: 10 µm.

f Double IF staining of SPEN-GFP (αGFP, green) and H3K27me3 (red) at day 0, 2 and 5 of differentiation of a *Spen*-GFP ESC line. Early accumulation is indicated with a white arrowhead, late accumulations with an asterisk. DNA is stained with DAPI (blue). Scale bar: 10 µm.

g Quantification of **(e and f)**, showing the percentage of nuclei with early and late SPEN accumulations (green) and the percentage of nuclei with EZH2 (yellow) and H3K27me3 (light red) co-localizing with early or late SPEN accumulations. Average percentage ± SD, n=2 independent clones, 200-500 nuclei quantified per replicate.

Supplementary Fig. 5 Chromatin features of X-linked genes. |Related to **Fig. 5.**

a-b Genome browser tracks showing the allele-specific SPEN, HDAC3, H3K27ac and H3K27me3 binding for two X-linked genes: **a** *Rnf12* and **b** *Pgk1*. The top part (dark colour) of each track represents the Xi and the bottom (light colour) the Xa. SPEN CUT&RUN **(blue, top)** profile in *Xist* inducible undifferentiated ESCs (day 0) untreated or treated with doxycycline (24h)[25]. HDAC3 **(green, top-middle)** and H3K27ac ChIP-seq **(yellow, bottom-middle)** in *Xist* inducible undifferentiated ESCs (day 0) untreated or treated with doxycycline (24h)[30]. H3K27me3 ChIP-seq **(red, bottom)** in Wt and *Spen*$^{-/-}$ ESCs at day 0 and 3 of monolayer differentiation. The light-blue square highlights the promoter region of *Rnf12* and *Pgk1*.

Supplementary Fig. 6 Generation and characterization of *Spen*$^{-/-}$ in *Tsix* defective ESC lines. | Related to **Fig. 5.**

a Genotyping strategy to identify optimal *Spen*$^{-/-}$ clones in F1:129/Cast, *Tsix*-Stop and *Tsix*-Cherry ESC lines, done by PCR on gDNA. Specific 5' and 3' integrations of the PuroR cassette in the *Spen* locus to identify *Spen* knockout lines. RFLP analysis on *Spen* exon 13 to determine the absence of the *Spen* 129 and/or Cast allele. Verification of the presence of two X chromosomes per ESC line, making use of a length polymorphism on the X chromosome. *Tsix*-Stop line genotyping, using a primer pair across the triple poly(A) signal blocking *Tsix* transcription. *Tsix*-Cherry line genotyping by determining the loss of the Cast band of a specific length polymorphism, indicating proper mCherry integration downstream of the *Tsix* promoter. M = DNA ladder.

b SPEN western blot of Wt and *Spen*$^{-/-}$ F1:129/Cast, *Tsix*-Stop and *Tsix*-Cherry ESC lines. VCP was used as a loading control. M = High molecular weight protein ladder.

c Relative allele-specific *Tsix* expression at day 0 and 5 of monolayer differentiation of Wt and *Spen*$^{-/-}$ F1:129/Cast, *Tsix*-Stop and *Tsix*-Cherry ESC lines, determined by RT-qPCR. Each individual clone was differentiated twice in biological duplicates. Average expression ± SD, n=2 biological replicates.

d Percentage of *Rnf12* allelic expression at day 0 and 5 of monolayer differentiation of Wt and *Spen*$^{-/-}$ F1:129/Cast, *Tsix*-Stop and *Tsix*-Cherry ESC lines, determined by RT-qPCR. Relative *Rnf12* allelic (129 and Cast) expression was normalized to *Rnf12* total expression and averaged ± SD, n=2 biological replicates.

Methods

Cell culture

Mouse ESCs were grown on male feeder cells and medium containing DMEM (Gibco), 15% Foetal Calf Serum (FCS), 0.1 mM non-essential amino acids (NEAA), 100 U mL^{-1} penicillin, 100 ug mL^{-1} streptomycin, 0.1 mM 2-mercaptoethanol (Gibco) and 1000 U mL^{-1} LIF. Previous to monolayer differentiation cells were plated in non-gelatinized plates to eliminate feeder cells. Then, plated at specific densities per time point in differentiation medium composed of IMDM-glutamax (Gibco), 15% FCS, 0.1 mM NEAA, 100 U mL^{-1} penicillin, 100 µg mL^{-1} streptomycin, 37.8 µL L^{-1} monothioglicerol and 50 mg mL^{-1} ascorbic acid. When appropriate, medium was supplemented with 2 µg mL^{-1} doxycycline.

Gene editing using the CRISPR/Cas9 technology

Different female F1 2–1 hybrid (129/Sv-Cast/Ei) ESC lines with different genetic modifications to interrogate various aspects of XCI were targeted in this study (Supplementary Table 2). To generate *Spen* heterozygous and homozygous knockout clones two single-guide RNAs (sgRNA) targeting the 5' (5'-AGTGCGCTTCGTCACTGCAC-3') and 3' (5'-TCCTCCCGCCCCGACGCGGA-3') region of the *Spen* ORF were cloned in the Cas9-GFP pX458 vector (Addgene plasmid #48138). Compatible 5' and 3' Homology arms (HA) of approximately 500 bp were amplified by PCR from mouse gDNA and cloned in the pCR-Blunt-II-TOPO vector with a NdeI restriction site in-between the 5' and 3' HA. This site was used to insert a Puromycin resistance (PuroR) cassette flanked by loxP sites. The *ROSA26* locus was targeted using the pX458 vector coding for a sgRNA (5'-CGCCCATCTTCTAGAAAGAC-3') compatible with the pFD46 expression vector [25], coding for the *Spen* cDNA and a hygromycin resistance cassette. To create an endogenous Spen C-terminal enhanced GFP (eGFP) knock-in, a sgRNA targeting the 3' end of *Spen* ORF (5'-GATTGTCATTGCCTCGGTG-3') was cloned in the Cas9-PuroR pX459 vector (Addgene plasmid #62988). The donor template was made using a gblock from Integrated DNA Technologies coding for compatible 5' and 3' HA of 600 bp with a NheI and AscI restrictions sites in-between the 5' and 3' HA, which were used to insert an eGFP in frame with the *Spen* coding sequence. The appropriate plasmid combinations were transfected into ESCs using lipofectamine 2000 and plated at low density to obtain single colonies, when appropriate medium was supplemented with 1 µg mL^{-1} puromycin (Sigma-Aldrich, P8833) for 48-72 hours or 250 µg mL^{-1} hygromycin B (Invitrogen, 10687010) for 7 days. Colonies were screened by PCR for correct integration of the desired construct (Supplementary Table 3). Positive clones were further characterized by western blot, RT-qPCR and/or FACS, and the presence of 2 X chromosomes and correct karyotype was also assessed.

673 Protein extraction and western blot

674 To prepare nuclear extracts all procedures were done at 4°C and buffers supplemented with 1x protease
675 inhibitors (Roche, 4693132001), 15 µM MG-132 (Sigma-Aldrich, C2211) and 0.5 mM DTT. Cells were
676 harvested by scraping in cold PBS, collected and centrifuged (1500 rpm, 5 min, 4°C). The pellet was
677 incubated in 5x times the pellet volume of Buffer A (10 mM Hepes pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl) for
678 10 min, vortexed (30 sec) and centrifuged (3000 rpm, 5 min, 4°C). Then, the pellet was resuspended in
679 1,5x times Buffer C (20 mM Hepes pH 7.6, 25% glycerol, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA) and
680 rotated (30 min, 4°C). The solution was centrifuged (14000 rpm, 10 min, 4°C) and the supernatant
681 collected as nuclear extract. The protein concentration was measured using Nanodrop and all samples
682 diluted to the same concentration using Buffer C. NuPAGE™ LDS Sample Buffer (4X) (Thermo Scientific,
683 NP0007) containing 5% b-mercaptoethanol were added to nuclear extracts, and boiled (95°C, 5 min). For
684 western blot analysis, the NuPAGE™ 3 to 8% Tris-Acetate gels (Invitrogen, EA03755) were used with Tris-
685 Acetate SDS Running Buffer (pH 8.24) and the HiMark Pre-Stained Protein Standard (Invitrogen, LC5699).
686 Wet transfer on a PVDF membrane was done overnight at 4 mA, with the NuPAGE Transfer Buffer
687 (Invitrogen #NP00061), containing 10% methanol and 0.01% SDS. After blocking, the membrane was
688 incubated with the appropriate antibodies: SPEN antibody (Abcam, ab290, 1:2000) and VCP antibody
689 (Abcam, ab11433, 1:20000).

690 RT-qPCR

691 Total RNA was isolated from cell pellets using the ReliaPrep RNA Cell Miniprep System (Promega, Z6012)
692 and reversed transcribed using Superscript III (Invitrogen, 18080093) and random hexamers (Invitrogen,
693 N8080127), following the manufacturer's instructions. All RT-qPCRs were done using the GoTaq qPCR
694 Master Mix (Promega, A6002) in a CFX384 real-time PCR detection system (Bio-Rad). Hist2h2aa1 was used
695 as a normalization control, except in the RNA stability assay were β-actin was used. All expression primers
696 are listed in Supplementary Table 4. The optimal allele-specific primer pair concentration to amplify the
697 129 and Cast allele at the same efficiency was optimized using pure 129, Cast and 129-Cast gDNA.

698 RNA stability assay

699 Wt and *Spen*$^{-/-}$ ESC lines with a doxycycline-responsive *Xist* promoter were cultured in medium
700 supplemented with doxycycline to induce *Xist* expression for 4 days. While in doxycycline treatment, 5 µg
701 mL^{-1} actinomycin D (Sigma-Aldrich, A1410) was added for different time (t= 0, 2, 4, 6 and 10 hours). Before
702 collection, ESCs were plated in non-gelatinized plates to remove the feeders and harvested for RNA
703 isolation, cDNA synthesis and RT-qPCR. Total *Xist* levels were determined and normalized to β-actin and

the percentage of remaining *Xist* RNA was calculated by dividing *Xist* expression levels at the different times of collection relative to t=0. Using linear regression analysis the RNA decay rate constant (k_{decay}) was calculated from the slope of the curve that best fitted our data and the RNA half-life ($t_{1/2}$) obtained with the following formula $t_{1/2} = \ln 2/(k_{decay})$.

Immunofluorescence (IF)

Undifferentiated ESCs were attached to slides using a cytospin, while differentiating cells were grown on coverslips coated with gelatin or laminin. Cells were fixed with 4% paraformaldehyde (PFA) in PBS (10 min, RT), permeabilized with 0.4% Triton-X100 and 5% goat serum in PBS (15 min, on ice) and blocked with 10% goat serum in PBST, composed of 0.05% Tween-20 in PBS (30 min, RT). Primary antibodies incubation was done in blocking buffer (2h, RT), after three washings with PBST the slides were incubated with the secondary antibody (1h, RT). After three washings, the second washing containing DAPI, the slides were mounted with ProLong Gold antifade mounting medium (Invitrogen, P36930). The used primary antibodies were the following: αGFP (Abcam, ab290, 1:500), H3K27me2/3 (Active motif, 39535, 1:100) and EZH2(BD43 clone, kindly provided by Dr. Kristian Helin, 1:100). Images were taken using a ZEISS Axio Imager M2 including digital microscopy camera AxioCam 503 and analysed with ImageJ.

IF combined with RNA Fluorescent in Situ Hybridization (FISH)

The *Xist* FISH probe was made from 2 µg of a 5.5 kb DNA fragment of mouse *Xist* comprising exons 3 to 7. Fluorescent labelling was done with dUTP SpectrumRed using the nick translation kit (Abbott, 07J00-001) overnight at 16°C. The probe was purified with ProbeQuant G-50 Micro Columns (GE, GE28-9034-08), combined with 100 µg mouse tRNA, 20 µg mouse Cot-1 DNA and 100 µg salmon sperm DNA and precipitated with 2 M NaAc (pH 5.6) and 100% EtOH. The pellet was resuspended in 50 µL hybridization mix (50% formamide, 10% dextran sulphate in 2xSSC) and stored at -20°C. Before use, 20 uL of probe was pre-hybridized (10 min, 75°C) with 0.5 µg mouse Cot-1 DNA supplemented with 10 mM vanadyl ribonucleoside complex (VRC) (NEB, S1402S) and 0.2 U μL^{-1} RNAseOUT (Invitrogen, 10777019).

Cells on coverslips or slides were fixed with 4% PFA in PBS (10 min, RT), permebilized with 0.5% Triton-X100 in PBS and blocked in TS-BSA buffer (0.1 M Tris-HCl (pH 7.5), 0.15 M NaCl, 2 mg mL^{-1} BSA (Jena Bioscience, BU-102) in H_2O). The αGFP (Abcam, ab290, 1:500) primary antibody was incubated in blocking buffer (30 min, 37°C), washed in PBS (3x) and then incubated with the secondary antibody (30 min, 37°C). All solutions were supplemented with 10 µM VRC and 0.2 U μL^{-1} RNAseOUT. Slides were washed in PBS (3x), post-fixed with 4% PFA (10 min, RT) and washed again in PBS (x3). Dehydration was done with

734 increasing ethanol concentration (70%, 90% and 100%). The pre-hybridized probe was added on top of
735 the coverslips or slides (20h, 37°C, humid chamber). Various washings were done with 50% formamide/2x
736 SSC (5 min, 37°C, 2x), 2x SSC (5 min, 37°C, 2x) and TS buffers (5 min, RT, 2x). Then the slides were mounted
737 and visualized as explained in the IF section.

738 FISH

739 The FISH only protocol was done as described in the IF-FISH section, skipping the IF section. Directly after
740 the permeabilization, the dehydration step was performed.

741 RNA-seq

742 Doxycycline treated cells were sorted to isolate dsRed positive cells upregulating *Xist* from the desired
743 allele. RNA was isolated using the ReliaPrep RNA Cell Miniprep System. A total of 16 DNA libraries were
744 created, according to the Smart-seq2 protocol [49], using the Nextera DNA Flex library prep kit (Illumina) to
745 create a library from full-length cDNA. Samples were sequenced on a HiSeq2500 sequencer (50 bp single-
746 end reads).

747 Chromatin immunoprecipitation (ChIP)-seq

748 50x10^6 cells were trypsinised, resuspended and fixed in 50 mL warm medium and 1% PFA for 10 min at
749 37°C. 2.5 mL Glycine 2.5 M were added to the cells (final concentration 0.125 M) to quench the PFA, 5
750 min RT on a rotator. All buffers from now on contain protease inhibitors (Roche, 4693132001). Cells were
751 washed twice with cold PBS. Then 1x in 10 mL Buffer 1 (10 mM Hepes pH 7.5, 10 mM EDTA, 0.5 mM EGTA,
752 0.75% Triton X-100) and 1x in Buffer 2 (10 mM Hepes pH 7.5, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA),
753 10 min rotating at 4°C. Nuclei were then resuspended in Lysis/Sonication Buffer (150 mM NaCl, 25 mM
754 Tris-HCl pH 7.5, 5 mM EDTA, 1% Triton, 0.1% SDS, 0.5% Sodium deoxycholate) and incubated 30 min on
755 ice. Nuclei were then sonicated 2x 15 min (30'' ON/OFF, max input, ice cold water) in a Bioruptor. A small
756 fraction of the lysate was run on a gel to confirm size population of 100-500 bp. 1 µg of antibody (Cell
757 Signaling, 9733S) was conjugated with 25 µL magnetic beads (Life Technologies, 10004D) for 3h rotating
758 at 4°C. 25 ug of chromatin was IP'd with the Ab-beads overnight rotating at 4°C. Beads were washed 2x in
759 standard RIPA buffer (140 mM NaCl, 10 mM Tris-HCl pH 7.5150 mM NaCl, 1 mM EDTA pH8.0, 0.5 mM
760 EGTA pH 8.0, 1% Triton, 0.1% SDS, 0.5% Sodium deoxycholate), 1x in High Salt RIPA (same as standard
761 RIPA but with 500 mM NaCl), 1x LiCl RIPA (same as standard RIPA but with 250 mM LiCl instead of NaCl)
762 and rinsed once with TE, 10 min 4°C each wash. Chromatin was eluted with 450 µL of Elution Buffer (1%
763 SDS; 0.1M $NaHCO_3$ in H_2O) with 22 µl protease K (10 mg mL^{-1}) and 5ul RNAse A (10 mg mL^{-1}) and shaken

764 at 1000 rpm for 2 hours at 37°C first and then 65°C overnight. DNA was then Phenol-Chloroform extracted
765 and resuspended in 20 µL H_20. 0.25 µL was used per PCR to confirm the ChIP's success. The concentration
766 was then measured and a ChIP-seq library was prepared following the manufacturer's instructions
767 (ThruPLEX DNA, Takara Bio) and sequenced on an Illumina HiSeq 2500 sequencer (50 bp paired-end
768 reads).

769 NGS data analysis: allele-specific RNA-seq and ChIP-seq

770 Both the RNA-seq as the ChIP-seq data were processed allele-specifically. The single nucleotide
771 polymorphism (SNPs) in the 129/Sv and Cast/Ei lines were downloaded from the Sanger institute (v.5
772 SNP142)[50]. These were used as input for SNPsplit v0.3.4[51], to construct an N-masked reference genome
773 based on mm10 in which all SNPs between 129/Sv and Cast/Ei were masked. The 50 bp single-end RNA-
774 seq and 50 bp paired-end ChIP-seq reads were mapped to this N-masked reference genome using the
775 default settings of hisat2 v2.2.1 and bowtie2 v2.4.1, respectively[52,53]. SNPsplit (--paired for the ChIP-seq
776 analysis) was then used to assign the reads to either the 129/Sv or Cast/Ei bam file based on the best
777 alignment or to a common bam file if mapping to a region without allele-specific SNPs. The allele-specific
778 and unassigned bam files were sorted using samtools v1.10[54].

779 For the RNA-seq, the number of mapped reads per gene were counted for both alleles separately using
780 HTSeq v0.12.4 (--nonunique=none -m intersection-nonempty)[55] based on the gene annotation from
781 ensembl v98. For each condition, genes with more than 20 allele-specific reads across both replicates
782 were used to calculate the allelic ratio, defined as Xi/(Xi+Xa). For the day 0 and day 7, Cast/Ei and 129/Sv
783 were used as the Xi and active allele (Xa), respectively, whereas for the day 3, Cast/Ei and 129/Sv were
784 used as Xa and Xi, respectively. The allelic ratios of X-linked genes were visualized as violin plots with
785 boxplots of the same data on top. Allelic ratios of individual genes were plotted along the X chromosome.
786 Different conditions were plotted together showing only the X-linked genes that had at least 20 allele-
787 specific reads in both conditions. We visualized differences in allelic ratios of X-linked genes between
788 conditions by plotting both ratios on the different axes of a scatter plot. Genes were highlighted when
789 they were identified as lowly silenced genes in *Spen*$^{-/-}$ ESCs from[27], which were defined as genes showing
790 $-0.05 < z < -0.2$ where z (gene silencing) = [Xi/(Xi+Xa)]dox – [Xi/(Xi+Xa)]noDox. Significant differences
791 between conditions were tested using a Mann-Whitney test with P-value < 0.05.

792 The allele-specific ChIP-seq bam files were normalized using the 'callpeak' and 'bdgcmp' functions of
793 MACS2 v2.2.7.1[56]. We called broad peaks (-f BAMPE --broad --bdg) and used the Poisson P-value as
794 method for normalizing the tracks. The input-normalized tracks were visualized using pyGenomeTracks

v3.4.[57]. For validation, we downloaded several publicly available datasets. The SPEN CUT&RUN data (SRX5903674, SRX5903675, SRX5903676, SRX5903677, SRX5903678, SRX5903679, SRX5903682, SRX5903683)[25], was processed similar to our analysis using a C57BL/6NJ-Cast/Ei reference genome. However, the allele-specific tracks were normalized based on the total number of mapped reads per sample. The scaling factor was calculated as 10^6 / total number of mapped reads and used as parameter --scaleFactor to both allelic tracks using deepTools bamCoverage v3.5.0.[58]. A binsize of 1 was used and paired-end reads were extended. The allele-specific tracks from HDAC3 and H3K27Ac (SRX4384412, SRX4384420, SRX4384476, SRX4384484, SRX4887836, SRX4887839) were downloaded from[30]. For all datasets, replicates for each condition were averaged using deepTools bigwigCompare v.3.5.0 with the settings '--operation mean --binSize 1'[58]. In the genome browser overview showing the allele-specific tracks, the y-axis was scaled for each group of samples separately.

Fig. 1

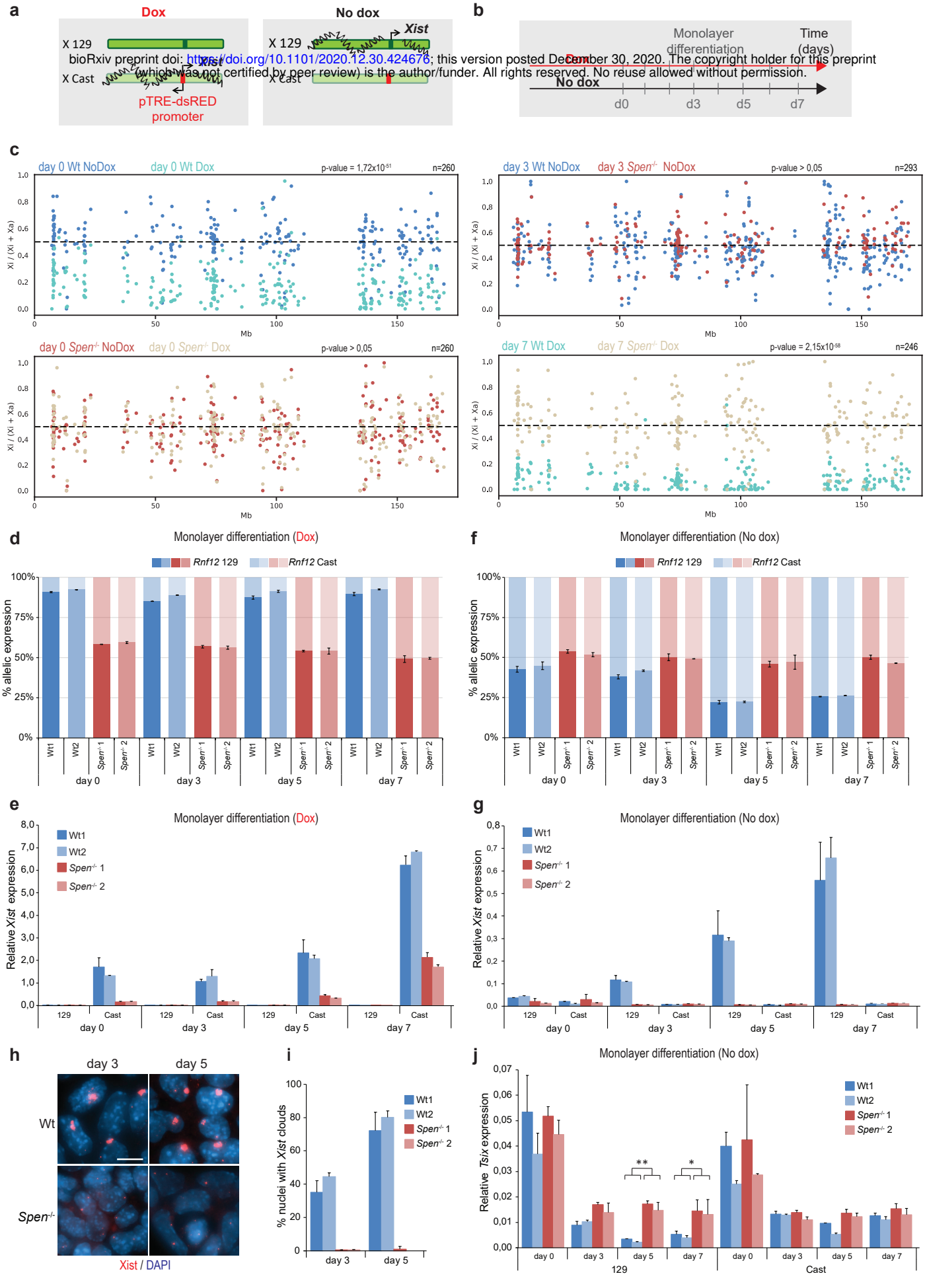


Fig. 2

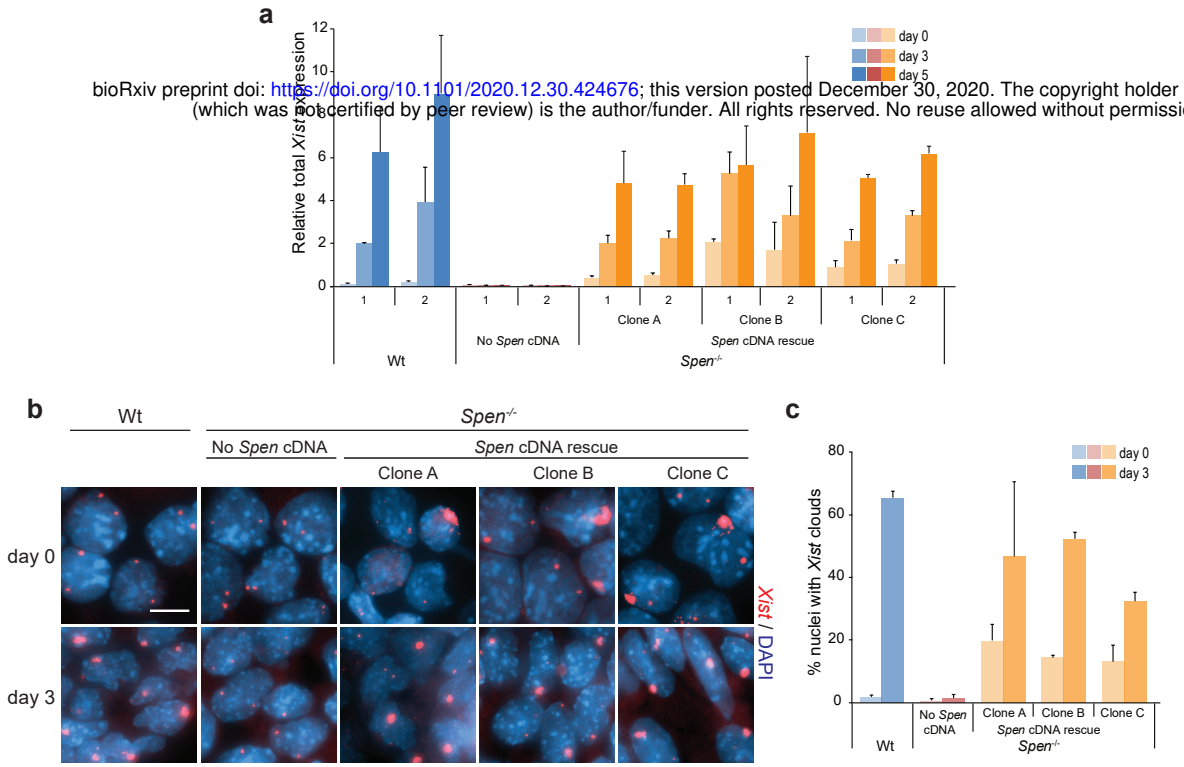


Fig. 3

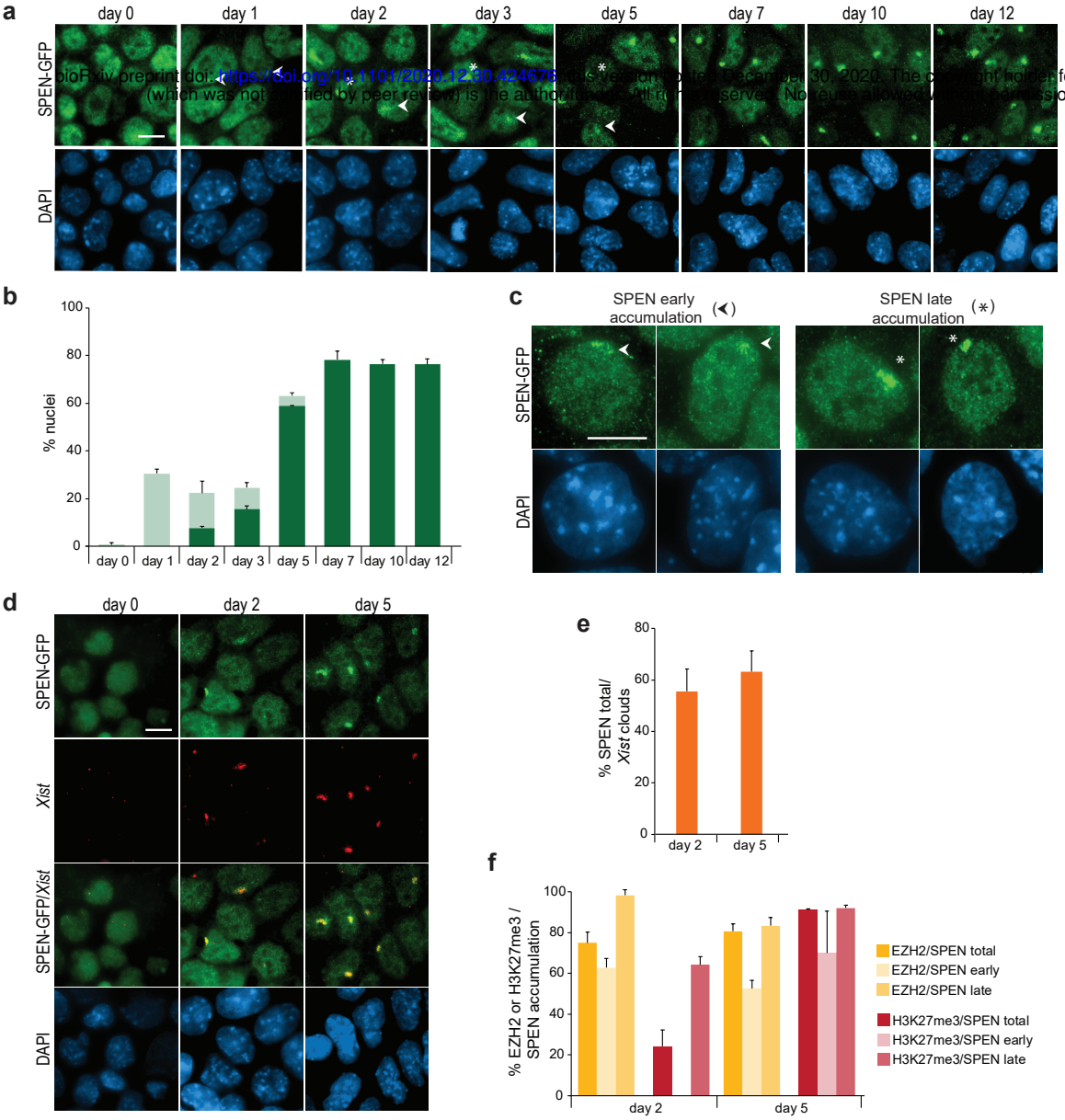


Fig. 4

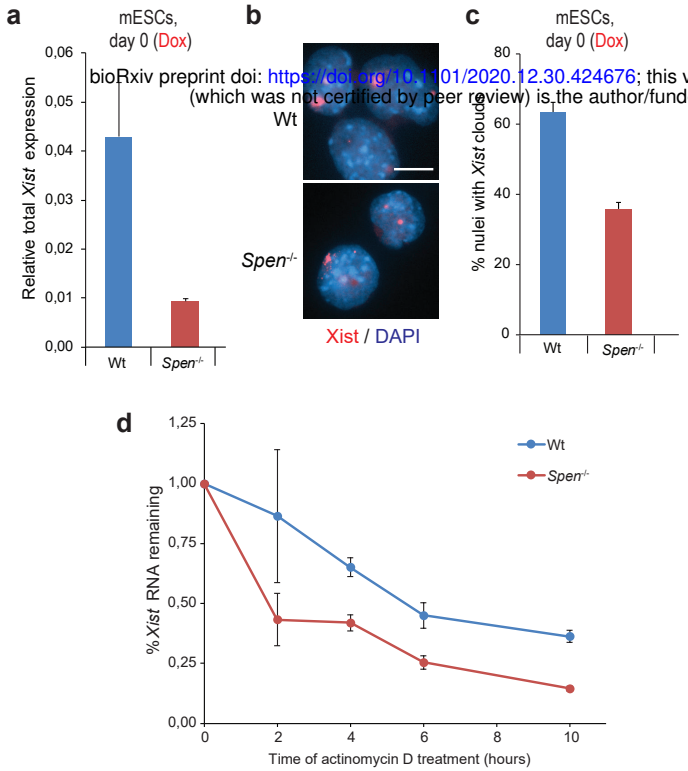


Fig. 5

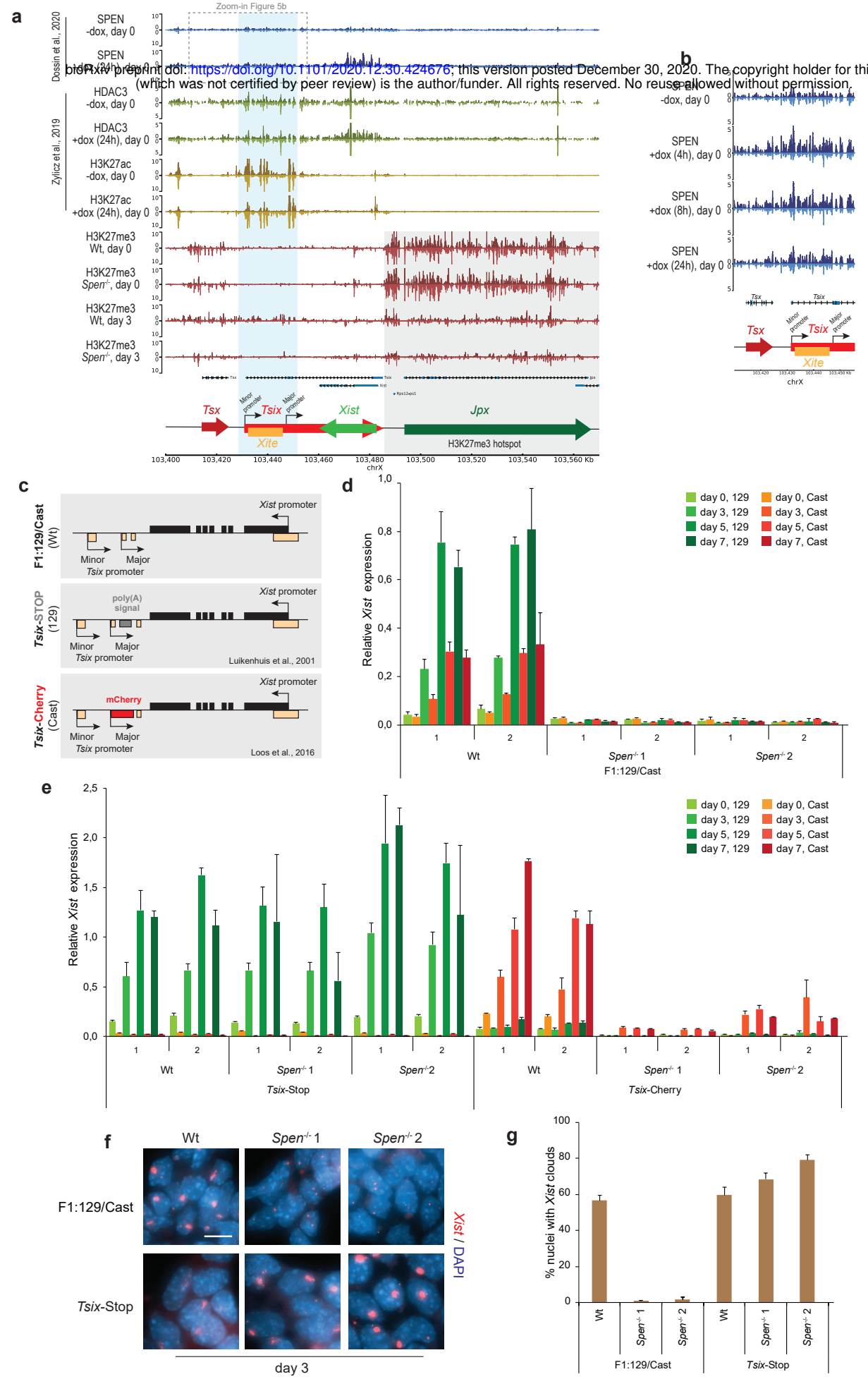
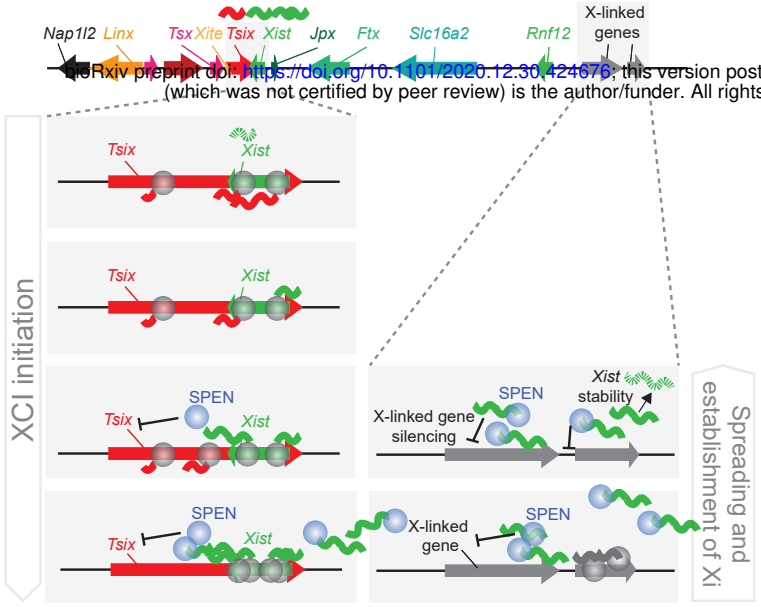
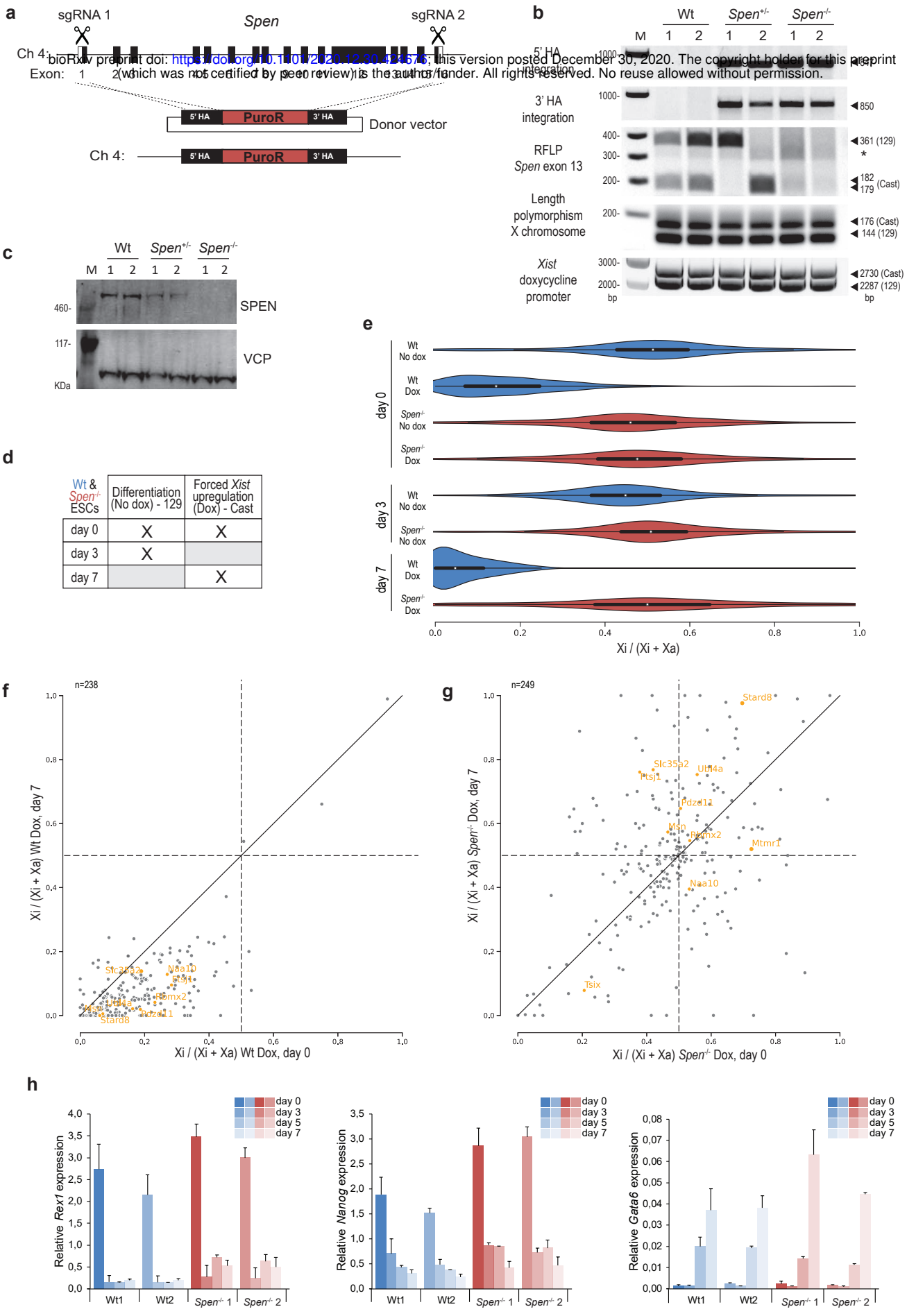


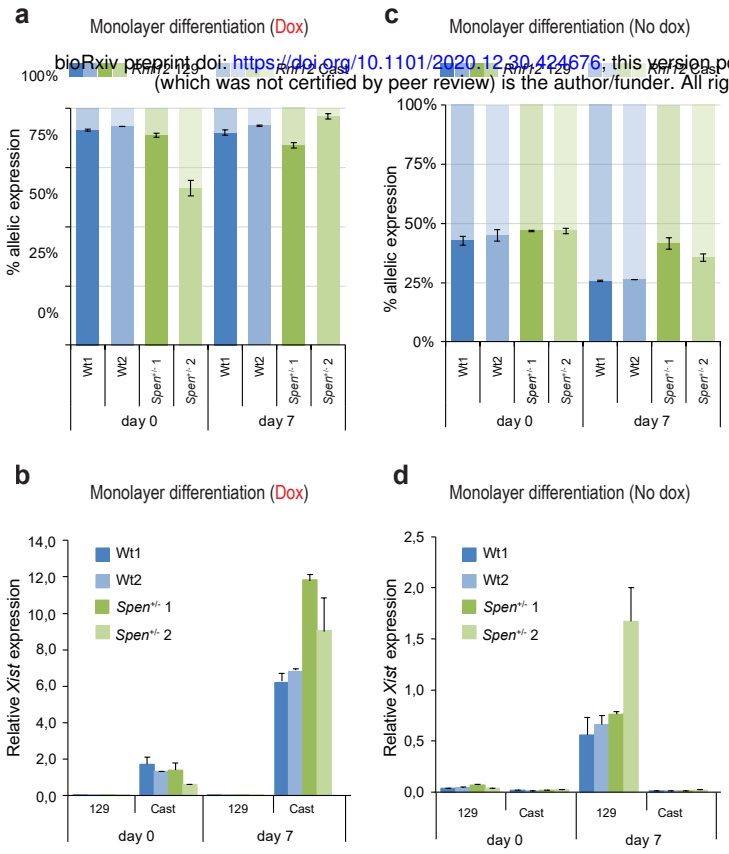
Fig. 6



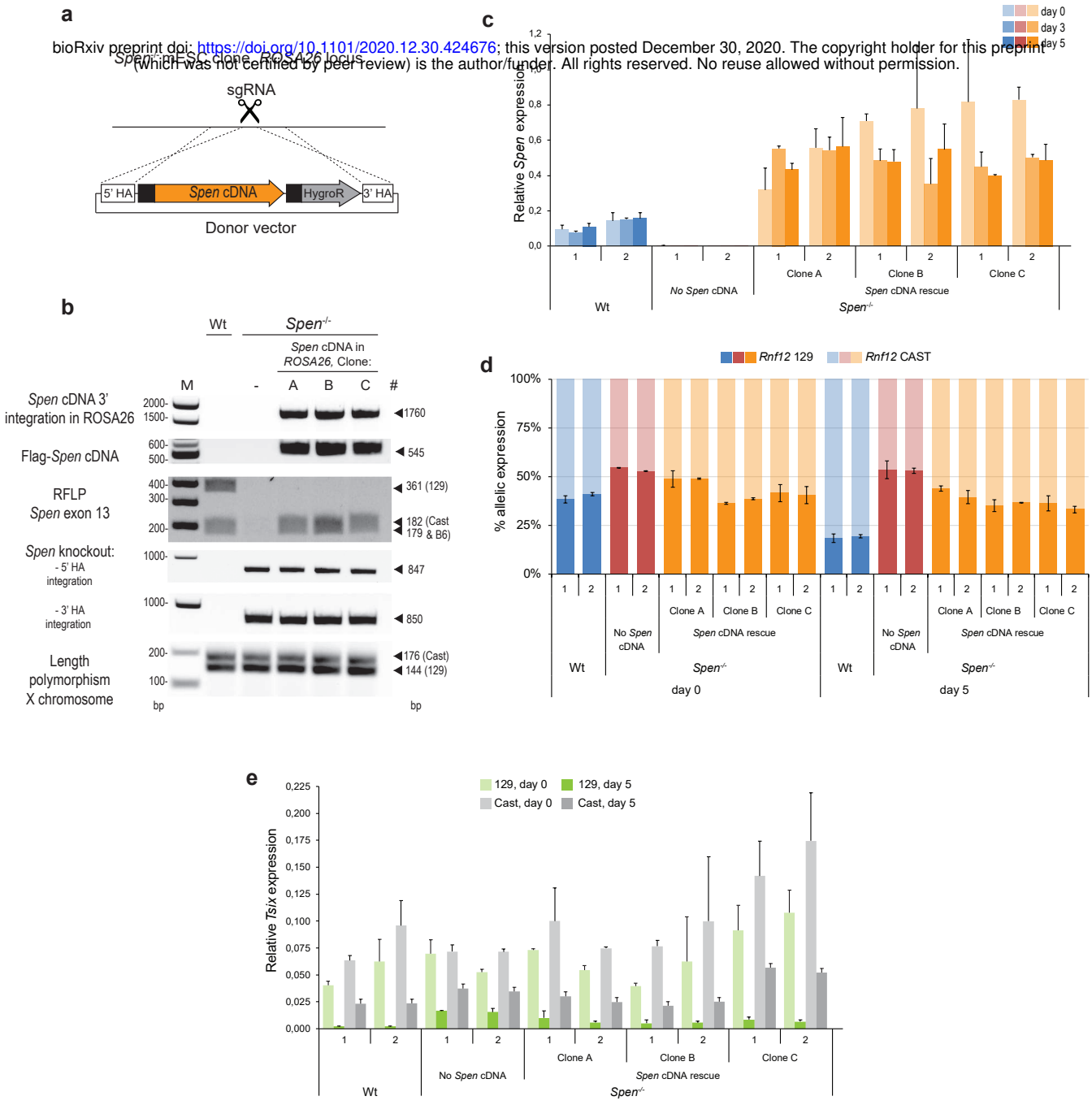
Supplementary Fig. 1



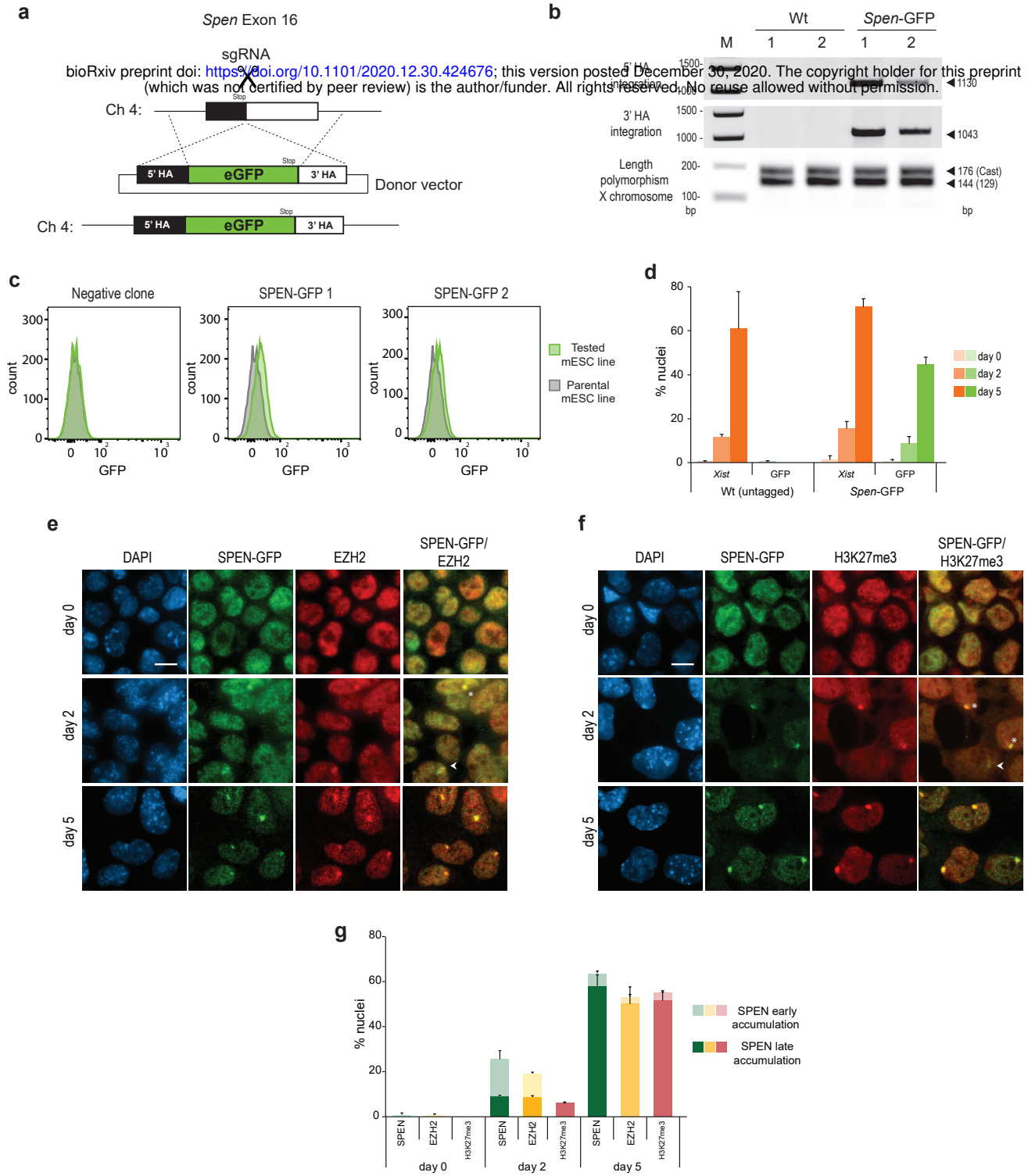
Supplementary Fig. 2



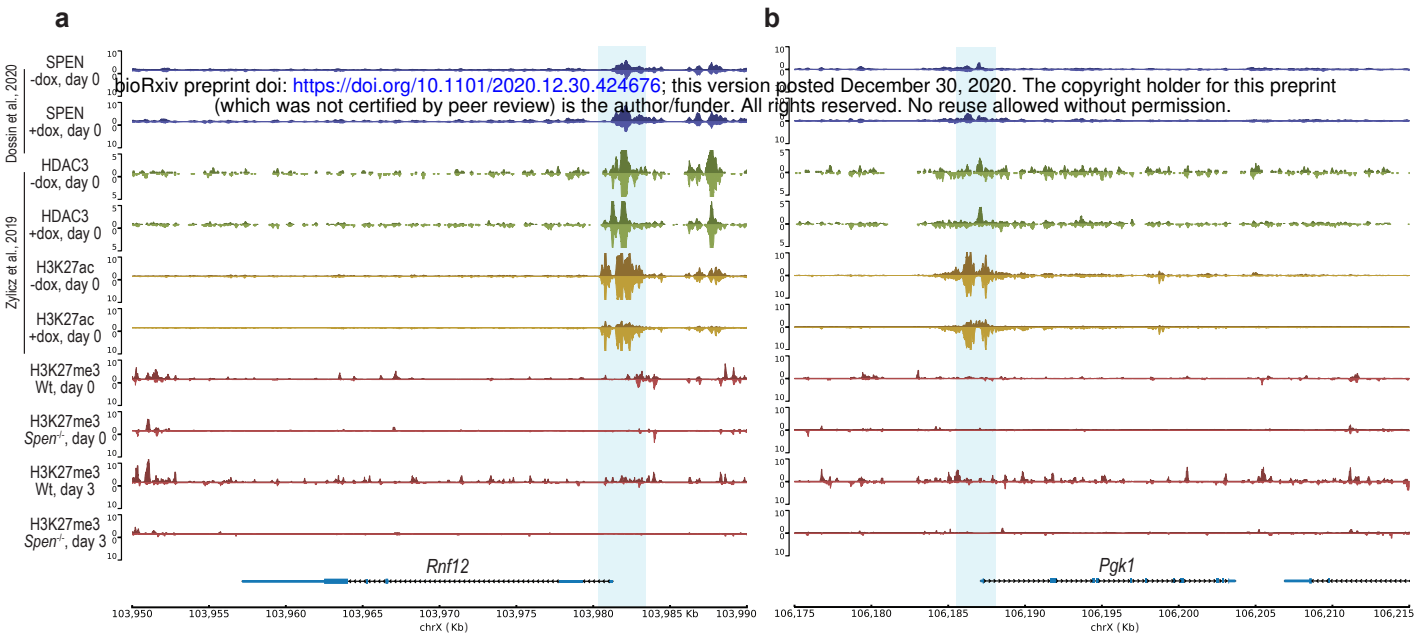
Supplementary Fig. 3



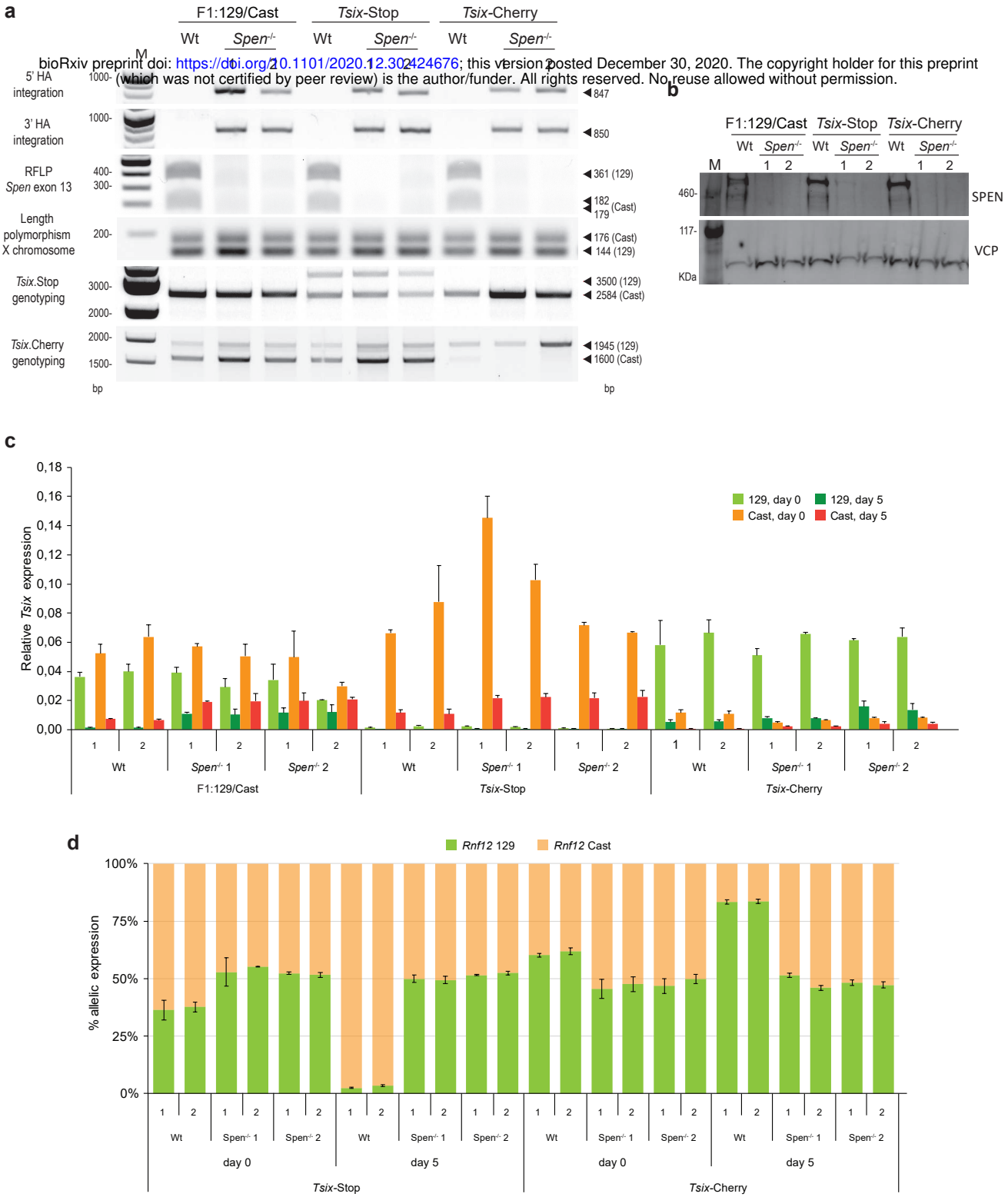
Supplementary Fig. 4



Supplementary Fig. 5



Supplementary Fig. 6



Appendix B

Long résumé en français

Chez les mammifères de la sous-classe des Thériens – comprenant les marsupiaux et les placentaires –, le dimorphisme sexuel est déterminé génétiquement et repose sur une paire de chromosomes sexuels appelés X et Y [Marshall Graves and Shetty, 2001]. Chez ces mammifères, les femelles sont porteuses de deux exemplaires du chromosome X (notées XX). Les mâles au contraire n'en portent qu'un seul, mais sont en revanche porteur du chromosome Y (notés XY). Bien qu'issus d'une même paire d'autosomes, les chromosomes X et Y sont génétiquement très différents, et ce en conséquence de plusieurs centaines de millions d'années d'évolution [Graves, 2006]. Alors que le chromosome X humain est long de plus de 160 mégabases, présentant plus de 800 gènes codant des protéines, le chromosome Y est nettement plus petit (60 mégabases), et ne porte que quelques dizaines de gènes codant des protéines [Graves et al., 2006].

Cette différence de taille entre les deux chromosomes sexuels implique un fort déséquilibre d'expression des gènes liés aux chromosomes sexuels chez les mammifères. En effet, alors que les mâles n'expriment qu'une seule copie des >800 gènes liés à l'X, les femelles ont quant à elles le potentiel d'exprimer ces même gènes à un niveau deux fois plus élevé. Un tel déséquilibre de d'expression génique n'est pas tolérable pour l'espèce. Ainsi, l'évolution a sélectionné un mécanisme de compensation de dosage de l'expression des gènes liés à l'X, appelé "inactivation du chromosome X" (ICX). Ce phénomène, qui n'a lieu que chez les femelles, au cours du développement embryonaire précoce, est caractérisé par l'inactivation transcriptionnelle complète de la quasi-totalité des gènes présents sur un des deux chromosomes X. A l'issu de cette ICX au sein des cellules femelles, le chromosome X qui a été inactivé (noté Xi) est maintenu à l'état inactif au cours de toutes les divisions cellulaires futures. Les cellules femelles n'expriment alors qu'un seul chromosome X, atteignant ainsi le dosage d'expression génique observé chez les mâles.

L'ICX a été mise en évidence pour la première fois par la généticienne britannique

Mary Lyon, en 1961 [Lyon, 1961]. Depuis, de nombreux laboratoires se sont intéressés à ce processus, et plusieurs décennies de recherche ont permis de mieux décrire et comprendre celui-ci. Dans les années 1990, plusieurs groupes ont mis en évidence que l'ICX est entièrement dépendante de l'expression du gène *Xist* [Brown et al., 1991; Brockdorff et al., 1991]. *Xist* est un gène lié à l'X transcrit en un long (15 kilobases) ARN non-codant [Brockdorff et al., 1992]. L'ARN de *Xist* s'accumule en *cis*, autour du chromosome X à partir duquel il est transcrit [Clemson et al., 1996].

La propagation de *Xist* le long du chromosome X entraîne la répression transcriptionnelle des gènes codés par ce dernier [Penny et al., 1996]. Cette répression s'accompagne de nombreux changements le long du chromosome. Les marques d'histone associées à l'activation de la transcription sont graduellement perdues, tandis que les marques d'histone associées à la répression de la transcription s'accumulent [Żylicz et al., 2019]. Parallèlement, le chromosome X se réorganise complètement au sein du nucléoplasme, formant un compartiment duquel l'ARN polymérase II est exclue [Chaumeil et al., 2006]. Le chromosome X inactif adopte alors une structure tridimensionnelle unique, s'organisant en deux mégadomaines mal structurés, séparés par le macrosatellite *Dxz4* qui joue un rôle de frontière topologique [Darrow et al., 2016; Deng et al., 2015; Giorgetti et al., 2016; Minajigi et al., 2015; Rao et al., 2014].

Etant donnée l'ampleur des changements à l'échelle de la transcription, de la chromatine, et de la structure chromosomique qui s'opèrent au cours de l'ICX, l'ICX est un modèle de grand intérêt pour l'étude de la régulation épigénétique, particulièrement afin de caractériser les interactions fonctionnelles entre la régulation de la transcription, de la chromatine, et de la structure 3D du génome. Les changements subis par l'Xi au cours de l'ICX sont particulièrement bien décrits et caractérisés. En revanche, les mécanismes par lesquels l'ARN de *Xist* entraîne l'inactivation transcriptionnelle du chromosome entier demeurent très peu compris (voire inconnus). En 2015, trois groupes de recherche ont indépendamment caractérisé l'interactome protéique de l'ARN de *Xist*, et ainsi identifié de nombreux facteurs pouvant être impliqués dans l'activité répressive de *Xist* [Minajigi et al., 2015; McHugh et al., 2015; Chu et al., 2015].

Parmi les facteurs identifiés, un seul l'a été invariablement au sein de chacune de ces trois études, et correspond à la protéine SPEN [Minajigi et al., 2015; McHugh et al., 2015; Chu et al., 2015]. De plus, SPEN a également été rapporté comme étant important pour l'inactivation des gènes au cours de l'ICX lors de deux cribles génétiques menés dans des cellules souches embryonaires murines [Monfort et al., 2015; Moindrot et al., 2015]. SPEN est une très grosse protéine, dont le poids moléculaire avoisine les 400kDa chez la souris et l'humain. SPEN possède quatre "RNA recognition motifs" (RRMs) à son extrémité amino-terminale, lesquels interagissent directement avec l'ARN de *Xist*

[Monfort et al., 2015; Lu et al., 2016; Carter et al., 2020]. L'interaction a lieu au niveau d'une séquence de *Xist* appelée *A-repat*, qui assure la fonction répressive de *Xist* [**?**]. A son extrémité carboxy-terminale, SPEN possède un domaine appelé SPOC (SPEN paralogue and orthologue C-terminal), capable d'interagir avec les corépresseurs NCoR et SMRT afin de réprimer la transcription, notamment par l'intermédiaire de certaines HDACs [Shi et al., 2001; Ariyoshi and Schwabe, 2003; Guenther et al., 2001].

En Octobre 2017, au démarrage de mon travail doctoral chez Edith Heard, la protéine SPEN avait été impliquée dans l'inactivation transcriptionnelle d'une dizaine de gènes au cours de l'ICX [Chu et al., 2015; McHugh et al., 2015; Monfort et al., 2015; Moindrot et al., 2015]. Aussi, une étude proposait que SPEN jouait un rôle crucial dans la répression transcriptionnelle des gènes liés à l'X en recrutant la protéine HDAC3 [McHugh et al., 2015]. En revanche, aucune étude n'avait à ce jour précisément établi à quel point SPEN est impliqué dans l'inactivation des gènes à l'échelle du chromosome X entier, ni par quel(s) mécanisme(s) un tel rôle pouvait être expliqué. C'est afin de répondre à ces thématiques que j'ai entrepris mon travail doctoral.

Dans un premier temps, j'ai, grâce à l'outil CRISPR/Cas9, introduit la technique dite du dégron inductible à l'auxine (DIA) [Nishimura et al., 2009]. Appliquée à SPEN, cette approche permet d'induire la dégradation protéique de SPEN, et ce de façon rapide, conditionnelle et réversible. En introduisant cette technique dans des cellules souches embryonaires murines (CSEM) capables d'induire l'expression de *Xist* et donc de procéder à l'ICX (cellules dites TX1072, [Schulz et al., 2014]), j'ai pu dans un premier temps interroger le degré d'importance de SPEN au cours de ce processus. J'ai ainsi pu observer qu'après dégradation complète de SPEN, aucun gène lié à l'X n'est capable d'être réprimé transcriptionnellement par *Xist*, démontrant que SPEN est requis pour l'inactivation de tous les gènes au cours de l'ICX. De plus, j'ai aussi pu confirmer ce rôle majeur de SPEN *in vivo*, par l'étude transcriptionnelle d'embryons de souris *Spen* KO, eux aussi incapables d'inactiver le chromosome X. J'ai pu, au cours de ces expériences, observer que l'expression de *Xist* ainsi que sa capacité à envelopper le chromosome X ne sont pas impactées par l'absence de SPEN, suggérant ainsi que ce dernier agit en aval de *Xist* au cours de l'ICX.

Ensuite, grâce à des techniques d'imagerie couplées à des systèmes génétiques ingénieux, j'ai pu suivre la dynamique simultanée des molécules de SPEN et de l'ARN de *Xist*, dans des cellules vivantes. Ces approches ont permis de rapporter que SPEN est recruté sur le chromosome X par *Xist*, aussitôt que celui-ci est exprimé. Cette information fournit une explication quant au rôle majeur que SPEN semble jouer au cours de l'initiation précoce de l'ICX.

En revanche, la déplétion de SPEN dans des cellules progénitrices neurales (dans

lesquelles l'ICX a eu lieu, et l'Xi est maintenu au stade inactif par des mécanismes épigénétiques), m'a permis de démontrer que SPEN ne joue pas un rôle essentiel dans la maintenance du chromosome X. En effet, après deux jours complets de déplétion de SPEN, aucune augmentation significative de l'expression des gènes complètement inactivés n'a été observée par RNAseq. Aussi, aucun changement significatif dans la structure tridimensionnelle du Xi n'ont été détectés par HiC dans ces mêmes conditions. En revanche, les quelques gènes échappant à l'inactivation du chromosome X, c'est à dire ceux conservant une expression significative à partir du Xi dans les cellules somatiques, montrent une expression encore plus élevée après déplétion de SPEN. Ainsi, bien qu'inutile pour réprimer les gènes dont la transcription est déjà inexistante après l'ICX, SPEN semble important pour tamponner le niveau d'expression des gènes qui demeurent actifs sur l'Xi.

L'avènement de la technique CUT&RUN [Skene and Henikoff, 2017], et son optimisation afin de cartographier les sites de liaison à la chromatine de SPEN m'ont ensuire permis d'identifier que *Xist* recrute SPEN strictement au niveau des enhancers et des promoteurs, et ce spécifiquement pour les gènes actifs. Aussi, cette analyse a révélé que les gènes dont les promoteurs sont les plus enrichis pour SPEN au cours de l'ICX sont aussi ceux qui sont inactivés le plus rapidement après l'expression de *Xist*. Enfin, j'ai également pu constater que SPEN se dissocie de la chromatine une fois que la transcription a été éradiquée des gènes ciblés par SPEN. Ainsi, cette expérience montre un lien inattendu entre la transcription et la capacité de SPEN à cibler la chromatine, et indique que les enhancers et les promoteurs sont les substrats privilégiés où se déroule l'action répressive qui permet l'ICX.

En utilisant un fond génétique SPEN-dégron, j'ai ensuite pu interroger la capacité de différents mutants tronqués de SPEN à complémenter fonctionnellement l'ICX. Au cours de cette expérience de complémentation, seuls deux domaines ont été détectés comme jouant un rôle crucial pour la fonction de SPEN au cours de l'ICX. Le premier domaine fonctionnel, composé de trois RRMs (RRM2, RRM3 et RRM4) s'est révélé nécessaire pour la capacité de SPEN à s'accumuler sur l'Xi. Cette découverte est cohérente avec des études antérieures ayant montré que ces RRMs interagissent directement avec l'ARN de *Xist*, *in vitro* [Monfort et al., 2015; Lu et al., 2016].

A *contrario*, l'ablation du domaine SPOC (que celle-ci soit endogène ou performée sur une construction cDNA surexprimée) n'empêche pas SPEN d'être recrutée par *Xist* sur le chromosome X. En revanche, mon travail démontre que SPOC est absolument essentiel pour la fonction répressive médiée par SPEN au cours de l'ICX. La quasi-totalité de la fonction répressive de SPEN semble d'ailleurs dépendante de SPOC. Aussi, en contraignant "artificiellement" le domaine SPOC (seul) à être recruté sur le chromosome

X (dans un contexte où la protéine SPEN est absente), j'ai observé que SPOC est alors largement suffisant pour médier la répression transcriptionnelle le long du chromosome X entier. Ainsi, SPOC est nécessaire et suffisant pour l'ICX une fois que ce dernier est localisé sur le chromosome X.

Ces résultats suggèrent également que des facteurs interagissant avec SPOC sont très probablement impliqués dans les mécanismes d'extinction de la transcription qui opèrent pendant l'ICX. Afin d'identifier ces mécanismes, j'ai alors caractérisé l'interactome protéique de SPOC par une approche d'immunoprécipitation combinée à de la spectrométrie de masse. Cette expérience a révélé une multitude de facteurs appartenant à différents complexes multiprotéiques impliqués dans la régulation de l'expression génique. Parmi ces complexes, il convient de citer l'ARN polymérase II, le complexe NuRD, le complexe NCoR/SMRT (avec HDAC3), ainsi que le complexe de méthylation m^6A des ARN.

Ainsi, SPOC apparaît comme un hub, capable de promouvoir l'interaction avec différents facteurs, impliquant ainsi différents niveaux de contrôle de la transcription, afin d'assurer une inactivation efficace et robuste du chromosome X.

Pendant la dernière année de mon travail doctoral, je me suis concentré sur l'interaction observée entre SPOC et la machinerie de transcription, en l'occurrence, la sous-unité principale de l'ARN polymérase II, POLR2A. Au travers d'une collaboration avec le laboratoire de Christoph Müller à l'EMBL, et grâce à la technique de titration calorimétrique isotherme, j'ai pu montrer que SPOC interagit directement avec le domaine C-terminal (DCT) de POLR2A. Ce domaine, composé d'une multitude de séquences peptidiques "YSPTSPS" répétées en tandem, est soumis à de nombreuses phosphorylations *in vivo*, notamment au niveau des trois résidus de Sérine. L'interaction entre SPOC et le DCT de POLR2A n'a lieu uniquement lorsque ce dernier est phosphorylé au niveau du résidu de sérine n°5.

Après cristallisation du complexe SPOC/POLR2A, j'ai utilisé la technique de diffractométrie de rayons X pour résoudre la structure de ce complexe. Cette structure (résolue à 1.8 Å) montre que SPOC reconnaît le groupe phosphate porté par la phosphoSérine 5 du motif "YSPTSPS" au niveau d'une surface basique. L'interaction implique trois résidus basiques (deux lysines, une arginine).

En conclusion, l'ensemble de mon travail doctoral a permis la caractérisation fonctionnelle de SPEN au cours de l'ICX, à la fois par des modèles *in vitro* et *in vivo*. Les mécanismes par lesquels SPEN entraîne la répression transcriptionnelle tout au long du chromosome X ont pu être en partie révélés.

RÉSUMÉ

Chez les mammifères placentaires femelles, la transcription d'un des deux chromosomes X est inactivée (ICX) au cours du développement embryonnaire. Cette inactivation est permise par *Xist*, un ARN non-codant qui recouvre le chromosome X à partir duquel il est exprimé, induisant ensuite l'extinction transcriptionnelle de tout ce chromosome. Les mécanismes moléculaires par lesquels *Xist* permet une telle répression des gènes liés à l'X demeurent globalement incompris. En 2015, la protéine SPEN a été identifiée comme interagissant directement avec l'ARN de *Xist*, mais sa fonction précise ainsi que son mécanisme d'action au cours de l'ICX restaient à découvrir.

Au cours de mon travail de thèse, j'ai utilisé la technique du dégron inductible à l'auxine, une approche de type perte de function, permettant de dégrader SPEN à façon dans des cellules souches embryonnaires de souris (CSES) en cours d'ICX. Grâce à cette technique, je démontre que SPEN est absolument necessaire pour la repression transcriptionnelle de tout le chromosome X pendant l'ICX. Aussi, en ayant recours à des embryons de souris *Spen* KO, je montre que SPEN is tout autant essentiel pour l'inactivation du chromosome X paternel au cours de l'ICX soumise à empreinte. En revanche, la perte de SPEN dans des cellules différenciées, où l'ICX est déjà établie, révèle que SPEN n'est plus requis ni pour maintenir les gènes éteints, ni pour préserver l'organisation tridimensionnelle du chromosome X inactif.

Par des approches de microscopie en cellules vivantes, je rapporte ensuite que SPEN est colocalisé avec l'ARN de *Xist* immédiatement après l'expression de ce dernier, suggérant que SPEN peut initier la répression transcriptionnelle très tôt pendand l'ICX. La caractérisation des sites de liaisons de SPEN à la chromatine révèle que SPEN est recruté uniquement au niveau des promoteurs et des enhancers des gènes actifs. Aussi, la magnitude du recrutement de SPEN aux promoteurs liés à l'X prédit la rapidité avec laquelle les gènes sont inactivés au cours de l'ICX. Enfin, une fois les genes inactivés, SPEN se dissocie de la chromatine, ce qui indique qu'une activité transcriptionnelle est requise pour l'association de SPEN à la chromatine. Par complémentation fonctionnelle, le domaine SPOC est identifié comme l'effecteur principal de l'activité répressive de SPEN pendant l'ICX, et le recrutement « forcé » de SPOC sur l'ARN de Xist suffit à entraîner l'inactivation des gènes à l'échelle du chromosome entier.

L'identification de l'interactome protéique de SPOC par spectrométrie de masse révèle que SPOC interagit avec de nombreux complexes impliqués dans la répression transcriptionnelle : NCoR/SMRT (désacétylation des histones), NuRD (remodelage de la chromatine et désacétylation des histones), la machinerie de méthylation m6A des ARN, ainsi qu'avec la machinerie de transcription (Pol2).

En utilisant des approches de biophysique et de biologie structurale, je montre que SPOC interagit directement et spécifiquement avec le domaine C-terminal (CTD) de Pol2, seulement quand ce dernier est phosphorylé sur la Sérine 5. Ces résultats suggèrent que SPEN peut réprimer la transcription directement en interférant avec les évènements transcriptionnels dépendant de Pol2-CTD Ser5P.

Ainsi, mon travail de these souligne l'essentialité de SPEN pour éteindre la transcription à l'échelle du chromosome X entier au cours de l'ICX, aussi bien *in vitro* que *in vivo*. Immédiatement après l'expression de *Xist*, SPEN est recruté aux promoteurs et enhancers de gènes actifs, réprime la transcription, puis se dissocie de la chromatin une fois les gènes éteints. Grâce à ses domaines RRMs et SPOC, SPEN joue un rôle d'intégrateur, asociant *Xist* à des désacétylases des histones, des remodeleurs de la chromatine, mais surtout, à la machinerie de transcription.

MOTS CLÉS

épigénétique - inactivation du chromosome X, SPEN/SPOC, génomique fonctionnelle, biologie structurale, cellules souches embryonnaires murines, CRISPR/Cas9

ABSTRACT

In female placental mammals, dosage compensation of X-linked gene expression is achieved early during development through transcriptional inactivation of one of the two X chromosomes (XCI). This process is dependent on Xist, a long non-coding RNA which coats and silences the X chromosome from which it is transcribed. The mechanisms through which *Xist* initiates transcriptional silencing during XCI remain however completely unknown. In 2015, several studies identified that the SPEN protein binds *Xist* RNA directly, and its implication in mediating gene silencing was reported. However, its precise function and mechanism(s) of action during XCI are unclear.

During my PhD, I made use of a conditional loss of function approach, the auxin inducible degron, to acutely deplete SPEN in mouse embryonic stem cells (mESCs) undergoing XCI. Using this approach, I demonstrate that SPEN is absolutely necessary for chromosome-wide *Xist*-mediated gene silencing during initiation of XCI. Furthermore, using conditional *Spen* KO mouse embryos, I show that SPEN is also required for the transcriptional inactivation of the paternal X chromosome during imprinted X inactivation. Depleting SPEN in differentiated cells, in which XCI has been established, reveals that SPEN is neither required to maintain gene silencing nor to preserve the spatial organization of the inactive X chromosome.

By combining fixed and live cell imaging of *Xist* and SPEN, I show that SPEN colocalizes with *Xist* RNA, and accumulates on the X chromosome, immediately upon *Xist* upregulation, suggesting that SPEN can initiate gene silencing very early on during XCI. Profiling SPEN chromatin binding sites reveals that SPEN is recruited to promoters and enhancers of active genes specifically. The magnitude of SPEN recruitment to X-linked promoters dictates the efficiency with which these genes will be silenced. Remarkably, SPEN disengages from chromatin after gene silencing, indicating that active transcription required for SPEN's association with chromatin.

Using a functional complementation approach, I identify the SPOC domain as the effector of SPEN's gene silencing activity during XCI. Artificial tethering of SPOC to Xist RNA results in transcriptional repression along the entire X chromosome, demonstrating that SPOC contains all the sufficient potential to instruct gene silencing during XCI. I further characterize the protein interactors of SPOC using mass spectrometry and reveal that SPOC interacts with several protein complexes involved in repressing transcription, including the NCoR/SMRT (histone deacetylation), the NuRD (nucleosome remodeling/histone deacetylation) and the m6A writing (governing mRNA fate) complexes. Finally, several transcription initiation and elongation factors are found to interact with SPOC, as well as the RNA polymerase II (RNAPII) transcription machinery.

I identify that SPOC interacts directly and specifically with the C-terminal domain (CTD) of RNAPII only when the latter is phosphorylated on Ser5, and determine the 3D structure of the SPOC/RNAPII-CTD Ser5-P complex at 1.8Å resolution. These results suggest that SPEN could directly repress transcription during XCI by interfering with RNAPII-CTD Ser5-P templated processes.

Altogether, my PhD work reveals that SPEN is essential for chromosome-wide transcriptional silencing during XCI, both in mESCs and *in vivo*. Following *Xist* upregulation, SPEN is immediately recruited to active gene promoters and enhancers, silences transcription, and subsequently disengages from chromatin. Through its RRMs and SPOC domains, SPEN acts as a molecular integrator, bridging *Xist* with histone deacetylases, nucleosome remodelers, RNA methyltransferases and most importantly, the transcription machinery.

KEYWORDS

epigenetics - X chromosome inactivation, SPEN/SPOC, functional genomics, structural biology, mouse embryonic stem cells, CRISPR/Cas9