Title
Mechanisms of X-chromosome Regulation During Mammalian Development

Author
Sahakyan, Anna

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Mechanisms of X-chromosome Regulation During Mammalian Development

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Molecular Biology

by

Anna Sahakyan

2018

ABSTRACT OF THE DISSERTATION

Mechanisms of X-chromosome Regulation

During Mammalian Development

by

Anna Sahakyan

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles

Professor Kathrin Plath, Chair

The mammalian blastocyst forms several days after one of the smallest cells - the sperm - fertilizes one of the largest cells -the egg. Depending on sex chromosome contribution from the sperm, either a female (XX genotype) or a male (XY genotype) embryo develops. To compensate for the X chromosome genetic imbalance between males and females, female cells transcriptionally silence one of their two X chromosomes. This phenomenon of X-chromosome inactivation (XCI) occurs in the blastocyst inner cell mass (ICM) cells – which form the embryo – as it implants to the uterine wall. The transition of the naïve pluripotent ICM cells with two active X chromosomes to primed pluripotency with an active and an inactive X chromosome (Xa and Xi, respectively) is mediated by the lncRNA *Xist*. While both mouse and human cells arising after implantation have an *Xist*-expressing Xi and an *Xist*-negative Xa, the X-chromosome state in the naïve pre-implantation development is rather different in the two species. In mice, *Xist* is not

expressed in the ICM cells with two active X chromosomes since *Xist* expression invariably leads to XCI. Yet, human pre-implantation blastocysts, including cells of the ICM, express *XIST* from active X chromosomes. We demonstrate that the presence of an *XIST*-expressing Xa, which is unique to human pre-implantation development, is a robust marker of human naïve pluripotency. We utilize this marker to identify a culture condition that, for the first time, allows detailed molecular studies of X-chromosome regulation of human pre-implantation development using cultured human pluripotent stem cells (hPSCs). We demonstrate that naïve hPSCs, despite having two active X chromosomes, compensate expression of X-chromosome genes via chromosome-wide transcriptional dampening and mediate *XIST*-mediated XCI upon differentiation to a somatic lineage. Lastly, we determine that naïve culture conditions obliterate the epigenetic abnormalities of the Xi characteristic to conventional hPSCs of developmentally advanced – primed – pluripotent state.

Once established, the Xi is kept transcriptionally silent for the life of the cell and its progeny, namely due to the many epigenetic layers forming the facultative heterochromatin of the Xi. In addition to covalent modifications of histone proteins, CpG islands of X-chromosome genes are methylated at cytosine residues to keep this silent state. To uncover the mechanisms which work in synergy with DNA methylation to maintain the inactive state of the Xi, we performed an RNAi and small chemical screen using engineered mouse cells with a reporter gene on the silent X chromosome. Our screen identified a synergistic combination of two FDA-approved chemicals that together lead to increased DNA demethylation not only of the Xi, but genome-wide, and works in synergy to reduce viability of leukemic cells. Thus, understanding XCI at the molecular level can be used to optimize the epigenetic activity of drug combinations.

The dissertation of Anna Sahakyan is approved.

Siavash K. Kurdistani

Amander T. Clark

Tracy L. Johnson

William E. Lowry

Kathrin Plath, Committee Chair

University of California, Los Angeles

2018

To my parents

TABLE OF CONTENTS

FIGURES AND TABLES

ACKNOWLEDGEMENTS

I am forever grateful to my mentor, Dr. Kathrin Plath, for her infectious enthusiasm for science, insightful guidance, and unparalleled support. Thanks for leading by example and helping me develop as an independent thinker.

I am proud to have been a member of the Plath lab. Thank you all for your mentorship and friendship, scientific support and constructive criticism. Many thanks to Drs. Costas Chronis, Amy Pandya-Jones, Giancarlo Bonora, and Iris Dror for their guidance and collaboration. I am thankful to Dr. Rachel Kim and Shan Sabri for their contribution to the work covered in Chapter 3. Thanks to Dr. Alissa Minkovsky for working together at the start of my PhD. Thanks to Yihao Yang for being an amazing student and having the magic 'touch' in molecular biology. Robin McKee does a tremendous job in balancing order and fun in the lab, which I acknowledge and respect.

I am thankful to my committee members for their scientific input and continuous support. I would like to acknowledge the Molecular Biology Institute, the Whitcome Fellowship, and Dr. Fowler, for financial support of my studies. I am indebted to the Mangasar M. Mangasarian scholarship for supporting higher education of people from the Armenian descent. My gratitude also extends to the Graduate Division at UCLA, Iris Cantor-UCLA Women's Health Center at UCLA, and the National Institute of Health for financial support.

I am forever grateful to my parents for going through many obstacles and making personal sacrifices to establish a new life for our family in the United States. I am lucky to have an elder brother who instilled the love of science in me and tutored me in chemistry and biology.

I have been fortunate to share this journey, with all its ups and downs, with the most supportive, encouraging, and adventurous person I know – my husband Jacob.

Chapter 1 is a version of a peer-reviewed review paper in press in the journal *Trends in Cell Biology*. It is authored by Sahakyan, A., Yang, Y., and Plath, K.

Chapter 2 is a reformatted version of a publication in the journal *Epigenetics Chromatin* reporting findings from a collaborative project with equal effort from Alissa Minkovsky, Giancarlo Bonora, and me. Its authors are Minkovsky, A., Sahakyan, A., Bonora, G., Damoiseaux, R., Dimitrova, E., Rubbi, L., Pellegrini, M., Radu, C.G., and Plath, K.

Chapter 3 is a reformatted version of our publication in the journal *Cell Stem Cell* authored by Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., and Plath, K.

Chapter 4 is a peer-reviewed review manuscript authored by Sahakyan, A., Plath, K., and Rougeulle, C. and published in *Philosophical Transactions of the Royal Society B: Biological Sciences.*

VITA

University of California, Los Angeles

2008 – 2010	Bachelor of Science Microbiology, Immunology and Molecular Genetics Summa Cum Laude, Departmental Honors
2010-2012	Staff Research Associate School of Nursing, AIDS Institute
2014	Teaching Assistant, Life Sciences
2017 – 2018	Writing Consultant Graduate Writing Center

Academic Honors

2018	MBIDP Dissertation Year Award, UCLA
2016- 2017	Dissertation Year Fellowship, UCLA
2017	Fowler Fellowship, Molecular Biology Institute, UCLA
2015 – 2017	Ruth L. Kirschstein NRSA F31 Individual Pre-Doctoral Fellowship, NIH
2015 -2016	Iris Cantor Women's Health Pilot Award, UCLA
2014 -2015	Philip Whitcome Pre-Doctoral Fellowship, UCLA
2014 -2015	Mangasar M. Mangasarian Scholarship, UCLA

Selected Publications

- Sahakyan, A. *et al.* Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* **20,** 87–101 (2017).
- Sahakyan, A., Plath, K. & Rougeulle, C. Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372,** 20160363 (2017).
- Patel, S., Bonora, G., Sahakyan, A. *et al.* Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. *Cell Reports* **18,** 54–67 (2017).
- Sahakyan, A. & Plath, K. Transcriptome Encyclopedia of Early Human Development. *Cell* **165,** 777–779 (2016).
- Gu, W., Gaeta, X., Sahakyan, A. *et al.* Glycolytic Metabolism Plays a Functional Role in Regulating Human Pluripotent Stem Cell State. *Cell Stem Cell* **19,** 476–490 (2016).
- Minkovsky, A.*, Sahakyan, A.*, Bonora, G*. *et al.* A high-throughput screen of inactive X chromosome reactivation identifies the enhancement of DNA demethylation by 5-aza-2'-dC upon inhibition of ribonucleotide reductase. *Epigenetics Chromatin* **8,** 42 (2015). * Equal contribution.
- Minkovsky, A., Sahakyan, A. *et al.* The Mbd1-Atf7ip-Setdb1 pathway contributes to the maintenance of X chromosome inactivation. *Epigenetics Chromatin* **7,** 12 (2014).

CHAPTER 1

The Role of *Xist* in X-chromosome Dosage Compensation

X-chromosome dosage compensation

Male (XY) and female (XX) eutherian mammals have equivalent expression levels of most X-chromosome genes despite the presence of an extra X chromosome in females. This X-chromosome dosage compensation is due to the phenomenon of X-chromosome inactivation (XCI), which refers to the transcriptional silencing and heterochromatinization of one of the two X chromosomes in females early in embryonic development [1]. Most of our knowledge of XCI is based on mouse studies, where two types of XCI exist: imprinted and random (Figure 1-1A).

In imprinted XCI, which initiates in all cells of the female mouse four- to eight-cell stage pre-implantation embryo, the paternally inherited X chromosome (Xp) undergoes inactivation, while the maternally inherited X chromosome (Xm) remains active [2,3]. As pre-implantation development progresses to form the blastocyst, cells of the trophectoderm layer, which give rise to extra-embryonic tissues (e.g. the placenta), maintain their imprinted XCI state. In contrast, epiblast cells of the blastocyst, which give rise to the embryo proper, reactivate the inactive Xp, re-establishing a state with two active X-chromosomes. The biallelic X-linked gene expression of epiblast cells is resolved again via XCI, but in this second wave of XCI either the Xp or the Xm is chosen at random for inactivation. Random XCI is maintained in all descendent somatic cells throughout life, resulting in adult mice that are a mosaic of cells expressing either maternal or paternal alleles of X-linked genes [4,5]. A group of X-linked genes express both the maternal and the paternal allele in each cell since these genes escape XCI and are thus the exception to the rule (reviewed by [6]). The chromosome-wide inactivation of the X chromosome, both in

imprinted and random XCI, appears to always be governed by the lncRNA X inactive specific transcript (*Xist*), which is encoded in the X-inactivation center (XIC) of the X chromosome [7].

The occurrence of both imprinted and random XCI in the same species, as is the case in mouse, may not be very common. Most mammals studied utilize only one form of XCI for X-chromosome dosage compensation. In marsupials, only imprinted XCI is observed where the Xp is exclusively chosen for inactivation [8] (Figure 1-1C). Contrary to this, imprinted XCI does not occur in rabbit, pig, [9,10] horse, or human development [11,12,13] based on analysis of pre-implantation blastocysts [9,10,13] or placental tissues [11,12]. In human post-implantation development, both extra-embryonic and embryonic lineages dosage-compensate via random XCI [12,14] (Figure 1-1B). However, in the first week of human development, prior to implantation and XCI, the existence of a novel gene-dosage regulation has recently been uncovered [13]. Here, both X chromosomes remain active from the onset of zygotic gene activation until the blastocyst stage [9,13] (Figure 1-1B). However, transcription from both X chromosomes is tuned down, or dampened, resulting in a net reduction of X-linked gene expression in female blastocyst cells [13]. X-chromosome dampening (**XCD**) has not been observed in any other mammal yet, but it has been reported in the nematode *Caenorhabditis elegans* [15], although the underlying mechanism in human and nematode may differ. In the XX hermaphrodite *C. elegans*, the 3D conformation of the X chromosomes is remodeled to reduce chromosome-wide gene expression by half in order to achieve gene-expression balance between XX hermaphrodites and XO males [16]. 3D chromosome conformation also differs between the active and inactive X chromosomes in mammals [17], suggesting that chromosome confirmation remodeling might also be at play in human XCD. However, unlike the mammalian inactive X-chromosome (Xi), lncRNAs have not

been reported to regulate X-chromosome dose in the *C. elegans*. Instead, dampening in the nematode is carried out by the dosage compensation complex, a condensin-containing multi-subunit protein assembly that binds at multiple sites along the X chromosome and leads to chromosome-wide compaction and gene repression [18].

Whether XCD observed in human pre-implantation development and *C. elegans* are related at the molecular level needs further investigation. Moreover, when single cell RNA-sequencing data of pre-implantation human blastocysts are analyzed using different bioinformatics tools and approaches, the dosage compensation observed in human pre-implantation embryos has been interpreted as initiation of XCI rather than dampening of both X chromosomes [19,20]. Fortunately, naïve human embryonic stem cells (hESCs), which are the *in vitro* counterparts of the pluripotent cells in the human pre-implantation embryo, exhibit XCD and thus can be used as a model system to address XCD and its relationship to the initiation of XCI further [21].

In this review, we discuss the experimental evidence examining the role of *Xist* in X-chromosome dosage compensation via imprinted and random XCI in mouse. We also consider *XIST* function in early human development and in human pluripotent stem cells (PSCs), reflecting on potential molecular mechanisms which might regulate context-dependent *XIST* function.

Long non-coding RNAs are key players in X-chromosome regulation

An intriguing fact about X-chromosome dosage regulation in all mammals is the utilization of long

non-coding RNAs (lncRNAs) such as *Jpx*, *Ftx*, *Tsix*, and *XACT,* most of which are located in the XIC of the X chromosome [20,22,23-28] (Figure 1-2). When expressed, *Xist* is exclusively associated with the X-chromosome from which it is expressed, acting only *in cis* [29]. The lncRNAs *Jpx* and *Ftx* exert their function by acting as activators of *Xist* to fine-tune *Xist* expression and thus indirectly regulate XCI. The *Jpx* lncRNA product acts either *in cis* or *in trans* [23] and binds the *Xist* repressor CTCF, taking away repression of *Xist* transcription [22]. Contrary to this, the *Ftx* transcript itself is not required for *Xist* regulation: it is the act of transcription of the Ftx locus that leads to Xist expression *in cis* [25]. Moreover, in mice, the lncRNA *Tsix*, which is transcribed in anti-sense orientation to Xist*,* represses Xist expression, thus ensuring Xist induction on the Xi and protecting the **Xa** (active X chromosome) from ectopic silencing by *Xist* [26,27]. *Tsix* is not expressed in human pre-implantation development [13], and the roles of *Jpx* and *Ftx* are yet to be examined in humans. A recently discovered lncRNA called *XACT* (X active coating transcript) that is unique to human PSCs, but does not reside in the XIC, may aid in maintaining transcriptional activity of the X chromosome from which it is expressed by counteracting *XIST* [20,28]. Interestingly, correlative studies suggest that *XIST* RNA, in addition to XCI, might also mediate the dampening of the transcriptional output of both X chromosomes of female human pre-implantation blastocysts [13,21] (discussed below). While *Xist* is unique to placental mammals, marsupials also use a cis-acting lncRNA encoded on the X chromosome, termed Rsx (RNA-on-the-silent X), which in many ways appears to act like *Xist* in XCI [8] (Figure 1-1). Taken together, it is rather interesting that different lncRNAs have evolved to regulate gene expression chromosome-wide *in cis*. This is perhaps due to the unique ability of lncRNAs to bind distant sites on chromatin while still tethered to their transcription loci. *Xist* [30] and other lncRNAs such as

HOTTIP (HOXA transcript at the distal tip), which is encoded on mouse chromosome 6 and activates genes in its neighborhood [31], reach their target chromatin sites simply by proximity – by being close to these sites in 3D space due to the folding of chromatin within the nucleus (reviewed and illustrated in [32]). Understanding how X-chromosome dosage is regulated via lncRNAs will thus not only shed light onto X-chromosome biology but can serve as a starting point in understanding how lncRNAs localize to and act on chromatin in general.

Xist is required for XCI in mouse

Xist is the best studied lncRNA to date. As the name suggests, *XIST* was discovered due to its association with the inactive X-chromosome – X inactive specific transcript [33-37]. The requirement of *Xist* for XCI was first directly implicated using female mouse embryonic stem cells (mESCs) with a mutated Xist gene lacking the first two thirds of exon 1 [38]. When induced to differentiate, both *in vitro* and *in vivo* (using aggregation chimeras), the X chromosome bearing the mutant truncated Xist gene was always spared from inactivation while *Xist* was expressed from the wild type X chromosome, causing non-random silencing [38]. This study demonstrated that Xist is required for choosing the chromosome for inactivation, and suggested, without direct evidence, that *cis*-expression of *Xist* is required for XCI.

Since mESCs are derived from the epiblast cells of the blastocyst, the role of *Xist* in pre-implantation development, from zygote to blastocyst formation, cannot be studied using these cells. Therefore, to further investigate the role of *Xist* in early development, mESCs were used to

generate chimeric mice containing cells with a large deletion of Xist, which were then mated with wild type mice to generate hemizygote males or heterozygote females [39]. When the mutant Xist was inherited from the mother, both normal female and male pups were born. However, female embryos with a paternally-inherited mutant Xist had severe prenatal growth defects and survived until approximately embryonic day 10.5 (E10.5). In these embryos, the extraembryonic tissues failed to develop due to the lack of imprinted XCI of the paternally inherited X chromosome in these tissues [39] (Figure 1-3). Due to the pre-determined choice of the paternal X chromosome for silencing in imprinted XCI, this experiment is the first clear demonstration that XCI, specifically imprinted XCI, cannot be initiated without *Xist*. Note that there is no paternal inheritance of mutant Xist to be studied in males since males inherit a Y chromosome but no X chromosome from their father. The inability of the extraembryonic tissue to support the growth of the embryo in the absence of X-chromosome dosage compensation was investigated further and this failure was attributed to exhaustion of the extra-embryonic ectoderm due to premature cell differentiation [40].

A recent study used single cell RNA-sequencing to provide high temporal and chromosome-wide resolution of X-linked gene silencing in mouse pre-implantation development and its dependence on *Xist* RNA [3]. Comparisons of female wild type to mutant embryos carrying a paternal *Xist* deletion at the 8-, 16-, 32-cell, and blastocyst stages validated the need for *Xist* in initiating imprinted XCI [3], which was also demonstrated recently using RNA-sequencing of single embryos [2]. Such high-resolution data were important to clarify the role of *Xist* in imprinted XCI and to rule out prior arguments for an *Xist*-independent imprinted XCI [41]. The dynamics of silencing during the initiation of imprinted XCI revealed that genes are silenced with

different kinetics. Genes silenced at an earlier stage of pre-implantation development were those near the XIC (where the Xist gene is located) in 3D space [3,30]. Intriguingly, this recapitulates the finding that at the initiation of random XCI, *Xist* first contacts the sites of the chromosome closest to its site of transcription in 3D, rather than linear space [30]. This correlation between the kinetics of imprinted XCI and the proximity to the Xist locus in 3D space independently supports the notion that imprinted XCI is mediated by *Xist*, and additionally suggests that the mechanism of *Xist* spreading *in cis* during the initiation of imprinted and random XCI is conserved.

Requirement of XCI in the development of the mouse embryo proper

Although the above-mentioned paternal Xist deletion experiment demonstrated that *Xist* is required for imprinted XCI, the early embryonic lethality due to malfunctioning extraembryonic tissues have made addressing the requirement of *Xist* in the embryo itself unfeasible with a germline mutation of Xist. A homozygous Xist knockout is required to address this question, which, when using a germline mutation, affects imprinted XCI as well (Figure 1-3). However, the maternal germline deletion of Xist allele was valuable to demonstrate the requirement of a functional Xist allele for choosing the X chromosome for XCI *in vivo* [42], extending the prior mESC study [38]. In the meantime, experiments demonstrating the sufficiency of *Xist* RNA for silencing was demonstrated. Particularly, ectopic expression of Xist cDNA from autosomes or the X chromosome [43,44], or activation of the endogenous Xist allele from the single X chromosome in male mESCs with an inducible promoter [30] demonstrated that *Xist* expression is sufficient to

cause silencing *in cis.* These gain-of-function experiments also opened the way for dissection of *Xist* RNA for its functional units [43].

Studying the need for *Xist* in random XCI requires a unique approach that specifically deletes Xist in embryonic tissues. A recent study took on the challenge to assess the importance of X-chromosome dosage compensation by random XCI in mouse embryonic development. To silence Xist specifically in the embryo while sparing the extra-embryonic tissue, *Xist* was conditionally deleted in the epiblast lineage [45]. While most mutant mice died *in utero,* surprisingly some mice survived to term, but exhibited growth retardation with reduced body size, dying within one month after birth (Figure 1-3). Only one mouse, which was a mosaic of XX and XO cells – cells that had lost one of their two X chromosomes - survived to adulthood. These data indicate that *Xist* and random XCI are required for normal embryonic development. However, given the survival to term, the phenotype appeared much weaker than expected and was a surprise – since XCI occurs soon after implantation the expectation was that *Xist* loss should have led to early embryonic lethality. This can perhaps be explained by another unexpected observation, namely that X-chromosome dosage compensation was not completely wiped out upon deletion of Xist. The authors reported partial X-chromosome dosage compensation in the absence of *Xist*, concluding that an *Xist*-independent mechanism was responsible [45].

This phenotype of partial X-chromosome dosage compensation in *Xist* mutant mice was observed due to less-than-expected increase in average expression levels of all genes by RNA-sequencing when compared to wild-type female mice, and due to the presence of mono-allelic gene expression in some, but not all cells at the single cell level of a few X-linked genes, assessed

by fluorescent *in situ* hybridization [45]. However, understanding the Xist mutant mouse model system used [45] can perhaps better explain these above-mentioned observations. Deletion of Xist in the epiblast lineage was accomplished with the Sox2 promoter-driven Cre recombinase [45]. However, the efficiency of Xist excision by the Sox2-driven Cre was not measured in the developmental interval around the induction of random XCI [46], as embryos were only harvested at E8.5 to confirm Xist deletion, a time point at which XCI has already occurred [45]. Originally, Sox2-driven Cre recombinase activity has been shown to occur in blastocyst outgrowths in culture and in all cells of the epiblast in E6.5 embryos *in vivo*, by assessing the removal of a 'stop' cassette in front of a beta-galactosidase reporter gene integrated into the ROSA26 locus [47,49]. Together, these findings suggest that the ROSA26 reporter recombined before E6.5 and possibly before XCI would be initiated *in vivo*, indicating that the Sox2-driven Cre-recombinase may be ideally suited to delete *Xist* before induction of random XCI. However, since chromatin accessibility may be different for the Xist locus in comparison to ROSA26, a region on mouse chromosome 6 identified because of its high recombination frequency [49], the Cre-mediated excision kinetics may differ for the two loci and thus need to be independently determined for the Xist locus. Additionally, the DNA segment flanked by LoxP sites, which need to come together for Cre-mediated recombination, is significantly longer in the Xist deletion construct [45] compared to the ROSA26 beta-galactosidase reporter system [48], and therefore potentially less favorable for deletion, again arguing for the need to establish *in vivo* Xist deletion kinetics in this mouse model. Hence, it cannot be ruled out that Xist deletion may have occurred *after* initiation/completion of random XCI.

Previous studies have shown that deletion of Xist has no dramatic *short-term* effect on the silent state of genes on the X chromosome when it occurs after the inactive X chromosome is fully established [50,51] (see below). Depending on the proportion of cells undergoing random XCI prior to Xist deletion, the embryos would then survive to term and demonstrate incomplete dosage compensation at organism level due to the mixture of XaXa and XaXi cells (Xa is the active X chromosome, Xi is the inactive X chromosome). Xist deletion post XCI establishment would also explain the low number of pups surviving to term, since if not enough cells per embryo undergo X-chromosome dosage compensation prior to Xist deletion, the embryo would not be viable, arguing for the importance of Xist in XCI and embryonic development. Taken together, additional experiments with a temporally precise Xist deletion (and confirmation of Xist deletion before induction of random XCI) are required to dissect whether a novel embryonic X-chromosome dosage compensation mechanism in the absence of Xist or the delayed deletion of Xist relative to the onset of XCI explain the surprisingly weak consequences of the current Xist deletion in the embryo [45]. For instance, using homozygous Xist knockout mESCs in tetraploid complementation assays [52] could provide wild-type extraembryonic tissues capable of supporting normal development while all cells of the epiblast, derived from mESCs, would lack Xist, ruling out the possibility of random XCI occurring in any fraction of the cells. It would be interesting to see if/when development would fail in this scenario, but our hypothesis is that no viable pups would be obtained from such mice. Regardless, the current data argue that X-chromosome dosage compensation mediated by *Xist* is critical for embryonic development.

The influence of XCI on the developmental potential of female cells has been shown with mESCs, as the double dose of X-linked genes delays the differentiation of these cells due to its

stabilizing effect on the naïve pluripotent state [55]. This stabilization is achieved via inhibition of the MAPK and Gsk3 pathways and stimulation of the Akt pathway, and XCI is needed to properly exit naïve pluripotency [53]. The delayed exit from pluripotency in the presence of two active X chromosomes may also occur *in vivo* [53], since embryos with a single X chromosome undergo accelerated development [54].

***Xist* executes XCI by recruiting a diverse set of proteins**

During initiation of XCI, *Xist* recruits numerous silencing factors to the X chromosome to establish facultative heterochromatin, also known as the Barr body [55]. This is accompanied by epigenetic changes including substitution of certain core histones, covalent modifications of histone tails, and promoter CpG methylation (reviewed by [1]). Although some of these *Xist*-induced epigenetic remodeling steps were discovered years ago, most of the proteins binding *Xist* directly and indirectly were identified only recently using mass spectrometry-based approaches and genetic screens [56-60]. While some of the *Xist*-binding proteins influence histone modifications (via the activation of the histone deacetylase HDAC3 through the engagement of SPEN by the 5' end of *Xist*) [57] or nuclear positioning of the *Xist*-coated Xi (through the binding of lamin B receptor (LBR) to *Xist*) [61], others induce RNA modification of adenosine methylation (m6A) on *Xist* to influence its silencing ability [62]. Identification of the *Xist* interactome – i.e. the proteins that directly or indirectly bind to *Xist* - has created a newfound appreciation of the multiple roles *Xist* plays in XCI, spanning from orchestrating chromosome-wide silencing, localizing itself to chromatin, altering chromatin state, and remodeling the 3D chromosome architecture, recently

reviewed in [55]. While the primary sequence of *Xist* RNA is not well conserved between mouse and human, the gene structure (exons/introns) as well as the presence of key repeat regions [37] are conserved in the two species. Some of these repeat regions are important for *Xist* function since they are the sites where proteins that directly interact with *Xist* bind to [56,57,63]. Therefore, although the mass-spectrometry based unbiased approaches have identified the *mouse Xist* interactome [56-58], it is safe to predict that the *human XIST* interactome will largely overlap with that of mouse [64].

Xist is required for long-term maintenance of random XCI

The Xi with *Xist* expression remains inactive in all somatic progeny of cells. In short-term *in vitro* studies (days – weeks), *Xist* does not seem to play a major role in the maintenance of the silent status of genes in random XCI in somatic cells [44,50]. This appears to also be the case in maintenance of imprinted XCI in the vole *Microtus Levis*, where ablation of *Xist* expression via deletion of its promoter region in trophoblast stem cells, which have already undergone imprinted XCI, does not lead to transcriptional reactivation or loss of repressive chromatin marks of the Xi [65]. However, a longer-term *in vivo* mouse study suggests that the prolonged absence of *Xist* in mice, initiated in the blood lineage using a tissue-specific Cre-recombinase, induces at least partial reactivation of genes on the X chromosome [51]. Notably, the experimentally induced deletion of *Xist* in hematopoietic cells in mice results in poor postnatal survival and development of myelodysplasia and various cancers of the blood with 100% penetrance [51] (Figure 1-3). The inevitable development of cancer in the absence of *Xist* clearly labels *Xist* as a

potent tumor-suppressor, most likely due to its requirement in the maintenance of gene silencing in somatic cells. In agreement with this *in vivo* mouse study, abnormal reactivation of the Xi has also been reported in human breast cancer cells, although here an extra dose of X-linked genes is either due to Xi erosion or loss of an Xi combined with an Xa duplication [66]. However, the connection between *XIST*-dependent maintenance of XCI and cancer formation in humans needs to be further explored. Since the importance of *Xist* in maintaining XCI only became obvious from mouse *in vivo* studies, it is critical to address the role of *XIST* in human cancers with carefully designed experiments.

There are two instances in mouse development that require reactivation of the Xi: once in cells of the inner cell mass (ICM) of the blastocyst, when imprinted XCI needs to be reversed prior to induction of random XCI, and once more in the development of primordial germ cells, prior to meiosis [67]. In both cases, shutdown of *Xist* expression from the existing Xi precedes the removal of chromosome-wide transcriptional repression [68,69]. Xi reactivation is also observed *in vitro*, when female mouse somatic cells such as mouse embryonic fibroblasts are reprogrammed to form induced pluripotent stem cells (**iPSCs**). Whereas the necessity of *Xist* loss in Xi reactivation is difficult to address in the *in vivo* scenarios described above, the *in vitro* reprogramming system has allowed detailed studies of this relationship. Using ectopic maintenance of expression or the deletion of Xist in a reprogramming experiment, it was demonstrated that Xist loss is necessary, but not sufficient, for Xi reactivation in iPSC generation, reviewed in more detail in [67]. During reprogramming to iPSCs, Xi reactivation is one of the last steps of reprogramming, requiring DNA demethylation in addition to *Xist* RNA loss [70]. Similarly, in somatic cells, DNA demethylating agents, such as 5-aza-dC, induce reactivation of genes of the

somatic Xi, albeit in a small proportion of cells, via induction of global DNA demethylation [71,72]. DNA methylation works in synergy with *Xist* RNA and histone hypoacetylation [71], as well as the H3K9 trimethylation pathway [72,73] in maintaining the inactive state of the somatic Xi. In fact, it takes the synergism of triple-drug combinations targeting DNA methylation, topoisomerase activity (involved in relieving torsional stress during DNA replication and transcription) combined with knockdown of an *Xist*-interacting protein to obtain dramatic re-activation of the Xi, and even then the re-activation is not for all silenced genes [58]. Complete chromosome-wide reactivation of all silenced X-linked genes in somatic cells has not been reported thus far, highlighting the unbreachable nature of the multiple epigenetic layers protecting the Xi.

Xi reactivation in human pluripotent stem cells

Studies of the relationship between *XIST* and Xi reactivation in human cells are not as straightforward as in mouse, mainly because 1) there is no imprinted XCI to be reversed in human pre-implantation development [9], 2) reactivation of the Xi in human primordial germ cells is difficult to study due to the hurdles associated with obtaining appropriate tissue samples and the lack of a human germ cell culture system that recapitulates Xi reactivation, and 3) reprogramming of human somatic cells under standard conditions does not lead to Xi reactivation [74] as it results in iPSCs that are in a developmentally advanced – primed – pluripotent state [75]. However, when conventional human iPSCs and ESCs in the primed pluripotent state are expanded in culture, *XIST* expression becomes gradually lost, which is accompanied by methylation of the XIST promoter [74,76-78]. The *XIST* loss in these pluripotent cells is usually accompanied by partial

reactivation of the Xi, where transcriptional repression of some, but not all genes on the Xi goes away, hence the overall inactive state of the Xi erodes, a phenomenon coined Xi erosion [77-80] (Figure 1-4). Importantly, erosion differs from escape of XCI as the genes undergoing erosion are initially silent on the Xi in early passage hPSCs and become reactivated upon extended passaging of these cells [80], whereas escape is not passage-dependent and rather includes genes on the Xi in chromosome regions with reduced *Xist* occupancy [30,81]. The extent of erosion of the Xi, i.e. the number of genes affected by this process, varies between individual human pluripotent stem cell (hPSC) lines; however, *XIST* loss occurs in nearly all hPSC lines studied over time in culture and often leads to Xi erosion [78,80] (Figure 1-4). Currently it remains to be tested whether loss of *XIST* expression causes Xi erosion, but the fact that no Xi erosion is observed while *XIST* is expressed in newly derived human iPSC lines [74] suggests such a causative relationship. While *XIST* may have a protective role in preventing Xi erosion, another X-linked lncRNA, *XACT* (X active coating transcript), has been implicated in driving Xi erosion in primed hPSCs [79]. While the eroded Xi does not interfere with hPSC growth or ability to differentiate, it does modulate these processes [80,82]. Moreover, when primed hPSCs with Xi erosion are differentiated, the reactivated genes on the Xi do not get re-silenced, resulting in somatic cells that at least partially lack dosage compensation of X-linked genes [80] (see [83] and [84] for detailed review) (Figure 1-4). Methods of repairing or preventing Xi erosion of female hPSCs are needed for their use in disease modeling [77] and regenerative medicine, particularly when considering X-linked diseases. For instance, iPSCs or iPSC-derived neurons from female patients with Lesch-Nyhan syndrome, a devastating disease affecting neurologic, cognitive, and behavioral functions [85], can be used to model the disease only in the presence of a faithfully silenced Xi. This is because

the disease phenotype is caused by a heterozygous mutation in the X-linked HPRT1 gene, leading to HPRT1 insufficiency in cells where the non-mutant HPRT1 resides on the Xi. When the region of the Xi harboring the HPRT1 gene undergoes erosion, it results in expression of the non-mutant HPRT1 gene product, over-riding HPRT1 insufficiency. Thus, the Lesch-Nyhan diseases phenotypes can no longer be faithfully recapitulated with cultured iPSCs or iPSC-derived neurons in the presence of Xi erosion [77]. Additionally, in regenerative medicine such as cell replacement therapies, introducing cells with an Xi erosion into a patient may be treacherous because these cells lack proper dosage compensation of X-linked genes, a phenomenon observed in cancers [66].

The role of *XIST* in early human development

It is interesting to note that primed hESCs do not reflect the X-chromosome state of the human pre-implantation embryos from which they are derived: all cells of a female human blastocyst, including those of the epiblast lineage, have two active X chromosomes and simultaneously express *XIST* [9,13], (Figure 1-4). The recent discovery of this non-silencing *XIST* in early human development has intrigued many researchers who study X-chromosome dosage compensation, including us. The two immediate questions regarding this unusual X-chromosome state are, 1) what role, if any, does *XIST* have, and 2) what is the molecular mechanism disabling *XIST* from silencing the X chromosome(s). These and many other questions cannot be addressed with *in vitro* studies of conventional (primed) hESCs since their X-chromosome state is different from the cells of the blastocyst from which they are derived, most plausibly due to suboptimal cell culture

composition used (reviewed in more detail in [83]). However, recently devised cell culture conditions, which have been formulated to support cells in a naïve (pre-implantation) pluripotent state, allow growth of hESCs that better resemble the pluripotent state of cells in the pre-implantation blastocyst from which they are derived [86,87]. Most importantly, the X-chromosome state of these naïve hESCs recapitulates many aspects of the human blastocyst, where female cells have two active X chromosome and express *XIST* [20,21]. While most of the cells in a pre-implantation female human blastocyst express *XIST* bi-allelically, this pattern is a minority in naïve hESCs which exhibit mostly mono-allelic *XIST* expression [21]. Hence, naïve hESCs resemble the blastocyst, but not perfectly, as there is still room for improvement in the naïve culture media formulation. The molecular mechanism behind the non-silencing *XIST* is currently not understood, but investigating *XIST*-interacting proteins and *XIST* RNA modifications, which have recently been demonstrated to be crucial for *Xist*'s silencing role in the mouse [55], warrant further investigation. Current naïve culture conditions will allow such studies since naïve hESCs exhibit non-silencing *XIST,* albeit mostly mono-allelically [21].

In addition to recapitulating the X-state of the pre-implantation blastocyst, naïve hPSCs allow *XIST*-mediated induction of XCI upon differentiation [21]. When primed hPSCs with large regions of Xi erosion are adapted to naïve pluripotency and then differentiated, the erosion is, for the first time, reversed and replaced with XCI [21] (Figure 1-4). Hence, the transition to the naïve state resets the X-chromosome abnormalities of the primed pluripotent state. However, when primed hPSCs are adapted to naïve pluripotency, the memory of the starting Xi does not get lost in the naïve transition, since upon differentiation the starting Xi becomes silenced despite the presence of *de novo* XCI [21]. Therefore, although naïve hPSCs allow studies of de novo XCI

in humans for the first time, they cannot be used to study choice of XCI since the process is non-random. The epigenetic memory of the starting Xi is unlikely due to DNA methylation, since the naïve state results in robust hypo-methylation of DNA [21,88,89], but may be due to the presence of histone modifications. For instance, it is possible that tri-methylation of histone H3 lysine 27 (H3K27me3), which was recently shown to regulate *Xist* imprinting in mice [90], marks the inactive or the active X chromosome through the transitions from primed to naïve pluripotency and eventually differentiation. It is, however, not clear whether naïve hESCs directly derived from the blastocyst or somatic cells directly reprogrammed to the naïve state can undergo random XCI upon differentiation.

Since there is no imprinted XCI in early human development, it has been unclear how X-chromosome dosage is compensated prior to onset of random XCI. Single cell RNA-sequencing of human pre-implantation embryos demonstrates gradual and time-dependent reduction of X-linked gene expression from both X chromosomes in embryonic days 4 to 7 in development [13]. This gradual dampening of X-linked gene expression correlates with upregulation of *XIST* [13]. X-chromosome dampening has also been observed in naïve *XIST*-expressing hESCs, further suggesting a novel role of *XIST* in human naïve pluripotency [21]. Independent analysis of the sequencing data from the pre-implantation blastocyst [13] and naïve hESC [21] studies has instead suggested the presence of XCI instead of XCD in human pre-implantation development [19]. If *XIST* is truly initiating XCI in the human blastocyst, given the fact that it is expressed from both X chromosomes in most cells, there must be a critical time-point at which point the cell decides to limit *XIST*'s silencing function to a single X chromosome, since silencing both X chromosomes is lethal due to phenotypic nullisomy of most X-chromosome genes [44,91].

Interestingly, blastocyst outgrowth studies demonstrated the presence of a *XIST*-negative transitionary state between the *XIST*-expressing blastocyst cells and the *XIST*-expressing XCI cells [80]. In the transition from XCD to XCI in hPSCs an *XIST*-negative state is also observed [21]. These data suggest that X-chromosome dosage compensation via XCD does not lead to the initiation of XCI.

If *XIST* is responsible for XCD in naïve pluripotency, it might do so by mediating accumulation of some, but perhaps not all chromatin modifications that are also responsible for XCI. For instance, H3K27me3 [92] accumulates on the *XIST*-coated active X chromosome in naïve hESCs [21], which might be responsible for dampening of X-linked gene expression. Another hypothesis is that expression of the lncRNA *XACT* may counteract some but not all functions of *XIST*, thereby achieving dampening instead of silencing. Indeed, it has recently been shown that *XACT* prevents accumulation of *Xist* when ectopically expressed on the mouse X chromosome [20], consistent with the idea that *XACT* can limit *XIST*'s activity in naïve hPSCs. Interestingly, in a fraction of cells of rabbit blastocysts, *Xist* gets expressed from both X chromosomes, initiating silencing of both X-chromosomes before resolving to mono-allelic XCI via unknown mechanisms [9]. It is possible that the human scenario derives from such a mode of initiation of XCI and that XACT has evolved in primates to alleviate the detrimental consequences of inactivating both X-chromosomes for too long or in too many cells.

Regardless, naïve hPSCs, for the first time, allow detailed molecular studies of *XIST* and XCD, as well as the transition to XCI as cells exit pluripotency. Moreover, by studying these cells

we can now gain insight into early human pre-implantation development and understand how it compares to what we already know in the mouse model organism.

Concluding Remarks and Future Directions

The biology of *Xist* unites researchers from multiple disciplines, including but not limited to those studying sex-chromosome dosage regulation, epigenetics (lncRNAs and chromatin remodeling), cell fate changes (reprogramming), cancer biology, disease pathogenesis (X-linked disorders), as well as development. *Xist*'s ability to recruit such a diverse group of researchers has enabled rapid advancement in understanding how *Xist* functions at the molecular level. We now understand that *Xist* acts as a scaffold to bring proteins to their site of action, effectively increasing the concentration of these proteins in a localized manner. Thanks to its very long-lived outcome and ability to be used in an allele-specific manner, *Xist*'s ability to silence genes can be mined for therapeutic purposes to balance gene expression in trisomic diseases such as Down Syndrome [93]. Furthermore, increased understanding of *Xist* can help engineer variants of this lncRNA to silence smaller and specific regions of a chromosome, increasing its therapeutic potential to silence mutant genes in an allele-specific manner.

Xist's unique expression pattern in human pre-implantation development (expression without silencing) is a great marker of human naïve pluripotency [21] which can be used to develop new and improved naïve culture conditions in the future. Lastly, from an evolutionary perspective, *Xist* is an interesting lncRNA to study since it carries out similar functions in mouse

and human despite lack of sequence conservation, but also seems to have evolved extra functions in a context-dependent manner, which requires further investigation. Both mouse and human *Xist* serve as a wonderful model for expanding our knowledge on lncRNA function, while learning about development and dosage regulation.

FIGURE LEGENDS

Figure 1-1: X-chromosome inactivation in different mammals

The X-chromosome states of eutherian (A and B) and **metatherian** (C) female mammals are shown in embryonic development. A) Imprinted XCI occurs in mouse pre-implantation development, but it is re-set in the cells that develop into the embryo to give way to random XCI, resulting in a mosaic adult female mouse. Both imprinted and random XCI in the mouse are regulated by the lncRNA *Xist.* B) Humans have evolved away from imprinted XCI as they dosage compensate in pre-implantation development by turning down transcription from both X chromosomes via XCD. Moreover, *XIST* is expressed on both dampened X chromosomes, where its functional role remains to be determined. In post-implantation development, similar to the mouse, human females display random XCI mediated by *XIST.* C) Metatherians, such as the marsupial opossum (*Monodelphis domestica*) dosage-compensate by inactivating the paternally-inherited X chromosome using the lncRNA *Rsx*. This imprinted dosage-compensation is maintained throughout marsupial development, resulting in a female adult with a transcriptionally inactive paternal X chromosome in all of its cells. XCI = X chromosome inactivation, *Xist / XIST* = X inactive specific transcript, Xa = active X chromosome, Xi = inactive X chromosome, Xm = maternal X chromosome, Xp = paternal X chromosome, XCD = X chromosome dampening, *Rsx* = RNA on the silent X.

Figure 1-2: Long non-coding RNAs involved in X-chromosome dosage regulation in mouse and human

The X inactivation center (XIC) is located on the X chromosome and harbors the master regulator of XCI – the long non-coding RNA (lncRNA) Xist. A) In the mouse, *Xist* itself is positively regulated by the lncRNAs *Jpx* and *Ftx*, which are also encoded in the XIC, upstream of the *Xist* gene. The lncRNA *Tsix*, which is anti-sense to *Xist*, has a mutually exclusive expression pattern with *Xist*: it is expressed bi-allelically from both X chromosomes prior to XCI. *Tsix* 'protects' the active X chromosome in pluripotency from being silenced by *Xist* upon induction of XCI, and is thought to be a repressor of *Xist*.

B) The human *XIST* gene is also encoded in the XIC of the human X chromosome. The lncRNAs *JPX* and *FTX* are also upstream of *XIST* in the human XIC, similar to mouse. The role of *JPX* and *FTX* in regulating *XIST* expression in human is speculated based on mouse studies. Unlike mouse, the human XIC does not contain the *XIST* anti-sense lncRNA *TSIX*, since *TSIX* expression is not detected in pre-implantation blastocysts or human embryonic stem cells. A novel, human-specific lncRNA, X active coating transcript (*XACT*), however, is encoded about 40Mb upstream of the human XIC and seems to antagonize *XIST* in naïve pluripotency. Similar to mouse *Tsix,* expression of *XACT* is unique to pluripotent cells and not detected in somatic cells.

Figure 1-3: Lack of *Xist* at various developmental time-points highlights its importance in normal development

Summary of key studies addressing the role of *Xist* in mouse development from fertilization to birth and into adulthood. (i) When a zygote is formed with a maternally deleted Xist (inherited

from the egg), mouse development progresses normally and results in non-mosaic adults where all cells inactivate the paternally inherited X chromosome (since only that X has the only functional *Xist* allele). (ii) However, when *Xist* is deleted from both X chromosomes (from the egg and the sperm), or from only the paternal X chromosome (sperm), extra-embryonic tissues fail to develop in the absence of imprinted XCI since this process requires paternally-inherited *Xist*, and thus mouse development halts 5-7 days post implantation. (iii) Conditional Xist deletion from both X chromosomes in epiblast cells that give rise to the embryo is often embryonically lethal, and if pups are born, they display partial loss of X-chromosome silencing and do not survive to adulthood. (iv) When Xist is deleted several days post-implantation, specifically in hematopoietic stem cells (HSCs), after the establishment of the Xi, pups are born but succumb to Multilineage Dysplasia as early as 1.5 months after birth. E = embryonic day, *Xist* = X inactive specific transcript, iXCI = imprinted X-chromosome inactivation, HSCs = hematopoietic stem cells.

Figure 1-4: The X-chromosome state of naïve and primed human pluripotent stem cells

(i) Female human somatic cells have an active and an *XIST*-expressing inactive X chromosome (Xa and Xi). (ii) Reprogramming of these cells to **primed pluripotency** does not change the X-chromosome state. (iii) Similarly, derivation of hPSCs from a pre-implantation blastocyst stabilizes the post-XCI state in primed pluripotent culture conditions. (iv) Over time in culture, the Xi loses expression of *XIST* and undergoes epigenetic erosion, resulting in partial reactivation and thus double-dose of the X-linked genes that fall in these eroded regions in primed hPSCs. (v)

Although these cells can differentiate into somatic lineages, the resulting differentiated cells maintain the eroded X (Xe).

Female pre-implantation blastocysts have two active X chromosomes and express *XIST*, serving as a unique scenario where *XIST* expression does not cause XCI. (vi) When hESCs are derived under naïve pluripotent culture conditions, or when primed hPSCs are adapted to such naïve conditions, the X-chromosome state of resulting hPSCs resembles that of pre-implantation blastocyst. (vii) Similar to normal development, differentiation of naïve hPSCs results in *XIST*-mediated XCI. *denotes the state found in majority of cells. hPSCs = human pluripotent stem cells, *XIST* = X inactive specific transcript, Xi = inactive X chromosome, Xa = active X chromosome, Xe = eroded X chromosome.

Figure 1-1: X-chromosome inactivation in different mammals

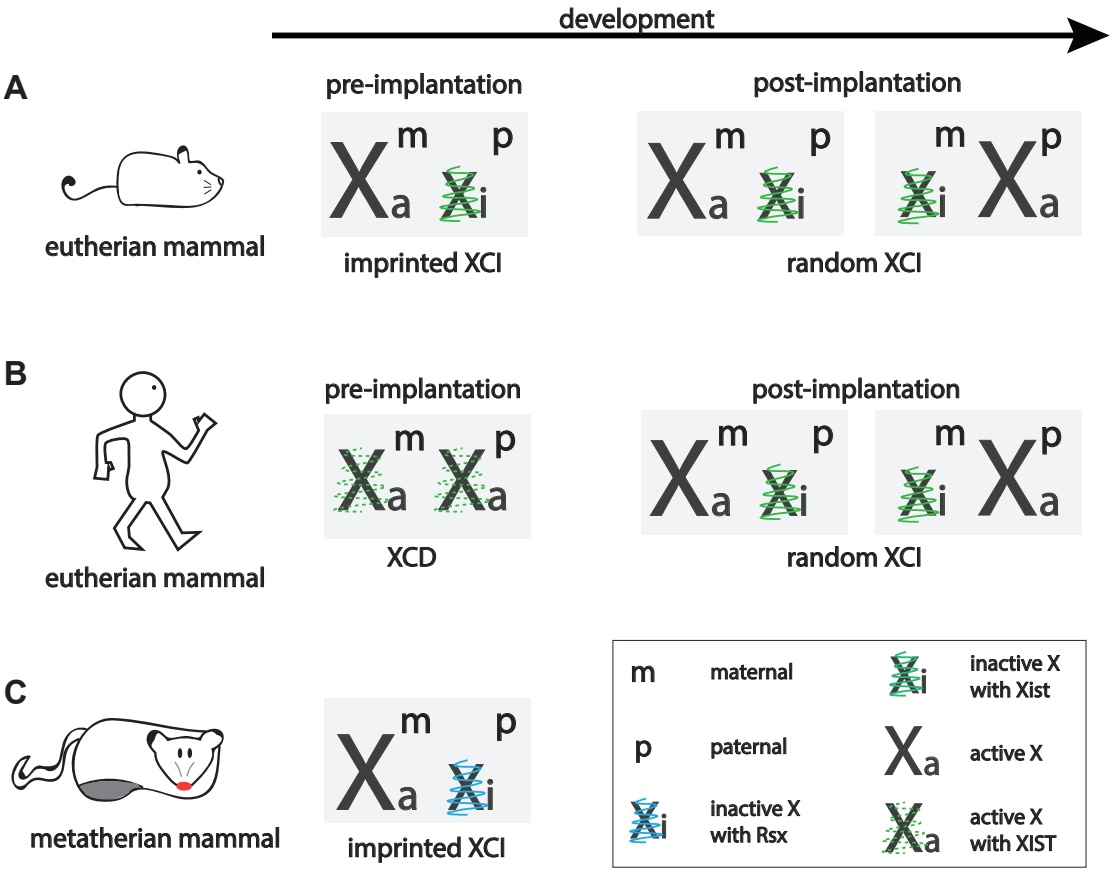


Figure 1-2: Long non-coding RNAs involved in X-chromosome dosage regulation in mouse and human

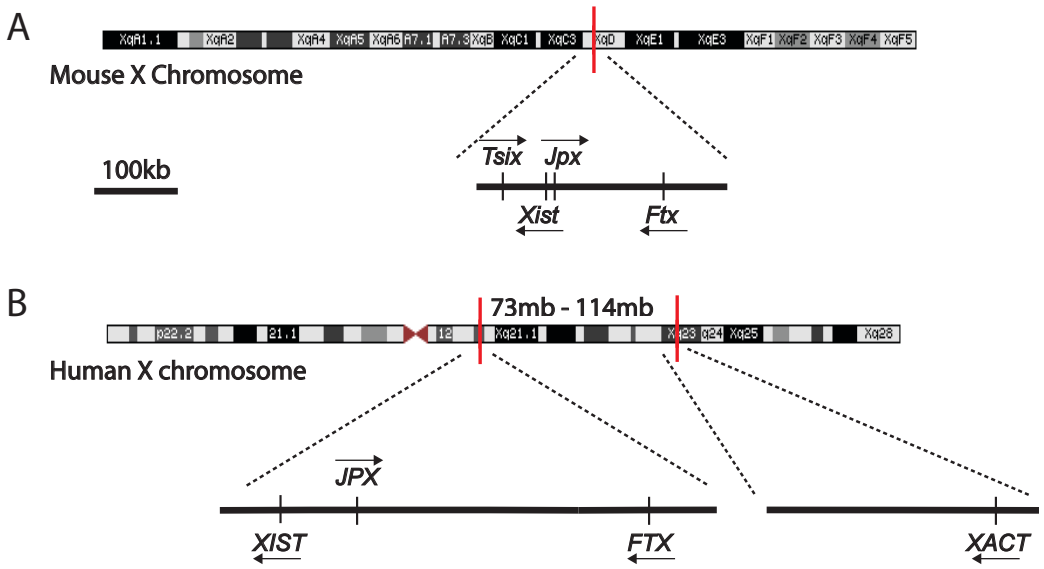


Figure 1-3: Lack of *Xist* at various developmental time-points highlights its importance in normal development

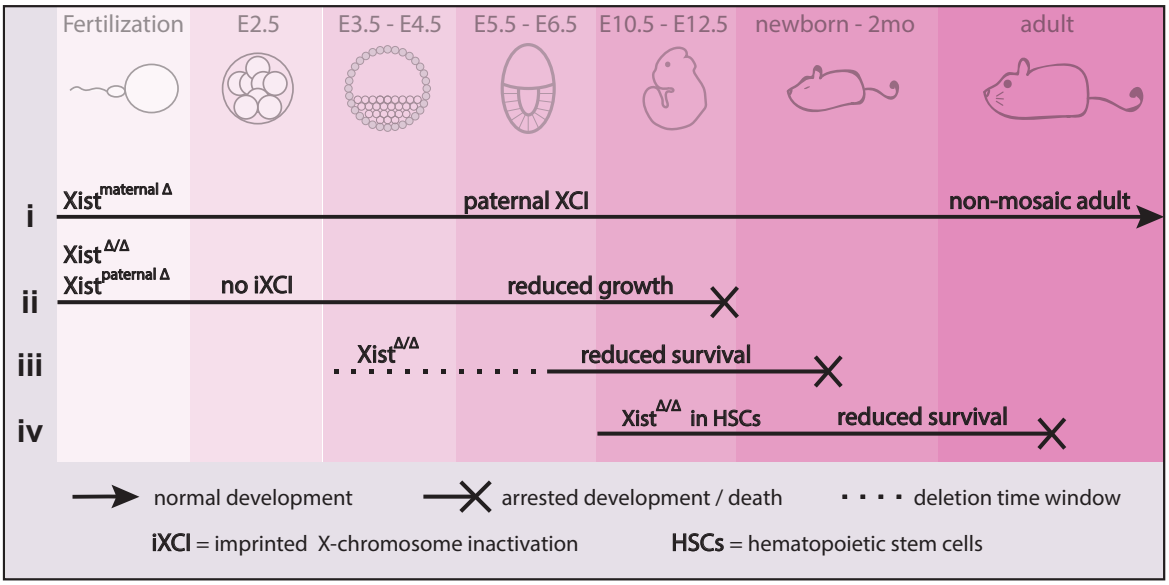
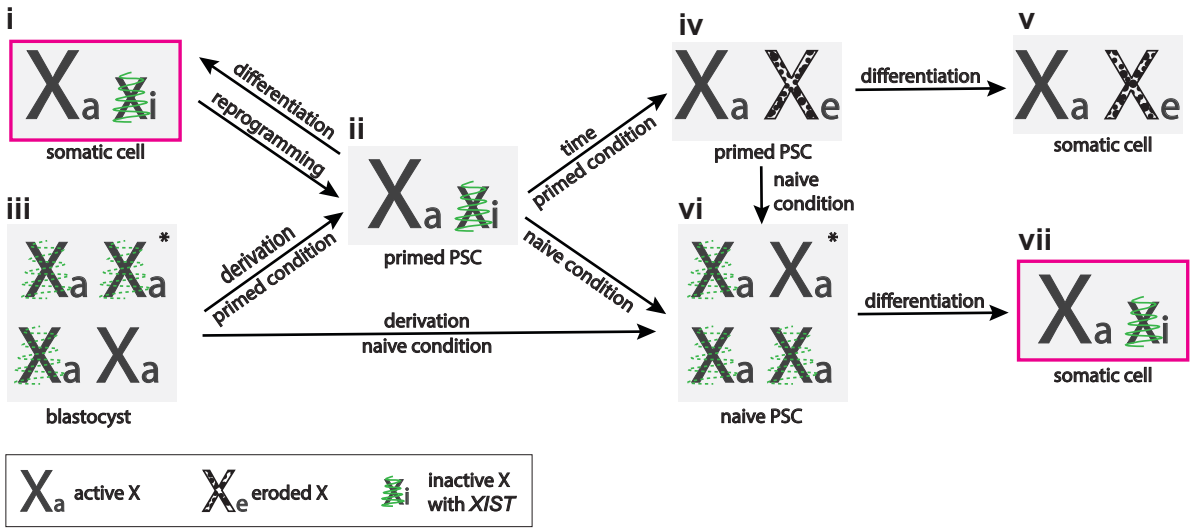


Figure 1-4: The X-chromosome state of naïve and primed human pluripotent stem cells



REFERENCES

1. Dixon-McDougall, T. & Brown, C. The making of a Barr body: the mosaic of factors that eXIST on the mammalian inactive X chromosome [1]. *Biochem. Cell Biol.* **94,** 56–70 (2016).

2. Wang, F. *et al.* Regulation of X-linked gene expression during early mouse development by Rlim. *eLife* **5,** (2016).

3. 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).

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

5. Wu, H. *et al.* Cellular Resolution Maps of X Chromosome Inactivation: Implications for Neural Development, Function, and Disease. *Neuron* **81,** 103–119 (2014).

6. Balaton, B. P. & Brown, C. J. Escape Artists of the X Chromosome. *Trends Genet.* **32,** 348–359 (2016).

7. Gendrel, A.-V. & Heard, E. Noncoding RNAs and Epigenetic Mechanisms During X-Chromosome Inactivation. *Annu. Rev. Cell Dev. Biol.* **30,** 561–580 (2014).

8. Grant, J. *et al.* Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487,** 254–258 (2012).

9. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472,** 370–374 (2011).

10. Hwang, J. Y., Oh, J.-N., Park, C.-H., Lee, D.-K. & Lee, C.-K. Dosage compensation of X-chromosome inactivation center-linked genes in porcine preimplantation embryos: Non-chromosome-wide initiation of X-chromosome inactivation in blastocysts. *Mech. Dev.* **138,** 246–255 (2015).

11. Wang, X., Miller, D. C., Clark, A. G. & Antczak, D. F. Random X inactivation in the mule and horse placenta. *Genome Res.* **22,** 1855–1863 (2012).

12. Moreira de Mello, J. C. *et al.* Random X Inactivation and Extensive Mosaicism in Human Placenta Revealed by Analysis of Allele-Specific Gene Expression along the X Chromosome. *PLoS ONE* **5,** e10947 (2010).

13. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165,** 1012–1026 (2016).

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

15. Meyer, B. J. Targeting X chromosomes for repression. *Curr. Opin. Genet. Dev.* **20,** 179–189 (2010).

16. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523,** 240–244 (2015).

17. Pandya-Jones, A. & Plath, K. The 'lnc' between 3D chromatin structure and X chromosome inactivation. *Semin. Cell Dev. Biol.* **56,** 35–47 (2016).

18. Albritton, S. E. & Ercan, S. Caenorhabditis elegans Dosage Compensation: Insights into Condensin-Mediated Gene Regulation. *Trends Genet.* **34,** 41–53 (2018).

19. Moreira de Mello, J. C., Fernandes, G. R., Vibranovski, M. D. & Pereira, L. V. Early X chromosome inactivation during human preimplantation development revealed by single-cell RNA-sequencing. *Sci. Rep.* **7,** (2017).

20. Vallot, C. *et al.* XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell* **20,** 102–111 (2017).

21. Sahakyan, A. *et al.* Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* **20,** 87–101 (2017).

22. Sun, S. *et al.* Jpx RNA Activates Xist by Evicting CTCF. *Cell* **153,** 1537–1551 (2013).

23. Carmona, S., Lin, B., Chou, T., Arroyo, K. & Sun, S. LncRNA Jpx induces Xist expression in mice using both trans and cis mechanisms. *PLOS Genet.* **14,** e1007378 (2018).

24. Chureau, C. *et al.* Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum. Mol. Genet.* **20,** 705–718 (2011).

25. Furlan, G. *et al.* The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products. *Mol. Cell* **70,** 462–472.e8 (2018).

26. 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).

27. Gayen, S., Maclary, E., Buttigieg, E., Hinten, M. & Kalantry, S. A Primary Role for the Tsix lncRNA in Maintaining Random X-Chromosome Inactivation. *Cell Rep.* **11,** 1251–1265 (2015).

28. Vallot, C. *et al.* XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45,** 239–241 (2013).

29. Jonkers, I. *et al.* Xist RNA Is Confined to the Nuclear Territory of the Silenced X Chromosome throughout the Cell Cycle. *Mol. Cell. Biol.* **28,** 5583–5594 (2008).

30. Engreitz, J. M. *et al.* The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* **341,** 1237973–1237973 (2013).

31. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472,** 120–124 (2011).

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

33. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349,** 38–44 (1991).

34. Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351,** 329–331 (1991).

35. Borsani, G. *et al.* Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351,** 325–329 (1991).

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

37. 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).

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

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

40. Mugford, J. W., Yee, D. & Magnuson, T. Failure of extra-embryonic progenitor maintenance in the absence of dosage compensation. *Dev. Camb. Engl.* **139,** 2130–2138 (2012).

41. Kalantry, S., Purushothaman, S., Bowen, R. B., Starmer, J. & Magnuson, T. Evidence of Xist RNA-independent initiation of mouse imprinted X-chromosome inactivation. *Nature* **460,** 647–651 (2009).

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

43. 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).

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

45. Yang, L., Kirby, J. E., Sunwoo, H. & Lee, J. T. Female mice lacking Xist RNA show partial dosage compensation and survive to term. *Genes Dev.* **30,** 1747–1760 (2016).

46. Rastan, S. Timing of X-chromosome inactivation in postimplantation mouse embryos. *J. Embryol. Exp. Morphol.* **71,** 11–24 (1982).

47. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21,** 70–71 (1999).

48. Hayashi, S., Lewis, P., Pevny, L. & McMahon, A. P. Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. *Mech. Dev.* **119 Suppl 1,** S97–S101 (2002).

49. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5,** 1513–1523 (1991).

50. Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. & Jaenisch, R. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat. Genet.* **22,** 323–324 (1999).

51. Yildirim, E. *et al.* Xist RNA Is a Potent Suppressor of Hematologic Cancer in Mice. *Cell* **152,** 727–742 (2013).

52. Nagy, A. *et al.* Embryonic stem cells alone are able to support fetal development in the mouse. *Dev. Camb. Engl.* **110,** 815–821 (1990).

53. 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).

54. Thornhill, A. R. & Burgoyne, P. S. A paternally imprinted X chromosome retards the development of the early mouse embryo. *Dev. Camb. Engl.* **118,** 171–174 (1993).

55. da Rocha, S. T. & Heard, E. Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nat. Struct. Mol. Biol.* **24,** 197–204 (2017).

56. Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161,** 404–416 (2015).

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

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

59. 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).

60. 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).

61. Chen, C.-K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354,** 468–472 (2016).

62. Patil, D. P. *et al.* m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* (2016). doi:10.1038/nature19342

63. Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell* **68,** 955–969.e10 (2017).

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

65. Shevchenko, A. I. *et al.* Impact of Xist RNA on chromatin modifications and transcriptional silencing maintenance at different stages of imprinted X chromosome inactivation in vole Microtus levis. *Chromosoma* **127,** 129–139 (2018).

66. Chaligné, R. & Heard, E. X-chromosome inactivation in development and cancer. *FEBS Lett.* **588,** 2514–2522 (2014).

67. Pasque, V. & Plath, K. X chromosome reactivation in reprogramming and in development. *Curr. Opin. Cell Biol.* **37,** 75–83 (2015).

68. Mak, W. *et al.* Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303,** 666–669 (2004).

69. de Napoles, M., Nesterova, T. & Brockdorff, N. Early Loss of Xist RNA Expression and Inactive X Chromosome Associated Chromatin Modification in Developing Primordial Germ Cells. *PLoS ONE* **2,** e860 (2007).

70. Pasque, V. *et al.* X Chromosome Reactivation Dynamics Reveal Stages of Reprogramming to Pluripotency. *Cell* **159,** 1681–1697 (2014).

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

72. Minkovsky, A. *et al.* The Mbd1-Atf7ip-Setdb1 pathway contributes to the maintenance of X chromosome inactivation. *Epigenetics Chromatin* **7,** 12 (2014).

73. Keniry, A. *et al.* Setdb1-mediated H3K9 methylation is enriched on the inactive X and plays a role in its epigenetic silencing. *Epigenetics Chromatin* **9,** (2016).

74. Tchieu, J. *et al.* Female Human iPSCs Retain an Inactive X Chromosome. *Cell Stem Cell* **7,** 329–342 (2010).

75. Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell* **4,** 487–492 (2009).

76. Shen, Y. *et al.* X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proc. Natl. Acad. Sci.* **105,** 4709–4714 (2008).

77. Mekhoubad, S. *et al.* Erosion of Dosage Compensation Impacts Human iPSC Disease Modeling. *Cell Stem Cell* **10,** 595–609 (2012).

78. Nazor, K. L. *et al.* Recurrent Variations in DNA Methylation in Human Pluripotent Stem Cells and Their Differentiated Derivatives. *Cell Stem Cell* **10,** 620–634 (2012).

79. Vallot, C. *et al.* Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. *Cell Stem Cell* **16,** 533–546 (2015).

80. Patel, S. *et al.* Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. *Cell Rep.* **18,** 54–67 (2017).

81. Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504,** 465–469 (2013).

82. Bruck, T., Yanuka, O. & Benvenisty, N. Human pluripotent stem cells with distinct X inactivation status show molecular and cellular differences controlled by the X-Linked ELK-1 gene. *Cell Rep.* **4,** 262–270 (2013).

83. Sahakyan, A., Plath, K. & Rougeulle, C. Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. *Philos. Trans. R. Soc. B Biol. Sci.* **372,** 20160363 (2017).

84. Geens, M. *et al.* Female human pluripotent stem cells rapidly lose X chromosome inactivation marks and progress to a skewed methylation pattern during culture. *Mol. Hum. Reprod.* **22,** 285–298 (2016).

85. Bell, S., Kolobova, I., Crapper, L. & Ernst, C. Lesch-Nyhan Syndrome: Models, Theories, and Therapies. *Mol. Syndromol.* **7,** 302–311 (2016).

86. Theunissen, T. W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15,** 471–487 (2014).

87. Takashima, Y. *et al.* Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* **158,** 1254–1269 (2014).

88. Pastor, W. A. *et al.* Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell* **18,** 323–329 (2016).

89. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* (2016). doi:10.1016/j.stem.2016.06.011

90. Inoue, A., Jiang, L., Lu, F. & Zhang, Y. Genomic imprinting of *Xist* by maternal H3K27me3. *Genes Dev.* **31,** 1927–1932 (2017).

91. Guyochin, A. *et al.* Live Cell Imaging of the Nascent Inactive X Chromosome during the Early Differentiation Process of Naive ES Cells towards Epiblast Stem Cells. *PLoS ONE* **9,** e116109 (2014).

92. Plath, K., Zhang, Y., Panning, B. & Otte, A. P. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science* **300,** 131–135 (2003).

93. Jiang, J. *et al.* Translating dosage compensation to trisomy 21. *Nature* **500,** 296–300 (2013).

CHAPTER 2

Synergistic Action of Ribonucleotide Reductase Inhibition

and DNA Demethylation Identified by an

X-chromosome Reactivation Screen

SUMMARY

Background: DNA methylation is important for the maintenance of the silent state of genes on the inactive X chromosome (Xi). Here, we screened for siRNAs and chemicals that reactivate an Xi-linked reporter in the presence of 5-aza-2′-deoxycytidine (5-aza-2′-dC), an inhibitor of DNA methyltransferase 1, at a concentration that, on its own, is not sufficient for Xi-reactivation.

Results: We found that inhibition of ribonucleotide reductase (RNR) induced expression of the reporter. RNR inhibition potentiated the effect of 5-aza-2′-dC by enhancing its DNA incorporation, thereby decreasing DNA methylation levels genome-wide. Since both 5-aza-2′-dC and RNR-inhibitors are used in the treatment of hematological malignancies, we treated myeloid leukemia cell lines with 5-aza-2′-dC and the RNR-inhibitor hydroxyurea, and observed synergistic inhibition of cell growth and a decrease in genome-wide DNA methylation.

Conclusions: Taken together, our study identifies a drug combination that enhances DNA demethylation by altering nucleotide metabolism. This demonstrates that Xi-reactivation assays can be used to optimize the epigenetic activity of drug combinations.

BACKGROUND

X-chromosome inactivation (XCI) is a program of transcriptional gene silencing that occurs on one of two X chromosomes in female mammalian cells to equalize gene dosage of X-linked genes to male cells. The inactive X chromosome (Xi) is a striking example of developmentally regulated heterochromatin formation in mammals. XCI has served as paradigm for understanding factors with generalized roles in gene silencing genome- wide such as DNA methylation and Polycomb protein-mediated histone methylation [1–3]. The Xi is established early in female embryonic development through a series of stepwise molecular changes that cooperate to ensure stable chromosome-wide gene silencing. Once established, the Xi is inherited through all somatic cell divisions and adult life [1–3]. XCI is initiated by the upregulation of the long noncoding RNA *Xist* from the maternal or paternal X chromosome early in embryonic development [1–3]. *Xist* coats the X chromosome from which it is expressed and initiates a cascade of events including exclusion of RNA polymerase II, changes in histone marks, and recruitment of structural chromo- some proteins [1–3]. Accumulation of the histone variant macroH2A1 and gain of CpG island methylation characterize the transition to the maintenance phase of XCI, which is marked by resistance to X chromosome reactivation (XCR) upon deletion of *Xist* [4–9]. Thus, *Xist* is absolutely required for the initiation of XCI, but later is largely dispensable for the maintenance of the Xi, due to the presence of various other repressive chromatin marks [8, 9]. Notably, complete XCR is induced in vivo during pre-implantation and germ line development and in vitro by reversing cellular identity to the pluripotent state [10–13].

Despite the observation that many repressive chromatin factors are implicated in Xi establishment and maintenance, interference with DNA methylation has thus far shown the largest effect on eliciting loss of gene silencing on the Xi [5, 9, 14]. It is therefore thought that DNA

methylation may uniquely 'lock-in' the silenced state and execute a greater influence on the robust nature of Xi maintenance than other repressive regulatory mechanisms [9]. DNA methylation concentrates on CpG islands in the course of XCI with redistribution away from intra- genic and intronic CpGs relative to the active X chromosome [5, 7, 15, 16, 17]. CpG island methylation on the Xi is established by the de novo methyltransferase DNMT3B and is subsequently propagated by the maintenance methyltransferase DNMT1 [5, 9, 15]. Interference with DNA methylation by deletion of *Dnmt1* or treatment with 5-aza-2′-deoxycytidine (5-aza-2′-dC, also called decitabine) has been shown to induce the reactivation of an Xi-linked reporter gene and endogenous X-linked genes in a proportion of female somatic cells [9]. 5-aza-2′-dC is a deoxycytidine analog that upon phosphorylation incorporates into DNA and irreversibly inhibits DNMT1 [18]. Subsequent rounds of DNA replication therefore lead to passive DNA demethylation due to the absence of DNMT1 activity [19]. Together these findings indicate that Xi reporter systems permit the functional analysis of gene silencing, and that in addition to DNA methylation various other mechanisms contribute to Xi silencing. Therefore, XCI is an attractive model system to probe therapeutic approaches to the reactivation of silenced genes.

In the field of cancer biology, there is growing appreciation that abnormalities in histone modification and DNA methylation pathways can drive tumorigenesis across many cancer types and there is promise for improved therapies aimed at reversal of gene silencing [20]. In this study, we bridge the study of the Xi with the development of strategies to more efficiently demethylate and reactivate silenced genes. 5-aza-2′-dC is used clinically in the setting of hematologic malignancies with the rationale of reactivating silenced genes [19]. The drug is currently approved for the treatment of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) [20]. Several studies have confirmed that 5-aza-2′-dC at low doses elicits genome-wide DNA

demethylation in AML patient samples [21–23]. One approach to increase the epigenetic activity of 5-aza-2′-dC in myeloid malignancy is to use it in combination with other agents known to elicit reactivation of silenced genes, such as histone deacetylase inhibitors [20]. Notably, for the Xi, such co-treatment approaches increase the rate of XCR in cell culture systems [9]. The similar efficacy of 5-aza-2′-dC alone or in combination with other chromatin-modifying agents in Xi-linked genes and in myeloid leukemia supports the translation of findings from X-chromosome inactivation to epigenetic cancer therapies.

Here, we set out to find additional pathways that in combination with 5-aza-2′-dC, elicit XCR. Specifically, we applied high-throughput siRNA and chemical screening to identify factors that could reactivate a silent reporter transgene that is specifically located on the Xi. Our screen employed treatment with a low dose of 5-aza-2′-dC to sensitize somatic cells for DNA demethylation and XCR, which on its own is not sufficient to induce these effects. We identified that inhibition of the ribonucleotide reductase protein complex significantly enhances the DNA demethylation action of 5-aza-2′-dC and hence the activity of the Xi-reporter. We characterize the mechanism of action as increasing DNA incorporation of 5-aza-2′-dC and thus its demethylating activity. While our approach initially centered on the Xi, we found a pathway that altered DNA methylation levels genome-wide. Our study therefore demonstrates that assays of XCR can be adapted to optimize the epigenetic activity of a DNA demethylating drug combination.

RESULTS AND DISCUSSION

An siRNA Screen for XCR in the Presence of a Low 5-aza-2′-dC Dose Identifies the Ribonucleotide Reductase Pathway

Previous work from our lab has shown that an Xi-linked, CAG promoter-driven luciferase transgene in the *Hprt* locus (Xi-luciferase) is a sensitive reporter of gene silencing on the Xi when tested in primary mouse embryonic fibroblasts (MEFs) [24]. Our Xi-luciferase MEFs faith- fully inactivate the luciferase-bearing X chromosome in female embryonic development rather than under- going random XCI because an *Xist* deletion on the other X chromosome forces XCI on the chromosome carrying the wild-type *Xist* allele [25] (Fig. 2-1a). The Xi-luciferase gene body and promoter are highly methylated at the DNA level and Xi-luciferase reporter MEFs increase luciferase activity in a dose-dependent fashion in response to 5-aza-2′-dC treatment [24]. Here, we used Xi-luciferase reporter MEFs to screen for gene knock- downs or chemicals that could elicit XCR.

In order to perform a high-throughput screen for XCR, we established an siRNA knockdown assay in 384-well format, with each individual siRNA tested in a single well. As positive control, we chose knockdown of *Dnmt1* since interference with *Dnmt1* either by knockout or 5-aza-2′-dC treatment has previously been described to elicit XCR in MEFs [9, 24]. Initially, we tested increasing concentrations of 5-aza-2′-dC in combination with *Dnmt1* knockdown to determine a 5-aza-2′-dC concentration for which the depletion of *Dnmt1* by siRNAs yielded robust reporter reactivation, but where 5-aza- 2′-dC treatment alone does not elicit reactivation. Our titration experiment demonstrated that the combination of 5-aza-2′-dC at a concentration ranging from 0.1 to 0.2 μM, along with si*Dnmt1* treatment, enhanced luciferase activity in the 384-well

format. Importantly, 5-aza-2′-dC treatment in this concentration range or knockdown of *Dnmt1* alone did not induce a significant difference in luciferase signal compared to untreated wells (Fig. S2-1A). The requirement for 5-aza-2′-dC co-treatment with *Dnmt1* knockdown to detect XCR likely reflects that *Dnmt1* knockdown alone does not lead to sufficient levels of XCR detected in the small 384-well format assay. By comparison, a higher dose of 5-aza-2′-dC (1 μM) elicited strong reactivation of the Xi-linked luciferase reporter on its own that was not as dramatically enhanced by si*Dnmt1* treatment (Fig. S2-1A). Thus, a low dose of 5-aza- 2′-dC has a sensitizing effect on eliciting XCR by *Dnmt1* knockdown. The interaction of 5-aza-2′-dC with other Xi maintenance factors indicates a similar sensitizing effect with respect to XCR. For instance, the knock- down of the candidate Xi-maintenance factor *Atf7ip* or deletion of *Xist* produces a low rate of XCR that is significantly boosted by the addition of 5-aza-2′-dC [9, 24]. Therefore, we extended the low concentration 5-aza- 2′-dC treatment (0.2 μM) to the entire genome-wide siRNA screen with the rationale that knockdown of other chromatin-modifying factors may require concur- rent DNA demethylation to produce strong Xi-luciferase reporter reactivation.

We performed a genome-wide mouse siRNA screen with 51,150 siRNAs against 21,114 genes on 153 384-well plates (see “Methods” section for details on the library used) in duplicate using female Xi-linked luciferase reporter MEFs in the presence of low 5-aza-2′-dC. We measured luciferase levels 72 h after siRNA transfection (Fig. 2-1b). To eliminate batch effects, we normalized luminescence data by 384-well plate, and then analyzed the data by prioritizing gene hits with multiple active siRNAs by redundant siRNA activity (RSA) analysis [26] (Fig. 2-1c and 2-S2). Notably, *Dnmt1* was the top hit in our genome-wide screen, which provided internal validation of the method (Fig. 2-1c). Further support came from another hit, identified as *Atf7ip*, which our group recently reported as a maintenance factor in XCI [24]. As with other previously

described maintenance factors, we found that the Xi-luciferase signal in response to knock- down of *Atf7ip* was greatly increased by low 5-aza-2′-dC (0.2 μM) co-treatment [24]. Identification of *Atf7ip* in the screen supports the strategy of 5-aza-2′-dC co-treatment to unmask functional contribution of Xi-maintenance factors.

To select novel hits, we chose the top 54 genes from the RSA analysis with at least two unique active siRNAs inducing an increase in luciferase levels in the 384-well screen, omitting genes we deemed irrelevant such as those for olfactory receptors, and retested the active siRNAs sequences from the library. Several of these siRNAs showed reproducible increases in luciferase activity in the validation assay (Fig. 2-S3). We decided to focus on *Rrm2* as a hit since one siRNA against it had produced the next highest level of luciferase activity in the validation assay after the siRNAs targeting *Atf7ip* or *Dnmt1*. Follow-up assays with a greater number of starting cells demonstrated an increase in luciferase activity for each of our three different siRNAs against the ribonucleotide reductase (RNR) M2 subunit gene (*Rrm2*) (Fig. 2-1d). The luminescence generated with si*Rrm2* treatment was in proportion to individual extent of *Rrm2* knockdown, suggesting specificity of *Rrm2* targeting for the XCR effect (Fig. 1d).

As part of the RNR enzyme complex, RRM2 catalyzes the conversion of ribonucleoside 5′-disphosphates to their 2′-deoxyribonucleoside form in the rate-limiting step of de novo dNTP biosynthesis [27]. The RRM2 subunit, which was identified in this siRNA screen, is specifically upregulated at S phase of cell cycle and is necessary for the activity of the RNR complex [27]. Since we identified si*Rrm2* in combination with 5-aza-2′-dC (0.2 μM) in the genome-wide screen, we next asked if knockdown of *Rrm2* could also elicit XCR in the absence of 5-aza- 2′-dC, since interference with *Atf7ip* or *Xist* produces a low rate of XCR that is significantly boosted by the addition of 5-aza-2′-dC [9, 24]. However, unlike previously described Xi-maintenance factors, we

did not find that si*Rrm2* produced luciferase activity in the absence of 5-aza-2′-dC (Fig. 2-1d). These results were reproduced with an Xi-linked fluorescent reporter (Fig. S2-4B/C) [9]. We conclude that low doses of 5-aza- 2′-dC are necessary for the XCR effect of si*Rrm2* identified by our genome-wide screen for factors involved in the maintenance of Xi silencing.

Chemical Inhibitors of Ribonucleotide Reductase Elicit XCR in the Presence of 5-aza-2′-dC

We used a complimentary approach to further probe the pathways contributing to Xi maintenance by performing a companion screen analogous to the siRNA screen but instead using a collection of annotated chemicals (Fig. 2-1b and S2-5A). In the screen, we found that resveratrol, a chemical agent known for mimicking cellular effects of caloric restriction, demonstrated the potential to activate the Xi-luciferase reporter (Fig. 2-1e) [28]. To validate the screening result, we tested the effect of various resveratrol concentrations on the Xi-luciferase reporter. The bell-shaped dose–response activity of resveratrol in combination with fixed, low concentration of

5-aza-2′-dC (0.2 μM) indicated a maximal XCR activity at a concentration of 20 μM (Fig. 2-1f). In order to confirm an XCR-specific effect, we tested whether the combination of resveratrol with 5-aza-2′-dC could reactivate different Xi-linked reporters. We found that 20 μM resveratrol and 5-aza-2′-dC (0.2 μM) together could also reactivate two fluorescent reporters including the CAG- driven H2B citrine transgene within the *Hprt* locus and a distal CAG-driven GFP transgene (Fig. S2-4) [9, 24]. As previously observed, the proportion of cells expressing the reporter differs for Xi-CAG-H2B- citrine and Xi-GFP MEFs likely owing to different silencing requirements of the two loci on the Xi [24].

Resveratrol is a naturally-occurring polyphenolic molecule believed to have numerous direct intracellular protein targets [29]. It is described to mediate its metabolic effects through direct and indirect activation of the histone deacetylase SIRT1 though no specific role in reversal of chromatin silencing or effects on the Xi has been described [30–32]. Of note, we did not find that knockdown of *Sirt1* attenuated the ability of resveratrol with 5-aza-2′-dC to elicit Xi-luciferase reactivation (not shown). A further search for the cellular target of resveratrol in XCR led us to a study that described resveratrol as an inhibitor of RNR, the same enzyme complex that we identified as a hit in the genome-wide siRNA screen for XCR described above [33]. This link between our complimentary screening approaches pointed to resveratrol's role in XCR in the presence of 5-aza-2′-dC by means of RNR inhibition.

In order to further investigate whether RNR is the target of resveratrol in eliciting XCR, we tested a well- characterized inhibitor of RNR, hydroxyurea (HU), and found that it also increased Xi-luciferase activity in the presence of a low dose of 5-aza-2′-dC (0.2 μM) (Fig. 2-1g). From the titration, HU had a maximum effect on Xi-luciferase reactivation at 50 and at 200 μM and a fading effect at 20 μM (Fig. 2-1g). HU was not detected from the chemical library because it was assayed at a screening concentration of 10 μM, which was probably insufficient concentration to detect activity. 50 μM HU treatment in combination with low 5-aza-2′-dC also induced reactivation of the Xi-linked GFP (Fig. S2-4B/C). We reasoned that if resveratrol and HU converge on inhibition of RNR, that the XCR effect of resveratrol and HU should require the co-treatment with 5-aza-2′-dC as seen for the *Rrm2* knockdown. Indeed, we found that similar to the si*Rrm2* condition, resveratrol and HU treatment demonstrated a complete dependence on low levels of 5-aza-2′-dC to elicit XCR (Fig. 2-1d/f/g and S2-4B/C). We conclude that RNR inhibition alone by these various means does not increase Xi-luciferase activity.

Since the Xi-luciferase reporter assay is not reflective of cell number, we measured protein concentration in luciferase assay lysates to rule out variable cell number due to different treatments as an explanation for lack of Xi-reporter activation in the absence of 5-aza-2′-dC (Fig. S2-6A). Variations in protein lysate concentration were minor across all treatments, indicating RNR inhibition requires the presence of 5-aza-2′-dC to elicit its effect on XCR. The identification of chemicals with RNR-inhibiting activity added support to the XCR role of inhibiting this pathway in the presence of 5-aza-2′-dC.

We further investigated the relationship between 5-aza-2′-dC and RNR inhibition in XCR by querying whether 5-aza-2′-dC can be replaced by knockdown of *Dnmt1*. In previous studies where 5-aza-2′-dC had a sensitizing effect towards XCR, the effect is attributable to interference with *Dnmt1* [24]. For instance, 5-aza-2′-dC treatment could be substituted by knock- down of *Dnmt1* to elicit synergistic XCR by *Atf7ip* knockdown [24]. Contrary to these prior findings, we found that *Dnmt1* depletion by siRNAs did not replace the contribution of 5-aza-2′-dC to XCR induced by RNR-inhibition via resveratrol (Fig. 2-1h). Together, these findings suggest a mechanism of action whereby RNR inhibition specifically affects the action of cytidine analog 5-aza-2′-dC.

RNR Inhibition Increases Incorporation of 5-aza-2′-dC into DNA

Next, we sought to understand how RRM2 inhibition interacts with low amounts of 5-aza-2′-dC to elicit XCR. The pool of dNTPs in the nucleus is tightly regulated and studies have speculated that RNR inhibition can increase the likelihood of nucleoside analog DNA incorporation by reducing the pools of endogenous nucleotide concentrations [27, 34]. Accordingly, we postulated

that RRM2 inhibition may increase 5-aza-2′-dCTP concentration in the nucleus relative to the endogenous dCTP pool, leading to more 5-aza-2′-dCTP DNA incorporation (Fig. 2-2a). Higher rates of 5-aza-2′-dC incorporation into DNA subsequently could lead to greater DNA demethylation and XCR (Fig. 2-2a).

Consistent with this model, we observed that knock- down of *Rrm2* or resveratrol treatment reproducibly increased the amount of tritiated 5-aza-2′-dC incorporated into DNA approximately by two-fold (Fig. 2-2b). We further tested the role of the ratio of 5-aza-2′-dCTP to endogenous dCTP by the converse manipulation of increasing dCTP relative to 5-aza-2′-dC. This experiment was performed by adding increasing concentrations of deoxycytidine (dC) into media, which is metabolized to dCTP within the cell, in the presence of 5-aza-2′-dC with resveratrol treatment (Fig. 2-2c) or *Rrm2* knockdown (Fig. 2-2d). Importantly, dC does not require the action of RNR for DNA incorporation. Our expectation was that an increase in dCTP levels in the cell would reduce the incorporation of 5-aza-2′-dC into the DNA, and there- fore reduce the reactivation of the Xi-linked luciferase reporter. As expected, the luciferase signal decreased in a dose-dependent fashion when exogenous deoxycytidine was supplied in the media (Fig. 2-2c/d). The loss of the Xi-reporter reactivation is consistent with the notion that the relative nuclear concentration of 5-aza-2′-dCTP to dCTP is shifted by the addition of an exogenous nucleotide substrate to reduce the effective concentration of the 5-aza-2′-dC analog (Fig. 2-2c/d).

An alternate explanation for the observed decrease in luciferase signal upon addition of dC is a reduction in cell number. To rule out possible nucleotide treatment- dependent cell growth effects, we confirmed that protein concentrations in lysates were similar for the various treatment conditions (Fig. S2-6B/C). Furthermore, we used uridine as a control because it is a nontoxic precursor of pyrimidine synthesis that, like deoxycytidine, can be taken up by cells and used as a

substrate via the nucleoside salvage synthetic pathway (Fig. 2-2c/d) [35]. However, unlike deoxycytidine, uridine requires reduction by RNR in order to contribute to dNTP pools [35]. Increasing levels of uridine did not alter Xi-luciferase levels and thereby XCR in the presence of 5-aza-2′-dC with *Rrm2* knockdown and resveratrol treatment, respectively, compared to control (Fig. 2-2c/d and S2-6B/C). These results support the role of deoxycytidine in reversing the XCR effect downstream of RNR.

In summary, RRM2/RNR inhibition was identified in the XCR screen because it augmented 5-aza-2′-dC DNA incorporation. This mechanism is consistent with the observation that RNR inhibition alone, i.e. in the absence of 5-aza-2′-dC, did not produce measurable Xi-reporter reactivation in prior assays.

RRM2 Inhibition Enhances Genome-wide Demethylation Caused by 5-aza-2′-dC in MEFs

If RRM2 inhibition potentiates low dose 5-aza-2′-dC action to increase XCR by increasing the incorporation of 5-aza-2′-dC, then DNA methylation levels in cells treated with a low dose of 5-aza-2′-dC with RRM2 inhibition should approximate those of cells treated with a high dose of 5-aza-2′-dC. We investigated DNA methylation patterns at genome-scale by reduced representation bisulfite sequencing (RRBS) [36]. Specifically, MEFs were treated with si*Rrm2* or resveratrol alone, low or high doses of 5-aza-2′-dC, and combinations of si*Rrm2* or resveratrol with a low dose of 5-aza-2′-dC (Fig. 2-3, S2-7, and S2-8).

As expected, based on global methylation averages, hierarchical clustering, and methylation distributions, the treatment of MEFs with a low dose of 5-aza-2′-dC (0.2 μM) induced

a smaller reduction in the level of genome-wide methylation than the high dose of 5-aza- 2′-dC (10.0 μM), which resulted in marked demethylation compared to control samples (Fig. 3a/b/d/ and S2-7C/D). We found that treatment with si*Rrm2* or resveratrol alone (without 5-aza- 2′-dC) marginally increased global DNA methylation levels compared to untreated samples (Fig. 2-3a/b/d and S2-7C/D). Notably, the combination of *Rrm2* knockdown or resveratrol with the low dose 5-aza-2′-dC reduced global methylation to a similar extent as the high dose 5-aza-2′-dC treatment (Fig. 2-3a/b/d and S2-7C/D). These effects on the methylation profile were similar for autosomes and the X chromosome at the global scale (Fig. 2-3b and S2-8A) as well as on promoters and CpG islands (Fig. S2-7A and S2-8B). These findings are consistent with a genome-wide effect on DNA methylation rather than an Xi-specific mechanism, owing to increased DNA incorporation of 5-aza- 2′-dC under RRM2 inhibition conditions. We observed that CpGs with the highest levels of methylation in the control samples showed the most dramatic 5-aza-2′-dC- induced demethylation (Fig. S2-7B). For CpGs with lower methylation levels in the untreated conditions, demethylation due to 5-aza-2′-dC incorporation is still visible but less extensive (Fig. 2-3c and S2-7B). We believe that the greater apparent effect in highly methylated regions does not represent a predilection of 5-aza-2′-dC for highly methylated regions, as has been previously suggested [21], but rather that the random incorporation of 5-aza-2′-dC disproportionately affects the methylation estimates of highly methylated sites.

We also extracted the available methylation data for the Xi-linked luciferase reporter gene to determine whether the methylation levels correlated with the extent of Xi-luciferase reactivation in the various conditions. We found that CpG sites within the luciferase reporter gene followed the genome-wide methylation changes, and that the low 5-aza-2′-dC treatment together with RNR inhibition, by either *Rrm2* knock- down or resveratrol, induced similar demethylation as the high

dose of 5-aza-2′-dC (Fig. S2-9). The concordant behavior of Xi-luciferase reporter CpG sites supports the conclusion that the augmentation of DNA incorporation of 5-aza-2′-dC describes the *Rrm2* result in our Xi-reporter reactivation screen.

Taken together, our genome-wide methylation analysis for low dose 5-aza-2′-dC with RRM2 inhibition supports the idea that RRM2 inhibition increases the effective concentration of 5-aza-2′-dC and thereby its DNA incorporation, leading to global DNA demethylation.

Hydroxyurea and 5-aza-2′-dC Synergistically Inhibit Myeloid Leukemia Cell Line Proliferation in a Dose-dependent Fashion

Given that RRM2 inhibition increases DNA incorporation of 5-aza-2′-dC, we next applied the combination of RRM2 inhibition and 5-aza-2′-dC to a disease model in which 5-aza-2′-dC has therapeutic relevance. 5-aza-2′-dC is an FDA-approved drug and commonly used off-label in the setting of acute myeloid leukemia (AML) [20]. Therefore, we tested the drug combination in four myeloid leukemia cell lines (THP1, U937, K562, HL60) (Fig. 2-4 and S2-10). We hypothesized that, since RRM2 inhibition increased DNA incorporation of 5-aza-2′-dC, the combination of RRM2 inhibition with 5-aza-2′-dC could improve the therapeutic index of 5-aza-2′-dC, allowing lower doses to maximize demethylation activity with fewer cytotoxic off-target effects. We chose to use HU as the form of RRM2 inhibition because it also is an FDA-approved agent commonly used off-label for cyto- reductive purposes, also in the setting of AML [37].

To assess the effect of combining HU and 5-aza-2′-dC on myeloid leukemia cell line proliferation, we applied a luminescence-based cell viability assay that linearly scales with cell

number and titered HU and 5-aza-2′-dC individually to determine IC50 values for each cell line (Fig. S2-10A). For 5-aza-2′-dC, IC50 values were difficult to approximate given a plateau in cell proliferation changes at higher concentrations (not shown). Thus we chose the 5-aza-2′-dC concentration corresponding to halfway to the point of plateau effect. We then combined HU and 5-aza-2′-dC at fixed ratios, empirically determined for each of the four myeloid leukemia cell lines (Fig. 2-4a/b and S2-10B/C). In each of the four cell lines tested, the combination treatment inhibited cell proliferation more than either treatment alone. In order to make a quantitative determination of the drug interaction, we calculated Chou-Talalay Combination Indices (CI) where CI <1, =1, >1 indicate synergism, additive effect, and antagonism, respectively [38]. The combination of HU and 5-aza-2′-dC demonstrated evidence of drug synergism across a range of fixed drug concentration ratios in the four cell lines tested (Fig. 2-4a/b and S2-10B/C). We repeated the drug treatments with K562 cells in a soft agar assay and confirmed the syn- ergistic effect of HU and 5-aza-2′-dC on clonal cell expansion (Fig. S2-10D). Consistent with the proliferation studies, the combination HU and 5-aza-2′-dC reduced colony formation to a greater extent compared to either treatment alone. Together, these results demonstrate a synergistic interaction between HU and 5-aza- 2′-dC in the control of cell proliferation.

We next assessed whether DNA demethylation related to the synergistic drug effect observed. Specifically, we determined the DNA methylation profile of K562 cells treated at a low, mid, and high concentration of 5-aza- 2′-dC and HU at a fixed ratio by RRBS (Fig. 2-4c and S2-11). The low average genome-wide CpG methylation levels of approximately 35 % in K562 cells with few highly methylated CpGs is consistent with a prior studies reporting overall global hypomethylation inherent to K562 cells (Fig. S2-11, DMSO-treated control conditions) [39]. Nonetheless, treatment with a fixed ratio of low HU and 5-aza-2′-dC concentrations, that induced

a synergistic effect on cell growth (Fig. 2-4a, low condition), reduced DNA methylation compared to the low 5-aza-2′-dC treatment alone (Fig. 2-4c). As expected, HU treatment alone did not alter

DNA methylation levels (Fig. 2-4c). As with MEFs, filtering by CpGs that are highly methylated in control conditions best displayed the enhancing effect of low HU to low 5-aza-2′-dC concentrations (Fig. 2-4c and S2-11A). Unexpectedly, methylation levels did not appreciably decrease and even increased with the higher dose combinations of HU and 5-aza-2′-dC (mid and high treatment combinations) (Fig. 2-4c). Particularly at the high concentration combination, HU addition almost completely blunted the effect of 5-aza-2′-dC on methylation (Fig. 2-4c and S2-11). We hypothesized that the differing effects of the low and high concentration combinations may be due to interference of cell cycle progression with increasing concentrations of HU, which in turn interferes with the incorporation of 5-aza-2′-dC into DNA during DNA replication. Accordingly, flow cytometry analysis revealed a significant cell-cycle arrest of K562 cells at the high HU concentration, but not at the low concentration (Fig. 2-4d).

Our data suggest that at lower concentrations, HU and 5-aza-2′-dC act synergistically on cell growth, at least partially via DNA demethylation, while at higher concentrations, direct effects on cell cycle progression inhibit cell growth not allowing DNA demethylation via 5-aza-2′-dC, which is replication dependent. Regard- less, these data indicate that the combination of HU and 5-aza-2′-dC synergistically decreases cell proliferation of the four myeloid leukemia cell lines tested. Moreover, the mechanism of action of this synergistic drug combination changes in a dose-dependent fashion.

CONCLUSIONS

Using an Xi-linked luciferase reporter sensitized to reactivate by low concentration 5-aza-2′-dC treatment, we screened genome-wide siRNA and chemical libraries for reactivation activity. We found that inhibition of the RRM2 subunit of the ribonucleotide reductase enzyme increases rates of Xi-linked reporter reactivation. We attribute the effect of RRM2 inhibition on the Xi in MEFs to augmentation of 5-aza-2′-dC incorporation into DNA, which in turn induces increased genome-wide DNA demethylation in a pattern similar to a high dose 5-aza-2′-dC treatment alone. Moreover, treatment of myeloid leukemia cells with 5-aza- 2′-dC and the RRM2-inhibitor HU together synergistically inhibited cell proliferation and altered DNA methylation levels in these cancer cell lines in a dose-dependent manner. These findings suggest RRM2-inhibitors improve the demethlyation activity of 5-aza-2′-dC and may have clinical benefit if used in combination.

Our screen utilized a single copy Xi-linked reporter to identify the effect of RRM2 inhibition, which was then characterized as a genome-wide effect of augmenting 5-aza-2′-dC-mediated demethylation. The extension of our findings from a single gene reporter on the Xi to a genome-wide effect indicates that the Xi can be used as a model system for identifying and targeting general mechanisms of gene silencing. The optimal dose-schedule of 5-aza-2′-dC remains to be determined and the most effective epigenetic therapy will likely require use of 5-aza-2′ dC in combination with other epigenetic agents [40]. The XCR assay may be helpful to accomplish these objectives. The robust nature of Xi silencing in differentiated cells, however, contributes to one of the challenges of high-throughput screening with this model: XCR is partial and occurs at low rates, thus XCR assays must be optimized in sensitivity. Previous Xi maintenance screens have used pooled shRNA libraries in combination with immortalized Xi-GFP transgene-bearing reporter fibro- blasts [41, 42]. The list of Xi-maintenance candidate factors from

these prior studies is distinct from ours, with the exception of *Dnmt1*, for several potential reasons. First, previous approaches did not screen in the presence of 5-aza-2′-dC and are thus not expected to find 5-aza-2′- dC-augmenting pathways such as RRM2-inhibition. Second, cell line immortalization has the potential to create aberrancies in chromatin silencing pathways that deviate from normal development, as in cancer. Therefore, using primary MEFs, as in our screen, may more closely reflect in vivo silencing contributions of Xi maintenance path- ways. However, our single-well format using individual siRNAs presents challenges in detecting rare Xi reactivation events, even if adapted to a more sensitive luciferase reporter gene. Our screen was likely underpowered to identify novel high-confidence Xi silencing pathways as reflected by a low signal-to-noise margin of the assay, expressed as a low Z-factor of 0.11 (Fig. S2-1B). Improvement of the assay using different co-treatments (besides 5-aza-2′-dC) to increase rates of XCR, may lead to identification of different classes of Xi maintenance factors and minimize screening false negatives and positives [43].

Regardless, the adoption of 5-aza-2′-dC in the optimization of this screen in order to sensitize for DNA demethylation ultimately led to identification of a 5-aza- 2′-dC-interacting pathway with therapeutic relevance. From the standpoint of optimizing epigenetically acting drugs, monitoring gene reactivation from the Xi can therefore provide a readout of chromatin reprogramming with immediate effects on gene expression.

We used cell proliferation assays and genome-wide methylation level estimates in myeloid leukemia cell lines to gauge the activity of 5-aza-2′-dC. Our data suggest that at a low concentration of 5-aza-2′-dC, the addition of low dose HU, increases the fraction of 5-aza-2′-dC that is incorporated into DNA and available to inhibit DNMT1. This DNA incorporation augmentation effect has the potential to represent a therapeutic advantage. RRM2- inhibitors such

as resveratrol and hydroxyurea improve the demethylation activity of 5-aza-2′-dC and may have clinical benefit if used in combination. The clinical use of 5-aza-2′-dC is hampered by incomplete disease response in AML and MDS and by high rates of adverse effects [18, 44, 45]. Its mechanism of action in patients is most likely due to a combination of demethylating and direct cytotoxic actions that differ in their relative contribution according to disease context and 5-aza-2′-dC concentration. At higher doses, 5-aza-2′-dC is thought to form DNA adducts leading to DNA synthesis arrest, which inhibits its DNA incorporation [20, 46]. Higher doses therefore contribute to higher rates of adverse reactions including hematologic toxicities [20]. Accordingly, lower doses have been favored in more recent clinical trials and have shown greater likelihood in eliciting gene expression changes as well as producing clinic responses in AML and even solid tumors [20, 22, 46]. Thus, increasing DNA incorporation of 5-aza-2′-dC at low doses is a promising strategy to increase its therapeutic index by biasing its activity profile towards DNA demethylation.

In this study, we observed synergistic an anti-proliferative effect of 5-aza-2′-dC in combination with HU, how- ever, did not capture genome-wide methylation changes at all concentrations to explain this effect. The anti-proliferative effect in the absence of global DNA methylation changes is likely secondary to cytotoxic effects such as DNA adduct formation and DNA synthesis arrest. Alter- natively, it is possible that differentially methylated loci are preferentially demethylated by 5-aza-2′-dC at lower concentrations and expression of these genes drives the phenotypic effects of inhibiting proliferation, even when mean global methylation levels are not affected.

Previous studies have reported that 5-aza-2′-dC and HU drug combination is antagonistic to DNA methylation based on bisulfite sequencing analysis of three loci in two other cancer cell lines [47]. Our data support these findings at high concentration HU with 5-aza-2′-dC in K562

cells but shows a synergistic effect on DNA demethylation at lower doses of HU and with RNR inhibition. The extent of RNR inhibition is likely critical for a synergistic interaction with 5-aza-2′-dC as too little RNR inhibition will not increase DNA incorporation of 5-aza- 2′-dC and too much RNR inhibition with lead to S-phase arrest and interfere with 5-aza-2′-dC-mediated passive DNA demethylation (see model Fig. 2-2a).

Another relevant disease model to test a potential therapeutic benefit of the combination of 5-aza-2′-dC and HU may be sickle cell anemia. Current therapies to treat the genetic defect in adult hemoglobin are aimed at reactivating the fetal hemoglobin gene [48]. Hydroxyurea is a standard therapy that when administered at cytotoxic doses to patients severely affected with sickle cell anemia increases fetal hemoglobin levels, but only in a subset of patients for unknown reasons [48]. As opposed to myeloid leukemia, where the efficacy of 5-aza-2′dC is partially attributable to demethylation, in sickle-cell anemia clinical responses to 5-aza-2′-dC do correlate with demethylation of the fetal hemoglobin gene and increases in hemoglobin levels [48–50]. Thus it is appealing to explore modified dosing schedules of HU and 5-aza-2′-dC for sickle cell patients already receiving these therapies in order to potentially exploit some synergistic effect of combination therapy for raising hemoglobin levels.

METHODS

Genome-wide siRNA Library Plate Preparation

The Silencer Mouse Druggable siRNA Library V3 and Extension set V3 (Ambion) were provided as 250 pmol of lyophilized powder in a total of 153 382-well source plates, containing one siRNA per well except in columns 23 and 24, which were reserved for controls. Each of 21,114 genes is represented by mostly 3 unique (some 2 unique) siRNAs on different 384-well plates. Plates were centrifuged at 1700×*g*, 50 µl of nuclease-free water was added to each well, sealed and briefly vortexed to resus- pend the siRNAs in individual wells. RNA concentrations were confirmed by measuring 1 µl of siRNA solution from 14 randomly chosen wells by NanoDrop spectro- photometer (Thermo Scientific). 2 µl of siRNA diluted to 0.5 pmol/µl from each source plate was stamped in duplicate onto Matrix white opaque 384-well tissue culture-treated plates (Thermo Scientific) by BenchCel 4X system with a PlateLoc plate sealer, Vcode Barcode Printer, and Vprep pipette fitted with a 96 LT head (all from Agilent Technologies) and stored in −80°.

Derivation of MEFs

Xi reporter MEFs were derived from a cross between transgenic male mice bearing a CAG promoter-driven luciferase, H2B-Citrine allele in the *Hprt* locus, and the X-linked GFP, respectively, and transgenic female mice heterozygous for an *Xist* knockout allele [24]. MEFs were derived at embryonic day 14.5 and cultured in MEF media (DMEM supplemented with 10 % FBS, nonessential amino acids, l-glutamine, penicillin–streptomycin, β-mercaptoethanol) following standard procedures. The reporter MEFs with genotypes $Xi^{CAG\text{-}Luciferase}Xa^{\Delta Xist}$, $Xi^{CAG\text{-}}$

$^{H2BCitrine}Xa^{\Delta Xist}$, and $Xi^{GFP}Xa^{\Delta Xist}$ were obtained at expected Mendelian ratios of 1 out of 4 embryos and identified by PCR genotyping for presence of an *Xist* knockout allele, presence of a FLP-Frt recombination production in the *Hprt* locus and GFP, respectively, and lack of Y chromosome gene *Zfy* [24].

High-throughput Screening siRNA and Chemical Screening Assays

The screening assay was optimized to maximize the Z-factor statistical measure of signal-to-noise ratio between the positive control of *Dnmt1* knockdown and negative control or no siRNA mock-transfected cells [51]. Pilot experiments sequentially tested individual variables of the assay such as incubation times and reagent types to increase the Z-factor of the assay. The 5-aza-2′-dC concentration of 0.2 uM used in the screen was determined in this empiric fashion, by titrating a range of 5-aza-2′-dC concentrations to determine which would maximally increase the signal separation between *Dnmt1* knock- down and control samples, calculated as the Z-factor of the assay. The Z-factor of the finalized screening assay was 0.11 (Fig. S2-1B) [51]. Screening data analysis was performed by first normalizing raw luminescence values by robust z-score which is the number of median absolute deviations for a given well luminescence value from the plate median luminescence value [52].

Primary MEFs from four female Xi-luciferase reporter embryos were thawed in 15 cm^2 plates, passaged twice at a 1:6 split, pooled to ensure a homogeneous cell population, and then frozen into 144 vials for use in screening and hit validation. For the large-scale screen, for each batch of 30 plates carrying the genome-wide siRNA library, 2 vials of cells were thawed in MEF media. After 1 day in culture, adherent cells were trypsinized, live cells excluding Trypan blue

were counted using a hemocytometer and brought up in suspension with MEF media agitated by a stir bar.

Meanwhile, a batch of 30 plates including duplicates from 15 source plates of 384-well siRNA library were thawed at room temperature, centrifuged, and cleaned with RNAse-reducing solution (Life Technologies). A positive control siRNA targeting *Dnmt1* (Ambion AM161526) was stamped by BenchCel 4X system with an 8 channel LT head (Agilent Technologies) into 16 wells of column 24 of each library plate by adding 4ul of nuclease-free water containing 1 pmol of si*Dnmt1* to each well. The 16 wells of the column 23 were reserved as negative control and contained no siRNA. Transfection was initiated by adding 20 μl of Opti-MEM (Life Technologies) and 0.05 μl RNAimax (Life Technologies) per well by Multidrop 384 (Thermo Scientific) and incubating for 20 min to 1 h. 20 ul of cell suspension containing 2000 cells with 5-aza-2′-dC (0.4 μM, Sigma) was added to the transfection mix, bringing the final 5-aza-2′-dC concentration to 0.2 μM. Cells were incubated for 3 days in a humidified 37° incubator at 5 % CO_2. 20 μl of media was then aspirated off using an ELx 405 plate washer (BioTek Instruments) and 20 μl of One-Glo luciferase assay reagent (Promega) was added using the Multidrop 384 and incubated for 20 min. As luminescence data were collected on an Acquest reader (Molecular Devices), quality control for each plate was performed by visual inspection of positive and negative controls on the heat map during data collection.

Chemical screening was performed analogously with several exceptions: 384-well plates were not pre-treated. Rather, 50 μl of cell suspension with 2000 MEFs and 5-aza-2′-dC (0.2 μM) were plated in fifteen 384-well plates. A positive control mixture was distributed to a row of wells on each plate by mixing 50 μl of cell sus- pension with 2000 cells per well in 1× MEF media with high concentration 5-aza-2′-dC (10.0 μM). The screening compounds were added to all but

positive control wells as 0.5 µl of 1 mM stock in DMSO by Biomek FX (Beck- man Coulter). After 72 h incubation, 30 µl of media were aspirated off, and the luciferase assay was performed as described for the siRNA screen. Libraries screened include 4266 compounds from Microsource (2000), Bio- mol enzyme inhibitor (337) and bioactive lipid libraries (203), Prestwick chemical library (1120), and NIH clinical collections (606) at the UCLA MSSR [53]. The 30 chemi- cals producing highest luciferase values were chosen for subsequent validation.

High-throughput siRNA Screening Analysis

Genome-wide siRNA screen hits were identified by Redundant siRNA Activity (RSA) analysis using robust z-scores as the input values [26]. The R script provided by Konig et al. was used with minor modifications to adapt it for our workflow (http://carrier.gnf.org/publications/ RSA). RSA works by ranking hits in order of activity then assigning P values for genes based on whether their siRNAs rank higher than would be expected by chance. We obtained two activity measurements for each siRNA since the siRNA library was screened in duplicate, and treated these data points as independent measurements with regard to the analysis. Therefore, most genes were represented by six data points (and some with four data points) in the RSA analysis.

Cell Culture and Treatment Methods

For subsequent Xi-reactivation/validation assays, MEFs at passage 1 or 2 post-derivation were seeded at a density of 6.0×10^4 cells per 12-well well and chemicals in MEF media and/or siRNAs in Opti-MEM media (Gibco) were added and incubated for 72 h. For 5-aza-2′-dC (Sigma), which

was resuspended in DMSO and stored at −80 °C, final DMSO concentration on the cells was kept below 0.1%.TotalvolumesofMEFand/orOpti-MEMmedia were normalized across samples when different treatments were used. Hydroxyurea and resveratrol (Sigma) were resuspended in DMSO and Uridine and Deoxycytidine (Sigma) were resuspended in water and stored at −20 °C. K562, HL60, U937, and THP1 cells were purchased from ATCC. K562, U937, and THP1 cells were cultured in RPMI media (Gibco) with 10 % FBS and HL60 cells were cultured in IMDM (Gibco) with 20 % FBS. ATCC culture method suggestions were followed for expanding the cells. The soft agar assay was performed by mixing of 1.2 % nobel agar (Sigma) in water with 2X RPMI to achieve final concentration of 0.6 % agar for the bottom layer. After this solidified in 6-well plates, top soft agar was prepared at final 0.3 % nobel agar concentration containing K562 cells to achieve 4.0×10^4 cells per well. DMSO or 5-aza-2′dC (0.05 μM) and/or HU (0.05 mM) were added to both bottom and top agar layers. This 1000:1 ratio of HU to 5-aza-2′-dC was determined to be optimal for the soft agar assay, which is different from the 4000:1 optimal ratio used in CellTiter Glo assay. Small colonies started appearing 4 days after plating. On day 8, colonies were stained with 0.01 % crys- tal violet for 1 h, washed with PBS, and the plates were scanned to obtain images.

Luciferase Assay

For each luciferase assay, MEF Xi-luciferase reporter treatments were performed in triplicate 12-well wells for 72 h and lysed with 200 μl passive lysis buffer (PLB, Promega) for 20 min at room temperature on an orbital shaker. Lysates were cleared by 30 s of centrifugation and 20 μl were assayed for luciferase activity with 50 μl of LARI reagent (Promega) on a GloMax microplate lumi- nometer (Promega). Protein concentration measure- analyzed by interpolating to standard

curve according to the manufacturer's instruction. For the proliferation TM ments were performed on corresponding PLB lysates by Quick Start Bradford Protein Assay Kit (Bio-Rad) and assays of leukemia cell lines, 100 µl of well-suspended cells were mixed with 100 µl of CellTiter Glo® reagent (Promega), incubated at room temperature for 20 min, and luciferase units were measured using a GloMax microplate luminometer (Promega).

RT-qPCR Analysis

Cells were harvested from a 6-well format in TRI- zol (Invitrogen) and RNA purification was performed with the RNeasy kit (Qiagen) according to manufac- turer's instructions with on-column DNAse treatment. cDNA was prepared using SuperScript III (Invitrogen) with random hexamers and RT-qPCR was performed using a M×3000 thermocycler (Stratagene) with prim- ers for *Rrm2* (F-GCACTGGGAAGCTCTGAAAC, R-GGCAATTTGGAAGCCATAGA), *Dnmt1* (F-CATGAATTCCTGCAAACAGAA, R-TTGACTTTAGCCAG GTAGCC), or *Gapdh* (F-GGCCTTCCGTGTTCCT, R-GCCTGCTTCACCACCTTCT). Results were normalized to *Gapdh* by the ΔCt method.

Knockdowns in Follow-up Experiments

Knockdowns by siRNA were performed by reverse trans- fection at 25 nM final concentration of siRNA. Briefly, a cell suspension was added to a pre-incubated mixture of Lipofectamine RNAimax, 100 µl of reduced serum Opti- MEM media, and siRNA. The siRNAs used were *Rrm2* [Ambion, 150659 (A), 64497 (B), 150661 (C)], *Dnmt1* (Ambion, 161526), and, as negative

controls, *Scramble* (Ambion, 4636), *Luciferase* (Dharmacon, D-001210-02), *Aurkb* (Dharmacon, D-063793-01), and *GFP* (Dhar- macon, P-002048-01). For *Rrm2* knockdown where the siRNA is not specified, siRNA 66497 was used.

Flow Cytometry

Flow cytometry for measuring the reactivation of the Xi- linked H2B Citrine and Xi-GFP reporters was performed as described previously [24]. For the cell cycle measure- ment with K526 cells, 5.0 × 10^6 cells (determined by trypan blue exclusion assay) were taken from each treat- ment condition, washed once with PBS, and stained with propidium iodide buffer (3 mM EDTA pH 8.0, 0.05 % NP40, 50 μg/ml PI, 1 mg/ml RNaseA in PBS) for 30 min at room temperature. Stained cells were passed through a strainer and analyzed by FACSDiva (BD Biosciences) with FlowJo software (Tree Star, Inc.).

3H Decitabine Incorporation

This assay was analogous to the reactivation treatment assays with a few modifications: assays were scaled 2.5- fold to 6-well format, 1 μl (1 μCi) of tritiated 5-aza-2′-dC (3H-Decitabine, Moravek Biochemicals Inc.) was added instead of cold 5-aza-2′-dC, and samples were harvested after 48 h of incubation. Cells were trypsinized, genomic DNA isolated using the Quick-gDNA MinPrep kit (Zymo Research), and measured by QuBit fluorometer (Life Technologies). Tritium content of 25 μl of genomic DNA was measured using a scintillation counter and normal- ized to the measured DNA concentration.

Reduced Representation Bisulfite Sequencing

Primary Xi-reporter MEFs were subjected to the same chemical treatment as used for the luciferase assays, but in 6-well format, and a fraction of the cells was taken to confirm appropriate luciferase reporter activ- ity. Genomic DNA was isolated using the Blood and Cell Culture Mini Kit (Qiagen) with RNase A treat- ment (Life Technologies). The RRBS libaries were gen- erated at previously described by Orozco et al. with minor modifications [54]. DNA purifications for each enzymatic reaction was carried out using AMPure XP beads (Beckman Coulter). Bisulfite conversion was performed using the Epitect kit (Qiagen) twice com- pared to manufacturer's instruction to optimize the efficiency. Bisulfite-converted libraries were amplified using MyTaq Mix (Bioline) with the following pro- gram:(98°Cfor15s,60°Cfor30s,72°Cfor30s)12 cycles, 72 °C for 5 min, 4 °C storage. DNA Methyla- tion calling was performed using BS-Seeker2 (2.0.32) using Bowtie (0.12.9) for read alignment on the UCLA Hoffman2 computer cluster [55]. Reads with adapter contamination were trimmed. The adapter sequence used for the contamination check was as follows: CGAGATCGGAAGAGCACACGTC, i.e. meCG end repair ± A tail ± 10 bp of Illumina adapter sequence. CpG islands (CGIs) were obtained from UCSC (http:// genome.ucsc.edu) and CGI tracks were based on meth- ods by Gardiner-Garner and Frommer [56]. Promot- ers were defined as the region transcription start site (TSS) minus 1 kb to TSS for all UCSC genes. Only sites covered by at least five reads across all samples under consideration were used in an effort to obtain reliable methylation levels. The methylation levels of samples were hierarchically clustered using complete linkage and the Euclidean distance metric. Statistical analysis, clustering, and heat map generation were performed using custom R scripts [57] (R core team, http://www.r-project.org).

Public Availability of Data

All genome-wide data are available from the GEO resource at http://www.ncbi.nlm.nih.gov/geo/query/acc. cgi?acc=GSE72295.

FIGURE LEGENDS

Figure 2-1. High-throughput siRNA and Chemical Screens Identify RRM2 Depletion and Resveratrol as Mediators of XCR

a Schematic of the X chromo- somes in female reporter MEFs carrying the luciferase reporter transgene in the *Hprt* locus specifically on the Xi. The *Xist* deletion on one of the chromosomes skews X-inactivation to the wild-type *Xist*-bearing X chromosome.

b Diagram of the screening workflow. siRNAs from the mouse genome-wide library and selected chemical libraries were assayed in 384-well plates containing a column of positive and negative controls. Xi-luciferase reporter MEFs were added and incubated for 72 h in the presence of 5-aza-2′-dC (0.2 μM) prior to a luciferase assay.

c Gene activity dis- tribution plot ranked by the –log of the p-value obtained with the redundant siRNA activity (RSA) assay from duplicate genome-wide siRNA screens following transformation of the luminescence activity values into robust z-scores. The top validated hits, *Dnmt1*, *Atf7ip,* and *Rrm2,* are labeled.

d (*i*) Graph depicting Xi-luciferase reporter reactivation upon knockdown of *Rrm2* with the three siRNAs (A, B, C) obtained from the genome-wide library in the presence or absence of 5-aza-2′-dC (0.2 μM) in the 12-well format. Luminescence was measured 72 h after the start of the treatment. *Error bars* indicate standard deviation of luminescence unit values from three individual wells with a given treatment in one experiment. (*ii*) RT-qPCR for RNA levels of *Rrm2* normalized to si*GFP* control and *Gapdh* expression. RNA was harvested in parallel to luciferase assays shown in (*i*). *Error bars* indi- cate standard deviation from three measurements in one experiment.

e Activity of chemicals in the chemical screen in the presence of 5-aza-2′-dC (0.2 μM), ranked by luminescence unit with the value corresponding to resveratrol designated.

f Xi-luciferase reporter assay as described in (di) titrating the resveratrol concentration with or without 5-aza-2′-dC (0.2 μM).

g Xi-luciferase reporter assay as in (di) titrating hydroxyurea (HU) with or without (untreated) 5-aza-2′-dC (0.2 μM). The result for resveratrol treatment in the same experiment is given for comparison. **h** (*i*) Xi-luciferase reporter assay as in (di) comparing the consequences of 0.2 μM 5-aza-2′-dC treatment and siRNA-mediated knockdown of *Dnmt*1 to elicit reporter reactivation by 20 μM resveratrol. (*ii*) RT-qPCR for *Dnmt1* RNA levels normalized to si*GFP* control and *Gapdh* expression in the same experiment as (*i*).

Figure 2-2. Inhibition of RNR Enhances DNA incorporation of 5-aza-2′-dC to Elicit XCR

a Illustration of model in which (*1*) inhibition of ribonucleotide reductase (RNR) by various means leads to (*2*) increased relative dCTP utilization for DNA synthesis from salvage pathways which are supplemented with exogenous 5-aza-2′-dC. (*3*) DNMT1 inhibition occurs upon binding to DNA-incorporated 5-aza-2′-dC leading to (*4*) increased loss of DNA methylation with successive cell divisions.

b Quantification of 3H-5-aza-2′-dC (3H-Decitabine) incorporation into genomic DNA upon either resvera- trol treatment or *Rrm2* knockdown for 48 h. Genomic DNA was isolated and an equal volume measured for 3H-Decitabine incorporation (disin- tegrations per minute, DPM), then normalized to the amount of DNA loaded (μg). *Error bars* indicate standard deviation from three

independent treatment wells. Asterisks indicate $p < 0.01$ by Student's T-test.

c Xi-luciferase reactivation assay as in Fig. 2-1d(i) in the presence of 0.2 μM 5-aza-2′-dC and 20 μM resveratrol and increasing concentrations of deoxycytidine (dC) or uridine. ALU represents luminescence unit values.

d As in **c** except with si*Rrm2* in the place of resveratrol.

Figure 2-3. RNR Inhibition Increases 5-aza-2′-dC-mediated Genome-wide DNA Demethylation in MEFs

a Bar chart displaying average genome-wide CpG methylation levels for the indicated 72 h treatments filtered for CpGs with at least 5X sequencing coverage by RRBS across all samples. Label color reflects the various treatment groups. *Subscripts* (A and B) indicate replicates where applicable. Treatment concentrations are: LowAza (5-aza- 2′-dC 0.2 μM), HighAza (5-aza-2′-dC 10.0 μM), and Resv (resveratrol 20 μM).

b Heat map of unsupervised hierarchical clustering of DNA methylation levels for all autosomal CpGs assayed by RRBS in MEFs treated with the indicated chemicals for 72 h as in **a** with at least 5X sequencing coverage across all samples. A methylation level of 1 indicates 100 % methylation, while 0 represents complete absence of methylation.

c Heat maps as in **b** but for subsets of autosomal CpG sites partitioned into four groups representing different DNA methylation levels in the untreated control samples; (*i*) 0.75–1.0, (*ii*) 0.50–0.75, (*iii*) 0.25–0.50, and (*iv*) 0–0.25. In each case, the combination of RNR inhibition with 0.2 μM 5-

aza-2′-dC clusters away from all other samples, but together with the high dose of 5-aza-2′-dC, and is more demethylated.

d Histograms display DNA methylation distributions for all autosomal CpGs as in **b** for indicated treatments and replicates. Data for additional replicates can be found in Fig. S2-7C/D.

Figure 2-4. Hydroxyurea and 5-aza-2′-dC Treatment of Myeloid Leukemia Cell Lines

a Dose–response curves measuring viable K562 cells using the Cell-Titer Glow Assay. Chemical treatments were performed with HU or 5-aza-2′-dC (Aza) alone or in combination at a fixed concentration ratio of 4000:1 HU:Aza (based on individual IC50 values). *Red lines* indicate concentrations assayed in the subsequent RRBS analysis displayed in **c**. The accompanying table depicts the Chow-Talalay analysis of the Combination Index (CI) at the given treatment combination concentrations. CI values <1.0 indicates synergy, 1= additive effect, and >1.0 antagonism of the combination drug effect.

b As in **a** except dose–response curves for HL60 cells using a fixed concentration ratio of 150:1 HU:Aza in the combined treatments.

c Heat map showing an unsupervised hierarchical clustering of X chromosome CpG methylation levels in K562 cells treated with the indicated chemicals for 72 h. Only those CpGs with greater than 0.75 methyla- tion level in the DMSO-treated samples are displayed, as these are the CpGs most dramatically affected by 5-aza-2′-dC treatment (see Fig. 2-3). CpGs were filtered for at least 10× sequencing coverage across all samples. DNA methylation levels were profiled for the treatments with the HU/aza concentration combinations indicated with *red lines* in **a** (labeled Low,

Mid, High). In addition, we treated cells with either 5-aza-2′-dC or HU at the respective concentrations (Low HU (0.16 mM), Mid HU (0.4 mM), High HU (0.80 mM), and Low Aza (0.04 μM), Mid Aza (0.1 μM), High Aza (0.2 μM)). Aza was resuspended in DMSO, thus DMSO only controls are matched to volume of Aza added in corresponding Aza samples.

d Flow cytometry analysis of propidium iodide-stained K562 cells treated with low and high HU concentrations as described in **c** compared to low or high DMSO control treatment.

Figure 2-1. High-throughput siRNA and Chemical Screens Identify RRM2 Depletion and Resveratrol as Mediators of XCR

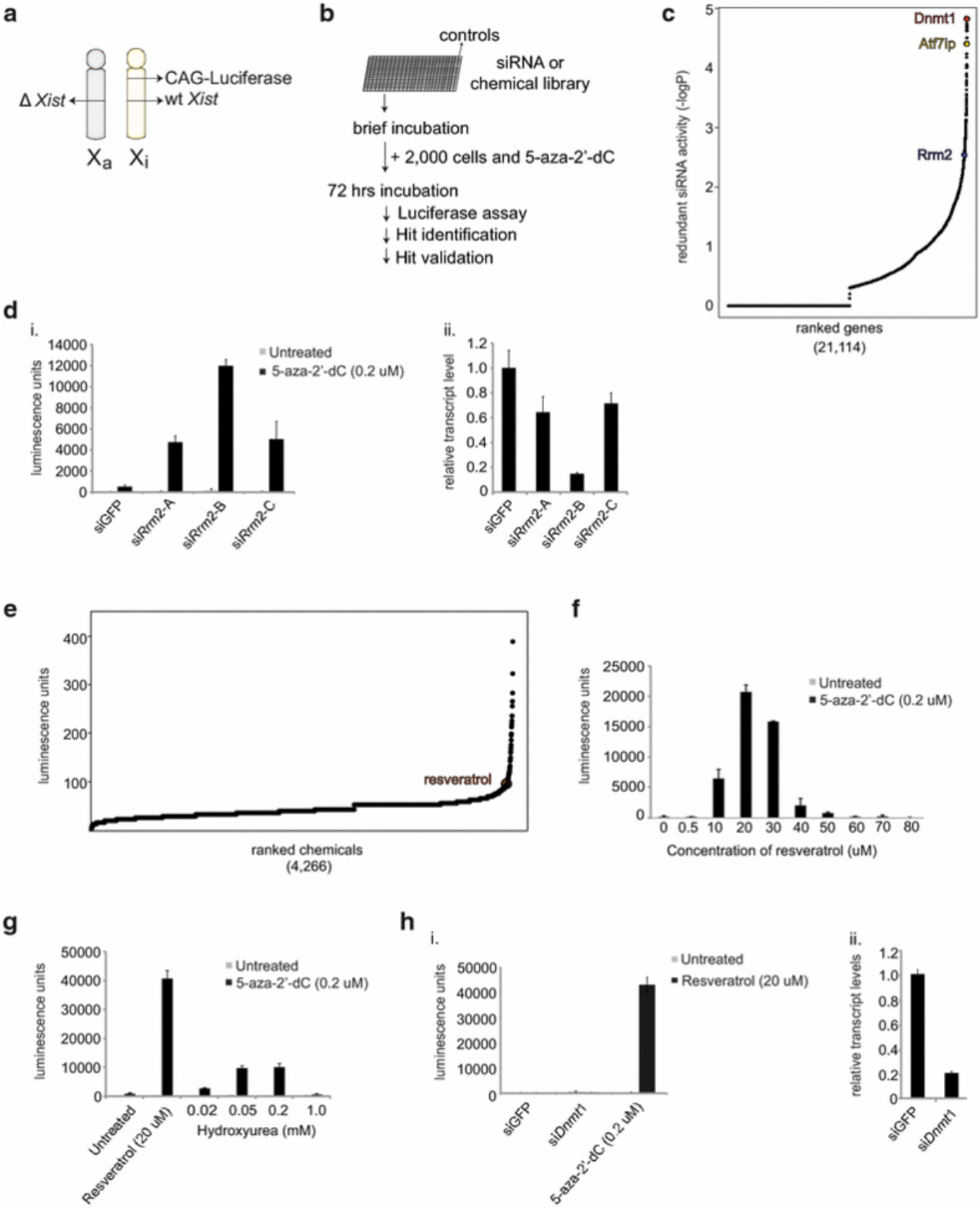


Figure 2-2. Inhibition of RNR Enhances DNA incorporation of 5-aza-2′-dC to Elicit XCR

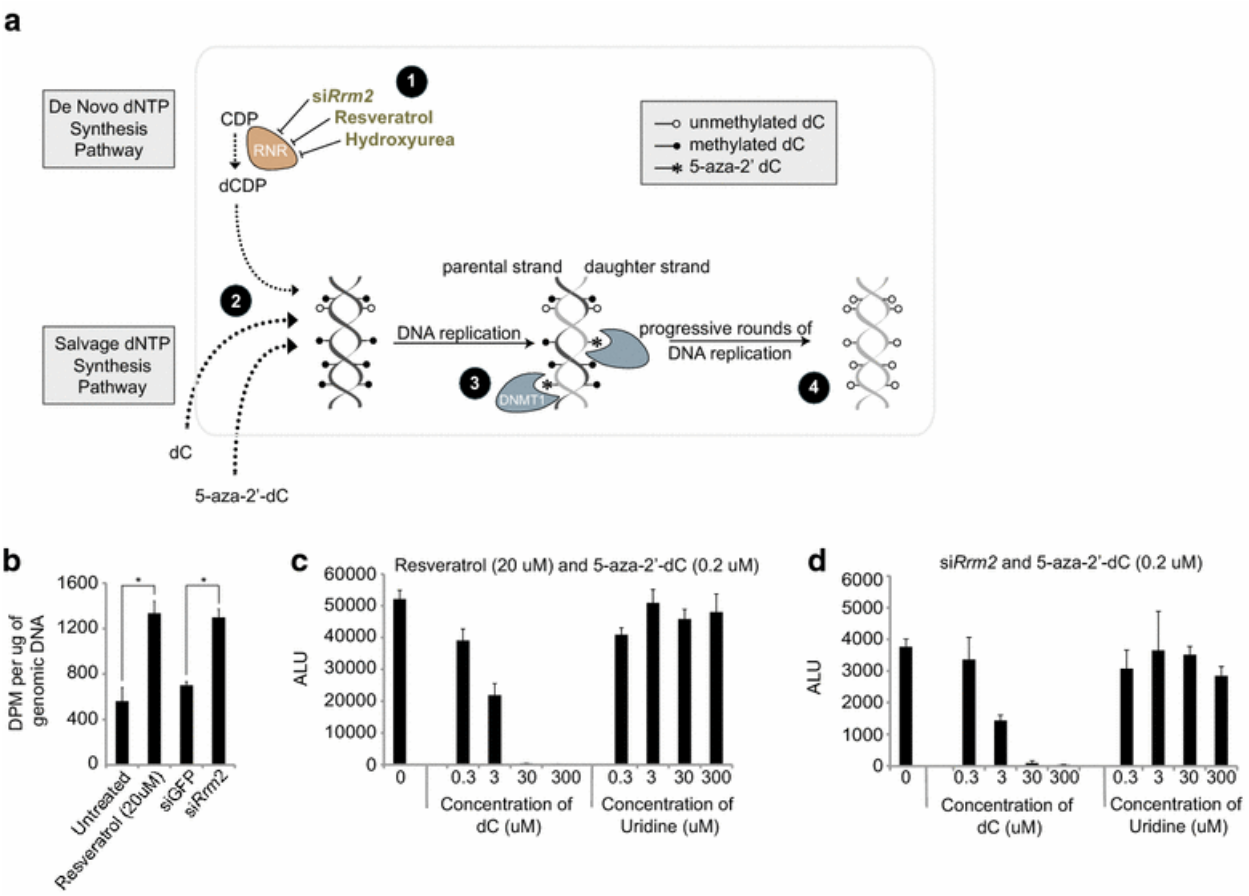


Figure 2-3. RNR Inhibition Increases 5-aza-2′-dC-mediated Genome-wide DNA Demethylation in MEFs

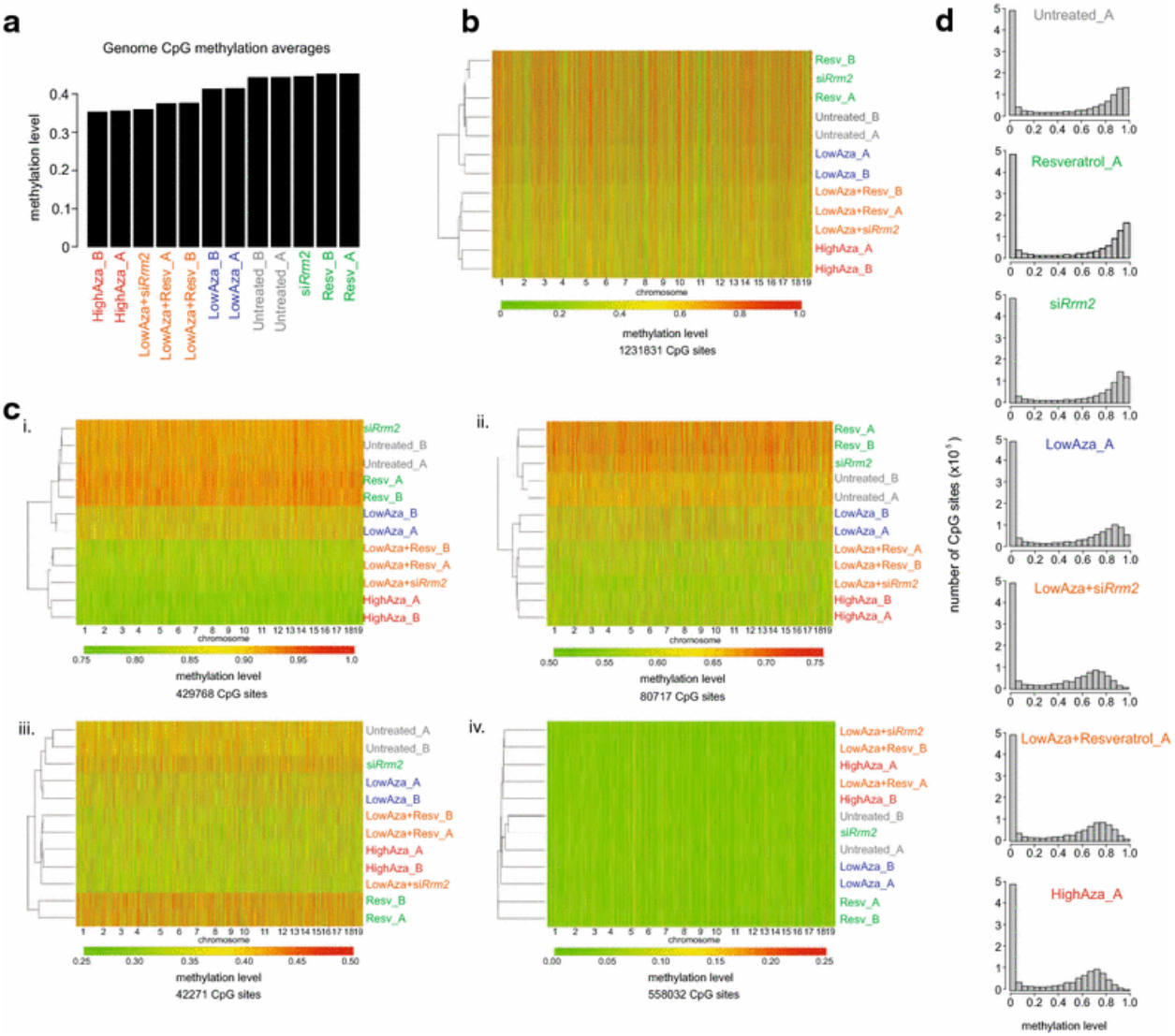
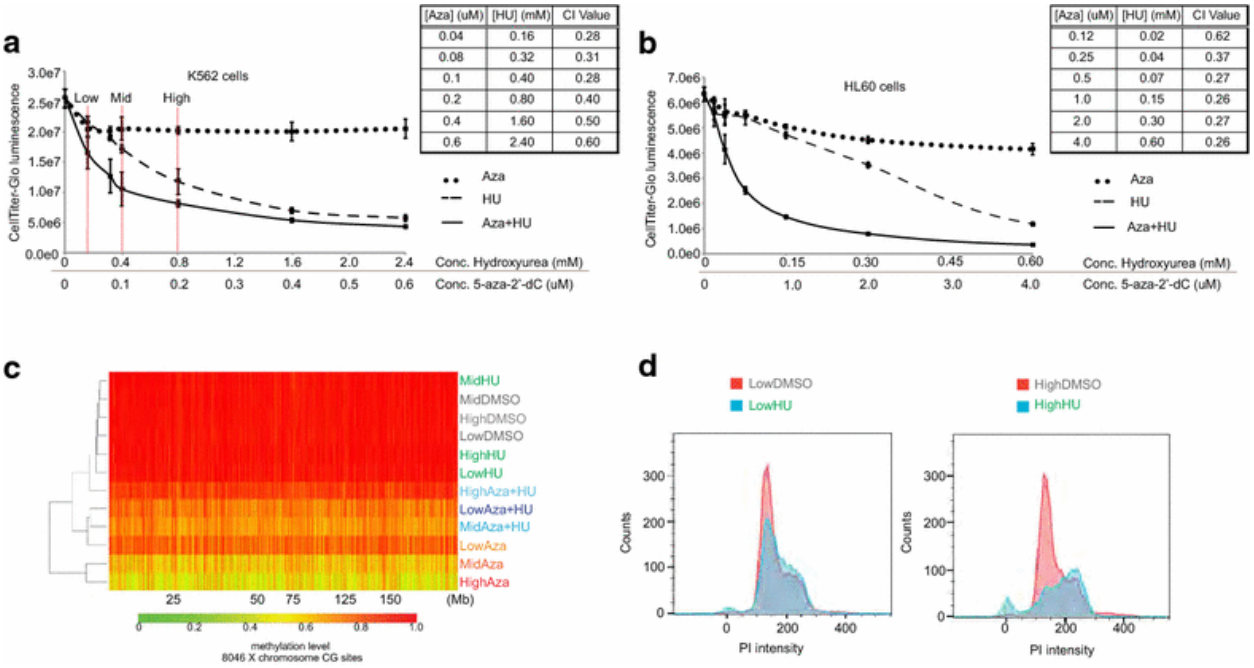


Figure 2-4. Hydroxyurea and 5-aza-2′-dC Treatment of Myeloid Leukemia Cell Lines



SUPPLEMENTAL FIGURE LEGENDS

Figure S2-1. Optimization of 5-aza-2'-dC concentration for the genome-wide siRNA screen

A. Bar chart illustrating luciferase activity from Xi-reporter MEFs upon knockdown of *Dnmt1* and treatment with varying concentrations of 5-aza-2'-dC in 384-well format for 72 hours. Error bars indicate standard deviation from eight measurements in one experiment. Aterisks indicate $p < 0.01$ by Student's T-test.

B. Scatterplot of luminescence values from the optimized Xi-reactivation screening assay in 384-well format in the presence of 5-aza-2'-dC (0.2 μM) with si*Dnmt1* (red) or negative control si*Aurkb* (Aurora kinase B, blue). The Z-factor, a measure of separation between positive and negative control populations used in the assessment of high-throughput assays, is shown [52].

Figure S2-2. Batch effects of genome-wide siRNA screening and robust z-score normalization

A. Box plot of all raw luciferase measurements distributions per individual 384-well plate from one of the duplicates of the siRNA screen. These plates were prepared and assayed in 30-plate batches according to their numerical order in the source library plates, keeping duplicate plates together.

B. As in (A) except each measurement was normalized by the robust z-score (median absolute deviations from the plate median [52].

Figure S2-3. Validation of gene hits identified by genome-wide siRNA screening

The chart displays the luminescence for the Xi-luciferase assay in 24-well format with knockdown by the indicated siRNAs, chosen as top hits of the genome-wide screen, in combination with 5-aza-2'-dC (0.2 μM) for 72 hours. For each gene hit, siRNAs were re-ordered to match the sequences of the 2 or 3 active siRNA identified by RSA activity analysis of the genome-wide siRNA screen. Error bars indicate one standard deviation from duplicate wells. si*Dnmt1* positive control is shown in red.

Figure S2-4. Validation of the resveratrol result with different Xi-reporter lines

A. Diagram of MEF Xi-H2B Citrine reporter genotype. As in Fig. 1A, except the Xi is bearing a CAG-driven histone H2B-Citrine reporter gene instead of luciferase in the *Hprt* locus The chart summarizes flow cytometry analysis of Xi-H2B Citrine reporter MEFs treated with resveratrol (20 μM) and/or 5-aza-2'-dC (0.2 μM or 10 μM) for 72 hours.

B. Diagram of MEF Xi-GFP reporter genotype. The Xi is bearing a randomly integrated CAG-driven GFP allele near the centromere [58]. The chart summarizes flow cytometry analysis of GFP reporter MEFs treated with siRrm2, resveratrol (20 uM) or HU (0.05 mM), and DMSO or 5-aza-2'-dC (0.2 μM). Error bars represent standard deviation from triplicate wells.

C. Representative flow cytometry dot plots of GFP reporter MEFs from part B.

Figure S2-5. Chemical screen results and validation

A. Box plot of all raw luciferase measurements from the chemical screen by individual 384-well plate, demonstrating lack of obvious batch effect. Chemical library plates were prepared and assayed as one batch of 15 plates.

B. Chart displaying results from the Xi-luciferase assay in the 24-well format upon treatment with various chemicals (at 10 μM) in the presence of 5-aza-2'-dC (0.2 μM) for 72 hours. Error bars indicate one standard deviation from duplicate wells except for negative control 5-aza- 2'-dC (0.2 μM) alone (n=16) and positive control 5-aza-2'-dC (10.0 μM) alone (n=16). Resveratrol is indicated with an asterisk.

Figure S2-6. Protein concentration measurements for Xi-luciferase reactivation assays

A. Chart depicts protein concentration of cell lysates corresponding to luciferase measurements in (1G). Error bars indicate standard deviation from three individual wells.

B. As in (A) but protein concentrations of cell lysates corresponding to luciferase meas- urements for (2C).

C. As in (A) but protein concentrations of cell lysates corresponding to luciferase measurements for (2D).

Figure S2-7. Analysis of autosomal DNA methylation in MEFs treated with combinations of RNR inhibition and 5-aza-2'-dC

A. (i) Heat map of the unsupervised hierarchical clustering as in Fig. 2-3B but only for autosomal CpG sites within CpG islands (CGIs). Genomic locations of CpG islands were obtained from UCSC Genome browser (see "Methods" section). Constitutively hypermethylated (>0.75) and hypomethylated (<0.15) sites were filtered out to improve contrast. (ii) Heat map of the unsupervised hierarchical clustering as in Fig. 2-3B but only for autosomal CpG sites within promoters. As in (i), constitutively hypermethylated (>0.75) and hypomethylated (<0.15) sites were filtered out to improve contrast. Promoters were defined as the region 1 kb upstream of the TSS for all UCSC genes.

B. Heat maps of the unsupervised hierarchical cluster- ing as in Fig. 3C for autosomal CpGs within CGIs with at least 5X coverage by RRBS across samples, but filtered for sites with methylation levels in the untreated sample of either (i) 0.75–1.0 (ii) 0.50–.75 (iii) 0.25–0.50 or (iv) 0–0.25.

C. As in Fig. 2-3D but for replicate samples.

D. Pairwise significance test results conducted using a two-sample Kolmogorov-Smirnov test ('KS- test stat' and 'KS-test p-value' columns) between the distributions of auto- somal CpG methylation in Fig. 2-3D and S2-7C, as well as two measures of effect size: Cohen's *d* and the differences between these 'upper modes' between the comparisons ('Delta upper mode' column).

Figure S2-8. Analysis of DNA methylation status on the X chromosome in MEFs treated with combinations of RNR inhibition and 5-aza-2'-dC

A. Heat map of the unsupervised hierarchical clustering of CpG methylation levels in MEFs as in Fig. 2-3B, except that the data for X chromosome CpG sites are shown.

B. (i) Heat map of the unsupervised hierarchical clustering as in Fig. 2-3B but only for CpG sites within CpG islands on the X chromosome. Constitutively hypermethylated (>0.75) and hypomethylated (<0.15) sites were filtered out to improve contrast. (ii) As in (i) except for CpG sites within promoters on the X chromosome. Again, constitutively hypermethylated (>0.75) and hypomethylated (<0.15) sites were filtered out to improve contrast. Promoters were defined as the region 1 kb upstream of the TSS for all UCSC genes.

Figure S2-9. DNA methylation status of the luciferase transgene in MEFs treated with combinations of RNR inhibition and 5-aza-2'-dC

A. Bar chart displaying average CpG methylation levels for the indicated treatments filtered by CpGs with at least 5X sequencing cover- age by RRBS across all samples as in Fig. 3A, but only considering the CpGs in the luciferase reporter gene/promoter.

B. Heat map of unsupervised hierarchical clustering of CpG methylation levels as in Fig. 2-3B except for CpG sites in the luciferase reporter gene/promoter.

C. As in (B), except for CpG sites within the luciferase reporter with a methylation level greater than 0.75 in both of the untreated samples.

D. Histograms showing the distribution of CpG methylation levels within the luciferase reporter gene.

Figure S2-10. Synergistic effect of HU and 5-aza-2'dC on myeloid leukemia cell line proliferation

A. Graphs represent cell counts measured with the hemocytometer after trypan blue staining compared to viable cell number measurement determined by CellTiter-Glo reagent (Promega) for four myeloid leukemia cell lines. High correlation coefficient, R^2, demonstrates linear relationship.

B. Dose response curves as in Fig. 2-4A, except for THP1 cells using a fixed concentration ratio of 1000:1 HU:Aza. **C.** Dose response curves as in Fig. 2-4A, except for U937 cells using a fixed concentration ratio of 300:1 HU:Aza. **D.** Soft agar assay of K562 cells plated in DMSO or 5-aza-2'dC (0.05 uM) and/or HU (0.05 mM) in a final concentra- tion of 3% agar and stained with crystal violet after 8 days of growth.

Figure S2-11. Extended data on the methylation analysis of K562 cells

A. Heat map showing an unsupervised hierarchical clustering of X chromosome CpG methylation in K562 cells treated with the indicated chemicals for 72 hours as in Fig. 2-4C but for all X chromosome CpGs.

B. CpG methylation distribution along the X chromosome in K562 cells, for CpG sites with at least 10X coverage across all samples as determined by RRBS. Chemical treatments are as shown in Fig. 2-4A/C.

Figure S2-1. Optimization of 5-aza-2'-dC concentration for the genome-wide siRNA screen

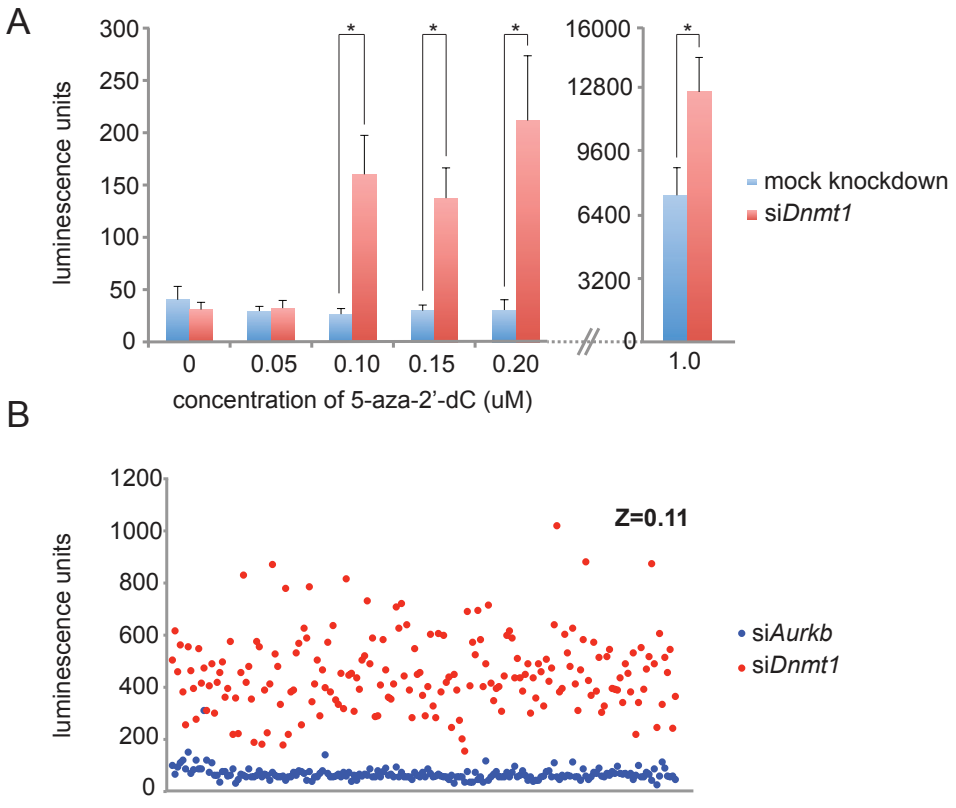


Figure S2-2. Batch effects of genome-wide siRNA screening and robust z-score normalization

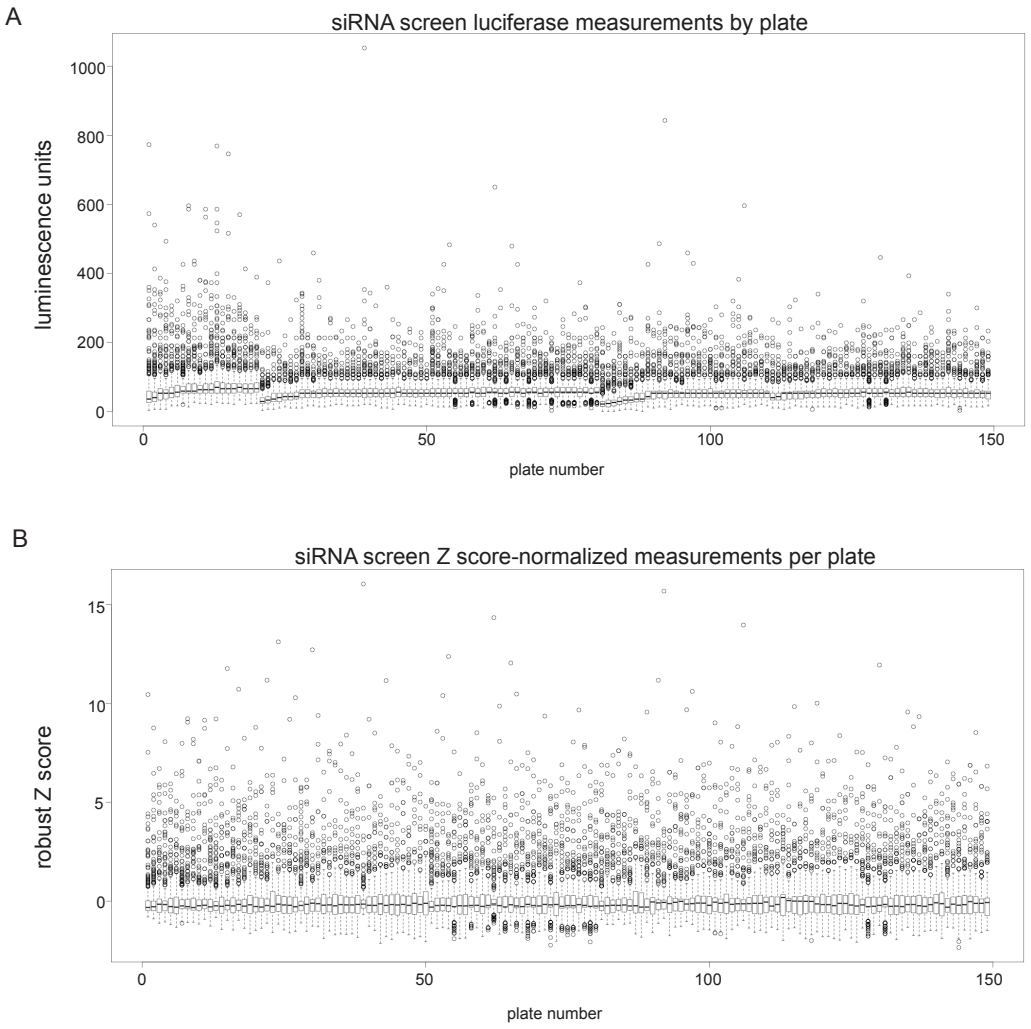


Figure S2-3. Validation of gene hits identified by genome-wide siRNA screening

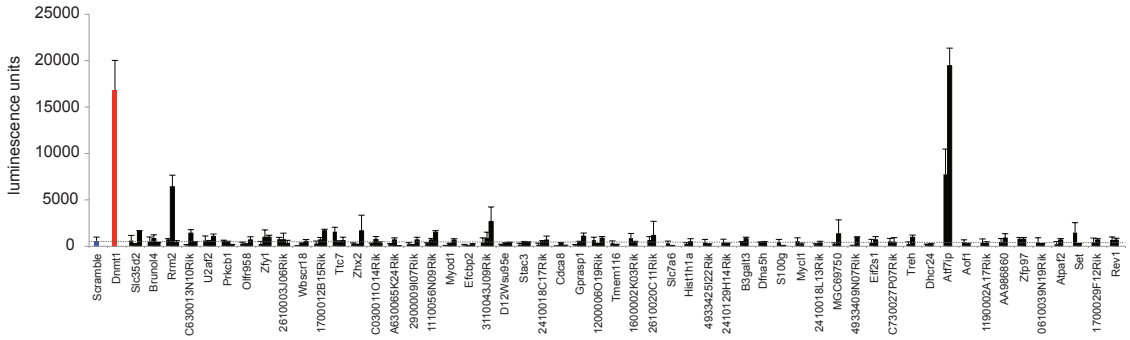


Figure S2-4. Validation of the resveratrol result with different Xi-reporter lines

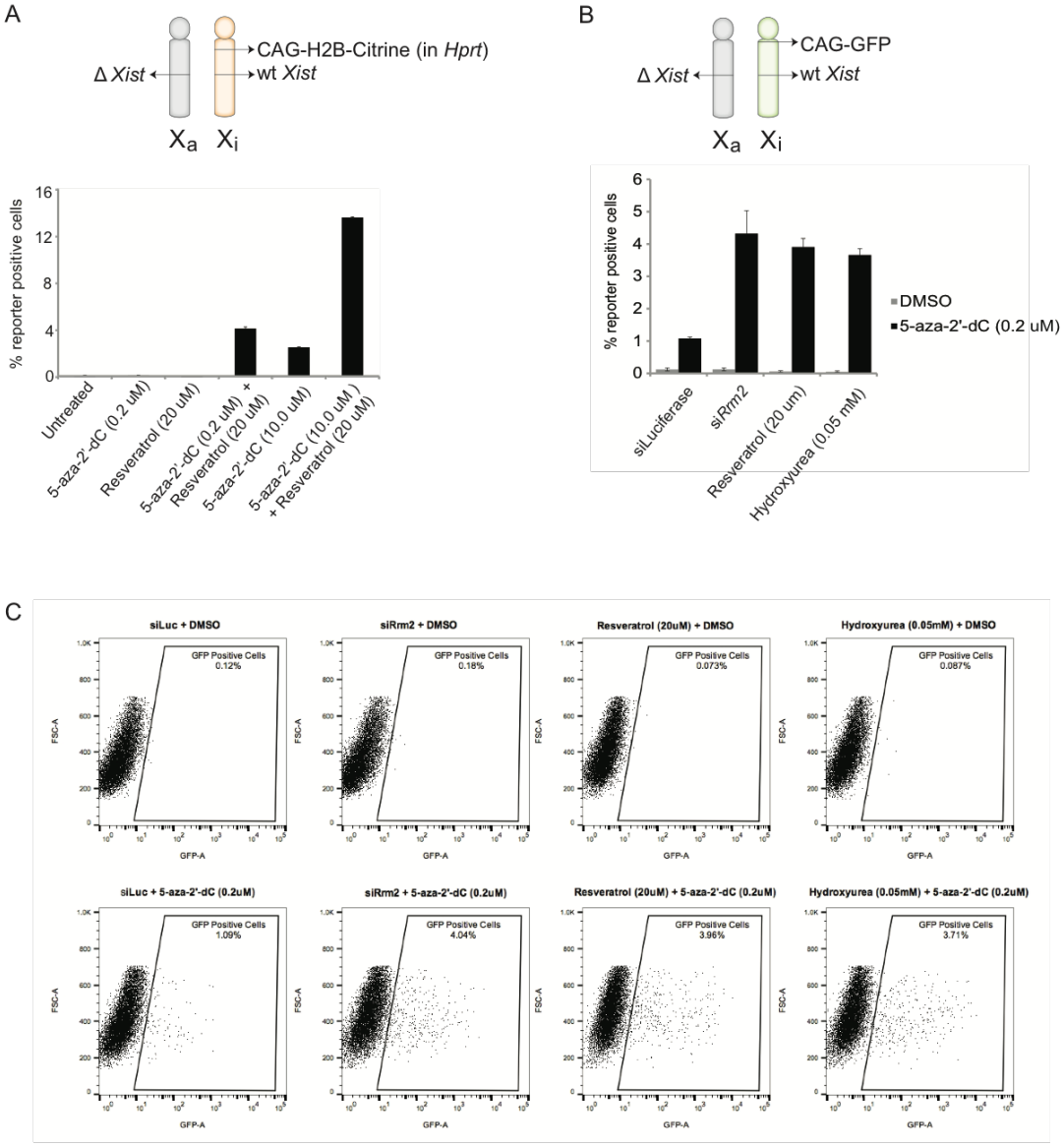


Figure S2-5. Chemical screen results and validation

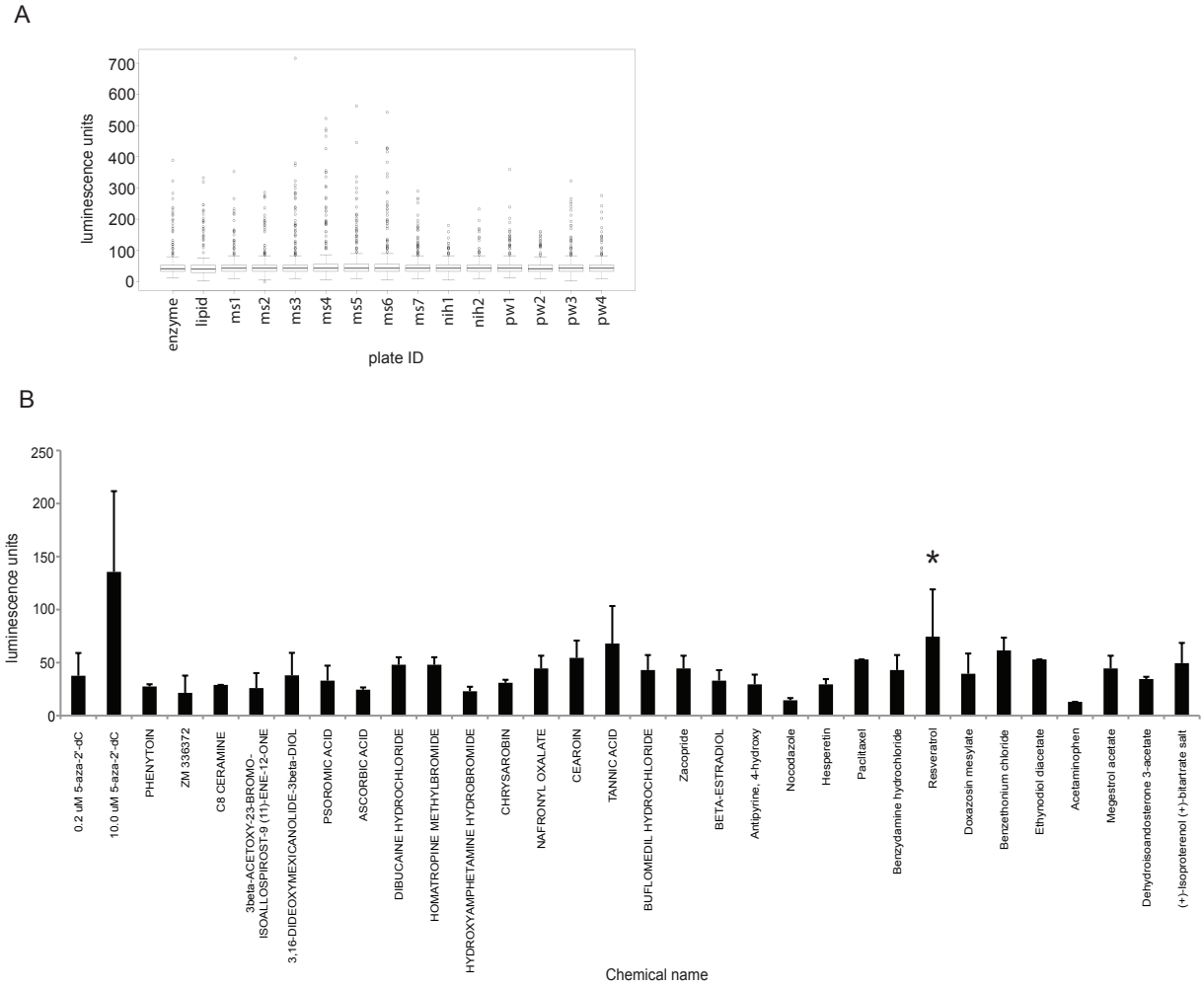


Figure S2-6. Protein concentration measurements for Xi-luciferase reactivation assays

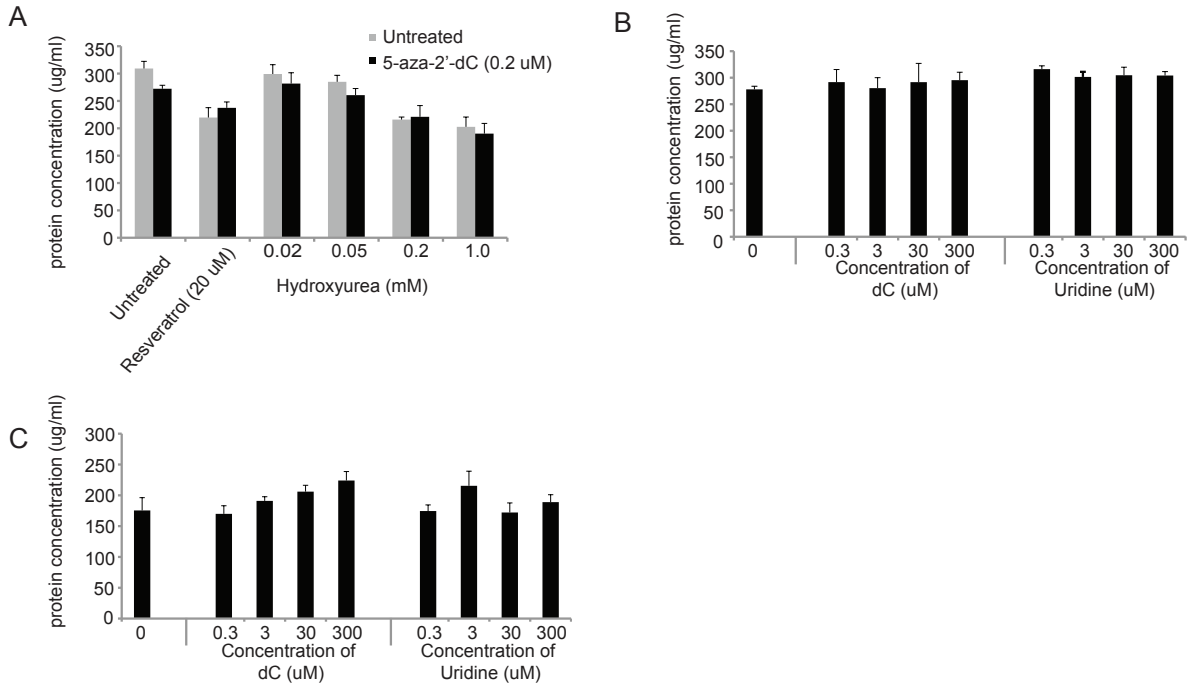
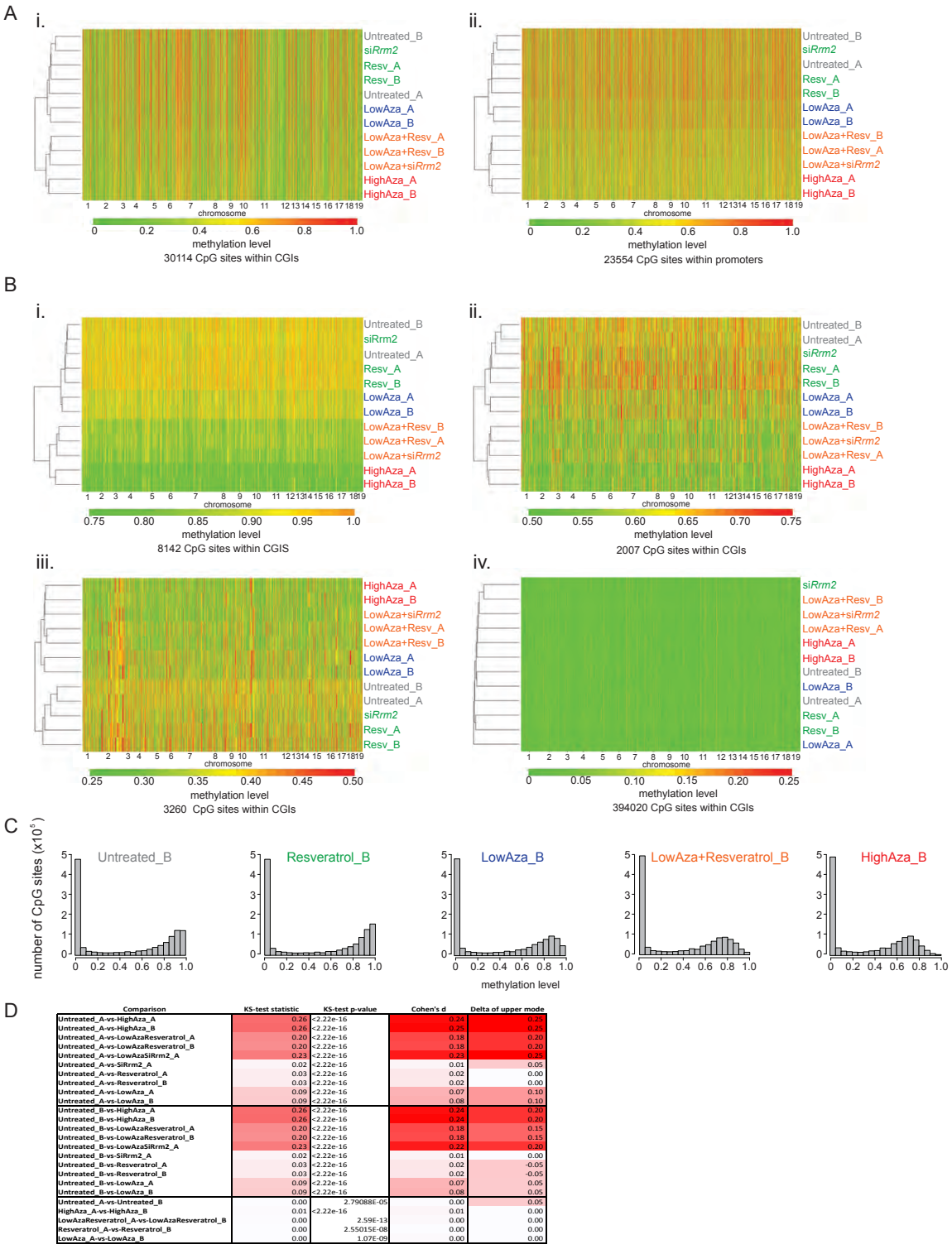


Figure S2-7. Analysis of autosomal DNA methylation in MEFs treated with combinations of RNR inhibition and 5-aza-2'-dC



Comparison	KS-test statistic	KS-test p-value	Cohen's d	Delta of upper mode
Untreated_A-vs-HighAza_A	0.26	<2.22e-16	0.24	0.25
Untreated_A-vs-HighAza_B	0.26	<2.22e-16	0.25	0.25
Untreated_A-vs-LowAzaResveratrol_A	0.20	<2.22e-16	0.18	0.20
Untreated_A-vs-LowAzaResveratrol_B	0.20	<2.22e-16	0.18	0.20
Untreated_A-vs-LowAzaSiRrm2_A	0.23	<2.22e-16	0.23	0.25
Untreated_A-vs-SiRrm2_A	0.02	<2.22e-16	0.01	0.05
Untreated_A-vs-Resveratrol_A	0.03	<2.22e-16	0.02	0.00
Untreated_A-vs-Resveratrol_B	0.03	<2.22e-16	0.02	0.00
Untreated_A-vs-LowAza_A	0.09	<2.22e-16	0.07	0.10
Untreated_A-vs-LowAza_B	0.09	<2.22e-16	0.08	0.10
Untreated_B-vs-HighAza_A	0.26	<2.22e-16	0.24	0.20
Untreated_B-vs-HighAza_B	0.26	<2.22e-16	0.24	0.20
Untreated_B-vs-LowAzaResveratrol_A	0.20	<2.22e-16	0.18	0.15
Untreated_B-vs-LowAzaResveratrol_B	0.20	<2.22e-16	0.18	0.15
Untreated_B-vs-LowAzaSiRrm2_A	0.23	<2.22e-16	0.22	0.20
Untreated_B-vs-SiRrm2_A	0.02	<2.22e-16	0.01	0.00
Untreated_B-vs-Resveratrol_A	0.03	<2.22e-16	0.02	-0.05
Untreated_B-vs-Resveratrol_B	0.03	<2.22e-16	0.02	-0.05
Untreated_B-vs-LowAza_A	0.09	<2.22e-16	0.07	0.05
Untreated_B-vs-LowAza_B	0.09	<2.22e-16	0.08	0.05
Untreated_A-vs-Untreated_B	0.00	2.79088E-05	0.00	0.05
HighAza_A-vs-HighAza_B	0.01	<2.22e-16	0.01	0.00
LowAzaResveratrol_A-vs-LowAzaResveratrol_B	0.00	2.59E-13	0.00	0.00
Resveratrol_A-vs-Resveratrol_B	0.00	2.55015E-08	0.00	0.00
LowAza_A-vs-LowAza_B	0.00	1.07E-09	0.00	0.00

Figure S2-8. Analysis of DNA methylation status on the X chromosome in MEFs treated with combinations of RNR inhibition and 5-aza-2'-dC

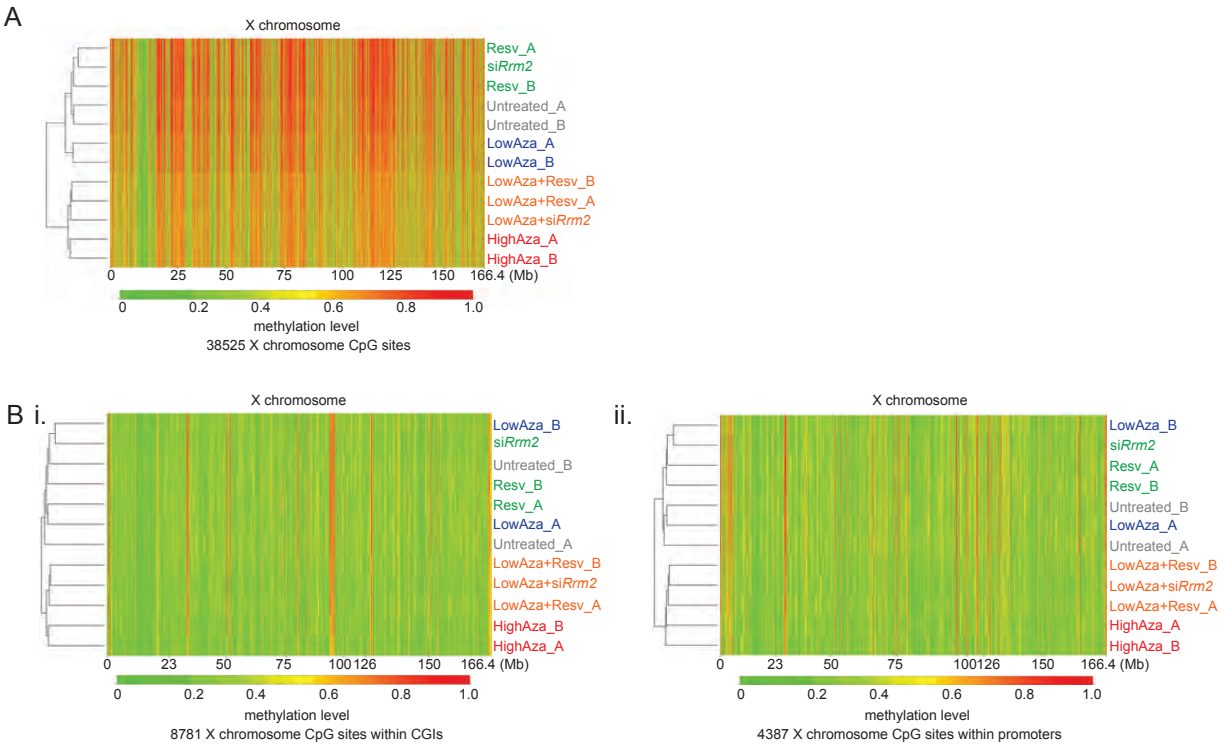


Figure S2-9. DNA methylation status of the luciferase transgene in MEFs treated with combinations of RNR inhibition and 5-aza-2'-dC

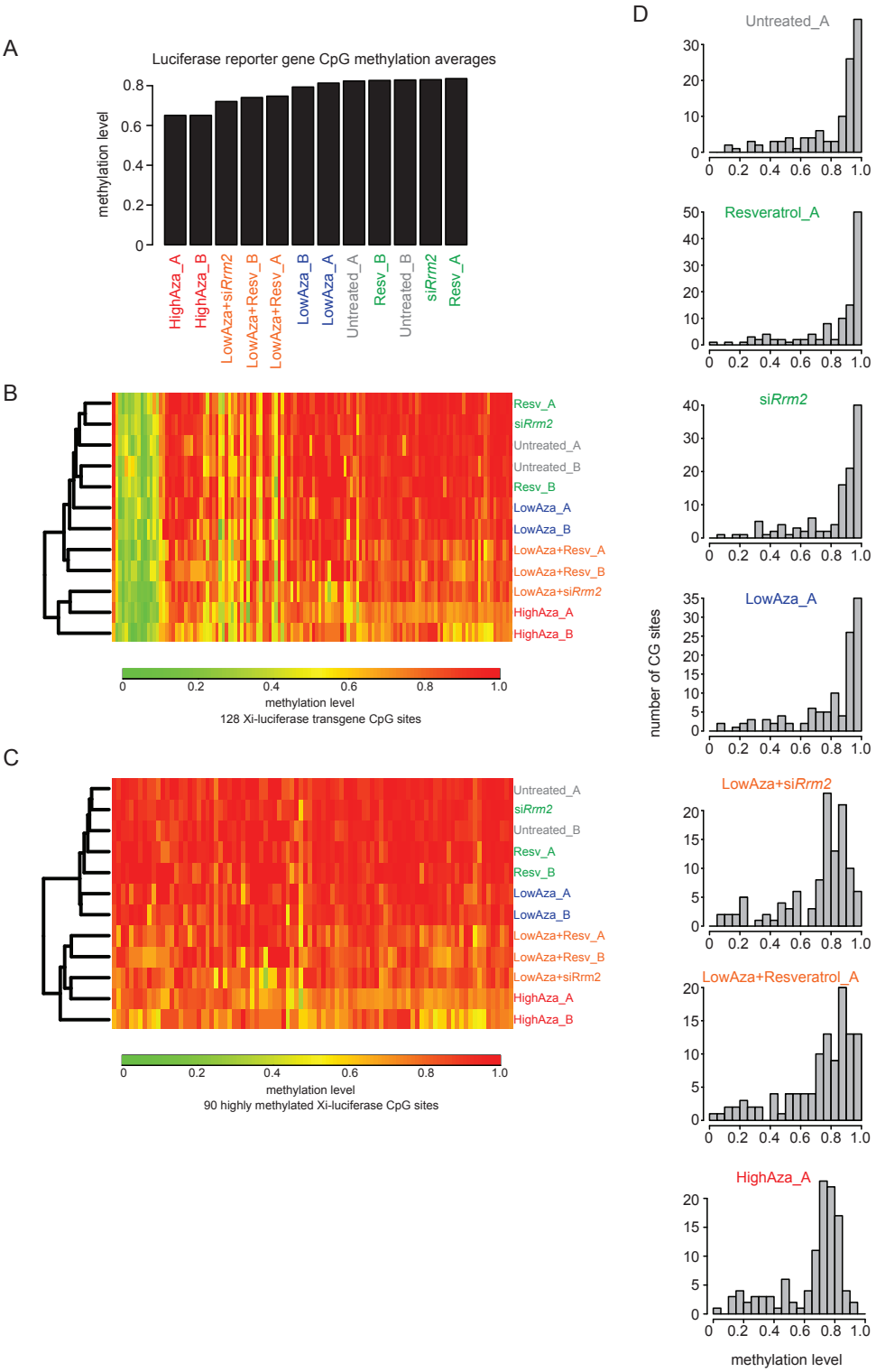


Figure S2-10. Synergistic effect of HU and 5-aza-2'dC on myeloid leukemia cell line proliferation

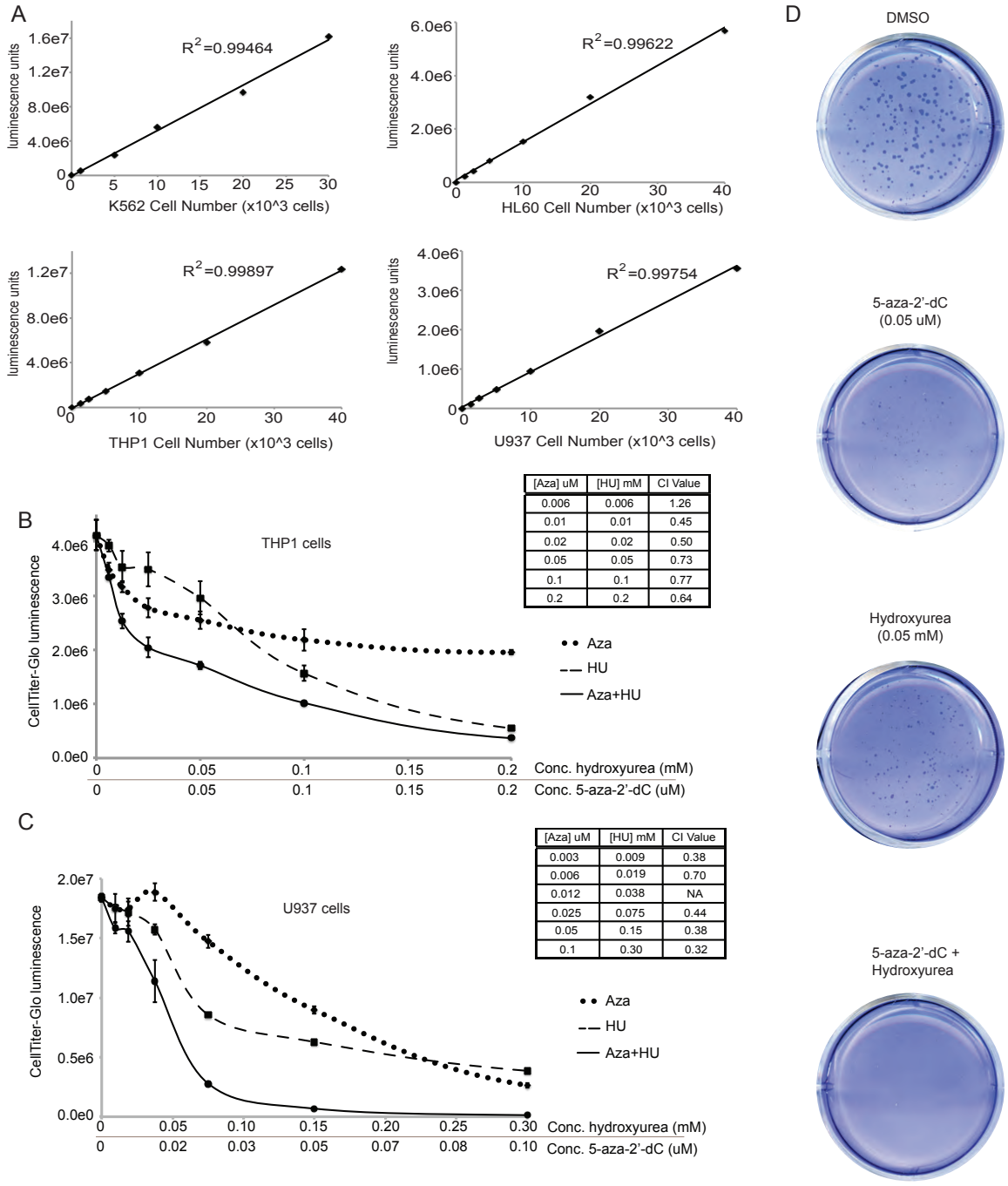
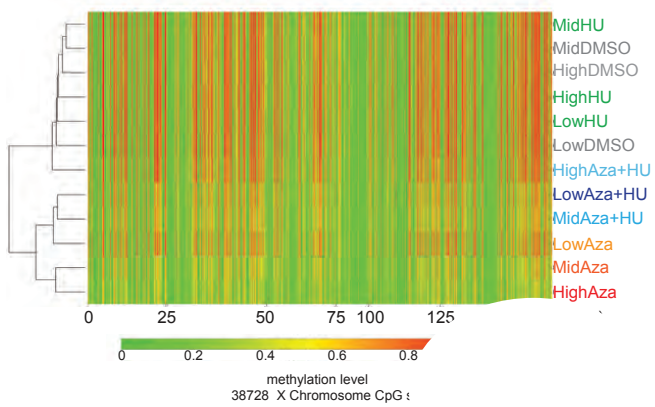
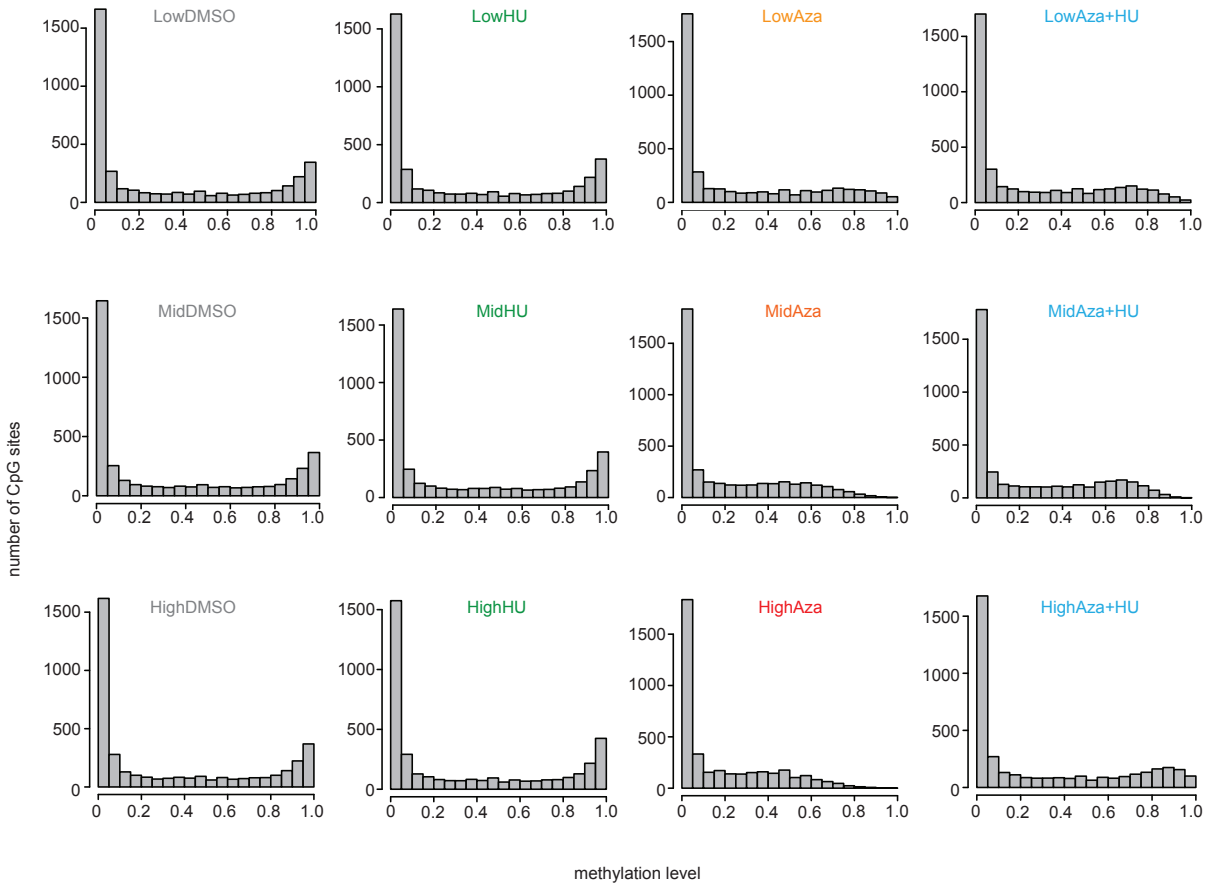


Figure S2-11. Extended data on the methylation analysis of K562 cells

A.



B.



REFERENCES

1. Brockdorff N. Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns. Dev Camb Engl. 2011;138:5057–65.

2. Lee JT. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nat Rev Mol Cell Biol. 2011;12:815–26.

3. Pollex T, Heard E. Recent advances in X-chromosome inactivation research. Curr Opin Cell Biol. 2012;24:825–32.

4. Costanzi C, Pehrson JR. Histone macroH2A1 is concentrated in the inac- tive X chromosome of female mammals. Nature. 1998;393:599–601.

5. Gendrel A-V, Apedaile A, Coker H, Termanis A, Zvetkova I, Godwin J, Tang YA, Huntley D, Montana G, Taylor S, Giannoulatou E, Heard E, Stancheva I, Brockdorff N. Smchd1-dependent and -independent pathways deter- mine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev Cell. 2012;23:265–79.

6. Grant M, Zuccotti M, Monk M. Methylation of CpG sites of two X-linked genes coincides with X-inactivation in the female mouse embryo but not in the germ line. Nat Genet. 1992;2:161–6.

7. Lock LF, Takagi N, Martin GR. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell. 1987;48:39–46.

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

9. Csankovszki G, Nagy A, Jaenisch R. Synergism of Xist RNA, DNA methyla- tion, and histone hypoacetylation in maintaining X chromosome inacti- vation. J Cell Biol. 2001;153:773–84.

10. Mak W, Nesterova TB, de Napoles M, Appanah R, Yamanaka S, Otte AP, Brockdorff N. Reactivation of the paternal X chromosome in early mouse embryos. Science. 2004;303:666–9.

11. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X inactivation during early mouse development. Science. 2004;303:644–9.

12. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R, Plath K, Hochedlinger K. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell. 2007;1:55–70.

13. Pasque V, Tchieu J, Karnik R, Uyeda M, Sadhu Dimashkie A, Case D, Papp B, Bonora G,

Patel S, Ho R, Schmidt R, McKee R, Sado T, Tada T, Meissner A, Plath K. X chromosome reactivation dynamics reveal stages of repro- gramming to pluripotency. Cell. 2014;159:1681–97.

14. Blewitt ME, Gendrel A-V, Pang Z, Sparrow DB, Whitelaw N, Craig JM, Apedaile A, Hilton DJ, Dunwoodie SL, Brockdorff N, Kay GF, Whitelaw E. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet. 2008;40:663–9.

15. Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol. 2000;225:294–303.

16. Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. Science. 2007;315:1141–3.

17. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005;37:853–62.

18. Lyko F, Brown R. DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. J Natl Cancer Inst. 2005;97:1498–506.

19. Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. Cell. 1980;20:85–93.

20. Azad N, Zahnow CA, Rudin CM, Baylin SB. The future of epigenetic therapy in solid tumours–lessons from the past. Nat Rev Clin Oncol. 2013;10:256–66.

21. Klco JM, Spencer DH, Lamprecht TL, Sarkaria SM, Wylie T, Magrini V, Hundal J, Walker J, Varghese N, Erdmann-Gilmore P, Lichti CF, Meyer MR, Townsend RR, Wilson RK, Mardis ER, Ley TJ. Genomic impact of transient low-dose decitabine treatment on primary AML cells. Blood. 2013;121:1633–43.

22. Tsai HC, Li H, Van Neste L, Cai Y, Robert C, Rassool FV, Shin JJ, Harbom KM, Beaty R, Pappou E, Harris J, Yen RW, Ahuja N, Brock MV, Stearns V, Feller-Kopman D, Yarmus LB, Lin YC, Welm AL, Issa JP, Minn I, Matsui W, Jang YY, Sharkis SJ, Baylin SB, Zahnow CA. Transient low doses of DNA- demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. Cancer Cell. 2012;21:430–46.

23. Yan P, Frankhouser D, Murphy M, Tam H-H, Rodriguez B, Curfman J, Trimarchi M, Geyer S, Wu Y-Z, Whitman SP, Metzeler K, Walker A, Klisovic R, Jacob S, Grever MR, Byrd JC, Bloomfield CD, Garzon R, Blum W, Caligiuri MA, Bundschuh R, Marcucci G. Genome-wide methylation profiling in decitabine-treated patients with acute myeloid leukemia. Blood. 2012;120:2466–74.

24. Minkovsky A, Sahakyan A, Rankin-Gee E, Bonora G, Patel S, Plath K. The Mbd1-Atf7ip-

Setdb1 pathway contributes to the maintenance of X chromosome inactivation. Epigenetics Chromatin. 2014;7:12.

25. Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 1997;11:156–66.

26. Konig R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, Bergauer T, Orth A, Krueger U, Zhou Y, Chanda SK. A probability-based approach for the analysis of large-scale RNAi screens. Nat Methods. 2007;4:847–9.

27. Nordlund P, Reichard P. Ribonucleotide reductases. Annu Rev Biochem. 2006;75:681–706.

28. Wood JG, Rogina B, Lavu S, Howitz K, Helfand SL, Tatar M, Sinclair D. Sirtuin activators mimic caloric restriction and delay ageing in metazoans. Nature. 2004;430:686–9.

29. Britton RG, Kovoor C, Brown K: Direct molecular targets of resveratrol: identifying key interactions to unlock complex mechanisms. Ann N Y Acad Sci 2015;1348:124–33.

30. Hubbard BP, Gomes AP, Dai H, Li J, Case AW, Considine T, Riera TV, Lee JE, SY E, Lamming DW, Pentelute BL, Schuman ER, Stevens LA, Ling AJY, Armour SM, Michan S, Zhao H, Jiang Y, Sweitzer SM, Blum CA, Disch JS, Ng PY, Howitz KT, Rolo AP, Hamuro Y, Moss J, Perni RB, Ellis JL, Vlasuk GP, Sinclair DA. Evidence for a common mechanism of SIRT1 regulation by allosteric activators. Science. 2013;339:1216–9.

31. Park S-J, Ahmad F, Philp A, Baar K, Williams T, Luo H, Ke H, Rehmann H, Taussig R, Brown AL, Kim MK, Beaven MA, Burgin AB, Manganiello V, Chung JH. Resveratrol ameliorates aging-related metabolic phenotypes by inhibiting cAMP phosphodiesterases. Cell. 2012;148:421–33.

32. Sajish M, Schimmel P. A human tRNA synthetase is a potent PARP1-acti- vating effector target for resveratrol. Nature 2014;519:370–3.

33. Fontecave M, Lepoivre M, Elleingand E, Gerez C, Guittet O. Resvera- trol, a remarkable inhibitor of ribonucleotide reductase. FEBS Lett. 1998;421:277–9.

34. Clouser CL, Chauhan J, Bess MA, van Oploo JL, Zhou D, Dimick-Gray S, Mansky LM, Patterson SE. Anti-HIV-1 activity of resveratrol derivatives and synergistic inhibition of HIV-1 by the combination of resveratrol and decitabine. Bioorg Med Chem Lett. 2012;22:6642–6.

35. Loffler M, Fairbanks LD, Zameitat E, Marinaki AM, Simmonds HA. Pyrimi- dine pathways in health and disease. Trends Mol Med. 2005;11:430–7.

36. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high- resolution DNA methylation analysis. Nucleic Acids Res. 2005;33:5868–77.

37. Ganzel C, Becker J, Mintz PD, Lazarus HM, Rowe JM. Hyperleukocyto- sis, leukostasis and leukapheresis: practice management. Blood Rev. 2012;26:117–22.

38. Chou T-C. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. Pharmacol Rev. 2006;58:621–81.

39. Shipony Z, Mukamel Z, Cohen NM, Landan G, Chomsky E, Zeliger SR, Fried YC, Ainbinder E, Friedman N, Tanay A. Dynamic and static mainte- nance of epigenetic memory in pluripotent and somatic cells. Nature. 2014;513:115–9.

40. Issa JPJ, Kantarjian HM. Targeting DNA methylation. Clin Cancer Res Off J Am Assoc Cancer Res. 2009;15:3938–46.

41. Bhatnagar S, Zhu X, Ou J, Lin L, Chamberlain L, Zhu LJ, Wajapeyee N, Green MR. Genetic and pharmacological reactivation of the mammalian inactive X chromosome. Proc Natl Acad Sci USA. 2014;111:12591–8.

42. Chan KM, Zhang H, Malureanu L, van Deursen J, Zhang Z. Diverse factors are involved in maintaining X chromosome inactivation. Proc Natl Acad Sci. 2011;108:16699–704.

43. Minajigi A, Froberg JE, Wei C, Sunwoo H, Kesner B, Colognori D, Lessing D, Payer B, Haas W, Lee JT. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Sci- ence. 2015;349:aab2276–2276.

44. Kantarjian H, Oki Y, Garcia-Manero G, Huang X, O'Brien S, Cortes J, Faderl S, Bueso-Ramos C, Ravandi F, Estrov Z, Ferrajoli A, Wierda W, Shan J, Davis J, Giles F, Saba HI. Issa J-PJ: Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome and chronic myelomonocytic leukemia. Blood. 2007;109:52–7.

45. Kantarjian HM, Thomas XG, Dmoszynska A, Wierzbowska A, Mazur G, Mayer J, Gau J-P, Chou W-C, Buckstein R, Cermak J, Kuo C-Y, Oriol A, Ravandi F, Faderl S, Delaunay J, Lysák D, Minden M, Arthur C. Multicenter, randomized, open-label, phase III trial of decitabine versus patient choice, with physician advice, of either supportive care or low-dose cytarabine for the treatment of older patients with newly diagnosed acute myeloid leukemia. J Clin Oncol Off J Am Soc Clin Oncol. 2012;30:2670–7.

46. Issa JP, Garcia-Manero G, Giles FJ, Mannari R, Thomas D, Faderl S, Bayar E, Lyons J, Rosenfeld CS, Cortes J, Kantarjian HM. Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza- 2′-deoxycytidine (decitabine) in hematopoietic malignancies. Blood. 2004;103:1635–40.

47. Choi SH, Byun HM, Kwan JM, Issa JP, Yang AS. Hydroxycarbamide in combination with azacitidine or decitabine is antagonistic on DNA meth- ylation inhibition. Br J Haematol. 2007;138:616–23.

48. Sauntharajah Y, Hillery CA, Lavelle D, Molokie R, Dorn L, Bressler L, Gavazova S, Chen YH, Hoffman R, DeSimone J. Effects of 5-aza- 2′-deoxycytidine on fetal hemoglobin

levels, red cell adhesion, and hematopoietic differentiation in patients with sickle cell disease. Blood. 2003;102:3865–70.

49. DeSimone J, Koshy M, Dorn L, Lavelle D, Bressler L, Molokie R, Talischy N. Maintenance of elevated fetal hemoglobin levels by decitabine during dose interval treatment of sickle cell anemia. Blood. 2002;99:3905–8.

50. Saunthararajah Y, Molokie R, Saraf S, Sidhwani S, Gowhari M, Vara S, Lavelle D, DeSimone J. Clinical effectiveness of decitabine in severe sickle cell disease. Br J Haematol. 2008;141:126–9.

51. Zhang Chung. Oldenburg: a simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen. 1999;4:67–73.

52. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, Smith Q, Beijersbergen RL, Ghazal P, Shamu CE. Statistical methods for analysis of high-through- put RNA interference screens. Nat Methods. 2009;6:569–75.

53. Damoiseaux R. UCLA's molecular screening shared resource: enhancing small molecule discovery with functional genomics and new technology. Comb Chem High Throughput Screen. 2014;17:356–68.

54. Orozco LD, Morselli M, Rubbi L, Guo W, Go J, Shi H, Furlotte NA, Lopez D, Bennett BJ, Farber, Charles R, Ghazalpour A, Zhang MQ, Bahous R, Rozen R, Lusis AJ, Pellegrini M: Epigenome-wide association of complex meta- bolic traits in mice. Cell Metab. 2015;21:905–17.

55. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory- efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

56. Gardiner-Garden M, Frommer M. CpG Islands in vertebrate genomes. J Mol Biol. 1987;196:261–82.

57. Team RC: R: A Language and Environment for Statistical Computing. Vienna, Austria, 2011. http://www.R-Proj.Org 2013.

58. Hadjantonakis AK, Cox LL, Tam PP, Nagy A. An X-linked GFP transgene reveals unexpected paternal X-chromosome activity in trophoblastic giant cells of the mouse placenta. Genes N Y N. 2000;2001(29):133–40.

Chapter 3

Human Naive Pluripotent Stem Cells Model

X-chromosome Dampening and Inactivation

SUMMARY

Naive human embryonic stem cells (hESCs) can be derived from primed hESCs or directly from blastocysts, but their X chromosome state has remained unresolved. Here, we show that the inactive X chromosome (Xi) of primed hESCs was reactivated in naive culture conditions. Like cells of the blastocyst, the resulting naive cells contained two active X chromosomes with *XIST* expression and chromosome-wide transcriptional dampening and initiated *XIST*-mediated X inactivation upon differentiation. Both establishment of and exit from the naive state (differentiation) happened via an *XIST*-negative XaXa intermediate. Together, these findings identify a cell culture system for functionally exploring the two X chromosome dosage compensation processes in early human development: X dampening and X inactivation. However, remaining differences between naive hESCs and embryonic cells related to mono- allelic *XIST* expression and non-random X inactivation highlight the need for further culture improvement. As the naive state resets Xi abnormalities seen in primed hESCs, it may provide cells better suited for downstream applications.

INTRODUCTION

X chromosome inactivation (XCI) is the silencing of one of the two X chromosomes in placental female mammals. XCI initiates early in embryonic development during the transition from naive to primed pluripotency upon implantation of the blastocyst (Minkovsky et al., 2012). Loss- and gain-of-function experiments have demonstrated that the long non-coding RNA (lncRNA) *Xist*, itself encoded on the X chromosome, is the master regulator of XCI. Currently, mechanistic studies of XCI initiation by *Xist* are performed with mouse embryonic stem cells (mESCs), as these cells capture the naive pluripotent state of epiblast cells of the pre-implantation blastocyst from which they originate. Accordingly, mESCs carry two active X chromosomes (XaXa) that lack *Xist* expression and, upon differentiation, upregulate *Xist* on one randomly chosen X chromosome to initiate XCI (XaXi^{XIST+}, where Xi denotes inactive X chromosome; Minkovsky et al., 2012). Unlike mESCs, conventional human ESCs (hESCs) do not resemble their embryonic cells of origin, likely as a consequence of culture induced changes during their derivation. Molecular characteristics, such as their post-XCI state (XaXi) and resemblance to mouse post-implantation epiblast stem cells, classify hESCs as primed pluripotent (Nichols and Smith, 2009). Therefore, studies of XCI initiation in the human system are currently not feasible (Patel et al., 2016) and require the establishment of hESCs that recapitulate the pre-XCI state of the pre-implantation embryo.

An additional limitation of female hESCs, as well as human induced pluripotent stem cells (hiPSCs), is the epigenetic instability of the Xi over time in culture, which is characterized by loss of *XIST* RNA and partial transcriptional reactivation, leading to an eroded Xi (Xe) (Shen et al., 2008; Silva et al., 2008; Tchieu et al., 2010; Mekhoubad et al., 2012; Nazor et al., 2012). Because loss of *XIST* and Xi erosion cannot be reversed upon differentiation (Mekhoubad et al., 2012;

Nazor et al., 2012; Patel et al., 2016), downstream applications of primed female human pluripotent stem cells (hPSCs) are adversely affected by the lack of proper X chromosome dosage compensation. Hence, hESCs that recapitulate the pre-XCI state of the pre-implantation blastocyst are perhaps better for basic research and therapeutic applications.

Recently, multiple culture conditions have been devised to promote the establishment and maintenance of hPSCs in a naive pluripotent state, either by converting primed hPSCs to the naive state or by maintaining the naive state during derivation from the blastocyst (Hanna et al., 2010; Gafni et al., 2013; Chan et al., 2013; Takashima et al., 2014; Ware et al., 2014; Theunissen et al., 2014). To date, the X chromosome state of naive hPSCs has remained controversial (Davidson et al., 2015). Molecular characterization of these cells suggests that the diverse culture conditions applied establish pluripotency states of different developmental stages. Notably, two of these protocols (Takashima et al., 2014; Theunissen et al., 2014) achieve a global gene expression profile most similar to cells of human pre-implantation embryos (Huang et al., 2014). These findings raise the possibility that the pre-XCI state of the blastocyst could be captured under these culture conditions.

A distinct characteristic of pluripotent cells of the human—but not mouse—blastocyst is the expression of *XIST* from both active X chromosomes ($Xa^{XIST+}Xa^{XIST+}$; Okamoto et al., 2011), indicating uncoupling of *XIST* from XCI. In addition, a recent single-cell RNA sequencing study of human pre-implantation embryos described a downregulation, or dampening, of X-linked genes in female pre-implantation embryos (Petropoulos et al., 2016; Sahakyan and Plath, 2016). Thus, in early human development, an X chromosome dosage compensation process different from conventional XCI is in play. It is currently unclear whether the presence of two active yet *XIST*-expressing X chromosomes with lowered X-linked gene expression can be captured in cultured

female hPSCs.

Here, we studied the X chromosome state of cells derived by one of the naive hPSC culture protocols known to closely resemble the pre-implantation state, the 5iLAF-based culture method (Theunissen et al., 2014). 5iLAF media contains small-molecule inhibitors targeting MEK, B-Raf, GSK3beta, Src, and ROCK and the growth factors LIF, ActivinA, and FGF2. We discovered that the conversion of primed XaXi hPSCs to the naive state established cells with two active X chromosomes and expression of *XIST* RNA. The transition to the pre-XCI state was gradual and involved an *XIST*-negative intermediate state with two active X chromosomes. X-linked genes became down-regulated when cells transitioned from this intermediate to the *XIST*-expressing X_aX_a state, suggesting that dosage compensation of X-linked gene expression by dampening (Petropoulos et al., 2016; Sahakyan and Plath, 2016) occurs in 5iLAF-cultured hPSCs. Naive hPSCs initiated *XIST*-mediated XCI upon differentiation. These features uniquely resemble the X chromosome state of the human pre-implantation blastocyst (Okamoto et al., 2011; Petropoulos et al., 2016). Even though the bi-allelic *XIST* expression pattern of the human blastocyst was consistently captured in naive hPSCs, the majority of naive cells typically expressed *XIST* from only one of their two active X chromosomes. Moreover, instead of random XCI, only the prior Xi underwent XCI in differentiating naive hPSCs, indicating the presence of an epigenetic memory of the primed state in naive hPSCs, demonstrating the need for further culture modifications. Regardless, we demonstrate that the conversion from primed to naive pluripotency and subsequent differentiation provide an opportunity to reverse Xi erosion of primed hPSCs. In summary, our work identifies a cell culture system that enables reversal of Xi erosion and studies of *XIST* function from an active X chromosome, X chromosome dampening, and initiation of XCI. Our findings also establish the *XIST*-expressing active X chromosome as a defining feature of human

naive pluripotency both in vitro and in vivo.

RESULTS

Primed to naive conversion leads to Xi reactivation

We converted the primed female hESC line UCLA1 to naive pluripotency using the 5iLAF approach (Theunissen et al., 2014). As previously described, 5iLAF media initially resulted in considerable cell death followed by the emergence of dome-shaped colonies that could be passaged as single cells and induced the upregulation of key naive pluripotency markers (Theunissen et al., 2014; Figures S3-1A and S3-1B). To examine the expression status of the X chromosomes, we utilized RNA fluorescent in situ hybridization (FISH) to capture sites of nascent transcription at single-cell resolution. The primed UCLA1 hESC line used here carried a mostly silent *XIST*-negative Xi with slight erosion due to loss of *XIST* over time in culture (Patel et al., 2016). The silencing of the Xi was demonstrated by one nascent transcription spot per nucleus for the X-linked genes *HUWE1*, *ATRX*, and *THOC2* (Figures 3-1A–3-1C) and Xi erosion by expression of the lncRNA *XACT* (Vallot et al., 2015) from both X chromosomes (Figure S3-1C). Because Xi erosion was very limited in primed UCLA1, we considered this line to be $Xa^{XIST-}Xi^{XIST-}$ in this study.

Upon 19 passages (P19) in 5iLAF media, we detected bi-allelic expression of *HUWE1*, *ATRX*, and *THOC2* in all cells (Figures 3-1D–3-1F), indicating that the transition from primed to naive pluripotency resulted in X_i reactivation. Expression of *XACT* remained bi-allelic (Figure S3-1C), consistent with transcriptional activity of both X chromosomes. Cells with abnormal X chromosome count were present in a small proportion of naive UCLA1, in agreement with previously reported karyotypic abnormalities (Pastor et al., 2016). However, we analyzed only naive cells with two transcription foci of the X-linked gene *UTX*, which escapes XCI (Balaton et al., 2015), to ensure that naive cells with abnormal X chromosome count were not included in our

quantifications.

Naive hPSCs express *XIST* from an active X chromosome

At P19 in 5iLAF media, we detected cells with *XIST* expression from both active X chromosomes (bi-allelic *XIST*, $Xa^{XIST+}Xa^{XIST+}$; Figures 3-1F and 3-1G), perfectly recapitulating the blastocyst pattern (Okamoto et al., 2011; Petropoulos et al., 2016). However, the majority of naive cells displayed *XIST* expression from only one of the two active X chromosomes (mono-allelic *XIST*, $Xa^{XIST-}Xa^{XIST+}$; Figures 3-1D–3-1F). *XIST* RNA formed a cloud that was often diffuse and appeared more punctate than on a somatic Xi in cells with mono- and bi-allelic *XIST* expression (Figures 3-1D, 3-1E, and 3-1G), reminiscent of the human pre-implantation blastocyst pattern (Okamoto et al., 2011). These findings were reproduced in four independently performed conversions of UCLA1. Together, these results revealed the presence of an *XIST*-expressing Xa in cultured human cells.

Interestingly, naive pluripotent cells generated from the primed XaXi H9 hESC line by inducible expression of NANOG and KLF2 and adaptation to another naive culture media (t2iL+Gö media; Takashima et al., 2014) also displayed X_i reactivation (Figure S3-1D). As in 5iLAF-cultured UCLA1, we observed *XIST* expression from an X_a in t2iL+Gö-adapted H9; however, it was only the mono-allelic (and not bi-allelic) pattern in 3% of all cells, even after 14 passages post-t2iL+Gö adaptation ($Xa^{XIST-}Xa^{XIST+}$; Figure S3-1D). Detection of *XIST* from an Xa by two independent naive culture approaches further supports the presence of the unique non-silencing *XIST* expression in cultured naive cells. Because the 5iLAF culture condition induced upregulation of *XIST* in most cells, we focused our further studies only on naive hESCs obtained

by this method.

Primed to naive conversion occurs via an *XIST*-negative XaXa intermediate

To better understand the dynamics of Xi reactivation and *XIST* upregulation in the transition from primed to naive pluripotency, we analyzed the X chromosome state during the initial seven passages of UCLA1 in 5iLAF media by RNA FISH for *HUWE1* and *XIST* (Figure 3-2A). The Xi remained inactive in most cells during the first three passages but reactivated in all cells by passage 4 (P4). At this time point, *XIST* became induced only in a subset of XaXa cells, but continued passaging largely resolved the heterogeneity of *XIST* expression as most cells transitioned to the $Xa^{XIST-}Xa^{XIST+}$ state (Figures 3-1F and 3-2A). Bi-allelic *XIST*-expressing cells were generally detected together with mono-allelic ones but did not increase in proportion over time (Figures 3-1F and 3-2A). These results indicated that cells gradually change their X chromosome state in 5iLAF media, progressing from the XaXi state to the *XIST*-expressing end state with two active X chromosomes via an *XIST*-negative XaXa intermediate. Consistent with this, we found that *XIST*-negative X_aX_a cells obtained by subcloning of an early-passage heterogeneous naive UCLA1 culture (clone 4) converted to the $Xa^{XIST-}Xa^{XIST+}/Xa^{XIST+}Xa^{XIST+}$ state with passaging, where again the bi-allelic *XIST*-expressing cells were a minority (Figures S3-1E and S3-1F). Conversely, naive *XIST*-positive XaXa sub-clones (clones 9 and 12) did not change their *XIST* expression status and stably maintained their predominantly mono-allelic and some bi-allelic *XIST* expression over time. The *XIST*-negative XaXa intermediate was also observed when primed hiPSCs carrying an *XIST*-positive Xi (prior to *XIST* loss and X_i erosion) were converted to naive pluripotency, indicating that the presence of *XIST* in the primed state did not interfere with the sequence of events leading

to Xi reactivation (Figure S3-1G). Thus, the transition from the *XIST*-expressing Xi to the *XIST*-expressing Xa involves silencing and re-expression of *XIST* (see Table 3-1 for a summary of X chromosome states of hPSCs used in this study).

Xi reactivation in naive hESCs is chromosome-wide

To address whether Xi reactivation happened chromosome-wide, we assessed the expression of multiple X-linked genes in an allele-specific manner based on single nucleotide polymorphisms (SNPs). We applied either Sanger sequencing of RT-PCR products or RNA sequencing (RNA-seq) to measure the proportion of transcripts containing the reference and alternate SNP for a given X-linked gene. In primed UCLA1, almost all tested X-linked genes normally subject to XCI expressed only one of the two alleles, i.e., either solely the reference or the alternate allele (Figures 3-2B, S3-2A, and S3-2E), demonstrating that the same X chromosome was inactive in all cells. This finding was not surprising because hESC lines tend to be clonal (Shen et al., 2008). Only one gene normally subject to XCI displayed expression of both alleles (*TCEAL4*; SNP rs11010) similar to genes known to escape XCI (Figures 3-2B and S3-2A), likely due to the slight Xi erosion in primed UCLA1.

The non-random XCI state of primed UCLA1 line allowed us to determine at the population level whether the entire Xi reactivated in the naive state, which would lead to the expression of both alleles. We defined the allelic expression pattern in our three XaXa sub-clones of naive UCLA1: clone 4 at early passage when it was largely *XIST* negative, and the *XIST*-positive clones 9 and 12. Our analysis showed that all X-linked genes with mono-allelic expression in primed UCLA1 were bi-allelically expressed in the naive clones (Figures 3-2B and S3-2B–S3-

2E). These findings were corroborated by RNA-seq at the single-cell level (Figure 3-2C). In *XIST*-negative and *XIST*-positive single cells of naive UCLA1 (at early and late passage post-5iLAF adaptation, respectively), single-cell RNA-seq detected reads from both X chromosomes for genes that had only the reference or the alternate allele expressed in the starting primed cells (Figure 3-2C). These data demonstrated the reactivation of a larger number of X-linked genes, distributed across the entire X chromosome, in both the *XIST*-negative and *XIST*-positive naive hESCs, supporting and extending our RNA FISH-based conclusion of Xi reactivation.

The single-cell RNA-seq analysis also allowed us to examine the allelic origin of *XIST* expression in naive UCLA1. The majority of the *XIST*-positive X_aX_a cells (26 of 46; 57%) expressed only the alternate allele of *XIST* (based on SNP rs1620574), which was corroborated by Sanger sequencing of the *XIST* RT-PCR product at the population level (Figure S3-2F). This allele was expressed from the Xi in early passage primed UCLA1, when *XIST* was still on. However, 28% of single naive cells (13 of 46) expressed both alleles of *XIST*, perfectly recapitulating the blastocyst pattern (Figure S3-2F). We had estimated this to be closer to 5% by RNA FISH (Figures 3-1 and 3-2A), most likely due to lower sensitivity of RNA FISH compared to single-cell RNA-seq. Interestingly, 15% of cells (7 of 46) expressed only the *XIST* carrying the reference allele of SNP rs1620574 (Figure S3-2F), indicating that either of the two X chromosomes has the ability to upregulate *XIST* in naive cells.

X-inactivation-specific de-methylation occurs faster than global hypo-methylation in the primed to naive transition

DNA methylation of CpG islands (CGIs) is a key characteristic of XCI (Sharp et al., 2011). We

utilized reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2005) to examine methylation levels in primed UCLA1 and the naive clones 4, 9, and 12 as expression-independent evidence of Xi reactivation. The intermediary methylation typical for Xi-linked CGIs was observed in primed UCLA1 but largely absent in the naive clones, where it was reduced to the level of a male control (XaY; primed UCLA10 hESC line), consistent with chromosome-wide Xi reactivation in naive cells with and without *XIST* expression (Figure 3-2D).

CpGs outside of CGIs are not subject to regulation by XCI (Sharp et al., 2011). Notably, the naive clones displayed significant de-methylation of X-linked and autosomal CpGs outside the context of CGIs, consistent with reaching the globally hypo-methylated state of the human pre-implantation epiblast (Smith et al., 2014; Pastor et al., 2016; Theunissen et al., 2016; Figures 3-2E, S3-3A, and S3-3B). However, in this context, the *XIST*-negative XaXa clone 4 was not as dramatically de-methylated as the *XIST*-expressing clones 9 and 12 (Figures 3-2E, S3-3A, and S3-3B). This was also true for imprint control regions where hypo-methylation was detectable in the naive state but was much less prominent in clone 4 than clones 9 and 12 (Figure S3-3C), suggesting that global de-methylation and imprint de-methylation occurred more slowly than X-linked CGI de-methylation associated with Xa reactivation. Indeed, the quantification of average methylation levels demonstrated that the methylation loss within X-linked CGIs of clone 4 was far more significant compared to non-CGI CpGs both in chromosome X and autosomal context (Figures 3-2F and S3-3D). Together, these results hint at the presence of an active de-methylation process of CGIs on the X_i in the transition from primed to naive pluripotency. The X-specific nature of this process was further highlighted by the fact that autosomal CGIs became de-methylated at the global hypo-methylation rate of non-CGIs (Figures 3-2F, S3-3A, and S3-3D).

Transition from the *XIST*-negative to the *XIST*-Positive naive state is accompanied by dampening of X-linked gene expression

Given the changes of *XIST* expression and DNA methylation from primed to early- and late-passage naive hESCs, we asked whether gene expression also changed with this progression (Figure S3-3E). We compiled lists of genes known to be up- or downregulated in naive versus primed pluripotent cells based on single-cell RNA-seq data of epiblast cells of human blastocysts and primed hESCs (Yan et al., 2013; Table S2). We found that the naive clone 4 at early passage, when it was still *XIST* negative, exhibited an incomplete downregulation of genes that have higher expression in primed than naive state (Figure S3-3F). However, on average, it achieved upregulation of naive-specific genes as efficiently as the *XIST*-positive naive clones 9 and 12 (Figure S3-3G). The incomplete hypo-methylation as well as downregulation of primed-specific genes in clone 4 supported the classification of the *XIST*-negative XaXa state as an intermediate in the transition from primed to naive pluripotency.

Although both *XIST*-positive and negative naive clones harbored two active X chromosomes, we found that the *XIST*-negative clone 4 exhibited significantly higher X-linked, but not autosomal, gene expression compared to the *XIST*-positive clones 9 and 12 (Figure 3-2G). Importantly, upon transition to the *XIST*-positive state, clone 4 specifically reduced the expression of genes on the X chromosome, but not autosomes (Figure 3-2G), indicating that this difference could not be simply explained by clonal differences. This was also the case for an early, *XIST*-negative passage of naive cells derived from a different primed female hESC line, UCLA4, compared to late-passage *XIST*-positive cells of the same line (Figure 3-2H). Thus, the decrease of X-linked transcript levels consistently occurred during the transition from the *XIST*-negative to the

XIST-positive naive XaXa state.

The lowered expression of X-linked genes in *XIST*-positive naive cells mirrored the dampened X-linked gene expression recently described in female human pre-implantation embryos, where X chromosome dampening (XCD) also coincided with up-regulation of *XIST* during pre-implantation development (Petropoulos et al., 2016). The comparison of *XIST*-expressing XaXa UCLA1 naive clones to the starting XaXi primed UCLA1 cells revealed similar extent of X-linked gene expression in both, even though the former had two active X chromosomes and the latter only one (Figure S3-3H).

H3K27me3 accumulates on the *XIST*-expressing Xa in naive cells

A common feature of the *XIST*-expressing Xi is the accumulation of the repressive histone modification H3K27me3 and exclusion of RNA polymerase II in immunofluorescence experiments (Plath et al., 2003). To understand the function of *XIST* on the X_a further, we examined the distribution of H3K27me3 by immunostaining coupled to *XIST* RNA FISH in naive UCLA1 clone 12 cells and in normal female $XaXi^{XIST+}$ fibroblasts as control. We detected an enrichment of H3K27me3 on the *XIST*-expressing Xa similar to fibroblasts (Figure 3-3A). Notably, H3K27me3 enrichment is not sufficient for silencing the X chromosome in mouse (Plath et al., 2003), consistent with the notion that the accumulation of this mark can occur on an active X. In agreement with the *XIST*-expressing X chromosome being active, we failed to detect clear exclusion of RNA polymerase II from the *XIST*-expressing Xa in naive UCLA1 clone 12, whereas the exclusion was very obvious in normal female fibroblasts (Figure 3-3B).

Blastocyst-derived naive hESCs capture the *XIST*-positive, but not the *XIST*-negative, XaXa State

For late-passage naive cells obtained from UCLA1 and hiPSCs, the majority of cells expressed *XIST* from one active X chromosome and XaXa cells expressing *XIST* bi-allelically were consistently a minority (Figures 3-1, 3-2A, and S3-1G). This result was also obtained for all other primed to naive cell line conversions performed in this study (Table 3-1). In addition, we confirmed that the originally described naive WIBR3 hESC line (Theunissen et al., 2014) showed reactivation of the Xi and induction of *XIST* from one Xa and more seldom from both X chromosomes (Figures 3-3C and S3-4). These proportions did not change when primed hESCs were converted to naive pluripotency without FGF2 (5iLA) or without the GSK3beta inhibitor IM-12 and FGF2 (4iLA; Figures 3-3C and S3-4; Theunissen et al., 2016). Elimination of other inhibitors of the 5iLAF media interfered with cell growth and/or morphology.

An intriguing question was what the X chromosome state would be like in naive cells derived directly from pre-implantation blastocysts in 5iLAF media. We found that the $Xa^{XIST-}Xa^{XIST+}$ pattern was also dominant in two female naive hESC lines, UCLA19n and UCLA20n, which we directly derived from human blastocysts (Pastor et al., 2016). Interestingly, the proportion of cells expressing *XIST* bi-allelically was raised to over 30% at the expense of mono-allelic *XIST*-expressing cells (Figure 3-3D). Another interesting observation was that *XIST*-negative XaXa cells were largely lacking in embryo-derived lines, even at early passage (Figure 3-3D). We conclude that the 5iLAF culture condition supports the *XIST*-positive XaXa state regardless of cell origin, that the transient *XIST*-negative XaXa state is unique to the primed to naive transition, and that naive $Xa^{XIST+}Xa^{XIST+}$ cells are stabilized more effectively when naive

hESCs are derived directly from the embryo. Similar to naive cells derived from primed hESCs, blastocyst-derived naive cells displayed H3K27me3 accumulation and lacked exclusion of RNA polymerase II on the *XIST*-coated Xa (Figures 3-3A and 3-3B).

Naive hESCs induce XCI upon differentiation

Next, we asked whether naive hESCs were capable of undergoing XCI upon differentiation. Direct differentiation from the 5iLAF condition resulted in extensive cell death; hence, we re-adapted the naive cells to the primed culture condition (called re-primed state) before inducing differentiation. When starting from naive $Xa^{XIST-}Xa^{XIST+}$ UCLA1, *XIST* expression was lost in the re-primed state whereas both X chromosomes remained active and displayed a higher dose of gene expression (Figures 3-4A and S3-3I).

Upon differentiation of re-primed cells, one of the two X chromosomes upregulated *XIST* in around 80% of cells and the majority of cells with an *XIST* RNA cloud displayed silencing as measured by RNA FISH for the X-linked gene *HUWE1* (Figures 3-4A and 3-4B). The *XIST*-coated X chromosome was still actively expressing *HUWE1* in a fraction of differentiated cells, but the nascent *HUWE1* transcription focus was smaller compared to that on the *XIST*-negative X chromosome (Figure 3-4B), suggesting that XCI was initiated, but not completed. Re-priming and differentiation of the blastocyst-derived naive hESC line UCLA20n also resulted in loss of *XIST* in the re-primed state followed by *XIST*-mediated XCI upon differentiation (Figure 3-4C). Thus, we describe a pre-XCI state in cultured hPSCs that can faithfully induce *XIST*-mediated XCI. Additionally, the data suggest that the developmental path from the *XIST*-positive XaXa naive state to XCI occurs by first downregulating *XIST* on the Xa and then inducing it again on the future

Xi, reverting the sequence of events that took place in establishing naive pluripotent cells from the primed state.

XCI in differentiating naive hPSCs is non-random

To address whether XCI occurred randomly on one of the two X chromosomes, as normally observed in development, we analyzed the allelic expression of X-linked genes in single cells after 7 days of differentiation from the re-primed state of UCLA1. We found that almost all individual differentiated cells expressed only one allele of *XIST* and of the X-linked genes normally subject to XCI (Figures 3-5A–3-5C). For *XIST* as well as the X-linked genes (with one exception), it was the allele previously expressed in the starting primed cells. Genes escaping XCI expressed both alleles in the primed, naive, and differentiated states, validating the robustness of our assays (Figures S3-2, S3-5A, and S3-5B). These data demonstrated that XCI in differentiating naive hESCs is non-random, suggesting the presence of an epigenetic memory of the Xi inherited from the primed state.

Next, we tested whether non-random XCI also is a feature of differentiating naive cells that were derived from a population of cells known to undergo random XCI. We generated primed hiPSC lines from Rett syndrome patient fibroblasts harboring a heterozygous mutation in the X-linked gene *MECP2* that leads to the absence of the MeCP2 protein product (Lee et al., 2001). Consequently, the presence of the MeCP2 protein indicated that the wild-type allele was on the X_a, whereas lack of MeCP2 indicated that the mutant allele was on the X_a. In resulting hiPSC lines, all cells either expressed the MeCP2 protein (clone 16; $Xa^{MECP2wt}$ $Xi^{MECP2mut}$) or did not (clone 17; $Xa^{MECPmutt}$ $Xi^{MECP2wt}$; Figures 3-5D and S3-5C), consistent with the clonal nature of

hiPSCs with respect to XCI status (Tchieu et al., 2010). Due to its MeCP2-negative starting state, clone 17 was most informative for our analysis. Xi reactivation was captured with the detection of the MeCP2 protein in most cells of clone 17 at early passage in 5iLAF media (P3; Figure S3-5C) and was maintained in the re-primed state (Figure 3-5D). Subsequent differentiation, confirmed by expression of the negative pluripotency cell-surface marker CD44, induced complete loss of MeCP2 in clone 17—and not mosaic MeCP2 expression, which would be expected for random XCI—reverting to the original state where the X chromosome bearing the wild-type *MECP2* allele was inactivated ($Xa^{MECPmut}$ $Xi^{MECP2wt}$; Figures 3-5D and S3-5D).

As observed in UCLA1, silencing of the X chromosome in differentiated cells correlated with *XIST* upregulation (Figure S3-5E). For clone 16, all cells maintained MeCP2 expression in the naive, reprimed, and differentiated states (Figures 3-5D and S5C–S5E). We conclude that naive clones obtained from a population of randomly inactivated cells underwent non-random XCI upon differentiation and that the epigenetic memory for the Xi was maintained in naive hESCs as well as in hiPSCs (Figure 3-5E). Future experiments will demonstrate whether the Xi memory is due to residual DNA methylation or histone modifications that persist in the naive and re-primed states or perhaps involves other mechanisms.

Xi abnormalities of primed hESCs are erased in the naive state

We tested whether the naive state would correct the epigenetic abnormalities of the Xi prevalent in female primed hPSCs. *XIST* typically becomes silenced in primed hPSCs over time in culture, which is often associated with Xi erosion (Shen et al., 2008; Silva et al., 2008; Tchieu et al., 2010; Mekhoubad et al., 2012; Nazor et al., 2012; Vallot et al., 2015; Patel et al., 2016). Importantly, the

X chromosome state does not change upon differentiation of eroded (XaXe) primed hPSCs, giving rise to differentiated daughter cells with improper X chromosome dosage compensation. Primed hESC lines can also be in an extremely aberrant XaXa state that cannot undergo XCI upon differentiation (Patel et al., 2016). Currently, methods are lacking to overcome these epigenetic abnormalities of the X chromosome in primed hPSCs and their differentiated daughter cells. Because our experiments with UCLA1 demonstrated that *XIST*-mediated XCI was present in differentiated cells only after transition through the naive state but never when starting from primed cells (Patel et al., 2016; Figures 3-4A, 3-4B, 3-5A–3-5C, and S306A), we decided to examine the X chromosome state in hESC lines with extreme epigenetic abnormalities of the X.

In the primed state, the hESC line UCLA9 carried two active X chromosomes without *XIST* expression and did not display any evidence of XCI upon differentiation from the primed state, as shown by RNA FISH for the X-linked genes *HUWE1*, *ATRX*, and *THOC2* (Patel et al., 2016; Figures 3-6A, 3-6B, and S3-6B–S3-6D). Upon adaptation to the 5iLAF condition, the X-linked genes remained bi-allelically expressed and *XIST* was gradually induced (Figure 3-6C). Similar to UCLA1, at later passage, the majority of naive UCLA9 cells expressed *XIST* mono-allelically, albeit some bi-allelic *XIST*-expressing cells were consistently detected (Figure 3-6C and 6D). Upon transition to the re-primed state, *XIST* was silenced again and bi-allelic expression of X-linked genes was maintained (Figures 3-6D and 3-6E). Importantly, differentiation from the re-primed state induced XCI with *XIST* expression shown by silencing of *HUWE1*, *ATRX*, and *THOC2* (Figures 3-6D, 3-6F, S3-6E, and S3-6F). Thus, the naive state allowed the differentiation product of starting primed cells with abnormal XaXa state to have a proper Xi with *XIST* expression (Figure 3-6G).

We extended this analysis to the hESC line UCLA4. Primed UCLA4 and its differentiated

daughter cells carried an *XIST*-negative Xe (Patel et al., 2016). We captured the XaXe state by the differential expression of the X-linked genes *HUWE1* and *ATRX* (Figures 7A and S7A–S7C). Specifically, *HUWE1*, which was not affected by Xi erosion, was mono-allelically expressed in primed UCLA4 and its differentiated product. Conversely, *ATRX* had bi-allelic expression in primed UCLA4 due to Xi erosion and retained this expression pattern upon differentiation from the primed state. *HUWE1* became bi-allelically expressed in the naive states, consistent with reactivation of the Xe, whereas *ATRX* retained its bi-allelic expression (Figures 7B and 7C). The transition to the naive state was accompanied by *XIST* induction mostly in its mono-allelic form, albeit some bi- allelic cells were again observed (Figures 3-7B and 3-7C). We confirmed that at P9 naive UCLA4 hESCs were largely karyotypically normal (Figure S3-7D). Re-priming was associated with the maintenance of bi-allelic X-linked gene expression and loss of *XIST* (Figures 3-7C and 3-7D). Upon differentiation from the re- primed state, both *HUWE1* and *ATRX* were silenced on one X chromosome and *XIST* was expressed from the Xi (Figures 3-7C–3-7E). Together, these data demonstrate the ability of differentiating naive cells to undergo XCI, even when derived from primed XaXa and XaXe hESCs (Figures 6G and 7F).

DISCUSSION

A key difference between mouse and human development is the X chromosome state of the female pre-implantation embryo. In the mouse, the paternally inherited X chromosome be- comes silenced at the four- to eight-cell stage due to imprinted XCI. At the blastocyst stage, the imprinted Xi is reactivated specifically in epiblast cells to allow random XCI in embryonic cells upon implantation (Minkovsky et al., 2012). Intriguingly, imprinted XCI does not occur in human pre-implantation embryos (Petropoulos et al., 2016). Another distinguishing feature between mouse and human is that epiblast cells of the mouse blastocyst harbor two active X chromosomes that do not express *Xist*, whereas, in human pre-implantation embryos, both active X chromosomes are marked by *XIST* expression (Okamoto et al., 2011). Because naive hPSCs of the mouse resemble the X chromosome state of the epiblast cells of the *mouse* blastocyst, in our study we addressed whether human naive hPSCs can capture the unique X chromosome state of the *human* blastocyst. We applied independent approaches, such as RNA FISH and SNP-based allelic expression profiling by single-cell RNA sequencing, to robustly assess the X chromosome state in naive hPSCs at single-cell resolution, as well as bulk RNA sequencing and DNA methylation analyses at the cell population level.

Our data demonstrate that the naive 5iLAF culture condition captures an X chromosome state resembling that of the human blastocyst (Okamoto et al., 2011; Petropoulos et al., 2016). Specifically, we found that 5iLAF-cultured female naive cells carry two active X chromosomes, express *XIST* from the active X chromosome(s), display dampening of X-linked gene expression, and undergo de novo X inactivation upon differentiation. These results were robust and highly reproducible, regardless of whether naive hPSCs were derived directly from the blastocyst or primed hESCs or hiPSCs or established in different laboratories (K.P. and R.J.). Also, our results

are entirely consistent with data published recently (Theunissen et al., 2016). Therefore, our work identifies a cell-culture system for the mechanistic study of the unique and human-specific form of X chromosome dosage compensation that results in dampening, but not silencing, of X-linked gene expression early in human development and for the function of *XIST* on an active X chromosome. Moreover, our findings enable the molecular dissection of the transition from XCD to XCI and of the mechanisms of XCI initiation in human embryonic development. Most of our current understanding of how *XIST* initiates XCI comes from studies in mouse ESCs, but it is yet to be seen whether human *XIST* functions similarly to its mouse homolog in XCI initiation. In addition, our findings clarify the question of the X chromosome pattern in naive hPSCs, which has remained controversial (Davidson et al., 2015).

Despite the similarities of the X chromosome state in 5iLAF-cultured naive hPSCs and cells of the blastocyst, we found several key differences. First, the mono-allelic *XIST*-expressing XaXa state was predominant in naive hPSCs. The Xa^{XIST+} Xa^{XIST+} pattern, perfectly recapitulating that of the blastocyst, was consistently observed, yet only in a subset of cells. Second, the observation of non-random XCI upon differentiation from the naive state does not reflect the randomness of the process normally seen in development. Moreover, the fact that it is the same inactive X chromosome of the starting primed hESCs that be-comes inactivated in differentiated naive cells indicates the existence of an epigenetic memory for the prior Xi that does not get erased in the naive state. Third, we observed an accumulation of the repressive H3K27me3 histone modification on the *XIST*-coated Xa that has not been described for the blastocyst (Okamoto et al., 2011). However, it is unclear whether the lack of H3K27me3 accumulation on the X in cells of the blastocyst is due to differential staining of cultured cells and embryos or a reflection of a different X chromosome state. Regardless, together these findings indicate that additional

modifications to the naive culture media are required to reset these features to the most faithful state. However, it is noteworthy that naive hESCs directly derived from the blastocyst in 5iLAF media had higher proportion of Xa^{XIST+} Xa^{XIST+} cells. Additionally, our data suggest that the non-randomness of XCI and the memory of the prior Xi from the primed state in differentiating naive cells are not linked to the mono-allelic expression status of *XIST* in the naive state. Single-cell and population analyses revealed that either of the two X chromosomes is capable of upregulating *XIST* in naive cells (Figure S3-3-2F), yet upon differentiation, there is very heavy skewing toward inactivating the prior Xi, resulting in non-random XCI. Moreover, *XIST* expression is silenced in the transition from primed to naive and prior to differentiation in the re-primed state, arguing against *XIST*'s direct involvement in Xi memory.

Xi reactivation was not limited to the 5iLAF culture condition. Interestingly, we found that the t2iL+ Gö culture media (Takashima et al., 2014) also induced Xi reactivation in primed hPSCs in all cells and mono-allelic *XIST* expression from an active X chromosome in a subset of cells. This result is in agreement with global gene expression studies that have placed naive cells generated by the 5iLAF and t2iL+Gö conditions closest to the blastocyst state (Huang et al., 2014). The detection of the *XIST*-expressing Xa in two independent naive culture conditions establishes the *XIST*-expressing active X chromosome as a hallmark of human naive pluripotency in cultured cells. The Xa^{XIST+} provides a straightforward readout for characterizing the constantly evolving naive media formulations, which will immensely help in improving naive culture conditions. Interestingly, we found that, in t2iL+Gö condition (Takashima et al., 2014), most cells did not express *XIST* despite harboring two active X chromosomes, consistent with the notion that the different media compositions for naive cells may capture pluripotent states at slightly different stages of development (Huang et al., 2014).

Another interesting outcome of our work is the finding that the transition from an *XIST*-expressing Xa in the naive state to the *XIST*-expressing Xi in differentiated cells occurs through an intermediate in which both X chromosomes are actively transcribed but *XIST* expression is off. Although both *XIST*-negative and positive naive cells had the ability to undergo de novo XCI upon differentiation, gene expression and DNA methylation analysis indicated *XIST* expression from an active X chromosome as a more stringent marker of the blastocyst-like naive pluripotent state than the presence of two active X chromosomes alone. The *XIST*-negative XaXa intermediate was also observed during Xi reactivation in the reverse process during the transition from primed XaXi hPSCs to naive *XIST*-expressing XaXa cells. Hence, we speculate that, in the developing human embryo, *XIST* expression turns off during the transition from XCD to XCI. Intriguingly, during derivation of primed hESC from blastocysts, we found that cells rapidly lost *XIST* expression from both active X chromosomes, yielding a mixture of XaXa cells without *XIST* and Xa^{XIST-} Xi^{XIST+} cells, where *XIST* is expressed from the Xi in a post-XCI manner (Patel et al., 2016). This further supports our classification of the *XIST*-negative XaXa cells as a developmental intermediate in the transition from pre- to post-XCI.

In both the primed to naive transition in vitro and human pre-implantation development, the dampening of X-linked gene expression correlated with upregulation of *XIST*, suggesting that this unique form of X chromosome dosage compensation may be *XIST* dependent. If this was the case, dampening should only affect one X chromosome in naive hPSCs with mono-allelic *XIST* and both X chromosomes in cells with bi-allelic *XIST*. Understanding this requires in-depth examination of X-linked allele-specific gene expression patterns at the single-cell level so that the expression of the *XIST*-expressing X can be compared to that of the *XIST*-negative one in the same cell. Such analyses require multiple cells with deeply sequenced high-quality data for meaningful

conclusions, and unfortunately, our single-cell analyses were not sufficient for resolving this question. The dampened expression of X-linked genes in naive hPSCs may explain why originally described naive female hESCs were thought to harbor an Xi based on global gene expression analysis (Theunissen et al., 2014). Hence, methods beyond global gene expression changes should be used to assess the X chromosome status of hESCs.

Lastly, our findings demonstrated that the exposure to the naive state reverses the epigenetic abnormalities of the X_i prevalent in female primed hiPSC and hESC lines, including loss of *XIST*, Xi erosion, and the inability to induce X inactivation upon differentiation. Because erroneous X chromosome dosage compensation has been linked to developmental abnormalities and cancer in both mouse and human (Schulz and Heard, 2013), this result is critical for the application of female hPSCs, as it enables the generation of differentiated cells with properly dosage-compensated X chromosomes. One concern is that imprint methylation erasure and karyotypic abnormalities are observed in the 5iLAF-based naive culture condition. We observed faithful XCI in hPSCs that were kept in the naive condition for only few passages, and these cells remain karyotypically normal and maintain methylation at imprint control regions (Theunissen et al., 2014; Pastor et al., 2016; Figure S3-3-3C). Hence, our work suggests a path toward hPSCs without undesirable genetic and epigenetic abnormalities, which is more desirable for cell replacement therapies and more apt for studies of basic development and diseases.

EXPERIMENTAL PROCEDURES

Cell culture

Primed hPSCs were cultured on feeder cells and passaged with collagenase IV every 6 or 7 days. Primed hPSCs were converted to 5iLAF naive condition as described (Theunissen et al., 2014). Briefly, 2 3 10^5 primed single cells were plated on feeders in primed media with ROCK inhibitor Y-27632 for 2 days before switching to 5iLAF. Ten days later, surviving cells were passaged as single cells. For re-priming, 5iLAF media was changed to primed when naive colonies were of medium size, and thereafter, cells were treated exactly as primed cultures. For differentiation, primed or re-primed hPSC cultures were depleted of feeder cells and plated as single cells on Matrigel-coated plates or coverslips for 7 days in fibroblast media. Human embryo studies in this work received the approval of the UCLA Institutional Review Board (IRB#11- 002027) and the UCLA Embryonic Stem Cell Research Oversight (ESCRO) Committee (2008-015 and 2007-009) (see Supplemental Experimental Procedures for more details).

RNA FISH and microscopy

Cells grown on gelatinized glass coverslips were fixed with 4% formaldehyde, permeabilized with 0.5% Triton X-100 (10 min each), and serially dehydrated with 70%–100% ethanol. Coverslips were hybridized with labeled DNA probes generated from bacterial artificial chromosomes (BACs). The Imager M1 microscope (Zeiss) was used for acquiring and ImageJ software (NIH) for processing z stack images (see Supplemental Experimental Procedures for more details).

RNA sequencing

RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina). The C1 Single-Cell Auto Prep System (Fluidigm) was used for single-cell RNA-seq. Allelic expression proportions were determined either by read counts (RNA-seq) or Sanger sequencing of exonic SNPs of X-linked genes (see Supplemental Experimental Procedures for more details).

DNA methylation

Genomic DNA was harvested from replicate samples, and libraries for RRBS were created as previously described (Meissner et al., 2005). Only CpG sites covered by at least five reads across all samples under consideration were used for data presentation (see Supplemental Experimental Procedures for more details).

Accession numbers

The accession number for the RNA-seq and RRBS data reported in this paper is GEO: GSE87239.

FIGURE LEGENDS

Figure 3-1. The Xi of primed hESCs reactivates in the 5iLAF culture condition

(A) Representative RNA FISH images for primed UCLA1 at passage 16 (P16), detecting *XIST* and nascent transcription foci of *UTX* (escapes XCI) and *HUWE1* (subject to XCI). Panels show *HUWE1* and *UTX* expression without *XIST* (left), *XIST* only (middle), and all three channels together (right) in DAPI-stained nuclei (blue). A nucleus with the most-prevalent pattern is highlighted with a dotted box and enlarged, and its pattern is depicted by the cartoon on the right.

(B) As in (A) but for the X-linked genes *ATRX* (top) or *THOC2* (bottom), showing only the merged panel with *XIST* and *UTX*. (C) Quantification of the RNA FISH patterns for *XIST* and the X-linked genes *HUWE1*, *ATRX*, and *THOC2* in primed UCLA1 as shown in (A) and (B). Only cells with bi-allelic *UTX* expression were considered.

(D) Representative RNA FISH images of the most- prominent X chromosome state in naive UCLA1 at P19 in 5iLAF media, detecting *HUWE1*, *XIST*, and *UTX*, similar to (A).

(E) As in (D) but detecting the X-linked genes *ATRX* (top) or *THOC2* (bottom). (F) Quantification of the RNA FISH patterns for *XIST* with *HUWE1*, *ATRX*, and *THOC2*, respectively, in naive UCLA1 as shown in (D) and (E) in cells with bi-allelic *UTX* expression.

(G) Representative RNA FISH images similar to (D) but for the bi-allelic *XIST* (top) and *XIST*-negative (bottom) X_aX_a patterns.

Figure 3-2. Xi reactivation leads to an *XIST*-negative XaXa intermediate before induction of *XIST* expression from the Xa

(A) Quantification of the RNA FISH patterns for *XIST* and *HUWE1* as UCLA1 progressed from the primed state at P12 to the naive pluripotent state up to P7 in 5iLAF media. All counts were in naive cells with two *UTX* foci.

(B) Allelic expression proportions of X-linked genes in the primed UCLA1 population, the naive UCLA1 clone 4 (at early-passage, *XIST*-negative state), and the naive *XIST*-positive UCLA1 clones 9 and 12, based on reads covering indicated SNPs, ordered along the X chromosome (depicted with red lines in the X chromosome image below), in replicate RNA-seq data (rep1 and rep2; closely spaced bars). SNPs located in the same gene were placed next to each other and marked with an asterisk (*). ND, not determined due to insufficient read coverage.

(C) Analysis of allelic expression proportions of X-linked genes in single *XIST*-negative and *XIST*-positive naive UCLA1 cells. Single cells of naive UCLA1 at early passage, when the population was still largely *XIST* negative, and at later passage, when most cells were *XIST* positive, were grouped into *XIST*-positive and *XIST*-negative cells based on their *XIST* expression levels. The two graphs on the right show the allelic expression proportion of X-linked genes for each single cell with sufficient SNP coverage in these two groups, with each circle representing the proportion of one SNP in one single cell. Highlighted with yellow is the 20%–80% range region, where we considered X-linked gene expression from both X chromosomes to take place. For comparison, the graph on the left plots the allelic ratio of the same SNPs used in the right in primed UCLA1 as described in (B).

(D) Heatmap of unsupervised hierarchical clustering of RRBS-based methylation levels of CpGs within X-linked CGIs. The number of CpGs considered is given. Mb (megabase) indicates the relative position on the X chromosome.

(E) Heatmap as in (C) but for all X-linked CpGs covered by RRBS.

(F) Methylation averages of X-linked and autosomal CpGs outside (top) and within CGIs (bottom) in primed UCLA1 and the naive clones 4 (*XIST* negative), 9, and 12 (*XIST* positive).

(G) Empirical cumulative distribution functions of X-linked gene expression for naive UCLA1 clone 4 at early (*XIST*-negative) and late (*XIST*-positive) passage and the *XIST*-positive clones 9 and 12, based on replicate RNA-seq datasets where available (rep1 and rep2). The inset shows autosomal expression data for the same samples in the same format. Asterisk (*) indicates statistically significant difference between the distributions of X-linked gene expression of both replicates of early-passage (*XIST*-negative) clone 4 compared to all other samples ($p < 0.0035$ by Wilcoxon rank sum test with continuity correction). The X-linked and all autosomal distributions for all other samples were not significantly different from each other.

(H) Same as in (G) but for early (*XIST*-negative) and late (*XIST*-positive) passage naive UCLA4 hESCs. X-linked, but not autosomal, gene expressions of both replicates of early-passage (*XIST*-negative) samples were significantly different from replicate 2 (but not replicate 1) of late-passage (*XIST*-positive) cells ($p < 0.05$).

Figure 3-3. The XIST-expressing XaXa state is inherent to naive hESC lines, regardless of source

(A) Immunofluorescence detection of H3K27me3 (red) combined with RNA FISH for *XIST* (green) in normal female fibroblasts (top), the naive UCLA1 clone12 (middle), and blastocyst-derived naive UCLA20n (bottom) cells.

(B) As in (A), except for RNA polymerase II (RNA PolII) and *XIST*. Yellow arrowheads demonstrate regions in nuclei devoid of RNA PolII signal under the *XIST* signal.

(C) Quantification of RNA FISH patterns of transcription foci of *HUWE1*, *ATRX*, and *THOC2*, respectively, and of *XIST* in primed, 5iLA-, and 4iLA-cultured naive WIBR3 hESCs.

(D) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in the blastocyst-derived naive hESC lines UCLA19n and UCLA20n. RNA FISH patterns were stable throughout passaging and in the graph are quantified at P3 for both lines.

Figure 3-4. Naive hESCs initiate XCI upon differentiation

(A) Representative RNA FISH images for *XIST*, *HUWE1*, and *UTX* in naive UCLA1 at P42 in 5iLAF, in the re-primed condition at P7, and after 7 days of differentiation (Diff) from the re-primed state. A nucleus with the most-prevalent pattern is highlighted with a dotted box, enlarged, and depicted by the cartoon on the right.

(B) Quantification of *XIST*-expressing differentiated cells with different *HUWE1* transcription

patterns (bi-allelic with equal transcription foci on both X chromosomes, bi-allelic with a smaller transcription focus on the *XIST*-coated chromosome, and mono-allelic) in cells with bi-allelic *UTX* expression.

(C) Representative RNA FISH images as in (A) but for the naive hESC line UCLA20n at P8 (naive), P4 (re-primed), and after 7 days of differentiation of re-primed state.

Figure 3-5. XCI in differentiating naive cells is non-random

(A) Representative electropherograms from Sanger sequencing of SNP-containing regions in the *HUWE1* and *XIST* cDNA obtained from the primed UCLA1 population (at P19 for *HUWE1* and P4 for *XIST* analysis; top) and from an individual cell differentiated from the re-primed state after transition through the naive state (bottom). Pie charts summarize the Sanger sequencing results for additional *XIST*-positive, single-differentiated cells (43 individual cells for *HUWE1*; 47 for *XIST*), considering three expression categories: expression of only reference (Ref) allele; only alternate (Alt) allele; or both alleles.

(B) Allelic expression proportions of *XIST* based on RNA-seq reads (number of reads indicated on top) that covered the indicated SNP in single cells differentiated from naive UCLA1 after transition through the re-primed state. The x axis refers to single-cell coordinates in a 96-well plate. For comparison, the allelic *XIST* proportion in primed UCLA1 was estimated from the Sanger sequencing electropherograms in (A) (top).

(C) As in (B) but for the X-linked genes *RBM3* (two different SNPs), *IDS*, *MPP1*, and *TAF9B*. Only *XIST*-expressing single cells were analyzed. Allelic expression proportions in primed

UCLA1 were calculated from replicate RNA-seq data of the primed cell population (Pri1 and Pri2).

(D) Representative immunofluorescence images detecting MeCP2 and CD44 in hiPSC clones 16 and 17 harboring an Xa with wild-type and mutant *MECP2*, respectively. Shown are primed hiPSCs (P17; top), re-primed cells (P3) after transition through the naive 5iLAF for three passages (middle), and the day 7 differentiation product of the re-primed state (bottom).

(E) Schematic summary of X chromosome dynamics as primed XaXi cells with or without *XIST* expression from the Xi transition to naive pluripotency and differentiate.

Figure 3-6. Naive pluripotency enables XCI in primed hESCs with an aberrant XaXa state

(A) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in primed UCLA9 (P15) and upon 7 days of differentiation. A nucleus with the most-prevalent pattern is highlighted with a dotted box, enlarged, and depicted by the cartoon.

(B) Quantification of RNA FISH patterns of *XIST* in combination with *HUWE1*, *ATRX*, and *THOC2*, respectively, in the primed state and its differentiated daughter cells. Only cells with bi-allelic *UTX* expression were quantified.

(C) Quantification of RNA FISH patterns of *XIST* and *HUWE1* in UCLA9 at indicated passages post- 5iLAF conversion. Only cells with bi-allelic *UTX* expression were quantified.

(D) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in UCLA9 in the naive state (P22), the re-primed state obtained by culturing naive cells in the primed culture condition for three passages, and in day 7 differentiated cells derived from re-primed cells. In each row, a

nucleus with the most-prevalent pattern is highlighted with a dotted box, enlarged, and depicted by the cartoon on the right.

(E) As in (B) but for re-primed state.

(F) As in (B) but for differentiated UCLA9 described in (E).

(G) Schematic summary of the X chromosome state of UCLA9 in indicated states.

Figure 3-7. Naive pluripotency erases Xi erosion of primed hESCs

(A) Quantification of RNA FISH patterns of *XIST* and *HUWE1* or *XIST* and *ATRX* in primed UCLA4 and upon 7 days of differentiation from the primed state. Only cells with bi-allelic *UTX* expression were quantified.

(B) Quantification of RNA FISH patterns of *XIST* and *HUWE1* in UCLA4 at indicated passages post- 5iLAF conversion. Only cells with bi-allelic *UTX* expression were quantified.

(C) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in UCLA4 in the naive state (P7), in the re-primed state obtained from the naive state, and in day 7 differentiated cells derived from re-primed cells. A nucleus with the most-prevalent pattern is highlighted with a dotted box, enlarged, and depicted by the cartoon.

(D) Quantification of RNA FISH patterns in re-primed UCLA4 and their differentiated progeny in cells with bi-allelic *UTX* expression as in (A). Only *XIST*-expressing cells were considered in the differentiated state.

(E) Representative RNA FISH image detecting *XIST* RNA and nascent transcription foci of *ATRX* in differentiated cells originating from re-primed cells derived from naive UCLA4.

(F) Schematic summary of the X chromosome states of UCLA4 in indicated states.

Table 3-1. Summary of the X chromosome state in primed, naïve, re-primed and differentiated cells described in this study

Summary of the X chromosome states for cell lines analyzed in this study in primed, naïve, re-primed and differentiated conditions, as applicable.

Figure 3-1. The Xi of primed hESCs reactivates in the 5iLAF culture condition

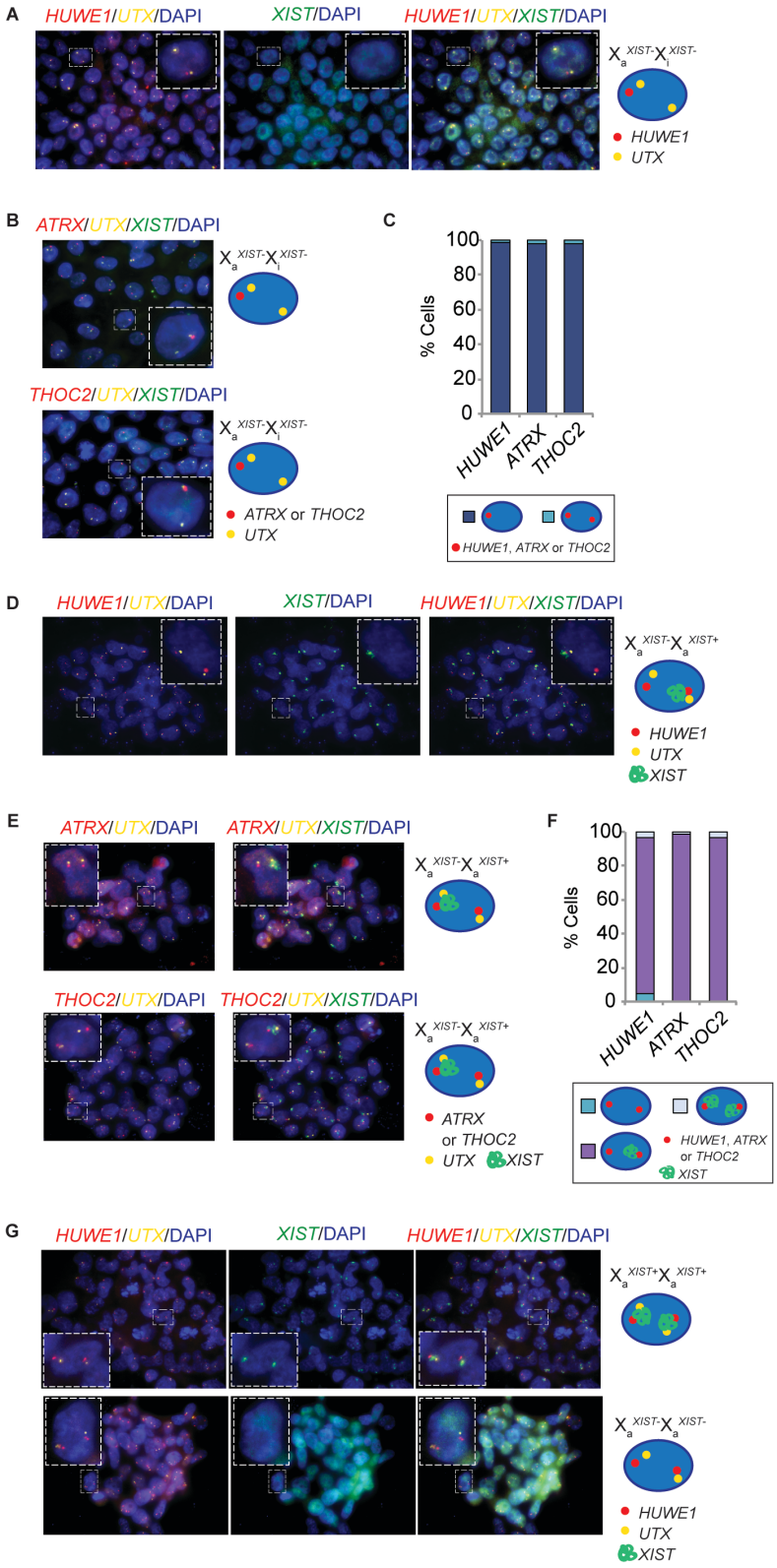


Figure 3-2. Xi Reactivation leads to an *XIST*-negative XaXa intermediate before induction of *XIST* expression from the Xa

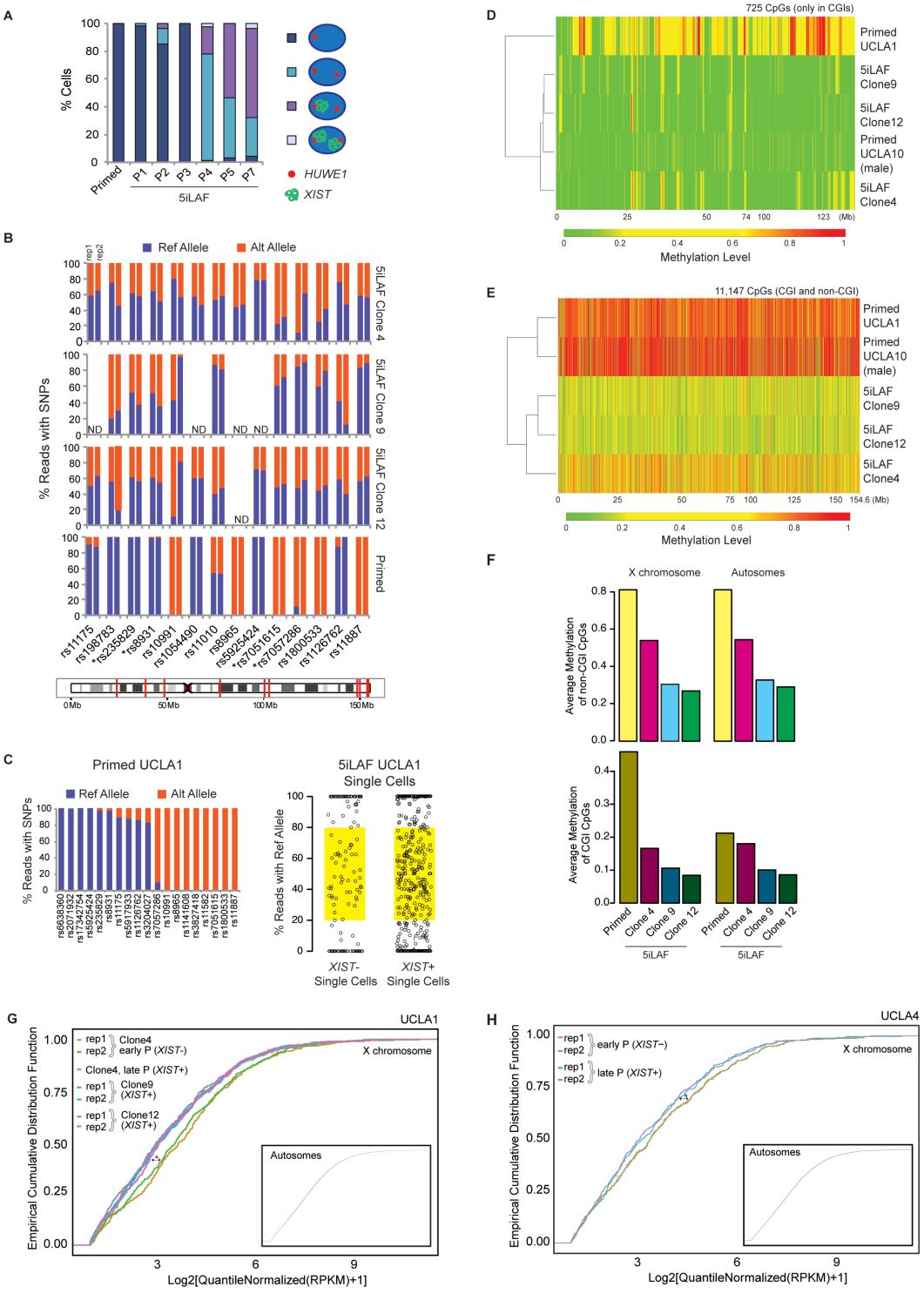


Figure 3-3. The XIST-expressing XaXa state is inherent to naive hESC lines, regardless of source

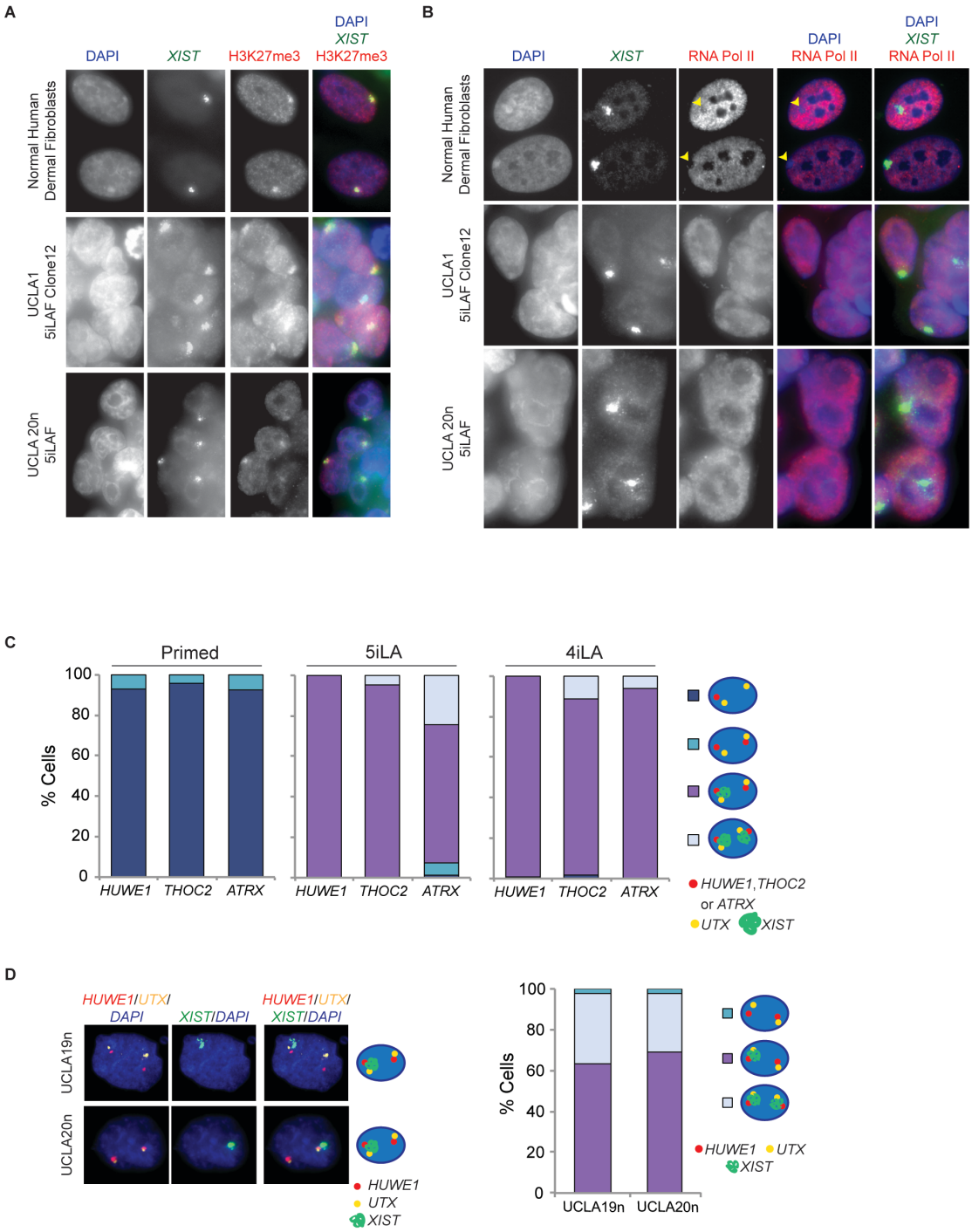


Figure 3-4. Naive hESCs initiate XCI upon differentiation

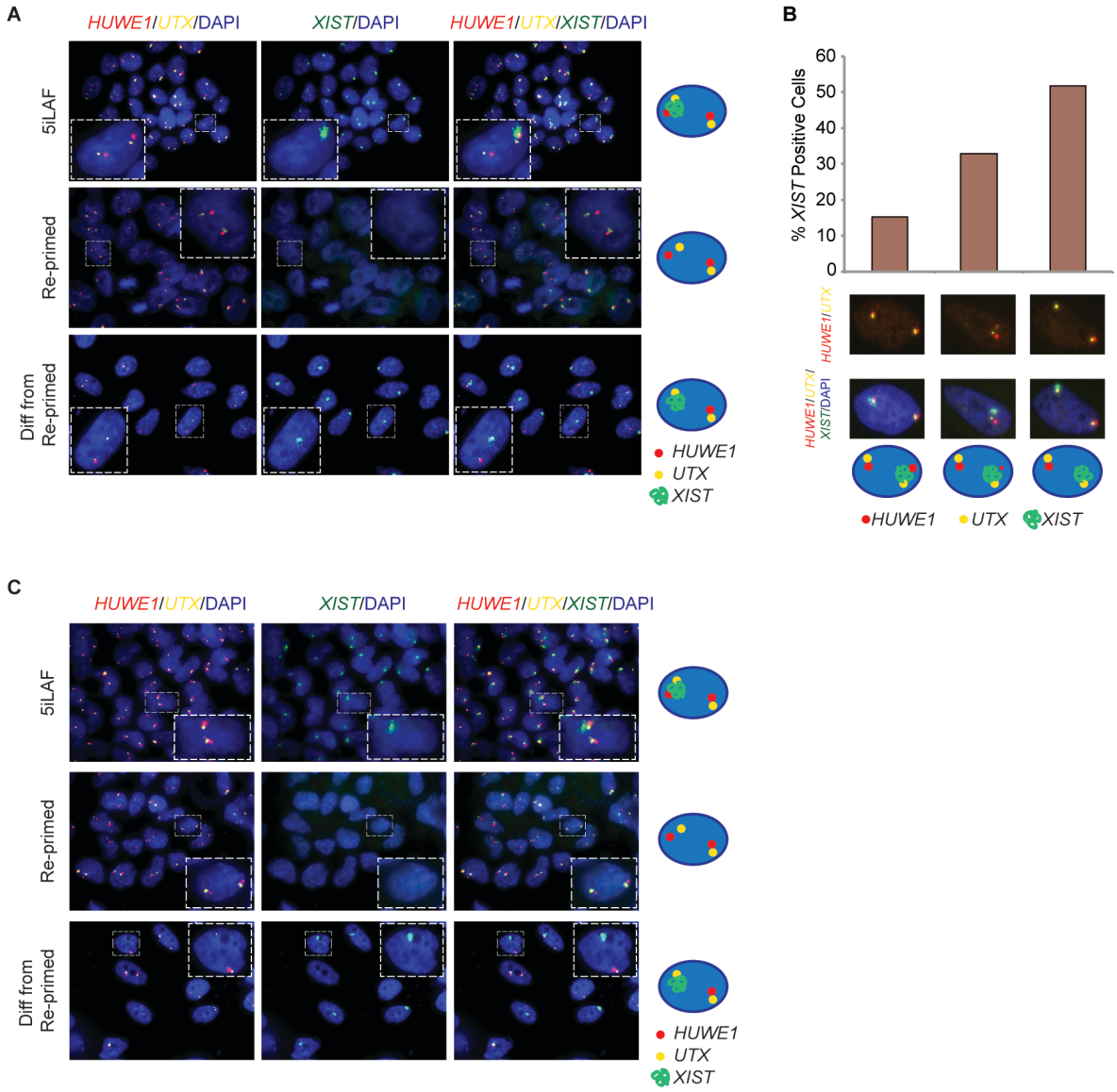


Figure 3-5. XCI in differentiating naive cells is non-random

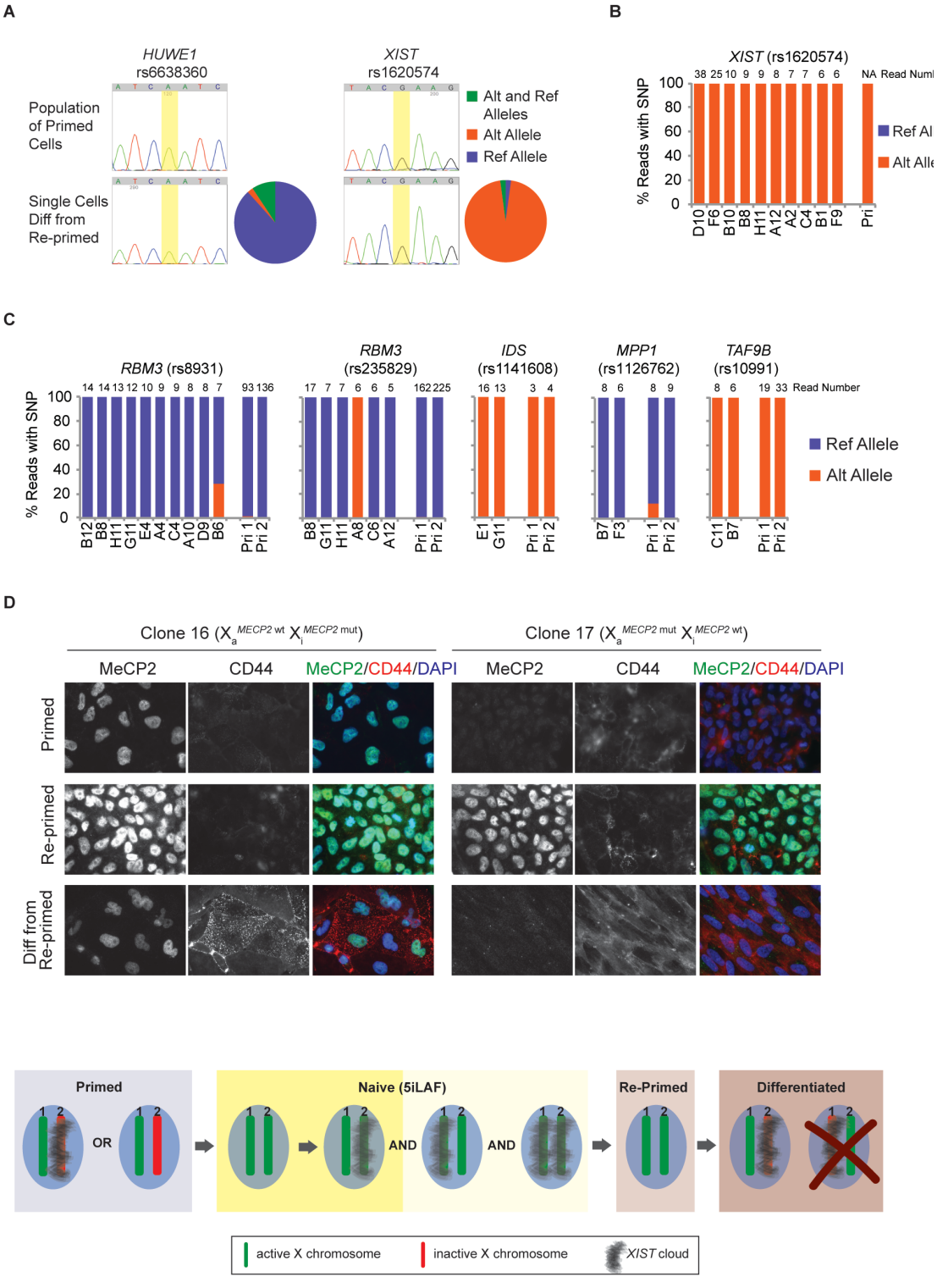


Figure 3-6. Naive pluripotency enables XCI in primed hESCs with an aberrant XaXa state

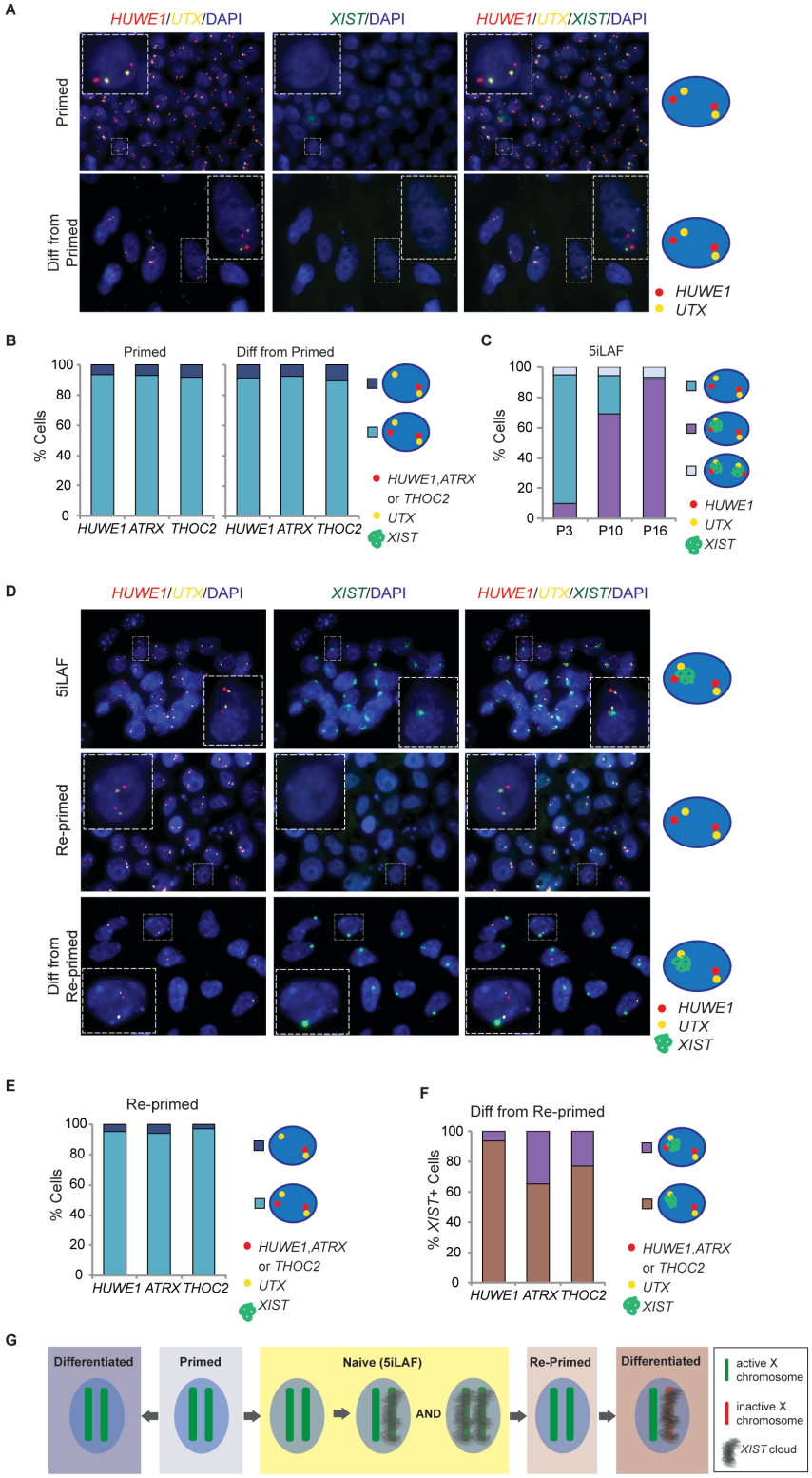


Figure 3-7. Naive pluripotency erases Xi erosion of primed hESCs

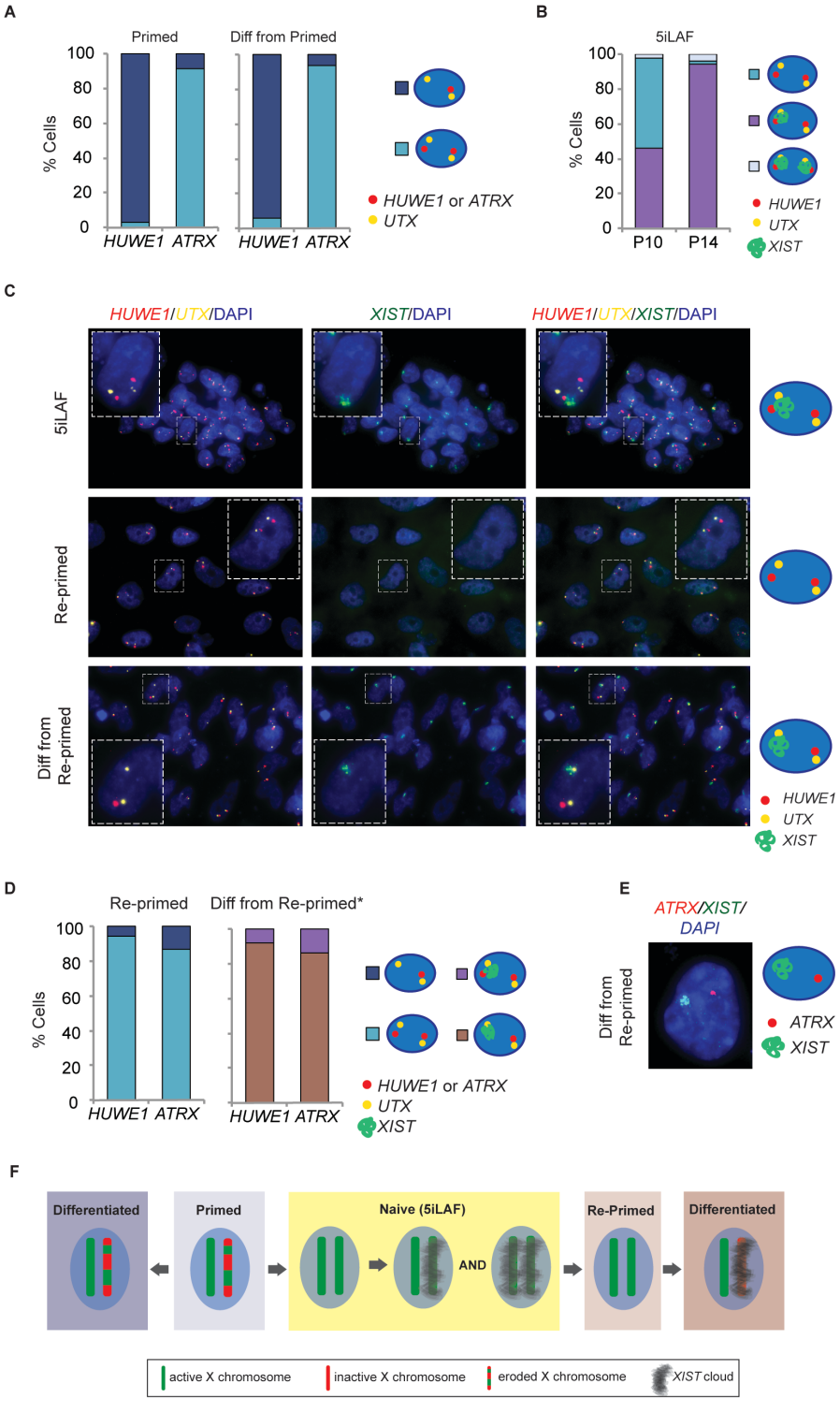


Table 3-1. Summary of the X chromosome state in primed, naïve, re-primed and differentiated cells described in this study

	Cell Line	Primed	
		Early Passage	Late Passage
Derived from blastocyst in the primed state, then converted to naïve pluripotency in culture	UCLA1	XaXi, XIST from Xi#	XaXi*, no XIST
	UCLA4	N/A	XaXe, no XIST
	UCLA9	N/A	XaXa, no XIST
	WIBR3	N/A	XaXi, no XIST
Derived from blastocyst in the naïve state	UCLA 19n	N/A	N/A
	UCLA 20n	N/A	N/A
Reprogrammed from dermal fibroblasts to the primed state, then converted to naïve pluripotency in culture	iPSC #1001	XaXi, XIST from Xi	N/A
	iPSC Clone16 (MECP2 het)	N/A	XaXi, no XIST
	iPSC Clone17 (MECP2 het)	N/A	XaXi, no XIST

	Naïve	
	Early Passage	Late Passage
Derived from blastocyst in the primed state, then converted to naïve pluripotency in culture	XaXa, no XIST	XaXa, mono- and bi-allelic XIST
	XaXa, no XIST	XaXa, mono- and bi-allelic XIST
	XaXa, no XIST	XaXa, mono- and bi-allelic XIST
	N/A	XaXa, mono- and bi-allelic XIST
Derived from blastocyst in the naïve state	XaXa, mono- and bi-allelic XIST	XaXa, mono- and bi-allelic XIST
	XaXa, mono- and bi-allelic XIST	XaXa, mono- and bi-allelic XIST
Reprogrammed from dermal fibroblasts to the primed state, then converted to naïve pluripotency in culture	XaXa, no XIST	XaXa, mono- and bi-allelic XIST
	XaXa, no XIST	N/A
	XaXa, no XIST	N/A

	Re-primed	Differentiated	
	(from Naïve)	**From Late Primed**	**From Naïve/Re-primed**
Derived from blastocyst in the primed state, then converted to naïve pluripotency in culture	XaXa, no XIST	XaXi*, no XIST	XaXi, XIST from Xi
	XaXa, no XIST	XaXe, no XIST	XaXi, XIST from Xi
	XaXa, no XIST	XaXa, no XIST	XaXi, XIST from Xi
	N/A	N/A	N/A
Derived from blastocyst in the naïve state	XaXa, no XIST	N/A	XaXi, XIST from Xi
	XaXa, no XIST	N/A	XaXi, XIST from Xi
Reprogrammed from dermal fibroblasts to the primed state, then converted to naïve pluripotency in culture	N/A	N/A	N/A
	XaXa, no XIST	XaXi, no XIST	XaXi, XIST from Xi
	XaXa, no XIST	XaXi, no XIST	XaXi, XIST from Xi

REFERENCES

Balaton, B.P., Cotton, A.M., and Brown, C.J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol. Sex Differ. *6*, 35. Chan, Y.-S., Go¨ke, J., Ng, J.-H., Lu, X., Gonzales, K.A.U., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., and Ng, H.-H. (2013). Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplanta- tion epiblast. Cell Stem Cell *13*, 663–675.

Davidson, K.C., Mason, E.A., and Pera, M.F. (2015). The pluripotent state in mouse and human. Development *142*, 3090–3099.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben- Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282–286.

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc. Natl. Acad. Sci. USA *107*, 9222–9227.

Huang, K., Maruyama, T., and Fan, G. (2014). The naive state of human plurip- otent stem cells: a synthesis of stem cell and preimplantation embryo tran- scriptome analyses. Cell Stem Cell *15*, 410–415. Lee, S.S., Wan, M., and Francke, U. (2001). Spectrum of MECP2 mutations in Rett syndrome. Brain Dev. *23* (*Suppl 1*), S138–S143.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high- resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877. Mekhoubad, S., Bock, C., de Boer, A.S., Kiskinis, E., Meissner, A., and Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. Cell Stem Cell *10*, 595–609.

Minkovsky, A., Patel, S., and Plath, K. (2012). Concise review: pluripotency and the transcriptional inactivation of the female Mammalian X chromosome. Stem Cells *30*, 48–54. Nazor, K.L., Altun, G., Lynch, C., Tran, H., Harness, J.V., Slavin, I., Garitaonandia, I., Mu€ller, F.-J., Wang, Y.-C., Boscolo, F.S., et al. (2012). Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell *10*, 620–634.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell *4*, 487–492.

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

Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., and Clark, A.T. (2016). Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. Cell Stem Cell *18*, 323–329.

Patel, S., Bonora, G., Sahakyan, A., Kim, R., Chronis, C., Langerman, J., Fitz- Gibbon, S., Rubbi, L., Skelton, R.J.P., Ardehali, R., et al. (2016). Human embryonic stem cells do not change their X-inactivation status during differen- tiation. Cell Reports *18*, in press. Published online December 15, 2016. http:// dx.doi.org/10.1016/j.celrep.2016.11.054.

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

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 his- tone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

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

Schulz, E.G., and Heard, E. (2013). Role and control of X chromosome dosage in mammalian development. Curr. Opin. Genet. Dev. *23*, 109–115.

Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of hu- man active and inactive X chromosomes. Genome Res. *21*, 1592–1600.

Shen, Y., Matsuno, Y., Fouse, S.D., Rao, N., Root, S., Xu, R., Pellegrini, M., Riggs, A.D., and Fan, G. (2008). X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. Proc. Natl. Acad. Sci. USA *105*, 4709–4714.

Silva, S.S., Rowntree, R.K., Mekhoubad, S., and Lee, J.T. (2008). X-chromo- some inactivation and epigenetic fluidity in human embryonic stem cells. Proc. Natl. Acad. Sci. USA *105*, 4820–4825.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611–615.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254– 1269.

Tchieu, J., Kuoy, E., Chin, M.H., Trinh, H., Patterson, M., Sherman, S.P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J.A., et al. (2010). Female hu- man iPSCs retain an inactive X chromosome. Cell Stem Cell *7*, 329–342.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive hu- man pluripotency. Cell Stem Cell *15*, 471–487.

Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J., Wang,

H., Iouranova, A., Imbeault, M., et al. (2016). Molecular criteria for defining the naive human pluripotent state. Cell Stem Cell *19*, 502–515.

Vallot, C., Ouimette, J.-F., Makhlouf, M., Feˊraud, O., Pontis, J., Coˆme, J., Martinat, C., Bennaceur-Griscelli, A., Lalande, M., and Rougeulle, C. (2015). Erosion of X chromosome inactivation in human pluripotent cells initiates with XACT coating and depends on a specific heterochromatin landscape. Cell Stem Cell *16*, 533–546.

Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., et al. (2014). Derivation of naive human embryonic stem cells. Proc. Natl. Acad. Sci. USA *111*, 4484–4489.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

SUPPLEMENTAL INFORMATION: FIGURE LEGENDS

Figure S3-1. Characterization of naïve hESCs and the primed to naïve transition

(A) Phase-contrast images of primed and 5iLAF-adapted UCLA1 hESCs.

(B) Fold-change of transcript levels between the naïve UCLA1 clones 4 (early passage), 9, 12 and the starting primed UCLA1 population for genes known to be up-regulated in the naïve compared to the primed state (Theunissen et al., 2014). For comparison, averaged expression ratios for the same genes for originally described 5iLAF-cultured hESCs were plotted (obtained from Table S2 of Theunissen et al., 2014). *ZFP42* is an outlier in our data because primed UCLA1 expressed *ZFP42* more highly than the primed WIBR2 and WIBR3 hESC lines used by Theunissen et al., (2014), as it is among the top 3% of expressed genes in UCLA1 vs. the bottom 2% in WIBR2 and WIBR3.

(C) Representative RNA FISH images detecting the lncRNAs *XACT* and *XIST* in primed and naïve UCLA1 cells at P15 and P24, respectively.

(D) Representative RNA FISH images detecting *XIST*, *HUWE1* and *UTX* in primed (P62) and naïve t2iL+Gö (P14 in the naïve media) H9 hESCs, and for *XIST, XACT* and *UTX* in primed H9. The UTX, typically escaping XCI, displayed only mono-allelic pattern in the primed H9 cells, possibly due to it not escaping in this particular cell line at the primed state (see *HUWE1/UTX/XIST* stain in primed H9). However, the *XACT* RNA-FISH in primed H9 is expressed from two sites due to Xi-erosion, which demonstrates the presence of two X chromosomes in these cells.

(E) Representative RNA FISH images detecting *XIST*, *HUWE1* and *UTX* in the naive UCLA1 clone4 at early passage (P3, *XIST*-negative) and after it became predominantly *XIST*-positive

(P12). The most prevalent RNA-FISH pattern in each condition is depicted by the cartoons on the right.

(F) RPKM values of *XIST* normalized to that of *ACTB* in primed UCLA1 and naïve clones 4, 9 and 12. Both early passage (replicates rep1 and rep2, P6) and late passage (P16) clone4 were included, demonstrating the transition of this clone from *XIST*-negative to the *XIST*-positive state, in agreement with data shown in (E).

(G) The early passage (P6) primed female hiPSC line #1001 (Karumbayaram et al., 2012) with about 55% of cells carrying an *XIST*-coated Xi, was converted to the naïve state with the 5iLAF approach. RNA FISH staining patterns detecting *XIST*, *HUWE1* and *UTX* were quantified as cells progressed from primed to naïve pluripotency at the indicated passages. All counts were in cells with only two *UTX* foci. Representative RNA FISH images of single nuclei capturing the different RNA FISH patterns of X chromosome in primed hESCs (top) and upon conversion to the naïve state (bottom) are shown, with corresponding cartoons.

Figure S3-2. The transition from primed to naïve pluripotency leads to Xi-reactivation

(A) Allelic expression proportions of X-linked genes in primed UCLA1 based on ten or more reads covering indicated SNPs in replicate RNA-seq data (rep1, rep2, closely-spaced bars). SNPs located in the same gene were placed next to each other and marked with an asterisk (*). Genes known to be subject to XCI, to escape from XCI, to be located in pseudo-autosomal regions (which is not subject to XCI), or found to display conflicting XCI-states in different studies (discordant) were color-coded as by Balaton et al. (2015). Chromosomal position of the SNP-carrying genes is given

along the X chromosome on the bottom of the bar graph.

(B) Allelic expression proportions of X-linked genes in naïve UCLA1 clone4 (early passage, P6), as described in (A).

(C) Allelic expression proportions of X-linked genes in naïve UCLA1 clone9, as described in (A).

(D) Allelic expression proportions of X-linked genes in naïve UCLA1 clone12, as described in (A).

(E) Electropherograms from Sanger sequencing of cDNA obtained from the primed UCLA1 population and the naïve UCLA1 clone12, detecting the indicated SNP in the X-linked genes *MID1IP1*, *RBM3*, *HUWE1, UBL4A* (subject to XCI) and *HDHD1* (escapes XCI). Chromosomal position of the SNP- containing genes along the X chromosome is given below. These electropherograms for the *HUWE1* SNP in primed UCLA1 was also used in main Figure 5A, and for the *HDHD1* SNP in primed UCLA1 in Figure S5A, to make data comparison easy in those figures.

(F) Electropherogram of Sanger sequencing of cDNA obtained from late passage naïve UCLA1 cells demonstrating the detection of both the reference (A) and alternate (G) SNP rs1620574 (in the last exon of *XIST*). The graph below gives allelic expression proportions of the same SNP in *XIST*-expressing single cells of naïve UCLA1 cells. Each bar represents a single cell from single cell RNA-seq of early and late passage naïve UCLA1 (two different single cell RNA-seq experiments (exp1 or 2, marked along the X-axis with labels of each cell coordinate in a 96 well plate format). Experiment 1 (exp1) contained cells from late passage naïve culture, whereas experiment 2 (exp2) was from an early passage (before all cells became *XIST*-positive). Twenty-

six single cells expressed solely the alternate SNP (orange; mono-allelic *XIST*), 7 cells solely the reference SNP (blue; mono-allelic *XIST*), and 13 cells both the reference and alternate SNPs (part orange and part blue, evidence for bi-allelic expression of *XIST* at the single-cell level).

Figure S3-3. Naïve *XIST*-positive hESCs resemble the human blastocyst more closely than the intermediate *XIST*-negative cells

(A) Heatmap of unsupervised hierarchical clustering of RRBS-based methylation levels of covered CpGs (number is given) in autosomal CGIs in indicated cell lines and states.

(B) Heatmap of DNA methylation as in (A), but of all covered autosomal CpGs.

(C) Violin plots of methylation levels of covered CpGs within primary imprinted control regions (Okae et al., 2014) in indicated cell lines and states, based on RRBS data. UCLA10 is a male primed hPSC line included for comparison.

(D) Empirical cumulative distribution functions of differential methylation values of each covered CpG in pairwise comparisons between primed UCLA1 and naïve UCLA1 clone4 (at early passage when it was largely *XIST*-negative), and the *XIST*-positive naïve UCLA1 clones 9 and 12, and among the clones, for CpGs within and outside of CGIs, separated by X chromosome and autosomes. P-values for the difference test of these pairwise comparisons were determined using the Kolmogorov-Smirnov (KS) test and shown below (p=0 is red, p=1 is green). Note that when distributions including clone4 (*XIST*-negative) were compared to any other pair-wise distributions, the p-values were much higher only for X-linked CGIs, indicating that the methylation loss in clone4 relative to primed UCLA1 was far more striking in X-linked CGIs compared to non-CGIs

in both chromosome X and autosomal contexts as well as autosomal CGIs.

(E) K-means clustering of 3701 genes differentially expressed among primed UCLA1 and naïve clones 4 (early passage), 9 and 12 (see Table S2 for a list of these genes). The average expression level for all genes in a cluster is shown and the number of genes (n) in each cluster is given on the right.

(F) Expression levels in replicate RNA-seq data (rep1, rep2) of primed UCLA1 and the naïve clones 4 (early passage), 9 and 12 for genes significantly down-regulated in epiblast cells of the human blastocyst compared to early passage primed hESC lines defined based on published single cell RNA-seq data (Yan et al., 2013), visualized with violin plots (see Table S2 for a list of these genes). KS test p-values of pair-wise comparisons are given below and color-coded based on significance.

(G) As in (F), but for genes significantly up-regulated in epiblast cells of the human blastocyst compared to early passage primed hESC lines.

(H) Empirical cumulative distribution functions of X-linked gene expression for primed UCLA1 and naïve UCLA1 clones 4 (early passage, *XIST*-negative), 9 and 12, from replicate RNA-seq data sets (rep1, rep2). The inset shows autosomal expression data for the same samples at the same scale. X-linked but not autosomal expression distributions of both replicates of early passage clone4 were the only samples that were significantly different from the distributions of any other sample (*=$p<0.006$ by Wilcoxon rank sum test with continuity correction). X-linked gene expression from primed UCLA1 cells did not differ from that of *XIST*-positive naïve clones 9 and 12 ($p>0.39$).

(I) Same as in (H), but for replicates of primed UCLA1 and re-primed cells of naïve UCLA1 after 30 passages in the naïve condition and 7 passages in the re-primed state. X-linked but not autosomal gene expressions of re-primed cells were significantly different from both replicates of primed cells ($p<0.003$ by Wilcoxon rank sum test with continuity correction).

Figure S3-4. Originally described naïve WIBR3 hESCs have two active X chromosomes with predominantly mono-allelic *XIST* expression

(A) Representative RNA FISH images detecting *XIST, HUWE1* and *UTX* in primed WIBR3 (P24) and WIBR3 in two modified naïve culture conditions (5iLA - without FGF2 at P15, and 4iLA - without IM12 and FGF2 at P12; Theunissen et al., 2014; Theunissen et al., 2016). A single nucleus with the most prevalent pattern is highlighted with a dotted box and enlarged for ease of viewing, and also depicted by the cartoon on the right.

(B) Representative RNA FISH images detecting *XIST*, *THOC2* and *UTX* in WIBR3 hESCs as described in (A).

(C) Representative RNA FISH images detecting *XIST*, *ATRX* and *UTX* in WIBR3 hESCs as described in (A).

Figure S3-5. Non-random XCI in differentiating naïve hESCs and hiPSCs

(A) Representative electropherograms from Sanger sequencing of a SNP-containing region in the X- linked gene *HDHD1*, known to escape XCI, on cDNA obtained from the primed UCLA1

population at P19 (top), and from an individual cell differentiated from the re-primed state after transition through the naïve state (bottom). The pie chart summarizes the Sanger sequencing results for this SNP in 24 *XIST*-positive, single cells differentiated from the re-primed state, considering three categories: when only the reference (Ref) allele, only the alternate (Alt) allele, or both alleles are expressed.

(B) Allelic expression proportions of X-linked genes known to escape XCI (*PLS3*, *CD99*) based on RNA- seq reads (number of reads indicated on top) from single *XIST*-expressing cells differentiated from naïve UCLA1 after transition through the re-primed state. The alphanumeric labels along the X-axis refer to single cell coordinates in a 96-well plate. For comparison, allelic proportions in primed UCLA1 were calculated from two replicates of population RNA-seq data of primed UCLA1 (Pri1/Pri2). **(C)** Representative immunofluorescence images of MeCP2 in combination with RNA FISH for *XIST* in primed hiPSCs derived from Rett syndrome patient fibroblasts, bearing the wild-type (clone16) or mutant (clone17) *MECP2* allele on the X_a, and upon conversion to the naïve state (P3 in 5iLAF media). The appearance of the MeCP2 protein in almost all cells of clone17 at P3 in naïve culture media captured the reactivation of the X_i, which occurred without induction of *XIST* expression consistent with sequential order of X_i-reactivation and *XIST* induction from the Xa.

(D) Representative immunofluorescence images of cells differentiated from primed hiPSC clones 16 and 17 described in (C) detecting MeCP2 and the negative pluripotency marker CD44 (Quintanilla et al., 2014).

(E) Representative RNA FISH images detecting *XIST*, *HUWE1* and *UTX* in re-primed cells (P3) obtained from naïve hiPSC clones 16 and 17 described in (C), and upon differentiation of these re-

primed cells. Both re-primed clones 16 and 17 displayed biallelic *HUWE1* expression, indicating the XaXa state of re-primed cells. Differentiation was accompanied by *XIST*-mediated XCI demonstrated by mono-allelic *HUWE1* focus not overlapping with the *XIST* cloud.

Figure S3-6. Characterization of the XCI status in primed UCLA1 and UCLA9 and upon differentiation

(A) Representative RNA FISH images detecting *XIST*, *HUWE1* and *UTX* in primed UCLA1 (P16) and after seven days of differentiation from the primed state. A single cell is enlarged in each row for ease of viewing of the prevalent RNA FISH pattern (dotted box), also depicted by the cartoon on the right. The primed image is the same one as shown in Figure 1A and is included here to make data comparison with that of differentiated cells easy.

(B) Representative phase-contrast images of UCLA9 in the primed state (P15) and after seven days of differentiation from the primed state.

(C) Representative RNA FISH images detecting *XIST*, *ATRX* and *UTX* in primed UCLA9 and after seven days of differentiation from the primed state. The absence of *XIST* expression and lack of silencing upon induction of differentiation of primed UCLA9 is described and extensively discussed by Patel et al. (2016).

(D) As in (C), but for *XIST*, *THOC2* and *UTX*.

(E) Representative RNA FISH images detecting *XIST* and *ATRX* in differentiated cells derived from naïve

then re-primed UCLA9. The single cell shown depicts the prevalent RNA FISH pattern in differentiated cells, which is also depicted by the cartoon on the right. Contrary to the differentiated cells originating from

the original primed UCLA9 hESC line, these cells, which are the progeny of naïve then re-primed UCLA9, demonstrated XCI with *XIST* expression.

(F) As in (E), but for *XIST* and *THOC2*.

Figure S3-7. UCLA4 maintains an eroded Xi upon differentiation from the primed state and is karyotypically normal in early passage naïve state

(A) Representative phase-contrast images of UCLA4 in the primed state (P14) and upon 7 days of differentiation from the primed state.

(B) Representative RNA FISH images detecting *XIST* RNA and the nascent transcription foci of *HUWE1* along with *UTX* in primed UCLA4 and their day seven differentiated product. A single nucleus representing the predominant X chromosome pattern is highlighted with a dotted box and enlarged for ease of viewing, and its X-pattern is depicted by the cartoon on the right.

(C) Representative RNA FISH images detecting *XIST*, *ATRX* and *UTX* as described in (B).

(D) A representative metaphase chromosome spread and summary of the cytogenetic analysis of naïve UCLA4 at P9.

Figure S3-1. Characterization of naïve hESCs and the primed to naïve transition

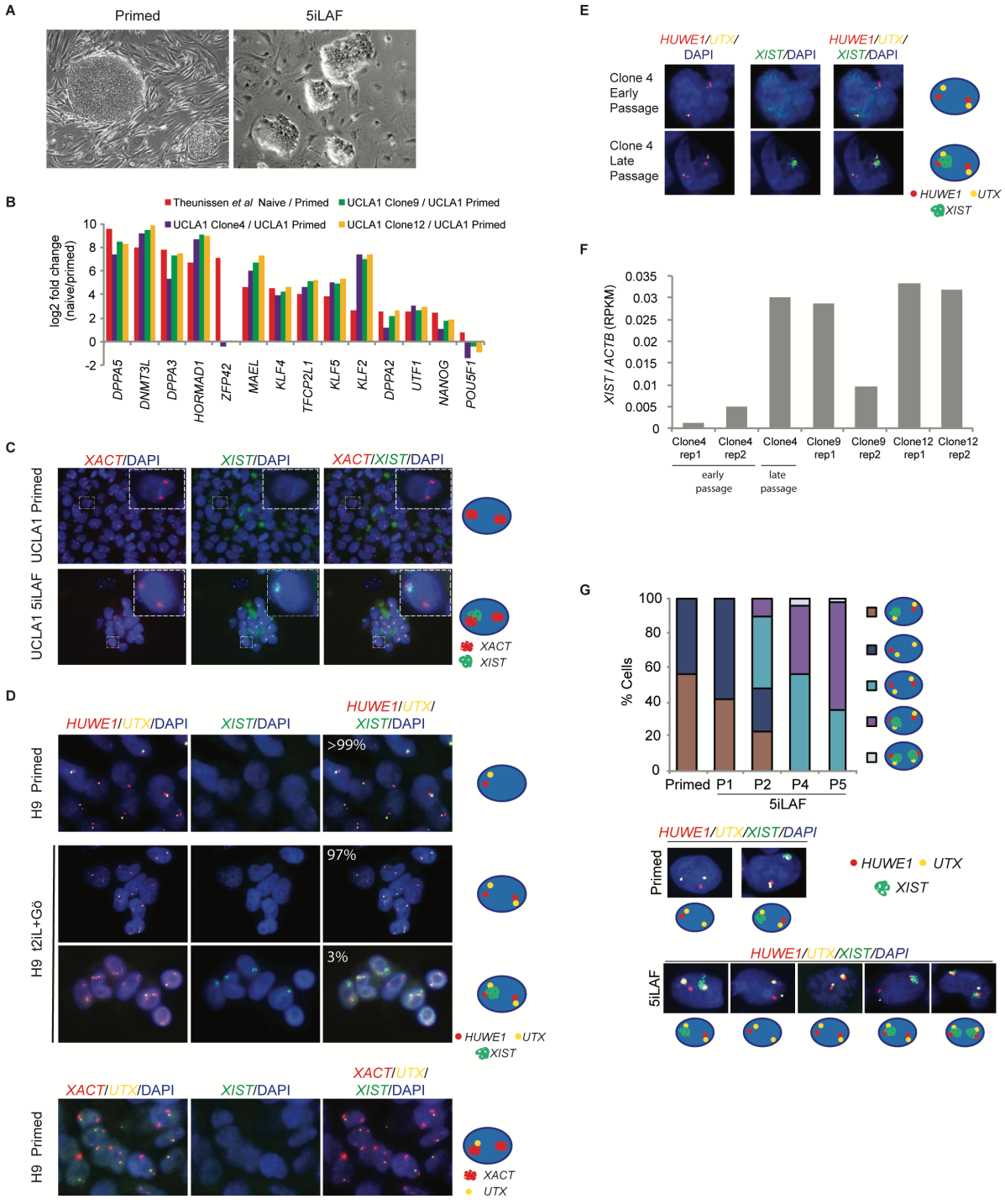


Figure S3-2. The transition from primed to naïve pluripotency leads to Xi-reactivation

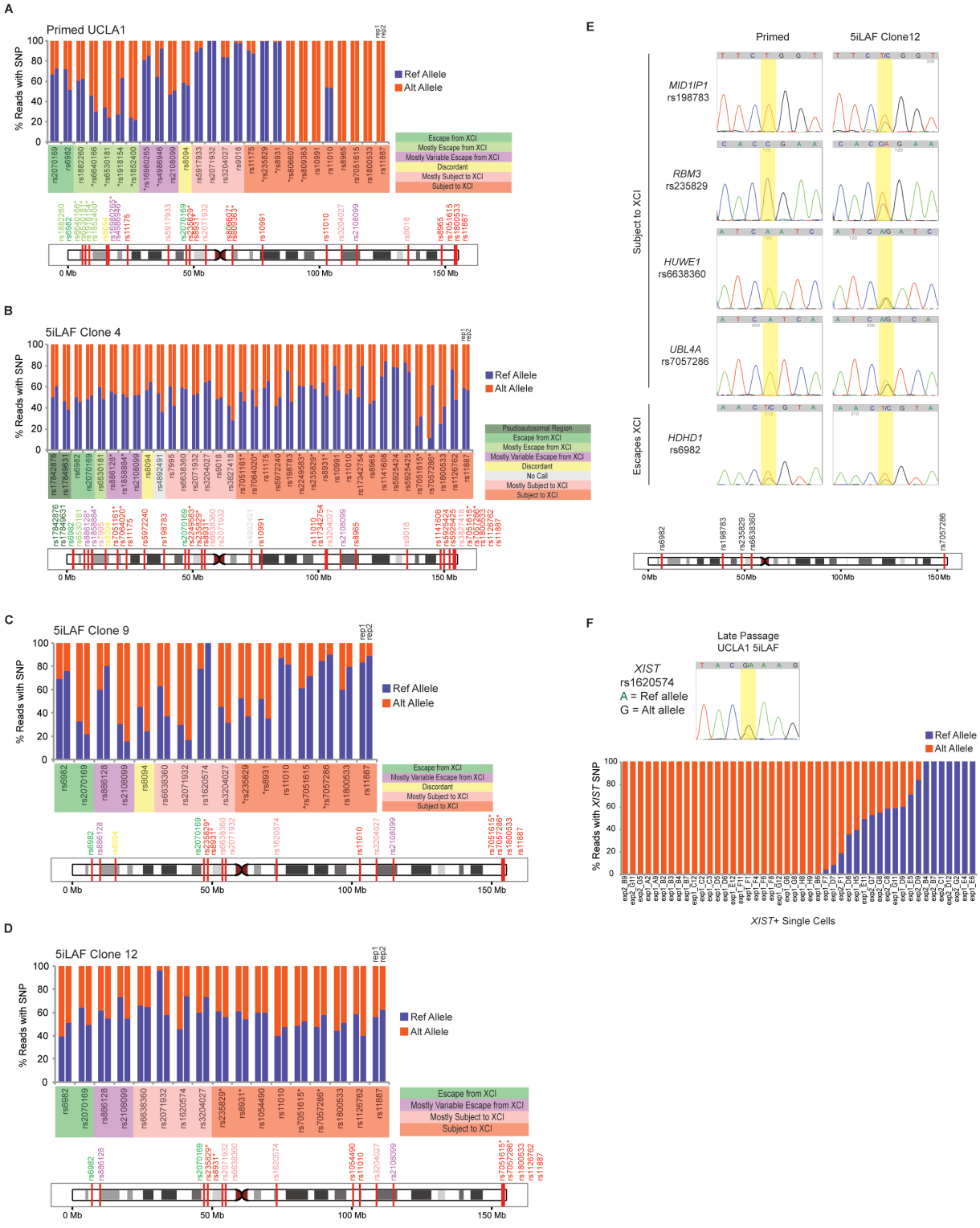


Figure S3-3. Naïve *XIST*-positive hESCs resemble the human blastocyst more closely than the intermediate *XIST*-negative cells

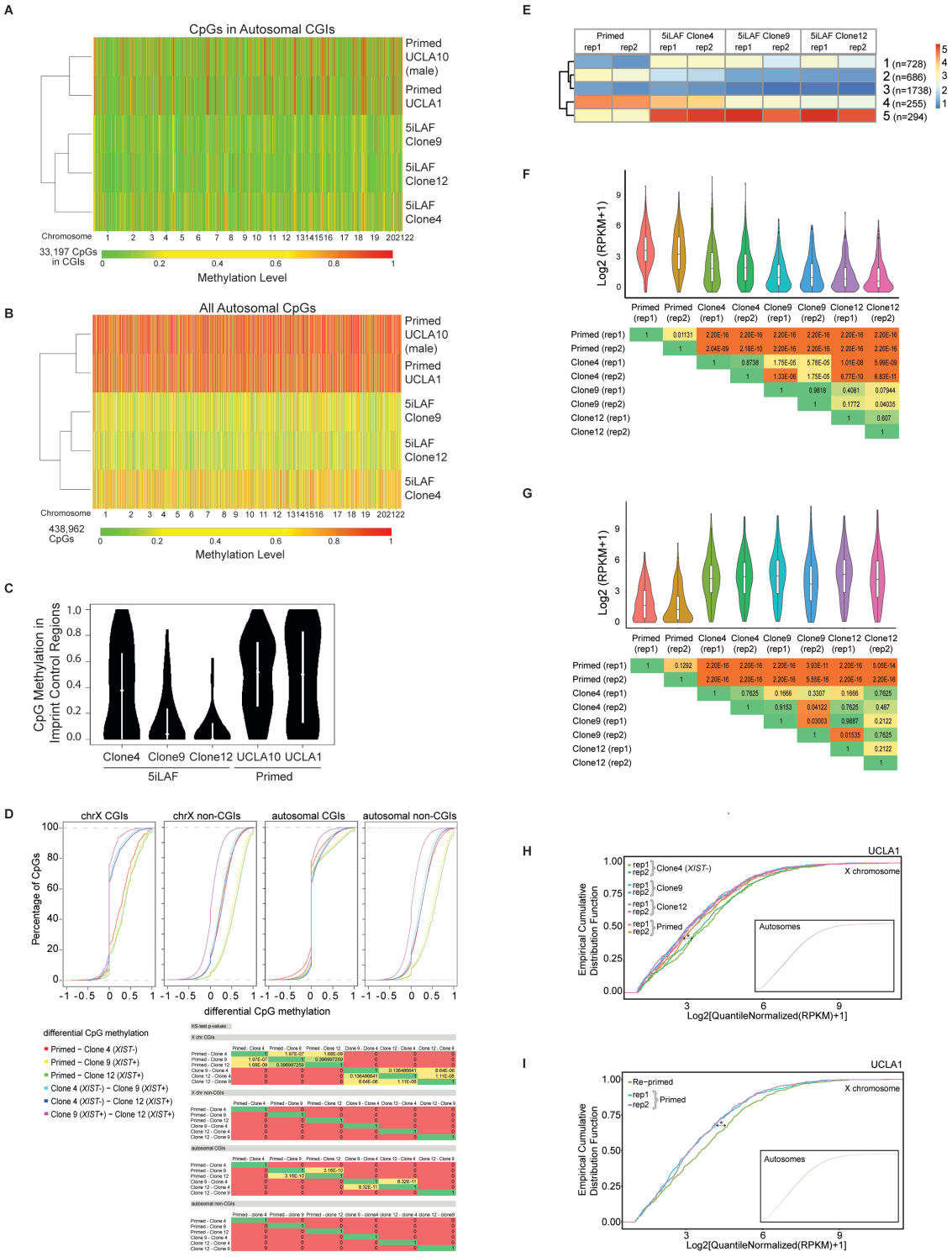


Figure S3-4. Originally described naïve WIBR3 hESCs have two active X chromosomes with predominantly mono-allelic *XIST* expression

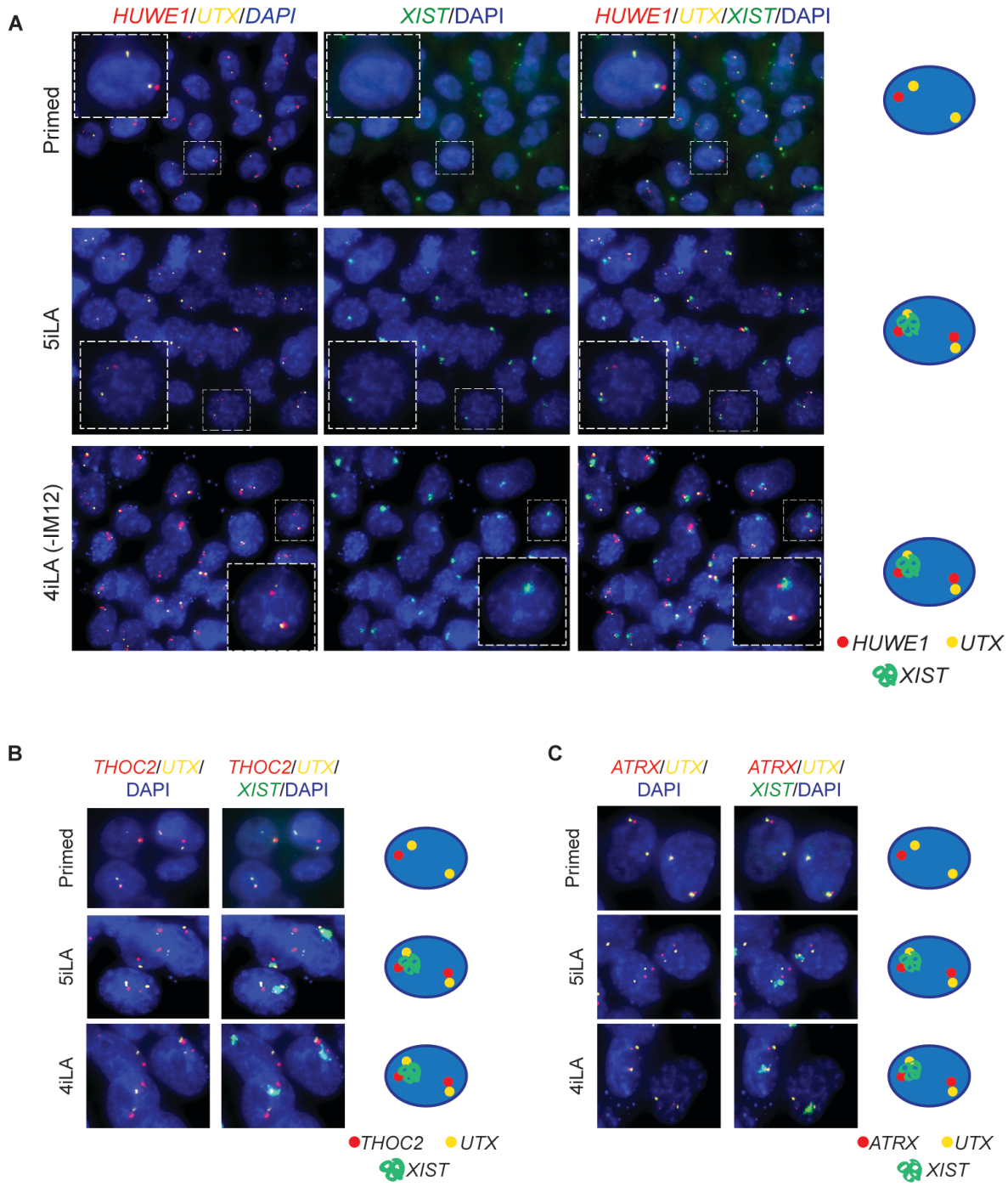


Figure S3-5. Non-random XCI in differentiating naïve hESCs and hiPSCs

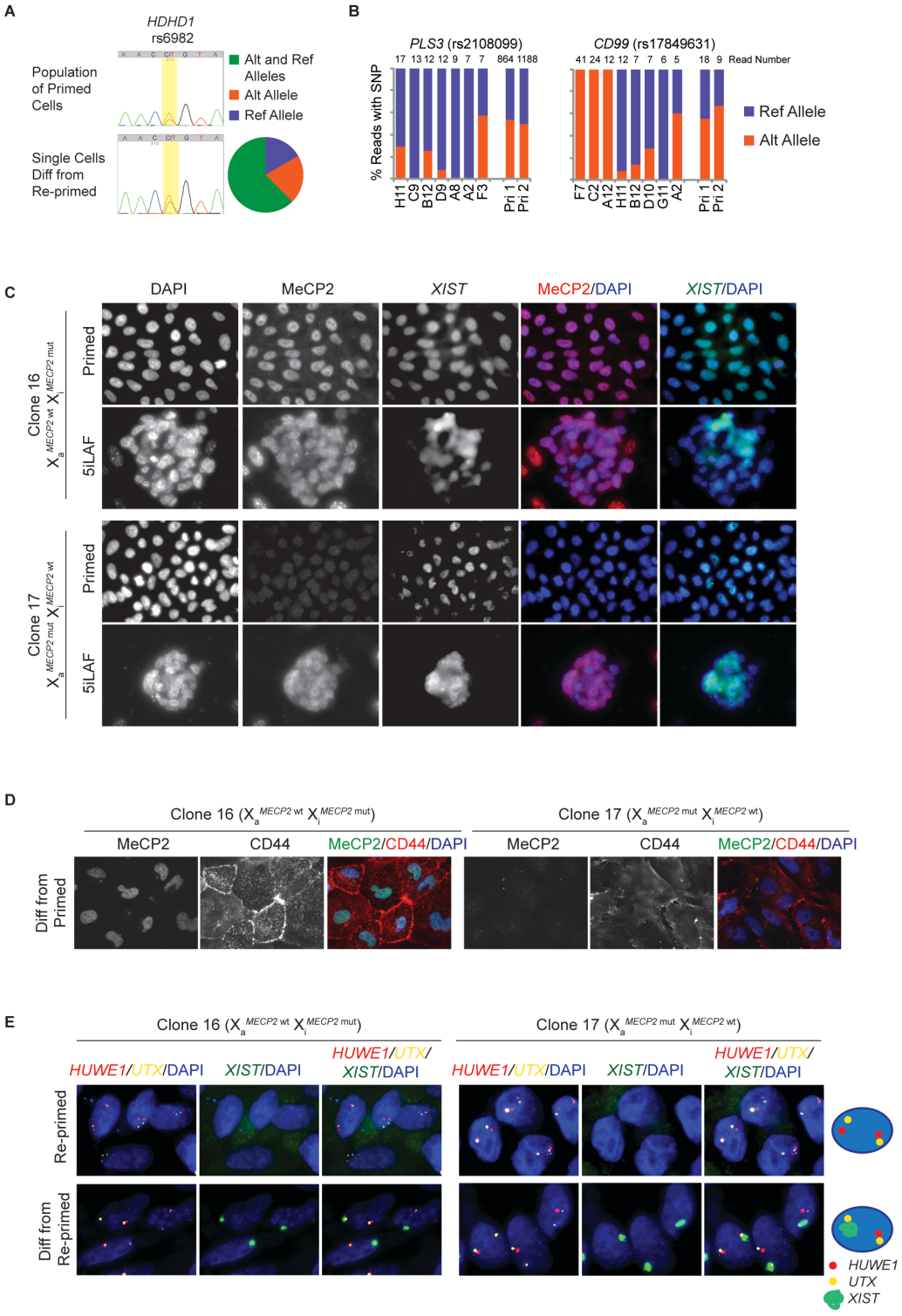


Figure S3-6. Characterization of the XCI status in primed UCLA1 and UCLA9 and upon differentiation

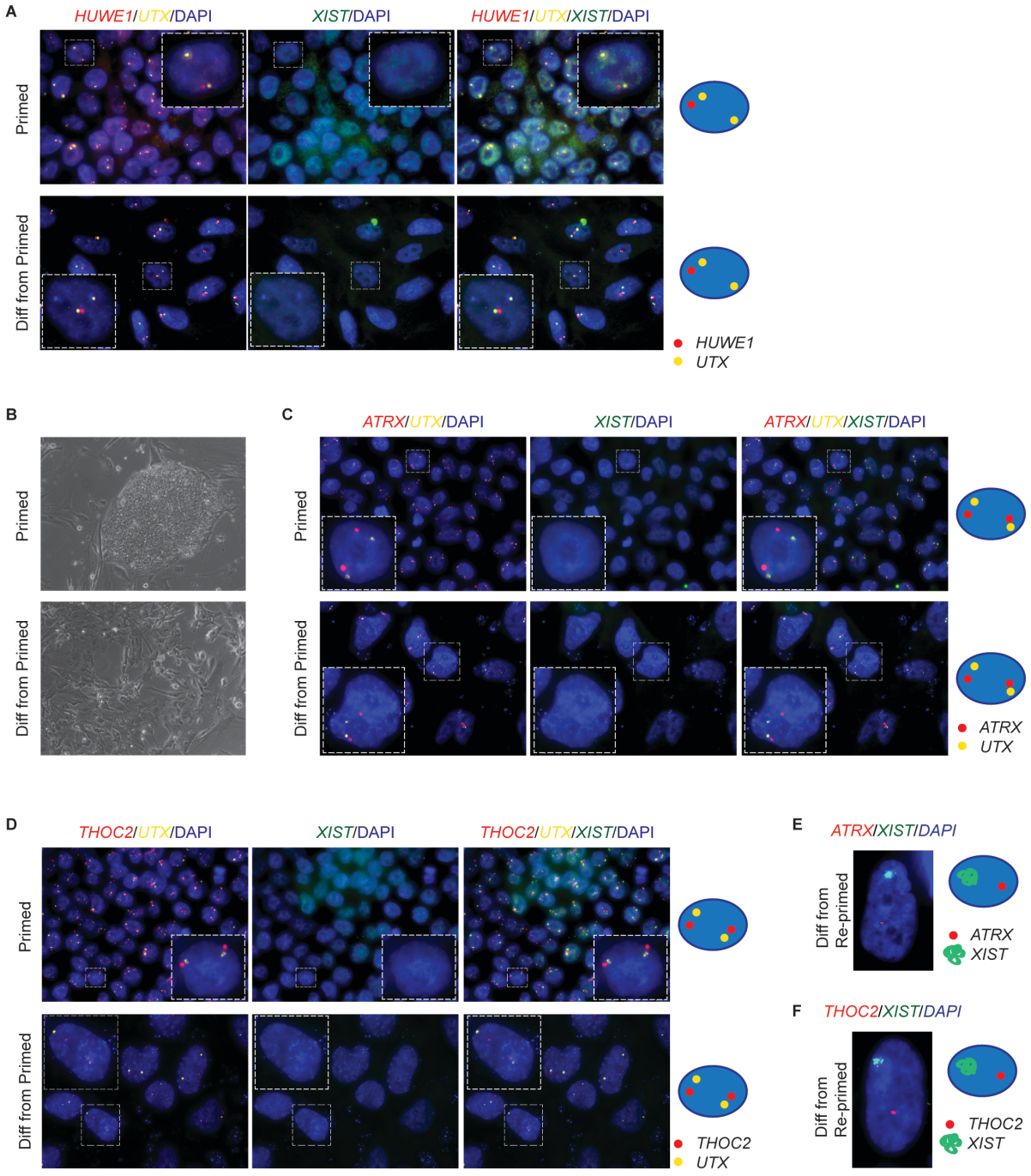
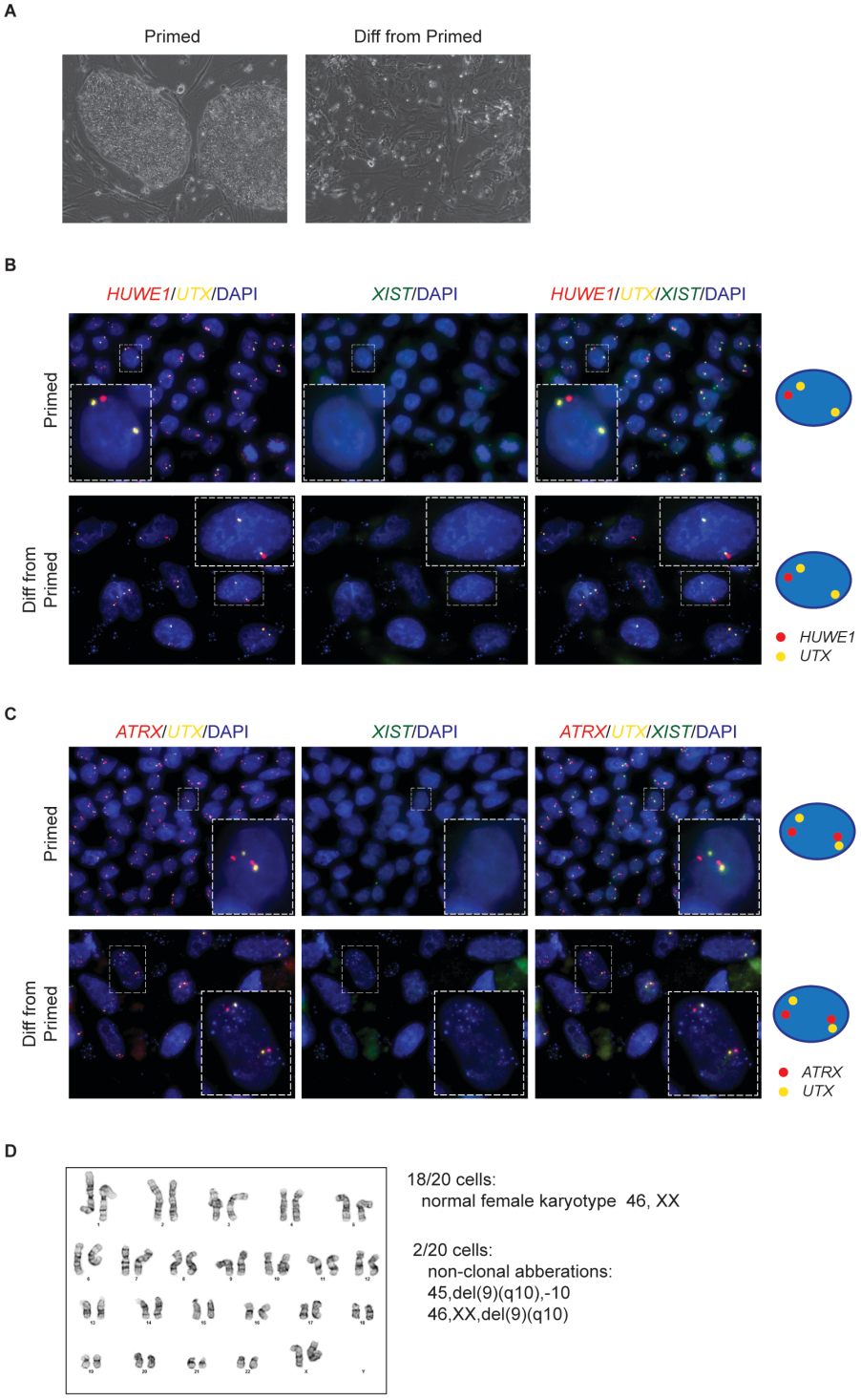


Figure S3-7. UCLA4 maintains an eroded Xi upon differentiation from the primed state and is karyotypically normal in early passage naïve state



SUPPLEMENTAL INFORMATION: EXPERIMENTAL PROCEDURES

Cell culture

Primed hPSCs were obtained from the Human Embryonic and Induced Pluripotent Stem Cell Core at UCLA and maintained in primed media consisting of 20% KnockOut Serum Replacement (KSR; Life Technologies) in DMEM/F12 (Life Technologies) supplemented with 1x penicillin/streptomycin (Life Technologies), 1x nonessential amino acids (Life Technologies), 0.5x GlutaMAX (Life Technologies), 0.1mM β-mercaptoethanol (Sigma), and 10ng/ml FGF2 (Peprotech). Cells were passaged every 6-7 days by detaching colonies with 1mg/ml collagenase IV (Life Technologies) at 37°C for 5−15 minutes (min), followed by manually breaking up colonies via pipetting. Conversion to the naïve state with the 5iLAF culture protocol was done as described previously (Theunissen et al., 2014). Briefly, two days post passage of primed hPSCs, media was changed to modified primed media which is made exactly as primed media, but instead of 20% KSR contained 5% KSR and 15% FBS (Omega Scientific). On day 7 post passage, cells were dissociated into a single cell suspension with 0.05% trypsin-EDTA at 37°C for 3 min and passed through a 40μm strainer. $2x10^5$ single cells were plated in one well of a 6-well plate in modified primed media in the presence of 10μM ROCK inhibitor Y-27632 (Stem Cell Technologies). Two days post passage, 5iLAF media was applied, which consisted of a 1:1 mixture of DMEM/F12 and Neurobasal (Life Technologies), supplemented with 1x N2 (Life Technologies), 1x B27 (Life Technologies), 1x penicillin/streptomycin, 1x nonessential amino acids, 0.5x GlutaMAX, 0.5% KSR, 0.1mM β- mercaptoethanol, 50μg/ml bovine serum albumin (Sigma), 20ng/ml rhLIF (EMD Millipore), 20ng/ml Activin A (Peprotech), 8ng/ml FGF2, 1μM MEK inhibitor PD0325901 (Stemgent or Bio-Techne), 0.5μM B-Raf inhibitor SB590885 (Bio-Techne), 1μM GSK3β inhibitor IM-12 (Enzo), 1μM Src inhibitor WH-4- 023 (A Chemtek), and

10μM ROCK inhibitor Y-27632. At about 11 to 12 days post plating, cells were dissociated by a 3 min treatment with StemPro Accutase (Life Technologies) at 37°C and re-plated after passing through a 40μm cell strainer in 5iLAF medium. Naïve hESCs were passaged as single cells every 5−6 days.

For sub-cloning of the 5iLAF UCLA1 hESC line at early passage, individual dome-shaped colonies that arose after single-cell plating were manually picked and expanded (into clones 4, 9 and 12). Conversion in 4iLA and 5iLA media, lacking both FGF2 and IM-12 or only FGF2, respectively, was done following the steps described for the 5iLAF protocol (Theunissen et al., 2016). In general, primed and naïve hPSCs were grown on irradiated CF-1 or DR4 mouse embryonic fibroblast feeder cells and maintained in a humidified 37°C incubator at 5% CO_2 and atmospheric oxygen levels. Conversion from primed UCLA1 to 5iLAF naïve pluripotency was also done at 5% O_2, but there was no change in cell growth rate, colony morphology, or X chromosome state when compared to cells at atmospheric oxygen.

For the transition of naïve cells to the primed state (re-priming), 5iLAF media was switched to primed media 3−4 days after splitting the naïve cells. Two−three days later, we observed significant cell death and naïve colonies became flat, visually resembling a primed hPSC culture. Re-primed cells were passaged with collagenase IV as described for primed hPSCs.

For differentiation, primed and re-primed hPSC lines were detached briefly with collagenase IV, transferred into fibroblast media and collected at the bottom of a conical tube by gravity to remove feeder cells. The colonies collected at the bottom were washed with Dulbecco's phosphate-buffered saline (DPBS), incubated with StemPro Accutase at 37°C for 2 min, passed through a 40μm strainer and plated at a final 1:2 split ratio in fibroblast media supplemented with

10μM ROCK inhibitor Y-27632 on plates or glass coverslips coated with matrigel (Fisher Scientific). Subsequently, the media was changed after two days and then every other day, which excluded the ROCK inhibitor. Differentiation was carried out for seven days.

The primed hESC lines WIBR3, UCLA1, UCLA4 and UCLA9 were described and characterized previously (Lengner et al., 2010;Diaz Perez et al., 2012; Patel et al., 2016). Similarly, naïve WIBR3 and UCLA1 lines, the 5iLAF naïve hESC lines UCLAn19 and UCLAn20 (directly derived from blastocysts), and the naïve t2iL+Gö adapted H9 cells were reported previously (Theunissen et al., 2014; Pastor et al., 2016; Takashima et al., 2014). t2iL+Gö-adapted naïve H9 cells and their primed counterparts were grown at 5% O_2 and 7% CO_2 conditions. The low passage *XIST*-expressing X_aX_i primed iPSC line #1001 used for the conversion to the naïve state in 5iLAF was generated and analyzed previously (Karumbayaram et al., 2012). Conversion of primed hESC line UCLA1 to 5iLAF naïve state was repeated four times and each time similar timing of epigenetic events and morphological changes were observed. Conversions of the primed hESC lines UCLA4 and UCLA9 were repeated three times each, again with similar timing of events and morphological changes. Naïve UCLA4 hESCs were sent to Cell Line Genetics (Madison, WI) for G-banding karyotype analysis at passage 9 and karyotypes for naïve UCLA1 and the blastocyst-derived naïve hESC lines UCLA19n and UCLA20n have been reported (Pastor et al., 2016) Re-priming and differentiation of all naïve hPSC lines was repeated at least twice.

Rett syndrome hiPSCs were generated from GM07982 female fibroblasts (NIGMS Human Genetic Cell Repository) heterozygous for the 705delG frame-shift mutation in *MECP2*, which leads to a premature stop codon in the mRNA and therefore absence of a full length MeCP2 protein product (Lee et al., 2001). Before reprogramming, the presence of the mutation was verified by PCR of genomic DNA. For reprogramming to hiPSCs, 100,000 fibroblasts were plated in one well

of a 6-well plate and one day later infected overnight with 20ul of concentrated (about $5x10^8$ TU/ml) STEMCCA lentivirus, a single polycistronic lentiviral vector encoding Oct4, Klf4, Sox2 and c-Myc under control of the constitutive EF1a promoter (Sommer et al., 2009) in 1ml fibroblast media (10% FBS in DMEM supplemented with 1x penicillin/streptomycin, 1x nonessential amino acids, 1x GlutaMAX and 0.1mM β-mercaptoethanol) with 5μg/ml polybrene. On day 5 post-infection, cells were trypsinized and re-plated on feeders. The next day media was changed to primed hESC media and replaced daily. Human ESC-like colonies were picked between weeks 2−3 post infection and enzymatically passaged with collagenase IV. Several clones were analyzed for their XCI state as described (Tchieu et al., 2010) to demonstrate their $XaXi^{XIST+}$ state at early passage and the transition to the *XIST*-negative X_aX_i state over subsequent passages. The location of the mutant or wild-type *MECP2* allele, respectively, on the X_a was determined by Sanger sequencing of RT-PCR products, which allowed the classification of clone16 as $Xa^{MECP2\text{wt}}$ $Xi^{MECP2\text{mut}}$ and clone17 as $Xa^{MECP2\text{mut}}$ $Xi^{MECP2\text{wt}}$. In addition, the expression of endogenous pluripotency genes and the silencing of the ectopic reprogramming cassette were confirmed by RT-PCR, cytogenetic analysis was performed by Cell Line Genetics to demonstrate a karyotypic normal state, and teratoma formation assays were conducted by testis injection following standard procedures to demonstrate ability to differentiate into all three germ layers, as described (Tchieu et al., 2010).

RNA fluorescent in situ hybridization (FISH)

Cells were grown on gelatinized 18mm circular glass coverslips (Fisher Scientific, 12-545-100), washed with DPBS, fixed with 4% formaldehyde for 10 min, permeabilized with cold (4°C) 0.5%

Triton X-100 in DPBS for 10 min, and serially dehydrated with cold (4°C) 70-100% ethanol. Coverslips were air dried and hybridized with labeled DNA probes in a chamber humidified with 50% formamide in 2x SSC at 37°C for 24−36 hours, washed for three 5-min intervals with 50% formamide in 2x SSC, 2x SSC, then 1x SSC at 37°C, then mounted with ProLong Gold Antifade reagent containing DAPI (ThermoFisher). Double- stranded DNA probes were generated from BACs with the BioPrime Array CGH Genomic Labeling System and fluorescently labeled ChromaTide nucleotides (ThermoFisher). The BACs used include *XIST* (RP11-13M9), *XACT* (RP11-35D3), *HUWE1* (RP11-975N19), *UTX* (RP11-256P2)*, ATRX* (RP11-1145J4) and *THOC2* (RP11-121P4). The labeled DNA pellet, along with salmon sperm DNA and human Cot1 DNA (ThermoFisher), was stored at -80°C after re-suspension in deionized formamide (VWR) and 2x hybridization buffer (0.2g/ml dextran sulfate, average M_W >500,000 (Sigma) in 4x SSC and 0.1M NaH_2PO_4). Every new batch of probes was tested on normal human dermal fibroblasts before use in experiments. Every 5iLAF naïve hESC line generated was subjected to RNA FISH analysis of its X chromosome state three or more times at different passages and the data obtained were consistent for each time point.

Cells were grown on coverslips and fixed as for RNA FISH, but after the 10 min incubation in 0.5% Triton X-100 in DPBS, coverslips were incubated in 0.2% TWEEN-20 in DPBS for 10 min. Coverslips were then incubated with blocking buffer (5% donkey serum (Fisher Scientific), 0.2% gelatin from cold water fish skin (Sigma), 0.2% TWEEN 20 for 30 min in a humidified chamber. This was followed by incubation with primary antibody, diluted in blocking buffer, overnight at 4°C (anti-MeCP2: Diagenode, C15410052, used at 50ng/ml; anti-CD44: also called Hermes-1, deposited to the Developmental Studies Hybridoma Bank by E.C. Butcher, used at 1:10;

Quintanilla et al., 2014). Coverslips were then washed three times (5 min each) with 0.2% TWEEN-20 containing DPBS and incubated with secondary antibody diluted in blocking buffer for 1hr at room temperature. Coverslips were washed again as after the primary antibody incubation and mounted with ProLong Gold Antifade reagent containing DAPI. For MeCP2, H3K27me3 or RNA-PolII co- detection with *XIST* RNA, immunofluorescence (IF) was followed by RNA FISH. IF was done as described above, but the primary antibody incubation time was reduced to 1hr at room temperature and RNase-out (ThermoFisher) was added to the blocking buffer at 1:200. Anti-H3K27me3 antibody was used at 1:400 (Active Motif, 39155) and the anti-RNA PolII antibody at 1:1000 (Millipore 05-623, clone CTD4H8). After secondary antibody washes, the RNA FISH protocol was carried out as described above, starting from the ethanol dehydration steps.

Microscopy and image analysis

Images were taken as Z-stacks with the Imager M1 microscope (Zeiss) at 630x magnification using the Axio Vision software. All image processing was done with the ImageJ software (NIH). Z-stack images were merged using maximum intensity on gray scale images, and the color merge function was used for overlaying merged Z-stacks of different channels. For FISH images, brightness and contrast was adjusted for each channel after merging of Z-stacks but before overlaying, to remove background signal. RNA FISH signals were quantified by eye either from images or through the eyepiece of the microscope. For IF and IF/RNA-FISH, the exposure time was kept constant across all samples during image acquisition in all channels except for DAPI and the images were processed without any changes to brightness or contrast. A minimum of 100 cells from at least

five different colonies (in the case of hPSCs) or fields (in the case of differentiated cells) was analyzed.

RNA-sequencing

For strand-specific RNA sequencing of the primed hESC line UCLA1, the naïve UCLA1 clones 4 (at early (*XIST*-negative, P6) and late (*XIST*-positive, P16) passage), clones 9 and 12 (P8), re-primed UCLA1, and naïve UCLA4 at early (P5) and late passage, cells were harvested, washed with DPBS, and collected in Trizol (ThermoFisher). Where indicated, two replicates were obtained by harvesting cells from two different wells of a 6-well plate and independent processing through all downstream steps. Only for late passage naïve UCLA4 (*XIST*-positive), the replicates represent two independent conversions from the primed to the naïve state at P22 and P17, respectively. RNA was isolated after chloroform extraction using the RNease Mini Kit (Qiagen). Before elution of RNA, the column was subjected to DNase digestion (Qiagen) for 30 min at room temperature to remove any contaminating genomic DNA. RNA quality and amount were determined using a NanoDrop spectrophotometer ND-1000 (Thermo Scientific). Four μg of total RNA were then used for mRNA isolation and library preparation using the TruSeq Stranded mRNA Library Prep Kit (Illumina) following the manufacturer's Sample Preparation Guide. All libraries were amplified for 15 cycles and the final PCR product was run on a low-melt agarose gel and DNA with a median of 250bp was extracted from gel slices. The resulting libraries were sequenced as single-end 50bp reads at the UCLA Broad Stem Cell Center High-Throughput Sequencing facility. Reads were mapped using TopHat v2.0.13 and assigned to genes with HTSeq-0.6.1 (using the Illumina iGenomes UCSC hg19 assembly). Reads Per Kilobase per Million mapped reads (RPKM) values

were calculated (Table S2) and differential gene expression analysis for the replicates of primed UCLA1 and naïve UCLA1 clones 4, 9 and 12 was preformed using the DESeq2 (version 1.6.2) R package for each pairwise comparison among the two replicates of primed hESC line UCLA1 and its naïve clones 4 (early passage), 9 and 12. Genes differentially expressed (p-value < 0.01 and abs(Log_2 fold-change) > 2) among all comparisons (n=3,701) (Table S2 were concatenated and used for k-means clustering (k=5) after log_2 transformation of RPKM values. The heatmap (version 1.0.8) R package was used to construct the heatmap.

For comparing the expression pattern of primed UCLA1 cells and naïve clones 4, 9 and 12 to that of epiblast cells of the human pre-implantation blastocyst, we used the 1,008 genes differentially expressed between human epiblast cells of the blastocyst and newly derived primed hESCs defined by Yan et al., (2013) that satisfied a significance threshold (p-value < 0.01 and abs(Log_2 fold-change) > 2), and divided them into ordered sets of up- and down-regulated genes according to fold-change and directionality (Table S2). In these two gene sets, after Log_2 transformation of the RPKM expression values, the geom_violin function of ggplot2 (version 2.1.0) R package was used to generate violin plots, and the Kolmogorov- Smirnov (KS) test to identify significantly different distributions. Again, only UCLA1 samples with replicates were used in this analysis.

For globally comparing X-linked and autosomal gene expression across all samples, the complete gene set (n=26,364) was subset into autosomal (n=25,265) and X-linked (n=1,099) genes. The Empirical Cumulative Distribution Function (ECDF) (computed using the stat_ecdf() function of ggplot2 (version 2.1.0) R package; http://www.r-project.org (Dean and Nielsen, 2007) was used to plot the gene expression distribution for both subsets across all samples in each of

indicated comparison groups and the Wilcoxon rank sum test with continuity correction (performed with the wilcox.test() function of the stats (version 3.3.0 R package) was used to identify significantly different distributions.

For single cell RNA-seq, we used early and late passage UCLA1 5LAF naïve hESCs (not sub-cloned) to capture *XIST*-negative (enriched at early passage) and *XIST*-positive (predominant at late passage) naïve cells, respectively, as well as UCLA1 cells differentiated for 7 days from the re-primed state (after transition through the naïve state for over 20 passages). Cells were detached with 0.05% trypsin-EDTA (differentiated) or Accutase (naïve) at 37°C for 2–3 min, passed through a 40μm strainer, and counted using a hemacytometer. Cell concentration was adjusted to 200,000 cells/ml before loading on a Fluidigm C1 Single-Cell Auto Prep System (Fluidigm) following the manufacturer's guidelines. For differentiated cells we used the Single Cell Preamp IFC 17–25 μm, for naïve cells the Single Cell Preamp IFC 10–17 μm. cDNAs were made on-chip with the Clontech SMARTer Ultra Low RNA kit for Illumina using protocols provided by Fluidigm. Libraries were constructed in 96-well plates using the Illumina Nextera XT DNA Sample Preparation kit according to the standard protocol supplied by Fluidigm, and sequenced as paired-end 100 base pair reads at the UCLA Broad Stem Cell Center High-Throughput Sequencing facility. 1/10 of the produced cDNA was diluted 5-fold and used for experiments with direct Sanger sequencing. Microsoft Excel and BoxPlotR (http://boxplot.tyerslab.com) were used for generating figures describing the single cell data.

Determination of allelic expression of X-linked genes based on SNPs

To determine the allelic expression state of X-linked genes, heterozygous SNPs were defined

genome-wide using the Affymatrix SNP6.0 array at the UCLA Clinical Microarray Core, on genomic DNA of the primed hESC line UCLA1. Taking advantage of SNPs in exons of X-linked genes, allelic expression was assessed using either population or single cell RNA-seq data, or SANGER sequencing of RT-PCR products, as indicated. For population RNA-seq data, the proportion of reads covering the reference or alternative SNP was graphed for those SNPs common to primed and naïve sub-clones with 5 or more reads (Figure 2B), or for all SNPs in primed or each naïve sub-clone with 10 or more reads (Figures S2A–S2D). For differentiated UCLA1 single cell RNA-seq data, the proportion of reads covering the reference or alternative SNP was graphed for SNPs with 5 ore more reads in each cell expressing *XIST,* which was confirmed by PCR from single cell cDNA. Only SNPs with coverage in two or more individual cells were graphed (Figures 5A–5C, S5A and S5B). For naïve early- and late- passage UCLA1 single cell RNA-seq data, cells were classified as *XIST*-negative and *XIST*-positive based on *XIST* RNA counts (if the log2- depth normalized count value was < 2, a single cell was considered *XIST*-negative). As expected, early passage naïve UCLA1 had more *XIST*-negative single cells than late passage naïve UCLA1. For uncovering the allelic expression status of *XIST*, *XIST*-positive single cells with 10 or more reads spanning the *XIST* SNP rs1620574 were graphed in Figure S2F. For Figure 2C, the allelic expression status of X- linked genes normally subject to XCI covered by 10 or more reads at informative SNPs was determined. Each circle in the figure represents the result for a particular SNP in a single cell.

In case of Sanger sequencing, we used either cDNA from single differentiated cells from re-primed UCLA1 hESCs or from populations of primed UCLA1 at P19 and naïve clone12. Only for the *XIST* expression analysis in primed UCLA1, early passage cells were employed (P4), since *XIST* was silenced at P19. cDNA was synthesized using SuperScript III First-Strand Synthesis

SuperMix (ThermoFisher), and amplified with primers spanning the SNP of interest using KAPA HiFi PCR polymerase (KAPA Biosystems). The following forward and reverse primers were used: 5'CCATAGCTGACCAAGGCCAG3' and 5'CGGCAGCACCGAGATAAAAGG3' for *MID1IP1* (SNP ID rs198783), 5'TGGTTATGACCGCTACTCAGG3' and 5'CTGCCCCCACTTTTAATTTGC3' for *RBM3* (SNP ID rs235829), 5'CTACCCGTGAAGTCCTTGGC3' and 5'CGTTCCTCTGTACCAACAACC3' for *HUWE1* (SNP ID rs6638360), 5'CAGCAGGGTCCTGGAACAG3' and 5'CAGTGCTGGGGATGAGGAC3' for *UBL4A* (SNP ID rs7057286), 5'GTTTGCTACCTCACAACAACC3' and 5' GCAGACATATATTCAGGCCATC3' for *HDHD1* (SNP ID rs6982), 5'CATTGCTAGGCATTGGGGATG3' and 5'CCAGGAAGCATGTATCTTCTGG3' for *XIST* (SNP ID rs1620574). PCR products were run on an agarose gel, expected sized bands gel-eluted using the MinElute Gel Extraction Kit (Qiagen), and sent for Sanger Sequencing (Retrogen Inc., San Diego, CA) using the forward PCR primer. The 4Peaks software was used to visualize the electropherograms of sequencing results.

DNA methylation

Genomic DNA was harvested from primed hESCs and UCLA1 naïve clones using the DNAeasy blood and tissue kit (Qiagen). For each cell population two replicates were obtained by harvesting cells from two different wells of a 6-well plate and independent processing. Libraries for Reduced Representation Bisulfite Sequencing (RRBS) were created as previously described (Meissner, 2005), and size-selected between 50 and 500bp. DNA methylation analysis was performed using BS-Seeker2 (2.0.32) (Guo et al., 2013) using Bowtie (0.12.9) (Lee et al., 2001) for read alignment

to the human genome (hg19) on the UCLA Hoffman2 computer cluster. Reads with adapter contamination were trimmed. CpG island (CGI) coordinates were obtained from UCSC (http://genome.ucsc.edu). Only CpG sites covered by at least five reads across all samples under consideration were used in an effort to obtain reliable methylation levels. Since replicate samples showed good correlation (data not shown), we merged replicates by summating counts at each CpG site.

Methylation data were hierarchically clustered using complete linkage and the Euclidean distance metric. Statistical analysis, clustering, and heat map generation were performed using custom R scripts. For the presentation of X-linked CGIs in Figure 2, CpGs were additionally filtered and required to have less then 20% methylation in the male primed hESC line UCLA10 (Patel et al., 2016), to emphasize the methylation state due to XCI. For the analysis of methylation in imprint control regions, all CpGs were filtered for minimum 5-fold coverage from merged replicates of RRBS data, then selected for overlap with maternal and paternal non-placental imprinted regions (Okae et al., 2014). Methylation was visualized using the R-software package. For Figure S3D, the pairwise distributions of differential DNA methylation was plotted using CDFs for X-linked and autosomal CpGs within and outside of CGIs with at least 5x coverage and the Kolmogorov-Smirnov test was used to test the difference of pairwise comparisons between the distributions.

SUPPLEMENTAL INFORMATION: REFERENCES

Balaton, B.P., Cotton, A.M., and Brown, C.J. (2015). Derivation of consensus inactivation status for X- linked genes from genome-wide studies. Biol. Sex Differ. *6*, 35.

Dean, C.B., and Nielsen, J.D. (2007). Generalized linear mixed models: a review and some extensions. Lifetime Data Anal. *13*, 497–512.

Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., and Clark, A.T. (2012). Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. Hum. Mol. Genet. *21*, 751–764.

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.-Y., and Pellegrini, M. (2013). BS- Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics *14*, 774.

Karumbayaram, S., Lee, P., Azghadi, S.F., Cooper, A.R., Patterson, M., Kohn, D.B., Pyle, A., Clark, A., Byrne, J., Zack, J.A., et al. (2012). From Skin Biopsy to Neurons Through a Pluripotent Intermediate Under Good Manufacturing Practice Protocols. Stem Cells Transl. Med. *1*, 36–43.

Lee, S.S., Wan, M., and Francke, U. (2001). Spectrum of MECP2 mutations in Rett syndrome. Brain Dev. *23 Suppl 1*, S138–S143.

Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J., et al. (2010). Derivation of Pre-X Inactivation Human Embryonic Stem Cells under Physiological Oxygen Concentrations. Cell *141*, 872–883.

Meissner, A. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., et al. (2014). Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet. *10*, e1004868.

Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., and Clark, A.T. (2016). Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. Cell Stem Cell *18*, 323–329.

Patel, S., Bonora, G., Sahakyan, A., Kim, R., Chronis, C., Langerman, J., Fitz-Gibbon, S., Rubbi, L., Skelton, R.J.P., Ardehali, R., Pellegrini, M., Lowry, W.E., Clark, A.T., Plath, K. (2016). Human embryonic stem cells do not change their X-inactivation status during differentiation. *Cell Reports*, in press.

Quintanilla, R.H., Asprer, J.S.T., Vaz, C., Tanavde, V., and Lakshmipathy, U. (2014). CD44 Is a Negative Cell Surface Marker for Pluripotent Stem Cell Identification during Human Fibroblast Reprogramming. PLoS ONE *9*, e85419.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells Dayt. Ohio *27*, 543–549.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell *158*, 1254–1269.

Tchieu, J., Kuoy, E., Chin, M.H., Trinh, H., Patterson, M., Sherman, S.P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J.A., et al. (2010). Female Human iPSCs Retain an Inactive X Chromosome. Cell Stem Cell *7*, 329–342.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., et al. (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. Cell Stem Cell.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

CHAPTER 4

Regulation of X-chromosome Dosage Compensation in Human:

Mechanisms and Model Systems

The X-chromosome state of the human pre-implantation embryo

Somatic cells of an adult female human have two X chromosomes, but most genes on one of them are silenced at the level of transcription, so that the X-chromosome gene dosage in female XX cells is equal to that of male XY cells. The silenced X chromosome can be either the paternally or the maternally inherited one, making the adult female a natural mosaic. This random pattern of X-chromosome inactivation (XCI) is established in early embryogenesis. The X chromosomes inherited from the egg (maternal) and the sperm (paternal) are both active in very early female development [1,2] before each cell commits to transcriptionally silencing one X chromosome for the rest of the cell's and its progeny's life. It is not known exactly when this choice is made in human development, but based on mouse studies it is hypothesized to happen shortly after the embryo implants [3]. Surplus pre-implantation embryos from *in vitro* fertilization clinics donated to research have made *ex vivo* studies of human pre-implantation development possible. Combined with advances in single-cell transcriptome profiling, these have recently enabled a closer look at the X-chromosome biology in early human development [1,2,4–6].

Petropoulos and colleagues studied the transcriptome of the largest number of human pre-implantation embryos reported to date, and performed sex-specific analysis of human development at days 3–7 post fertilization (E3–E7) at the single-cell level [2]. Their analysis revealed that immediately after zygotic gene activation (ZGA) at E4, female embryos had almost double expression of X-linked genes compared with males, consistent with females having two active X chromosomes (4-1). However, with increasing developmental time from E4 to E7, this roughly 2 : 1 female : male ratio decreased, reaching nearly 1 : 1 in all cells of the embryo at E7 (4-1), just in time for the commencement of implantation. Surprisingly, this drop in X-linked gene expression level was not due to the onset of X-chromosome-inactivation, because allelic expression analysis

by single-cell RNA-sequencing revealed that both X chromosomes were active at all times [2]. Evidence for the presence of two active X chromosomes in female human pre-implantation embryos was extended further by RNA fluorescent *in situ* hybridization (RNA-FISH) [1,2,5,6]. Thus, Petropoulos *et al*. uncovered a novel mechanism of X-chromosome dosage compensation, at the mRNA level, in human pre-implantation development where female to male expression is equalized not by inactivating one of the two X chromosomes in the female, but rather by dampening the expression of both female X chromosomes (4-1). This X-chromosome dampening (XCD), which has not been observed in mice, is reminiscent of the dosage compensation system occurring in a model organism further removed from the human on the evolutionary scale—the roundworm *Caenorhabditis elegans.* Both X chromosomes of XX hermaphrodite *C. elegans* undergo condensin-mediated three-dimensional structural remodelling, resulting in reduced transcriptional output to match X-linked gene dosage to that of the single X in XO males [7,8]. However, whether XCD in human and *C. elegans* are mechanistically similar remains an open question. In any case, together these findings indicate that X-chromosome dosage compensation in human is regulated by two different and sequential processes: first XCD and later XCI. Interestingly, moderate but significant expression asymmetry between the two X chromosomes was detected from E5, suggesting that X-linked gene silencing may initiate in a progressive manner at this developmental stage [5].

XIST expression correlates with X-chromosome dampening

A hallmark of the inactive X chromosome (Xi) is expression and accumulation of the *cis*-acting long non-coding RNA (lncRNA) *XIST* (X inactive specific transcript) [9–11], which, as its

name suggests, was thought until recently to always correlate with the inactive status of the X chromosome. However, an unexpected finding was made in 2011, when Edith Heard's group used RNA-FISH to demonstrate that both male and female human pre-implantation embryos express the lncRNA *XIST* without any evidence of X-inactivation (4-1) [1]. This was the first report of long-term expression (over several days) and accumulation of *XIST* RNA that does not lead to chromosome-wide silencing, and was indeed very intriguing. This finding inspired further studies of the X-chromosome state in the human pre-implantation embryo, which validated the presence of *XIST*-expressing active X chromosomes [2,4–6]. While *XIST* was expressed from both X chromosomes in the majority of cells in female blastocysts, a proportion of the cells, however, displayed mono-allelic *XIST* expression pattern [1,2,5]. In RNA-FISH studies, *XIST* was also found accumulating on the single X in male embryos, although contrasting results were obtained between studies in the proportion of *XIST*-expressing cells—from a majority of male cells in the blastocyst expressing *XIST* [1,5] to most cells being devoid of *XIST* expression [2]. This discrepancy is perhaps due to differences in the sensitivity of the RNA-FISH assays employed, and might be related to the fact that *XIST* was found at much lower amounts in male cells compared with female cells in RNA-sequencing experiments [2].

Human *XIST* expression initiates as early as at the 4–8-cell stage of the embryo and coincides with the onset of ZGA [2,4,5]. *XIST* levels increase over time up to E7, in a manner that correlates with X-linked dampening (4-1). This correlation is also observed in naive human embryonic stem cells (hESCs), where cells with two active chromosomes and no *XIST* expression have overall higher X-linked gene expression compared with cells with two active Xs and *XIST* expression [12]. Whether XCD in human is mediated by *XIST* remains an open question, but in the worm other mechanisms are involved as *XIST* is not conserved beyond placental mammals

[13]. Should *XIST* mediate XCD in the female human pre-implantation embryo, one would have to assume that the lower level of *XIST* in male embryos is not sufficient for the induction of XCD on the male single X chromosome.

Differences between mouse and human XCI

In contrast to the human, mouse embryos are more easily attainable in larger numbers; hence our understanding of mouse pre- and post-implantation development, including the regulation of X-chromosome dosage, is more advanced. It is well established that female mice undergo X-chromosome dosage compensation via XCI in two waves. At the 4-cell stage mouse embryos initiate paternally imprinted XCI, which is completed by the morula stage; hence only the maternally inherited X chromosome is active in all cells (4-1; reviewed by Takagi [14]). Imprinted XCI is maintained in the cells of the trophectoderm, which will eventually give rise to extra-embryonic tissues such as the placenta [15]. By contrast, as the embryo develops into the mid-stage blastocyst, the inactive X chromosome is reactivated in cells of the inner cell mass (ICM) that give rise to the epiblast [16–18], resulting in cells with two active X chromosomes (4-1). These cells then undergo a second wave of XCI, which is not imprinted, but rather the maternally or the paternally inherited X chromosome is chosen at random. Both imprinted and random XCI depend on *Xist*, which acts *in cis* in both cases to silence the X chromosome from which it is expressed [19–21], and the reactivation of the imprinted Xi is accompanied by *Xist* silencing (4-1) [16–18].

Early reports addressing the question of whether human early development follows what is observed in the mouse with respect to imprinted XCI have been mixed, but recent studies using

more advanced techniques and larger sample sizes agree that human pre-implantation embryos lack imprinted XCI [1,2,22], and that, instead, human pre-implantation embryos reduce X-linked gene dosage by XCD on both X chromosomes [2]. Thus, in addition to XCD and *XIST* expression from an active X chromosome, the lack of imprinted XCI in human pre-implantation embryos is a key difference between mouse and human embryonic development. Interestingly, *XIST* expression and lack of imprinted XCI are also observed in rabbit pre-implantation development, despite the closer evolutionary distance between mouse and rabbit compared with rabbit and human [1].

Another distinguishing feature between mouse and human in the epigenetic regulation of the X chromosome is the presence of the long non-coding RNA *Tsix* in mice but not in humans. *Tsix* is transcribed antisense to *Xist* and, in imprinted XCI, is expressed from the active, maternal X chromosome in mouse pre-implantation embryos and extra-embryonic annexes, where it is required to maintain *Xist* repressed on this chromosome [23,24]. Similar to imprinted XCI, *Tsix* represses *Xist* expression from the active X chromosome during random XCI [25,26]. Despite the role of *Tsix* in both imprinted and random XCI, there is a *Tsix*-independent repression of *Xist* at play during embryo cleavage stages of mouse development, because the maternal *Xist* is repressed in the absence of *Tsix* expression [24]. Although a *TSIX* gene has been annotated in the human genome, a recent study shows that it is not transcribed in human pre-implantation embryos [2]. The lack of *TSIX* expression and function may be related to the expression of *XIST* from the active X chromosomes in the human pre-implantation embryo. Thus, human cells seem to have evolved a different mechanism to control the function of *XIST* during the initiation of random XCI and to cope with *XIST* expression in the pre-implantation embryo: it is the silencing ability of *XIST* rather than *XIST* expression that is prevented in these cells. This contrasts to the mouse, where *Xist* expression systematically leads to silencing, unless certain regions of the *Xist* gene are

deleted [27]. A strong candidate for repressing *XIST*'s ability to silence the X chromosomes in the pre-implantation embryo is the recently identified human- and pluripotency-specific lncRNA *XACT* (X active coating transcript) [28].

Mouse ESCs perfectly recapitulate the X-chromosome state of the mouse blastocyst

Much of our understanding of XCI comes from mouse studies mainly because mouse embryonic stem cells (mESCs), derived from the pre-implantation blastocyst, perfectly capture the X-chromosome state of *in vivo* development [29]. Cells of the ICM and mESCs have two active X chromosomes and, upon implantation *in vivo* or differentiation *in vitro*, *Xist* expression is induced from one of the two X chromosomes, chosen at random, which leads to chromosome-wide inactivation *in cis* (4-1). The *in vitro* model system has been ideal for unravelling the molecular mechanism behind the initiation of random XCI and the transition from the XaXa (Xa for active X chromosome) to the $XaXi^{Xist+}$ state (Xi for inactive X chromosome). For instance, mESCs were used to perform extensive *Xist* RNA domain deletion studies that suggested a modular structure of *Xist* RNA, with different RNA domains mediating different functions [27]. More recently, mESCs were used to reveal that, at the onset of XCI, *Xist* spreads to regions on the X chromosome spatially closest to the *Xist* transcription locus, highlighting the importance of three-dimensional modelling of the X chromosome [30]. Moreover, two groups independently identified protein partners of *Xist* at the onset of XCI [31,32], beginning to provide a detailed mechanistic understanding of how *Xist* function is mediated [31–34].

The mouse model has also contributed immensely to our understanding of pluripotency—the ability to differentiate into all three germ layers. Pluripotent cell identity is not fixed but rather

represents a spectrum of states, perhaps because pluripotency *in vivo* spans multiple days of development instead of a fixed singular time point [35]. This became obvious when pluripotent stem cells (PSCs) with characteristics rather distinct from those of mESCs were isolated from the mouse post-implantation epiblast (EpiSCs for epiblast stem cells) [36,37]. Although both are pluripotent, mESCs capture the naive pluripotent state of the pre-implantation blastocyst and EpiSCs the developmentally more advanced primed pluripotent state of the post-implantation embryo [35].

Limitations of conventional human ESCs in modelling the pre-implantation X-chromosome state and initiation of XCI

Unlike mESCs, conventional hESCs, which are derived in the presence of basic fibroblast growth factor, do not recapitulate the X-chromosome state of the naive pluripotent cells in the human blastocyst. When comparing to what we know from mouse studies, conventional hESCs resemble mouse EpiSCs instead of naive mESCs, although, like mESCs and unlike mouse EpiSCs, they are derived from the pre-implantation and not the post-implantation blastocyst (see [38] for a detailed review). This resemblance extends to cell morphology, signaling pathway dependence with global transcriptional signature, and the post-XCI state [38]. Hence, similar to mouse EpiSCs, conventional hESCs are in primed pluripotency [35].

The X-inactivation status of hESCs has been very controversial, likely due to the epigenetic instability of the inactive X chromosome in primed hESCs. Two X-chromosome patterns can be observed at early passage, when hESCs are derived from the pre-implantation embryo in conventional conditions: one with two active X chromosomes (XaXa) and no *XIST* expression,

and one with one active and one inactive X chromosome from which *XIST* is expressed (XaXi^{XIST+}), the latter being more frequent (figure 4-2) [6,39–43]. The *XIST*-negative XaXa state was initially reported to be the pristine state, due to its resemblance to the mouse situation [40,41]. However, in our study, induction of differentiation of XaXa hESCs is accompanied neither by *XIST* induction nor by XCI [6] (figure 4-2). Because of this, we classified this *XIST*-negative XaXa state as an abnormal state, probably due to the permanent silencing of the *XIST* gene during the derivation of primed hESCs [6]. Previous studies contradicting this conclusion and reporting de novo XCI from such cells [40,41] may be explained by the heterogeneity of most hESC lines, with both XaXa and XaXi^{XIST+} cells present in the same culture before induction of differentiation. Following cells through the derivation process from human blastocysts by the analysis of a few time points suggested that the transition from the pre-implantation embryo state with two active, *XIST*-expressing X chromosomes to a post-XCI state involves transient silencing of *XIST* on both X chromosomes and its subsequent reactivation from one X only, to induce XCI [6] (figure 4-2). In this model, it may be possible that effective upregulation of *XIST* is only possible in a brief developmental window and, in cases when this window is missed *in vitro*, both *XIST* alleles become permanently silenced, leading to the stabilization of the XaXa state without *XIST* expression.

The other, more common XaXi^{XIST+} state in early passage hESC lines appears to resemble the post-XCI state of somatic cells, as shown, for example, by the occurrence of methylation of CpG islands on the Xi [6,39,44]. However, it changes in culture over time: in nearly all cases, *XIST* expression on the Xi is gradually lost in these cells, and the inactive X is partially reactivated, resulting in double dosage of a subset of X-linked genes (figure 4-2) [6,39,44–47]. This erosion of XCI is accompanied by the loss of DNA methylation specifically in the CpG islands of affected

genes, and its extent varies in different cell lines, ranging from only a handful of genes to almost the entire inactive X chromosome [6,44]. The determinants of XCI erosion are currently poorly understood, but certain regions on the Xi are more likely to erode than others [44]. Interestingly, chromatin signatures, such as H3K27me3 and H3K9me3 modifications, are good predictors of erosion, with genes enriched for H3K27me3 and relatively depleted for H3K9me3 on the Xi having an increased likelihood of reactivation upon XCI erosion [47]. A defining feature of XCI erosion is that it cannot be undone, even during differentiation [6,44,46] (figure 4-2). In other words, the aberrant X-chromosome state of these cells is locked in place so that, upon differentiation, the reactivated parts of the inactive X chromosome cannot be re-silenced, resulting in differentiated cells with a double dose of the X-linked genes that fall in eroded regions. This has not only been problematic for basic researchers who wish to study the onset of XCI in the human system, but also influences studies of X-linked diseases and use of female induced pluripotent stem cells (iPSCs) for disease modelling (see below) [46]. Furthermore, XCI erosion may affect cell replacement and regenerative therapies, because inappropriate dosage compensation of X-linked genes is a hallmark of female-specific cancers [48].

The X-chromosome state of human iPSCs, and whether reprogramming of somatic cells to pluripotency is accompanied by Xi-reactivation, has been heavily debated in the literature. Data from us and others argue that human iPSCs are XaXi with *XIST* at early passage, but over time in culture *XIST* expression is lost and the Xi is partially reactivated due to XCI erosion, similar to *XIST*-expressing XaXi hESCs [6,44,45,49,50]. Thus, despite various reports of complete Xi-reactivation in human iPSCs [51–54], our data suggest that the XaXa state is not achieved in human iPSC cultures but is unique to hESCs, consistent with the idea that it is due to the expansion of this transient state unique to the transition from the blastocyst to primed pluripotency [6].

Naive human PSCs capture features of the X Chromosome of the blastocyst

Mouse PSCs can transition from one pluripotent state to the other *in vitro*. For instance, over-expression of specific transcription factors, such as Klf4 [55] or deriving stem cells from post-implantation epiblasts in leukemia inhibitory factor (LIF) and fetal calf serum [56] achieves primed to naive conversion. The ability to convert mouse cells *in vitro* from one pluripotent state to the other inspired researchers to screen for naive culture conditions appropriate for hESCs, with the idea that establishment of the primed pluripotent state was due to culture conditions and not intrinsic to the pre-implantation human blastocysts from which these cell lines are derived. Different approaches were used in the search for media formulations supporting naive pluripotency, with most of them using small molecule inhibitors, building upon naive condition of the mouse. Hanna and co-workers [57] demonstrated that the serum-free naive culture formulation for mESCs on its own—inhibition of both glycogen synthase kinase 3 beta and extracellular signal-regulated kinase 1/2 in combination with LIF (2i/LIF)—was not enough to support human naive PSCs, and constant expression of the pluripotency transcription factors OCT4, SOX2 and KLF4 was required in combination with 2i/LIF to support naive-like human PSCs (hPSCs). They screened for small molecule inhibitors of additional pathways that could stabilize the naive-like state in the absence of exogenous OCT4, SOX2 and KLF4 expression and formulated the first naive hPSC condition termed NHSM (naive human stem cell medium) [57]. This was followed by the development of several other formulations based on different combinations of small molecule inhibitors and cytokines [58–60]. Each newly devised culture condition resulted in cells with transcriptional profiles different from the human primed PSCs and similar, to various degrees, to the naive PSCs of the human pre-implantation blastocyst, likely reflecting the stabilization of various pluripotency states by each method. To address this systematically, Huang *et al*. [61] used

an unbiased approach of comparing the transcription signature of each of these naive *in vitro* states to that of early human pre-implantation development, including oocyte, 1-, 2-, 4-, 8-cell stage embryos, morula and the blastocyst. In this analysis, two of the naive conditions—devised by Takashima *et al*. [58] and Theunissen *et al*. [59]—had the most significant gene expression overlap with the human blastocyst. Moreover, we demonstrated that the X-chromosome state of hESCs in these two culture conditions resembles that of the blastocyst, where *XIST* is expressed and accumulates on active X chromosomes [5,12]. Furthermore, the naive condition devised by Theunissen *et al.* allowed direct derivation of naive hESC lines from pre-implantation blastocysts [59], and the stabilization of the blastocyst X-chromosome state in culture [12]. Hence, we conclude that the X-chromosome state—mainly expression of *XIST* from active X chromosomes—is a reliable way of testing for true naivety of hPSCs that should be employed in assessing new naive formulations in the future. Importantly, the ability to capture the naive status of *XIST* expression in hPSCs provides a unique system to investigate the inability of *XIST* to silence the X chromosome.

When primed hPSCs harboring one active and one inactive X chromosome (with or without *XIST* expression from the Xi) are converted to naive pluripotency, the inactive X reactivates first, giving rise to XaXa cells, and only after several passages does *XIST* become expressed from either one or both X chromosomes, although the mono-allelic *XIST* pattern is dominant [12]. Interestingly, XaXa *XIST*-positive naive hESCs exhibited overall dampened X-linked gene expression levels compared with those not expressing *XIST* [12]. Thus, the correlation of XCD and *XIST* observed in human pre-implantation embryos appears to be recapitulated *in vitro* in the transition from primed to naive hESCs. These observations suggest that naive hESCs

will also serve as model system for further exploring the novel X-linked gene dosage compensation mechanism of XCD.

In addition to serving as an *in vitro* model of the pre-implantation human embryo, naive culture conditions also provide a means of overcoming the XCI anomalies observed in primed PSCs (discussed above). When primed hESCs with either a slight or very high degree of XCI erosion, or even those that are trapped in the *XIST*-negative XaXa state, are adapted to the naive culture condition described by Theunissen *et al*. [59], and then subjected to differentiation, regardless of the starting primed XCI state, all of them result in cells with the proper somatic-like X-chromosome state: with an Xa and an *XIST*-expressing Xi [12]. These findings demonstrate that XCI erosion in primed hPSCs is truly just an anomaly caused by imperfect culture conditions and can be reversed given the right media formulation. Moreover, the ability to induce de novo XCI upon differentiation of naive hESCs (figure 4-2) [12] now opens opportunities of studying this epigenetic process in the human system for the first time.

The novel lncRNA XACT and its potential role in regulating human-specific aspects of X-chromosome dosage compensation

The puzzling differences in the way dosage compensation is established in the human compared with the mouse raise the intriguing hypothesis that some regulators of the process may differ between species. *Tsix*, the *Xist* antisense transcript identified in the mouse and described above, is one such example, having an important contribution to the regulation of murine XCI and no functional orthologue in the human. More recently, through RNA-sequencing analysis, we

identified a novel X-linked lncRNA—*XACT*—which shares with *XIST* the capacity to accumulate on the chromosome from which it is expressed [28]. The appearance of *XACT* seems to be a recent event on the evolutionary scale, which took place in the higher primate branch, suggesting that it might fulfil primate (or human)-specific function [28].

Insights into such function came from the analysis of hPSCs with various X-chromosome states. In fact, expression of *XACT* is restricted to pluripotent cells: *XACT* gets silenced when the cells are induced to differentiate and reactivates upon induction of pluripotency (figure 4-2) [28]. In primed XaXi^{XIST+} cells, *XACT* is expressed from the active X only, while in *XIST*-negative XaXi cells, *XACT* is accumulating on both X chromosomes (figure 4-2). While this shift in *XACT* expression profile could simply reflect the partial reactivation of the Xi that characterizes XCI erosion, capturing the transition between the two states suggested an alternative scenario. Indeed, re-expression of *XACT* from the Xi undergoing erosion occurs before loss of *XIST* expression and prior to extended X-chromosome reactivation [47]. *XACT* reactivation from the Xi is thus not a mere consequence of erosion but is instead one of the earliest markers of this phenomenon. Pushing the reasoning further, *XACT* could causally participate in the erosion, by interfering with *XIST* expression or accumulation. In agreement with this hypothesis, when *XACT* was artificially inserted onto one X chromosome in female mESCs, XCI was biased towards the untargeted X chromosome. In other words, forced expression of *XACT* from one X reduced the likelihood of *Xist* accumulating on the very same chromosome, at least in a heterologous system [5].

What about *XACT* in the human embryo? Combining analysis of multiple datasets of single-cell RNA-sequencing and RNA-FISH confirmed that *XACT* is not an artefact of hPSC in culture, and that it is expressed in pre-implantation embryos [5]. Its expression is in fact strongly

correlated to that of *XIST* in the early developmental stages (up to early E5), where it accumulates, together with *XIST*, on every X chromosome in both male and female embryos (figure 4-2). This pattern of active X chromosomes simultaneously decorated by *XIST* and *XACT* is recapitulated to some extent in naive hPSCs derived either in 5iLAF or in t2iL + Gö conditions [5,12,58,59], further reinforcing the idea that these naive conditions indeed bookmark the *in vivo* situation. Intriguingly, in both cases *XIST* RNA was found more dispersed in the nucleus compared with cells in which *XIST* coats the Xi [5,12]. This altered distribution of *XIST* might be linked to its inability to properly silence X chromosome at these stages. As it correlates with the simultaneous presence of *XACT*, it is also tempting to speculate that *XACT* might impair proper *XIST* accumulation in human cells, as it does in the heterologous mouse system described earlier.

Other potential mechanisms preventing *XIST*-mediated silencing

XACT is one strong candidate for preventing *XIST* from silencing the X chromosome during human pre-implantation development, but additional, non-mutually exclusive scenarios can be envisioned based on recent advances in studying the mechanism of action of mouse *Xist*. For instance, Patil *et al.* [34] demonstrated that a reversible RNA modification of adenosine residues—N^6-methyladenosine (m^6A)—is enriched on *Xist* and required for its silencing ability. Differences in this or perhaps even other RNA modifications or downstream readers of such modifications in early pre-implantation versus later post-implantation stages of human development might contribute to the functional differences of *XIST*.

RNA antisense purification followed by next-generation sequencing has allowed mapping of chromatin contacts made by mouse *Xist* at the onset of XCI, and combined with chromosome conformation studies, uncovered that *Xist* first contacts distal regions on the X chromosome that are spatially close to the *Xist* transcription locus [30]. Hence one can postulate that the three-dimensional structure of the X chromosome is important when considering how *Xist* can spread along the X chromatin. Therefore, another speculation is that, due to different three-dimensional folding of the X chromosome in the pre-implantation embryo and/or expression of *XACT*, the chromatin structures might be unfavorable for *XIST* spreading and thus silencing of the X chromosome in naive pluripotency.

Several independent groups recently confirmed known and identified novel proteins that bind to mouse *Xist* RNA at the onset of XCI initiation or on the already established Xi [31–33,62–65]. Functional experiments have demonstrated that some of these *Xist* binding proteins are absolutely required for *Xist*-mediated silencing of the X chromosome. Hence it is plausible that one or more of key *XIST* interacting proteins required for its silencing ability are simply not expressed in the naive context, or that *XIST* is somehow unable to bind to such key protein factors, due to alternative splicing, the presence of competing proteins/RNAs or to chemical modifications.

CONCLUSION

The emerging studies of XCI in the human revealed a quite surprising flexibility in the way dosage compensation is established in various mammalian species [1,2,5,12]. Not only does XCI differ in kinetics and parental origin between human and mouse, but the strategies *per se* by which X-chromosome dosage imbalance is compensated for follow different routes, even if only transiently. XCD is reminiscent of the worm dosage compensation system, but the underlying mechanisms in human are still largely mysterious. We have seen that there are good reasons to believe that *XIST* could also contribute to this process. In this context, *XACT* could act as a switch for *XIST* function, from dampening X-chromosome expression (when *XACT* is present) to fully silencing it (in the absence of *XACT*). Our understanding of human dosage compensation has for long been impaired by the paucity of relevant biological material. Recent developments in the field of human naive pluripotency will help in uncovering molecular mechanisms, although it should be kept in mind that the field in still in its infancy. In this context, and as mentioned above, we believe that rigorous assessment of the X-chromosome status through monitoring *XIST* and *XACT* expression will be instrumental in assessing true naivety of hPSCs and identifying novel conditions to robustly trigger and, importantly, maintain naive pluripotency.

FIGURE LEGENDS

Figure 4-1: X-chromosome dosage compensation in mouse and human

In human pre-implantation development, XIST becomes expressed from all X chromosomes upon zygotic gene activation. As pre-implantation development progresses, XIST expression from both female X chromosomes increases, but remains low in males. The former correlates with dampened gene dosage from both X chromosomes of the blastocyst, equalizing X-linked gene dosage of females to that of males. Upon implantation, all cells undergo random XCI, again resulting in dosage-compensation. In mice, XCI happens in two waves. First, Xist is induced only on the paternally inherited X chromosome (P), causing imprinted XCI in all cells of early pre-implantation embryos (morula). As the blastocyst forms, Xist expression becomes sup- pressed in the ICM cells (but not in the TE), and the Xi reactivates, leading to increased X-linked gene dosage in females compared to males. As the embryo implants, the maternal or the paternal X chromosome becomes randomly chosen to undergo XCI, similar to humans.

Figure 4-2. X-chromosome states of human pluripotent stem cells

The X chromosome state of conventional (primed) hESCs differs from the ICM of human pre-implantation blastocysts from which they are derived. Primed hESCs are in a post-XCI state with an XIST-coated Xi. Over time in culture, the Xi loses expression of XIST and partially reactivates, undergoing XCI erosion. Primed hESCs with two active X chromosomes can also be derived from ICM outgrowths (far right), potentially capturing an intermediate state of the X chromosome in the transition to XCI. Differentiation does not change the X-chromosome state of any of these

primed hESCs. When hESCs are derived from the blastocyst under naïve culture conditions, or when primed hESCs, regardless of their X state, are converted to naïve pluripotency, the X-chromosome state resembles that of the blastocyst, with two active X chromosomes and XIST expression (on one or both X chromosomes). Like normal development, differentiation of naïve hESCs induces XCI. Similar to primed hESC derivation, an XIST-negative state with two active X chromosomes is an intermediate in the primed to naïve hESC conversion, suggesting stepwise reversal of events. The lncRNA XACT is co-expressed with XIST in naïve pluripotency and might be responsible for inhibiting XIST-mediated silencing. XIST and XACT occupy non-overlapping territories on the active X chromosome (green and purple) in naïve hESCs. XACT is also expressed in primed hESCs both from active and eroding/eroded X chromosomes, and it might be driving erosion by interfering with XIST expression or accumulation. XACT is not expressed in differentiated cells.

Figure 4-1: X-chromosome dosage compensation in mouse and human

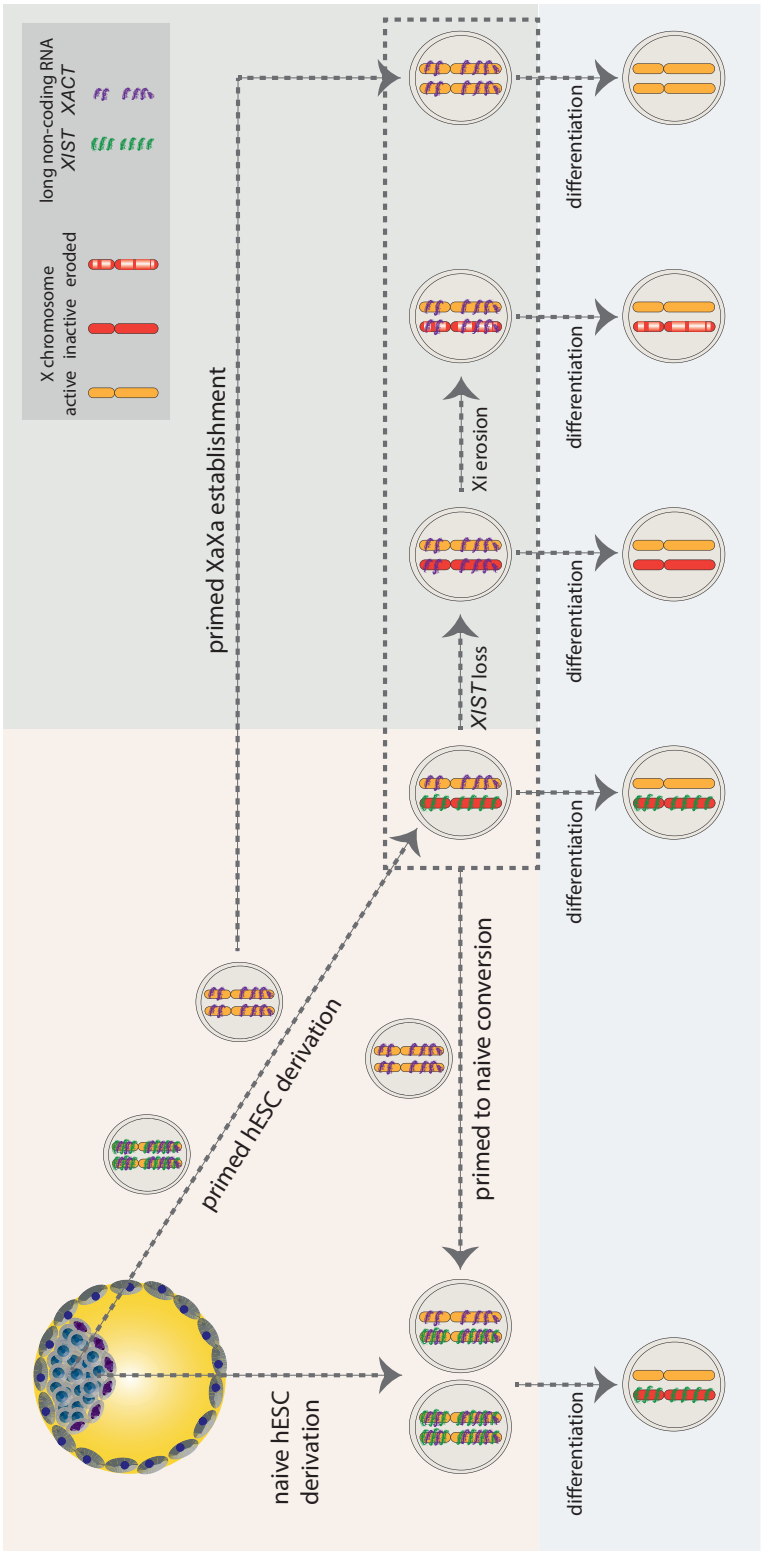
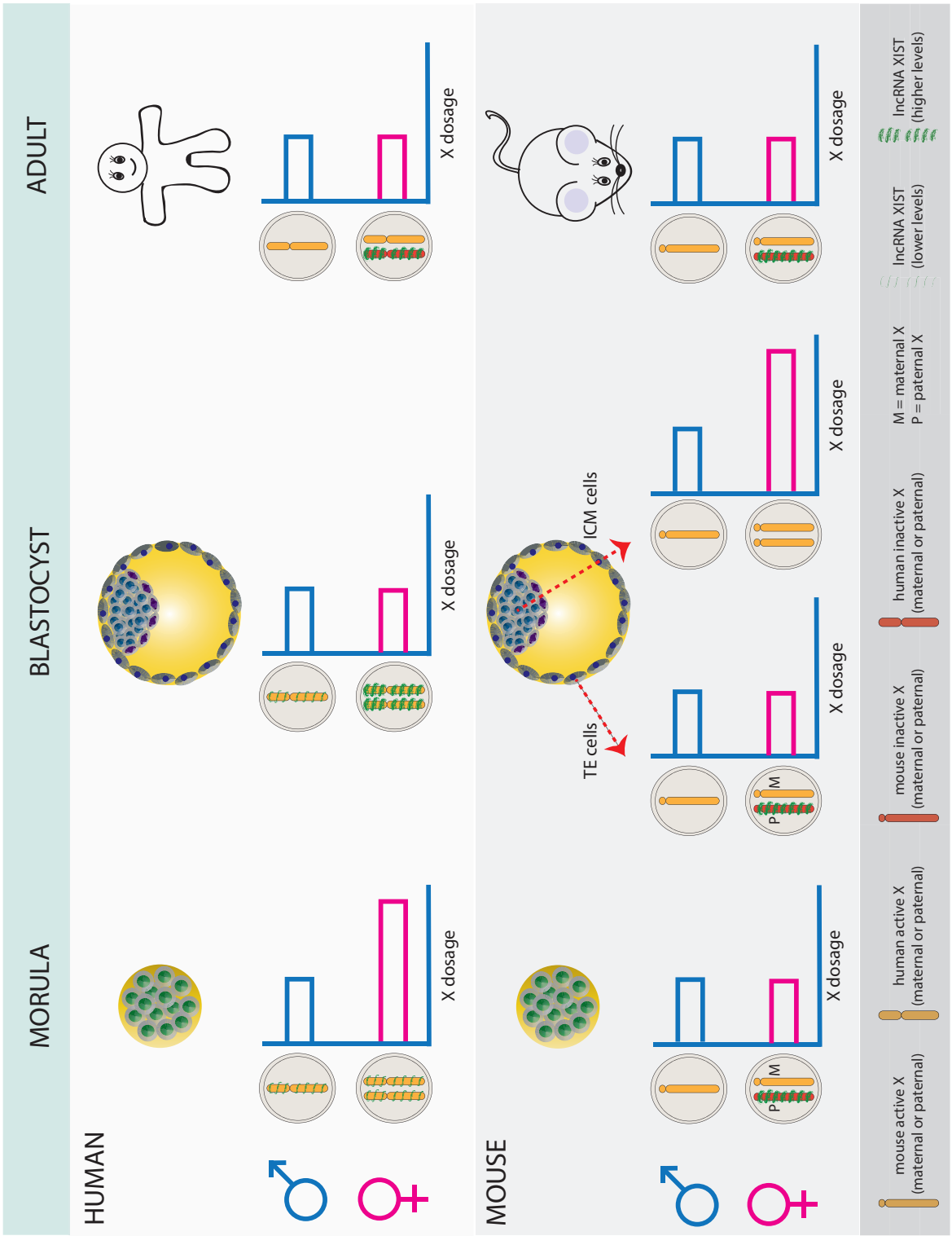


Figure 4-2. X-chromosome states of human pluripotent stem cells



REFERENCES

1. Okamoto I et al. 2011 Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature 472, 370 – 374. (doi:10.1038/nature09872)

2. Petropoulos S et al. 2016 Single-cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell 165, 1012 – 1026. (doi:10.1016/j.cell.2016.03.023)

3. Monk M, Harper MI. 1979 Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. Nature 281, 311 – 313. (doi:10.1038/281311a0)

4. Briggs SF, Dominguez AA, Chavez SL, Reijo Pera RA. 2015 Single-cell XIST expression in human preimplantation embryos and newly reprogrammed female induced pluripotent stem cells. Stem Cells 33, 1771 – 1781. (doi:10.1002/stem.1992)

5. Vallot C et al. 2017 XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. Cell Stem Cell 20, 102 – 111. (doi:10.1016/j.stem.2016.10.014)

6. Patel S et al. 2017 Human embryonic stem cells do not change their X inactivation status during differentiation. Cell Rep. 18, 54 – 67. (doi:10.1016/j. celrep.2016.11.054)

7. Meyer BJ, Casson LP. 1986 Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. Cell 47, 871 – 881. (doi:10.1016/0092- 8674(86)90802-0)

8. Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ. 2015 Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature 523, 240 – 244. (doi:10.1038/nature14450)

9. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. 1991 A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349, 38 – 44. (doi:10.1038/ 349038a0)

10. Brockdorff N et al. 1991 Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature 351, 329 – 331. (doi:10.1038/351329a0)

11. Borsani G et al. 1991 Characterization of a murine gene expressed from the inactive X chromosome. Nature 351, 325 – 329. (doi:10.1038/351325a0)

12. Sahakyan A et al. 2017 Human naive pluripotent stem cells model X chromosome dampening and X inactivation. Cell Stem Cell 20, 87 – 101. (doi:10. 1016/j.stem.2016.10.006)

13. Duret L, Chureau C, Samain S, Weissenbach J, Avner P. 2006 The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science 312, 1653 – 1655. (doi:10.1126/ science.1126316)

14. Takagi N. 2003 Imprinted X-chromosome inactivation: enlightenment from embryos in vivo. Semin. Cell Dev. Biol. 14, 319 – 329. (doi:10.1016/j. semcdb.2003.09.027)

15. Takagi N, Sasaki M. 1975 Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature 256, 640 – 642. (doi:10.1038/256640a0)

16. Mak W, Nesterova TB, de Napoles M, Appanah R, Yamanaka S, Otte AP, Brockdorff N. 2004 Reactivation of the paternal X chromosome in early mouse embryos. Science 303, 666 – 669. (doi:10. 1126/science.1092674)

17. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. 2004 Epigenetic dynamics of imprinted X inactivation during early mouse development. Science 303, 644 – 649. (doi:10.1126/science. 1092727)

18. Williams LH, Kalantry S, Starmer J, Magnuson T. 2011 Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass. Development 138, 2049 – 2057. (doi:10.1242/dev. 061176)

19. Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. 1996 Requirement for Xist in X chromosome inactivation. Nature 379, 131 – 137. (doi:10.1038/379131a0)

20. Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. 1997 Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 11, 156 – 166. (doi:10.1101/gad.11.2. 156)

21. Borensztein M et al. 2017 Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. 24, 226 – 233. (doi:10.1038/nsmb.3365)

22. Moreira de Mello JC, Arau´jo E´SS, Stabellini R, Fraga AM, Souza JES, Sumita DR, Camargo AA, Pereira LV. 2010 Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. PLoS ONE 5, e10947. (doi:10.1371/ journal.pone.0010947)

23. Lee JT, Davidow LS, Warshawsky D. 1999 Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. 21, 400 – 404. (doi:10.1038/7734)

24. Sado T, Wang Z, Sasaki H, Li E. 2001 Regulation of imprinted X-chromosome inactivation in mice by Tsix. Dev. Camb. Engl. 128, 1275 – 1286.

25. Lee JT, Lu N. 1999 Targeted mutagenesis of Tsix leads to nonrandom X inactivation. Cell 99, 47 – 57. (doi:10.1016/S0092-8674(00)80061-6)

26. Gayen S, Maclary E, Buttigieg E, Hinten M, Kalantry S. 2015 A primary role for the Tsix lncRNA in maintaining random X-chromosome inactivation. Cell Rep. 11, 1251 – 1265. (doi:10.1016/j.celrep. 2015.04.039)

27. Wutz A, Rasmussen TP, Jaenisch R. 2002 Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. 30, 167 – 174. (doi:10.1038/ng820)

28. Vallot C, Huret C, Lesecque Y, Resch A, Oudrhiri N, Bennaceur-Griscelli A, Duret L, Rougeulle C. 2013 XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. Nat. Genet. 45, 239 – 241. (doi:10.1038/ng.2530)

29. Monk M. 1981 A stem-line model for cellular and chromosomal differentiation in early mouse- development. Differ. Res. Biol. Divers. 19, 71 – 76. (doi:10.1111/j.1432-0436.1981.tb01131.x)

30. Engreitz JM et al. 2013 The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341, 1237973. (doi:10.1126/science.1237973)

31. McHugh CA et al. 2015 The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521, 232 – 236. (doi:10.1038/ nature14443)

32. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. 2015 Systematic discovery of Xist RNA binding proteins. Cell 161, 404 – 416. (doi:10.1016/j.cell. 2015.03.025)

33. Chen C-K et al. 2016 Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science 354, 468 – 472. (doi:10.1126/ science.aae0047)

34. Patil DP, Chen C-K, Pickering BF, Chow A, Jackson C, Guttman M, Jaffrey SR. 2016 m^6A RNA methylation promotes XIST-mediated transcriptional repression. Nature 537, 369 – 373. (doi:10.1038/nature19342)

35. Nichols J, Smith A. 2009 Naive and primed pluripotent states. Cell Stem Cell 4, 487 – 492. (doi:10.1016/j.stem.2009.05.015)

36. Brons IGM et al. 2007 Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature 448, 191 – 195. (doi:10.1038/nature05950)

37. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RDG. 2007 New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 448, 196 – 199. (doi:10.1038/nature05972)

38. Davidson KC, Mason EA, Pera MF. 2015 The pluripotent state in mouse and human. Development 142, 3090 – 3099. (doi:10.1242/dev.116061)

39. Shen Y, Matsuno Y, Fouse SD, Rao N, Root S, Xu R, Pellegrini M, Riggs AD, Fan G. 2008 X-inactivation in female human embryonic stem cells is in a nonrandom pattern and

prone to epigenetic alterations. Proc. Natl Acad. Sci. USA 105, 4709 – 4714. (doi:10.1073/pnas.0712018105)

40. Silva SS, Rowntree RK, Mekhoubad S, Lee JT. 2008 X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. Proc. Natl Acad. Sci. USA 105, 4820 – 4825. (doi:10.1073/pnas. 0712136105)

41. Lengner CJ et al. 2010 Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. Cell 141, 872 – 883. (doi:10.1016/j.cell.2010.04.010)

42. O'Leary T et al. 2012 Tracking the progression of the human inner cell mass during embryonic stem cell derivation. Nat. Biotechnol. 30, 278 – 282. (doi:10. 1038/nbt.2135)

43. de Oliveira Georges JA et al. 2014 Aberrant patterns of X chromosome inactivation in a new line of human embryonic stem cells established in physiological oxygen concentrations. Stem Cell Rev. Rep. 10, 472 – 479. (doi:10.1007/s12015-014-9505-4)

44. Nazor KL et al. 2012 Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell 10, 620 – 634. (doi:10.1016/j.stem.2012.02.013)

45. Tchieu J et al. 2010 Female human iPSCs retain an inactive X chromosome. Cell Stem Cell 7, 329 – 342. (doi:10.1016/j.stem.2010.06.024)

46. Mekhoubad S, Bock C, de Boer AS, Kiskinis E, Meissner A, Eggan K. 2012 Erosion of dosage compensation impacts human iPSC disease modeling. Cell Stem Cell 10, 595 – 609. (doi:10. 1016/j.stem.2012.02.014)

47. Vallot C et al. 2015 Erosion of X chromosome inactivation in human pluripotent cells initiates with XACT coating and depends on a specific heterochromatin landscape. Cell Stem Cell 16, 533 – 546. (doi:10.1016/j.stem.2015.03.016)

48. Dunford A et al. 2016 Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat. Genet. 49, 10 – 16. (doi:10.1038/ng.3726)

49. Cheung AYL, Horvath LM, Grafodatskaya D, Pasceri P, Weksberg R, Hotta A, Carrel L, Ellis J. 2011 Isolation of MECP2-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. Human Mol. Genet. 20, 2103 – 2115. (doi:10.1093/hmg/ddr093)

50. Pomp O, Dreesen O, Leong DFM, Meller-Pomp O, Tan TT, Zhou F, Colman A. 2011 Unexpected X chromosome skewing during culture and reprogramming of human somatic cells can be alleviated by exogenous telomerase. Cell Stem Cell 9, 156 – 165. (doi:10.1016/j.stem.2011.06.004)

51. Anguera MC et al. 2012 Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. Cell Stem Cell 11, 75 – 90.

(doi:10.1016/j.stem.2012.03.008)

52. Tomoda K et al. 2012 Derivation conditions impact X-inactivation status in female human induced pluripotent stem cells. Cell Stem Cell 11, 91 – 99. (doi:10.1016/j.stem.2012.05.019)

53. Kim K-Y, Hysolli E, Tanaka Y, Wang B, Jung Y-W, Pan X, Weissman SM, Park I-H. 2014 X Chromosome of female cells shows dynamic changes in status during human somatic cell reprogramming. Stem Cell Rep. 2, 896 – 909. (doi:10.1016/j.stemcr.2014. 04.003)

54. Barakat TS et al. 2015 Stable X chromosome reactivation in female human induced pluripotent stem cells. Stem Cell Rep. 4, 199 – 208. (doi:10. 1016/j.stemcr.2014.12.012)

55. Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A. 2009 Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development 136, 1063 – 1069. (doi:10.1242/dev.030957)

56. Bao S, Tang F, Li X, Hayashi K, Gillich A, Lao K, Surani MA. 2009 Epigenetic reversion of post- implantation epiblast to pluripotent embryonic stem cells. Nature 461, 1292 – 1295. (doi:10.1038/ nature08534)

57. Gafni O et al. 2013 Derivation of novel human ground state naive pluripotent stem cells. Nature 504, 282 – 286. (doi:10.1038/nature12745) Takashima Y et al. 2014 Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell 158, 1254 – 1269. (doi:10.1016/j.cell.2014.08.029)

58. Theunissen TW et al. 2014 Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15, 471 – 487. (doi:10.1016/j.stem.2014.07.002)

59. Ware CB et al. 2014 Derivation of naive human embryonic stem cells. Proc. Natl Acad. Sci. USA 111, 4484 – 4489. (doi:10.1073/pnas. 1319738111)

60. Huang K, Maruyama T, Fan G. 2014 The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. Cell Stem Cell 15, 410–415. (doi:10.1016/j.stem.2014.09.014)

61. Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K, Nakagawa S. 2010 The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. Dev. Cell 19, 469–476. (doi:10.1016/j. devcel.2010.08.006)

62. Minajigi A et al. 2015 A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science 349, aab2276. (doi:10.1126/science. aab2276)

63. Moindrot B et al. 2015 A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. Cell Rep. 12, 562–572. (doi:10.1016/j.celrep.2015.06.053)

64. Monfort A, Di Minin G, Postlmayr A, Freimann R, Arieti F, Thore S, Wutz A. 2015 Identification of spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. Cell Rep. 12, 554–561. (doi:10.1016/j.celrep.2015.06.067)

CHAPTER 5

Concluding Remarks

The fact that X-chromosome inactivation (XCI) is conserved in all female mammals already speaks to the importance of this phenomenon. The inactivated X chromosome is rather remarkable, not only in appearance under the microscope, which was how it was first discovered [1], but also because it is a unique example of dramatic epigenetic reprogramming that affects life-long gene expression. Our understanding of the inactive X chromosome, both the steps leading to it and those maintaining it, have improved dramatically in the last few decades. Mainly, we now understand that initiation of XCI requires the long non-coding RNA (lncRNA) X inactive specific transcript (*Xist*), which is encoded on the X chromosome. *Xist* is a very interesting lncRNA since it can silence an entire chromosome – one of the larger chromosomes in the cell – without silencing its own expression. Furthermore, *Xist*-mediated silencing is long-lived, since the inactive X chromosome (Xi) remains transcriptionally silent for the rest of the cell's and its progeny's life.

Interestingly, findings based on studies of X-chromosome inactivation can inform us beyond the events of X-chromosome dosage regulation. This was the case with our study aimed at understanding the maintenance phase of XCI using engineered mouse somatic cells harboring a reporter gene on the Xi [2]. By subjecting the reporter cells to various RNAi or small chemical-based treatments along with low dose of DNA-demethylating agent, we discovered a cell intrinsic mechanism that, upon inhibition, enhanced DNA-demethylation not only of the X-chromosome, but genome-wide. Interestingly, by understanding the molecular pathway, we could predict a chemical combination affecting the very same pathway using FDA-approved dual drug combination. We extended our studies beyond the field of XCI research, and demonstrated that our identified dual drug combination, which previously had not been combined for cancer therapy, worked in synergy to inhibit human cancer cell growth *in vitro.* Thus, we aimed to understand the

maintenance phase of mouse XCI, and arrived at a novel drug combinatorial therapy for human cells – male or female.

XCI is governed by *Xist,* which is one of the best studied lncRNAs to date for multiple reasons. It 1) has a conserved gene structure in mouse and human, 2) has a dramatic outcome – chromosome silencing – when expressed, 3) serves as a model for understanding how lncRNAs influence gene regulation, and 4) has a short and critical developmental window to act early in embryogenesis. Hence, *Xist* unites biologists from multiple disciplines, including evolutionary biology (point 1 above), gene regulation (points 2 and 3 above), and developmental biology (point 4 above). Most of our understanding on XCI biology and *Xist* are based on mouse studies, mainly because mouse embryonic stem cells (mESCs), isolated from the inner cell mass (ICM) of the pre-implantation blastocyst which develop into the embryo, are an excellent model system as they perfectly recapitulate *in vivo* events *in vitro* with respect to X-chromosome dosage regulation. However, recent studies using human pre-implantation blastocysts have made it rather clear that the X-chromosome state of human pre-implantation development is 1) rather different from that in mouse and 2) is not represented with conventional human embryonic stem cells (hESCs) [3]. Furthermore, we and others have demonstrated that conventional hESCs, as well as human induced pluripotent stem cells (hiPSCs) derived by epigenetic reprogramming of somatic cells to pluripotency using transcription factors, are in a post-XCI state with an already established Xi [4 5]. However, over time in culture, the Xi in hESCs and hiPSCs (together hPSCs), loses expression of *XIST* and undergoes epigenetic erosion – de-repression of silencing of certain genes on the Xi [6]. When induced to differentiate, hPSCs have the ability to give rise to cells from any of the three germ layers (endoderm, ectoderm, mesoderm); however, unlike in normal development, the X-chromosome state does not change during differentiation [5]. This absence of change in X-

chromosome state upon exiting pluripotency in human cells raises two main issues: 1) hPSCs are not a useful model for studying initiation of XCI in human cells with cultured cells, and 2) hPSC-derived cells fail to have proper dosage compensation for genes that fall in eroded regions of the Xi, which is problematic since failure of proper dosage compensation correlates with female-specific cancers in humans [7].

By adapting conventional hPSCs to naïve culture conditions, we have demonstrated that both of the above-mentioned shortcomings of conventional hPSCs can be resolved. Interestingly, we used the X-chromosome state itself as a marker for true naïve pluripotency to find a culture condition that, while recapitulating the X-chromosome state of the human pre-implantation blastocyst, also resolves the epigenetic abnormalities of the Xi in conventional hPSCs [8]. Our readout was the presence of *XIST* lncRNA from an active X chromosome (Xa), since this non-silencing *XIST* expression was recently demonstrated to be unique to the human blastocyst [9]. We identified that two independently formulated naïve culture conditions [10 11 12] resulted in cells with transcriptionally active X chromosomes and *XIST* expression. Furthermore, differentiation of naïve hPSCs resulted in *XIST*-mediated XCI, modeling initiation of XCI in human cells *in vitro* for the very first time. Together, our findings open many research avenues, making it possible to 1) address how *XIST* does not lead to silencing in naïve pluripotency, 2) identify the steps of XCI initiation in human and compare them to mouse, and 3) discover is the molecular 'switch' that allows correction of the epigenetically abnormal Xi in conventional hPSCs upon adaptation to naïve condition.

Prior to our discovery, much of our understanding on the X-chromosome state of early human development was based on pre-implantation blastocysts which are a very scarce resource not available in many research institutions. Moreover, due to small cell numbers in each blastocyst

(several hundred), they cannot be used for many genomics experiments which require thousands of cells for high resolution understanding of the epigenetic state of these cells. Lastly, each blastocyst is unique, making data reproduction inherently complicated. Cultured hPSCs that truly recapitulate the blastocyst state, therefore, are an unparalleled tool since they are easily accessible, adaptable, expandable, and reproducible.

Despite the improvements over conventional hPSC culture conditions, there still needs to be further research to improve naïve culture conditions to better resemble the pre-implantation state of human development. Mainly, the naïve culture condition we use result in cells where majority have mono-allelic *XIST* expression from one of the two active X chromosomes, yet most cells in the blastocyst express *XIST* bi-allelically. While the mono-allelic expression can still be used to address how *XIST* does not lead to silencing, the bi-allelic state is more desirable since all the X chromosomes will be equivalent in that state, ruling out heterogeneity issues that might arise. Moreover, *XIST* expression correlates with reduced but not silenced gene expression – X chromosome dampening (XCD) – in naïve hPSCs, a phenomenon also observed in pre-implantation blastocysts by single cell RNA-sequencing studies [13] If *XIST* was expressed from both X-chromosomes and it is the regulator of XCD, then the XCD phenotype would be more robust in naïve hPSCs with bi-allelic *XIST* expression and more easily addressable for molecular understanding. Furthermore, while differentiation of naïve hPSCs results in *XIST*-mediated XCI, this event, unlike in the cells of the ICM in development, is not random [8]. Hence, there seems to be an epigenetic memory at play that is inherited from the conventional hPSC state, and the memory is either of the Xi, the Xa, or both. Understanding what this epigenetic memory is will aid in improving naïve culture conditions to allow for a formulation that would allow modeling of random XCI *in vitro.*

Our understanding of how mouse *Xist* functions to induce chromosome-wide silencing has improved dramatically in the last few years due to recently devised genomics techniques. These include next generation-based nucleic acid sequencing techniques that identify chromatin sites that *Xist* contacts upon initiation of XCI, as well as mass spectrometry-based approaches that identify direct and indirect protein partners of *Xist* [14]. Now, with the aid of naïve hPSCs, such studies can be extended to the human system since we can model *XIST*-mediated initiation of XCI. Moreover, we can use these techniques to understand how somatic *XIST* in human cells causes silencing, yet the very same molecule in naïve pluripotent cells does not. We can also further investigate the correlation between *XIST* expression and XCD in naïve pluripotency using CRISPR/Cas9-based genetic and epigenetic modifications of *XIST* and/or *XIST*-binding protein partners.

Overall, our ability to culture naïve hPSCs with a well-defined X-chromosome state is a big achievement towards deciphering how the conserved epigenetic process of X-chromosome inactivation occurs in human development. This will not only extend our knowledge of developmental biology, but also aid model X-linked diseases and develop safe cell-based therapies for more than half of our adult population – women.

REFERENCES

1. Ohno, S. & Hauschka, T. S. Allocycly of the X-chromosome in tumors and normal tissues. *Cancer Res.* **20,** 541–545 (1960).
2. Minkovsky, A. *et al.* A high-throughput screen of inactive X chromosome reactivation identifies the enhancement of DNA demethylation by 5-aza-2'-dC upon inhibition of ribonucleotide reductase. *Epigenetics Chromatin* **8,** 42 (2015).
3. Sahakyan, A., Plath, K. & Rougeulle, C. Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. *Philos. Trans. R. Soc. B Biol. Sci.* **372,** 20160363 (2017).
4. Tchieu, J. *et al.* Female Human iPSCs Retain an Inactive X Chromosome. *Cell Stem Cell* **7,** 329–342 (2010).
5. Patel, S. *et al.* Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. *Cell Rep.* **18,** 54–67 (2017).
6. Mekhoubad, S. *et al.* Erosion of Dosage Compensation Impacts Human iPSC Disease Modeling. *Cell Stem Cell* **10,** 595–609 (2012).
7. Chaligné, R. & Heard, E. X-chromosome inactivation in development and cancer. *FEBS Lett.* **588,** 2514–2522 (2014).
8. Sahakyan, A. *et al.* Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* **20,** 87–101 (2017).
9. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472,** 370–374 (2011).
10. Theunissen, T. W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15,** 471–487 (2014).

11. Takashima, Y. *et al.* Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* **158,** 1254–1269 (2014).
12. Guo, G. *et al.* Epigenetic resetting of human pluripotency. *Development* **144,** 2748–2763 (2017).
13. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **165,** 1012–1026 (2016).
14. da Rocha, S. T. & Heard, E. Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nat. Struct. Mol. Biol.* **24,** 197–204 (2017).