Title
X Chromosome Dosage Compensation During Human Early Embryonic Development

Author
Chitiashvili, Tsotne

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

X Chromosome Dosage Compensation During Human Early Embryonic Development

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy in

Molecular Biology

by

Tsotne Chitiashvili

2022

ABSTRACT OF THE DISSERTATION

X Chromosome Dosage Compensation During Human Early Embryonic Development

by

Tsotne Chitiashvili

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2022

Professor Kathrin Plath, Co-Chair

Professor Amander T. Clark, Co-Chair

Mammalian embryonic development is one of the most complex biological processes that involves multiple epigenetic events. One of the major epigenetic processes is X chromosome dosage compensation in female mammalian cells harboring two X chromosomes. The balance of X-linked gene levels between male and female cells can be achieved by X chromosome inactivation (XCI) or X chromosome dampening (XCD). Most of our knowledge about this complex process comes from mouse studies. Through decades of research, it has been established that the long-noncoding RNA *Xist* orchestrates XCI in mice. However, recent publications revealed intriguing epigenetic differences of the X chromosome and *Xist* function between mouse and human. In human, a different form of dosage compensation acts in preimplantation embryos and results in a dampening, but not silencing, of genes on both X chromosomes (XCD).

Additionally, the localization of *XIST* lncRNA to the dampened X chromosome, indicates that *XIST* RNA can be expressed without inducing silencing, which has never been observed in the mouse. Similar to the human pre-implantation embryo, *XIST* is also expressed in female human primordial germ cells (hPGCs) after the inactive X chromosome is reactivated. In our work, we also demonstrate that similar to preimplantation human blastocysts X chromosome dampening is taking place in hPGCs *in vivo*. Additionally, using single cell RNA sequencing we demonstrate that expression of *XIST* in female hPGCs correlates with downregulation of X linked genes. Lastly, we demonstrate that primate specific lncRNA *XACT,* that has been described in the literature as pluripotency specific lncRNA, is also explicitly expressed from both active X chromosomes in hPGCs.

To address the question whether *XIST* is mediating X chromosome dampening we have performed functional experiments in naïve human embryonic stem cell (ESC) lines that capture the preimplantation state of the X with *XIST* RNA expression from the active X chromosome and X chromosome dampening (XCD). We show that *XIST* deletion results in upregulation of dampened X linked genes. Additionally, we demonstrate for the first time that *XIST* spreads to specific autosomal regions in naïve pluripotent stem cells and regulate their levels. However, it still remains unclear whether *XIST* has a role in regulating autosomal genes during early development *in vivo*. Taken together, our studies reveal insights into how epigenetic mechanisms differ between mouse and human, increase our knowledge of X chromosome regulation, and create strong bases to understand transmission of X-linked diseases through the generations.

The dissertation of Tsotne Chitiashvili is approved.

William E. Lowry

April D. Pyle

Dana Leanne Jones

Siavash K. Kurdistani

Amander T. Clark, Committee Co-Chair

Kathrin Plath, Committee Co-Chair

University of California, Los Angeles

2022

To my family and friends

TABLE OF CONTENTS

Figures

CHAPTER 1

CHAPTER 2

ACKNOWLEDGEMENTS

From the early days of my undergraduate studies I was fascinated by non-coding RNAs and how they regulate gene expression in development, in particular, during germline formation. UCLA was a unique place for me where I could work on a project that I have been excited for many years. I will be forever grateful to my mentors, Dr. Kathrin Plath and Dr. Amander Clark for believing in me and supporting me throughout this journey. Their guidance and enthusiasm for science helped me develop not only as a scientist but a better person. Thanks for supporting my ideas and constructive criticism, that I also found extremely helpful. I am proud to have been a member of the Plath lab and the Clark lab.

I would like to thank all members of the Plath and Clark labs for their support, friendship and encouragement. I would like to specially thank Drs. Yolanda Markaki and Costas Chronis, for their help with microscopy and genomics experiments and for their friendship, their support has been invaluable. I would like to thank Drs. Iris Dror and Fei-man Hsu for their support with bioinformatics. I am thankful to Dr. Rachel Kim for her help with embryo work. Thanks to my bay-mates Shawn, Nivi and Amy for friendship and the fun atmosphere in our bay. Thanks to Dr. Anna Sahakyan for working together especially at the start of my PhD. Thanks to Rohan Chaudhari for being an amazing student and for his contributions to our work. Thanks to Drs. Di Chen and Grace Hancock for their help with PGCLC work. Robin McKee and Tim Hunt do a tremendous job in making sure Plath and Clark labs run smoothly. I am thankful to my committee members for their scientific input and continuous support throughout my PhD work.

I would like to acknowledge Boehringer Ingelheim Fonds for their generous support towards my PhD studies. My gratitude also extends to the Graduate Division at UCLA. I also would like to acknowledge the Molecular Biology Institute administration and staff for their constant support, special thanks to Ashley Straight and Stephanie Cuellar. I am grateful to all staff members of BSCRC FACS, microscopy and genomics cores, also genomics core at JCCC, their help has been invaluable throughout this work.

I am grateful to my parents, who have taught me about the importance of science from my early childhood days. I am thankful for their personal sacrifices especially during extremely difficult times back in my home country. I will always cherish and appreciate their support.

Last but not least I would like to thank love of my life - my wife Teo, who came with me across the world. I am extremely fortunate to have her in my life and without her constant support this work would be impossible.

Chapter 2 is peer-reviewed publication of our work in the journal *Nature Cell Biology* authored by **Chitiashvili T**, Dror I, Kim R, Hsu F, Chaudhari R, Pandolfi E, Chen D, Liebscher S, Schenke-Layland K, Plath K** and Clark A** (***Co-corresponding authors*)

Chapter 3 is a reformatted manuscript of our work currently under revision in the journal *Cell*. Authored by Dror I*, **Chitiashvili T***, Sahakyan A*, Tan S, Ding F, Miller J, Plath K. (**Equal contributions*)

Chapter 4 is a reformatted manuscript of our work currently submitted for publication in the journal *Stem Cell Reports.* Authored by **Chitiashvili T**, Hsu F, Dror I, Plath K and Clark A

VITA

Education

2017 **– PhD candidate** in the laboratories of Prof. Kathrin Plath and Prof. Amander Clark, *UCLA*

2015 **– M.Sc. in Biology,** *Ludwig Maximilians University, Munich, Germany*

2013 **– B.Sc. in Health Biotechnology,** *Tbilisi State University, Georgia*

2009 **– High-school Diploma***, Public School #42, Tbilisi, Georgia*

Publications

2022 – **Chitiashvili T**, Hsu F, Dror I, Plath K and Clark A "*FGFR3 is Expressed by Human Primordial Germ Cells and is Repressed after Meiotic Initiation to Form Primordial Oocytes"* ***Stem Cell Reports Submitted.***

2021 – Dror I*, **Chitiashvili T***, Sahakyan A*, Tan S, Ding F, Miller J, Plath K. "*XIST controls X chromosome dampening and autosomal genes in early human development"* ***Cell. Under revision.***

2021 – Guo J, Sosa E* & **Chitiashvili T***, Nie X, Rojas E, Oliver E, Connect D, Kim R, Carrell D, Plath K, Hotaling J, Stukenborg J, Clark A** and Cairns B**. "*Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment*" ***- Cell Stem Cell.*** DOI: 10.1016/j.stem.2020.12.004

2020 – **Chitiashvili T**, Dror I, Kim R, Hsu F, Chaudhari R, Pandolfi E, Chen D, Liebscher S, Schenke-Layland K, Plath K** and Clark A**. "*Human female primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation"* ***Nature Cell Biology.*** DOI: 10.1038/s41556-020-00607-4

2020 – Pandya-Jones A, Markaki Y, Serizay J, **Chitiashvili T**, Mancia W, Damianov A, Chronis C, Papp B, Chen C, McKee R, Wang X, Chau A, Leonhardt H, Zheng S, Guttman M, Black D. and Plath K. "*A protein assembly mediates Xist localization and gene silencing"* ***Nature.*** DOI: 10.1038/s41586-020-2703-0

2018 – Tao Y, Yen M, **Chitiashvili T**, Nakano H, Kim R, Hosohama L, Tan Y, Nakano A, Chen P, Clark A. "*TRIM28-Regulated Transposon Repression Is Required for Human Germline Competency and Not Primed or Naive Human Pluripotency*" ***Stem Cell Reports.*** DOI:10.1016/j.stemcr.2017.11.020

2017 – Kraus F, Miron E, Demmerle J, **Chitiashvili T**, Budco A, Alle Q, Matsuda A, Leonhardt H, Schermelleh L, Markaki Y. "*Quantitative 3D structured illumination microscopy of nuclear structures"* ***Nature Protocols.*** DOI:10.1038/nprot.2017.020

(Equal contribution, **Co-corresponding authors)*

Poster & Oral presentations

2020 – Oral presentation at **Cold Spring Harbor Laboratories (CSHL) Germ Cells** virtual conference

2019 – Oral presentation at **International Society for Stem Cell Research annual meeting** - ISSCR - Los Angeles

2019 – Oral and Poster presentations at **Germinal Stem Cell Biology Gordon Research** Conference - Hong Kong

2019 – Poster presentation at **UCLA Molecular Biology Institute annual retreat** - Arrowhead

2019 – Poster presentation at **15th Annual Stem Cell Symposium** - UCLA

2018 – Poster presentation at **UCLA Molecular Cell and Dev. Biology Department annual retreat** - Arrowhead

2012 – Poster at Amgen Scholars Symposium in Cambridge*, UK*

2012 – Poster at Amgen Scholars Closing Symposium in Munich*, Germany*

Awards and Distinctions

2019 – Poster prize at Molecular Biology Institute annual retreat, UCLA

2018 – Boehringer Ingelheim PhD Fellowship for research at UCLA

2018 – Poster prize at Molecular Cell and Developmental Biology retreat, UCLA

2015 – Lehre@LMU Funding Award for own research project

2013-15 – DAAD Scholarship for Master Studies in Germany

2012 – First prize at the Amgen Scholars Symposium in Cambridge

2009-13 – Awarded the Tbilisi State University stipend five times

2011 – Manana Gegechkori Scholarship for success in biology

Teaching

2020 – **Teaching assistant,** Developmental Biology (Upper division class MCDB 138). *University of California - Los Angeles.*

2018 – **Teaching assistant,** Molecular Biology of Cellular Processes (Upper division class MCDB 144). *University of California - Los Angeles*

2016 – **Seminar Series,** Model organisms and Stem cell biology. *Tbilisi State University, Tbilisi, Georgia*

2015 **– Teaching assistant,** Bio-imaging practical course for Master students. *LMU Bio-center, Munich, Germany*

2015 – **Teaching assistant,** iRES (hands on research) practical course. *LMU Bio-center, Munich Germany*

2014 – **Teaching assistant,** Model organisms in biological research theoretical and practical course. *Free University, Tbilisi, Georgia*

CHAPTER 1

X Chromosome regulation during human early embryonic development

X Chromosome dosage compensation

Numerous complex epigenetic events take place during mammalian embryonic development [1]. The interplay between DNA and chromatin modifications ensures switching on and off different sets of genes at certain time points of development. One of the major epigenetic processes, X chromosome inactivation (XCI), takes places specifically during early female embryonic development [2]. During mouse pre-implantation development, the paternal X chromosome becomes inactive in an "imprinted" fashion, while the maternal X chromosome remains active[3]. A little later, reactivation of the inactive X chromosome (the Xi) occurs in the epiblast cells of the blastocyst [1]. Upon implantation, with further development of the embryo, random XCI, affecting the maternally or paternally inherited X chromosome at random chance, is established in all cells of the developing embryo. While the Xi is maintained in all somatic cells, a second wave of Xi-reactivation takes place in female mouse primordial germ cells (PGCs). XCI, in the random and imprinted form is mediated by the long non-coding RNA (lncRNA) *Xist* that spreads from its transcription site in *cis* to coat the whole chromosome from which it is transcribed. *Xist* RNA is required and sufficient to induce XCI [4]. *Xist* expression has been widely used in mouse model as an indicator of inactive X chromosome.

Most of our knowledge about XCI and *Xist*'s functions are based on studies employing mouse embryonic stem cells (mESCs) [2]. These cells are derived from the pre-implantation blastocyst and capture the naïve pluripotent state of the epiblast cells in the pre-implantation blastocyst. Accordingly, mESCs harbor two active X chromosomes (Xa's) and lack *Xist* expression. Upon induction of differentiation, one of the randomly

chosen X chromosomes initiates expression of *Xist* and undergoes XCI [5]. Therefore, the *in vitro* differentiation of mESCs recapitulates embryonic development and has been largely used to understand epigenetic remodeling of X chromosome (**Fig. 1**).

Even though *Xist* is one of the well-studied lncRNAs to date [6], the precise molecular mechanisms of its function are yet to be determined. Studies employing mESCs revealed that *Xist* consists of various functional RNA domains [7], and that the RNA initially localizes to the DNA regions that most frequently contact the *Xist* genomic locus in 3-dimensional proximity, thereby exploiting 3D genome organization to find its regulatory targets across the X chromosome [8]. At each of these sites, *Xist* interacts with the nuclear matrix protein SAF-A, which provides the physical link that tethers the *Xist* lncRNA to genomic DNA [9]. Biochemical purification combined with mass spectrometry, identified SHARP as a direct *Xist*-interacting protein that is required for transcriptional silencing of genes on the X and functions with HDAC3 to trigger deacetylation on the X chromosome [10], [11]. The identification of *Xist*-interacting proteins has opened new horizons to address the mechanisms and functions of this lncRNA.

X Chromosome regulation during human preimplantation development

In contrast to mESCs, conventional human ESCs (hESCs) and induced pluripotent stem cells (hiPSCs) resemble mouse post-implantation epiblast stem cells (EpiSCs) with respect to their epigenetic and transcriptional landscape, and consequently harbor one active and one inactive X chromosome (XaXi^{XIST+}) (**Fig. 1**). These cells are also referred as primed pluripotent stem cells [12]. Thus, hESCs do not capture the epigenetic state of

the pre-implantation embryo as *in vitro* culture conditions during hESC derivation stabilize the primed pluripotent state (**Fig. 1**).

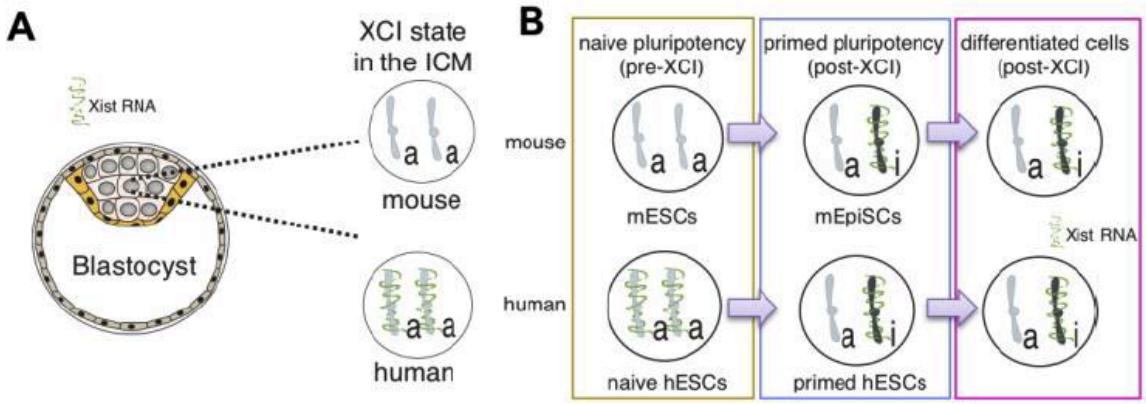


Figure 1. *XIST* expression and X chromosome differences in mouse and human

(A) At the blastocyst stage (*in vivo*), in mouse, female naïve pluripotent stem cells harbor two active X chromosomes and lack expression of *Xist*, whereas human naïve pluripotent cells carry two active X chromosomes, both of them expressing *XIST* (Dampened state). **(B)** In both human and mouse, female primed pluripotent and somatic cells one X chromosome is inactivated by *Xist*. Only naïve pluripotent stem cells in mouse and human resemble the blastocyst state. a – Active X chromosome (grey), I – inactive (dark).

Intriguingly, it has been shown that epiblast cells of the human blastocyst, which are in the naïve pluripotent state, harbor two active X chromosomes and express *XIST* from both alleles [13] (**Fig. 1**). This result is dramatically different from the mouse, where *Xist* has never been found expressed on an Xa. Since *Xist* has always been known to associate with XCI in mouse, this rather surprising finding in human embryos raises the question of what *XIST*'s function is on an active X chromosome and why it would not initiate XCI. Interestingly, recent studies also showed that *XIST* expression from active X chromosomes correlates with dampened X-linked gene expression [14], [15]. Petropoulos

and colleagues studied the transcriptome of the largest group of human pre-implantation embryos, performing sex-specific analysis of human development at days 3 to 7 post fertilization (E3-E7) at the single-cell level. Their analysis revealed that immediately after zygotic gene activation at E4, female embryos had almost double expression of X-linked genes compared to males, consistent with females having two active X chromosomes. However, with increasing developmental time from E4 to E7, this roughly 2:1 female to male ratio decreased, reaching nearly 1:1 in all cells of the embryo at E7. Surprisingly, this drop in X-linked gene expression level was not due to the onset of XCI, since allelic expression analysis by single cell RNA-sequencing revealed that both X chromosomes were active at all times. Thus, Petropoulos et al uncovered a novel mechanism of X chromosome dosage compensation in human pre-implantation embryos, 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. They also found that X chromosome dampening (XCD), which has not been observed in mice, correlates with *XIST* induction during pre-implantation development [14]. Thus, in early human embryonic development two different dosage compensation mechanisms are in play, first XCD and later XCI (**Fig 1**). In both cases *XIST* is expressed from the dosage compensated X chromosome(s). *However, whether XIST is responsible for XCD remains unclear. Furthermore, the reason or the mechanisms behind XIST expression from the active X chromosomes and why it cannot initiate silencing in naïve pluripotent cells, but can later in primed pluripotent cells, are still open questions.* In any case, the unique state of the X and *XIST* highlight the importance of mechanistic studies of human development.

Additional factor playing a role in species specific differences could be expression of primate specific lncRNA *XACT* in human preimplantation embryos [16]. More specifically *XACT* has evolved in higher order primates approximately 25 Million years ago and it is not present e.g. rhesus macaques and cynomolgus monkeys [17]. Using RNA-FISH, *XACT* has been found to be accumulating and co-activating alongside XIST on active X chromosomes in both human preimplantation embryos and in human naïve stem cells, although the two transcripts do not co-localize[14], [15], [18]. *XACT*, like *XIST*, is found in male cells on the active X chromosome and is found in either one or both X chromosomes in females. Additionally, in female embryos, *XACT*-negative cells are disproportionately from the TE, which could be due to TE cells going through XCI earlier than PE and epiblast cells. Between the 4-cell and 8-cell stages there was a large burst in expression of both *XACT* and *XIST* on the X chromosomes, suggesting that two lncRNAs may be co-regulated as well[18]. However, co-expression of *XACT* and *XIST* is not maintained in developmentally older embryos (E5-E7), which indicates that post-initiation the two lncRNA can be differentially regulated. Cells positive for *XIST* and *XACT* tend to display a dispersed pattern of *XIST* as opposed to the tight cloud that is usually seen[18]. Due to this, the current hypothesis is that *XACT* inhibits or disturbs *XIST's* role in silencing the X chromosome. To demonstrate this, Vallot *et al* knocked in a *XACT* transgene into mice and observed a bias in the localization of *XIST*, which preferentially transcribes on the X chromosome that does not express *XACT*. Knockdown of *XACT* on these cells resulted in XCI reverting to a random pattern. Although it has been suggested that *XACT* maintains X chromosomes active, whether accumulation of XACT on the X chromosome has direct impact on XCD and XIST localization remains unknown.

To establish and maintain hESCs in the naïve pluripotent state that capture cells the epigenetic and transcriptional state of the naïve epiblast cells of the human blastocyst, several research groups developed *in vitro* culture conditions [19],[7], [22]. Recent work from Dr. Plath's lab on female naïve hESCs obtained with the 5iLAF naïve culture condition developed by the Jaenisch lab [19], demonstrated that in these naïve hESCs, *XIST* is expressed from the active X-chromosomes with dampened expression specifically of X-linked genes, recapitulating the state of the human pre-implantation blastocyst [15] (**Fig. 1**). Moreover, these cells undergo XCI upon induction of differentiation [15]. Thus, it is now possible to examine the pre-XCI state, with XCD and *XIST* expression from an Xa, and the transition to XCI *in vitro,* allowing us to dissect molecular characteristics of human X chromosome dosage compensation mechanisms. For example, naïve hESCs now enable approaches that require large quantities of starting material (e.g. RAP-Seq, RAP-MS, see below), that were unfeasible before due to limiting number of cells from pre-implantation embryos.

X chromosome regulation during germline development

In human development, the Xi only reactivates again in one cell lineage, in human germ cell progenitors called Primordial Germ Cells (PGCs) [23]. Human germ cell formation begins with the specification of PGCs in the post-implantation embryo at the end of week 2 post-fertilization (pf) [24]. At the end of week 3 pf, a cluster of PGCs can be identified in the Yolk Sac endoderm, corresponding to specified PGCs that ultimately differentiate into eggs and sperm in the adult ovary or testis, respectively. Following specification,

PGCs migrate from the Yolk Sac niche, through the dorsal mesentery of the hind gut to colonize a new niche, the genital ridge epithelium (the future gonad) starting at around week 5 pf. Once colonized in the developing fetal gonads, PGCs are referred to as gonocytes, late PGCs, Fetal Germ Cells (FGCs) or in ovaries oogonia [25], [26]. While male PGCs do not start meiosis during prenatal development, in the embryonic ovary, PGCs initiate the process of meiotic differentiation at week 9 and will arrest in prophase I of meiosis I as primordial oocytes from week 13-16. Creation of primordial follicles composed of a primordial oocyte and its surrounding layer of squamous granulosa cells begins at approximately 20 weeks pf [27]. Overall, during human embryonic development, the PGC period lasts for at least 16 weeks, from week 3 to week 20 of embryonic development (6 days in mouse). Throughout this time PGCs undergo three major epigenetic and transcriptional waves, associated with early, late and advanced stages, respectively [25]. During the early to late transition (week 4), the Xi is reactivated, such that late PGCs harbor two active X chromosomes [23]. Intriguingly, the Clark lab demonstrated that *XIST* is expressed in late stage PGCs when the X chromosomes are both active [28]. This finding raises a unique parallel between PGC development and pre-implantation development, which (i) indicates that *XIST* is prevented from silencing in PGCs, similar to the state in naïve pluripotent cells, and (ii) suggests that XCD could be taking place in late-stage PGCs. X-chromosome dosage compensation by XCD in PGCs could be required since PGCs are present for an extended period of time compared to the mouse. Although *XIST* RNA can be detected in human PGCs by RNA-Seq, it is unclear whether *XIST* is expressed from both X chromosomes and whether *XIST*, and potentially XCD are a biological mechanism for regulating X-linked gene expression

between the generations. In any case, the finding of *XIST* expression from active X in PGCs, raises a unique parallel between PGC development and pre-implantation development in humans, where a unique state of the X chromosome and of *XIST* function appear to be at play.

Given the potential similarities between human PGCs and naïve pluripotent cells, one of our goals is to study the X chromosome state in human PGCs. In general, most of our knowledge about mammalian germline development comes from mouse studies, but recent work showed distinct molecular characteristics of human germline formation [29]. In this study, the authors identified SOX17 as a key regulator that determines the fate of human PGC during embryonic development. These discoveries emphasize the importance of understanding the distinct epigenetic regulation of the germ line in human. However, there are limiting factors to study human PGCs as these cells can currently not be cultured when derived from human fetuses and their derivation from hESCs is still not well defined. Isolation of PGCs from embryonic and fetal tissue using fluorescence activated cell sorting (FACS) has transformed our ability to study *in vivo* PGCs, and to enrich for PGC-like cells (PGCLCs) differentiated from human pluripotent stem cells *in vitro*. In addition, molecular analysis of FACS isolated *in vivo* PGCs has led to the creation of important benchmarks for comparing and staging *in vivo* PGCs with *in vitro* differentiated PGCLCs [23], [28]–[32]. So far, it is possible to efficiently differentiate hESCs to early-stage PGC- like cells (PGCLCs) [29], [33], [34], where Xi-reactivation has not yet occurred. Therefore, to study X dosage compensation and how it contributes to germline development the establishment of late-stage PGCLCs from hESCs is required.

To pave the way for understanding the epigenetic regulation of the X chromosome in PGC development, it is important to increase number of tools available to better understand PGC differentiation *in vivo* and advance *in vitro* systems for modeling germline development.

References

[1] H. J. Lee, T. A. Hore, and W. Reik, “Reprogramming the methylome: Erasing memory and creating diversity,” *Cell Stem Cell*, vol. 14, no. 6, pp. 710–719, 2014.

[2] A.-V. Gendrel and E. Heard, “Fifty years of X-inactivation research,” *Development*, vol. 138, no. 23, pp. 5049–5055, 2011.

[3] J. T. Lee and M. S. Bartolomei, “X-inactivation, imprinting, and long noncoding RNAs in health and disease,” *Cell*, vol. 152, no. 6, pp. 1308–1323, 2013.

[4] J. Jiang *et al.*, “Translating dosage compensation to trisomy 21,” *Nature*, vol. 500, no. 7462, pp. 296–300, 2013.

[5] A. Wutz, “Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation,” *Nat. Rev. Genet.*, vol. 12, no. 8, pp. 542–553, 2011.

[6] S. T. da Rocha and E. Heard, “Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation,” *Nat. Struct. Mol. Biol.*, vol. 24, no. 3, pp. 197–204, 2017.

[7] A. Wutz, T. P. Rasmussen, and R. Jaenisch, “Chromosomal silencing and

localization are mediated by different domains of Xist RNA," *Nat. Genet.*, vol. 30, no. 2, pp. 167–174, 2002.

[8] J. M. Engreitz *et al.*, "The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome," *Science (80-.).*, vol. 341, no. 6147, pp. 1237973–1237973, 2013.

[9] Y. Hasegawa, N. Brockdorff, S. Kawano, K. Tsutui, K. Tsutui, and S. Nakagawa, "The matrix protein hnRNP U is required for chromosomal localization of xist RNA," *Dev. Cell*, vol. 19, no. 3, pp. 469–476, 2010.

[10] C. Chu *et al.*, "Systematic discovery of Xist RNA binding proteins," *Cell*, vol. 161, no. 2, pp. 404–416, 2015.

[11] C. A. McHugh *et al.*, "The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3," *Nature*, vol. 521, no. 7551, pp. 232–236, Apr. 2015.

[12] J. Nichols and A. Smith, "Naive and Primed Pluripotent States," *Cell Stem Cell*, vol. 4, no. 6, pp. 487–492, 2009.

[13] I. Okamoto *et al.*, "Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development," *Nature*, vol. 472, no. 7343, pp. 370–374, Apr. 2011.

[14] S. Petropoulos *et al.*, "Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos," *Cell*, Apr. 2016.

[15] A. Sahakyan *et al.*, "Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation," *Cell Stem Cell*, vol. 20, no. 1, pp. 87–101, 2016.

[16] C. Vallot *et al.*, "XACT, a long noncoding transcript coating the active X

chromosome in human pluripotent cells," *Nat. Genet.*, vol. 45, p. 239, Jan. 2013.

[17] M. Casanova *et al.*, "A primate-specific retroviral enhancer wires the XACT lncRNA into the core pluripotency network in humans," *Nat. Commun.*, vol. 10, no. 1, p. 5652, 2019.

[18] C. Vallot *et al.*, "XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development," *Cell Stem Cell*, vol. 20, no. 1, pp. 102–111, 2017.

[19] T. W. Theunissen *et al.*, "Systematic identification of culture conditions for induction and maintenance of naive human pluripotency," *Cell Stem Cell*, vol. 15, no. 4, pp. 471–487, Oct. 2014.

[20] J. Hanna *et al.*, "Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 20, pp. 9222–9227, 2010.

[21] Y. Takashima *et al.*, "Erratum: Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human (Cell (2014) 158 (1254-1269))," *Cell*, vol. 162, no. 2, pp. 452–453, 2015.

[22] K. Huang, T. Maruyama, and G. Fan, "The naive state of human pluripotent stem cells: A synthesis of stem cell and preimplantation embryo transcriptome analyses," *Cell Stem Cell*, vol. 15, no. 4, pp. 410–415, 2014.

[23] F. Guo *et al.*, "The transcriptome and DNA methylome landscapes of human primordial germ cells," *Cell*, vol. 161, no. 6, pp. 1437–1452, 2015.

[24] D. Chen *et al.*, "Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors," *Cell Rep.*, vol. 29, no. 13, pp. 4568-4582.e5, Dec. 2019.

[25] D. Chen, J. J. Gell, Y. Tao, E. Sosa, and A. T. Clark, “Modeling human infertility with pluripotent stem cells,” *Stem Cell Res.*, vol. 21, pp. 187–192, 2017.

[26] W. W. C. Tang, T. Kobayashi, N. Irie, S. Dietmann, and M. A. Surani, “Specification and epigenetic programming of the human germ line,” *Nat. Rev. Genet.*, vol. 17, no. 10, pp. 585–600, 2016.

[27] I. Konishi, S. Fujii, H. Okamura, T. Parmley, and T. Mori, “Development of interstitial cells and ovigerous cords in the human fetal ovary: an ultrastructural study,” *J. Anat.*, vol. 148, pp. 121–135, Oct. 1986.

[28] S. Gkountela *et al.*, “DNA demethylation dynamics in the human prenatal germline,” *Cell*, vol. 161, no. 6, pp. 1425–1436, 2015.

[29] N. Irie *et al.*, “SOX17 is a critical specifier of human primordial germ cell fate,” *Cell*, vol. 160, no. 1–2, pp. 253–268, 2015.

[30] S. Gkountela *et al.*, “The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation,” *Nat. Cell Biol.*, vol. 15, no. 1, pp. 113–122, 2012.

[31] D. Chen *et al.*, “Germline competency of human embryonic stem cells depends on eomesodermin†,” *Biol. Reprod.*, vol. 97, no. 6, pp. 850–861, Oct. 2017.

[32] D. Chen *et al.*, “The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation,” *Cell Rep.*, vol. 25, no. 13, pp. 3591-3602.e5, Dec. 2018.

[33] K. Sasaki *et al.*, “Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells,” *Cell Stem Cell*, vol. 17, no. 2, pp. 178–194, 2015.

[34] F. Sugawa *et al.*, “Human primordial germ cell commitment in vitro associates

with a unique PRDM14 expression profile," *EMBO J.*, vol. 34, no. 8, pp. 1009–1024, 2015.

CHAPTER 2

Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation

ARTICLES
https://doi.org/10.1038/s41556-020-00607-4

nature cell biology

Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation

Tsotne Chitiashvili[1,2,3], Iris Dror[1], Rachel Kim[4], Fei-Man Hsu[2], Rohan Chaudhari[1,2], Erica Pandolfi[2], Di Chen[2], Simone Liebscher[5], Katja Schenke-Layland[5,6,7], Kathrin Plath[1,3,4,8] ✉ and Amander Clark[2,3,4,8] ✉

X-chromosome dosage compensation in female placental mammals is achieved by X-chromosome inactivation (XCI). Human pre-implantation embryos are an exception, in which dosage compensation occurs by X-chromosome dampening (XCD). Here, we examined whether XCD extends to human prenatal germ cells given their similarities to naive pluripotent cells. We found that female human primordial germ cells (hPGCs) display reduced X-linked gene expression before entering meiosis. Moreover, in hPGCs, both X chromosomes are active and express the long non-coding RNAs X active coating transcript (*XACT*) and X inactive specific transcript (*XIST*)—the master regulator of XCI—which are silenced after entry into meiosis. We find that *XACT* is a hPGC marker, describe XCD associated with *XIST* expression in hPGCs and suggest that XCD evolved in humans to regulate X-linked genes in pre-implantation embryos and PGCs. Furthermore, we found a unique mechanism of X-chromosome regulation in human primordial oocytes. Therefore, future studies of human germline development must consider the sexually dimorphic X-chromosome dosage compensation mechanisms in the prenatal germline.

Dosage compensation of genes on the X chromosome is an essential epigenetic event that equalizes the X-linked gene imbalance between males and females[1–8]. In mice, dosage compensation is mediated by XCI[1–7], which is established early during development, first in an imprinted form and, after a brief phase of reactivation in naive pluripotent epiblast cells, through the random form by silencing either the maternal or paternal X chromosome[1–7]. Molecularly, XCI is mediated by the long non-coding RNA (lncRNA) *XIST*, which coats the X chromosome in *cis* to establish and maintain a silencing compartment over the X-chromosome territory[1–7]. Once established, XCI is stably maintained in female somatic cells; however, in female mouse primordial germ cells (mPGCs), the inactive X chromosome (Xi) is reactivated, which coincides with global epigenetic reprogramming[9–11]. Following epigenetic reprogramming, and as mPGCs differentiate into meiotic cells in females or prospermatogonia in males, germ cells display X-chromosome dosage excess or X-chromosome dosage decompensation, respectively[12], highlighting the sexually dimorphic regulation of gene expression on the X chromosome in mouse germ cells.

Despite the conservation of *XIST* and XCI across placental mammals[13], it is now appreciated that there is no evidence for the imprinted form of XCI in human pre-implantation embryos[14–16]. Moreover, in human female pre-implantation blastocysts, *XIST* is uniquely expressed from both X chromosomes[14,15,17] and the expression of X-linked genes on both alleles is transcriptionally reduced but not silenced—a compensated state that is referred to as XCD[14]. Another striking difference in humans is the existence of *XACT*, a primate-specific lncRNA that is expressed from the active X chromosome(s) in both human pre-implantation embryos and pluripotent stem cells that can oppose the function of *XIST*[14,17,18]. The expression status of *XACT* in hPGCs is unknown.

The unique XCD state in human pre-implantation embryos has been puzzling and is speculated to resolve to XCI within 1–2 weeks[19]. It has therefore been postulated that the transient accumulation of *XIST* on both X chromosomes with XCD represents the initiating stages of the normal process of XCI[20]. This interpretation is consistent with the biallelic *XIST* expression that is observed in differentiating female mouse embryonic stem cells (ESCs) during the XCI initiation process[21], yet is inconsistent with other findings that suggest that there is a transient state with two active X chromosomes (Xa) without *XIST* expression in cells transitioning between XCD and XCI[22,23]. As mouse and human PGCs capture many of the epigenetic features of epiblast cells from the pre-implantation embryo[24–26], yet hPGCs stably persist for around two months during development, we hypothesized that female hPGCs could serve as an alternative model to evaluate the possibility that XCD is an independent and stable regulatory mechanism for X-chromosome dosage compensation in humans.

Results

Female hPGCs express *XACT* from two active X chromosomes. As the expression of the X-linked lncRNA *XACT* is associated

[1]Department of Biological Chemistry, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [2]Molecular Cell and Developmental Biology Department, University of California Los Angeles, Los Angeles, CA, USA. [3]Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA. [4]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA, USA. [5]Department of Women's Health, Research Institute for Women's Health, Eberhard Karls University Tübingen, Tübingen, Germany. [6]NMI Natural and Medical Sciences Institute, University Tübingen, Reutlingen, Germany. [7]Department of Medicine/Cardiology, Cardiovascular Research Laboratories, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [8]Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA, USA. ✉e-mail: kplath@mednet.ucla.edu; clarka@ucla.edu

with the unique X-chromosome state of human pre-implantation embryos[14,17], we performed RNA fluorescence in situ hybridization (FISH) analysis of *XACT* in prenatal ovaries together with immunostaining for the germ cell-specific protein deleted in azoospermia like (DAZL), which marks hPGCs[27]. We observed *XACT* transcripts in nearly all female hPGCs (DAZL$^+$ cells) at weeks 7–8 post-fertilization (p.f.), whereas somatic cells (DAZL$^-$) lacked *XACT* expression (Fig. 1a,b). We identified two *XACT* clouds in approximately 60–70% of hPGCs; each of these *XACT* clouds was probably associated with the X chromosome that the RNA was transcribed from (Fig. 1c). Starting from 10 weeks p.f., hPGCs heterogeneously differentiate into meiotic germ cells in females[28], repressing pluripotency-related genes such as *NANOG*[29]. We therefore also detected NANOG to distinguish between NANOG$^+$DAZL$^+$ hPGCs and NANOG$^-$DAZL$^+$ meiotic germ cells. We discovered that *XACT* is still expressed in the majority of NANOG$^+$DAZL$^+$ female hPGCs at week 14 p.f. (Fig. 1d,e). By contrast, the majority of NANOG$^-$DAZL$^+$ meiotic germ cells are *XACT* negative (Fig. 1d,e). Thus, *XACT* is expressed from both X chromosomes in hPGCs, and is not expressed by ovarian somatic cells.

To evaluate whether additional genes encoded on the X chromosome are biallelically expressed in hPGCs, we performed RNA FISH analysis of the X-linked genes *HUWE1* and *ATRX*, detecting their nascent transcription foci in week 8 p.f. ovaries. We found that the majority of hPGCs marked by *XACT* also biallelically expresses *HUWE1* and *ATRX* (Fig. 1f–h). By contrast, the somatic cells express these genes from a single allele due to XCI (Fig. 1f–h). These results are consistent with previous allelic analysis of X-linked gene expression from 53 sorted female germ cells[30], which revealed biallelic expression of X-linked genes (Extended Data Fig. 1a). As a consequence, we conclude that female hPGCs carry two active X chromosomes.

To confirm the active state of the X chromosome in female hPGCs, we assessed trimethylation of Lys 27 of histone H3 (H3K27me3), a chromatin mark that specifically accumulates on the Xi[31,32]. In mice, newly specified female PGCs are enriched in H3K27me3 on the Xi and progressively lose this mark as the cells undergo Xi reactivation[9,11]. Given that 23% of female hPGCs have one *XACT* cloud (Fig. 1c), we evaluated whether female hPGCs show any evidence of a nuclear H3K27me3 accumulation, indicative of an Xi[31,32]. At week 4 p.f., we identified early-stage OCT4$^+$ hPGCs entering the genital ridge epithelium before induction of DAZL expression (Extended Data Fig. 1b). We found that around 10% of OCT4$^+$DAZL$^-$ female hPGCs have an Xi-like nuclear accumulation of H3K27me3, which is reduced to less than 5% at weeks 7 and 12 p.f. once the hPGCs have settled in the ovaries (Extended Data Fig. 1b–e). The lack of an H3K27me3 Xi-like enrichment suggests that the mono-allelic expression of *XACT* observed in a portion of female hPGCs is not linked to residual XCI.

As a naive-like pluripotent expression program is common to both male and female hPGCs[33], we explored whether *XACT* also marks male hPGCs. Using published bulk RNA-seq data of sorted male and female hPGCs[24,25], we detected *XACT* expression in hPGCs but not somatic cells (Fig. 1i). We confirmed this result using RNA FISH on a fetal testis at 13 weeks p.f., detecting *XACT* in 80% of NANOG$^+$ hPGCs (Fig. 1j,k). Thus, as is the case for human pre-implantation embryos, both female and male hPGCs express *XACT*.

***XACT* is expressed in male and female hPGCLCs in vitro.** The expression of *XACT* in male and female hPGCs in vivo suggested that *XACT* may be expressed from the moment of hPGC specification. Given that hPGCs are thought to be specified between weeks 2–3 p.f.[34], it is not possible to study hPGC specification in vivo. We therefore modelled hPGC specification using the differentiation of hPGC-like cells (hPGCLCs)[35] from male (UCLA2 ESC[36]) and female (MZT04-J iPSC[37]) pluripotent stem cells (PSCs), by analysing *XACT* distribution in ITGA6$^+$EPCAM$^+$ hPGCLCs and somatic cells at day 4 of differentiation (Fig. 2a). We discovered that most male hPGCLCs had one *XACT* cloud, whereas most female hPGCLCs had two (Fig. 2b–g). By contrast, >90% of the somatic cells were *XACT* negative (Fig. 2b–g). The detection of biallelic *XACT* in female hPGCLCs is consistent with maintenance of the eroded Xi in differentiating hESCs[23]. Collectively, these data uncover that the lncRNA *XACT* is a new marker of male and female hPGCs in vivo and in vitro.

***XACT* expression is epiblast specific in human embryo attachment cultures.** Given that *XACT* is expressed in NANOG$^+$ hPGCs and hPGCLCs, we next investigated *XACT* distribution during the peri-implantation window of human development using human embryo attachment culture[38,39]. Specifically, we explored the pattern of *XACT* expression in NANOG$^+$ epiblast cells versus NANOG$^-$ trophectoderm (TE) and primitive endoderm (PE) cells. We also detected the lncRNA *XIST* to uncover changes in its distribution. First, we performed RNA FISH for *XACT* and *XIST* in combination with immunostaining for NANOG in female and male human blastocysts (day 6 p.f.). At day 6 p.f., both *XIST* and *XACT* clouds could be identified on all X chromosomes (one in males and two in females) in 75% of female and 55% of male NANOG$^+$ epiblast cells (Fig. 3a,b,d,e). Furthermore, 30% of male epiblast cells also express only *XACT* (Fig. 3e). By contrast, NANOG$^-$ TE and PE cells have more diverse *XIST* and *XACT* expression states—*XACT* is more often repressed (Fig. 3c,f). These observations indicate that *XACT* expression is more strongly associated with the NANOG$^+$ epiblast. Using human embryo attachment culture[38,39] grown to day 12 p.f.—that is, the legal limit for human embryo culture in California—we discovered that the majority of female NANOG$^+$ epiblast cells continued to display *XIST* and *XACT* clouds on both X chromosomes (Fig. 3g,h), whereas male epiblast cells maintained expression of *XACT* but not *XIST* (Fig. 3j,k). In the majority of female NANOG$^-$ cells, *XACT* was repressed and *XIST* was expressed from one X chromosome (Fig. 3i), suggesting that these cells have transitioned to the initiation of XCI. By contrast, in male embryos, both lncRNAs were silenced in 70% of NANOG$^-$ cells (Fig. 3l). Together, these data reveal that *XACT* is expressed by the majority of NANOG$^+$ epiblast cells in the pre-implantation and early peri-implantation stages of male and female human embryo development. By contrast, *XIST* is rapidly repressed in male NANOG$^+$ epiblast between days 6 and 12, highlighting differences in *XIST* regulation in male and female human peri-implantation development.

Female hPGCs dampen expression from the active X chromosomes. As female hPGCs express *XACT* from both X chromosomes, our next goal was to determine whether XCD occurs in hPGCs. We performed single-cell RNA sequencing (scRNA-seq) on single-cell suspensions of five prenatal ovaries and five prenatal testes from weeks 6 to 16 p.f. (Fig. 4a–c). This unbiased approach yielded a total of ~50,000 prenatal gonadal cells, including ~281 male and 1,938 female germ cells that were used to analyse the ratio of X-linked gene expression to autosome gene expression (X/A ratio; a summary of which is provided in Supplementary Table 1).

Calculating the X/A ratio across individual cells per developmental age revealed that female germ cells consistently had a higher X/A ratio compared with male germ cells (Extended Data Fig. 2a,b). This difference between female and male germ cells arises from the significantly lower X-linked gene expression in males (Extended Data Fig. 2c). By contrast, we found no difference in the X/A ratio between male and female gonadal somatic cells (Extended Data Fig. 2a,b). We confirmed these results using a published scRNA-seq dataset of c-KIT and size-selected germ cells[40], in which germline cells are called fetal germ cells (FGCs) (Extended Data Fig. 2d,e).

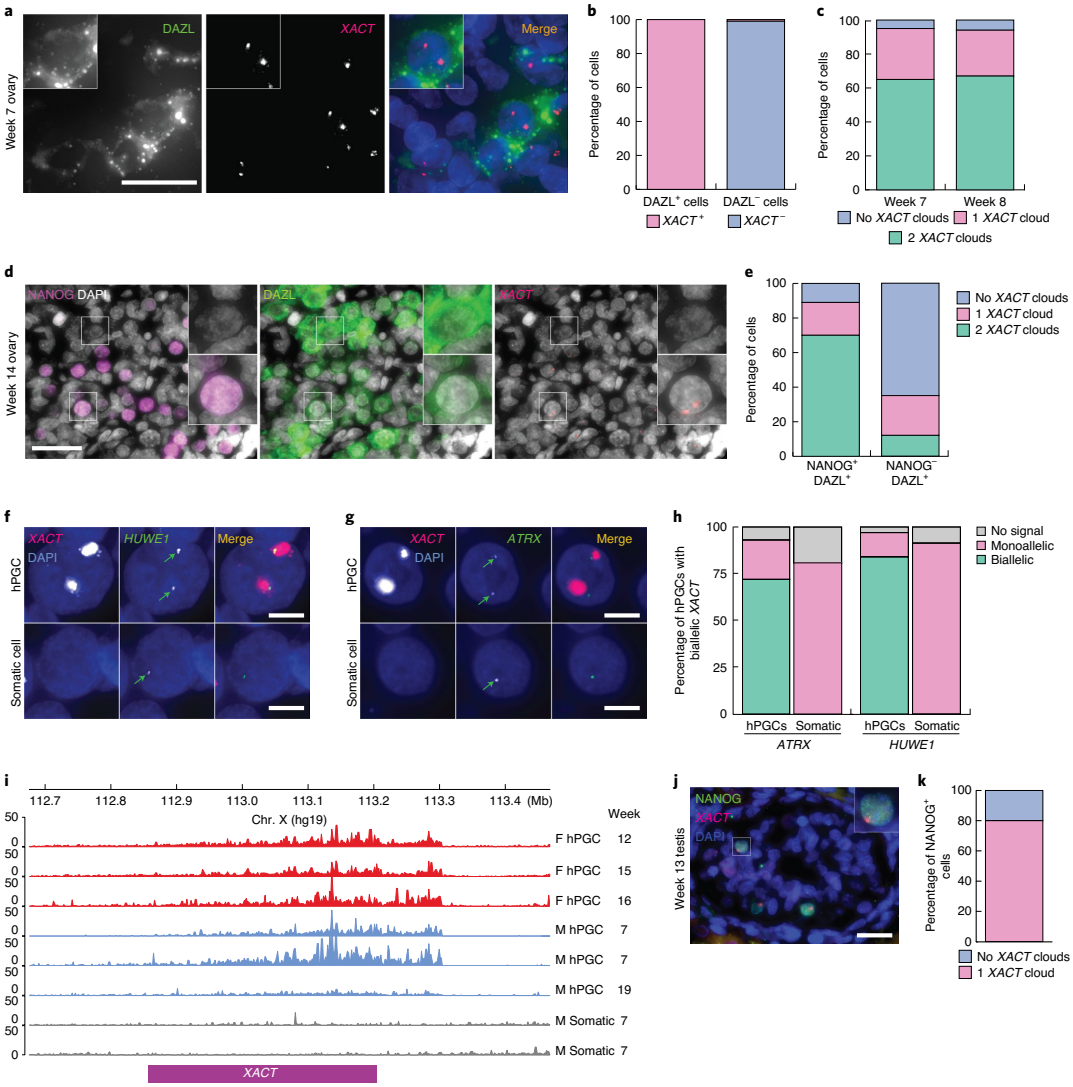


Fig. 1 | Male and female hPGCs express the lncRNA *XACT* and female hPGCs carry two active X chromosomes in vivo. a, Immuno-RNA FISH analysis of DAZL (green) and *XACT* (red) in a week 7 p.f. prenatal ovary with DAPI staining (blue) to detect nuclei. Scale bar, 30 µm. Inset zoom 1.2×. $n=1$ pair of ovaries. **b**, Quantification of cells with *XACT* clouds on the basis of the experiment shown in **a**. $n=100$ cells. **c**, Quantification of the number of *XACT* clouds in DAZL⁺ hPGCs at weeks 7 and 8 p.f. $n=100$ cells per time point. **d**, Immuno-RNA FISH analysis of, NANOG (magenta), DAZL (green), *XACT* (red) and DAPI (grey) in a week 14 p.f. fetal ovary. $n=1$ pair of ovaries. Insets: a NANOG⁻DAZL⁺ hPGC negative for *XACT* (top); and a NANOG⁺DAZL⁺ hPGC with two *XACT* clouds (bottom). Scale bar, 30 µm. Inset zoom 2.5×. **e**, Quantification of the proportion of cells with different *XACT* cloud patterns in the hPGCs (NANOG⁺DAZL⁺) and differentiating hPGCs (NANOG⁻DAZL⁺) from **d**. $n=92$ and $n=95$ cells, respectively, were assessed. **f**,**g**, Representative RNA FISH images for detecting nascent transcripts of the X-linked genes *HUWE1* (**f**) and *ATRX* (**g**), which are both normally subject to XCI, in a week 8 p.f. ovary. hPGCs are marked by *XACT* expression. The experiments were performed twice with similar results. Scale bar, 15 µm. **h**, Signal quantification for **g**. $n=60$ and $n=70$ cells for *ATRX* and *HUWE1*, respectively. **i**, Published bulk RNA-seq reads mapped to the *XACT* genomic locus in female (F) hPGCs (red, isolated using c-KIT^{bright} or using INTa6/EpCAM), male (M) hPGCs (blue, enriched for TNAP/KIT expression) and gonadal somatic cells (grey)[25,52,53]. Chr., chromosome. **j**, Immuno-RNA FISH of *XACT* (red), NANOG (green) and DAPI (blue) in fetal male testes at week 13 p.f. $n=1$ pair of testes. Scale bar, 30 µm. Inset zoom 2.5×. **k**, Quantification of the proportion of cells with one *XACT* cloud in NANOG⁺ male hPGCs from **j**. $n=75$ cells. Source data are available online.

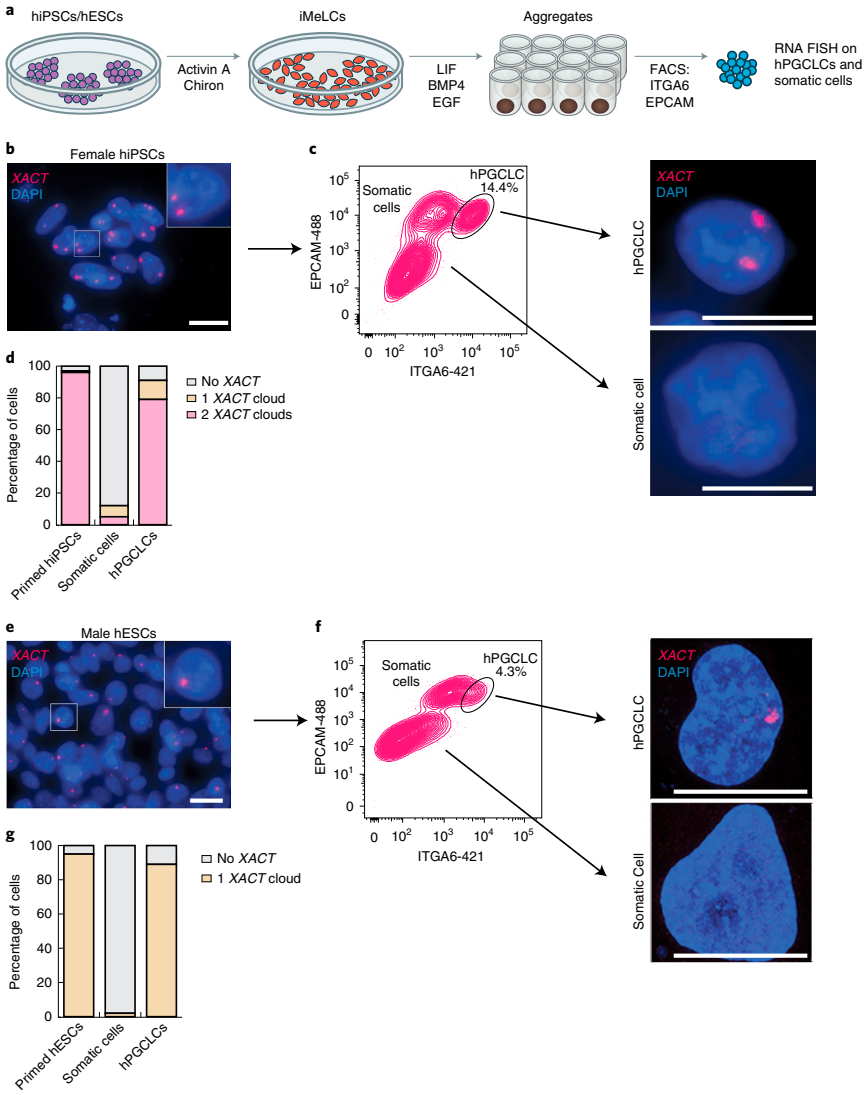


Fig. 2 | The lncRNA *XACT* is restricted to male and female hPGCLCs and is not expressed in somatic cells in vitro. a, Differentiation of hPGCLCs from human induced PSCs (hiPSCs) or hESCs through an incipient mesoderm-like cell (iMeLC) intermediate. PGCLCs and somatic cells within the aggregates were separated at day 4 by fluorescence-activated cell sorting (FACS) using antibodies against EPCAM and ITGA6. **b**, RNA FISH analysis of *XACT* in primed female hiPSCs (MZT04-J) that harbour an Xa and an eroded X chromosome with *XACT* (red) and DAPI (blue) staining. *XACT* clouds were detected from the Xa and eroded X chromosome. Scale bar, 20 µm. Inset zoom 2.5×. The experiments were performed twice with similar results. **c**, Female hiPSCs were differentiated to hPGCLCs and isolated from the aggregates using FACS at day 4 (left). The hPGCLC population is indicated. Right, *XACT* RNA FISH analysis of hPGCLCs and somatic cells. The experiments were performed twice with similar results. Scale bar, 10 µm. **d**, Quantification of the proportion of cells with different numbers of *XACT* clouds in starting hiPSCs ($n = 100$ cells), hPGCLCs ($n = 82$ cells) and somatic cells ($n = 100$ cells) from **c**. **e**, RNA FISH analysis of *XACT* (red) in male hESCs (UCLA2). Scale bar, 20 µm, inset zoom 2.5×, similar to **b**. **f**, FACS (left) and *XACT* RNA FISH (right) analyses as described in **c**, except for with UCLA2 hESCs; the experiments were performed twice with similar results. Scale bar, 10 µm. **g**, Quantification of the proportion of cells with a different number of *XACT* clouds in UCLA2 hESCs ($n = 100$ cells) and derived hPGCLCs ($n = 92$ cells) and somatic cells ($n = 100$ cells pooled from the two experiments). Source data are available online.

Thus, the X/A ratio is higher in female germ cells relative to male germ cells, whereas male and female gonadal somatic cells are equivalent.

As germ cell differentiation into meiotic cells is heterogeneous, we created an unsupervised developmental trajectory[41], which ordered female human germ cells across 11 clusters (Fig. 4d,e). Clusters 0–5 represent hPGCs expressing the transcription factors *NANOG* and *OCT4*, together with the hPGC markers *NANOS3*, *PRDM1* and *SOX17* (Fig. 4d). Starting in cluster 6, we observed downregulation of naive-like pluripotency genes and upregulation of the meiotic licensing gene *STRA8* (ref. [42]) and of genes encoding RNA-binding proteins such as *DAZL* and *DDX4* (also known as *VASA*; clusters 6–7). This was followed by expression of the meiotic prophase I genes *SPO11* and *SYCP1* in clusters 8–9 and, ultimately, in cluster 10, upregulation of primordial oocyte genes including the zona pellucida protein 3 (*ZP3*)[43] (Fig. 4d). Thus, consistent with previous reports[24–26,28,30,40], our data capture the heterogenous differentiation of female hPGCs into meiotic germ cells beginning around week 9–10 p.f., which results in a complex mixture of germ cells including hPGCs, meiotic germ cells and primordial oocytes in a given prenatal ovary (Fig. 4f).

Analysis of the X/A ratio in female germ cells along the developmental trajectory (Fig. 5a) showed that, as hPGCs begin differentiating into meiotic germ cells (cluster 6 onwards), the X/A ratio increases, reaching maximal levels in cluster 9 before precipitously dropping in primordial oocytes (cluster 10). Moreover, the X/A ratio in hPGCs (clusters 0–5) is lower than in gonadal somatic cells (Fig. 5a and Extended Data Fig. 3a–c). The median X/A ratio in hPGCs (clusters 0–5) is significantly lower compared with meiotic germ cells (clusters 6–9) but higher compared with primordial oocytes (cluster 10; Fig. 5b). These changes were largely due to changes in X-linked gene expression (Extended Data Fig. 3d,e). We confirmed these observations in the FGC scRNA-seq data[40]. Similar to our analysis, female FGCs displayed a slightly lower X/A ratio compared with the X/A ratio in gonadal somatic cells in the hPGC state, as well as a subtle, albeit significant, increase in the X/A ratio after expression of *STRA8* (Extended Data Fig. 3f,g). On the basis of these results, we conclude that the dosage of X-linked genes is dampened in female hPGCs and that this dosage compensation is erased as cells enter prophase I of meiosis I.

To evaluate X/A ratios in male prenatal germ cells, we ordered the cells along a developmental trajectory, which divided the male germ cells from week 6–16 p.f. gonads into five clusters (0–4)—the hPGC program corresponded to clusters 0–3 and cluster 4 captured differentiating germ cells (prospermatogonia; Extended Data Fig. 4a,b). In contrast to female germ cells that exhibit X-chromosome dosage excess with exit from the hPGC state, the X/A ratio does not change when male hPGCs differentiate into prospermatogonia (Extended Data Fig. 4c–h), which was confirmed with published datasets[40] (Extended Data Fig. 4i,j). Thus, an increase in X/A ratio as hPGCs differentiate is a female-specific phenomenon that is associated with entrance into prophase I of meiosis I.

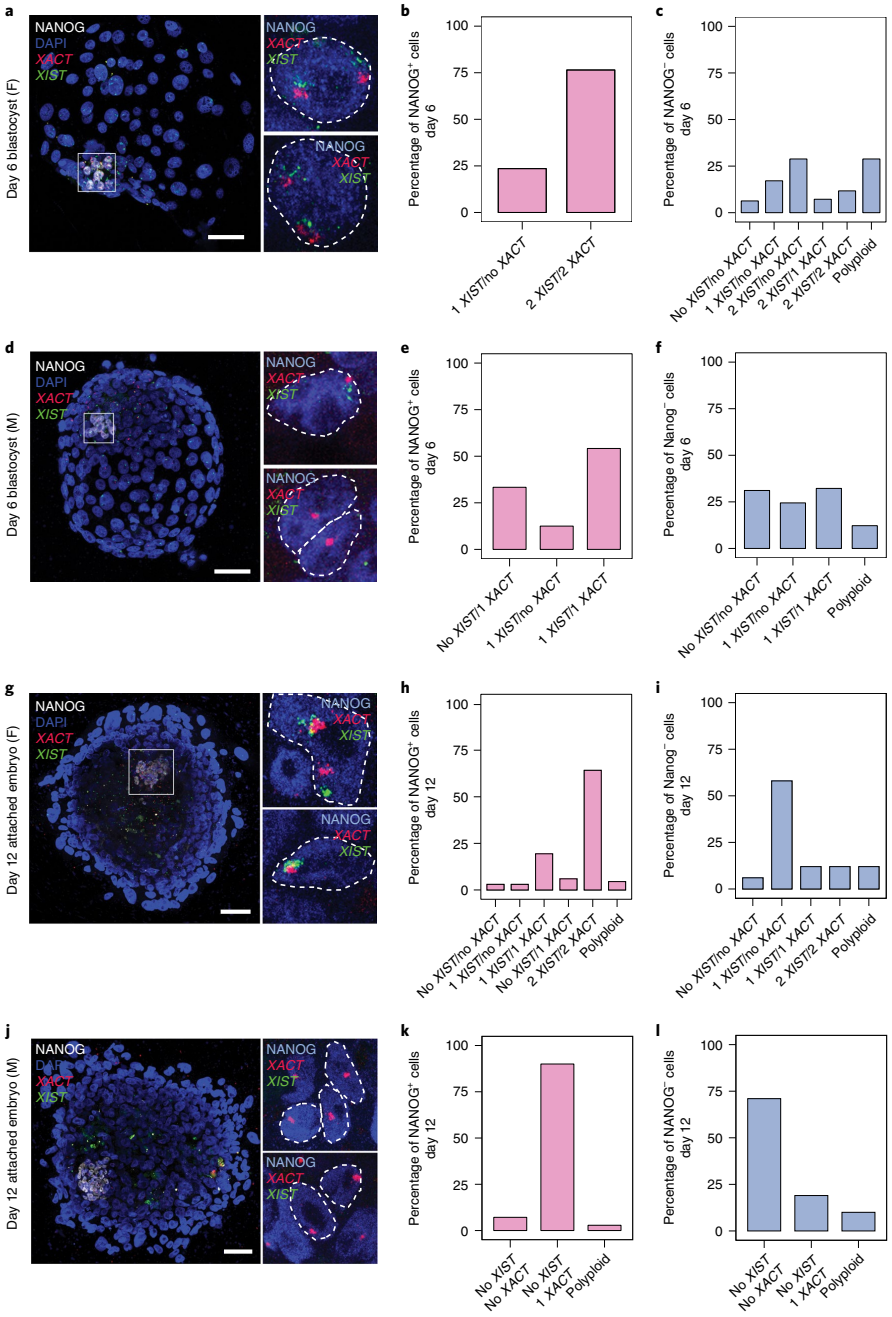
XCD in female hPGCs is associated with *XIST* expression. To evaluate whether XCD in female hPGCs is associated with the expression of *XIST*, as has been shown in female human pre-implantation embryos, we examined *XIST* in individual female germ cells along the developmental trajectory. We discovered that *XIST* expression is significantly higher in hPGCs (clusters 0–5) compared with meiotic germ cells (clusters 6–9) and primordial oocytes (cluster 10; Fig. 5c,d). This result was also validated in the female FGC dataset[40] (Extended Data Fig. 5a). In agreement with the low expression of *XIST* in primordial oocytes, the levels of the RNA are also low in adult oocytes[44] (Extended Data Fig. 5b). Thus, the increase in the X/A ratio from hPGCs to meiotic germ cells is accompanied by a decrease in *XIST* transcript levels (Fig. 5b,d and Extended Data Fig. 5c). Consistent with this result, when analysing hPGCs and meiotic germ cells together (clusters 0–9) on the basis of *XIST* expression, *XIST*$^+$ cells displayed a significantly lower X/A ratio compared with *XIST*$^-$ cells. Primordial oocytes (cluster 10) have an even lower X/A ratio, yet express *XIST* at the lowest level (Fig. 6b,d and Extended Data Fig. 5d).

Next, we evaluated the localization of *XIST* in female hPGCs using RNA FISH from week 7 to week 14 p.f., using *XACT* as a marker of hPGCs. We found that *XIST* is detectable in the vast majority of *XACT*-expressing hPGCs, with diverse patterns of the *XIST* signal (Fig. 6a–c). These include (1) an eroded cloud pattern, whereby *XIST* is restricted over one X chromosome in a pattern that is characteristic of Xi localization, yet less enriched compared with the Xi in somatic cells, combined with a nascent transcription spot of *XIST* on the second X chromosome; (2) a dispersed configuration whereby the *XIST* signal is detected throughout a large portion of the nucleus albeit in the vicinity of both *XACT* clouds, suggesting that *XIST* is expressed from both X chromosomes; (3) one dot in the vicinity of one of the two *XACT* signals, indicating expression from one X chromosome; and (4) two dots representing the nascent transcription sites of *XIST* on both X chromosomes (Fig. 6a). Quantification of the *XIST* expression patterns in female hPGCs with two *XACT* clouds—which enabled the localization of both X chromosomes—revealed that the majority of hPGCs (58%) at week 7 p.f. had a dispersed *XIST* signal (Fig. 6c). Around 6% of cells with biallelic *XACT* expression displayed an eroded *XIST* cloud pattern and 19% and 15%, respectively, exhibited mono- and biallelic nascent *XIST* transcription foci (Fig. 6c). At later stages of embryo development (week 8 and week 14 p.f.), *XACT*$^+$ female hPGCs displayed similar patterns of *XIST* RNA with an increasing fraction of *XIST*$^-$ cells (Fig. 6b,c).

As we observed an Xi-like distribution of *XIST* in 6% of week 7 and 8 p.f. hPGCs with two *XACT* clouds, we tested whether female

Fig. 3 | *XACT* is predominantly expressed in NANOG$^+$ pre- and post-implantation epiblast cells. a,**d**,**g**,**j**, Immuno-RNA FISH analysis of NANOG (white), *XIST* (green) and *XACT* (red) in day 6 p.f. intact female pre-implantation blastocysts (**a**) ($n=2$ blastocysts); day 6 p.f. male pre-implantation blastocysts (**d**) ($n=3$ blastocysts); female embryos cultured to day 12 p.f. using human embryo attachment culture (**g**) ($n=3$ embryos); and male human embryos cultured to day 12 p.f. using human embryo attachment culture (**j**) ($n=2$ embryos). Scale bars, 30 µm. For **a**, **d**, **g** and **j**, insets: NANOG$^+$ (blue) nuclei with *XIST* and *XACT* clouds. Inset zoom 7.5×. **b**, Quantification of the proportion of cells with different numbers of *XIST* and *XACT* clouds in NANOG$^+$ epiblast cells from the female blastocysts shown in **a**; 17 cells were analysed in 2 blastocysts. **c**, Quantification of the RNA pattern in NANOG$^-$ TE and PE cells from the female blastocysts shown in **a**. $n=111$ cells quantified from the 2 blastocysts. **e**, Quantification of the proportion of cells with different numbers of *XIST* and *XACT* clouds as described in **b** from the NANOG$^+$ epiblast cells from the male blastocysts shown in **d**. $n=24$ cells from 3 blastocysts counted. **f**, Quantification of the RNA pattern as described in **c**, except for with NANOG$^-$ TE and PE cells from the male blastocysts shown in **d**; $n=180$ cells from 3 blastocysts were counted. **h**, Quantification of the proportion of cells with different numbers of *XIST* and *XACT* clouds as described in **b** for the female day 12 embryos from **g**; $n=67$ cells from 3 embryos were counted. **i**, Quantification of the RNA pattern as described in **c** for the female day 12 embryos shown in **g**. $n=180$ cells from 3 embryos counted. **k**, Quantification of the proportion of cells with different numbers of *XIST* and *XACT* clouds as described in **b** for the male day 12 embryos from **j**. $n=70$ cells from 2 embryos were assessed. **l**, Quantification of the RNA pattern as described in **c** for the male day 12 embryos in **j**. $n=120$ cells from 2 embryos were quantified. Source data are available online.

hPGCs with one *XACT* cloud represent cells that have undergone XCI. In this case, *XACT* should be expressed from the Xa and *XIST* from the Xi. However, we found that *XIST* is typically expressed from the same X chromosome as *XACT* in cells with monoallelic *XIST*/*XACT* expression (Fig. 6d), providing additional evidence for XCD instead of XCI in female hPGCs.



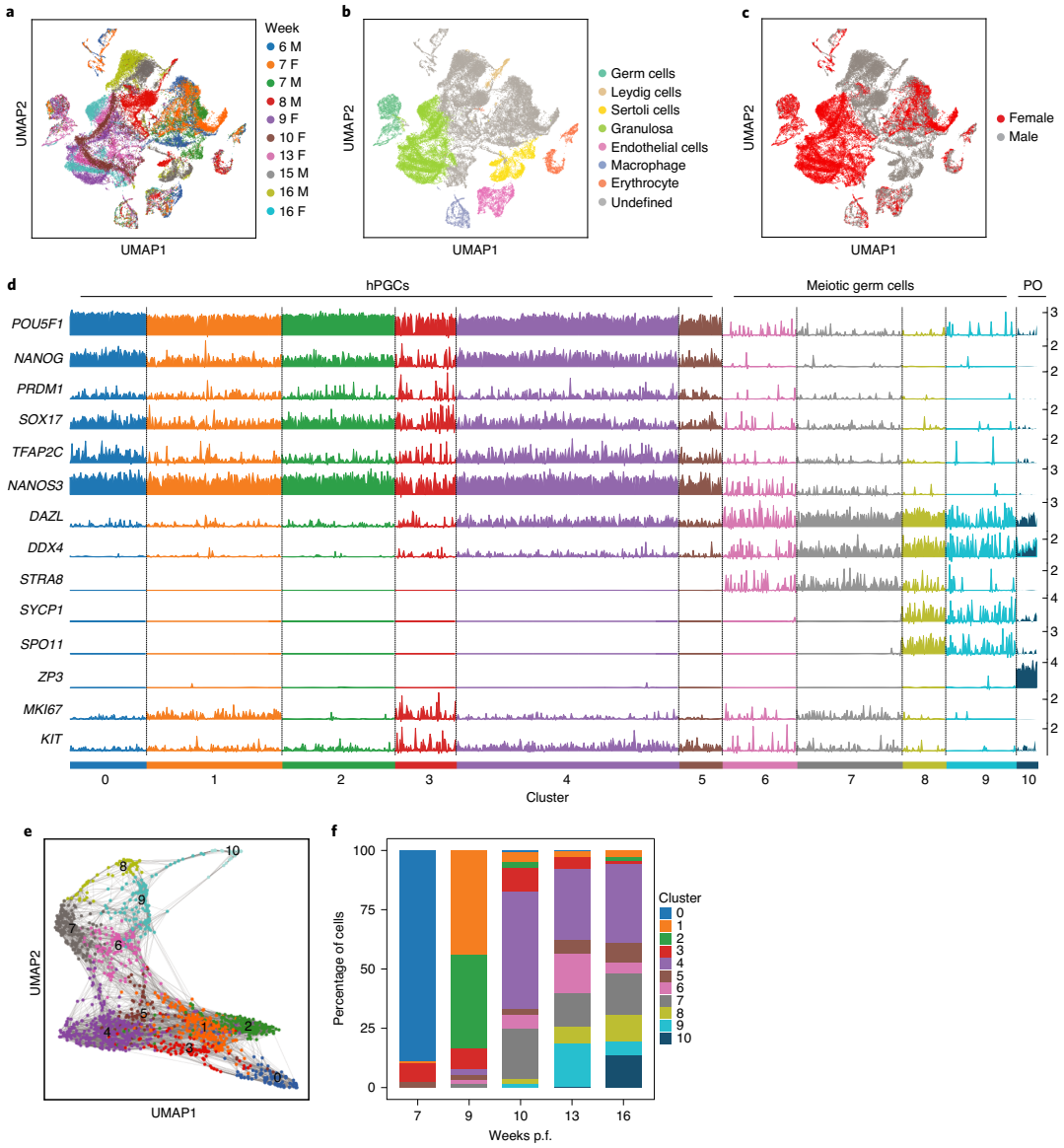


Fig. 4 | Female germ cells undergo dynamic and ordered transcriptional changes between 7-16 weeks p.f. a, The distribution of single-cell data derived from scRNA-seq from five prenatal male and five prenatal female gonads from 6–16 weeks p.f., displayed on a UMAP plot. $n = 49,674$ cells. **b**, Annotation of the gonadal cell types in the map in **a** on the basis of the expression of cell-type-specific markers. **c**, The distribution of male and female cells on the map from **a**. **d**, Ordering of female germ cells along the developmental trajectory from cluster 0 to cluster 10, with classification into hPGCs (clusters 0–5), meiotic germ cells (clusters 6–9) and primordial oocytes (PO, cluster 10) on the basis of diagnostic germ cell marker gene expression. Each cluster contains many individual cells (columns), for which expression of the indicated marker is given (rows). $n = 1,938$ cells from $n = 5$ samples. **e**, Female germ cells displayed on a UMAP plot, labelled by their cluster assignment from **d**. **f**, For each of the five female gonads described in **a**, the proportion of cells in the clusters defined in **d** is given. These data show that the repression of the pluripotency expression program and meiotic licensing (expression of *STRA8*) begins between weeks 9 and 10 p.f., and that all germ cells at week 7 are in the hPGC state. Source data are available online.

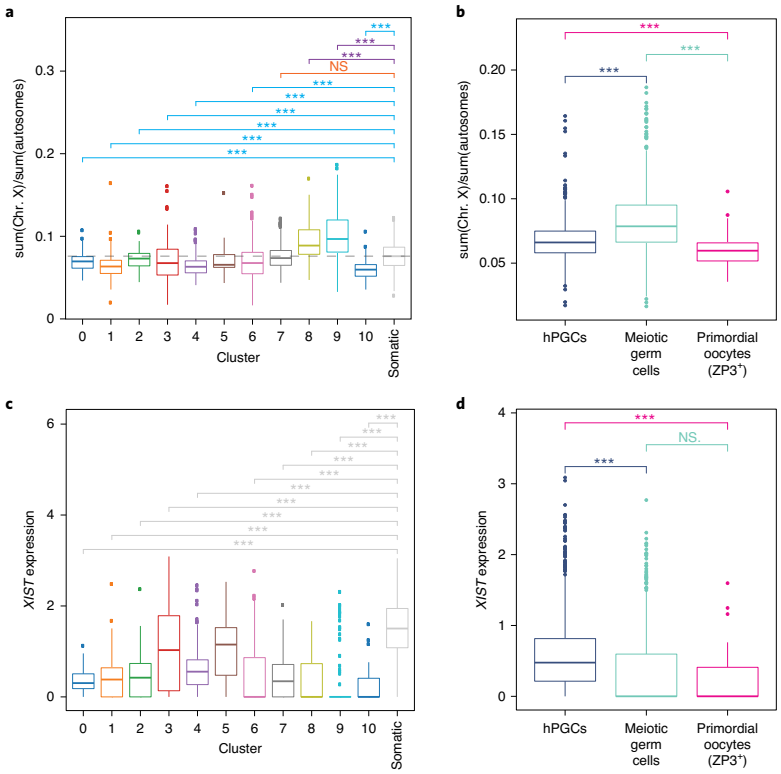


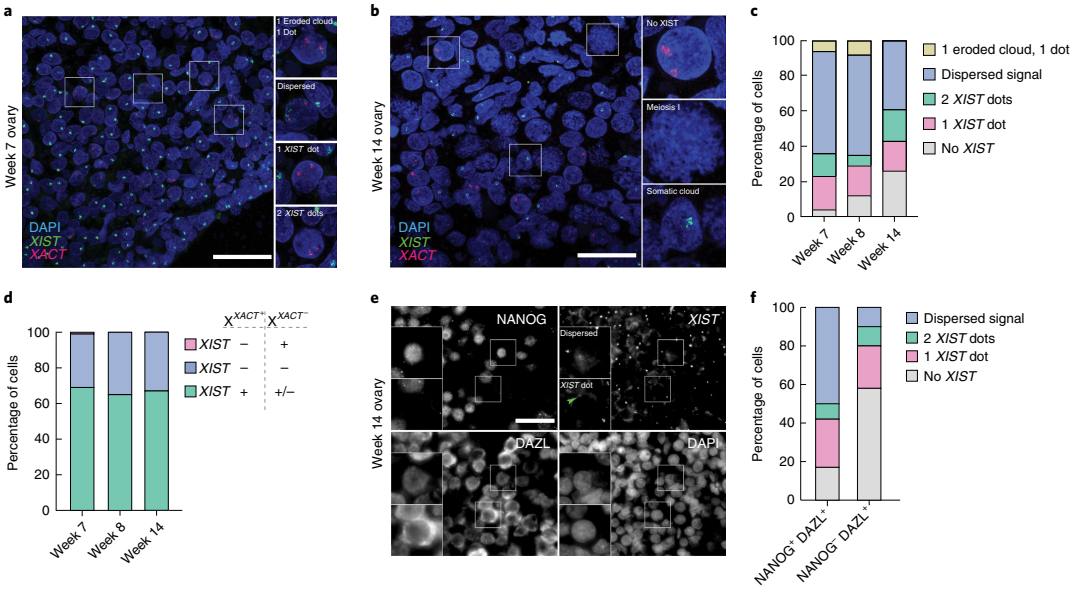
Fig. 5 | *XIST* expression is associated with XCD in female hPGCs. a, The X/A ratios in single female germ cells. $n=1{,}938$ cells for each cluster along the developmental trajectory described in Fig. 4d. Furthermore, the X/A ratio of gonadal somatic cells are given. The dotted line indicates the median level of the X/A ratio in female somatic gonadal cells. Significance testing shows the clusters in which the X/A ratio is significantly lower compared with that in the somatic cell cluster (blue lines); the orange line shows that the difference was not significant; and the purple lines show comparisons in which the X/A ratio was significantly higher than in in the somatic cell cluster. **b**, The X/A ratios of individual female germ cells in the hPGC state, meiotic germ cell state and ZP3$^+$ primordial oocytes, merged on the basis of developmental classification (hPGC state, clusters 0–5; meiotic germ cells, clusters 6–9; and ZP3$^+$ primordial oocytes, cluster 10). **c**, Normalized counts of *XIST* transcripts in individual female germ cells for each cluster along the developmental trajectory described in Fig. 4d and female gonadal somatic cells. These results show that female hPGCs express significantly lower levels of *XIST* and there is an abrupt loss of *XIST*, which coincides with the loss of the pluripotency program in cluster 6. **d**, Normalized *XIST* transcript counts in individual female germ cells, as defined in **b**. For **a**–**d**, statistical significance was assessed using Wilcoxon tests; ns, not significant; $^{***}P<0.001$. $n=1{,}938$ cells analysed across 5 independent experiments.

To confirm that the loss of *XIST* expression was associated with the differentiation of NANOG$^+$DAZL$^+$ hPGCs into NANOG$^-$DAZL$^+$ meiotic germ cells, we evaluated a fetal ovary at 14 weeks p.f. using RNA FISH. We found that *XIST* was detectable in the majority of NANOG$^+$DAZL$^+$ hPGCs, with the dispersed pattern being most prominent. By contrast, the majority of NANOG$^-$DAZL$^+$ germ cells were negative for *XIST* and the subset of cells with *XIST* expression displayed the dot-like distribution (Fig. 6e,f). The quantification suggests that cells transition from the dispersed *XIST* pattern to the two-*XIST*-dot and one-*XIST*-dot patterns before *XIST* is turned off during germ cell differentiation. Furthermore, evaluating *XIST* localization around the X chromosomes in female hPGCs relative to female human pre-implantation embryos revealed a higher degree of *XIST* dispersal in hPGCs (compare Figs. 3 and 6), suggesting that the association between *XIST* and chromatin differs between the two cell types.

In addition to detecting *XIST* expression in female hPGCs, *XIST* transcripts could also be detected in male hPGCs on the basis of scRNA-seq data, albeit at much lower level than in female cells (Extended Data Fig. 5e,f). Furthermore, a small but significant reduction in the X/A ratio in male hPGCs was correlated with a significant increase in *XIST* levels (Extended Data Figs. 4j and 5f). Thus, a subset of male hPGCs can express *XIST* and display slight dampening of X-chromosome dosage. Moreover, female germ cells that lacked *XIST* displayed an increase, albeit not significant, in the *XACT* cloud size compared with cells with *XIST* expression (Extended Data Fig. 5g). Taken together, these data suggest that *XIST* may mediate dampening of X-linked gene expression in hPGCs.

Discussion

Here, by analysing human pre-implantation embryos, human embryo attachment culture, hPGCLC differentiation in vitro and



hPGCs in vivo, we revealed that the lncRNA *XACT* is expressed in pluripotent epiblast cells and hPGCs/hPGCLCs (Fig. 6g and Extended Data Fig. 6). Mechanistically, this result may be explained by the presence of an enhancer that threads *XACT* into the pluripotency network common to these cell types[45]. From this analysis, we describe that *XACT* is a unique marker of hPGCs, and speculate that it could be used to trace hPGCs from the time of lineage specification using RNA FISH. Moreover, our RNA FISH analysis of the X-linked genes *XACT*, *ATRX* and *HUWE1*, together with absence of H3K27me3 accumulation in the nucleus of most female hPGCs demonstrates that female hPGCs harbour two Xa chromosomes from at least week 4 p.f. onwards.

Fig. 6 | *XIST* is repressed as hPGCs enter meiosis. a, RNA FISH analysis of *XIST* (green) and *XACT* (red) with DAPI (blue) identifying nuclei. $n=1$ pair of ovaries (week 7 p.f.). Scale bar, 30 µm. Inset zoom 2.1×. **b**, RNA FISH as described in **a**, except for week 14 p.f. ovaries. $n=1$ pair of ovaries, scale bar, 30 µm. Inset zoom 2.6×. **c**, The proportion of cells with the indicated *XIST* expression patterns in female hPGCs with two *XACT* clouds. For each time point, $n=100$ cells from 1 pair of ovaries. **d**, The proportion of cells with the indicated *XIST* expression patterns, similar to as described in **c** except in female hPGCs with a single *XACT* cloud. Cells with a dispersed *XIST* signal for which it was not clear whether *XIST* was expressed from one or both alleles were included in the green category. Total of $n=42$ cells from 3 ovaries. **e**, Immuno-RNA FISH analysis of NANOG, DAZL and *XIST* RNA in week 14 p.f. ovaries (1 pair of ovaries). Scale bar, 30 µm. **f**, The proportion of cells with the indicated *XIST* expression patterns similar to in **c** in female hPGCs ($NANOG^+DAZL^+$) and meiotic germ cells ($NANOG^-DAZL^+$) from **e**. $n=104$ and $n=98$ cells, respectively from 1 sample. **g**, Female $NANOG^+$ pre-implantation epiblast (pre-Epi) and post-implantation epiblast (post-Epi) cells predominantly express *XACT* and *XIST* from both X chromosomes. Although genes on both X chromosomes in pre-implantation epiblast cells are dampened due to XCD (Xd)[14], it is unclear whether XCD also occurs in female post-implantation epiblast cells. However, expression of both *XACT* and *XIST* from both X chromosomes indicates that XCI has not yet occurred. $NANOG^+$ female hPGCs, similar to pre-implantation epiblast cells, exhibit XCD with expression of *XIST* and *XACT* from both alleles. Notably, *XIST* is more dispersed in the nucleus of hPGCs (shown by the glow around the X chromosome) compared with pre/post-implantation epiblast cells. After advancement to meiosis, female germ cells silence both *XIST* and *XACT* and upregulate X-linked gene expression, transitioning to the Xa state. Next, in primordial oocytes, the X/A ratio is reduced to lower than in female hPGCs without expression of *XIST*. We call this *XIST*-independent repression of the X/A ratio oocyte-specific X-chromosome regulation (X^O). Male hPGCs harbour an Xa but exhibit a lower X/A ratio compared with male somatic cells, and this state is retained after differentiation into prospermatogonia. We refer to this state as Xa^*. Gonadal somatic cells display XCI at each stage analysed.

Although our imaging approaches demonstrated the presence of two active X chromosomes in female hPGCs, the scRNA-seq data revealed that the X/A ratio is reduced in female hPGCs compared with female meiotic germ cells (Fig. 6g). These results indicate that X-linked dosage compensation in female hPGCs is regulated by the XCD mechanism, similar to female naive human pluripotent stem cells[22] and female human pre-implantation embryos[14]. Although XCD is a transient state in pre-implantation embryos, in the case of hPGCs, we show that XCD is not a transitional state into XCI, but rather a stable state that lasts for at least 6 weeks until the point of meiotic initiation (Fig. 6g). Similar to female mouse PGCs[12], we also show that the X/A ratio excess occurs as female hPGCs initiate meiosis. After prophase I of meiosis I, the X/A ratio quickly declines coincident with primordial oocyte formation indicating a third unique state of X-chromosome dosage compensation, which we have called oocyte-specific X-chromosome regulation (X^O; Fig. 6g). Interestingly, the loss of XCD in the female germline after meiotic entry is linked to the silencing of *XIST* (Fig. 6g), suggesting that *XIST* is the mediator of XCD. By contrast, the further decline in the X/A ratio in primordial oocytes occurs in the absence of *XIST* expression. It remains unclear whether this regulation is achieved by XCD or other mechanisms.

Given that male hPGCs have a reduced X/A ratio compared with female and male somatic cells and female hPGCs, we refer to the active X chromosome in male prenatal germ cells as Xa^* (Fig. 6g). The lower X/A ratio in male hPGCs relative to female hPGCs could be due to inefficient dampening of X-linked gene expression from both X chromosomes by *XIST* in female hPGCs, such that the levels in female hPGCs are higher than males. In support of this hypothesis, *XIST* is highly dispersed in female hPGCs, which may lead to less efficient XCD compared with the more cloud-like distribution of *XIST* reported for the pre-implantation embryo[14,15]. By contrast, a higher X/A ratio in female and male somatic cells could be explained by upregulation of single Xa in somatic cells[12,46,47], which may not occur in female or male hPGCs. Combined with XCD occurring on the X chromosomes in female hPGCs, this alternative explanation would explain the lower X-linked gene expression in female hPGCs compared with somatic cells.

As germline specification in humans takes days compared with hours in mice, and hPGC development is a much lengthier process than in mice, it is conceivable that the maintenance of X-chromosome dosage compensation in germ cells between the two species is different. It is probable that a primate-specific X-chromosome regulation mechanism might be necessary to compensate dosage of X-linked genes in the human embryo during PGC specification and the first trimester of pregnancy. However, how *XIST* and *XACT* contribute to X-chromosome gene regulation in the developing human germline will need to be studied mechanistically. Achieving this goal will require new in vitro cell models of hPGC development that have the ability to reliably promote the differentiation of hPGCLCs into meiotic germ cells and primordial oocytes combined with functional approaches[48].

X-chromosome dosage regulation might be extremely important for patients with Turner (XO) and Klinefelter (XXY) syndrome, who have infertility associated with loss of germline cells[49,50]. Although germline development in fetuses diagnosed with Turner syndrome is morphologically normal, oocyte loss occurs within the first few months after birth[51]. Potentially, meiosis does not occur correctly in patients with Turner syndrome due to diminished levels of critical X-linked genes in differentiating XO hPGCs compared with XX hPGCs. Thus, upregulation of X-linked gene expression with entrance into meiosis may be necessary for the formation of mature oocytes.

In summary, with the demonstration of XCD, our research draws parallels between the X-chromosome state of human epiblast and hPGCs. Our study sheds light on mechanisms that regulate X-linked gene expression in hPGCs before meiosis, and reveals a unique X-chromosome state in oocytes, which could potentially be important for oocyte formation and zygote development downstream.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-020-00607-4.

Received: 24 February 2020; Accepted: 27 October 2020;
Published online: 30 November 2020

References

1. 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).
2. Payer, B. & Lee, J. T. X chromosome dosage compensation: how mammals keep the balance. *Annu. Rev. Genet.* **42**, 733–772 (2008).
3. Wutz, A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* **12**, 542–553 (2011).
4. Gendrel, A.-V. & Heard, E. Fifty years of X-inactivation research. *Development* **138**, 5049–5055 (2011).
5. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).

6. Robert Finestra, T. & Gribnau, J. X chromosome inactivation: silencing, topology and reactivation. *Curr. Opin. Cell Biol.* **46**, 54–61 (2017).
7. Brockdorff, N. Localized accumulation of Xist RNA in X chromosome inactivation. *Open Biol.* **9**, 190213 (2020).
8. Sahakyan, A., Plath, K. & Rougeulle, C. Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. *Philos. Trans. R. Soc. Lond. B* **372**, 20160363 (2017).
9. Chuva de Sousa Lopes, S. M. et al. X chromosome activity in mouse XX primordial germ cells. *PLoS Genet.* **4**, e30 (2008).
10. Sugimoto, M. & Abe, K. X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLoS Genet.* **3**, 1309–1317 (2007).
11. 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).
12. Sangrithi, M. N. et al. Non-Canonical and sexually dimorphic X dosage compensation states in the mouse and human germline. *Dev. Cell* **40**, 289–301 (2017).
13. Nesterova, T. B. et al. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* **11**, 833–849 (2001).
14. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* https://doi.org/10.1016/j.cell.2016.03.023 (2016).
15. Okamoto, I. et al. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
16. 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).
17. 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).
18. Vallot, C. et al. *XACT*, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.* **45**, 239–241 (2013).
19. Patrat, C., Ouimette, J.-F. & Rougeulle, C. X chromosome inactivation in human development. *Development* **147**, dev183095 (2020).
20. 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**, 10794 (2017).
21. Sousa, E. J. et al. Exit from naive pluripotency induces a transient X chromosome Inactivation-like state in males. *Cell Stem Cell* **22**, 919–928 (2018).
22. Sahakyan, A. et al. Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87–101 (2016).
23. Patel, S. et al. Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep.* **18**, 54–67 (2016).
24. Tang, W. W. C. et al. A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161**, 1453–1467 (2015).
25. Gkountela, S. et al. DNA demethylation dynamics in the human prenatal germline. *Cell* **161**, 1425–1436 (2015).
26. Guo, F. et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161**, 1437–1452 (2015).
27. Anderson, R. A., Fulton, N., Cowan, G., Coutts, S. & Saunders, P. T. K. Conserved and divergent patterns of expression of DAZL, VASA and OCT4 in the germ cells of the human fetal ovary and testis. *BMC Dev. Biol.* **7**, 136 (2007).
28. Gkountela, S. et al. The ontogeny of cKIT$^+$ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat. Cell Biol.* **15**, 113–122 (2012).
29. Perrett, R. M. et al. The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture. *Biol. Reprod.* **78**, 852–858 (2008).
30. Vértesy, Á. et al. Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells. *Nat. Commun.* **9**, 1873 (2018).
31. Plath, K. et al. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
32. Silva, J. et al. Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).
33. Chen, D. et al. The TFAP2C-regulated *OCT4* naive enhancer is involved in human germline formation. *Cell Rep.* **25**, 3591–3602 (2018).
34. Kobayashi, T. & Surani, M. A. On the origin of the human germline. *Development* **145**, dev150433 (2018).
35. Sasaki, K. et al. Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell* **17**, 178–194 (2015).
36. Diaz Perez, S. V. et al. 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 (2011).
37. Pandolfi, E. C. et al. Generation of three human induced pluripotent stem cell sublines (MZT04D, MZT04J, MZT04C) for reproductive science research. *Stem Cell Res.* **40**, 101576 (2019).
38. Deglincerti, A. et al. Self-organization of the in vitro attached human embryo. *Nature* **533**, 251–254 (2016).
39. Shahbazi, M. N. et al. Self-organization of the human embryo in the absence of maternal tissues. *Nat. Cell Biol.* **18**, 700–708 (2016).
40. Li, L. et al. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* 858–873 (2017); https://doi.org/10.1016/j.stem.2017.03.007
41. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
42. Anderson, E. L. et al. Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proc. Natl Acad. Sci. USA* **105**, 14976–14980 (2008).
43. Törmälä, R. M. et al. Zona pellucida components are present in human fetal ovary before follicle formation. *Mol. Cell. Endocrinol.* **289**, 10–15 (2008).
44. Zhang, Y. et al. Transcriptome landscape of human folliculogenesis reveals oocyte and granulosa cell interactions. *Mol. Cell* **72**, 1021–1034 (2018).
45. Casanova, M. et al. A primate-specific retroviral enhancer wires the XACT lncRNA into the core pluripotency network in humans. *Nat. Commun.* **10**, 5652 (2019).
46. Larsson, A. J. M., Coucoravas, C., Sandberg, R. & Reinius, B. X-chromosome upregulation is driven by increased burst frequency. *Nat. Struct. Mol. Biol.* **26**, 963–969 (2019).
47. Deng, X. et al. Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat. Genet.* **43**, 1179–1185 (2011).
48. Yamashiro, C. et al. Generation of human oogonia from induced pluripotent stem cells in vitro. *Science* **362**, 356–360 (2018).
49. Folsom, L. J. & Fuqua, J. S. Reproductive issues in women with Turner syndrome. *Endocrinol. Metab. Clin. North Am.* **44**, 723–737 (2015).
50. Franik, S. et al. Klinefelter syndrome and fertility—impact of X-chromosomal inheritance on spermatogenesis. *Andrologia* **50**, e13004 (2018).
51. Reynaud, K. et al. Number of ovarian follicles in human fetuses with the 45,X karyotype. *Fertil. Steril.* **81**, 1112–1119 (2004).
52. Irie, N. et al. SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253–268 (2015).
53. Chen, D. et al. Germline competency of human embryonic stem cells depends on eomesodermin. *Biol. Reprod.* **97**, 850–861 (2017).

Methods

Human fetal tissues. Prenatal gonads (4–16 weeks p.f.) were obtained from either the University of Washington Birth Defects Research Laboratory (BDRL) or the University of Tubingen. At the BDRL, prenatal gonads were obtained with regulatory oversight from the University of Washington IRB approved Human Participants protocol, combined with a Certificate of Confidentiality from the Federal Government. BDRL collected the fetal testes and ovaries and shipped them overnight in HBSS with an ice pack for immediate processing at UCLA. Prenatal samples from the University of Tübingen were delivered to UCLA 24–48 h after the procedure. The research project was also approved by the research ethics committee of the University of Tübingen (IRB, 584/2018BO2 and 634/2017BO1). All human fetal tissue used here was obtained following informed consent. The donated human fetal tissue sent to UCLA did not carry any personal identifiers. No payments were made to donors and the donors knowingly and willingly consented to provide research materials without restrictions for research and for use without identifiers. Developmental age was documented by the BDRL and the University of Tübingen as days p.f. using a combination of prenatal intakes and Carnegie staging. A total of 16 fetal samples was used for this study.

Human pre-implantation embryos. The use of human embryos in this research project followed California State law and was reviewed by the Institutional Review Board (IRB) and the human embryonic stem cell research oversight committee (ESCRO) at UCLA. The ESCRO committee at UCLA approves human pluripotent stem cell and human embryo work at UCLA according to 2016 ISSCR guidelines. Together, these committees approve the process of informed consent, and experiments using human embryos for research purposes on an annual basis. Patients were not paid for participating, and all of the donors were informed that the embryos would be destroyed as part of the research study. All research in this study using human embryos complied with the principles that are laid out in the International Society for Stem Cell Research. Frozen human blastocysts at days 6–7 p.f. were used in this study and thawed using the Vit Kit-Thaw (Irvine Scientific) according to manufacturer's protocol. After thawing, embryos were cultured overnight in 5% O_2, 6% CO_2 at 37 °C, and the zona pellucida was removed with Tyrode's acidified solution (Irvine Scientific). A total of 28 human blastocysts were used here. The sex of the blastocysts was determined by cloud counts of lncRNA *XIST* and *XACT* expression from a single (male) or both X chromosomes (female).

Tissue processing for scRNA-seq. Fetal tissues were processed 24–48 h after termination. On arrival, tissues were gently washed with PBS and placed in dissociation buffer containing of collagenase IV 10 mg ml^{-1} (Life Technologies, 17104-019), dispase II 250 µg ml^{-1} (Life Technologies, 17105041), DNase I 1:1,000 (Sigma-Aldrich, 4716728001), 10% fetal bovine serum (Life Technologies, 10099141) in 1× PBS. Tissues were dissociated for 15 min at 37 °C. After every 5 min, the tissues were pipetted against the bottom of Eppendorf tube. Cells were subsequently centrifuged for 5 min at 500*g*, resuspended in 1× PBS with 0.04% BSA, strained through a 40 µm strainer and counted using an automated cell counter (Thermo Fisher Scientific, Countess II). The cell concentration was adjusted to 800–1,200 cells per µl and immediately used for scRNA-seq.

scRNA-seq library preparation and sequencing. scRNA-seq libraries were generated using the 10x Genomics Chromium instrument and Chromium Single Cell 3′ Reagent Kit v2. Each individual library was designed to target 6,000 cells. Libraries were generated according to the manufacturer's instructions and library fragment size distribution was determined using a BioAnalyzer instrument. Libraries were pooled together and sequenced using an Illumina Novaseq 6000 platform, at an average depth of 400–420 million reads per sample.

scRNA-seq data analysis. scRNA-seq reads were aligned to the human hg38 genome assembly using 10x Genomics Cell ranger v.2.2. Expression matrixes generated by Cell Ranger were imported into Scanpy[54] for downstream analysis. First, all of the libraries were merged, and cells were filtered in the same manner. All of the genes that were expressed in less than five cells were discarded and cells with less than 250 detected genes were filtered out. The unique molecular identifier (UMI) counts were then normalized for each cell by the total expression, multiplied by 10,000 and log-transformed. Using Scanpy's default method, highly variable genes were identified, and data were scaled to regress out variation from UMI counts and mitochondrial genes. Cells were clustered using the Louvain algorithm[55] and the UMAP package was used to visualize cells in a two-dimensional plot[54]. Germ cell clusters were identified by expression of germ-cell-specific markers, such as *NANOS3*, *DAZL*, *DDX4* and *SYCP1*. Gonadal somatic cells were annotated by previously published literature[40]. The female and male germline trajectories were created by partition-based graph abstraction[41]. The dataset of Li at al.[40] was analysed through the same pipeline as described above. Gene expression matrixes of female and male germ cells were exported from Scanpy and X/A ratio per cell were calculated using a custom R script.

Tissue processing and cryo-sectioning. On arrival, fetal tissues were gently washed with 1× PBS and fixed with freshly made 4% paraformaldehyde in 1× PBS for 3–4 h on a rotator at room temperature. Tissues were washed with 1× PBS three times for 5 min and moved through increasing concentrations of sucrose—10% sucrose for 1 h, 20% for 1 h and 30% overnight at 4 °C. Tissues were next embedded in O.C.T. (Tissue-Tek) and sections (7 µm) were cut. Sections and tissue blocks were kept at −80 °C.

hESC culture. The hESC and hiPSC lines used in this study include UCLA2 (46, XY)[36] and MZTO4 iPSC (46, XX)[37]. hESCs and hiPSCs were cultured on mitomycin-C-inactivated mouse embryonic fibroblasts (MEFs) in hESC medium, which is composed of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100 mM L-glutamine (GIBCO, 25030-081), 1× MEM non-essential amino acids (NEAA) (GIBCO, 11140-050), 55 mM 2-mercaptoethanol (GIBCO, 21985-023), 10 ng ml^{-1} recombinant human FGF basic (R&D systems, 233-FB), 1× penicillin–streptomycin (GIBCO, 15140-122) and 50 ng ml^{-1} primocin (InvivoGen, ant-pm-2) in DMEM/F12 medium (GIBCO, 11330-032). The hESCs and iPSCs were split every 7–8 d using collagenase type IV (GIBCO, 17104-019). The hESC line used in this study is registered with the National Institute of Health Human Embryonic Stem Cell Registry and available for research use with NIH funds. Mycoplasma tests (Lonza, LT07-418) were performed every month for all cell lines used in this study. All of the experiments were approved by the UCLA ESCRO Committee.

hPGCLC differentiation. hPGCLCs were induced from primed hESCs and hiPSCs as described previously[53] with stem cell factor omitted from the differentiation medium and starting with human pluripotent stem cells grown on MEFs. In brief, hESCs and hiPSCs were dissociated into single cells with 0.05% trypsin-EDTA (GIBCO, 25300-054) and plated onto human-plasma-derived fibronectin-coated (Invitrogen, 33016-015) 12-well plates at a density of 200,000 cells per well in 2 ml per well of iMeLC medium (15% KSR (GIBCO, 10828-028), 1× NEAA (GIBCO, 11140-050), 0.1 mM 2-mercaptoethanol (GIBCO, 21985-023), 1× penicillin–streptomycin–glutamine (GIBCO, 10378-016), 1 mM sodium pyruvate (GIBCO, 11360-070), 50 ng ml^{-1} activin A (Peprotech, AF-120-14E), 3 mM CHIR99021 (Stemgent, 04-0004), 10 mM of ROCKi (Y27632, Stemgent, 04-0012-10) and 50 ng ml^{-1} primocin in Glasgow's MEM (GMEM) (GIBCO, 11710-035)). After 24 h, iMeLCs were dissociated into single cells with 0.05% trypsin-EDTA and plated into ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at a density of 3,000 cells per well in 200 ml per well of hPGCLC medium, which is composed of 15% KSR (GIBCO, 10828-028), 1× NEAA (GIBCO, 11140-050), 0.1 mM 2-mercaptoethanol (GIBCO, 21985-023), 1× penicillin–streptomycin-–glutamine (GIBCO, 10378-016), 1 mM sodium pyruvate (GIBCO, 11360-070), 10 ng per ml^{-1} human LIF (Millipore, LIF1005), 200 ng ml^{-1} human BMP4 (R&D systems, 314-BP), 50 ng ml^{-1} human EGF (R&D systems, 236-EG), 10 mM of ROCKi (Y27632, Stemgent, 04-0012-10) and 50 ng ml^{-1} primocin in GMEM (GIBCO, 11710-035).

hPGCLC sorting and preparation for RNA FISH. hPGCLC aggregates were dissociated 0.05% trypsin-EDTA (GIBCO, 25300-054) for 10 min at 37 °C. The dissociated cells were stained with conjugated antibodies, washed with FACS buffer (1% BSA in PBS) and resuspended in FACS buffer with 7-AAD (BD PharMingen, 559925) as viability dye. The conjugated antibodies used in this study include ITGA6 conjugated with BV421 (BioLegend, 313624, 1:60), EPCAM conjugated with 488 (BioLegend, 324210, 1:60). The single-cell suspension was sorted for further experiments using BD FACSAria FACS machine. FACS data were analysed using FlowJo v.10. Double-positive cells for ITGA6 and EPCAM (hPGCLCs) and negative cells (non-hPGCLCs) were collected in hPGCLC medium and plated on human-plasma-derived fibronectin-coated coverslips overnight. The next morning, RNA FISH was performed using the coverslips.

Immunofluorescence. Slides of paraffin-embedded sections were deparaffinized by successive treatment with xylene and 100%, 95%, 70% and 50% ethanol. Antigen retrieval was performed by incubation with 10 mM Tris pH 9.0, 1 mM EDTA, 0.05% Tween-20 at 95 °C for 40 min. The slides were cooled and washed with 1× PBS and 1× TBS (PBS + 0.2% Tween-20). Cryosections and blastocysts attached to Ibidi chambers were fixed with 4% PFA for 10 min and then washed with 1× PBS. Paraffin-embedded sections, cryosections and blastocysts were treated similarly. The samples were permeabilized with 0.5% Triton X-100 in 1× PBS, then washed with 1× TBS and blocked with 1% BSA in 1× TBS. Primary antibody incubation was conducted with 1% BSA for 1 h at room temperature. Samples were again washed with 3× TBS-Tween-20 and incubated with fluorescent secondary antibodies at 1:200 for 45 min, then washed and counterstained with DAPI for 5 min and mounted using Vectashield. A list of the primary antibodies used for immunofluorescence in this study is provided in Supplementary Table 1 under the antibody list tab. The secondary antibodies used in this study were all obtained from Life technologies and were used at 1:400 dilution. Images were taken using a LSM 880 Confocal Instrument (Zeiss) or Zeiss Axio Imager M1. For image processing and analysis, Fiji (ImageJ) was used. For signal quantification, images were converted into 8-bit images and then analysed using profile plot tool. Intensity values were exported as a CSV file and then R Studio and the ggplot2 package was used for plotting.

RNA FISH. After sorting, hPGCLCs and non-hPGCLCs were attached to fibronectin-coated 18 mm circular glass coverslips (Thermo Fisher Scientific, 12-545-100) overnight. The next morning, the coverslips were washed with DPBS, fixed with 4% formaldehyde for 10 min, permeabilized with cold (4 °C) 0.5% Triton X-100 in DPBS for 10 min and serially dehydrated with cold (4 °C) 70–100% ethanol. Coverslips were air dried and hybridized with labelled DNA probes in a humidified chamber at 37 °C overnight, washed for three 5 min intervals with 50% formamide in 2× SSC, 2× SSC, then 1× SSC at 42 °C, counterstained with DAPI and mounted with Vectashield (Vector Labs, H-1000). Double-stranded DNA probes were generated from full-length cDNA constructs or bacterial artificial chromosomes (BACs) as described previously[56]. The following BACs were used: *XIST* (RP11-13M9), *XACT* (RP11-35D3), *ATRX* (RP11-1145J4) and *HUWE1* (RP11-975N19). Every new batch of probes was tested on normal human dermal fibroblasts before it was used in experiments.

Immuno-RNA FISH. Immuno-RNA FISH on cryosections was performed as described previously described[56]. In brief, immunostaining was performed first on cryosections and blastocysts as described above. Samples were post-fixed with 4% PFA after immunofluorescence staining, and RNA FISH was performed after post-fixation as described in RNA FISH section above.

Immuno-RNA FISH analysis of human blastocysts was performed as described previously[23] with the following modifications: embryos at day 5 and day 6 p.f. were thawed for these experiments and cultured for 24 h before staining. First, zona pellucida was removed with Tyrode's acid and blastocysts were washed with 6 mg ml^{-1} BSA (Sigma-Aldrich) in RNase–DNase-free PBS. Blastocysts were then individually transferred to Ibidi chambers (Ibidi, 80827), which were coated with polylysine (Sigma-Aldrich, P4832-50ML). Fluid was aspirated until dry and the blastocysts were fixed with ice-cold 4% paraformaldehyde (PFA) for 10 min at room temperature. Immunostaining was performed as described above. In all buffers and antibody solutions, RNaseOUT 1:200 (Thermo Fisher Scientific, 10777019) was added to preserve RNA. Before performing RNA FISH, samples were post-fixed with 4% PFA for 10 min at room temperature. RNA FISH was then performed using DNA probes as described above.

Attached blastocyst culture. Human embryo attachment culture was performed according to previously described procedures[38,39]. In brief, cryopreserved human blastocysts were received vitrified from the IVF clinic following consent and thawed using Vit Kit-Thaw (Irvine Scientific) according to manufacturer's protocol. The embryos were cultured in drops of Continuous Single Culture Complete medium, which was presupplemented with HSA (Irvine Scientific) under mineral oil (Irvine Scientific) overnight at 37 °C, 6% CO_2 and 5% O_2. The zona pellucida was removed using Tyrode's solution acidified (Irvine Scientific) and plated onto an m-Slide eight-well chamber slide (Ibidi) in IVC-1 medium (Cell Guidance Systems) and incubated for 2 d at 37 °C and 5% CO_2 to allow for attachment. Medium was half replaced on the second and third day with IVC-1. From the fourth day onward, the medium was completely replaced with IVC-2 medium (Cell Guidance Systems) until the appropriate developmental day was reached up to a maximum of day 12, which includes the blastocyst stage plus days in culture. For these experiments, 9 blastocysts were cultured up to day 12 p.f., out of which 5 were used for immuno-RNA FISH experiments.

Bulk RNA-seq data analysis. Published raw population RNA-seq datasets[25,52,53] of male and female hPGCs and somatic cells were downloaded and realigned to the hg19 genome as described previously[17] for lncRNA *XACT* expression analysis. Expression tracks were generated using pyGenomeTracks package[57].

Statistics and reproducibility. In the quantitative data, significance was assessed using Wilcoxon tests; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; ns, non-significant. Statistical analyses are described in detail in the figure legends for each panel. Immuno-RNA FISH experiments were performed in two independent experiments with similar results, unless specified otherwise in the legends. scRNA-seq datasets were pooled from ten independent experiments. No statistical methods were used to predetermine the sample size; rather, sample size was limited by the availability of prenatal tissues. Signal intensity measurement details are described in the Immunofluorescence section. For plotting and statistical analysis of scRNA-seq datasets and immune-RNA FISH quantifications, ggsignif and ggplot2 R packages were used.

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

Data availability

The scRNA-seq data of prenatal tissues reported in this paper are available under the following accession numbers: GSE143380 (female cell data) and GSE143356 (male cell data). scRNA-seq datasets are also available online for interactive exploration (http://germline.mcdb.ucla.edu). Previously published RNA-seq data of male and female hPGCs and somatic cells[25,52,53] and single-cell RNA-seq data from female germ cells[30] and from female FCGs[40] that were reanalysed here are available under the following accession codes: GSE63392 (ref. [25]), GSE60138 (ref. [52]), GSE93126 (ref. [53]), GSE79280 (ref. [30]) and GSE86146 (ref. [40]), respectively. Human conceptus tissue requests can be made to bdrl@u.washington.edu. All other data supporting the findings of this study are available from the corresponding authors on reasonable request. Source data are provided with this paper.

Code availability

Custom scripts used for aligning population RNA-seq, scRNA-seq, data processing and plotting are available on request.

References

54. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
55. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
56. Solovei, I. Fluorescence in situ hybridization (FISH) on tissue cryosections. *Methods Mol. Biol.* **659**, 71–82 (2010).
57. Ramírez, F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 189 (2018).

Acknowledgements

We thank J. Scholes, F. Codrea and J. Calimlim for support with FACS; J. Tang for banking and culturing of the UCLA hESC lines; staff at the microscopy cores at the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Center (BSCRC) for help with imaging; staff at the Technology Center for Genomics and Bioinformatics at the UCLA Johnson Comprehensive Cancer Center (JCCC) and the Next Generation Sequencing core at BSCRC for help with genomics approaches; and staff at the Translational Pathology Core Laboratory for help with histology. T.C. was supported by a Boehringer Ingelheim PhD Fellowship. This work is supported by funds from the NIH to A.C. (R01HD079546). K.P. was supported by the BSCRC at UCLA, the David Geffen School of Medicine at UCLA, and the UCLA JCCC, the NIH (R01HD098387, P01GM099134) and a Faculty Scholar grant from the Howard Hughes Medical Institute. All human pre-implantation embryo and human embryo attachment culture studies were performed using funds from the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Innovation Award. No NIH funds were used for research with human pre-implantation embryos. Human fetal tissue research is supported by a grant to I. Glass at the University of Washington Birth Defects laboratory (5R24HD000836-53).

Author contributions

T.C., K.P. and A.C. designed the experiments. T.C. and R.C. conducted immuno-RNA FISH experiments on tissues and in vitro PSC-derived cells. T.C., R.K. and E.P. generated female hiPSC lines from fibroblasts and conducted human blastocyst experiments. D.C. contributed to hPGCLC differentiation experiments. F.-M.H. created an interactive web-site to explore our scRNA-seq datasets. S.L. and K.S.-L. provided the human fetal tissues. T.C. performed scRNA-seq experiments, and the resulting data were analysed by T.C and I.D.; T.C analysed bulk RNA-seq data. T.C., K.P. and A.C. interpreted all data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

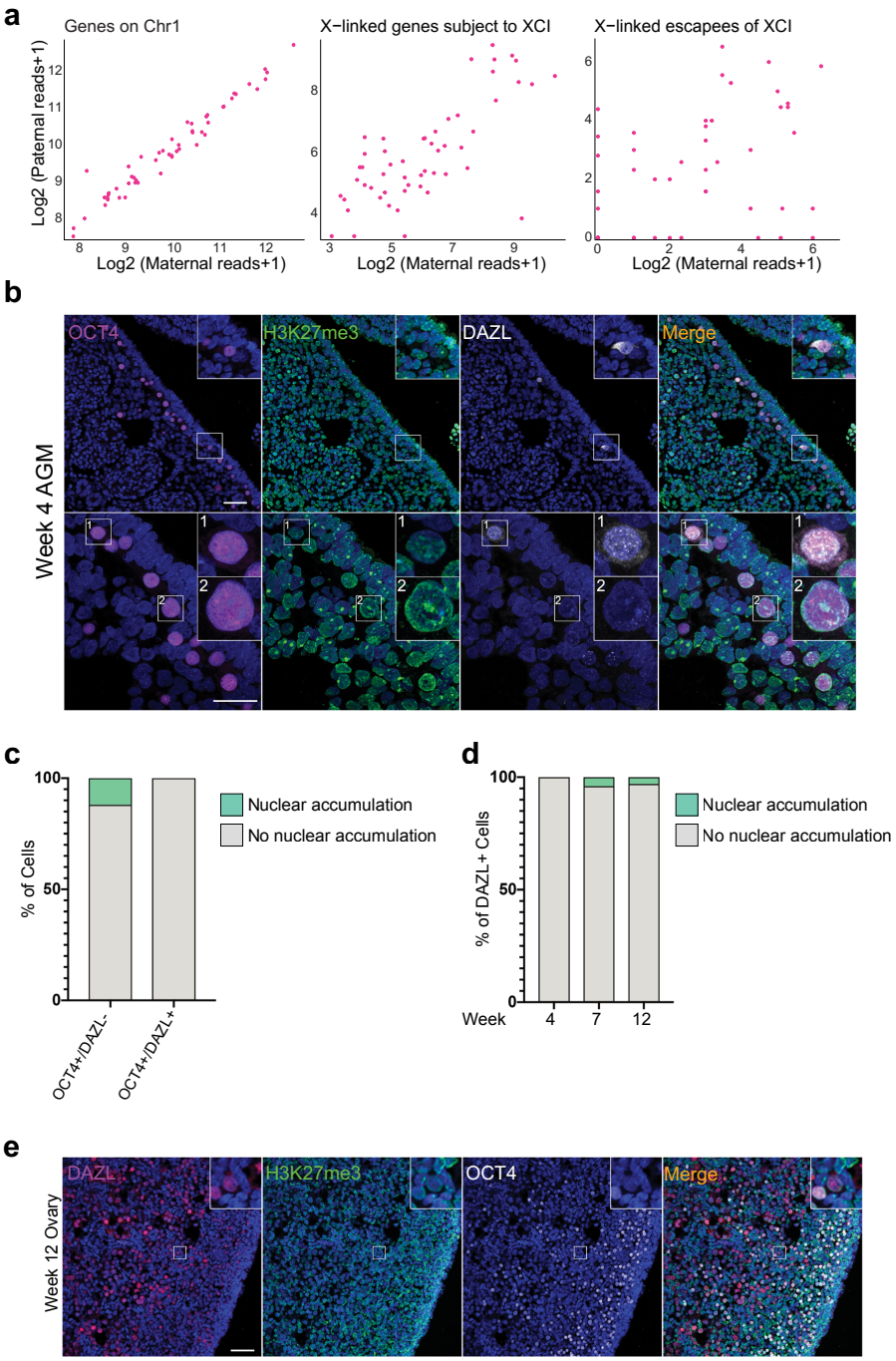
Extended data is available for this paper at https://doi.org/10.1038/s41556-020-00607-4.

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-020-00607-4.

Correspondence and requests for materials should be addressed to K.P. or A.C.

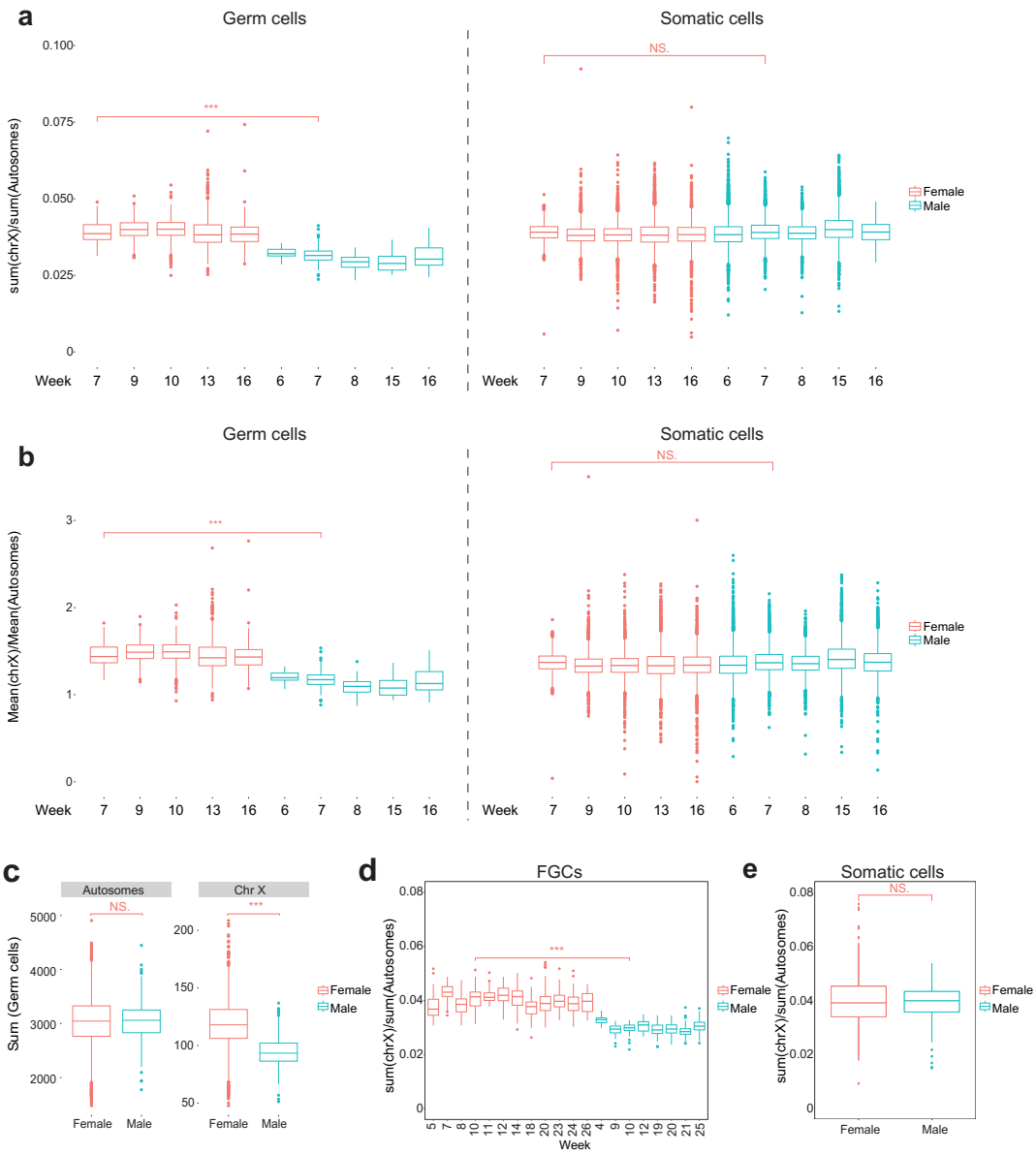
Peer review information: Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

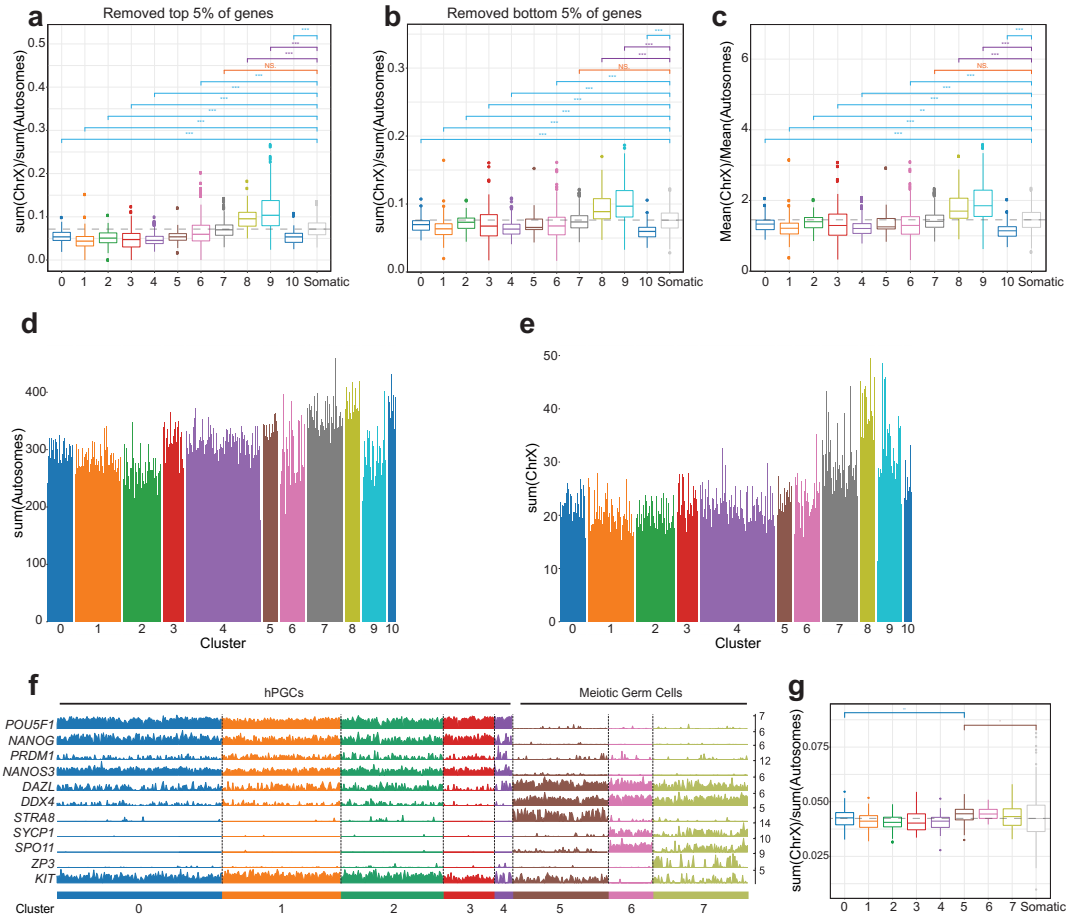


Extended Data Fig. 1 | See next page for caption.

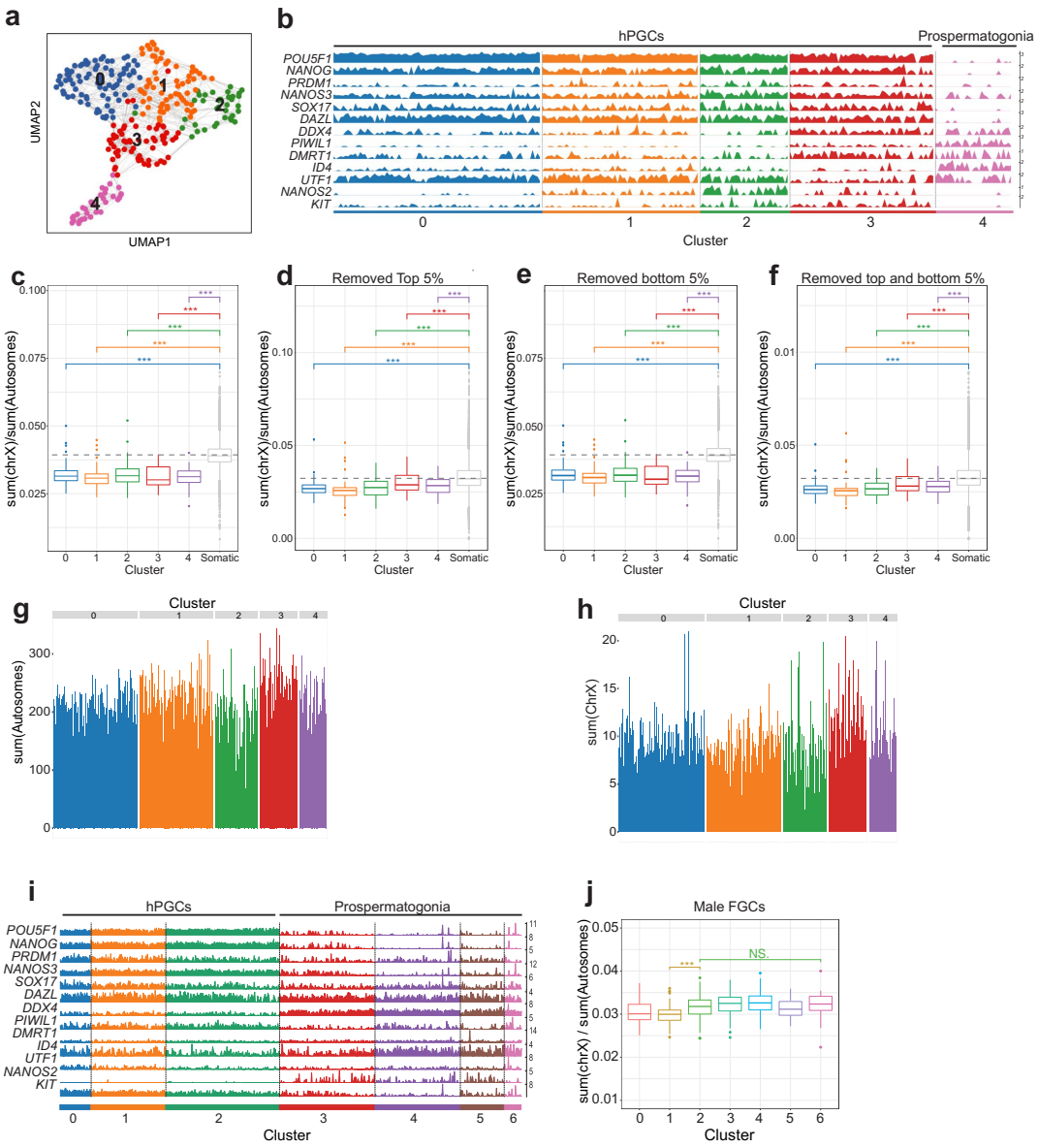
Extended Data Fig. 1 | Female hPGCs from week 4 pf ovaries have lost the H3K27me3 nuclear accumulation. a, Distribution of single nucleotide polymorphisms (SNPs) from maternally inherited (x-axis) and paternally inherited (y-axis) alleles in gene expression data of female hPGCs. Each dot represents sum of all detected SNPs per cell for genes on chromosome 1 (Chr1), X-linked genes subject to XCI and escapees of XCI, respectively, based on published scRNA-seq data[30]. **b**, Representative immunofluorescence staining of OCT4 (magenta), H3K27me3 (green), DAZL (grey) and DAPI (blue) on female hPGCs at week 4 pf prior to gonad formation, when hPGCs are migrating through the aorta-gonad-mesonephros (AGM) (1 sample was analyzed). Insets show a rare OCT4+/DAZL+ cell with no nuclear accumulation of H3K27me3 (inset 1) and an OCT4+/DAZL- cell with H3K27me3 accumulation (inset 2) along the genital ridge of the AGM. Scale bar upper panel 50 microns, lower panel 30 microns. **c**, Percentage of OCT4+/DAZL+ and OCT4+/DAZL- cells with an Xi-like nuclear accumulation of H3K27me3 from the experiment shown in (**c**); (n=58 cells from 1 AGM). **d**, Quantification of the proportion of DAZL+ female hPGCs at weeks 4, 7 and 12 pf with an Xi-like nuclear accumulation of H3K27me3 (n=50-100 cells per sample in 2 replicates). **e**, Representative immunofluorescence staining of a fetal ovary at week 12 pf with DAZL (magenta), H3K27me3 (green), OCT4 (grey) and DAPI (blue). Inset shows a DAZL+/OCT4 negative female germ cell that is negative for H3K27me3 (1 pair of ovaries were analyzed), scale bar 50 microns. Statistical source data are provided in Source Data Extended Data Fig. 1.



Extended Data Fig. 2 | The X/A ratio is higher in female germ cells than in male germ cells. a, Boxplots presenting the X/A ratio, calculated from the sum of X-linked linked gene expression and the sum of autosomal gene expression, of individual female (red) or male (cyan) germ cells (left panel) and surrounding somatic cells (right panel) obtained from gonads harvested from indicated developmental timepoints (week). **b**, As in (**a**), except that the X/A ratio was determined from the mean expression levels of X-linked and autosomal genes per cell. **c**, Boxplot showing the distribution of the sum of all autosomal (left) and X-linked (right) gene expression, respectively, in individual female and male germ cells across for developmental time points shown in (**a**). **d**, As in (**a**), except that the X/A ratios in female and male FGCs across developmental time from a published study are shown[40]. **e**, X/A ratio per single cell in female and male gonadal somatic cells from all developmental ages accompanying the data shown in (**d**). Wilcoxon statistical testing between age matched samples, NS- Not Significant, * p<0.05, ** p<0.01, *** p<0.001. a-c: n = 49528 cells analyzed across 10 independent experiments and d-e: 1016 cells analyzed from published dataset[40] in total.

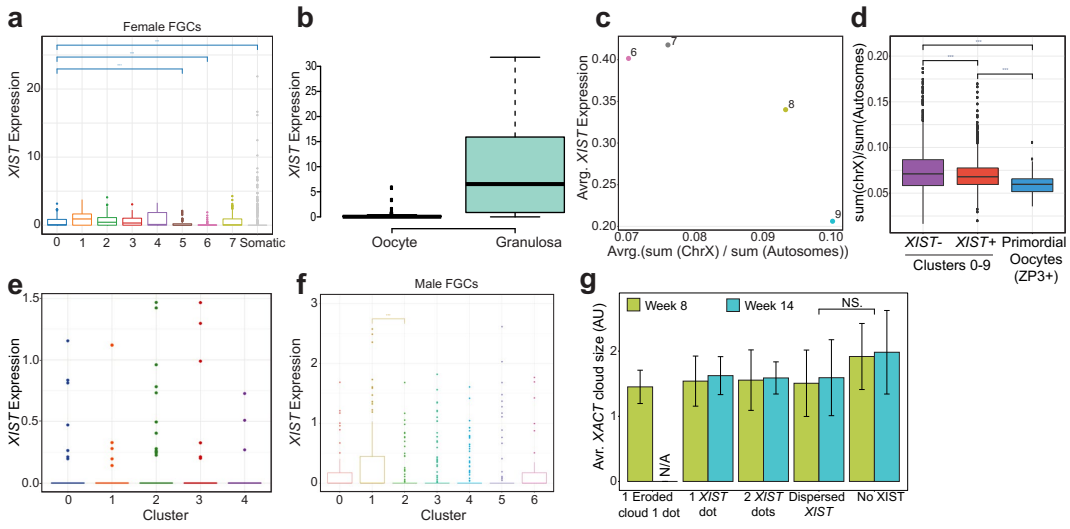


Extended Data Fig. 3 | Female hPGCs dampen X linked gene expression before entering meiosis. a, Boxplots of the X/A ratios of female germ cells along the developmental trajectory and in female gonadal somatic cells, as described in Fig. 5, except that the top 5% highest expressed genes were excluded from the analysis. **b**, As in (**a**), except that the bottom 5% of expressed genes were excluded from the analysis. **c**, Boxplots of the X/A ratios in female germ cells and female gonadal cells as described in Fig. 5, except that the X/A ratios were calculated from the mean value of X-linked and autosomal gene expression per cell. **d**, Sum of all autosomal gene expression normalized counts in female germ cells organized by clusters along the developmental trajectory. **e**, Sum of all X-linked gene expression in female germ cells organized by clusters along the developmental trajectory. X-linked gene expression increases in clusters 7-9 coincident with entrance into meiosis and repression of the naïve-like pluripotency program. **f**, Germline trajectory analysis of previously published scRNA-seq data from female FCGs[40]. The hPGCs state with pluripotency program expression is captured with clusters 0-4, and meiotic entry in cluster 5-7. **g**, Boxplots of the X/A ratios for female germ cells and female gonadal somatic cells for the data set shown in (**f**). From cluster 5 onwards, X/A ratios in differentiating female germ cells are higher than gonadal somatic cells. Wilcoxon statistical testing for (**a**), (**b**), (**c**), (**g**). NS- Not Significant, * p<0.05, ** p<0.01, *** p<0.001.

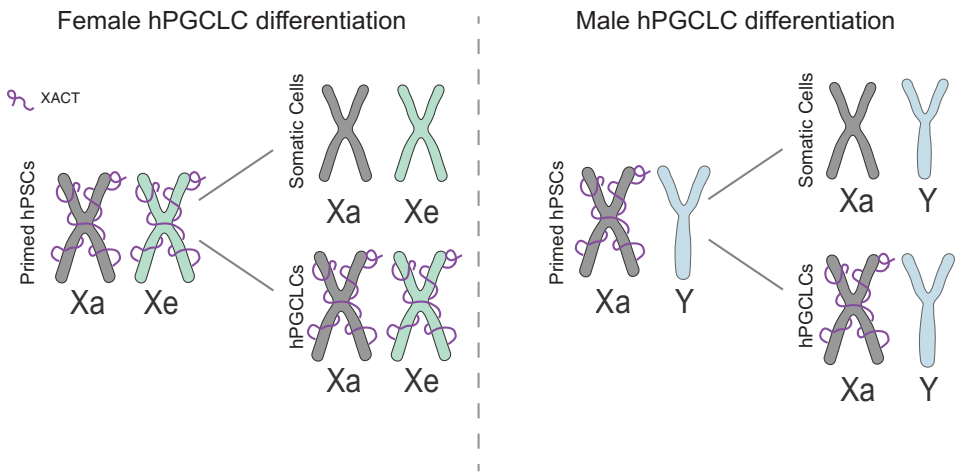


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Male hPGCs do not change X/A ratio upon sex specific differentiation. a, Male germ cells from the scRNA-seq data shown in Fig. 4a–c were projected along the developmental trajectory, and five clusters (0-4) were identified (n= 282 cells pulled from 5 samples). The pluripotency program is repressed in cluster 4, coincident with increased expression of prospermatogonia genes and exit from the cell cycle. **b**, Expression of marker genes along the developmental trajectory of male germ cells defined in (**a**). **c**, Box plots showing that X/A ratios in male germ cells along the developmental trajectory and in surrounding male somatic cells. **d**, As in (**c**), except that the top 5% highest expressed genes were excluded from the analysis. **e**, As in (**c**), except that bottom 5% of genes were excluded. **f**, As in (**c**), except that the top and bottom 5% of expressed genes were excluded. In total n = 24740 cells analyzed across 5 independent experiments in **c**-**f**. **g**, Sum of all autosomal gene expression per cell in male germ cells along the developmental trajectory, showing no dramatic differences across the clusters. **h**, As in (**g**), except for X-linked gene expression, showing no dramatic differences across the clusters. **i**, Germline trajectory analysis for male FGCs[40], identified 7 clusters (0-6). Marker gene expression is given for these clusters. **j**, Box plots of the X/A ratios in male FCGs along the developmental trajectory defined in (**i**), showing an increase in cluster 2 relative to cluster 1. In total, n = 779 cells analyzed from published dataset[40] in **j**, Wilcoxon statistical testing used for (**c**-**f**) and (**j**). NS- Not Significant, * p<0.05, ** p<0.01, *** p<0.001.



Extended Data Fig. 5 | *XIST* expression correlates with the X/A ratio. a, Boxplot depicting the expression of *XIST* in female germ cells organized by clusters along the developmental trajectory defined for the published FCG data set[40] in Extended Data Fig. 3f,g. *XIST* expression is significantly reduced from cluster 5 onwards. **b**, *XIST* expression in mature oocytes and granulosa cells from scRNA-seq data of adult ovary[44]. **c**, Scatter plot of average *XIST* expression (y-axis) and average X/A ratio (x-axis) for female germ cells clusters 6-9 (from Fig. 4d), capturing the entrance into meiosis. **d**, X/A ratios in female hPGCs and meiotic germ cells clustered based on expression of lncRNA *XIST*, *ZP3+* primordial oocytes are clustered separately. **e**, Boxplots depicting *XIST* expression in male germ cells ordered along the developmental trajectory defined in Extended Data Fig. 4a–h, indicating that *XIST* transcripts are rarely detected in male germ cells. **f**, As in (e), except for male FGCs[40] from Extended Data Fig. 4i,j. **g**, Average *XACT* cloud size in week 8 and 14 pf hPGCs with different patterns of *XIST* expression, error bars show standard deviation of the cloud sizes (76 cells analyzed in total from 2 independent experiments). Wilcoxon statistical testing used for (**a**), (**d**), (**f**), (**g**). NS- Not Significant, * p<0.05, ** p<0.01, *** p<0.001. Number of cells analyzed across 5 independent experiments: **a. n =** 1016 cells[40], **b. n =** 148 cells[44], **d. n =** 1938 cells, **f**, n = 779 cells[40].



Extended Data Fig. 6 | *XACT* marks male and female hPGCLCs *in vitro*. Summary of the hPSC differentiation figure shown in Fig. 2. Due to XCI erosion, *XACT* is expressed from the Xa and the eroded X-chromosome in primed, female human pluripotent stem cells. The Xe state is transmitted into differentiated cells and upon hPGCLC differentiation. Moreover, *XACT* is maintained in hPGCLCs, whereas somatic cells silence *XACT*. Consequently, female hPSC-derived PGCLCs carry two *XACT* clouds and male hPSC-derived PGCLCs one.

CHAPTER 3

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

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

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

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

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

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

co-first authors

* correspondence: kplath@mednet.ucla.edu

Summary

Female human pre-implantation embryos and naïve human pluripotent stem cells (hPSCs) equalize X-linked gene expression with males via X-chromosome dampening (XCD), a unique strategy of dosage compensation in mammals. The mechanisms controlling XCD are unknown. Here, we show that the long non-coding RNA XIST, which mediates X-chromosome inactivation (XCI), is required for XCD. XIST employs similar principles and protein partners, including SPEN, to execute XCD and XCI, but displays a lower accumulation and different distribution on the dampened versus the inactive X which might explain the differential gene regulation. Unexpectedly, XIST also spreads beyond the X chromosome territory to specific autosomal

regions and induces the downregulation of developmental autosomal genes in female naïve hPSCs and pre-implantation embryos. Thus, XIST balances X-linked gene expression but causes imbalances in autosomal gene expression between male and female cells early in human development.

Highlights

• XIST RNA controls X-chromosome dampening in female naïve hPSCs

• SPEN requirement, chromosome-wide localization of the RNA, and escape from repression for a subset of genes are common principles of both XCD and XCI

• XIST reproducibly localizes to a small number of autosomal target regions in naïve hPSCs likely/plausibly at the expense of X-chromosome binding

• Targeted autosomal genes are repressed by XIST in female naïve hPSCs and pre-implantation embryos, revealing a unique function of XIST *in trans* to its transcription locus

Introduction

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

Although XCI occurs in all placental mammals during *post*-implantation development, the regulation of X chromosome gene dosage in *pre*-implantation embryos differs significantly between the species (Mahadevaiah et al., 2020; Okamoto et al., 2011). For example, *mouse* pre-implantation embryos undergo an imprinted form of XCI, silencing specifically the paternally inherited X chromosome (Huynh and Lee, 2003; Lee and Bartolomei, 2013; Mak et al., 2004; Okamoto et al., 2004; Takagi and Sasaki, 1975). The silencing of the paternal X chromosome is maintained in the extra-embryonic lineages, whereas epiblast cells of the early blastocysts reactivate the paternal Xi and subsequently undergo random XCI (Lee and Bartolomei, 2013; Okamoto et al., 2011; Takagi and Sasaki, 1975; van den Berg et al., 2011). In contrast, in *human* pre-implantation embryos, both the paternally and maternally inherited X chromosomes remain active (Okamoto et al., 2011) and genes on both X chromosomes are downregulated by reducing transcription by about half, in a process referred to as X-chromosome dampening (XCD)

(Petropoulos et al., 2016; Sahakyan et al., 2017a, 2017b). Thus, X-chromosome dosage compensation in human development occurs by two different and sequential mechanisms, first XCD and later, upon implantation, XCI. Recently it was shown that dosage compensation in the human germ line is also achieved by XCD (Chitiashvili et al., 2020), suggesting that certain aspects of human development uniquely require two active, yet dosage compensated, X chromosomes. Since XCD does not occur in mice, which typically serve as model system for mammalian X chromosome dosage compensation, a mechanistic understanding of XCD is completely lacking.

It has been proposed that/the main suspect for initiating XCD is the master regulator of XCI,the long non-coding RNA (lncRNA) XIST (X-inactive specific transcript) (Borensztein et al., 2017; Brown et al., 1991; Leppig and Disteche, 2001; Marahrens et al., 1997; Penny et al., 1996; Wutz et al., 2002; Wutz and Jaenisch, 2000). In female human pre-implantation embryos, XCD occurs gradually from the 4-cell to the blastocyst stage, which correlates with the upregulation of XIST expression (Petropoulos et al., 2016).Dampening is not observed on the male X, which expresses XIST at a much lower level such that the XIST cloud is often not observed (Brown and Robinson, 1997; Daniels et al., 1997; Okamoto et al., 2011; Petropoulos et al., 2016). Similarly, germ cells express XIST and undergo XCD, while upon differentiation XIST is repressed, accompanied by an increase in X-linked transcript levels while global autosomal transcription remains unchanged (Chitiashvili et al., 2020). Despite the correlation between XIST expression and XCD it is still unclear if XIST mediate XCD as XIST expression generally initiate the completely inactive status of the X chromosome. Intriguingly, imaging approaches have shown that XIST expression in female human pre-implantation embryos and the female germline undergoing XCD is accompanied by the appearance of a distinct cloud on the active but dampened X chromosome (Xd) (Chitiashvili et al., 2020; Okamoto et al., 2011; Petropoulos et al.,

2016; Vallot et al., 2017), revealing a puzzling uncoupling of XIST expression from complete gene silencing and challenging the hypothesis of XIST role in this state.

In this study, we set out to explore the role of XIST in XCD. Recent reports have described culture conditions for the establishment and maintenance of naïve hPSCs that capture the state of pluripotent cells in the human pre-implantation embryo, including XCD as well as XIST expression (Chan et al., 2013; Gafni et al., 2013; Guo et al., 2016; Huang et al., 2014; Sahakyan et al., 2017a, 2017b; Takashima et al., 2014; Theunissen et al., 2014, 2014; Vallot et al., 2017). Specifically, naïve human embryonic stem cells (hESCs) can be derived directly from the human blastocyst or converted from developmentally advanced primed hESCs (Nichols and Smith, 2009; Patel et al., 2017; Rostovskaya et al., 2019; Theunissen et al., 2014). Similarly, naïve human induced pluripotent stem cells (hiPSCs) can be established by reprogramming from fibroblasts (Liu et al., 2017). Here, we generate and exploit naïve human pluripotent stem cell lines (hPSCs) to explore, for the first time, the mechanisms underlying XCD and the role of XIST in XCD and naïve pluripotency.

Applying a CRISPR-Cas9-based loss-of-function approach we found that the dampened X is reactivated in the absence of XIST, which establishes XIST as regulator of both XCD and XCI. Using genomic and imaging approaches, we discovered differential localization and lower accumulation of XIST on dampened versus inactive X chromosomes. Although XIST was predominantly enriched on the X chromosome, we identified a remarkable spreading of XIST beyond the X territory during XCD. We suggest that XIST's silencing function is attenuated in early human development due its decreased accumulation in the X territory. Unexpectedly, we found an additional new function of XIST. XIST spreading beyond the X territory was accompanied by its binding to specific autosomal genes. We provide evidence that the autosomal localization of XIST follows similar principles as on the X chromosome and that XIST

transcriptionally downregulates its autosomal target genes in female hPSCs. Thus, XIST is not completely retained on the chromosome in *cis* to its transcription locus in naïve human pluripotent cells and can function on chromosomes in *trans*. Importantly, we confirm that the female to male imbalance of the autosomal genes targeted by XIST is also observed in human pre-implantation embryos. Together, our study uncovers XIST as master regulator of both XCD and XCI and establishes a new role for XIST in the regulation of autosomal genes, which leads to gene expression differences between male and female human preimplantation embryos.

Results

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

To evaluate the role of XIST in the regulation of XCD, we derived several female naïve hPSC lines in 5iLAF culture medium and confirmed the presence of XIST transcript. Transcriptomics studies have shown that the 5iLAF culture method stabilizes the naïve pluripotent state described for epiblast cells of human pre-implantation embryos in cell cultured in the dish (Sahakyan et al., 2017a; Theunissen et al., 2016). Importantly, while 5iLAF-cultured naïve hPSCs display monoallelic expression of XIST, the XIST coated X is dampened but not silenced. Therefore, these cells are considered a model for studies of XCD mechanisms (Sahakyan et al., 2017a). Consistent with prior reports (Sahakyan et al., 2017a; Theunissen et al., 2016, 2014; Vallot et al., 2017), the two female naïve hESC lines, derived from the primed hESC lines UCLA1 and H9, and the naïve hiPSC line, expressed XIST from one of the two X chromosomes in most cells (Fig 1a,b). To explore the role of XIST in XCD we first examined the correlation/relation between XIST and XCD in the naïve hPSC line using bulk and single cell (sc) RNA-sequencing (RNA-seq) (Table S1) and RNA fluorescent in situ hybridization (FISH) as described below.

The conversion of primed to naïve hESCs occurs through an intermediate state in which both X-chromosomes are active and XIST is repressed (naïve hPSCs$^{pre\text{-}XIST}$) (Sahakyan et al., 2017a) (Fig S1a), providing an opportunity to evaluate X chromosome expression dynamics in relation to XIST. Based on bulk RNA-seq data, we observed that the expression ratio between genes on the X chromosome and on autosomes (X/A ratio) negatively correlates with XIST levels (Fig 1c). The lowered X/A ratio in XIST-expressing naïve cells is due to a decrease of X-linked gene expression in the absence of major changes in autosomal gene expression (Fig 1d). These changes bring X-linked gene expression in female naïve hPSCs closer to that of male naïve hESCs (WIN1), which carry a single XIST-negative X chromosome and do not display XCD (Fig 1a,b,e, S1a-c). Female naïve hPSCs did not completely reach the X-chromosome expression level of male naïve hESCs (Fig 1e), consistent with the idea that dampening of X-linked gene expression may only occur on one of the two X chromosomes. Single cell (sc) RNA-seq analysis of naïve H9 hESCs corroborated the negative correlation between XIST expression and the X/A ratio (Fig 1f, S1d). Together, these data confirm that X chromosome dosage compensation occurs in female naïve hPSCs and that higher XIST levels are associated stronger repression of X-linked gene.

To confirm that the dosage compensation process observed in naïve hPSCs is not XCI, we examined X chromosome expression at allelic resolution. Reads from the X-chromosome were split according to allele-specific single nucleotide polymorphisms (SNPs) and assigned to a given X-chromosome (X1 (Xa in the primed hPSCs) or X2 (Xi in the primed hPSCs)). Allelic analysis of the scRNA-seq data demonstrated biallelic expression of X-linked genes in individual naïve H9 cells (Fig 1g, S1e). Similarly, nascent RNA FISH for three X-linked genes that are normally subject to XCI (Tukiainen et al., 2017), and are also subject to XCD in epiblast cells (Fig S1f), *SMS*, *SMARCA1* and *GPC3,* revealed biallelic expression in naïve H9 cells (Fig S1g, h). Together, these findings demonstrate that both X chromosomes are active in our female naïve hPSC lines,

in agreement with previous studies (Okamoto et al., 2011; Petropoulos et al., 2016; Sahakyan et al., 2017a).

The monoallelic expression of XIST in naïve hPSCs uniquely enabled us to explore whether the XIST-positive X chromosome displays XCD compared to the XIST-negative X chromosome. To this end, we performed an allele-specific analysis of population RNA-seq data for different preparations of naïve UCLA1 hESCs (X1 = Xa and X2 = Xi in primed UCLA1 and H9 with non-random XCI, see methods) and found that when the population skews XIST expression towards one of the two X chromosomes, X-linked gene expression is biased towards the other X-chromosome (Fig 1h, S1i). Applying quantitative RNA FISH, we found that cells with biallelic expression of the X-linked genes *THOC2 and UTX* (*KDM6A*) display a 25% smaller nascent transcript signal on the XIST-positive versus XIST-negative X-chromosome (Fig 1i,j). Qualitatively, a similar result is obtained for the X-linked genes *SMS*, *SMARCA1* and *GPC3* (Fig S1g, h). Moreover, the lncRNA XACT, which associates with the X chromosome in *cis* (Vallot et al., 2017, 2013), displays a less intense cloud-like signal when transcribed on the same X chromosome as XIST (Fig 1k,l, S1j). Together, these results show that dampening occurs specifically on the XIST-coated X-chromosome and additionally suggest that XIST regulates XCD in naïve hPSCs.

XIST is required for XCD

To interrogate the function of XIST in XCD, we first examined the nature of XIST transcript in naïve hESCs and found that the transcriptional start site, 3' end of XIST, and exon/intron structure are the same in XCD and XCI(Fig S2a,b), suggesting that the differential function of XIST on the Xd and Xi is not explained by the predominant expression of alternative isoforms.

To directly test if XIST is required for XCD, we used CRISPR/Cas9 technology to homozygously excise a 2kb region from the promoter region and 5' end of XIST in the primed H9 hESC line (Fig 2a, Fig S2c,d). We converted two independent homozygous knockout (KO) clones (clones C7 and C18) and wildtype (WT) H9 control hESCs to the naïve state with 5iLAF medium (Fig 2a). As expected, XIST was detectable by RNA-seq and RNA FISH in control cells but not in targeted clones (Fig 2b-d, Fig S2e). Importantly, the induction of naïve pluripotency markers (Collier et al., 2017; Theunissen et al., 2016) confirmed the successful conversion to naïve pluripotency (Fig 2e). To explore the role of XIST at a single cell resolution we preformed scRNA-seq data for the control H9 hESCs and the deletion clone C18 and confirmed that naïve hESCs with and without XIST reach the same expression state (Fig S2f) and that both X-chromosomes are active in individual cells (Fig S2g). Based on scRNA-seq data, most WT and KO cells belong to the same cluster (cluster 0 with 91%, and 96% of WT and KO cells, respectively) (Fig 2f).

Since the naïve pluripotent state could be established and maintained in the absence of XIST, we were able to next examine the gene expression data for whether the loss of XIST resulted in reactivation of the dampened X. Based on population and scRNA-seq data, we found an increased X/A ratio upon XIST deletion (Fig 2g,h, S2h,i), which is due to a significant upregulation of X-linked genes compared to autosomal genes (Fig 2i, S2j). Thus, in the absence of XIST, the global level of X-linked transcripts is significantly increased in naïve hESCs, whereas global autosomal gene expression levels remain constant. Together, these data demonstrate that X-chromosome dosage in naïve hESCs is controlled by XIST, indicating that XIST controls XCD. Our results uncover XIST as regulator of not only XCI but also XCD in humans.

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

A key characteristic of XIST-mediated XCI is that a subset of X-linked genes escapes silencing (Carrel and Willard, 2005; Johnston et al., 2008; Navarro-Cobos et al., 2020; Tukiainen et al.,

2017; Zhang et al., 2011). Therefore, we explored whether X-linked genes can also escape XCD. Typically, allele-specific expression based on SNPs or the presence of biallelic nascent expression of X-linked genes by RNA-FISH are used to define XCI escapees, but these approaches cannot easily discern XCD escapees due to the inherent biallelic expression of X-linked genes (Fig 1h, S1g,h). Consequently, we defined XCD escapees as those genes with significantly higher expression in female versus male naïve hPSCs (see methods).

This analysis identified 157 (H9), 122 (UCLA1), and 67 (iPSCs) X-linked genes with significantly higher transcript levels in 5iLAF-cultured female hPSCs compared to the naïve male hESC line WIN1 (Table S2), defining 8-25% of X-linked genes as escapees (Fig S3a). XCD escapees are distributed over the entire X-chromosome (Fig S3b). 64 escapee genes were conserved in two female hPSC lines and 15 in all three, uncovering a conserved set of XCD escapee genes (Fig S3c, Table S2). XCD escapees also overlapped with genes that were unchanged or even upregulated upon transition from the pre-XIST to the stable XIST-positive naïve state (Fig S3d). Similarly, XCD escapees defined in 5iLAF-cultured naïve cells overlapped with XCD escapees defined in naïve hESCs that were derived directly from blastocysts with the t2iLGö culture method (Guo et al., 2017; Takashima et al., 2014) (Fig S3a,b,e,f). We conclude that XCD escape occurs in female naïve hPSCs regardless of cell line, derivation procedure (primed hESCs, fibroblasts, blastocyst), and culture condition. Notably, even upon removing the XCD escapees, X-linked genes were biallelically expressed in our scRNA-seq data (Fig S3g), providing further evidence that naïve cells achieved dosage compensation by XCD rather than by XCI.

Next, we determined whether the XCD escapees that we defined in female hPSCs capture also escape XCD in female pre-implantation embryos. Specifically, we compared the expression of X-linked genes between female and male epiblast cells from pre-implantation blastocysts at

day 7 of development (E7), taking advantage of previously published scRNA-seq data (Petropoulos et al., 2016). Due to XCD in female embryos, female and male epiblast cells express a comparable overall dose of X-linked genes (Petropoulos et al., 2016). Yet, we identified 86 genes as XCD escapees in E7 female epiblast cells (Fig S3a,b, Table S2) of which 7 overlapped with the 15 XCD escapees conserved among our three female 5iLAF-cultured hPSC lines (Fig S3h). Thus, escape from dosage compensation is a conserved feature of XCD *in vivo* and *in vitro*. XCD escapees are, on average, more highly expressed than the remaining X-linked genes even in cells that lack XCD (like in the male naïve hESC line WIN1 or in pre-XCI naïve UCLA1) (Fig S3i,j), suggesting that the higher gene expression level plays a role in the ability of a gene to escape XCD. Consistent with the regulation of both XCD and XCI by XIST, we also found significant overlap among XCI and XCD escapees (Fig S3k). For instance, the X-linked gene *UTX* (*KDM6A*) is a well-known XCI escapee that also escapes XCD in naïve UCLA1 hPSCs and the pre-implantation E7 epiblast (Fig S3m). Taken together, a conserved set of genes can escape both XCI and XCD consistent with the regulation of both processes by the XIST lncRNA.

Interestingly, our population and scRNA-seq data showed that both XCD subject and escapees were upregulated in the absence of XIST, yet genes subject to XCD displayed significantly higher upregulation than XCD escapees (Fig 2j, S2k,l). This result is consistent with the observations that the escapee *KDM6A* is repressed by 25% on the XIST-expressing X-chromosome compared to the other X-chromosome in the same cell (Fig 1i). In summary, these results demonstrate that a subset of highly expressed genes escape XCD at least partially, yielding higher expression level of these genes in female embryos and naïve hPSCs compared to their male counterparts.

XIST shows differential distribution on the Xd versus Xi

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

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

Although XIST was enriched across the entire X chromosome, its precise levels and relative distribution across the X differed between Xd versus Xi (Fig 3c,Fig S4c). We explored the genomic regions that are differentially enriched for XIST on the Xd and Xi and found that an 80 Mb region across the center of the X chromosome was relatively more enriched for XIST on the

Xi in fibroblasts whereas a 40 Mb region at the end of the X was relatively more enriched for XIST on the Xd in naive hPSC (Fig 3d, S4d). The switch from the Xi- to Xd- biased XIST enrichment occurred at the macro-satellite repeat locus *DXZ4* (Fig 3d, S4d), which is known to partition the Xi in somatic cells into two spatial super domains (Bonora et al., 2018; Darrow et al., 2016; Deng et al., 2015; Giorgetti et al., 2016; Rao et al., 2014; Wang et al., 2016) (Fig S4e). The spreading of XIST on the start of the X chromosome is similar on the Xd and Xi, which may be related to the fact that many XCI escapers are concentrated in this region (Tukiainen et al., 2017). Consistent with this, the intra-chromosomal interactions of the left arm of the X chromosome are similar between the Xa and Xi in somatic cells (Fig S4e). These results suggest that chromatin structure plays a role in controlling the differential distribution of XIST along the Xd and Xi and that the differential distribution may play a role in the differential gene regulation.

To further characterize the variation of XIST levels across the Xd and Xi, we correlated the enrichment of XIST with diverse genomic features (Table S4). On the Xd, XIST enrichment correlated best with gene density, SINE elements, mammalian-wide interspersed repeats (MIRs), and the LINE element L2 (Fig 3f). MIRs and L2 elements are often located in gene rich regions (Medstrand et al., 2002) and both have enhancer-like characteristics including high levels of H3K27ac, DNase I hypersensitivity (Cao et al., 2019), consistent with the positive correlation between XIST localization and gene density. The positive correlation with gene density and MIR elements also applied to XIST's association with the Xi, albeit it was much weaker compared to the Xd (Fig 3e, Table S4). XIST enrichment on both the Xd and Xi was negatively correlated with Long Terminal Repeats (*LTRs)* retrotransposons, particularly with ERVL-MaLR elements (Fig 3e, Table S4). Closer inspection of XIST localization at candidate escapees in female naïve hPSCs revealed an easily visible reduction of XIST levels over the gene bodies of some XCD escapees. For example, the XCD escapee UTX was in a genomic region with a depletion of XIST relative to neighboring intergenic regions (Fig 3f). Consistent with this result, a quantitative analysis showed

that genes subject to XCD tend to be more enriched for XIST than genes that escape XCD in naïve hPSCs (Fig 3g, Fig S4e) or in pre-implantation epiblasts (Fig 3h). XCI escapees behaved similarly (Fig 3f,i). We conclude that XIST preferentially localizes to gene-rich regions on the Xd in naïve hPSCs and the Xi in fibroblasts. Genes that escape from XCD or XCI are associated with reduced XIST localization on the Xd and Xi, respectively.

These data reveal that while the relative XIST localization on the X differ between Xd and Xi, its spreading follows similar principles, highlighting important parallels between XCD and XCI. Moreover, the X chromosome-wide localization of XIST on the Xd in naïve hPSCs reinforces the role of XIST as master regulator of XCD. Yet, at the global level, XIST localizes rather differently across the Xi and Xd, as described below.

XIST spreads beyond the X chromosome territory in naïve hPSCs

We surprisingly found that the proportion of reads aligned to the X in the XIST pulldown was reproducibly larger in fibroblasts than in naïve hPSCs (Fig 3a,b). Consequently, fewer genomic regions were highly enriched for XIST RNA on the Xd versus Xi. Specifically, 2-17% of windows showed a 20-fold enrichment of XIST on the Xd, whereas 41-46% of windows displayed a 20-fold enrichment of XIST on the Xi. Thus, the inability of XIST to induce complete gene silencing in naïve hPSCs is associated with decreased levels of the RNA on the X chromosome.

Given the lower overall enrichment of XIST on the Xd compared to the Xi (maybe: despite overall similar transcription levels (fig S56)?), we hypothesized that XIST spreads beyond the X chromosome in naïve hPSCs. In support of this idea, XIST formed a less compact cloud-like signal on the Xd than on the Xi (Okamoto et al., 2011; Petropoulos et al., 2016; Sahakyan et al., 2017a; Vallot et al., 2017)(Fig 4a). Moreover, DNA FISH with an X-chromosome paint followed by RNA FISH for XIST revealed that XIST RNA spreads beyond the Xd chromosome territory in

the naïve UCLA1 hESCs, whereas it is restricted to the Xi territory in fibroblasts (Fig 4b). To further characterize the unique localization pattern of XIST in naïve hPSCs, we inspected the XIST RAP-seq data for autosomal binding of the RNA.

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

Interestingly, in the three naïve hPSC samples, only 6-10% of 100 Kb autosomal windows display enrichment of XIST (Fig 4c), indicating that only a small portion of the autosomal genome is targeted by the RNA. Taken together, these data show that the reduced accumulation of XIST in the X chromosome territory in naïve hPSCs is accompanied by the localization of the RNA to a small set of autosomal regions. Moreover, autosomal targeting of XIST is a feature uniquely associated with the naïve pluripotent state.

XIST spreads to specific autosomal regions in naïve hPSCs

To explore the unprecedented localization of XIST to autosomal regions in naïve hPSCs, we defined XIST-enriched autosomal regions using the MACS2 broad peak calling algorithm in each RAP-seq replicate (Table S5). We identified thousands of XIST-enriched autosomal peaks in the

naive hPSCs (6873 in H9, 9835 in UCLA1, and 9325 in iPSCs) (Fig 4d,Table S5), targeting all the autosomes (Fig 4e, S5c/d). In contrast, in fibroblasts, where we expected no localization of XIST to autosomal regions, we observed 79 and 100 significant autosomal peaks, respectively, in the two RAP-seq replicates (q-value<0.05, Fig 4d/e, Fig S5c/d, Table S5). Close examination of XIST peaks in fibroblasts revealed that most fall into highly repetitive genomic regions such as those around centromeres and telomeres (Fig S5d), and therefore likely represent false alignment rather than real XIST localization.

771 of the XIST-enriched autosomal peaks were conserved among the RAP-seq data sets derived from the three different hPSC lines (UCLA1, H9 and iPSCs), and none of them were detected in fibroblast RAP-seq replicates (Fig 4f, S5e, Table S5). We therefore classified these as naïve hPSC-conserved XIST peaks. For comparison, only two conserved autosomal peaks were identified in the fibroblast replicates. In addition, 2355 peaks overlapped in 2 out of the 3 naive hPSC datasets (Fig S5e). The conservation of peaks correlates with the level of XIST enrichment, such that peaks present in all three naïve hPSC lines showed a higher level of XIST than those detected in only a subset of replicates (Fig 4g, Fig S5f). Naïve hPSC-specific autosomal peaks are on average 52Kb long on average (Fig S5g) and, although present on all autosomes, predominantly associated with chromosome (chr) 11 (106 peaks, 4.9% of the chromosome), chr20 (57 peaks, 3.3% of the chromosome), and chr 1 (101 peaks, 3.2% of the chromosome) (Fig 4f,h, Fig S5h).

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

XIST spreads to transcription regulatory regions on the autosomes

To get insights into the localization and function of XIST on autosomes, we explored the enrichment of diverse genomic and epigenomic features in the XIST-enriched autosomal regions. Similar to observations on the X chromosome, we found that autosomal peaks of XIST were best correlated with MIR and L2 repetitive elements (Fig 5a, Table S4), suggesting that similar mechanisms govern the localization of XIST on the X chromosome and on autosomes. Consistent with the association of MIRs and L2 elements with gene regulatory regions (Cao et al., 2019; Petri et al., 2019), we found that XIST localizes to autosomal regions enriched in gene bodies. Specifically, conserved XIST peaks in naïve hPSCs extend across 492 autosomal protein-coding genes (2.6% of all autosomal protein-coding genes) of which 250 (1.8%) are expressed in naïve hPSCs (Fig 5b-c, Table S6). These genes are significantly enriched for differentiation and developmental gene ontology (GO) terms (p-value < 0.000005, Fig 5d, Table S7) and contain genes encoding the transcription factors *ARNTL*, *CAMTA1*, *CTCFL*, *ESRRB*, *ETV5*, *FLI1*, *GLI2*, *HIVEP3*, *HOXA10*, *HOXA9*, *IRX6*, *KLF6*, *PKNOX2*, *RBPJ*, *TEAD1*, *TRERF1*, *ZNF423*, and *ZNF618.* Some of which play a role in development and pluripotency, as for example Estrogen Related Receptor Beta (ESRRB) which is known to play an essential role in early development and PSCs (Adachi et al., 2018; Benchetrit et al., 2019). Examples of autosomal XIST-covered genes shown in Figure 5e include the gene encoding Transcriptional-regulating factor 1 (*TRERF1*), a progesterone receptor coactivator encoded on chromosome 6 (Gizard et al., 2006), Transforming Growth Factor alpha (*TGF*□) located on chromosome 2, and Cluster of Differentiation 5 and 6 (*CD5* and *CD6*) involved in the activation and differentiation of lymphocytes located on chromosome 11 (Gimferrer et al., 2003).

Similar to genes on the X chromosome, autosomal genes targeted by XIST tend to be significantly more lowly expressed than those not associated with the RNA, in both female and

male naïve hPSCs (Fig 5f, Table S6). This result suggests that the gene expression state is one factor that guides XIST localization to specific autosomal regions. To identify potential protein mediators involved in the recruitment or functions of XIST to autosomes, we mined ENCODE ChIP-seq data across a large variety of cell types. We found that the binding sites of multiple regulators involved in transcriptional repression, including the Polycomb Repressive Complex 2 (containing SUZ12 and EZH2), NRSF and CtBP2, are enriched within autosomal XIST peaks (Fig 5g Table S7), consistent with their role in developmental regulation and the observed lower overall expression of genes under XIST peaks.

Our results reveal the unexpected and remarkable localization of XIST to a small subset of developmental regulators on autosomes in female naïve hPSCs. In mouse and human, Xist/XIST can silence autosomal genes when ectopically expressed from an autosome (Hall et al., 2002; Jonkers et al., 2009; Kelsey et al., 2015) or when spreading into autosomal regions on X:A translocations (Yang et al., 2011). For example, integration of XIST into one of the three chr21 copies in primed human iPSCs derived from Down Syndrome patients enables the silencing of the XIST-associated chromosome and corrected gene expression to nearly the normal disomic chr21 levels (Jiang et al., 2013). Although it was known that XIST is inherently able to regulate autosomal genes, the physiological relevance of this capacity was unclear. Our observations raise the interesting possibility that XIST may transcriptionally regulate autosomal developmental genes during early human development and hPSCs.

XIST mediates the repression of autosomal genes

To explore this exciting possibility, we compared gene expression levels between female naïve hPSCs with XIST (UCLA1, H9, and iPSCs) and the male naïve hESC line WIN1, which lacks XIST. We found that autosomal genes targeted by XIST had significantly lower expression in female versus male naïve hPSCs, whereas genes that do not overlap XIST peaks showed similar

expression in the four naïve hPSCs (Fig 6a). Similarly, XIST was more highly enriched on genes that were downregulated in female versus male naïve hPSCs, compared to genes with similar expression (Fig 6b).

We identified 87 autosomal genes that are located within conserved XIST peaks in naïve hPSC and downregulated in all three female naïve hPSC lines compared to WIN1 (Table S6). Among these genes is *SPON1* (Spondin 1), which is encoded on chr 11 and known to promote neuronal differentiation (Gyllborg et al., 2018) (Fig 6c). RNA FISH for the nascent transcripts of *SPON1* in female and male naïve hESCs revealed a higher proportion of female cells with monoallelic expression of this gene (Fig 6d), consistent with the lower expression of *SPON1* in female cells. 82 additional genes were downregulated in 2 of the 3 female naïve hPSC lines compared to the male WIN1 (Table S6), including *HUNK* (Hormonally Up-Regulated Neu-Associated Kinase) on chr 21 which has a role in the proliferation and *differentiation of epithelial cells* (Gardner et al., 2000) (Fig 6e). For both *HUNK* and *SPON1*, we found a negative correlation between their expression level and XIST transcript levels in naïve hPSCs (Fig 6f), which held true for other genes that overlapped with autosomal XIST peaks. Specifically, XIST targets had lower expression in cells where XIST was highly expressed and vice versa, whereas genes found in XIST-depleted autosomal regions showed no correlation to XIST transcript levels (Fig 6g). Overall, these data strongly support the notion that XIST mediates the transcriptional repression of autosomal genes in naïve hPSCs.

To experimentally examine the role of XIST in the regulation of autosomal genes, we interrogated the regulation of XIST-targeted autosomal genes upon loss of XIST, using our scRNA-seq expression data for XIST WT and KO H9 naïve hESCs (Fig 2). We found a small, yet significant, correlation between the degree by which an autosomal XIST associated gene is upregulated in the XIST KO and the enrichment of XIST around that gene (Fig 6h). More

importantly, genes located in XIST-enriched autosomal regions were significantly upregulated upon XIST deletion compared to genes in XIST-depleted regions (Fig 6i). Overall, our data show that XIST spreading beyond the X chromosome results in the downregulation of autosomal genes in naïve hPSCs, providing evidence of a role for endogenous XIST expression in controlling autosomal gene expression.

An important question is whether the regulation of autosomal genes by XIST is recapitulated during human pre-implantation development. Intriguingly, both *SPON1* and *HUNK* were also downregulated in female versus male epiblast cells of human pre-implantation embryos (Fig 6j). In addition, we found that genes in XIST-enriched autosomal regions displayed significantly lower expression in female compared to male epiblast cells in E6 and E7 pre-implantation embryos (Fig 6k). Cells from E5 embryos did not display this sex-specific difference (Fig 6l), consistent with the prior report that XIST is strongly upregulated in female embryos from E5 to E6 (Petropoulos et al., 2016). Similarly, genes with significantly lower expression in female compared to male epiblast cells *in vivo*, were associated with higher XIST enrichment compared to genes that showed no significant differences between female and male epiblasts (Fig 6l).

Taken together, these results show that the sex-specific expression of XIST-enriched autosomal genes observed in naïve hPSCs extends to the human embryo and point to a novel role of XIST in downregulation of autosomal genes during human embryonic development.

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

To further explore the mechanism by which XIST regulate XCD and autosomal genes in naïve hPSCs, we examined whether these XIST-mediated processes share additional features with the Xi, such as changes in chromosome organization and compaction compared to the Xa (Pandya-Jones and Plath, 2016). Indeed, combining RNA FISH of XIST with DNA FISH for specific

genomic locations on the X-chromosome, we found that high-order interactions in the XIST-coated Xd differ from the Xa, yet in a different way from the Xi (Fig S6a,b). In addition, Xist interacts with various proteins to establish gene silencing and the Xi compartment during XCI in mouse cells (Chu et al., 2015; Dossin et al., 2020; Graindorge et al., 2019; McHugh et al., 2015; Moindrot and Brockdorff, 2016; Nesterova et al., 2019). We therefore explored whether known interactors of XIST participate in XCD and autosomal gene regulation. First, we assessed the localization of the protein Cip1-interacting zinc finger protein 1 (CIZ1), which binds to the E-repeat sequence of Xist and restricts the localization of XIST to the X-territory during XCI (Pandya-Jones et al., 2020; Ridings-Figueroa et al., 2017; Sunwoo et al., 2017). Immunostaining for CIZ1 in naïve UCLA1 hESCs showed a strong enrichment on the Xd in most cells (Fig 7a), suggesting that XIST exploits CIZ1 for XCI and XCD.

SPEN, also known as SMRT and HDAC associated repressor protein (SHARP), binds the 5' A-repeat sequence of mouse Xist and is the key epigenetic repressor of the XCI process in mouse embryos and mouse ESCs (Dossin et al., 2020; McHugh et al., 2015). SPEN mediates gene silencing in XCI by interacting with co-repressors and activating HDAC3(Dossin et al., 2020; McHugh et al., 2015; Monfort et al., 2015). To examine whether SPEN is required for gene repression on the Xd and on autosomes, we used RNAi to knockdown SPEN in female naïve UCLA1 hESCs (Fig 7b, Fig S6c). As a control, we also depleted SPEN in male naïve WIN1 hESCs (Fig 7b, Fig S6c). SPEN knockdown resulted in an increase of global transcript levels from the X chromosome in naïve female cells and a much smaller increase in naïve male cells (Fig 7c,d, S6d-f). The female-biased upregulation of X-linked genes was not associated with major changes in steady-state XIST levels (Fig S6g) or in the expression of naïve pluripotency-related genes (Fig S6h) and affected genes subject to XCD more strongly than XCD escapees (Fig 7e). Thus, similar to XCI, SPEN is necessary for XCD, suggesting that XIST mediates XCD through its interaction with SPEN. The slight upregulation of X-linked genes upon SPEN depletion in naïve male hESCs

(Fig 7d,e) suggests that SPEN also regulates gene expression on the active X chromosome, consistent with a broader role of SPEN in gene regulation.

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

Discussion

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

Our findings reveal that XIST can generate functionally different outputs of gene expression as human development proceeds: first dampening and later silencing, and provide first insights into why XIST mediates dampening in one developmental stage and silencing in another. The remarkable plasticity of XIST's ability to regulate gene expression provides the

exciting possibility that nuclear compartments with a specific gene expression output can be rationally engineered. To our knowledge, XIST is the first lncRNA in which the same transcript can produce two distinct regulatory outputs during development. Prior work uncovered opposing gene regulatory roles for the lncRNA locus Haunt, yet these distinct roles arise from the function of the lncRNA transcript on one hand and an enhancer-like function of the genomic locus on the other hand (Yin et al., 2015). It is notable that Xist is normally not expressed in undifferentiated naïve mouse ESCs, yet the ectopic induction of Xist in these cells enables XCI and not XCD, suggesting that it is not simply the developmental state that determines the gene regulatory output of XIST. Therefore, it will be important to decipher whether the sequence of the RNA, which differs between mouse and human even within the conserved repeat sequence regions (Brockdorff et al., 1991; Nesterova et al., 2001), contributes to the differential ability of XIST in human versus mouse naïve hPSCs and pre-implantation embryos.

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

Our high-resolution mapping of XIST's chromatin association in naïve hPSCs showed that XIST localizes over the entire Xd, as seen on the Xi. However, despite this extensive spread, we found striking differences in XIST accumulation on the Xd and Xi. First, XIST enriches in different regions on the inactive and dampened X chromosome. The spatial organization of the X chromosome may contribute to the distinct enrichment pattern of XIST, since the differences in

XIST localization are linked to the macro-satellite repeat *DXZ4* that partitions the Xi into two spatial superdomains (Bonora et al., 2018; Rao et al., 2014). This conclusion is consistent with prior observations that Xist exploits the spatial organization of the X chromosome for the initial spread across the X chromosome during XCI initiation (Engreitz et al., 2013; Simon et al., 2013). Second, we observed reduced levels of XIST on the Xd compared to the Xi. Since reduced accumulation of XIST in the X-territory is associated with XCD, our results are consistent with a model where the concentration of XIST within the X-chromosome territory is the critical determinant for the magnitude of XIST-mediated gene repression (Fig 7I).

The reduced concentration of XIST on the Xd versus the Xi may in turn reduce the recruitment of effector proteins to the Xd, leading to XCD instead of XCI. Xist and its binding partners have been proposed to execute XCI via phase separation (Cerase et al., 2019; Pandya-Jones et al., 2020), which might be hindered by lower concentrations. Alternatively, proteins involved in regulating XIST localization and function may be present at different levels or differentially modulated in cells with XCD and XCI. For instance, the absence or decreased level of proteins such as PTBP1, MATR3, TDP-43, SAF-A, CIZ1, and CELF1, all involved in regulating the localization of mouse Xist (Chu et al., 2015; McHugh et al., 2015; Minajigi et al., 2015), or the modulation of their function, as for instance through posttranslational modifications, may alter XIST spreading and it's ability to regulate gene expression.

Consistent with prior observations (Patrat et al., 2020; Sahakyan et al., 2017a; Vallot et al., 2013), our imaging data revealed a dispersed cloud of XIST surrounding the Xd in human pre-implantation embryos and naïve hPSCs. Our genomic approach uncovered specific autosomal regions to which XIST localizes in cells with XCD. These results support the possibility that XIST association with chromatin beyond the X-chromosome territory results in a reduced local concentration on the Xd in naïve hPSCs. It is conceivable that XIST is lost from the X chromosome

and associates with chromatin regions on nearby chromosomes. Alternatively, autosomal regions may effectively compete for the capture of XIST and thereby reduce the accumulation of the RNA on the X chromosome. Regardless of the mechanism, the repression of autosomal genes by XIST requires SPEN and autosomal XIST localization occurs at genic regions that contain L2 and MIR transposable elements, as seen on the X chromosome. These observations indicate that autosomal gene repression by XIST follows similar principles as on the X chromosome.

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

The finding that XIST localizes to specific autosomal regions was surprising as Xist, in the mouse system, has always been shown to be confined to the X chromosome from which it is transcribed (Brockdorff, 2019; Jonkers et al., 2008). Multiple studies have reported that the ectopic expression of Xist/XIST from an autosome is sufficient to induce autosomal gene silencing in undifferentiated mouse ESCs. However, to our knowledge, autosomal gene regulation by endogenously expressed XIST has never been reported. Intriguingly, autosomal target genes of XIST are enriched for functions in developmental regulation. By comparing gene expression levels between females and male naïve hPSCs and human pre-implantation embryos, we discovered that 87 autosomal genes are more lowly expressed in female cells compared to males and also regulated by XIST. Although the differences in autosomal gene expression are relatively small between male and female cells or in female cells with and without XIST, they are consistently found in several comparisons. Further studies are necessary to define the consequences of the female-specific downregulation of autosomal genes. Overall, the small but

consistent downregulation of autosomal genes by XIST provides an unexpected and exciting new role for XIST during early human development.

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

In summary, with the demonstration that XIST mediates XCD, our work enables future studies of the role of dosage compensation by XCD in human pre-implantation embryos and the human germ line. Moreover, it raises the question of whether the autosomal gene regulation by XIST is critical for early human development and the differentiation of the germ line.

Acknowledgments

We are grateful to Robin Mckee, Yolanda Markaki, Amanda Collier, Anna Afasizheva and Konstantinos Chronis for help with experiments. We would like to thank the Guttman Lab at Caltech for providing XIST probes for RAP-seq experiments. We would like to thank Austin Smith and Rudolf Jaenisch labs, for kindly providing HNES ans WIN1 lines respectively. We would also like to thank the microscopy cores at the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Center (BSCRC) for help with imaging, the Technology Center for Genomics and Bioinformatics at the UCLA Johnson Comprehensive Cancer Center (JCCC) and the Next Generation Sequencing core at BSCRC for help with genomics approaches.

Author contributions

A.S., I.D. and K.P. conceptualized the project. A.S. and T.C. generated cell lines, conducted imaging experiments and bulk RNA-seq. T.C. conducted single cell RNA-seq, XIST KO, and siSPEN. A.S. and S.T. conducted RAP-seq experiments. D.F. analyzed FISH signal of THOC2 and UTX. J.M. preformed sequential RNA+DNA FISH. I.D. conducted all computational analysis. I.D. and K.P. designed the experiments, interpreted the data, contributed towards methodology and model creation, and wrote the manuscript.

Declaration of interests

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

Figure 1

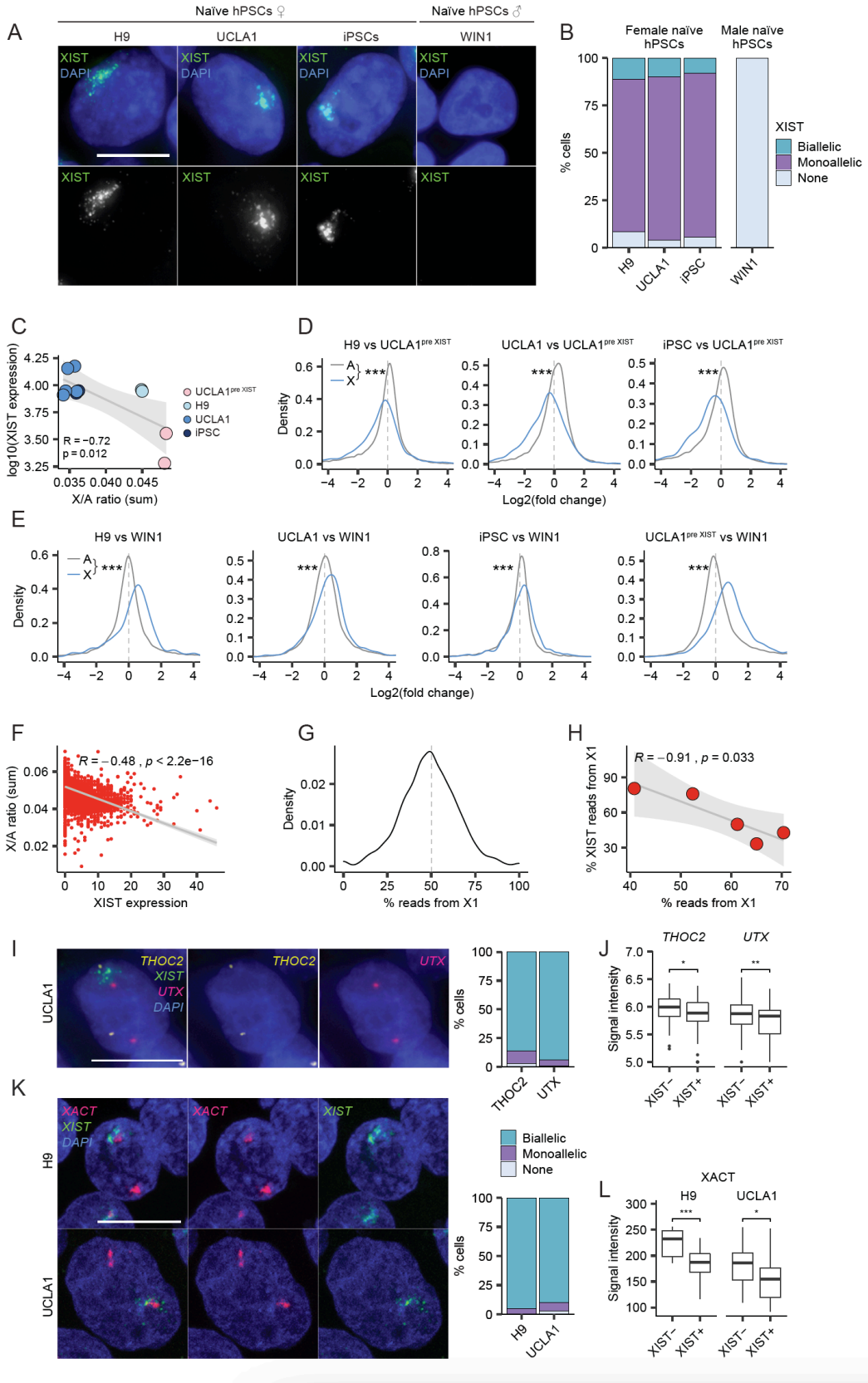

A
Naïve hPSCs ♀
Naïve hPSCs ♂
H9
UCLA1
iPSCs
WIN1
XIST
DAPI
B
Female naïve hPSCs
Male naïve hPSCs
% cells
XIST
Biallelic
Monoallelic
None
H9
UCLA1
iPSC
WIN1
C
log10(XIST expression)
X/A ratio (sum)
R = −0.72
p = 0.012
UCLA1pre XIST
H9
UCLA1
iPSC
D
H9 vs UCLA1pre XIST
UCLA1 vs UCLA1pre XIST
iPSC vs UCLA1pre XIST
Density
Log2(fold change)
E
H9 vs WIN1
UCLA1 vs WIN1
iPSC vs WIN1
UCLA1pre XIST vs WIN1
F
R = −0.48 , p < 2.2e−16
XIST expression
G
% reads from X1
H
R = −0.91 , p = 0.033
% XIST reads from X1
I
THOC2
XIST
UTX
DAPI
J
Signal intensity
XIST−
XIST+
K
XACT
L

Figure 1: XIST expression in naïve hPSCs correlates with XCD

A) Representative FISH images of XIST RNA for female naïve hPSC lines (UCLA1 and H9), female naïve human induced pluripotent stem cells (hiPSCs), and male naïve hPSC lines (WIN1) detecting XIST cloud (green), DAPI (blue). Scale bar = 10 microns.

B) Quantification of XIST RNA cloud in naïve female and male hPSCs (biallelic with equal XIST cloud on both X chromosomes, monoallelic with XIST cloud on one X chromosome, and none with no detected XIST cloud).

C) Scatter plot comparing the X/A ratios of the sum of read counts and XIST expression in each cell lines. Pearson correlation (R) and p-value are shown on the top.

D) Density plot of gene expression changes between a late XIST+ female naïve hPSCs (H9, UCLA1, iPSCs) and an early XIST- hPSC state (UCLA1pre XIST). X-linked genes are in blue, autosomal genes are in grey. Dashed line represents x=0 was added as a reference for no changes in gene expression. Wilcoxon test p-values are shown on top (* $P<0.05$, ** $P<0.01$, *** $P<0.001$).

E) Same as in (D) comparing between a late XIST+ female naïve hPSCs (H9, UCLA1, iPSCs) or an early XIST- female naïve hPSC (UCLA1pre XIST) with male naïve hPSCs (WIN1).

F) For single cell RNA-seq in female naïve hPSCs (H9), scatter plot showing the correlation between XIST expression in each cell and the X/A ratios (sum of read counts). Pearson correlation (R) and p-value are shown on the top.

G) Density plot showing the percentage of reads aligned to ChrX1 in each single cell of the female naïve hPSCs (H9). Dashed line represents x=0 was added as a reference for cells in which both X-chromosomes were active.

H) Percentage of reads aligned to ChrX1 for all X-linked genes (average, excluding XIST) vs XIST. Pearson correlation (R) and p-value are shown on the top.

I) Representative FISH images of XIST (green), THOC2 (yellow) and UTX (red) RNA in female naïve hPSC line (UCLA1). Scale bar = 10 microns. Quantification is shown on the right (biallelic

with transcription signal on both X chromosomes, monoallelic with signal on one X chromosome, and none with no detected transcription signal).

J) Signal intensity of THOC2 and UTX in female naïve hPSCs (UCLA1) under XIST cloud and away from XIST cloud. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

K) Representative FISH images of XIST (green) and XACT (red) for two female naïve hPSC line (UCLA1 and H9). Scale bar = 10 microns. Quantification of the RNA FISH patterns of XACT cloud is shown on the right (biallelic with XACT cloud on both X chromosomes, monoallelic with XACT cloud on one X chromosome, and none with no detected XACT cloud).

L) Average of top 10% intensities of XACT cloud signal per cell under XIST cloud and away from XIST cloud in H9 and UCLA1 hPSCs. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

Figure 2

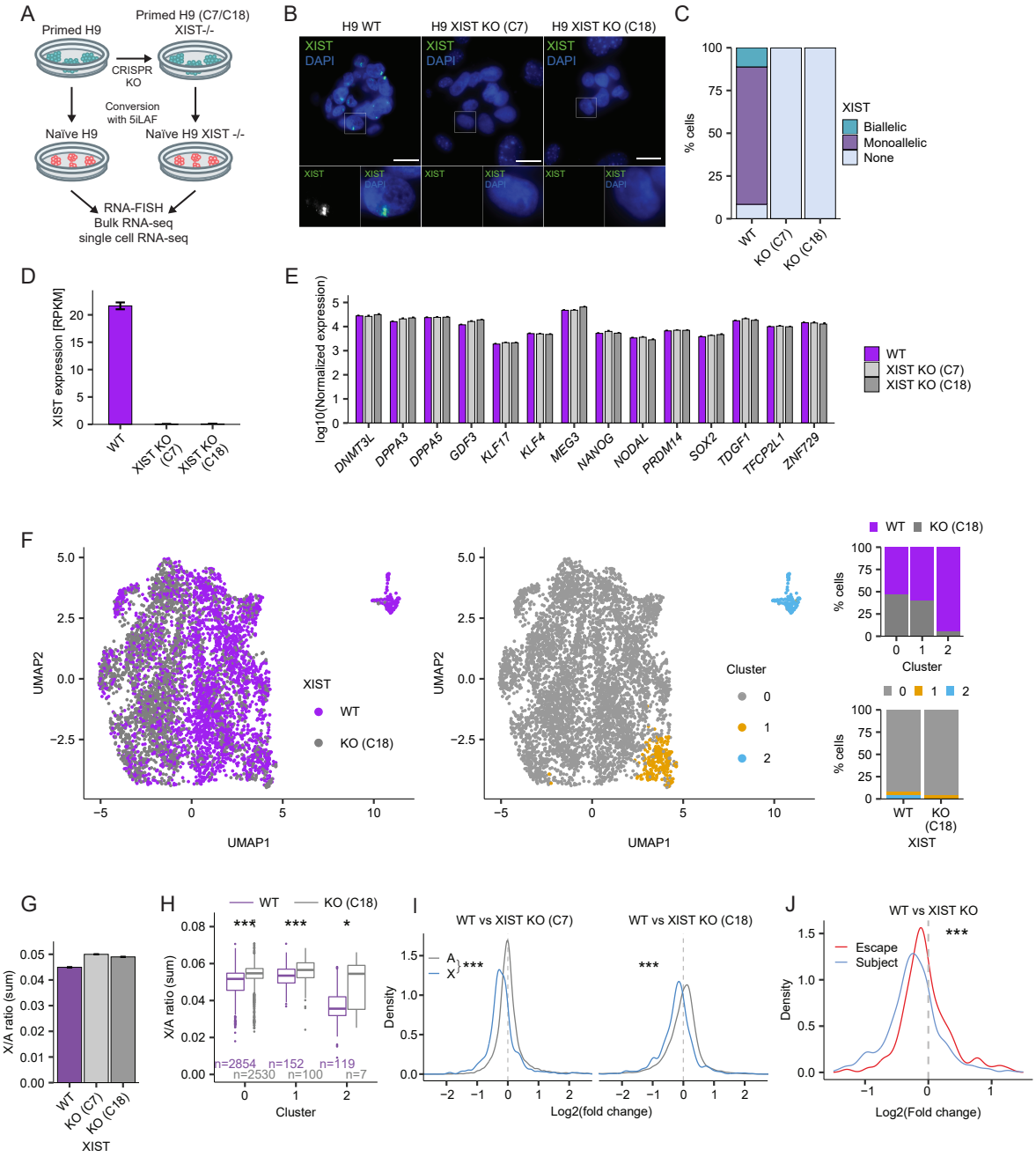

A
Primed H9
Primed H9 (C7/C18) XIST-/-
CRISPR KO
Conversion with 5iLAF
Naïve H9
Naïve H9 XIST -/-
RNA-FISH
Bulk RNA-seq
single cell RNA-seq
B
H9 WT
H9 XIST KO (C7)
H9 XIST KO (C18)
XIST
DAPI
C
% cells
XIST
Biallelic
Monoallelic
None
WT
KO (C7)
KO (C18)
D
XIST expression [RPKM]
WT
XIST KO (C7)
XIST KO (C18)
E
log10(Normalized expression)
DNMT3L
DPPA3
DPPA5
GDF3
KLF17
KLF4
MEG3
NANOG
NODAL
PRDM14
SOX2
TDGF1
TFCP2L1
ZNF729
WT
XIST KO (C7)
XIST KO (C18)
F
UMAP2
UMAP1
XIST
WT
KO (C18)
Cluster
0
1
2
% cells
G
X/A ratio (sum)
XIST
H
X/A ratio (sum)
Cluster
n=2854
n=2530
n=152
n=100
n=119
n=7
I
WT vs XIST KO (C7)
WT vs XIST KO (C18)
Density
Log2(fold change)
J
WT vs XIST KO
Escape
Subject
Density
Log2(Fold change)

Figure 2: XIST KO in naïve hPSCs results in reactivation of the dampened X chromosome

A) Scheme of XIST deletion.

B) Representative FISH images of XIST RNA for female naïve hPSC lines (H9) and in two different XIST KO clones (clone 7 and 18 respectively). Scale bar = 20 microns.

C) Quantification of the number of cells in which XIST was detected.

D) XIST expression in the WT (purple) and two XIST KO clones (gray). Error bars corresponding to the SD.

E) Gene expression of known naïve and preimplantation markers in the WT (purple) and two XIST KO clones (grey). Error bars corresponding to the SD.

F) UMAP of single cell female naïve hPSC lines (H9) color by right: WT (purple) and KO (grey), or by clusters. The percentage of cells found in each group is shown on the right.

G) X/A ratios (sum read counts) for WT (purple) and two XIST KO clones (grey). Error bars corresponding to the SD.

H) X/A ratios (sum read counts) for single cell RNA-seq in WT (purple) and XIST KO (grey). Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001). The number of cells in each cluster is represented bellow.

I) Density plot of the log2 fold change of gene expression between WT and XIST KO for X-linked (blue) and autosomal (grey) genes. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001). Dashed line represents x=0 was added as a reference for genes with similar expression in the WT and XIST KO.

J) Density plot of the log2 fold change of gene expression (using bulk RNA-seq) between WT and XIST KO for genes that escape (red) or subject to (blue) XCD. Dashed line represents x=0 was added as a reference for no changes in gene expression. Wilcoxon test p-values are shown on top.

Figure 3

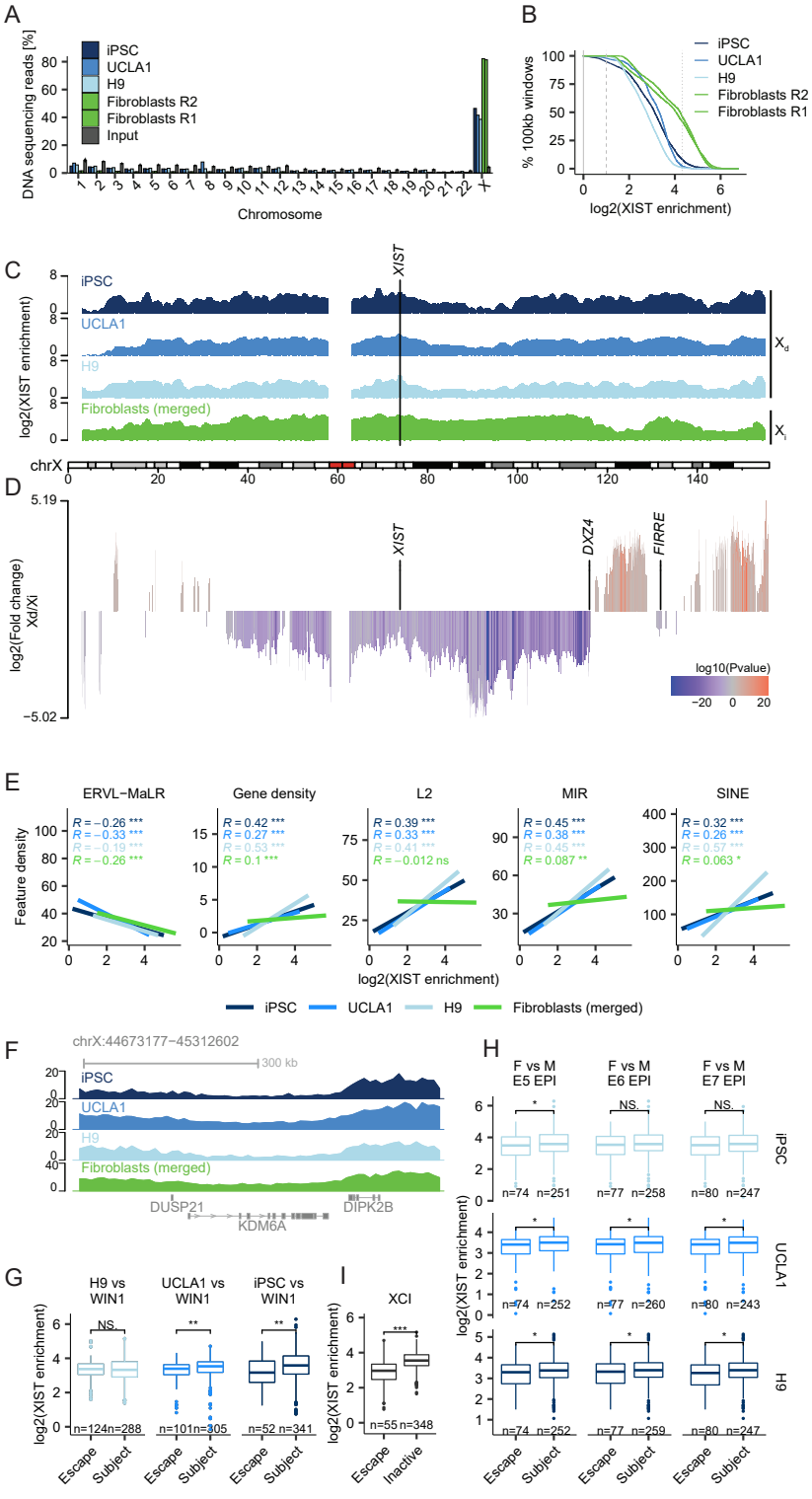

A
DNA sequencing reads [%]
iPSC
UCLA1
H9
Fibroblasts R2
Fibroblasts R1
Input
Chromosome
B
% 100kb windows
log2(XIST enrichment)
C
log2(XIST enrichment)
XIST
Fibroblasts (merged)
chrX
D
log2(Fold change) Xd/Xi
log10(Pvalue)
DXZ4
FIRRE
E
ERVL−MaLR
Gene density
L2
MIR
SINE
Feature density
log2(XIST enrichment)
F
chrX:44673177−45312602
300 kb
DUSP21
KDM6A
DIPK2B
G
H9 vs WIN1
UCLA1 vs WIN1
iPSC vs WIN1
Escape
Subject
I
XCI
Inactive
H
F vs M E5 EPI
F vs M E6 EPI
F vs M E7 EPI

Figure 3: XIST spreads across the dampened X chromosome

A) Percentage of DNA reads aligned to each chromosome for input genomic DNA (grey), somatic cells (green), female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue).

B) Cumulative distribution plot (greater than) showing XIST enrichment across all 100kb windows on the X chromosome. Color as in (A). Solid line mark x=0 and represent no difference between the input and pulldown. Dashed line mark x=log2(2) and represent two-fold enrichment. Dotted line showing mark x=log2(20) and represent 20-fold enrichment of pulldown over input.

C) XIST RAP-seq enrichment over input in somatic cells (green), female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue) along the X chromosome. The enrichment score at each window in the somatic cells was averaged across two replicates. Unmappable regions are masked.

D) Differential XIST enrichment along the X chromosome comparing naïve hESCs to somatic cells. Red represent genomic regions in which XIST was significantly more enriched in naïve hPSCs, blue represent genomic regions in which XIST was significantly more enriched in the somatic cells. Color represent log10(p-values) (using diffBind) and is shown as positive in naïve-enriched regions, or negative in somatic-enriched regions. Only regions with significant differences (p-value<0.01) are presented. XIST, DXZ4, and FIRRE genomic loci are marked.

E) Scatter plot showing linear regression line between XIST enrichment scores of each genomic region along the X chromosome, and the density of genes, ERVL-MaLR, L2 and SINE elements. Color as in (A). Pearson correlation (R) and p-value are shown on the top (* P<0.05, ** P<0.01, *** P<0.001).

F) XIST enrichment in naïve hESCs (H9 lightblue and UCLA1 in blue), naïve iPSC (dark blue), and somatic (green) across the genomic locus of KDM6A.

G) XIST enrichment in naïve hPSCs in genes subject to or escape XCD in each of the three female to male naïve hPSC comparisons. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

H) XIST enrichment in naïve hPSCs in genes subject to or escape XCD in pre-implantation Epiblasts. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

I) XIST enrichment in Fibroblasts in genes that escape or subject to XCI. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

Figure 4

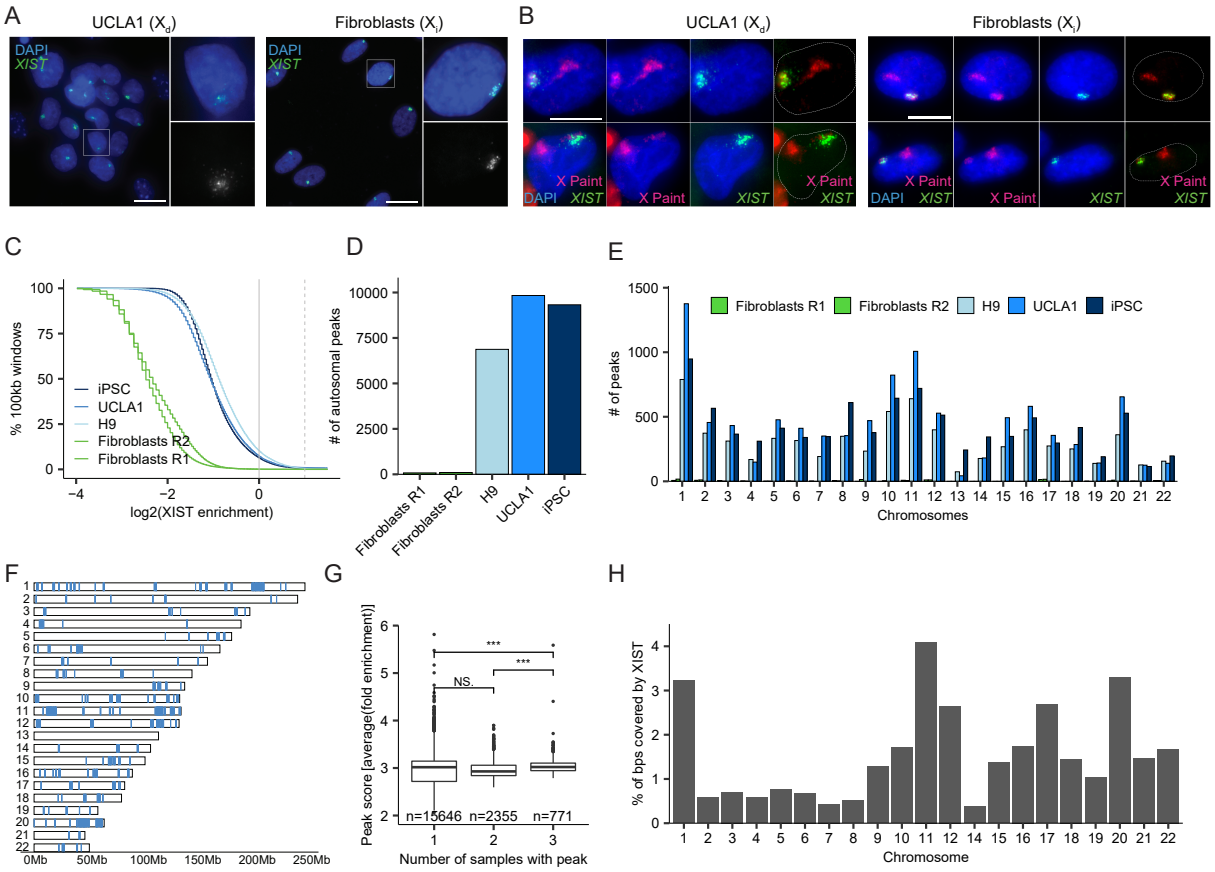


Figure 4: XIST spreads to specific autosomal regions in naïve hPSCs

A) Representative RNA FISH images for naïve hESCs (UCLA1, left) and somatic (NHDFs, right) detecting XIST cloud (green). Scale bar = 20 microns

B) Projections of XIST RNA cloud (green) on one of the two X-chromosomes (chromosome paint in pink) in female naïve hESCs (UCLA1) and somatic (NHDFs).

C) Cumulative distribution plot (greater than) showing XIST enrichment across all 100kb windows on autosomes for somatic cells (green), female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue). Solid line mark x=0 and represent similar enrichment of XIST in input and pulldown. Dashed line mark x=log2(2) and represent two-fold enrichment.

D) Number of significant peaks identified in autosomes for female somatic and naïve hPSCs. Colors as in (C).

E) Number of XIST peaks on each of the autosomes, colored as in (C). As comparisons, all peaks identified in the somatic cells are shown in green.

F) Karyotype plot showing naïve-conserved peaks along each autosome.

G) XIST enrichment scores (MACs log2(fold enrichment)) of autosomal peaks in the naïve hPSCs, shown for peaks that were identified in only one, two or three samples (x-axis). Wilcoxon test p-values are shown on top (* $P<0.05$, ** $P<0.01$, *** $P<0.001$). The number of peaks in each group is shown in the bottom

H) Percentage of bps in each chromosome cover by naïve-conserve peaks.

Figure 5

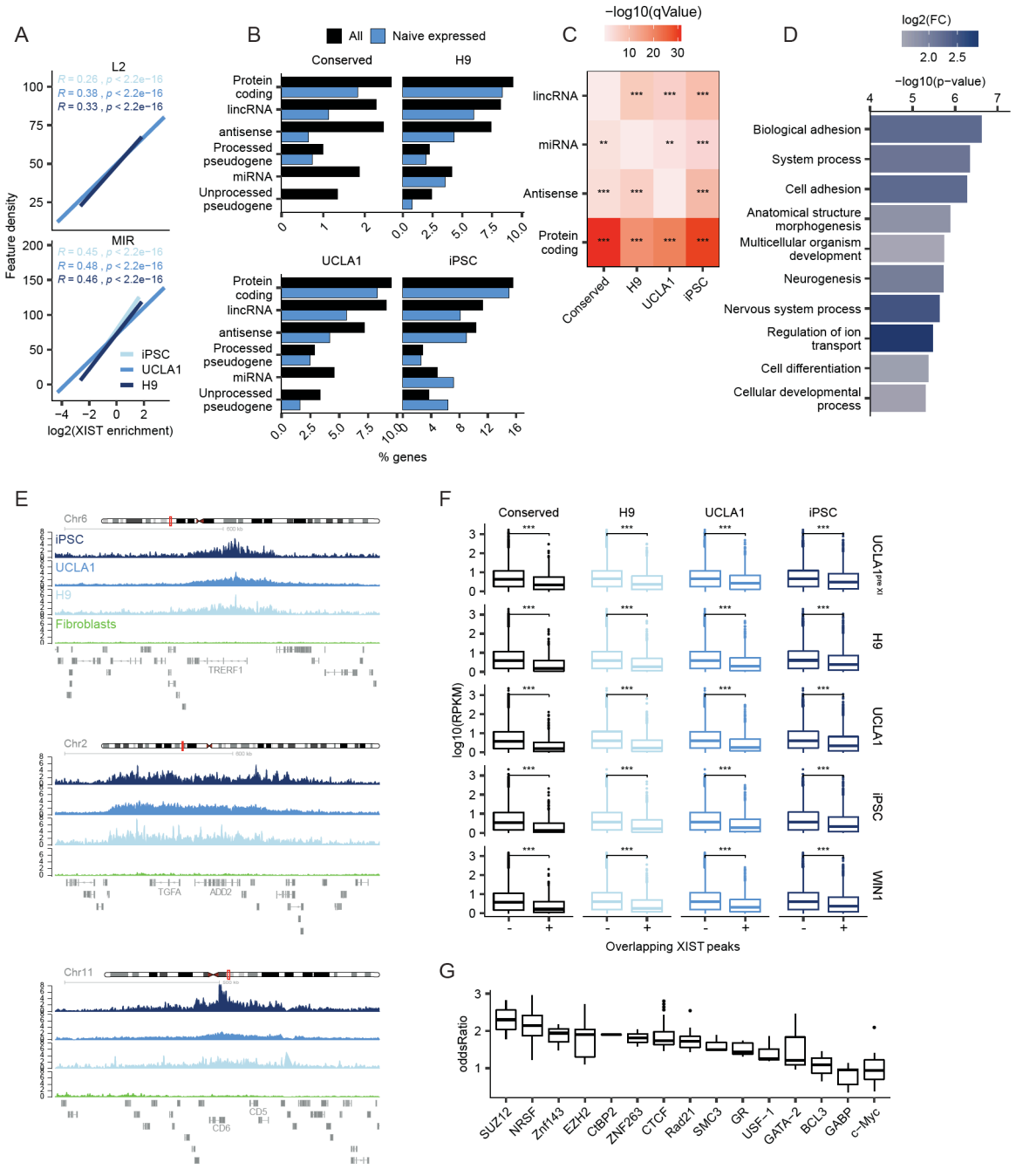

A
L2
R = 0.26 , p < 2.2e−16
R = 0.38 , p < 2.2e−16
R = 0.33 , p < 2.2e−16
MIR
R = 0.45 , p < 2.2e−16
R = 0.48 , p < 2.2e−16
R = 0.46 , p < 2.2e−16
iPSC
UCLA1
H9
Feature density
log2(XIST enrichment)
B
All
Naive expressed
Conserved
H9
UCLA1
iPSC
Protein coding
lincRNA
antisense
Processed pseudogene
miRNA
Unprocessed pseudogene
% genes
C
−log10(qValue)
lincRNA
miRNA
Antisense
Protein coding
Conserved
H9
UCLA1
iPSC
D
log2(FC)
−log10(p−value)
Biological adhesion
System process
Cell adhesion
Anatomical structure morphogenesis
Multicellular organism development
Neurogenesis
Nervous system process
Regulation of ion transport
Cell differentiation
Cellular developmental process
E
Chr6
iPSC
UCLA1
H9
Fibroblasts
TRERF1
Chr2
TGFA
ADD2
Chr11
CD5
CD6
F
Conserved
H9
UCLA1
iPSC
UCLA1pre XI
H9
UCLA1
iPSC
WIN1
log10(RPKM)
Overlapping XIST peaks
G
oddsRatio
SUZ12
NRSF
Znf143
EZH2
CtBP2
ZNF263
CTCF
Rad21
SMC3
GR
USF−1
GATA−2
BCL3
GABP
c−Myc

Figure 5: Features of XIST-enriched autosomal regions

A) Linear regression line between XIST enrichment scores along the autosomes, and the density of L2 and MIR elements, for female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue). Pearson correlation (R) and p-value are shown on the top.

B) Percentage of different gene types under XIST autosomal peaks, for all known genes (black), or genes expressed in naïve hPSCs (blue).

C) The enrichment (-log10(p-value)) of gene density for each gene type in XIST autosomal naïve peaks. Wilcoxon test p-values mark significant enrichment (* P<0.05, ** P<0.01, *** P<0.001).

D) The enrichment (-log10(p-value)) of GO term for genes detected under naïve-conserved autosomal peaks, colored by log2(observed/expected gene counts). Shown are the top enrichment terms (p-value<0.000005).

E) Three examples of XIST enrichment over autosomes. Showing XIST RAP-seq enrichment over input in somatic (green), female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue).

F) Gene expression in each of the female and male naïve hPSC (marked on the right) for genes overlapping (+) or not overlapping (-) with XIST peaks in each of the female naïve hPSCs (marked on top). Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

G) Enrichment (fisher's exact odd ratios) of the binding sites of TFs in diverse cell types and tissues, in XIST naïve-conserved autosomal peaks.

Figure 6

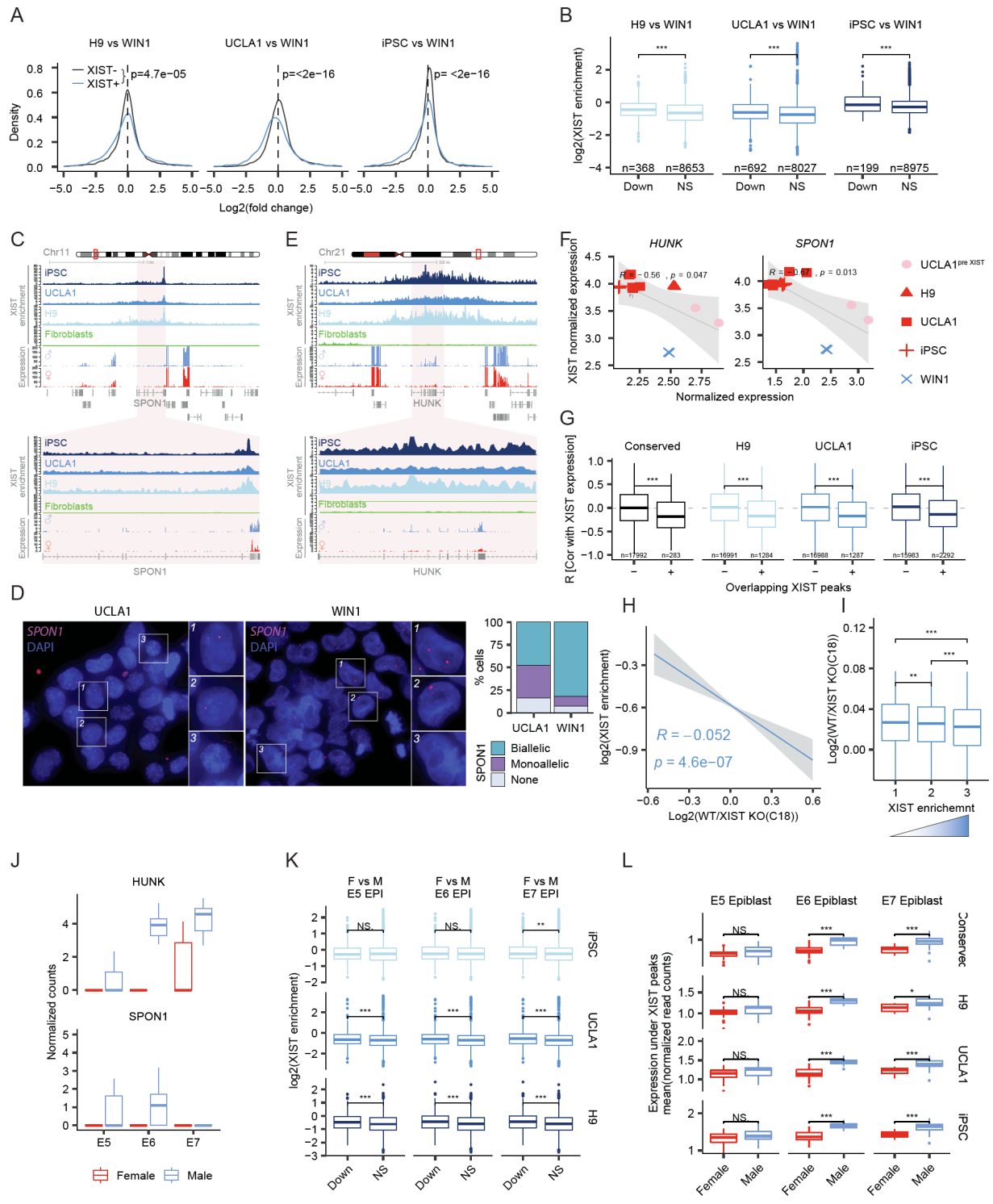

A
H9 vs WIN1
UCLA1 vs WIN1
iPSC vs WIN1
XIST-
XIST+
p=4.7e−05
p=<2e−16
p= <2e−16
Density
Log2(fold change)
B
H9 vs WIN1
UCLA1 vs WIN1
iPSC vs WIN1
log2(XIST enrichment)
n=368 n=8653
n=692 n=8027
n=199 n=8975
Down
NS
C
Chr11
iPSC
UCLA1
H9
Fibroblasts
XIST enrichment
Expression
SPON1
E
Chr21
HUNK
F
HUNK
SPON1
XIST normalized expression
Normalized expression
UCLA1pre XIST
H9
UCLA1
iPSC
WIN1
G
Conserved
H9
UCLA1
iPSC
R [Cor with XIST expression]
Overlapping XIST peaks
D
UCLA1
WIN1
SPON1
DAPI
% cells
Biallelic
Monoallelic
None
H
log2(XIST enrichment)
R = −0.052
p = 4.6e−07
Log2(WT/XIST KO(C18))
I
Log2(WT/XIST KO(C18))
XIST enrichemnt
J
HUNK
SPON1
Normalized counts
E5
E6
E7
Female
Male
K
F vs M
E5 EPI
E6 EPI
E7 EPI
iPSC
UCLA1
H9
log2(XIST enrichment)
Down
NS
L
E5 Epiblast
E6 Epiblast
E7 Epiblast
Conserved
H9
UCLA1
iPSC
Expression under XIST peaks
mean(normalized read counts)
Female
Male

Figure 6: XIST mediate downregulation of gene expression on autosomes

A) Density of autosomal gene expression changes (log2(fold change)) in the naïve female hPSCs (H9, UCLA1 and iPSCs) vs male naïve hPSCs for genes that are found overlapping XIST enriched peaks (in blue) or not overlapping (black). Wilcoxon test P-values indicate that gene expression is lower for genes under XIST peaks (* P<0.05, ** P<0.01, *** P<0.001). Dashed line mark x=0 and represent no differences in gene expression between females and males.

B) XIST scaled enrichment in each of the naïve hPSCs around autosomal genes that are significantly downregulated (Down) or don't change (NS) between female naïve hPSCs (H9, UCLA1 and iPSCs) and male naïve hPSCs (WIN1). Shown are RAP-seq enrichment scores of the corresponding female naïve hPSC in each comparison. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001). The number of genes in each group is shown below.

C) Example of XIST enrichment over the autosomal genes SPON1 (chromosome 11). Showing XIST RAP-seq enrichment over input in somatic (green), female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue) using 1Kb windows. RNA-seq read counts are shown on the bottom for males (blue), average naïve female hPSCs (red). Zoom in to the gene locus is shown below.

D) RNA FISH of SPON1 (pink) in female (UCLA1, left) and male (WIN1, right) naïve hPSCs. Quantification of the number of cells with no signal, monoallelic, or biallelic SPON1 signals is shown on the right.

E) Same as in (C) for the gene and HUNK (chromosome 21).

F) Scatter plot showing normalized expression of HUNK and SPON1 compared to XIST in each of the female and male naïve hPSCs. Pearson correlation (R) and p-value are shown on the top.

G) Boxplot of the correlations between XIST expression and genes overlapping (+) or not overlapping (-) XIST peaks. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

H) Scatter plot showing linear regression line between XIST scaled enrichment scores in female naïve hPSCs (H9) of each genomic region along the autosomes, and changes in gene expression (log2 fold change) between WT and XIST KO. Pearson correlation (R) and p-value are shown on the top.

I) Gene expression changes (log2(fold change)) between WT and XIST KO for three different XIST localization quantiles in naïve hPSCs (H9), marking low (1), middle (2), and high (3) XIST enrichment.

J) Single-cell normalized read counts of HUNK and SPON1 in females (red) and males (blue) epiblasts at different pre-implantation stages.

K) XIST scaled enrichment in naïve hPSCs around genes that are significantly downregulated (Down) or have no differences (NS) between female and male epiblasts at different pre-implantation stages. Wilcoxon test p-values are shown on top (* $P<0.05$, ** $P<0.01$, *** $P<0.001$).

L) Average expression (normalized read counts) of genes under XIST autosomal peaks in females (red) and males (blue) epiblasts at different pre-implantation stages. Autosomal peaks are defined in all naïve hPSC (conserved), H9, UCLA1 or iPSCs (on the right). Expression is shown for different stages pre-implantation (top).

Figure 7

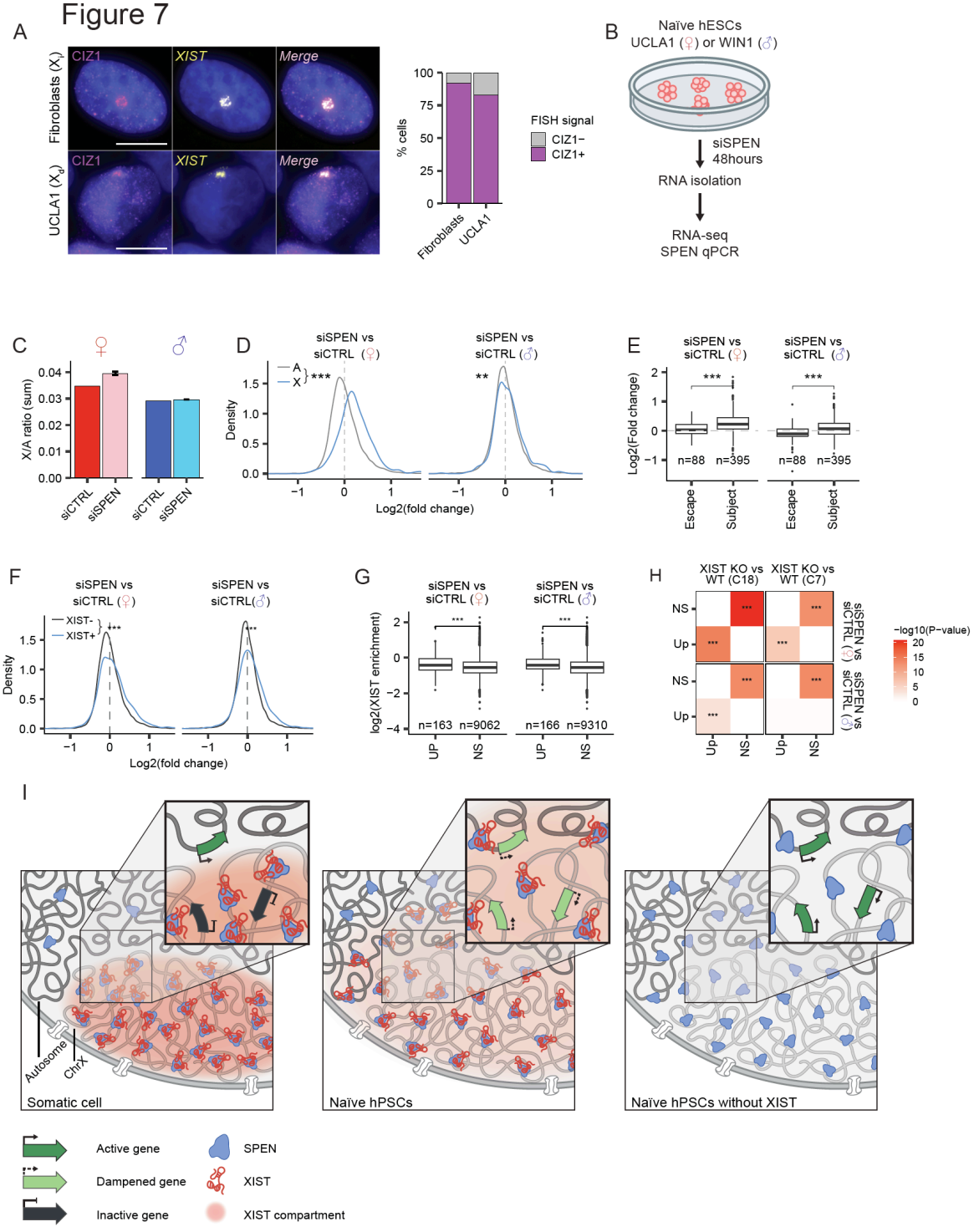

A
Fibroblasts (X)
UCLA1 (Xa)
CIZ1
XIST
Merge
% cells
FISH signal
CIZ1−
CIZ1+
Fibroblasts
UCLA1
B
Naïve hESCs
UCLA1 (♀) or WIN1 (♂)
siSPEN
48hours
RNA isolation
RNA-seq
SPEN qPCR
C
X/A ratio (sum)
siCTRL
siSPEN
D
siSPEN vs
siCTRL (♀)
siSPEN vs
siCTRL (♂)
Density
Log2(fold change)
E
Log2(Fold change)
n=88
n=395
Escape
Subject
F
XIST-
XIST+
G
log2(XIST enrichment)
n=163
n=9062
n=166
n=9310
UP
NS
H
XIST KO vs
WT (C18)
XIST KO vs
WT (C7)
−log10(P−value)
I
Autosome
ChrX
Somatic cell
Naïve hPSCs
Naïve hPSCs without XIST
Active gene
Dampened gene
Inactive gene
SPEN
XIST
XIST compartment

Figure 7: XIST mediates gene expression regulation through its interaction with SPEN

A) Immunoprecipitation of CIZ1 (pink) along with RNA FISH for XIST (yellow) in female somatic (top) and hESCs (UCLA1). Quantification of the percentage of cells showing CIZ1 are shown on the right.

B) Schematic representation of SPEN inhibition.

C) X/A ratios (sum read counts) for WT and siSPEN in female and male naïve hPSCs. Error bars corresponding to the SD.

D) Density of the log2 fold change of gene expression between siSPEN and WT (siCTRL) of X-linked (blue) and autosomal (grey) genes. Dashed red line represent no changes in gene expression between the siSPEN and siCTRL. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

E) Log2 fold change of genes expression between siSPEN and WT (siCTRL) for genes that escape or subject to XCD. Dashed red line represent no changes in gene expression between the siSPEN and siCTRL. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

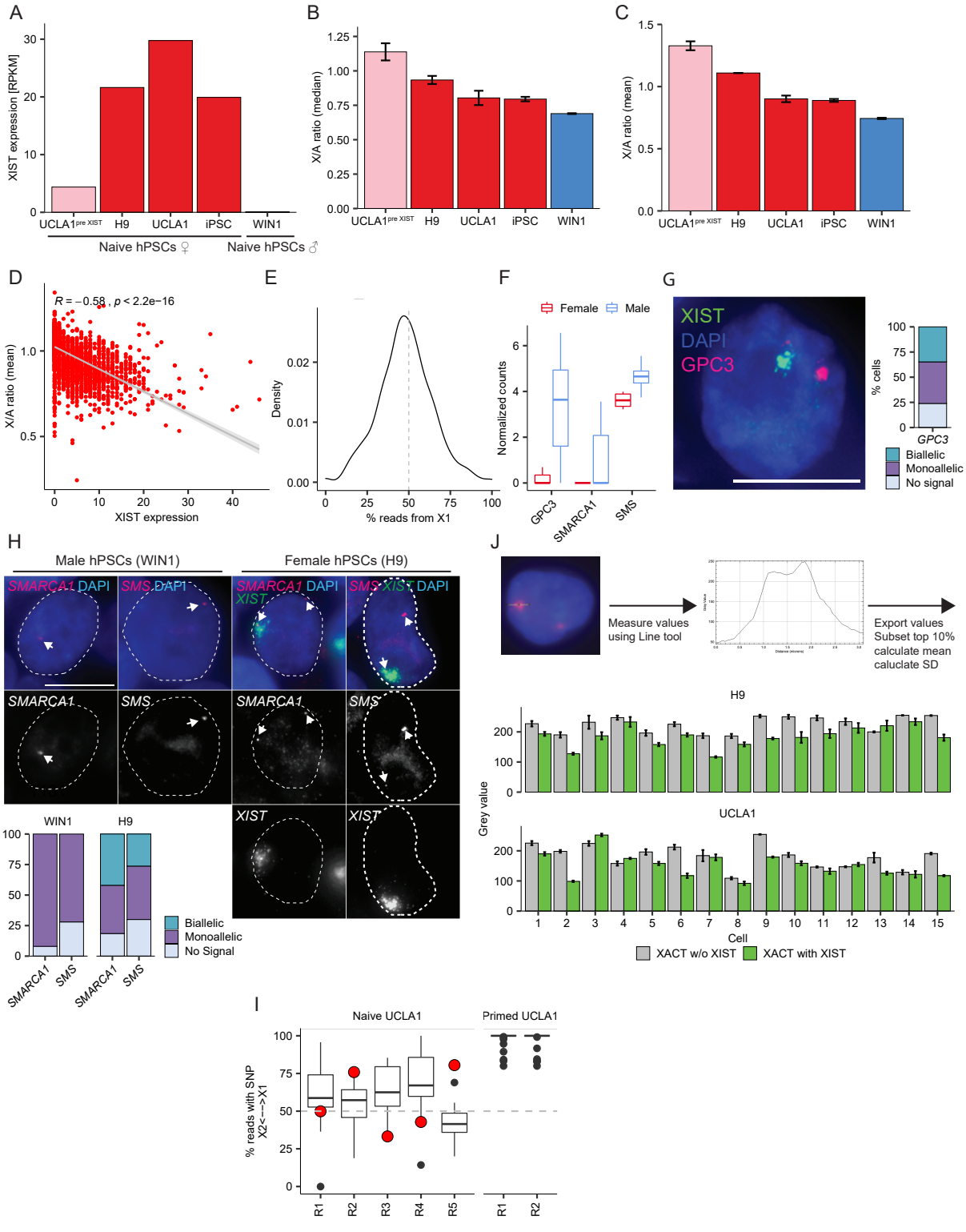
F) Density the log2 fold change of gene expression between siSPEN and WT (siCTRL) for genes that are found overlapping XIST enriched peaks (+, blue) or not overlapping (-, black). Wilcoxon test P-values indicate that genes under XIST peaks are upregulated upon siSPEN. Dashed line mark x=0 and represent no differences in gene expression between siSPEN and siCTRL.

G) XIST enrichment around autosomal genes that are significantly upregulated (Up) or don't change (NS) in siSPEN compared to WT. Shown are RAP-seq enrichment scores of the corresponding female naïve hPSC (UCLA1). Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001). The number of genes in each group is shown below.

H) Hypergeometric test (-log10(P-value) testing for significant overlap between the status of autosomal genes in each comparison. Asterisks represent significant p-values (* P<0.05, ** P<0.005, *** P<0.0005).

I) A model of XIST spreading and functions in female somatic (left), naïve (middle) and XIST KO (right). Zoom out shows XIST spreading and localization at each stage: during XCI, XIST spread over the X chromosome forming a dense cloud around the X territory. During XCD XIST spread beyond the X chromosome territory and into specific autosomal regions. Zoom in shows XIST role in regulating X and autosomal gene expression in each stage: In somatic cells XIST spreading on the X chromosome together with SPEN regulate gene inactivation. In naïve cells XIST binding along the X chromosome together with SPEN regulate gene dampening. In contrast to cells undergoing XCI, XIST KO in XCD results in the reactivation of the dampened X (right).

Figure S1


A
XIST expression [RPKM]
UCLA1 pre XIST
H9
UCLA1
iPSC
WIN1
Naive hPSCs ♀
Naive hPSCs ♂
B
X/A ratio (median)
C
X/A ratio (mean)
D
R = −0.58 , p < 2.2e−16
X/A ratio (mean)
XIST expression
E
Density
% reads from X1
F
Female
Male
Normalized counts
GPC3
SMARCA1
SMS
G
XIST
DAPI
GPC3
% cells
Biallelic
Monoallelic
No signal
H
Male hPSCs (WIN1)
Female hPSCs (H9)
SMARCA1 DAPI
SMS DAPI
SMARCA1 DAPI XIST
SMS XIST DAPI
SMARCA1
SMS
XIST
WIN1
H9
% cells
Biallelic
Monoallelic
No Signal
J
Measure values using Line tool
Export values Subset top 10% calculate mean caluclate SD
H9
UCLA1
Grey value
Cell
XACT w/o XIST
XACT with XIST
I
Naive UCLA1
Primed UCLA1
% reads with SNP X2<−−>X1
R1
R2
R3
R4
R5

Supplemental Figure S1: Characterization of XIST and XCD state in naïve hPSCs

A) XIST expression in early female naïve hPSCs (UCLA1pre XIST, pink), three late female naïve hPSC lines (H9, UCLA1 and iPSC, red), and male naïve hPSCs (WIN1) (blue).

B) Median expression (RPKM) X/A ratios. Colored as in (A). Error bars corresponding to the standard deviation (SD).

C) Average expression (RPKM) X/A ratios. Colored as in (A). Error bars corresponding to the SD.

D) For single cell RNA-seq in female naïve hPSCs (H9), scatter plot showing the correlation between XIST expression in each cell and the average X/A ratios. Linear regression line is shown in gray. Pearson correlation (R) and p-value are shown on the top.

E) Selecting only cells that express XIST, density plot showing the percentage of reads aligned to ChrX1 in each single cell of the female naïve hPSCs (H9). Dashed line represents x=0 was added as a reference for cells in which both X-chromosomes were active.

F) Gene expression of GPC3, SMARCA1 and SMS in female and male E7 epiblasts.

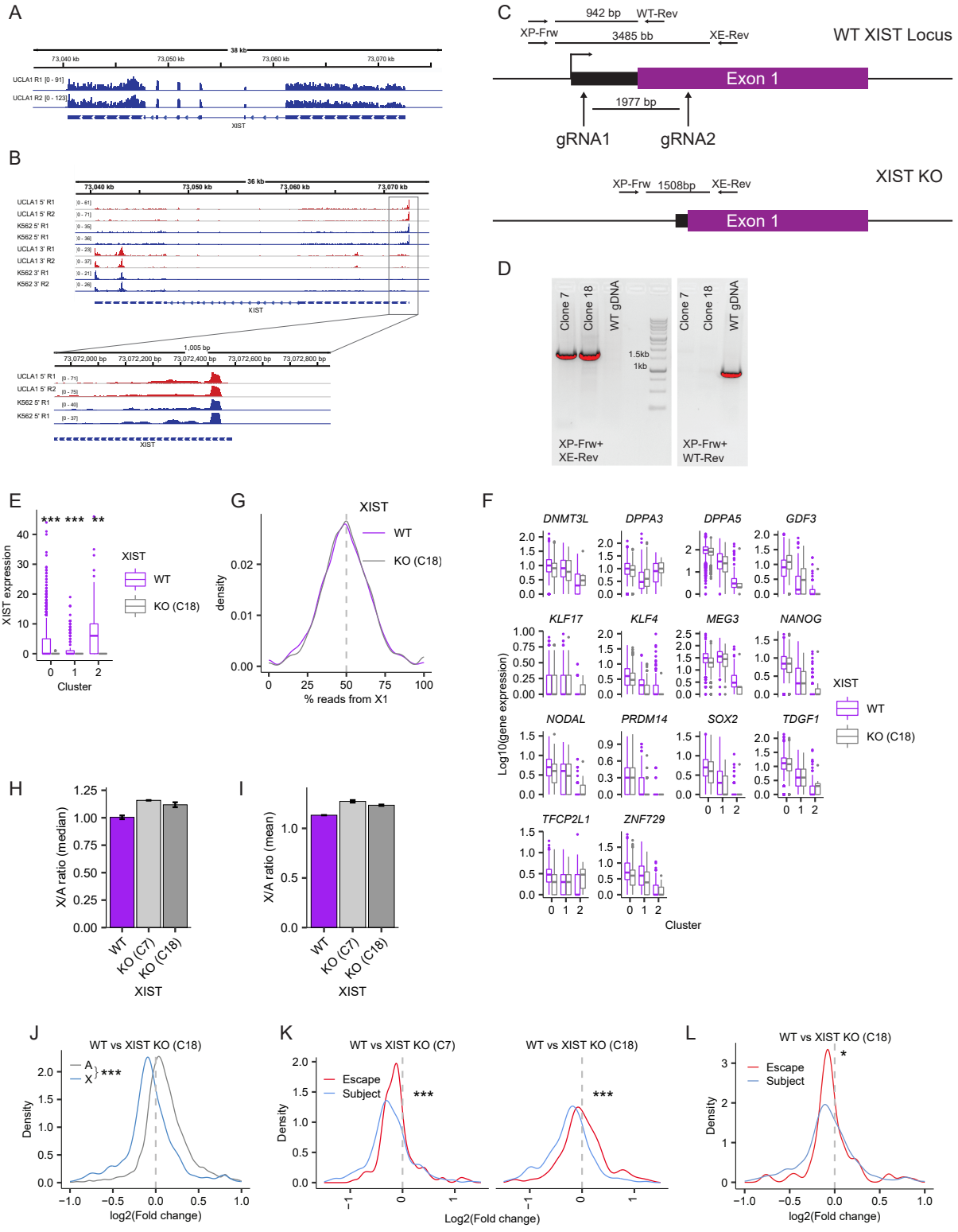
G) Representative RNA FISH image of XIST (green) and GPC3 (pink) in female naïve hPSCs (UCLA1). Quantification of the RNA FISH patterns of GPC3 is shown on the right. Scale bar = 10 microns.

H) Representative RNA FISH image of XIST, SMARCA1 (pink), or SMS (pink) in male (WIN1) and female (UCLA1) naïve hPSCs. Quantification of the RNA FISH patterns are shown on the bottom. Scale bar = 30 microns.

I) Boxplot of the percentage of reads aligned to ChrX1 in naïve and primed hPSCs (UCLA1). Red dots represent a SNP overlapping XIST locus (in cells where it was detected).

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

Figure S2


A
B
C
942 bp
WT-Rev
XP-Frw
3485 bb
XE-Rev
WT XIST Locus
Exon 1
1977 bp
gRNA1
gRNA2
XIST KO
XP-Frw
1508bp
XE-Rev
Exon 1
D
Clone 7
Clone 18
WT gDNA
1.5kb
1kb
XP-Frw+
XE-Rev
XP-Frw+
WT-Rev
E
XIST expression
Cluster
XIST
WT
KO (C18)
G
XIST
WT
KO (C18)
density
% reads from X1
F
DNMT3L
DPPA3
DPPA5
GDF3
KLF17
KLF4
MEG3
NANOG
NODAL
PRDM14
SOX2
TDGF1
TFCP2L1
ZNF729
Log10(gene expression)
Cluster
XIST
WT
KO (C18)
H
X/A ratio (median)
WT
KO (C7)
KO (C18)
XIST
I
X/A ratio (mean)
WT
KO (C7)
KO (C18)
XIST
J
WT vs XIST KO (C18)
A
X
Density
log2(Fold change)
K
WT vs XIST KO (C7)
WT vs XIST KO (C18)
Escape
Subject
Density
Log2(Fold change)
L
WT vs XIST KO (C18)
Escape
Subject
Density
log2(Fold change)

Supplemental Figure S2: Characterization of naïve hESCs with XIST KO

A) IGV plot showing RNA-seq reads in two replicates of WT female naïve hPSCs (H9) at the XIST locus.

B) Top: 5' (top four tracks) and 3' (bottom four tracks) RNA-seq at the XIST locus in female naïve hPSCs (UCLA1) in red, and K562 in blue. Zoom in to XIST transcription start site is shown on the bottom.

C) Schematic representation of XIST deletion.

D) PCR detection of deleted 2kb region in XIST locus. Upon deletion XP-Frw and XE-Rev primers shown in C, can only amplify 1.5 kb region from genomic DNA, while no amplicon is present in WT genomic DNA as the same primer set is not able to amplify 3.5kb region due to shorter extension time of PCR reaction. XP-Frw and WT-Rev primer set amplifies 942 bp region in WT genomic DNA, while WT-Rev complementary region is not present upon deletion, therefore no bend is present in homozygous deletions.

E) XIST expression in each cell of the WT (purple) and XIST KO (grey). Wilcoxon test p-values are shown on top (* $P<0.05$, ** $P<0.01$, *** $P<0.001$).

F) Gene expression of known naïve and preimplantation markers in each cell of WT (purple) and XIST KO (grey), shown separate for each cluster.

G) Density plot showing the percentage of reads from ChrX1 for each of the single cells of the WT (purple) and XIST KO (grey). Dashed line represents y=50 was added as a reference for biallelically expressed SNPs.

H) Median expression (RPKM) X/A ratios in WT (purple) and XIST KO (grey). Error bars corresponding to the SD.

I) Average expression (RPKM) X/A ratios in WT (purple) and XIST KO (grey). Error bars corresponding to the SD.

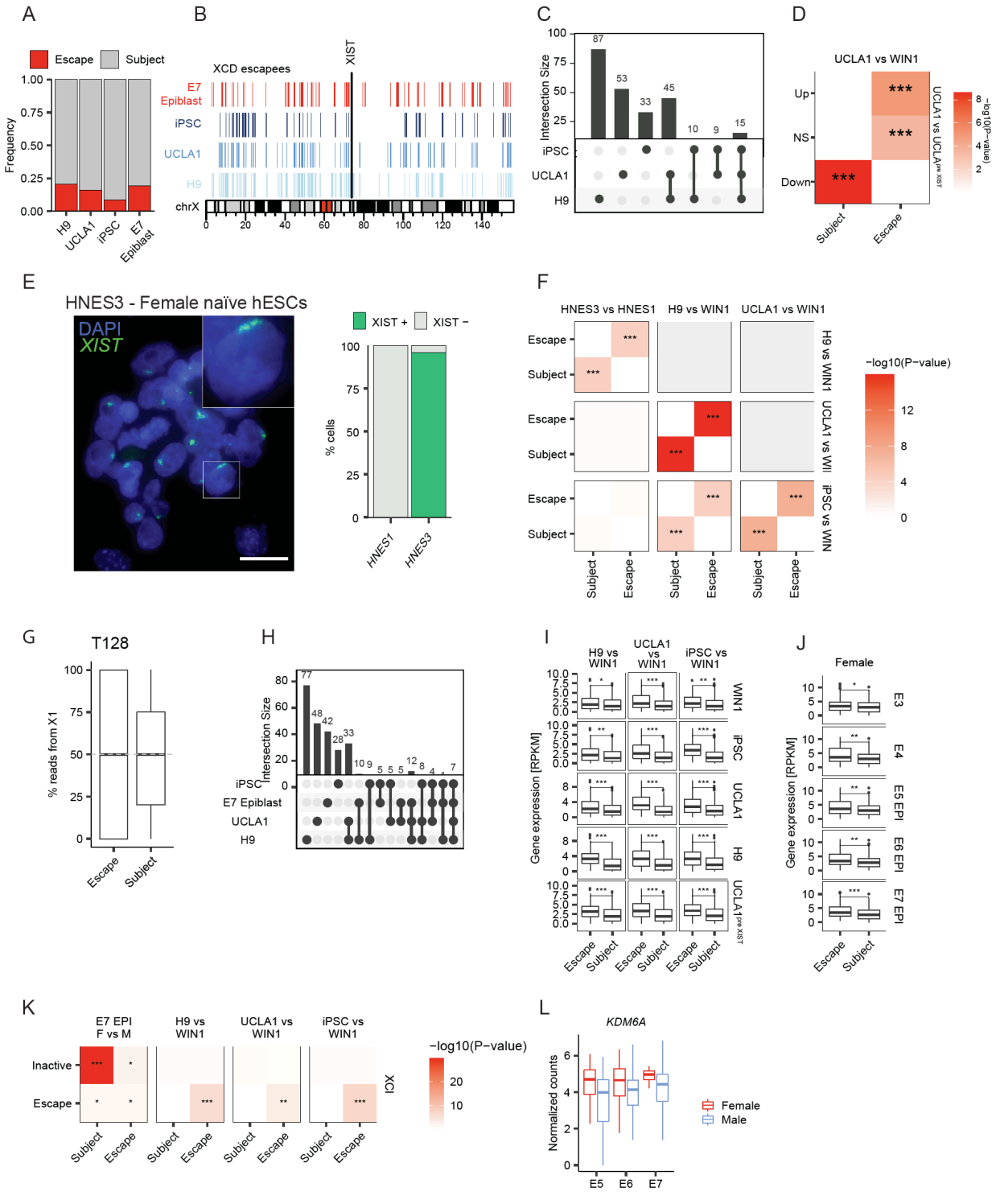
J) Density plot of the log2 fold change in gene expression between WT and XIST KO for X-linked (blue) and autosomal (grey) genes, using the average gene expression of single cells in cluster

0. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001). Dashed line represents x=0 was added as a reference for genes with similar expression in the WT and XIST KO.

K) Density plot of the log2 fold change in gene expression (using bulk RNA-seq) between WT and XIST KO. For genes that escape (in red) or subject to (blue) XCD. Dashed line represents x=0 was added as a reference for no changes in gene expression. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

L) Same as in (K) but for using the average gene expression of single cells in cluster 0.

Figure S3


A
Escape
Subject
Frequency
H9
UCLA1
iPSC
E7
Epiblast
B
XCD escapees
XIST
E7
Epiblast
iPSC
UCLA1
H9
chrX
C
Intersection Size
iPSC
UCLA1
H9
D
UCLA1 vs WIN1
Up
NS
Down
Subject
Escape
E
HNES3 - Female naïve hESCs
DAPI
XIST
XIST +
XIST −
% cells
HNES1
HNES3
F
HNES3 vs HNES1
H9 vs WIN1
UCLA1 vs WIN1
−log10(P−value)
G
T128
% reads from X1
Escape
Subject
H
Intersection Size
iPSC
E7 Epiblast
UCLA1
H9
I
H9 vs WIN1
UCLA1 vs WIN1
iPSC vs WIN1
Gene expression [RPKM]
J
Female
E3
E4
E5 EPI
E6 EPI
E7 EPI
K
E7 EPI F vs M
H9 vs WIN1
UCLA1 vs WIN1
iPSC vs WIN1
Inactive
Escape
−log10(P−value)
L
KDM6A
Normalized counts
Female
Male
E5
E6
E7

Supplemental Figure S3: Gene status in XCD and XCI

A) Frequency of genes the escape (red) or subject (grey) to XCD, defined by comparing female to male pre-implantation (E7) Epiblasts, or female naïve hPSCs (H9, UCLA1, iPSCs) to male naïve hPSC (WIN1).

B) XCD escapees plotted along the X-chromosome comparing female naïve hPSCs (H9, UCLA1, iPSCs) to male naïve hPSC (WIN1) or female to male pre-implantation (E7) Epiblasts.

C) UpSet diagram showing the intersection between XCD escapees as define in each female naïve hPSCs (H9, UCLA1, iPSCs) to male naïve hPSC (WIN1) comparison.

D) Hypergeometric test (-log10(P-value) testing for significant overlap between the status of XCD in each comparison. Asterisks represent significant p-values (* P<0.05, ** P<0.005, *** P<0.0005).

E) Representative FISH images of XIST RNA (green) for female naïve hPSC lines (HNES3). Quantification of the number of cells in which XIST was detected in female (HNES3) and male (HNES1) naïve hPSC lines is shown on the right.

F) Similar to (D) for genes that escape or subject to XCD between all female to male comparisons.

G) For the scRNA-seq (WT H9), the percentage of reads aligned to ChrX1 overlapping escape and subject to XCD genes (based on female naïve hPSCs H9 vs male naïve hPSCs WIN1). Dashed line represents y=50 was added as a reference for biallelically expressed SNPs.

H) UpSet diagram showing the intersection between XCD escapees as define in each female naïve hPSCs (H9, UCLA1, iPSCs) to male naïve hPSC (WIN1) comparison or between female to male pre-implantation (E7) Epiblasts.

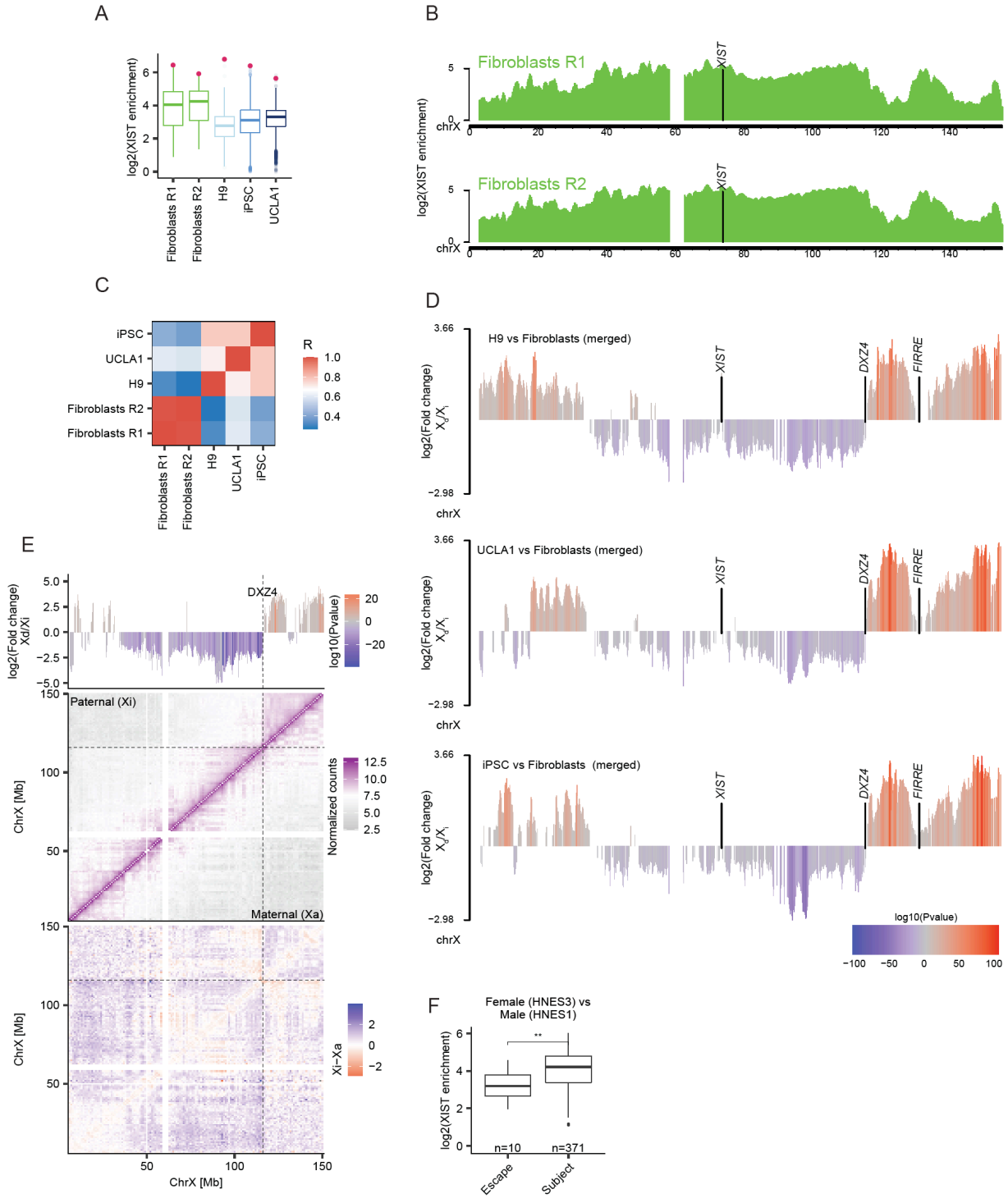
I) Gene expression of XCD escapees and subject (based on the three comparisons) in female and male naïve hPSCs.

J) Gene expression of XCD escapees and subject (E7 Epiblasts comparison) in female E7 Epiblasts (showing the average gene expression for each gene over all cells).

K) Similar to (D) comparing genes that escape XCD and XCI.

L) Gene expression in females and males in E5, E6, and E7 Epiblasts of the XCD escape KDM6A (defined as escapee in UCLA1 and E7 Epiblasts).

Figure S4


A
log2(XIST enrichment)
Fibroblasts R1
Fibroblasts R2
H9
iPSC
UCLA1
B
Fibroblasts R1
Fibroblasts R2
XIST
log2(XIST enrichment)
chrX
C
R
1.0
0.8
0.6
0.4
D
H9 vs Fibroblasts (merged)
UCLA1 vs Fibroblasts (merged)
iPSC vs Fibroblasts (merged)
log2(Fold change) Xd/Xi
DXZ4
FIRRE
chrX
log10(Pvalue)
E
log2(Fold change) Xd/Xi
log10(Pvalue)
Paternal (Xi)
Maternal (Xa)
ChrX [Mb]
Normalized counts
Xi−Xa
F
Female (HNES3) vs Male (HNES1)
log2(XIST enrichment)
n=10
n=371
Escape
Subject

Supplemental Figure S4: Comparison of XIST spreading in different cells

A) Boxplot of XIST RAP-seq enrichment over input for somatic cells (green), female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue), using 100Kb windows every 25Kb along the X chromosome. Enrichment at the XIST locus is marked in pink circles on top.

B) XIST RAP-seq enrichment over input for the somatic cell replicates along the X chromosome. Unmappable regions are masked.

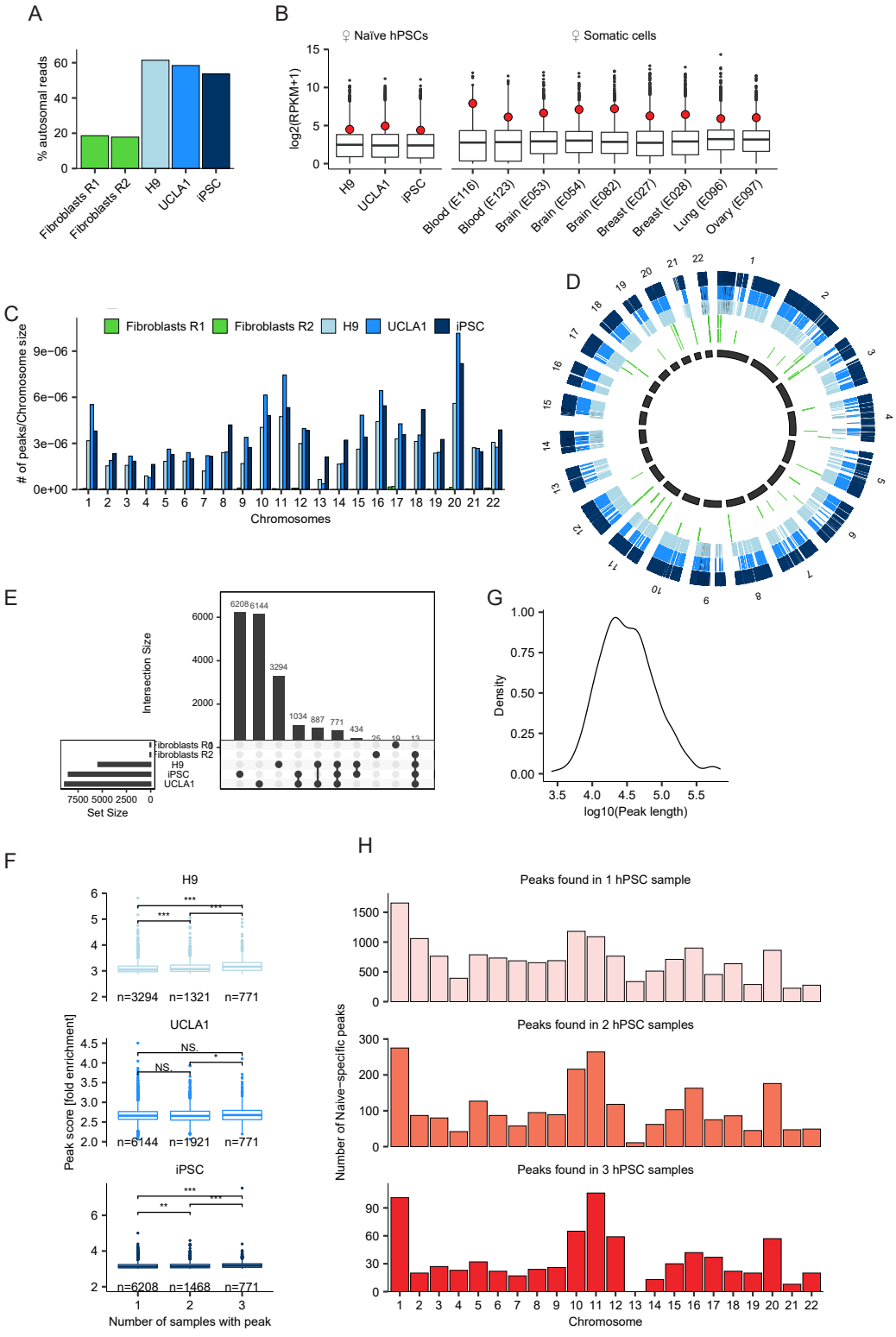
C) Heatmap showing Pearson correlation between XIST enrichment of each 100Kb windows every 25Kb along the X chromosome in each sample.

D) Differential XIST enrichment along the X chromosome comparing each of the naïve hPSC samples to somatic cells. Red represent genomic regions in which XIST was significantly more enriched in naïve hPSCs, blue represent genomic regions in which XIST was significantly more enriched in the somatic cells. Color represent log10(p-values) (using DESeq) and is shown as positive in naïve-enriched regions, or negative in somatic-enriched regions. Only regions with significant differences (p-value<0.01) are presented. XIST, DXZ4, and FIRRE genomic loci are marked.

E) Top: same as (E), using diffBind. DXZ4 genomic locus is marked. Middle: Hi-C normalized contacts in the paternal (Xi) (top half) and maternal (Xa) (bottom half) X chromosomes of GM12878. Bottom: Hi-C normalized contact differences between the paternal and maternal X chromosome.

F) XIST enrichment in naïve hESCs around genes subject to or escape XCD in female (HNES3) vs male (HNES1) naïve hPSC. Wilcoxon test p-values are shown on top (* P<0.05, ** P<0.01, *** P<0.001).

Figure S5


A
% autosomal reads
Fibroblasts R1
Fibroblasts R2
H9
UCLA1
iPSC
B
♀ Naïve hPSCs
♀ Somatic cells
log2(RPKM+1)
Blood (E116)
Blood (E123)
Brain (E053)
Brain (E054)
Brain (E082)
Breast (E027)
Breast (E028)
Lung (E096)
Ovary (E097)
C
# of peaks/Chromosome size
Chromosomes
D
E
Intersection Size
Set Size
G
Density
log10(Peak length)
F
H9
UCLA1
iPSC
Peak score [fold enrichment]
Number of samples with peak
H
Peaks found in 1 hPSC sample
Peaks found in 2 hPSC samples
Peaks found in 3 hPSC samples
Number of Naive−specific peaks
Chromosome

Supplemental Figure S5: Characterization of XIST enrichment on autosomes

A) Percentage of reads aligned to autosomes, for somatic cells (green), female naïve hESCs (H9 lightblue and UCLA1 in blue), and female naïve iPSC (dark blue).

B) Boxplot showing gene expression log2(RPKM) in naïve hPSCs and diverse female somatic tissues (extracted from the human Epigenome Roadmap). Pink represent XIST expression in each sample.

C) Number of significant peaks normalized by chromosome size. Colors as in (B).

D) XIST peaks across each autosome. Colors as in (B).

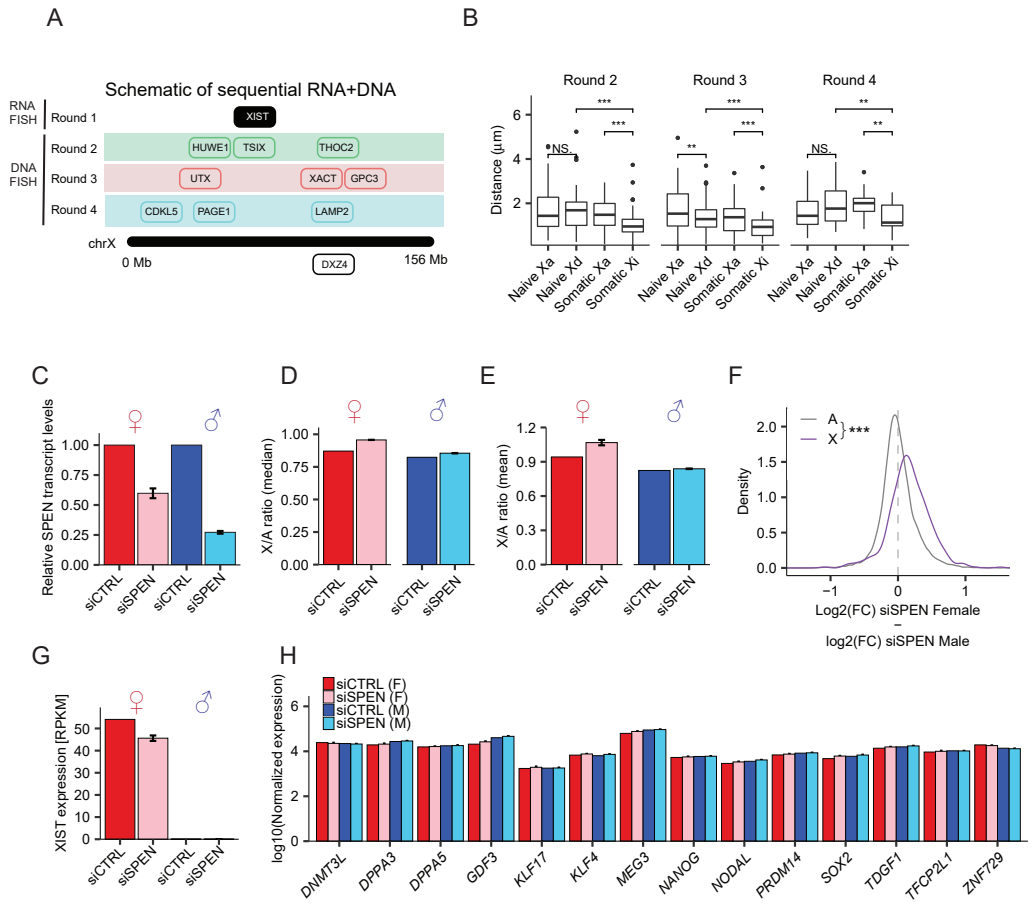
E) UpSet diagram showing the intersection between autosomal peaks detected in each female naïve hPSCs (H9, UCLA1, iPSCs) and somatic (Fibroblasts).

F) XIST enrichment scores (MACs log2(fold enrichment)) of autosomal peaks in each of the naïve hPSCs, shown for peaks that were identified in only one, two or three samples (x-axis). Wilcoxon test p-values are shown on top (* $P<0.05$, ** $P<0.01$, *** $P<0.001$).

G) Density plot showing the length (in log10), of the autosomal naïve-conserved peaks.

H) Number of naïve peaks in each of the autosomes, for peaks that were identified in only one of the naïve hPSCs sample (top), two sample (middle), or in all three naïve hPSCs.

Figure S6



Supplemental Figure S6: Characterization of naïve hPSCs with siSPEN

A) Schematic representation of sequential RNA+DNA. RNA FISH targeting XIST are first imaged to identify the Xd and Xi in each cell. Then, three rounds of DNA FISH are preformed targeting three different genomic loci along the X chromosome.

B) Distance between the DNA probes in each round of the sequential RNA+DNA, in naïve hPSCs Xa and Xd, and somatic Xa and Xi. Asterisks represent significant p-values (* P<0.05, ** P<0.005, *** P<0.0005).

C) qPCR detection of SPEN mRNA expression in the knockdown using targeted siRNA relative to female and male WT. Error bars corresponding to the SD.

D) X/A ratios (median gene expression) in each sample. Error bars corresponding to the SD.

E) X/A ratios (average gene expression) in each sample. Error bars corresponding to the SD.

F) Density plot of the fold change in gene expression upon siSPEN in female compared to male. X-linked genes are in purple, autosomal genes are in grey. Dashed line represents similar effect of siSPEN to gene expression in female and male.

G) XIST expression in female and male WT and siSPEN.

H) Gene expression of known naïve and preimplantation markers in female and male WT and siSPEN.

Methods

Cell lines and culturing conditions

Human ESC/hiPSC lines used in this study include UCLA1 (46, XX), iPSC (46, XX), H9 (46, XX), WIN1 (46, XY), HNES1(46, XY) and HNES3 (46, XX). Primed hESCs/hiPSCs were cultured on inactivated mouse embryonic fibroblasts (MEFs) in hESC media, which is composed of 20% knockout serum replacement (KSR) (GIBCO, 10828-028), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 55mM 2-Mercaptoethanol (GIBCO, 21985-023), 10ng/mL recombinant human FGF basic (R&D systems, 233-FB), 1x Penicillin-Streptomycin (GIBCO, 15140-122) in DMEM/F12 media (Sigma, D8437). Primed hESCs and iPSCs were split every 5-6 days using Collagenase type IV (GIBCO, 17104-019). 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).

Induced pluripotent stem cells (iPSCs) were generated by reprogramming NHDFs to pluripotency using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (ThermoFisher) following manufacturer's user guide for feeder-dependent reprogramming of fibroblasts. Briefly, 100,000 NHDFs (Lonza lot #472033) were plated in 6-well plates, and 24-hours later subjected to an overnight transduction with CytoTuneTM vectors. Cell-culture medium was renewed daily with fibroblast medium for one week. Seven days post transduction cells were harvested by TrypLE Express Enzyme (Thermofisher) and plated on irradiated monolayer of mouse embryonic fibroblasts (MEFs) in fibroblast medium. After one day, the medium was changed to naïve (5iLAF) or primed (with 20% KSR) human embryonic stem cell medium.

Somatic cell lines used in this study included normal human dermal fibroblasts – NHDFs (46,XX) (Lonza lot #472033). NHDFs were cultured in a fibroblast media that is composed of 10% FBS (Life technologies, 10099141), 100mM L-Glutamine (GIBCO, 25030-081), 1x MEM Non-Essential Amino Acids (NEAA) (GIBCO, 11140-050), 55mM 2-Mercaptoethanol (GIBCO, 21985-

023) in DMEM (Sigma, D6429). Mycoplasma test (Lonza, LT07-418) was performed routinely to all cell lines used in this study.

NHES1 and NHES3 lines were kindly provided by Austin Smith lab. Those cell lines were maintained on irradiated mouse embryonic fibroblast (MEF) feeder cells in N2B27 supplemented with 1 μM PD0325901, 10 ng/ml human LIF, 2 μM Gö6983 (Tocris Bio-Techne, 2285), and 2 μM XAV939 (Tocris Bio-Techne, 3748). ROCK inhibitor (10μM; Y-27632, Millipore) was added for 24h after passaging with StemPro Accutase.

Conversion of primed to naïve hPSCs

A detailed description of the conversion of primed human ESC line UCLA1 to its naïve state has been described in Sahakyan et al., (Sahakyan et al., 2017a). Briefly, the primed ESC line UCLA1 was obtained from the Human Embryonic and Induced Pluripotent Stem Cell Core at UCLA and plated on irradiated MEFs as single cells in the presence of 10uM ROCK inhibitor Y-27632 in DMEM/F12 medium supplemented with 5% KnockOut Serum Replacement (KSR), 15% heat inactivated FBS, penicillin/streptomycin, nonessential amino acids, GlutaMAX, 0.1mM β-mercaptoethanol, and 10ng/ml FGF2. After two days the medium was switched to the 5iLAF naïve ESC medium (1:1 mixture of DMEM/F12 and Neurobasal supplemented with N2, B27, penicillin/streptomycin, nonessential amino acids, GlutaMAX, 0.5% KSR, 0.1mM β-mercaptoethanol, 50ug/ml bovine serum albumin, 20ng/ml rhLIF, 20ng/ml Activin A, 8ng/ml FGF2, 1uM MEK inhibitor PD0325901 0.5uM B-Raf inhibitor SB590885, 1uM GSK3-β inhibitor IM-12, 1uM Src inhibitor WH-4-023, and 10uM ROCK inhibitor Y-27632). The established naïve UCLA1 cells were maintained by passaging every 5-6 days with StemPro Accutase onto freshly plated monolayer of irradiated MEFs. UCLA1 and H9 hESCs were converted and cultured either in 5% CO2, 5% O2 (Hypoxia) at 37C or in 5% CO2 and atmospheric oxygen level (Normoxia) at 37C.

Bulk RNA sequencing

For Bulk RNA sequencing, cells were washed with DPBS and dissociated with accutase. Harvested cells were lysed using Trizol reagent (Life technologies #15596018) and RNA was isolated using Qiagen RNAeasy kit (cat. #74104) according to manufacturer's instructions. RNA-seq libraries were prepared using the TrueSeq Stranded mRNA Library Prep Kit (Illumina 20020594) according to manufacturer's instructions. For strand-specific RNA sequencing of the naïve UCLA1 R1, R2, and R5, or early naïve state (UCLA1 pre XIST) cells were harvested, washed with DPBS, and collected in Trizol (ThermoFisher). RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina) as described in (Sahakyan et al., 2017a).

Bulk RNA alignment and gene expression quantification

RNA-sequencing reads were trimmed using trim_galore (https://github.com/FelixKrueger/TrimGalore) with default parameters to remove the standard Illumina adaptor sequence. Reads were then mapped to the human genome (hg38 assembly) using HISAT2 (Kim et al., 2019) with default parameters. Reads with mapping quality less than 30 were removed using samtools (Li et al., 2009). Read counts for each gene were calculated using HTSeq using the following parameters "--format=bam --order=pos --stranded=reverse --minaqual=0 --type=exon --mode=union --idattr=gene_name" (Anders et al., 2015). Genes with low read count were removed (keeping genes with counts-per-million>=0.5 in at least two samples). Regularized log transformation (rlog) of each gene in each sample was calculated using DESeq2 (Love et al., 2014), and was used as normalized gene expression unless noted otherwise. In addition, Reads Per Kilobase per Million mapped reads (RPKM) values were calculated for each gene. Differential gene expression analysis was done using DESeq2 (Love et al., 2014).

For gene expression in ENCODE cells and tissues, RNA-seq data (RPKMs) was downloaded from NIH Roadmap Epigenomics Mapping Consortium (using both coding and non

coding RNAs) (Davis et al., 2018). Genes with RPKM=0 across all cell types were removed. Cell types were filtered to select only female samples resulting in Blood, Breast, Lung, Brain, and Ovary.

Single cell RNA sequencing (scRNA-seq)

For scRNA-seq, naïve hESCs (H9) were dissociated with accutase for 5 minutes. Dissociated cells were resuspended in PBS+0.04% BSA and strained using 40micron strainer to avoid clumps. Cell concentration in PBS+0.04% BSA was adjusted to 800-1200cells/ul before loading cells on 10X chromoium instrument. scRNA-seq libraries were generated using the 10X Genomics Chromium instrument and Chromium single cell 3' reagent kit V3. Individual libraries were designed to target 10,000 cells. Libraries were generated following manufacturer's instructions and library fragment size distribution was determined by BioAnalyzer. Afterwards libraries were pooled together and sequenced on Illumina Novaseq 6000 platform.

Processing single-cell expression data

Reads were aligned using the CellRanger (https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome) count function against hg38 (GRCh38-3) genome assembly. To account for mouse feeder cells, we also aligned the reads against the mouse mm10 genome assembly.

The aligned reads were further processed using Seurat (Butler et al., 2018; Stuart et al., 2019) and filtered based on multiple criteria: 1) Initial filtering for genes detected in at least three cells, and cells where at least 200 detected genes. 2) To remove mouse feeder cells, the number of genes detected when aligning to the human and mouse genome was calculated (nFeatures), and cells with log2(nFeatures human/nFeatures mouse)>1 were kept for downstream analysis. 3) Mitochondrial RNA was quantified per cell. Quality control metric was calculated using R

function calculateQCMetrics (scater package (McCarthy et al., 2017)). The R function isOutlier (scater package (McCarthy et al., 2017)) was used to identify outliers based on the library size (nmads>3), number of genes (nmads>3), and the percentage of mitochondrial genes in each cell (nmads>1).

The data was then normalized and scaled, and highly variable genes were detected, using sctransform function (Seurat package (Butler et al., 2018; Stuart et al., 2019)) using default parameters. Principal Component Analysis was done using Seurat function RunPCA on the most highly variable genes. The top 20 principal components were used to find the 20 nearest neighbors of each cell (using FindNeighbors function from Seurat package (Butler et al., 2018; Stuart et al., 2019)), following by FindClusters function to identify cell clusters. Clusters were visualized by uniform manifold approximation and projection (UMAP) dimensional reduction technique using the Seurat RunUMAP function.

To explore fold change of genes that escape or are subject to XCD, the average gene expression for each scRNA-seq cluster was obtained using AverageExpression function in Seurat (Butler et al., 2018; Stuart et al., 2019), and the log fold change between the WT and XIST KO was calculated for each gene.

Single cell expression in pre-implantation male and female embryo was downloaded from Petropoulos et al., (Petropoulos et al., 2016). Read counts matrix was normalized and scaled using sctransform function (Seurat package (Butler et al., 2018; Stuart et al., 2019)) using default parameters. To explore fold change of genes that escape or are subject to XCD the mean RPKMs for each stage, tissue and sex was obtained from Petropoulos et al., (Petropoulos et al., 2016).

Variant Calling

For UCLA1 lines, a list of known X-linked Single-Nucleotide Polymorphisms (SNPs) was extracted from Sahakyan et al., (Sahakyan et al., 2017a). For H9 lines, a list of known X-linked SNPs in H9

was extracted from Collier et al., (Collier et al., 2017). The genotype coordinates (in hg19) in the original VCFs were liftOver from hg19 to hg38 using the CrossMap (Zhao et al., 2014).

Haplotype phasing

Since the SNP data is both UCLA1 and H9 were taken form an unphased genotypes, a read overlapping a known SNP could not be directly assigned to one of the two parental X chromosomes (haplotypes). To directly explore gene expression from the maternal and paternal X chromosomes, phasing haplotypes analysis, which identify the alleles that are co-located on the same chromosome was done, assigning each allele to either X1 or X2. This allowed us to explore expression of all SNPs between the maternal and paternal X chromosomes. Haplotype phasing is particularly important for SNP analysis of the single-cell RNA-seq data since in order to correctly asses if a SNP is bi or monoallelicaly expressed, a minimal number of RNA-seq reads need to overlap that SNP. However, in the single cell RNA-seq data, only a handful of SNPs overlapped by more than five RNA-seq reads. More so, these SNPs tent to be located in genes that escape XCD as XCD escapers tend to be more highly expressed. This can result in defining an X-chromosome as active, as SNPs overlapping genes subject to XCD will not be detected in the analysis. Phased haplotype enables the detection of all RNA reads coming from either the paternal or maternal allele (regardless of the number of reads overlapping a specific SNP), therefor allowing to overcome this problem. In order to assign a SNP to the paternal or maternal X-chromosome, RNA-seq data from Primed UCLA1 and H9 were used, in which one of the two X-chromosomes is silenced, and therefore all monoallelicaly expressed alleles are assigned to the Primed Xa, while the not expressed allele is assigned to the primed Xi. A step-by-step description of the Haplotype phasing follows: 1) Using the aligned reads in primed hESCs to determine allelic gene coverage. SAMtools mpileup tool (Li et al., 2009) was used to generate SNP coverage pileups for primed UCLA1 and H9, using min-BQ = 20, and max-depth = 1000000. For each SNP, the read coverage of the reference and alternative allele was calculated. This

provide a list of know SNPs and the read coverage of each allele. 2) Assigning each allele to the primed Xa and Xi. Focusing on SNPs with mono allelic expression (in which the total number of reads > 5 and the absolute ratio of reference to alternative allele < 0.2 (80% of reads aligned to either the reference or alternative allele). In cases where the reference allele was overlapped by more than 80% of reads, the reference allele was assigned to the Xa while the alternative allele was assigned to the Xi, and vice versa. This provide a list of alleles detected on the Xa and Xi for UCLA1 (39 X-chromosome phased SNPs) and H9 (524 X-chromosome phased SNPs). To prevent confusion, we refer to the primed Xa as chrX1 and the primed Xi as chrX2.

Determine allelic X-linked gene coverage

SAMtools mpileup tool (Li et al., 2009) was used to generate SNP coverage pileups from RNA-seq data of each cell line, or each single cell in the scRNA-seq data, using min-BQ = 20, and max-depth = 1000000. The number of reads covering the reference and alternative allele, along with the ratio of reference to alternative allele coverage, was calculated for each SNP. For the phased SNP analysis, the number of reads aligned to chrX1 and chrX2 was also calculates. When calling SNP coverage in the scRNA data, only cells with more than five read overlapping all phased SNPs were kept.

Define escape and subject to XCD

The result of the differential gene expression analysis as described above, was used to define genes that escape or are subject to XCD. An X-linked gene was defined as XCD escapee if it was significantly (log2FoldChange>0.5 and padj<0.01) upregulated in female naïve hPSCs (H9, UCLA1, and iPSC) vs male naïve hESC (WIN1), or in female naïve hESC (HNES3) vs male naïve hESC (HNES1). For the pre-implantation embryo definition, the log2 fold change of female to male E7 epiblasts as defined in Petropoulos et al., (Petropoulos et al., 2016) was used, where an X-linked gene was defined as XCD escapee using log2 fold change > 1. Tukiainen et al.,

(Tukiainen et al., 2017) was used for defining XCI subject and escapees. XIST was excluded from downstream analysis.

RAP sequencing

RNA anti-sense purification followed by DNA sequencing (RAP-seq) for human XIST was adapted from Engreitz et al., (Engreitz et al., 2013). Human XIST probes for RAP-seq were generously gifted from Guttman Lab at Caltech. After harvesting and washing with PBS, 10-30 million cells from confluent cultures were incubated with freshly-made 10ml 2mM DSG in PBS at room temperature for 45 minutes. Cells were further crosslinked with 10ml 3% formaldehyde for 10 minutes, and the reaction was stopped by the addition of 2ml 2.5M glycine. Cells were pelleted at 4˚C and subjected to lysate preparation as described by Engreitz et al., (Engreitz et al., 2013). Briefly, the fixed cell pellets were lysed in NP-40 containing cell lysis buffer either with glass dounce homogenizer (UCLA1, iPSCs and NHDFs) or without (H9), with further lysis of nuclei in a buffer with NP-40, sodium deoxycholate and N-lauroylsarcosine. Chromatin was solubilized by sonication and segmented by TURBO DNase digestion, followed by XIST RNA pulldown from 5 million cells using 1ug (UCLA1, iPSCs, NHDF R1), 5ug (NHDF R2) or 50pmol (H9) custom 162 non-overlapping 90nt long biotinylated oligonucleotides (Eurofins) and Streptavidin C1 beads. DNA was eluted by RNase H digestion, and crosslinking was reversed via proteinase K digestion of eluted DNA at 60˚C. DNA library was prepared using NEBNext Ultra End Rpair/dA-Tailing Module (NEB) and TruSeq DNA adapters (Illumina) were ligated using Quick Ligase (NEB). Libraries were amplified by KAPA HiFi Polymerase (Roche), pooled, and sequenced on the Illlumina HiSeq platform to generate 50bp single-end reads (UCLA1, iPSCs and NHDFs) or pair-end (H9).

RAP-seq alignment

DNA sequencing reads were trimmed using trim_galore (https://github.com/FelixKrueger/TrimGalore) with default parameters to remove the standard Illumina adaptor sequence. Bowtie2 (Langmead and Salzberg, 2012) was used to align reads to the human genome (hg38) with the default parameters. Reads with mapping quality less than 30 were removed using Samtools (Li et al., 2009), and Picard MarkDuplicates ("Picard Tools - By Broad Institute," n.d.) was used to mark PCR duplicates.

RAP-seq enrichment

RAP-seq read counts were calculated using two different approaches to define genomic regions: 1) at genomic intervals. bedtools makewindows (Quinlan and Hall, 2010) was used to create bed files of genomic intervals in either 100Kb windows every 25Kb (for enrichment analysis), 1Mb every 250Kb (for feature enrichment), or non-overlapping 100kb windows (for differential analysis) along the genome. 2) At 25Kb up and downstream of genes transcription start site (TSS) capturing gene bodies along with gene promoters (for comparing gene expression and XIST enrichment). bedtools intersect (Quinlan and Hall, 2010) was then used to count reads in each genomic region. To account for differences in sequencing depth, the read counts in each genomic region were normalized to the sum of all reads. To explore XIST enrichment on autosomes, a scaled enrichment score was also calculated by normalizing only autosomal windows to the sum of all autosomal reads. This is referred to as scaled enrichment.

RAP-seq enrichment scores in each genomic region were calculated using the ratio of the normalized read counts in the RAP-seq pulldown to the input of each sample. A region was defined as an unmappable region using the inputs of all samples. Specifically, the R function isOutlier (scater package (McCarthy et al., 2017)) was used to identify outliers (having less or more than expected read counts) based on the minimal number of reads across all input samples (nmads>4). Genomic regions identified as outliers were removed from all downstream analysis. These normalized enrichment ratios were used in all further computational analysis. We note that

while the read counts of genomic intervals were defined in 100Kb or 1Mb windows every 25Kb and 250Kb respectively along the genome, the enrichment scores were assigned to the 25Kb or 250Kb windows in the center of the 100Kb and 1Mb windows respectively. This was done to prevent overlapping windows in the downstream analysis.

RAP-seq peak calling

Peak calling was performed using MACS2 callpeak (Zhang et al., 2008) with max-gap=1000 using the input of each sample as a control sample. For visualization, MACS2 bdgcmp was also to generate fold-enrichment tracks. The normalized bedgraph files were then converted to tdf format using igvtools toTDF (Robinson et al., 2011). To compare peak scores between samples, bedtools merge (Quinlan and Hall, 2010) was used to merge peaks of all samples using either the narrow or broad peaks. bedtools intersect (Quinlan and Hall, 2010) was then used to intersect between the merged peaks, and the peaks identified in each sample.

Differential RAP-seq analysis

Two approaches were used to explore differential XIST localization on the X chromosome: 1) MACS2 (Zhang et al., 2008, p. 2) was used to call peaks in each of the samples, using the input of each sample as control as described above. The R package DiffBind (Ross-Innes et al., 2012) was then used to identify differential peaks between the three naïve female hPSCs and the two Fibroblast cells. Specifically, read counts of each sample were calculated for each of the merged peaks, filtering out intervals with low read counts (<10, using filterFun=sum) and keeping only X-chromosome peaks. We note that XIST is more enriched on the X-chromosome in fibroblast cells compared to the naïve cells and keeping only X-chromosome peaks allowed to identify differences in the relative enrichment in somatic and naïve hPSCs. 2) DESeq2 (Love et al., 2014) was used on the raw read counts in each 100Kb window. Regions with less than 10 reads in all

samples were removed. Significant differences were defined as 100kb regions with p-value<0.01. Similar to first approach, only reads on the X chromosome where selected.

To identify naïve-conserved autosomal peaks, broad peaks were used, filtering out non-significant peaks (using p-value < 0.05, and fold enrichment of >2 for all samples, and >3 for iPSCs). Short peaks (<500bps) were also remove. We note that the naïve hESC UCLA1 likely had the right arm of chromosome 8 duplicated (noted by the increased reads in the UCLA1 input), and therefor peaks on chr8 where peak start >121000000 were removed. Naïve-conserved peaks were defined as peaks identified in at all three naïve hPSCs and in none of the two fibroblast samples. Somatic-conserved peaks were defined as peaks identified in both somatic but none of the naïve samples.

Features enrichment analysis

Two different approaches were used to identify enrichment of different genomic features with XIST enrichment: 1) Correlation with XIST enrichment scores in genomic intervals and interval scores of different genomic features. Gene annotation (TxDb.Hsapiens.UCSC.hg38.knownGene) was downloaded from UCSC ("TxDb.Hsapiens.UCSC.hg38.knownGene," n.d.). DNA repeat annotations were downloaded from the UCSC Genome Browser, track RepeatMasker, table rmsk (Karolchik et al., 2004). The number of annotated genes and repeats in each genomic region was calculated and Pearson correlation was used to calculate the correlation between XIST enrichment score at each genomic region (1Mb windows every 250Kb), and each of the features. XIST locus (10Mb from each side) was removed similar to Engreitz et al., (Engreitz et al., 2013). 2) Enrichment of different features with naïve-conserved autosomal peaks. In addition to the features described above, LOLA Core database (Sheffield and Bock, 2016) were used. The R package LOLA (Sheffield and Bock, 2016) was then used to compute enrichment of the different features with autosomal naïve-conserved peaks. 3) Enrichment for gene ontologies (GO) was

perform using the R package topGO (Alexa et al., 2006) using Fisher test. GO enrichment was performed for genes found under Naïve-conserved XIST peaks.

Immunofluorescence

For immunofluorescent staining hESCs were seeded on MEF coated coverslips and fixed after 24-48 hours. Cells were fixed with 4% PFA for 10 minutes and washed with 1xPBS. Afterwards permeabilized with 0.5% Triton X-100 in 1xPBS and blocked with 1% BSA in 1xTBS. Primary antibody incubation was conducted 1% BSA for 1h at RT. Samples were again washed with 3xTBS-tween and incubated with fluorescent secondary antibodies at 1:400 for 45 min, then washed and counterstained with DAPI for 5 min and mounted using Vectashield. The secondary antibodies used in this study were all from Life technologies used at 1:400 dilution. Images were taken using LSM 880 Confocal Instrument (Zeiss) or Zeiss Axio Imager M1. For image processing and analysis Fiji (ImageJ) was used. For signal quantification, images were converted into 8-bit images and afterwards signal intensity was measured by drawing line over signal of interest and extracting pixel intensities with profile plot tool. Intensity values were exported as a csv file.

RNA fluorescent in situ hybridization (FISH)

For RNA FISH hESCs were seeded on a MEF coated coverslips 24-48 hours before fixation to keep hESC colonies smaller. Coverslips were washed with DPBS, fixed with 4% formaldehyde for 10 min, permeabilized with cold (4°C) 0.5% Triton X-100 in DPBS for 10 min, and serially dehydrated with cold (4°C) 70-100% ethanol for 10 minutes each step. Afterwards coverslips were air dried and hybridized with labeled DNA probes in a humidified chamber at 37°C overnight. Next day coverslips were washed with 50% formamide in 2x SSC, 2x SSC, then 1x SSC at 42°C for 20 minutes each step. For nuclei staining, coverslips were stained with DAPI and mounted with Vectashield (Vector labs: H-1000). Double- stranded DNA probes were generated from full length cDNA constructs or BACs as described previously(Solovei, 2010). The BACs used include XIST

(RP11-13M9), XACT (RP11-35D3), GPC3 (RP11-678F20), SMS (RP11-147O5) and SMARCA1 (RP11-137A15), THOC2 (RP11-121P4), UTX (RP11-256P2). Every new batch of probes was tested on normal human dermal fibroblasts before use in experiments. In case of X chromosome paints combined with XIST RNA FISH, first RNA-FISH was done as described above and afterwards same coverslips were post-fixed with 4% PFA for 10 minutes and X chromosome paint was done following to manufacturer's instructions (Metasystems Probes, cat:D-0323-100-OR).

Generation of XIST KO lines

To generate XIST KO lines we have electroporated H9 primed hESCs with two PX459 plasmids (#48139) carrying two different gRNAs, one targeting XIST promoter and one exon 1. gRNA sequences are added to STAR methods. Cell were electroporated using Lonza 4D-Nucleofector and P3 Primary Cell kit (Cat. V4XP-3024) according to manufacturer's instructions. Electroporated cells were seeded on MEF coated plated in Primed hESC media with 10uM ROCK inhibitor Y-27632. After 48h cells were selected with puromycin and surviving colonies were further propagated. To confirm the deletion, genomic DNA was isolated first from the bulk population using Zymo Quick-DNA isolation kit (Cat. D4069). PCR primers XP-Frw (CACAAAGATGTCCGGCTTTCA) and XE-Rev (CCTGCTGAATGCAAATGGGG) generate 1.5 kb bend upon deletion of 2kb. After this step, individual colonies were handpicked from the targeted population and screened again for homozygous and heterozygous deletions. To screen for WT allele with no deletion we used XP-Frw and WT-Rev (CTCTGCCAAAGCGGTAGGTAC) primers that amplifies 942bp region from WT allele and cannot amplify anything upon deletion, as WT-Rev complementary sequence is not present after 2kb deletion. With this screening strategy 2 Homozygous clones were selected for the experiments after PCR confirmation of presence of deletion indicating 1.5 kb bend and absence of WT 942 bp bend.

siRNA inhibition

For siRNA inhibition of SPEN we have used mix of 2 different siRNAs targeting human SPEN different exons. siRNAs were from Thermo Fisher #4427037, IDs: s22831, s22829. Equal amounts of siRNA were mixed prior to transfection. As a negative control scramble siRNA was used. Cells were transfected using Lipofectamine RNAiMAX reagent (Life Technologies 13778150). After 24h of transfection, cells were transfected again using the same siRNA mix. After 48 hours cells were harvested, and RNA was isolated using Qiagen RNAeasy kit (#74104) according to manufacturer's instructions. Afterwards RNA-seq libraries were constructed as described above.

Hi-C data analysis

Maternal and paternal Hi-C contacts (.hic files) in GM12878 cells were downloaded from Rao et al., (Rao et al., 2014) (GSE63525). Knight-Ruiz (KR) normalized observed and expected contact matrixes were generated using Juicebox dump command of the Juicebox tool (Durand et al., 2016). Distance normalized interaction signals (observed/expected) were calculated at 1Mb resolution for the X chromosome.

References

Adachi, K., Kopp, W., Wu, G., Heising, S., Greber, B., Stehling, M., Araúzo-Bravo, M.J., Boerno, S.T., Timmermann, B., Vingron, M., Schöler, H.R., 2018. Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. Cell Stem Cell 23, 266-275.e6. https://doi.org/10.1016/j.stem.2018.05.020

Alexa, A., Rahnenführer, J., Lengauer, T., 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22, 1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Augui, S., Nora, E.P., Heard, E., 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. 12, 429–442. https://doi.org/10.1038/nrg2987

Avner, P., Heard, E., 2001. X-chromosome inactivation: counting, choice and initiation. Nat. Rev. Genet. 2, 59–67. https://doi.org/10.1038/35047580

Benchetrit, H., Jaber, M., Zayat, V., Sebban, S., Pushett, A., Makedonski, K., Zakheim, Z., Radwan, A., Maoz, N., Lasry, R., Renous, N., Inbar, M., Ram, O., Kaplan, T., Buganim, Y., 2019. Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts. Cell Stem Cell 24, 983-994.e7. https://doi.org/10.1016/j.stem.2019.03.018

Bonora, G., Deng, X., Fang, H., Ramani, V., Qiu, R., Berletch, J.B., Filippova, G.N., Duan, Z., Shendure, J., Noble, W.S., Disteche, C.M., 2018. Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. Nat. Commun. 9, 1445. https://doi.org/10.1038/s41467-018-03694-y

Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.-B., Vassilev, I., Galupa, R., Servant, N., Barillot, E., Surani, A., Chen, C.-J., Heard, E., 2017. Xist -dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. 24, 226–233. https://doi.org/10.1038/nsmb.3365

Brockdorff, N., 2019. Localized accumulation of Xist RNA in X chromosome inactivation. Open Biol. 9, 190213. https://doi.org/10.1098/rsob.190213

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., Rastan, S., 1991. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature 351, 329–331. https://doi.org/10.1038/351329a0

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., Willard, H.F., 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature 349, 38–44. https://doi.org/10.1038/349038a0

Brown, C.J., Robinson, W.P., 1997. XIST expression and X-chromosome inactivation in human preimplantation embryos. Am. J. Hum. Genet. 61, 5–8.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. https://doi.org/10.1038/nbt.4096

Cao, Y., Chen, G., Wu, G., Zhang, X., McDermott, J., Chen, X., Xu, C., Jiang, Q., Chen, Z., Zeng, Y., Ai, D., Huang, Y., Han, J.-D.J., 2019. Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions. Genome Res. 29, 40–52. https://doi.org/10.1101/gr.235747.118

Carrel, L., Willard, H.F., 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434, 400–404. https://doi.org/10.1038/nature03479

Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., Tartaglia, G.G., 2019. Phase separation drives X-chromosome inactivation: a hypothesis. Nat. Struct. Mol. Biol. 26, 331–334. https://doi.org/10.1038/s41594-019-0223-0

Chan, Y.-S., Göke, J., Ng, J.-H., Lu, X., Gonzales, K.A.U., Tan, C.-P., Tng, W.-Q., Hong, Z.-Z., Lim, Y.-S., Ng, H.-H., 2013. Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. Cell Stem Cell 13, 663–675. https://doi.org/10.1016/j.stem.2013.11.015

Chitiashvili, T., Dror, I., Kim, R., Hsu, F.-M., Chaudhari, R., Pandolfi, E., Chen, D., Liebscher, S., Schenke-Layland, K., Plath, K., Clark, A., 2020. Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. Nat. Cell Biol. 22, 1436–1446. https://doi.org/10.1038/s41556-020-00607-4

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., Chang, H.Y., 2015. Systematic discovery of Xist RNA binding proteins. Cell 161, 404–416. https://doi.org/10.1016/j.cell.2015.03.025

Collier, A.J., Panula, S.P., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., Rugg-Gunn, P.J., 2017. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. Cell Stem Cell 20, 874-890.e7. https://doi.org/10.1016/j.stem.2017.02.014

Daniels, R., Zuccotti, M., Kinis, T., Serhal, P., Monk, M., 1997. XIST Expression in Human Oocytes and Preimplantation Embryos. Am. J. Hum. Genet. 61, 33–39. https://doi.org/10.1086/513892

Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.-C., Sanborn, A.L., Machol, I., Shamim, M., Seberg, A.P., Lander, E.S., Chadwick, B.P., Aiden, E.L., 2016. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. 113, E4504–E4512. https://doi.org/10.1073/pnas.1609643113

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., Onate, K.C., Graham, K., Miyasato, S.R., Dreszer, T.R., Strattan, J.S., Jolanki, O., Tanaka, F.Y., Cherry, J.M., 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 46, D794–D801. https://doi.org/10.1093/nar/gkx1081

Deng, X., Berletch, J.B., Nguyen, D.K., Disteche, C.M., 2014. X chromosome regulation: diverse patterns in development, tissues and disease. Nat. Rev. Genet. 15, 367–378. https://doi.org/10.1038/nrg3687

Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J.B., Blau, C.A., Shendure, J., Duan, Z., Noble, W.S., Disteche, C.M., 2015. Bipartite structure of the inactive mouse X chromosome. Genome Biol. 16, 152. https://doi.org/10.1186/s13059-015-0728-8

Dossin, F., Pinheiro, I., Żylicz, J.J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y., Dingli, F., Loew, D., Mercher, T., Dekker, J., Heard, E., 2020. SPEN integrates transcriptional and epigenetic control of X-inactivation. Nature 578, 455–460. https://doi.org/10.1038/s41586-020-1974-9

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., Aiden, E.L., 2016. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. 3, 99–101. https://doi.org/10.1016/j.cels.2015.07.012

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., Plath, K., Guttman, M., 2013. The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. Science 341, 1237973. https://doi.org/10.1126/science.1237973

Ercan, S., Dick, L.L., Lieb, J.D., 2009. The C. elegans Dosage Compensation Complex Propagates Dynamically and Independently of X Chromosome Sequence. Curr. Biol. 19, 1777–1787. https://doi.org/10.1016/j.cub.2009.09.047

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., Rais, Y., Shipony, Z., Mukamel, Z., Krupalnik, V., Zerbib, M., Geula, S., Caspi, I., Schneir, D., Shwartz, T., Gilad, S., Amann-Zalcenstein, D., Benjamin, S., Amit, I., Tanay, A., Massarwa, R., Novershtern, N., Hanna, J.H., 2013. Derivation of novel human ground state naive pluripotent stem cells. Nature 504, 282–286. https://doi.org/10.1038/nature12745

Galupa, R., Heard, E., 2015. X-chromosome inactivation: new insights into cis and trans regulation. Curr. Opin. Genet. Dev., Genome architecture and expression 31, 57–66. https://doi.org/10.1016/j.gde.2015.04.002

Gardner, H.P., Belka, G.K., Wertheim, G.B., Hartman, J.L., Ha, S.I., Gimotty, P.A., Marquis, S.T., Chodosh, L.A., 2000. Developmental role of the SNF1-related kinase Hunk in pregnancy-induced changes in the mammary gland. Dev. Camb. Engl. 127, 4493–4509.

Gendrel, A.-V., Heard, E., 2014. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu. Rev. Cell Dev. Biol. 30, 561–580. https://doi.org/10.1146/annurev-cellbio-101512-122415

Gimferrer, I., Farnós, M., Calvo, M., Mittelbrunn, M., Enrich, C., Sánchez-Madrid, F., Vives, J., Lozano, F., 2003. The accessory molecules CD5 and CD6 associate

on the membrane of lymphoid T cells. J. Biol. Chem. 278, 8564–8571. https://doi.org/10.1074/jbc.M209591200

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., Dekker, J., 2016. Structural organization of the inactive X chromosome in the mouse. Nature 535, 575–579. https://doi.org/10.1038/nature18589

Gizard, F., Robillard, R., Gross, B., Barbier, O., Révillion, F., Peyrat, J.-P., Torpier, G., Hum, D.W., Staels, B., 2006. TReP-132 Is a Novel Progesterone Receptor Coactivator Required for the Inhibition of Breast Cancer Cell Growth and Enhancement of Differentiation by Progesterone. Mol. Cell. Biol. 26, 7632–7644. https://doi.org/10.1128/MCB.00326-06

Graindorge, A., Pinheiro, I., Nawrocka, A., Mallory, A.C., Tsvetkov, P., Gil, N., Carolis, C., Buchholz, F., Ulitsky, I., Heard, E., Taipale, M., Shkumatava, A., 2019. In-cell identification and measurement of RNA-protein interactions. Nat. Commun. 10, 5317. https://doi.org/10.1038/s41467-019-13235-w

Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., Plath, K., Smith, A., 2017. Epigenetic resetting of human pluripotency. Dev. Camb. Engl. 144, 2748–2763. https://doi.org/10.1242/dev.146811

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., Nichols, J., 2016. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. Stem Cell Rep. 6, 437–446. https://doi.org/10.1016/j.stemcr.2016.02.005

Gyllborg, D., Ahmed, M., Toledo, E.M., Theofilopoulos, S., Yang, S., ffrench-Constant, C., Arenas, E., 2018. The Matricellular Protein R-Spondin 2 Promotes Midbrain Dopaminergic Neurogenesis and Differentiation. Stem Cell Rep. 11, 651–664. https://doi.org/10.1016/j.stemcr.2018.07.014

Hall, L.L., Byron, M., Sakai, K., Carrel, L., Willard, H.F., Lawrence, J.B., 2002. An ectopic human XIST gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. Proc. Natl. Acad. Sci. 99, 8677–8682. https://doi.org/10.1073/pnas.132468999

Huang, K., Maruyama, T., Fan, G., 2014. The Naive State of Human Pluripotent Stem Cells: A Synthesis of Stem Cell and Preimplantation Embryo Transcriptome Analyses. Cell Stem Cell 15, 410–415. https://doi.org/10.1016/j.stem.2014.09.014

Huynh, K.D., Lee, J.T., 2003. Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. Nature 426, 857–862. https://doi.org/10.1038/nature02222

Jiang, J., Jing, Y., Cost, G.J., Chiang, J.-C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., Pearl, J.R., Rebar, E.J., Byron, M., Gregory, P.D., Brown, C.J., Urnov, F.D., Hall, L.L., Lawrence, J.B., 2013. Translating dosage compensation to trisomy 21. Nature 500, 296–300. https://doi.org/10.1038/nature12394

Johnston, C.M., Lovell, F.L., Leongamornlert, D.A., Stranger, B.E., Dermitzakis, E.T., Ross, M.T., 2008. Large-Scale Population Study of Human Cell Lines Indicates that Dosage Compensation Is Virtually Complete. PLOS Genet. 4, e9. https://doi.org/10.1371/journal.pgen.0040009

Jonkers, I., Barakat, T.S., Achame, E.M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J.A., Gribnau, J., 2009. RNF12 Is an X-Encoded Dose-Dependent Activator of X Chromosome Inactivation. Cell 139, 999–1011. https://doi.org/10.1016/j.cell.2009.10.034

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., Gribnau, J., 2008. Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. 28, 5583–5594. https://doi.org/10.1128/MCB.02269-07

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., Kent, W.J., 2004. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 32, D493-496. https://doi.org/10.1093/nar/gkh103

Kelsey, A.D., Yang, C., Leung, D., Minks, J., Dixon-McDougall, T., Baldry, S.E.L., Bogutz, A.B., Lefebvre, L., Brown, C.J., 2015. Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. Genome Biol. 16, 208. https://doi.org/10.1186/s13059-015-0774-2

Khan, S.A., Audergon, P.N.C.B., Payer, B., 2017. X-chromosome activity in naive human pluripotent stem cells—are we there yet? Stem Cell Investig. 4. https://doi.org/10.21037/sci.2017.06.03

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915. https://doi.org/10.1038/s41587-019-0201-4

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Lee, J.T., Bartolomei, M.S., 2013. X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. Cell 152, 1308–1323. https://doi.org/10.1016/j.cell.2013.02.016

Leppig, K.A., Disteche, C.M., 2001. Ring X and other structural X chromosome abnormalities: X inactivation and phenotype. Semin. Reprod. Med. 19, 147–157. https://doi.org/10.1055/s-2001-15395

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, X., Nefzger, C.M., Rossello, F.J., Chen, J., Knaupp, A.S., Firas, J., Ford, E., Pflueger, J., Paynter, J.M., Chy, H.S., O'Brien, C.M., Huang, C., Mishra, K., Hodgson-Garms, M., Jansz, N., Williams, S.M., Blewitt, M.E., Nilsson, S.K., Schittenhelm, R.B., Laslett, A.L., Lister, R., Polo, J.M., 2017. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. Nat. Methods 14, 1055–1062. https://doi.org/10.1038/nmeth.4436

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Mahadevaiah, S.K., Sangrithi, M.N., Hirota, T., Turner, J.M.A., 2020. A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. Nature 586, 612–617. https://doi.org/10.1038/s41586-020-2629-6

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., Brockdorff, N., 2004. Reactivation of the paternal X chromosome in early mouse embryos. Science 303, 666–669. https://doi.org/10.1126/science.1092674

Marahrens, Y., Loring, J., Jaenisch, R., 1998. Role of the Xist Gene in X Chromosome Choosing. Cell 92, 657–664. https://doi.org/10.1016/S0092-8674(00)81133-2

Marahrens, Y., Panning, B., Dausman, J., Strauss, W., Jaenisch, R., 1997. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 11, 156–166. https://doi.org/10.1101/gad.11.2.156

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., Wills, Q.F., 2017. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33, 1179–1186. https://doi.org/10.1093/bioinformatics/btw777

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M.J., Shishkin, A.A., Su, J., Lander, E.S., Hess, S., Plath, K., Guttman, M., 2015. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature 521, 232–236. https://doi.org/10.1038/nature14443

Medstrand, P., van de Lagemaat, L.N., Mager, D.L., 2002. Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes. Genome Res. 12, 1483–1495. https://doi.org/10.1101/gr.388902

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., Lee, J.T., 2015. A comprehensive Xist

interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science 349, aab2276. https://doi.org/10.1126/science.aab2276

Minkovsky, A., Patel, S., Plath, K., 2012. Concise Review: Pluripotency and the Transcriptional Inactivation of the Female Mammalian X Chromosome. STEM CELLS 30, 48–54. https://doi.org/10.1002/stem.755

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J.C., Reik, W., 2017. Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. 20, 1215–1228. https://doi.org/10.1016/j.celrep.2017.07.009

Moindrot, B., Brockdorff, N., 2016. RNA binding proteins implicated in Xist-mediated chromosome silencing. Semin. Cell Dev. Biol., X chromosome inactivation 56, 58–70. https://doi.org/10.1016/j.semcdb.2016.01.029

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., Wutz, A., 2015. Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. Cell Rep. 12, 554–561. https://doi.org/10.1016/j.celrep.2015.06.067

Navarro-Cobos, M.J., Balaton, B.P., Brown, C.J., 2020. Genes that escape from X-chromosome inactivation: Potential contributors to Klinefelter syndrome. Am. J. Med. Genet. C Semin. Med. Genet. 184, 226–238. https://doi.org/10.1002/ajmg.c.31800

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., Zakian, S.M., 2001. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. Genome Res. 11, 833–849. https://doi.org/10.1101/gr.174901

Nesterova, T.B., Wei, G., Coker, H., Pintacuda, G., Bowness, J.S., Zhang, T., Almeida, M., Bloechl, B., Moindrot, B., Carter, E.J., Alvarez Rodrigo, I., Pan, Q., Bi, Y., Song, C.-X., Brockdorff, N., 2019. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. Nat. Commun. 10, 3129. https://doi.org/10.1038/s41467-019-11171-3

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

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., Heard, E., 2004. Epigenetic Dynamics of Imprinted X Inactivation During Early Mouse Development. Science 303, 644–649. https://doi.org/10.1126/science.1092727

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V., Heard, E., 2011. Eutherian

mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature 472, 370–374. https://doi.org/10.1038/nature09872

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W.R., Damianov, A., Chronis, C., Papp, B., Chen, C.-K., McKee, R., Wang, X.-J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, D.L., Plath, K., 2020. A protein assembly mediates Xist localization and gene silencing. Nature 587, 145–151. https://doi.org/10.1038/s41586-020-2703-0

Pandya-Jones, A., Plath, K., 2016. The “lnc” between 3D Chromatin Structure and X Chromosome Inactivation. Semin. Cell Dev. Biol. 56, 35–47. https://doi.org/10.1016/j.semcdb.2016.04.002

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., 2017. Human embryonic stem cells do not change their X-inactivation status during differentiation. Cell Rep. 18, 54–67. https://doi.org/10.1016/j.celrep.2016.11.054

Patrat, C., Ouimette, J.-F., Rougeulle, C., 2020. X chromosome inactivation in human development. Development 147. https://doi.org/10.1242/dev.183095

Payer, B., Lee, J.T., 2014. Coupling of X-Chromosome reactivation with the pluripotent stem cell state. RNA Biol. 11, 798–807. https://doi.org/10.4161/rna.29779

Payer, B., Lee, J.T., 2008. X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. 42, 733–772. https://doi.org/10.1146/annurev.genet.42.110807.091711

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., Brockdorff, N., 1996. Requirement for Xist in X chromosome inactivation. Nature 379, 131–137. https://doi.org/10.1038/379131a0

Petri, R., Brattås, P.L., Sharma, Y., Jönsson, M.E., Pircs, K., Bengzon, J., Jakobsson, J., 2019. LINE-2 transposable elements are a source of functional human microRNAs and target sites. PLoS Genet. 15. https://doi.org/10.1371/journal.pgen.1008036

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., Lanner, F., 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell 165, 1012–1026. https://doi.org/10.1016/j.cell.2016.03.023

Picard Tools - By Broad Institute [WWW Document], n.d. URL http://broadinstitute.github.io/picard/ (accessed 12.8.20).

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., Panning, B., 2002. Xist RNA and the Mechanism of X Chromosome Inactivation. Annu. Rev. Genet. 36, 233–278. https://doi.org/10.1146/annurev.genet.36.042902.092433

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Ridings-Figueroa, R., Stewart, E.R., Nesterova, T.B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., Bageghni, S.A., Albadrani, G., Mansfield, W., Roulson, J.-A., Brockdorff, N., Ainscough, J.F.X., Coverley, D., 2017. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. Genes Dev. 31, 876–888. https://doi.org/10.1101/gad.295907.117

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. https://doi.org/10.1038/nbt.1754

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., Ali, S., Chin, S.-F., Palmieri, C., Caldas, C., Carroll, J.S., 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393. https://doi.org/10.1038/nature10730

Rostovskaya, M., Stirparo, G.G., Smith, A., 2019. Capacitation of human naïve pluripotent stem cells for multi-lineage differentiation. Development 146. https://doi.org/10.1242/dev.172916

Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., Plath, K., 2017a. Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. Cell Stem Cell 20, 87–101. https://doi.org/10.1016/j.stem.2016.10.006

Sahakyan, A., Plath, K., Rougeulle, C., 2017b. Regulation of X-chromosome dosage compensation in human: mechanisms and model systems. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372. https://doi.org/10.1098/rstb.2016.0363

Schulz, E.G., Heard, E., 2013. Role and control of X chromosome dosage in mammalian development. Curr. Opin. Genet. Dev. 23, 109–115. https://doi.org/10.1016/j.gde.2013.01.008

Sheffield, N.C., Bock, C., 2016. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics 32, 587–589. https://doi.org/10.1093/bioinformatics/btv612

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., Lee, J.T., 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature 504, 465–469. https://doi.org/10.1038/nature12719

Solovei, I., 2010. Fluorescence in situ Hybridization (FISH) on Tissue Cryosections. Methods Mol. Biol. Clifton NJ 659, 71–82. https://doi.org/10.1007/978-1-60761-789-1_5

Strome, S., Kelly, W.G., Ercan, S., Lieb, J.D., 2014. Regulation of the X Chromosomes in Caenorhabditis elegans. Cold Spring Harb. Perspect. Biol. 6, a018366. https://doi.org/10.1101/cshperspect.a018366

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Sunwoo, H., Colognori, D., Froberg, J.E., Jeon, Y., Lee, J.T., 2017. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). Proc. Natl. Acad. Sci. 114, 10654–10659. https://doi.org/10.1073/pnas.1711206114

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

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., Reik, W., Bertone, P., Smith, A., 2014. Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell 158, 1254–1269. https://doi.org/10.1016/j.cell.2014.08.029

Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., Duc, J., Cohen, M.A., Wert, K.J., Castanon, R., Zhang, Z., Huang, Y., Nery, J.R., Drotar, J., Lungjangwa, T., Trono, D., Ecker, J.R., Jaenisch, R., 2016. Molecular Criteria for Defining the Naive Human Pluripotent State. Cell Stem Cell 19, 502–515. https://doi.org/10.1016/j.stem.2016.06.011

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., Lungjangwa, T., Imsoonthornruksa, S., Stelzer, Y., Rangarajan, S., D'Alessio, A., Zhang, J., Gao, Q., Dawlaty, M.M., Young, R.A., Gray, N.S., Jaenisch, R., 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15, 471–487. https://doi.org/10.1016/j.stem.2014.07.002

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., Cummings, B.B., Castel, S.E., Karczewski, K.J., Aguet, F., Byrnes, A., GTEx Consortium, Lappalainen, T., Aviv Regev, Ardlie, K.G., Hacohen, N., MacArthur, D.G., 2017. Landscape of X chromosome inactivation across human tissues. Nature 550, 244–248. https://doi.org/10.1038/nature24265

TxDb.Hsapiens.UCSC.hg38.knownGene [WWW Document], n.d. . Bioconductor. URL http://bioconductor.org/packages/TxDb.Hsapiens.UCSC.hg38.knownGene/ (accessed 12.8.20).

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

Vallot, C., Patrat, C., Collier, A.J., Huret, C., Casanova, M., Ali, T.M.L., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J., Rougeulle, C., 2017. XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. Cell Stem Cell 20, 102. https://doi.org/10.1016/j.stem.2016.10.014

van den Berg, I.M., Galjaard, R.J., Laven, J.S.E., van Doorninck, J.H., 2011. XCI in preimplantation mouse and human embryos: first there is remodelling…. Hum. Genet. 130, 203–215. https://doi.org/10.1007/s00439-011-1014-9

Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C., Zhuang, X., 2016. Spatial organization of chromatin domains and compartments in single chromosomes. Science 353, 598–602. https://doi.org/10.1126/science.aaf8084

Wutz, A., 2014. Haploid Mouse Embryonic Stem Cells: Rapid Genetic Screening and Germline Transmission. Annu. Rev. Cell Dev. Biol. 30, 705–722. https://doi.org/10.1146/annurev-cellbio-100913-012920

Wutz, A., Jaenisch, R., 2000. A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation. Mol. Cell 5, 695–705. https://doi.org/10.1016/S1097-2765(00)80248-8

Wutz, A., Rasmussen, T.P., Jaenisch, R., 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. 30, 167–174. https://doi.org/10.1038/ng820

Yang, C., Chapman, A.G., Kelsey, A.D., Minks, J., Cotton, A.M., Brown, C.J., 2011. X-chromosome inactivation: molecular mechanisms from the human perspective. Hum. Genet. 130, 175–185. https://doi.org/10.1007/s00439-011-0994-9

Yang, L., Kirby, J.E., Sunwoo, H., Lee, J.T., 2016. Female mice lacking Xist RNA show partial dosage compensation and survive to term. Genes Dev. 30, 1747–1760. https://doi.org/10.1101/gad.281162.116

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., Orkin, S.H., Shen, X., 2015. Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. Cell Stem Cell 16, 504–516. https://doi.org/10.1016/j.stem.2015.03.007

Zhang, Y., Klein, K., Sugathan, A., Nassery, N., Dombkowski, A., Zanger, U.M., Waxman, D.J., 2011. Transcriptional Profiling of Human Liver Identifies Sex-Biased Genes Associated with Polygenic Dyslipidemia and Coronary Artery Disease. PLOS ONE 6, e23506. https://doi.org/10.1371/journal.pone.0023506

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., Wang, L., 2014. CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinforma. Oxf. Engl. 30, 1006–1007. https://doi.org/10.1093/bioinformatics/btt730

CHAPTER 4

FGFR3 is Expressed by Human Primordial Germ Cells and is Repressed after Meiotic initiation to Form Primordial Oocytes

FGFR3 is Expressed by Human Primordial Germ Cells and is Repressed after Meiotic Initiation to Form Primordial Oocytes

Tsotne Chitiashvili[1,2,3], Fei-man Hsu[1], Iris Dror[2], Kathrin Plath[2,3,4,5] & Amander Clark[1,3,4,5*]

1. Molecular Cell and Developmental Biology Department - University of California Los Angeles, Los Angeles, CA, USA
2. Department of Biological Chemistry, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
3. Molecular Biology Institute – University of California, Los Angeles, Los Angeles, CA, USA
4. Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles Los Angeles, CA, USA
5. Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA, USA

*corresponding author: clarka@ucla.edu

Abstract

Human germ cell development is a highly regulated process beginning soon after embryo implantation with the specification of primordial germ cells (PGCs) and ending in adulthood with the differentiation of gametes. Here, we show that fibroblast growth factor receptor 3 (FGFR3) is expressed by human PGCs during the first and second trimester, becoming repressed as PGCs differentiate into primordial oocytes. Using fluorescence activated cell sorting (FACS) with antibodies that recognize FGFR3 followed by single cell RNA sequencing, we show that isolating FGFR3-positive cells enriches for human PGCs. Taken together, FGFR3 could be used in future studies as a strategy to identify maturing hPGCs.

Introduction

Human reproduction depends on the correct establishment and differentiation of germ cells. Each generation, germ cell formation begins with the specification of primordial germ cells (PGCs) in the post-implantation embryo at the end of week 2 post-fertilization (pf) [1]. At the end of week 3 pf, a cluster of PGCs can be identified in the Yolk Sac endoderm, corresponding to specified PGCs that ultimately differentiate into eggs and sperm in the adult ovary or testis, respectively. Following specification, PGCs migrate from the Yolk Sac niche, through the dorsal mesentery of the hind gut to colonize a new niche, the genital ridge epithelium (the future gonad) starting at around week 5 pf. Once colonized in the developing fetal gonads, PGCs are referred to as gonocytes, late PGCs, Fetal Germ Cells (FGCs) or in ovaries oogonia [2-3]. Sex determination of the genital ridge epithelial niche cells initiates at around week 6, with testicular PGCs beginning the process of differentiating into fetal Spermatogonia also called prospermatogonia or Fetal 0 (F0) at around week 14-16 [4]. In the embryonic ovary, PGCs initiate the process of meiotic differentiation at week 9 and will arrest in prophase I of meiosis I as primordial oocytes from week 13-16. Creation of primordial follicles composed of a primordial oocyte and its surrounding layer of squamous granulosa cells begins at approximately 20 weeks pf [5]. Given the importance of PGC formation to egg and sperm in the adult, the cell and molecular basis of human PGC development are of significant importance to our basic understanding of human fertility and reproduction.

Isolation of PGCs from embryonic and fetal tissue using fluorescence activated cell sorting (FACS) has transformed our ability to study *in vivo* PGCs*,* and to enrich for PGC-like cells (PGCLCs) differentiated from human pluripotent stem cells *in vitro*. In addition, molecular analysis of FACS isolated *in vivo* PGCs has led to the creation of important benchmarks for comparing and staging *in vivo* PGCs with *in vitro* differentiated PGCLCs [6–11].

So far, a small number of proteins expressed at the cell membrane of human PGCs has been used to enrich for PGCs from a single cell suspension of human embryonic and fetal tissue *in vivo*, as well as PGCLCs *in vitro*. These include cKIT, TNAP, PDPN, CD38, ITGA6 and EPCAM [6], [8], [12–14]. In most cases, combinations of antibodies that recognize two or more of these cell surface proteins are used in FACS strategies to isolate highly enriched PGC populations for further analysis. Therefore, the identification of additional surface molecules that could facilitate isolation of PGCs from somatic cells is warranted. A candidate new cell surface receptor identified previously as being expressed by human testicular fetal germ cells is Fibroblast Growth Factor Receptor 3 (FGFR3) [15]. However, it is not known whether FGFR3 protein is expressed by PGCs in the human fetal ovary, or whether anti-FGFR3 antibodies conjugated to fluorescent tags could be used to isolate *in vivo* PGCs from prenatal ovarian tissue consented to research or PGCLCs generated *in vitro*.

To dissect the expression of FGFR3 in embryonic and fetal ovaries, we utilized previously published scRNA sequencing data sets [16-17] as well as immunofluorescence. We show

that the *FGFR3* gene is transcribed, translated and present at the surface of embryonic PGCs from at least 5 weeks post-fertilization, and becomes repressed as ovarian meiotic germ cells upregulate *SCP3* and *SPO11* in prophase I of meiosis I. In addition, we show that FACS can be used to enrich for FGFR3 positive PGCs from the human embryonic and fetal ovary, but not PGCLCs differentiated from human pluripotent stem cells. Taken together, our data identifies FGFR3 as a new surface marker that can enrich for PGCs from a single cell suspension of embryonic ovarian cells and could be used in future studies to identify PGCLCs differentiated from human pluripotent stem cells that have progressed towards gonadal-stage PGCs.

Results

FGFR3 is expressed by germ cells in the fetal testis and ovary

Previous studies have shown that FGFR3 protein is dynamically expressed during fetal testicular germ cell development [15], however the expression of FGFR3 by fetal ovarian germ cells has not been reported. Utilizing a previously published single cell (sc) RNA-seq 10x Genomics data set of human fetal ovaries (Fig. 1a) and human fetal testes (Fig. S1a) from 6-16 weeks pf [16], we identify that *FGRF3* mRNA is most enriched in the germ cell population of each sex (Fig. 1a-d and Fig. S1a-b).

To identify the stage-specific expression of *FGFR3* mRNA in fetal ovarian germ cells, we defined the PGC population (positive for *NANOG, POU5F1, BLIMP1, TFAP2C* and *NANOS3*), the meiotic germ population (positive for *STRA8, ZGLP1, SPO11,* and *SCP1*) and primordial oocytes (positive for *ZP3*) (Fig. S2a-b). Using these annotations, we

discovered that *FGFR3* mRNA is expressed in PGCs and meiotic germ cells while being below the limit of detection in primordial oocytes (Fig. 1a-e). To confirm these results in a second data set, we examined *FGFR3* expression using the single cell SMART-seq data set published by Li et. al. which covers 5 – 26 weeks pf [17]. This analysis corroborated that *FGFR3* mRNA is expressed by PGCs and meiotic germ cells, while being below the limit of detection in the somatic cells of the fetal ovary (Fig. 1f).

Given that some cells within the meiotic cluster have low to no expression of *FGFR3* mRNA (Fig. 1c-e), we hypothesized that *FGFR3* may be repressed as the ovarian germ cells progress through meiosis. To address this, we separated the meiotic germ cell cluster into the initial Retinoic Acid (RA) responsive stage defined by expression of *ZGLP1* and *STRA8* [17] and prophase I of meiosis I, defined by expression of *SPO11* and *SYCP1* [16–18]. This analysis shows that as PGCs respond to retinoic acid to enter meiosis, *FGFR3* mRNA is still expressed and instead becomes repressed as the meiotic germ cells express *SPO11 and SYCP1*, which was also confirmed in the Li data set (Fig. 1g and h). In summary, utilizing scRNA-seq data from either the 10x Genomics [16] or SMART-seq [17] platform, we show that *FGFR3* mRNA is expressed by PGCs in the prenatal ovary and becomes repressed as the PGCs enter prophase I of meiosis I towards the formation of primordial oocytes

FGFR3 protein is expressed by ovarian PGCs and small VASA+ germ cells in ovarian cortical cords

Since RNA and protein may have different expression dynamics [19], we next evaluated expression of FGFR3 protein in fetal ovaries using immunofluorescence from week 7-14 pf (Fig. 2). This time window of prenatal life was chosen because it corresponds to the window of human development when PGCs initiate the process of meiotic entry within the cortical cords of the developing ovary (Fig. 2a) [6]. Human PGCs are detected by immunostaining for TFAP2C+/VASA^{low} [6], whereas meiotic germ cells in the cortical cords, are defined as TFAP2C-/VASA+ (Fig. 2a). Using these criteria at week 7, 13 and 14 pf, we show that >90% of TFAP2C+/VASA^{low} PGCs express FGFR3 (Fig 2b). In contrast, when focusing on the TFAP2C-/VASA+ germ cells in the cortical cords, ~60% of VASA+ germ cells are positive for FGFR3 (Fig. 2c). These results indicate that FGFR3 protein is expressed by the majority of PGCs in the fetal ovary between week 7-14pf becoming repressed in the VASA+ meiotic germ cells located in the cortical cords.

Upon closer inspection of VASA+ germ cells in the cortical cords, it appeared that the FGFR3+ germ cells were smaller than VASA+/FGFR3- germ cells (Fig 2d). This is important as the transition of germ cells into the pachytene stage of meiosis is associated with an increase in cell size[20]. To quantify this, we measured the size of VASA+/FGFR3- and VASA+/FGFR3+ germ cell nuclei in the cords. This approach revealed that the FGFR3+ germ cells are significantly smaller than the FGFR3-negative germ cells (Fig 2e). Taken together, these results show that FGFR3 protein is expressed by the majority of PGCs in the prenatal ovary from week 7-14pf. Moreover, during cortical cord formation, the smaller VASA+ germ cells continue to express FGFR3+, whereas the larger VASA+ germ cells are FGFR3 negative.

Flow cytometry for FGFR3 enriches for ovarian PGCs and meiotic germ cells

Given that FGFR3 defines PGCs in the prenatal ovary between 7-14 weeks, we next evaluated whether FGFR3 could be used to enrich for germ cells from single cell suspensions of embryonic and fetal ovary cells using fluorescence activated cell sorting (FACS). For this purpose, we dissociated ovaries from two different developmental time points, one at week 8pf and the second at week 13pf and stained each single cell suspension with antibodies that recognize FGFR3 and are conjugated to Phycoerythrin (PE) (Fig 3a). Dead cells were excluded from this analysis (by selecting 7AAD negative cells), and negative gates were set based on unstained cells from the same samples. Since the absolute number of female germ cells in a single ovary at week 8 pf or 13 pf is below the limit of cells required for a single 10X Genomics scRNA-seq run[6], we collected FGFR3+ cells from each sample and combined them after FACS. We reasoned that the samples could be demultiplexed during subsequent data analysis, using discriminating single nucleotide polymorphisms (SNPs) present between different donors. Since we did not have access to parental reference genomes, we combined the samples using a discordant ratio of 3 (week 13 pf) to 1 (week 8 pf) (Fig. 3a). Following common SNP analysis, the assignment of cells to our samples was in the predicted 3:1 ratio. Therefore, we assigned 6,633 FGFR3+ cells to the week 13pf ovary and 2,371 cells to the week 8pf ovary (Fig. 3a). 261 and 72 cells were designated as doublets or could not be assigned and were therefore excluded from further analysis.

As predicted from the immunofluorescence analysis, performing FACS of embryonic ovaries at week 8 pf and fetal ovaries at week 13 pf resulted in an enriched population of germ cells (5213 cells) which were of equivalent identity regardless of whether the cells were isolated at week 8 or week 13 (Fig. 3b-d and Fig. S3). In addition, non-germ cells (2447 cells) were isolated, with gene expression analysis predicting these cells correspond to pre-granulosa (*FOXL2*) and endothelial (*PECAM1*) cells (Fig. 3b-c and Fig. S3a-b). The mRNA expression levels of FGFR3 in these cells was on average lower than germ cells at both week 8 pf and 13pf samples (Fig. S3c-d). We therefore asked whether FGFR3 protein is detectable in granulosa cells of the fetal ovary. For this, we performed immunofluorescence for the granulosa cell marker FOXL2 together with FGFR3 and show that FOXL2+ cells are negative for FGFR3 (Fig. S3c-d). Taken together, FACS for FGFR3+ enriches for fetal ovarian germ cells, however using this approach, ~30% of analyze cells after scRNA-seq are granulosa and endothelial cells.

To evaluate whether the ovarian germ cells enriched by FACS correspond to both PGCs and retinoic acid responsive meiotic germ cells, we clustered FGFR3+ germ cells into 7 different groups (Fig 3d,e). Gene expression analysis revealed that clusters 0-4 correspond to PGCs which express genes such as *NANOG, PRDM1, TFAP2C, SOX17, DAZL* and *VASA* and gene ontology (GO) analysis indicating an enrichment in terms associated with mitotic cell cycle genes (Fig. 3f, g). In contrast, cells in clusters 5 and 6 correspond to the meiotic germ cells with GO analysis indicating an enrichment in terms associated with gamete generation (Fig. 3f, g). Rare populations of cells expressing the meiotic prophase I genes including SPO11 were also identified (Fig. 3F and S3b). In

contrast, primordial oocytes characterized by expression of *zona pellucida 3 (ZP3)*, were absent in FGFR3 sorted germ cells (Fig. 3f). Taken together, these results demonstrate that separation of prenatal ovarian cells using FACS from week 8 and 13 ovaries with antibodies that recognize FGFR3 results in the enrichment of female PGCs and meiotic germ cells but not ZP3+ primordial oocytes.

FGFR3 RNA and protein is expressed by early PGCs in vivo

The specification of PGCs begins at the end of week 2 pf [1]. This is approximately 3-4 weeks before the somatic cells of the ovary develop from the genital ridge epithelium at week 5-6 pf to form the ovaries. In order to evaluate FGFR3 expression in PGCs at the time of genital ridge colonization, we performed immunofluorescence of a single human embryo at week 5 where both early-PGCs defined as OCT4+/DAZL- and late-PGCs defined as OCT4+/DAZL+ are identified (Fig 4a). Evaluation of FGFR3 in this specimen revealed FGFR3 protein signal on all OCT4+/DAZL- PGCs (11 cells) as well as the single OCT4+/DAZL+ PGC (Fig. 4a). This result indicates that FGFR3 is expressed by early PGCs at the time of gonadal colonization. To confirm this result we examined week 4 and 5pf SMART-seq scRNA-seq data of Li et al [17] (Fig. 4b-c) and found *FGFR3* expression in the majority of *Pou5f1* (OCT4+) PGCs at this early gestational time point.

To evaluate the expression of FGFR3 in PGC like cells (PGCLCs) *in vitro*, we differentiated human embryonic stem cell (hESCs) using a two-step differentiation protocol involving an incipient mesoderm-like cell (iMELC) intermediate as previously described [13]. In this model, PGCLC induction begins 24-48 hours after BMP4 exposure

with fully specified PGCLCs identified at day 4 (D4) [1]. Examining three independent differentiation experiments of the hESC line UCLA1 at D4 using the 10X Genomics RNA-seq pipeline reveals a distinct cluster of PGCLCs defined as co-expressing *NANOS3, NANOG* and *SOX17* (Fig. 4d) [10]. Displaying *FGFR3* RNA expression on this data set shows that FGFR3 mRNA is detected in the PGCLC population, and the SOX17+ endodermal cells as well as sub-population of HAND1+ mesoderm cells (Fig. 4e-g). To evaluate protein expression, we performed immunofluorescence at day 4 following BMP4 exposure. Using immunofluorescence FGFR3 protein is not detectable in the clusters of OCT4+/TFAP2C+ PGCLCs at D4 (Fig. 4h and Fig. S4). Taken together, FGFR3 protein is expressed by PGCs *in vivo* from as early as week 5 pf and at the RNA level expressed in hPGCLCs at D4.

Discussion

In this work, we show that FGFR3 is expressed by PGCs in prenatal ovaries and testis. Additional characterization of FGFR3 expression by scRNA-seq and immunofluorescence revealed that female PGCs maintain FGFR3 expression into the initial stages of meiotic progression including the retinoic acid responsive stage of meiosis and the beginning of prophase I of meiosis I. However, ZP3+ primordial oocyte formation is associated with the repression of FGFR3.

In adult mouse ovaries, FGFR3 is expressed on the membrane of granulosa cells in the growing follicle, but not by the oocyte or the theca cells [21]. Consistent with this expression pattern, FGFR3 activating mutations in anchondroplastic female mice causes

loss of granulosa cells and ultimately infertility [22]. In adult bovine ovaries, FGFR3 was also detected in granulosa cells, with FGFR3 expression levels positively associated with increasing exposure to follicle stimulating hormone [23]. In the current study we show that the expression pattern of FGFR3 in the embryonic and fetal ovary is different from the adult with FGFR3 predominantly expressed by the PGCs and meiotic germ cells (Fig. 1). Our data suggest that FOXL2+ embryonic and fetal pre-granulosa cells do not express FGFR3(Fig. S3).

In the current study we show that FGFR3 is expressed by male PGCs and fetal Prospermatogonia (Fig. S1). Although the expression of FGFR3 had not been evaluated in the adult human testis, in the adult mouse testis, FGFR3 is expressed by Spermatogonia in vivo [23–25] and also *in vitro* [27]. As adult Spermatogonia are recruited into meiosis, FGFR3 is repressed [28]. This suggests that FGFR3 might have a similar role in the survival and proliferation of diploid germline cells prior to entering meiosis [28]. Indeed, rare gain of function mutations in FGFR3 gene have been associated with abnormal sperm formation and testicular cancer [29], [30].

In this work, we also demonstrate that sorting for FGFR3 positive cells, is a viable strategy that will enrich for human PGCs and meiotic germ cells from the embryonic and fetal ovary. In addition to enriching the germline, FGFR3+ FACS will also isolate somatic cells that could be due to wider FACS gates to increase number of cells and isolate recommended number of cells for 10X genomics scRNA-seq (Fig. S3). Therefore, moving forward, using anti-FGFR3 antibodies could be a powerful marker in combination with

other markers such as cKIT [6] to isolate PGCs and meiotic germ cells from prenatal ovaries.

In summary, FGFR3 is expressed predominantly by PGCs and meiotic germ cells in prenatal ovaries. It can be used to enrich for those germ cell populations and potentially used as a diagnostic surface marker for further improvements of PGCLC differentiation protocols.

Experimental procedures

Human fetal tissues

University of Washington Birth Defects Research Laboratory (BDRL) provided all prenatal gonads (5-15 weeks p.f.) for this study. At the BDRL, prenatal gonads were obtained with regulatory oversight from the University of Washington IRB approved Human Participants protocol, combined with a Certificate of Confidentiality from the Federal Government. BDRL collected the fetal ovaries and shipped them overnight in HBSS with an ice pack for immediate processing at UCLA. All human fetal tissues used here was obtained following informed consent. No personal identifiers were carried about the tissues sent to UCLA. Donors have not received any payments as they knowingly and willingly consented to provide research materials without restrictions for research and for use without identifiers. Developmental age was documented by the BDRL.

Immunofluorescence

Slides of paraffin-embedded sections were deparaffinized by successive treatment with xylene and 100%, 95%, 70% and 50% ethanol. Antigen retrieval was performed by incubation with 10 mM Tris pH 9.0, 1 mM EDTA, 0.05% Tween-20 at 95 °C for 40 min. The slides were cooled to room temperature and washed with 1× PBS and 1× TBS (PBS + 0.2% Tween-20). Afterwards, the samples were permeabilized with 0.5% Triton X-100 in 1× PBS, then washed with 1× TBS and blocked with 5% normal donkey serum in 1× TBS. Primary antibody incubation was conducted with 5% normal donkey serum overnight at 4C. Samples were again washed with 3× TBS-Tween-20 and incubated with fluorescent secondary antibodies at 1:100 for 1 hour, then washed and counterstained with DAPI for 5 min and mounted using Vectashield. A list of the primary antibodies used for immunofluorescence in this study is provided in Supplementary Table 1 under the antibody list tab. The secondary antibodies used in this study were all obtained from Life technologies and were used at 1:400 dilution. Images were taken using LSM 880 Confocal Instrument (Zeiss) or Zeiss Axio Imager M1. For image processing and analysis, Fiji (ImageJ) was used. For nuclear size quantification, images were converted into 8-bit images and then analyzed using profile plot tool. Intensity values were exported as a CSV file and then R Studio and the ggplot2 package was used for plotting.

PGCLC differentiation

PGCLCs were induced from primed UCLA1 and UCLA2 hESCs as described previously [10] starting with human pluripotent stem cells grown on MEFs. In brief, hESCs and hiPSCs were dissociated into single cells with 0.05% trypsin-EDTA (GIBCO, 25300-054)

and plated onto human-plasma-derived fibronectin-coated (Invitrogen, 33016-015) 12-well plates at a density of 200,000 cells per well in 2 ml per well of iMeLC medium (15% KSR (GIBCO, 10828-028), 1× penicillin–streptomycin–glutamine (GIBCO, 10378-016), 0.1 mM 2-mercaptoethanol (GIBCO, 21985-023), 1 mM sodium pyruvate (GIBCO, 11360-070), 1× NEAA (GIBCO, 11140-050), 3 mM CHIR99021 (Stemgent, 04-0004), 10 mM of ROCKi (Y27632, Stemgent, 04-0012-10), 50 ng ml^{-1} activin A (Peprotech, AF-120-14E), and 50 ng ml^{-1} primocin in Glasgow's MEM (GMEM) (GIBCO, 11710-035)). After 24 h, using 0.05% trypsin, iMeLCs were dissociated into single cells and plated into ultra-low cell attachment U-bottom 96-well plates (Corning, 7007) at a density of 3,000 cells per well in 200 ml per well of PGCLC medium, which is composed of 15% KSR (GIBCO, 10828-028), 1× NEAA (GIBCO, 11140-050), 0.1 mM 2-mercaptoethanol (GIBCO, 21985-023), 1 mM sodium pyruvate (GIBCO, 11360-070), 10 ng per ml^{-1} human LIF (Millipore, LIF1005), 1× penicillin–streptomycin-–glutamine (GIBCO, 10378-016), 200 ng ml^{-1} human BMP4 (R&D systems, 314-BP), 50 ng ml^{-1} human EGF (R&D systems, 236-EG), 10 mM of ROCKi (Y27632, Stemgent, 04-0012-10) and 50 ng ml^{-1} primocin in GMEM (GIBCO, 11710-035). Day 4 aggregates were collected and embedded in a paraffin block that was afterwards used for sectioning as described before [10].

Tissue processing for scRNA-seq

Fetal tissues were processed 24–48 h after termination. On arrival, tissues were washed with PBS and dissociated using mix of collagenase IV 10 mg ml^{-1} (Life Technologies, 17104-019), dispase II 250 μg ml^{-1} (Life Technologies, 17105041), DNase I 1:1,000

(Sigma-Aldrich, 4716728001), 10% fetal bovine serum (Life Technologies, 10099141) in 1× PBS. Tissues were dissociated for 15 min at 37 °C. In every 5 min, the tissues were pipetted against the bottom of Eppendorf tube using p1000 pipette. Afterwards, cells were centrifuged for 5 min at 500*g*, resuspended in 1× PBS with 0.04% BSA, strained through a 40 µm strainer to get rid of clumps and counted using an automated cell counter (Thermo Fisher Scientific, Countess II). Afterwards cells were used for FACS sorting.

FACS sorting

For FACS sorting tissues were dissociated as described above. The dissociated cells were stained with conjugated antibodies, washed with FACS buffer (1% BSA in PBS) and resuspended in FACS buffer with 7-AAD (BD PharMingen, 559925) as viability dye. The conjugated antibodies used in this study FGFR3 conjugated with PE (R&D systems FAB766P), 1:60 dilution. The single-cell suspension was sorted for further experiments using BD FACSAria FACS machine. FACS data were analyzed using FlowJo v.10. Positive cells for FGFR3 were collected in PSB+0.04% BSA. Cells from week 8 and week 13 tissues were mixed at 1:3 ratio to generate a single 10X library.

scRNA-seq library preparation

scRNA-seq libraries were generated using the 10x Genomics Chromium instrument and Chromium Single Cell 3′ Reagent Kit v3. Library was designed to target 10,000 cells and library was generated according to the manufacturer's instructions and library fragment size distribution was determined using a Tapestation instrument. Library was sequenced using an Illumina Novaseq 6000 platform, at an average depth of 300–350 million reads per sample.

scRNA-seq data analysis

scRNA-seq reads were aligned to the human hg38 genome assembly using 10x Genomics Cell ranger v3.4. Expression matrixes generated by Cell Ranger were imported into Seurat or Scanpy [31] for downstream analysis. First, all of the libraries were merged, and cells were filtered in the same manner. All of the genes that were expressed in less than five cells were discarded and cells with less than 250 detected genes were filtered out. The unique molecular identifier (UMI) counts were then normalized for each cell by the total expression, multiplied by 10,000 and log-transformed. Using Scanpy's default method, highly variable genes were identified, and data were scaled to regress out variation from UMI counts and mitochondrial genes. Cells were clustered using the Louvain algorithm and the UMAP package was used to visualize cells in a two-dimensional plot. Germ cell clusters were identified by expression of germ-cell-specific markers, such as *NANOS3*, *DAZL*, *DDX4* and *SYCP1*. Gonadal somatic cells were annotated by previously published literature[4], [16], [17].

Demultiplexing samples using SNPs

To demultiplex cells from scRNA-seq library and assign them to week 8 or week 13 sample, we have first aligned to data to human hg38 genome assembly using 10X Genomics Cellranger. Afterwards a bam file generated by cell ranger was used in cellSNP-lite tool that can detect expressed alleles in scRNA-seq datasets[32]. cellSNP-lite uses list of candidate SNPs from 1000 Genome Project. After creating the list of expressed SNPs output file was passed into Vireo to deconvolve the donors[33].

Following deconvolution of the donors we got 6633 cells from donor 1, 2370 cells from donor 2, while 261 cells were assigned as doublets and 72 cells were not assigned to any donor. Doubles and unassigned cells were filtered out from downstream analysis. Since we mixed the cells at 1:3 ratio at the beginning of the experiment we could assign that 6633 cells were from week 13 ovary and 2370 from week 8.

Data availability

The scRNA-seq data of prenatal tissues reported in this work are available under the following accession numbers: GSM5808297 (FGFR3 sorted cells) and GSM5808298 (UCLA1 D4 aggregate). Previously published scRNA-seq data from prenatal ovarian and testicular cells [16], FGCs [17] and D4 aggregates [1] under following accession numbers: GSE143380 (ovarian cells), GSE143356 (testicular cells)[16], GSE86146[17] and GSE140021[1].

Code availability

Custom scripts used for aligning scRNA-seq, data processing and plotting are available upon request.

Figure 1. FGFR3 mRNA is expressed by PGCs in the prenatal human ovary.

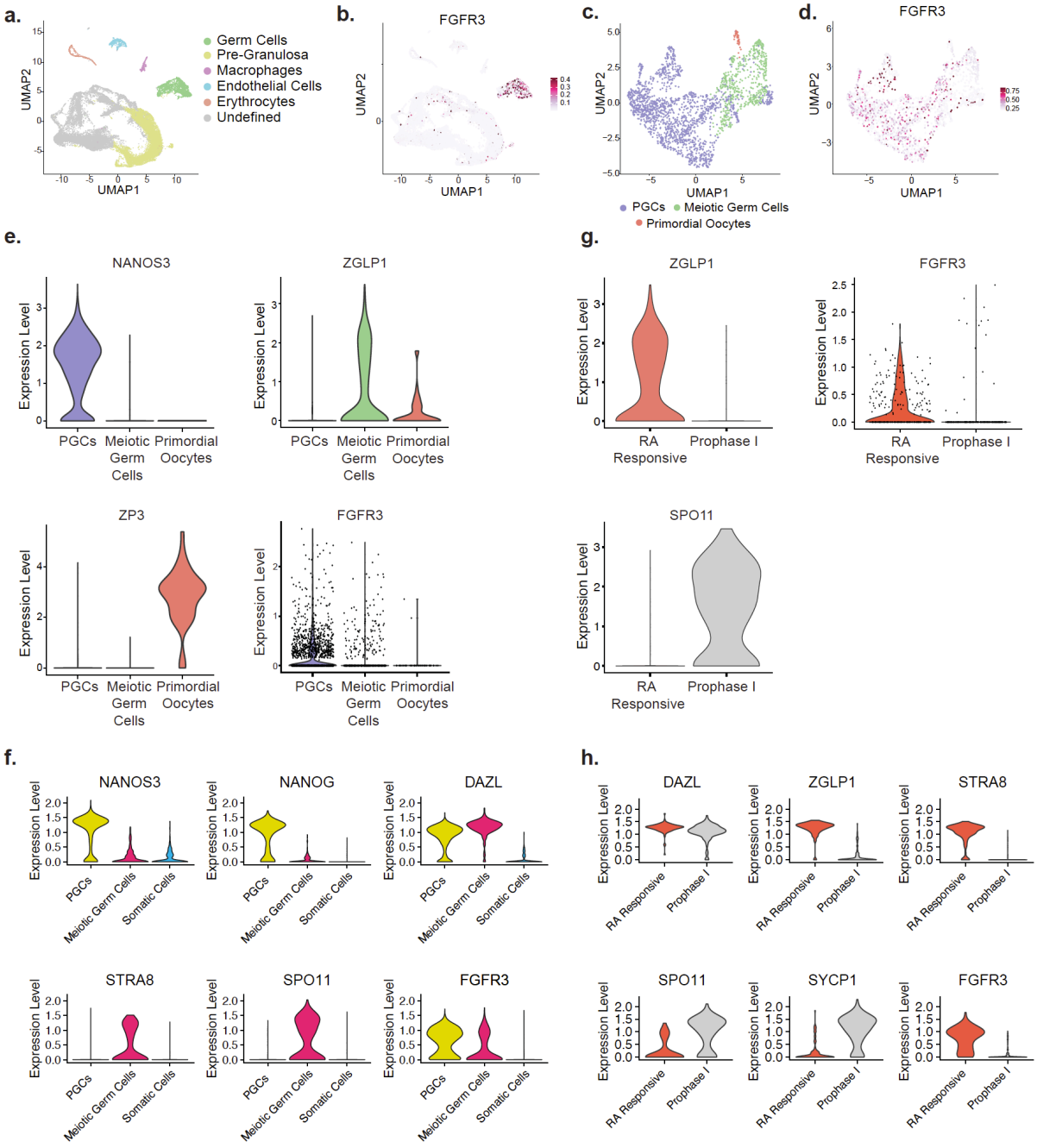


Figure 1. FGFR3 mRNA is expressed by PGCs in the prenatal human ovary

a. Annotation of ovary cell types based on expression of cell type-specific markers. **b.** Expression of FGFR3 in ovarian cells data from. **c.** Stage specific annotation of ovarian

germ cells. **d.** FGFR3 expression if female germ cells. **e.** Expression of PGC, meiotic germ cell and primordial oocyte specific markers together with FGFR3 in female germ cells. Data from Chitiashvili et al. [16] **f.** FGFR3 expression in ovarian germ and somatic cells. Data from Li et al [17] **g.** FGFR3 expression in sub-clusters of meiotic germ cells - RA responsive and Prophase I meiosis I germ cells, together with sub-cluster specific genes. Data from Chitiashvili et al [16]. **h.** Same as (g) except Li et al dataset [17].

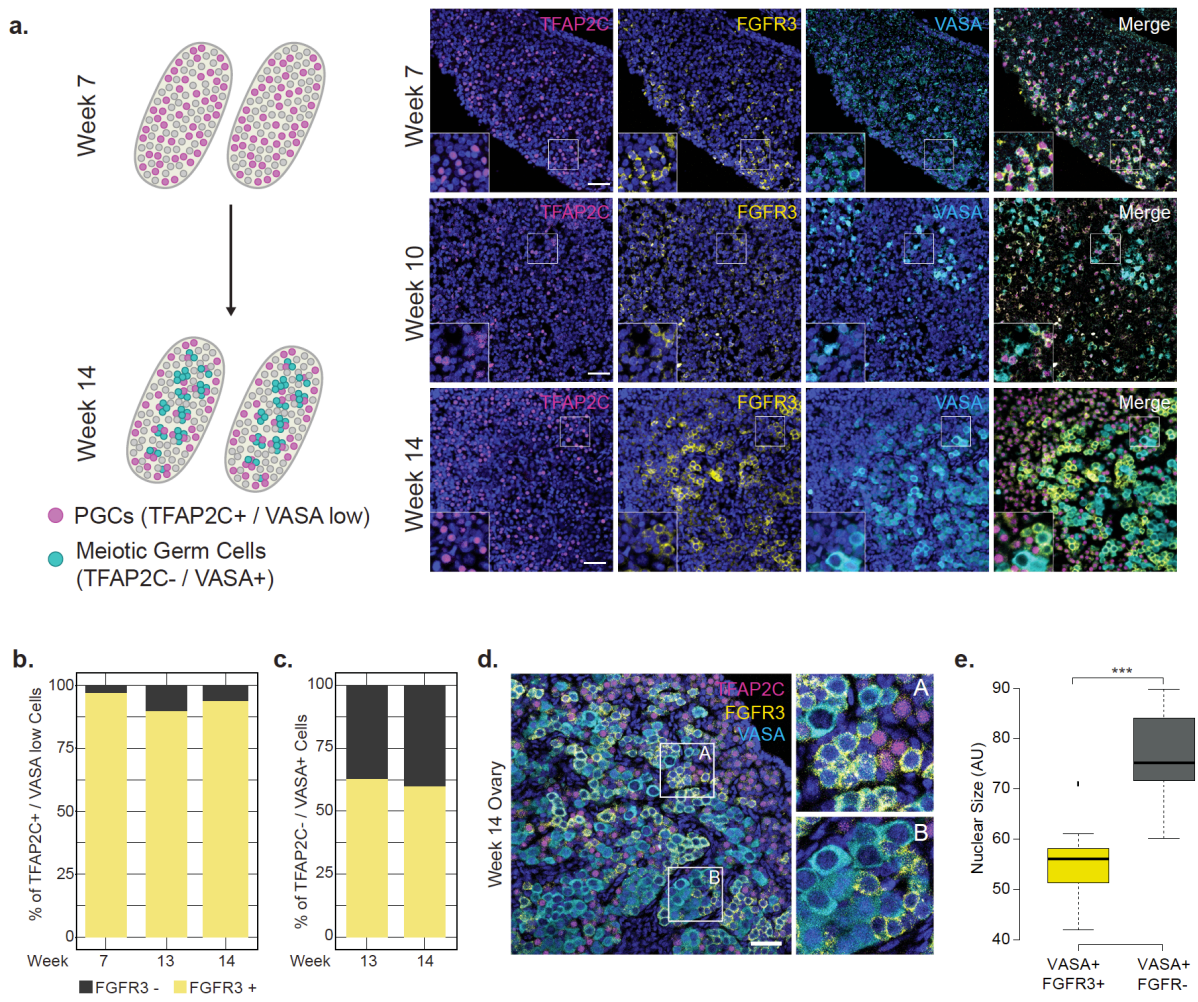


Figure 2. FGFR3 protein is expressed by PGCs in prenatal human ovary.

a. Schematic representation of germ cell differentiation in ovaries. TFAP2C+/VASA low PGCs mostly reside at the cortex of the ovary. With time, PGCs differentiate and advance towards meiosis by upregulating expression of VASA and silence TFAP2C. Right hand side immunostaining of ovarian tissues at week 7, 10 and 14 with TFAP2C (Magenta), FGFR3 (Yellow), VASA (Cayan) and DAPI (Blue). **b.** Quantification of the percentage of FGFR3 positive and negative cells that express TFAP2C and low levels of VASA (50 cells

were counted for each sample). **c.** Quantification of FGFR3 positive and negative cells that are negative for TFAP2C and positive for VASA (n=3 samples for each category, 50 cells were counted from each sample). **d.** Immunostaining of 14-week ovary with TFAP2C (magenta), FGFR3 (yellow), VASA (cyan) and DAPI (blue). **e.** Nuclei size of VASA+/FGFR3+ and VASA+/FGFR3- ovarian germ cells (25 cells were measured for each category from 14-week ovary). ***P<0.001, Statistical significance was assessed by Wilcoxon test.

Figure 3. FGFR3 can enrich for PGCs from single cells suspension of the prenatal ovary.

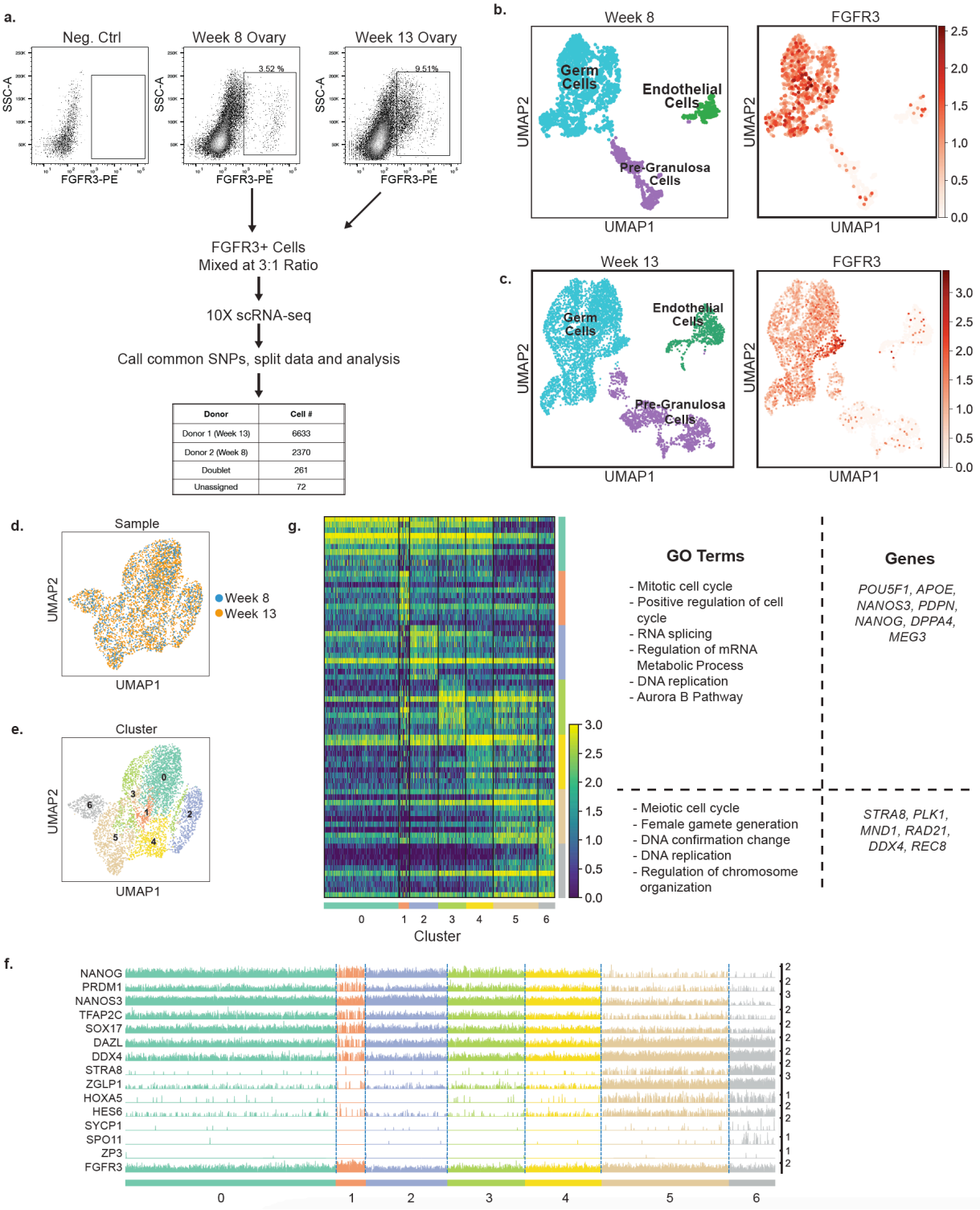


Figure 3. FGFR3 can enrich for PGCs from single cells suspension of the prenatal ovary.

a. Sorting strategy to isolate female germ cells based on FGFR3 staining. Cells from week 8 and 13 ovaries were mixed at 1:3 ratio accordingly to create one 10X genomics scRNA-seq library. After sequencing, library was split into two by identifying common SNPs. **b.** Annotation of FGFR3 sorted cells from week 8 ovary and expression of FGFR3 right-hand side. **c.** Similar to (b) for week 13 cells. **d.** UMAP clustering of germ cells from week 8 and week 13 samples. **e.** Sorted germ cells displayed on UMAP plot by their cluster numbers. **f.** Ordering of sorted germ cells along the developmental trajectory from cluster 0 to cluster 6, with classification into PGCs (clusters 0-4), meiotic germ cells (clusters 5-6), based on diagnostic germ cell marker expression. Each cluster contains individual cells (columns), for which expression of indicated marker is given (rows). **g.** Heatmap representation of top 10 expressed genes per cluster of germ cells from (f). Gene ontology analysis highlighting terms and top expressed genes per category.

Figure 4. FGFR3 mRNA and protein are expressed by PGCs before gonadal colonization.

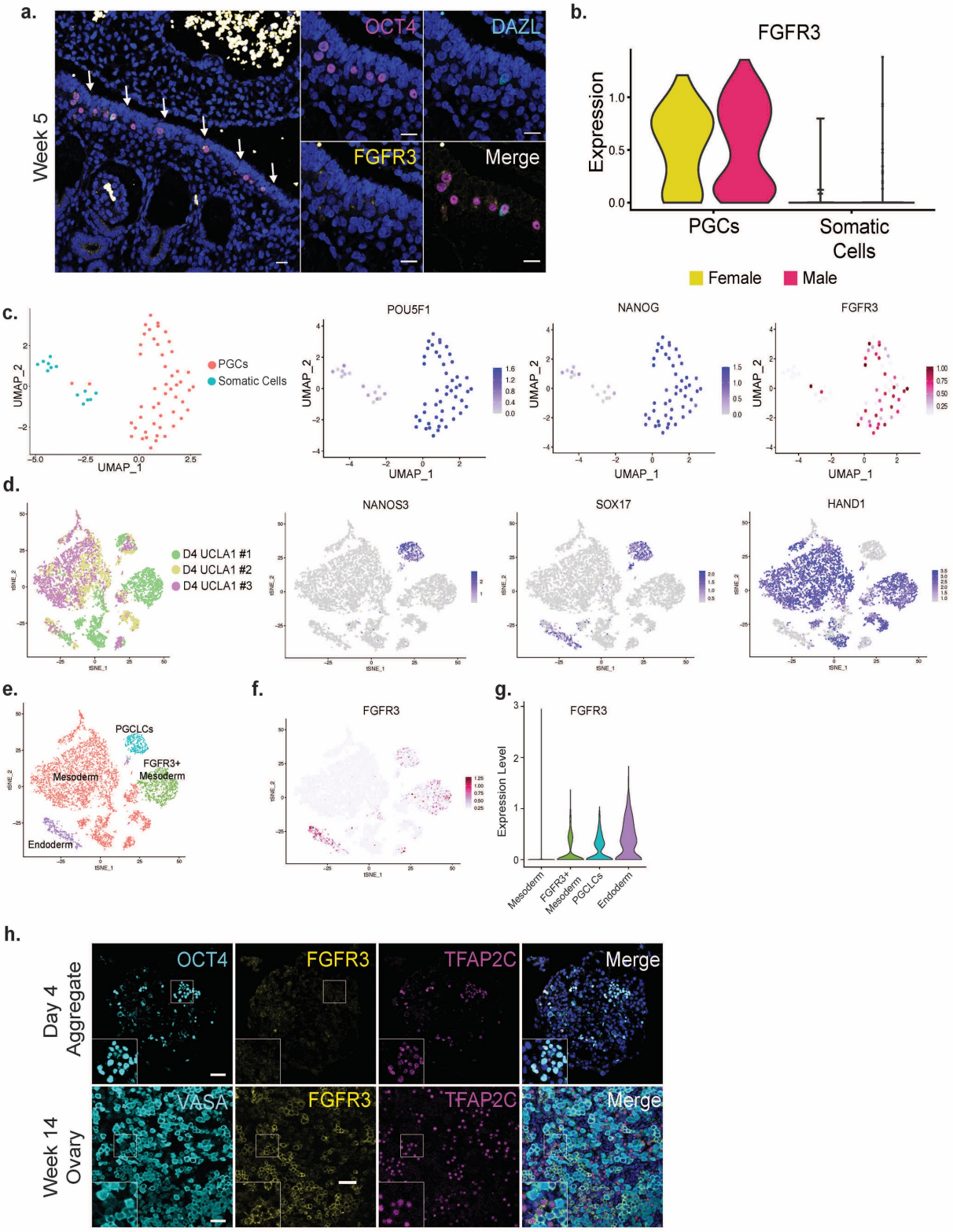
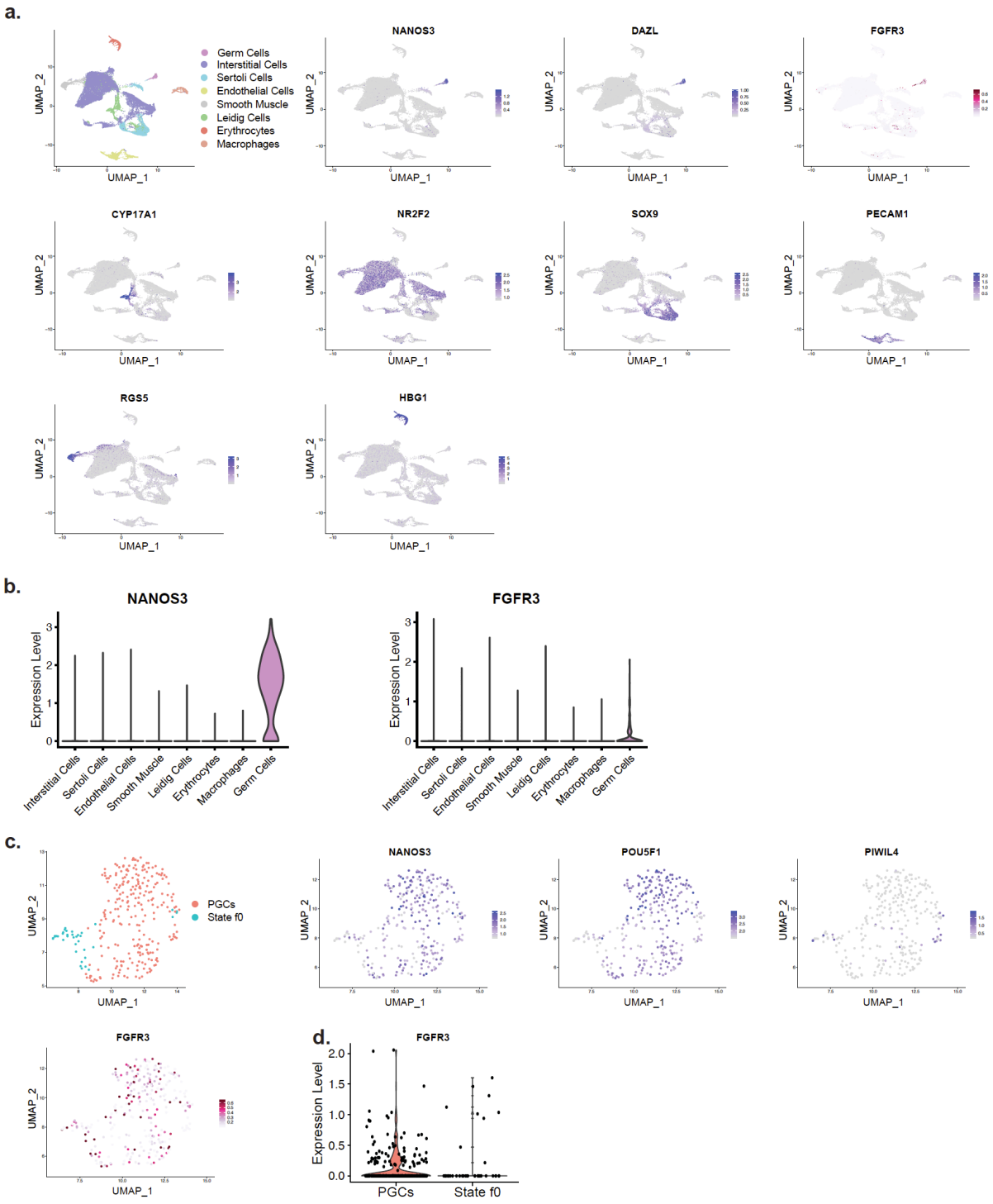


Figure 4. FGFR3 mRNA and protein are expressed by PGCs before gonadal colonization.

a. Immunostaining of week 5 aorta-gonad-mesonephros. OCT4 (magenta), DAZL (cyan), FGFR3 (yellow) and DAPI (blue), n=1 sample. **b.** Expression of FGFR3 in female and male PGCs at week 4 and 5 from Li et. al dataset [17]. **c.** Week 4 and 5 germ cells and somatic cells depicted on a UMAP and expression of POU5F1 (OCT4), NANOG and FGFR3. **d.** Day 4 PGCLCs in three replicates[1] differentiated from UCLA1 hESCs and expression of diagnostic markers: NANOS3 for PGCLCs, SOX17 – PGCLCs and endoderm cells, HAND1 for mesoderm cells of day 4 aggregates. **c.** Annotation of D4 aggregate cells based on marker expression from (d). **f.** Expression of FGFR3 depicted at tSNE maps and violin plots (**g**). **h.** UCLA1 female hESCs differentiated for 4 days to create PGCLC aggregates and immunostained for OCT4 (cyan), FGFR3 (yellow) and TFAP2C (magenta), DAPI (blue), (n=2 differentiation experiments). Lower panel week 14 ovary staining with VASA (cyan), FGFR3 (yellow) and TFAP2C (magenta) used as a positive control for FGFR3 staining.

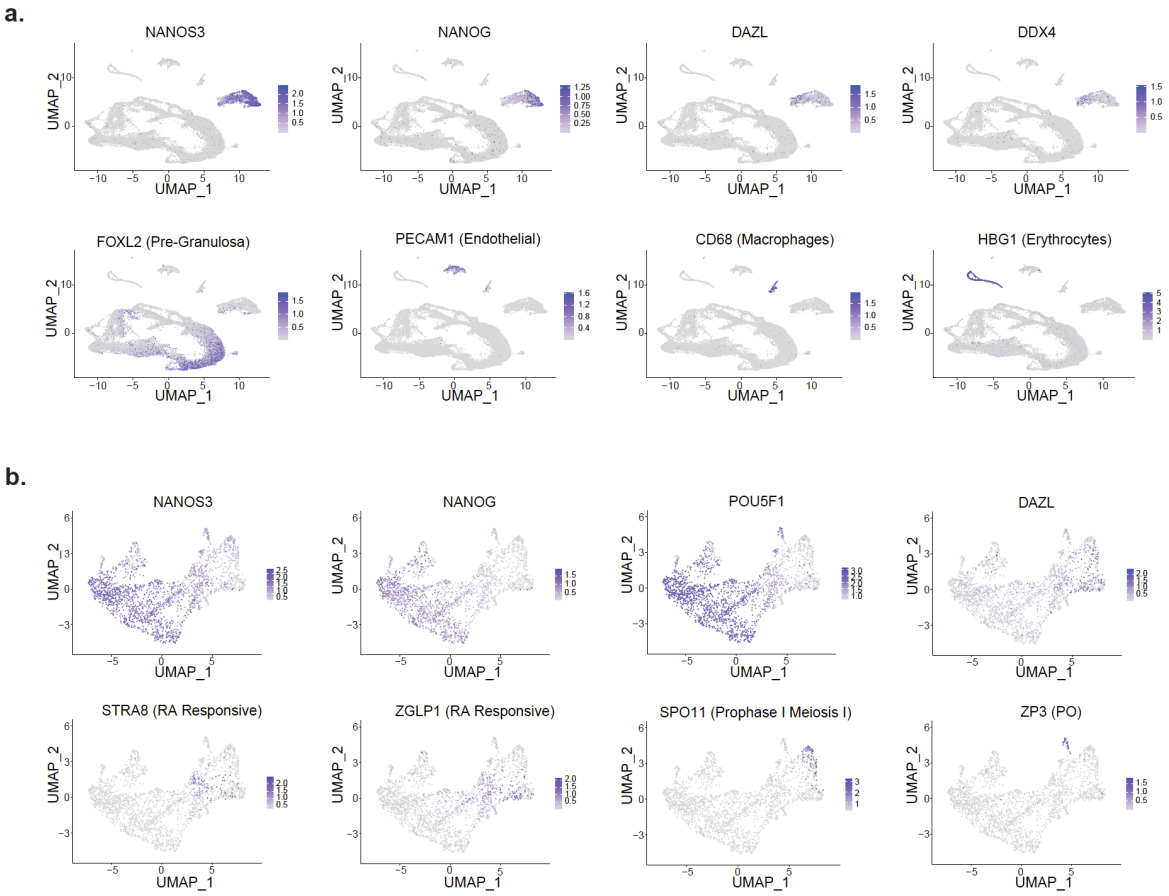
Figure S1. FGFR3 mRNA is expressed on germ cells in the prenatal human testis.



Supplemental Figure 1. FGFR3 mRNA is expressed on germ cells in the prenatal human testis.

a. Annotation of testis cell types based on diagnostic markers for each cell type. NANOS3 and DAZL for germ cells. NR2F2 marks interstitial cells, CYP17A1 marks leidig cells, PECAM1 endothelial cells, HBG1 erythrocytes, RGS5 smooth muscle cells, SOX9 sertoli cells. **b.** Expression of PGC specific marker NANOS3 and FGFR3 in testicular somatic cells and PGCs. **c.** Annotation of testicular germ cells based on their stage specific marker expression. NANOS3 and POU5F1 marking PGCs and PIWIL4 marking stage f0 prospermatogonia. FGFR3 expression in PGCs and state f0 bottom panel. **d.** Expression of FGFR3 in male PGCs and state f0 prospermatogonia.

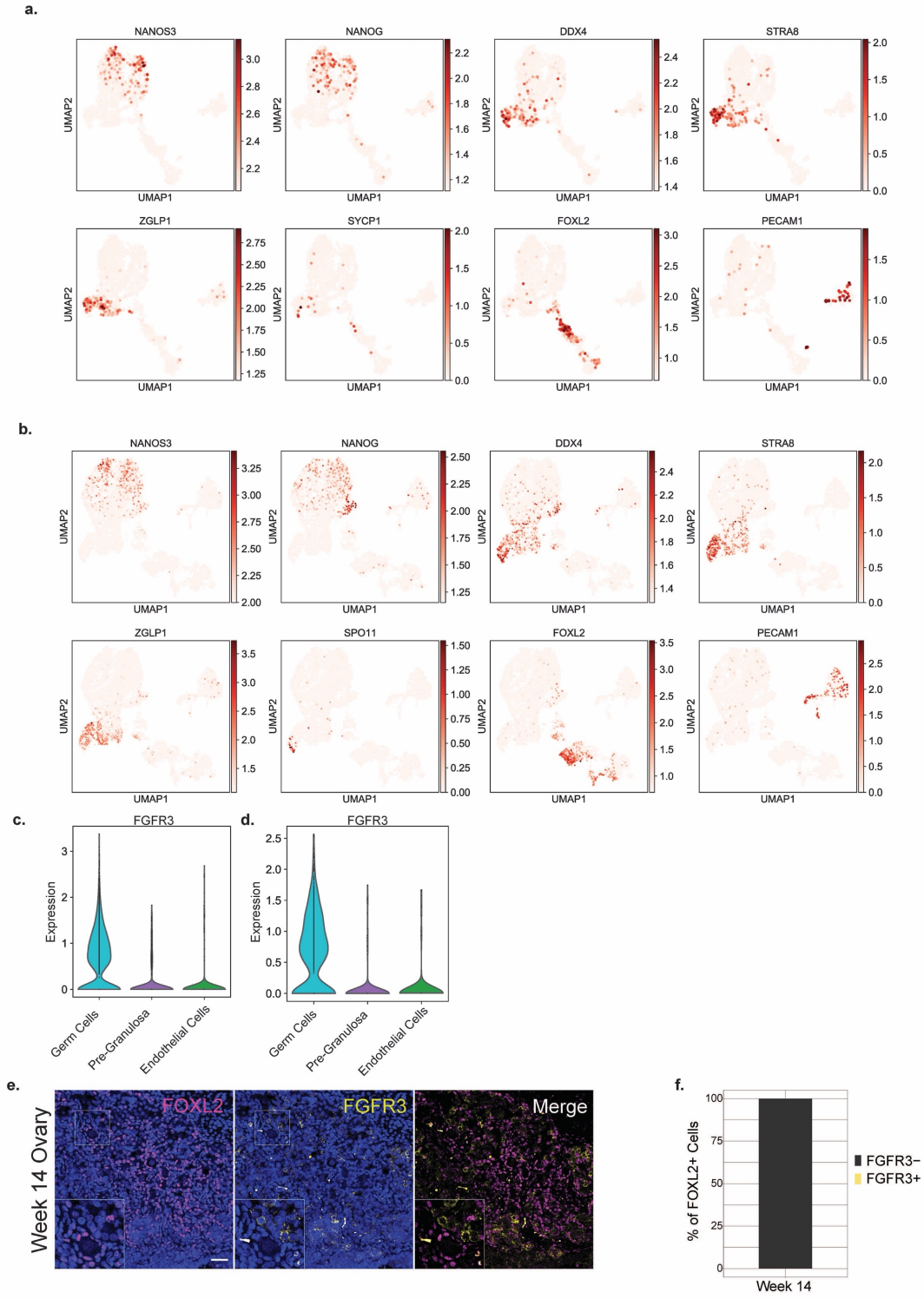
Figure S2. Cell type annotation of prenatal ovaries



Supplemental Figure 2. Cell type annotation of prenatal ovaries.

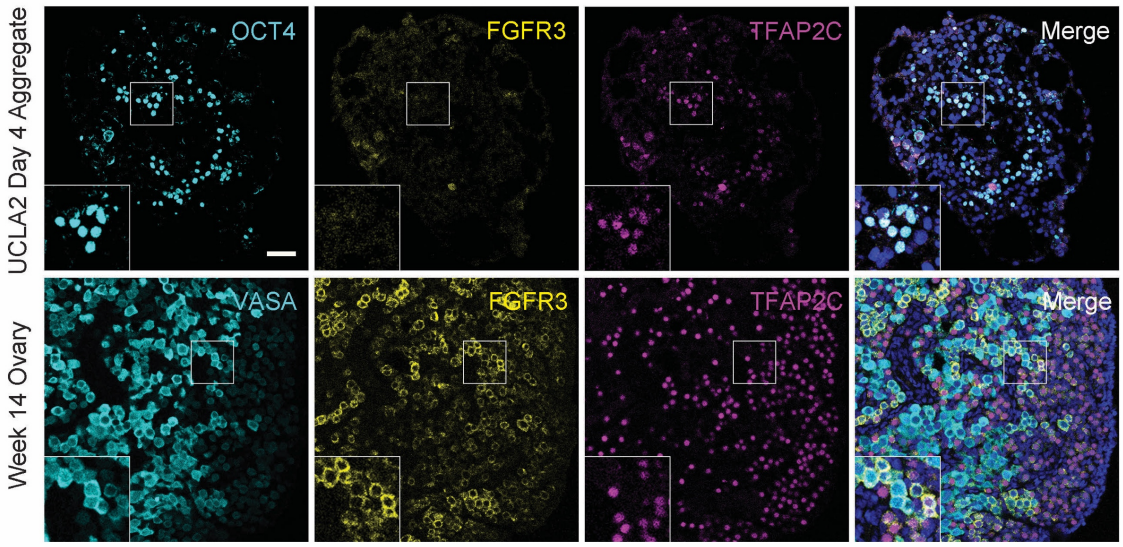
a. Expression of diagnostic markers for cell type annotation of ovarian cells. NANOS3, NANOG, DDX4 and DAZL marking germ cells, FOXL2 pre-granulosa cells, PECAM1 endothelial cells, CD68 macrophages and HBG1 erythrocytes. **b.** Expression of germ cell stage specific markers in female germ cells. NANOS3, NANOG, POU5F1 and DAZL expressed in PGCs. STRA8 and ZGLP1 in retinoic acid (RA) responsive meiotic germ cells. SPO11 in prophase I of meiosis I and ZP3 marker of primordial oocytes.

Figure S3. FGFR3 mRNA is PGC specific in the prenatal human ovary.



Supplemental Figure 3. FGFR3 mRNA is PGC specific in the prenatal human ovary. **a.** Expression of diagnostic markers for cell type annotation in week 8 and week 13 (**b**) ovarian samples. NANOS3 and NANOG marking PGCs, DDX4, STRA8, ZGLP1 and SYCP1 meiotic germ cells, while FOXL2 and PECAM1, pre-granulosa and endothelial cells accordingly. **c.** Expression of FGFR3 in week 8 Germ cells, Pre-granulosa and endothelial cells. **d.** Similar as c for week 13 cells. **e.** Immunostaining of Week 14 ovary for pre-granulosa marker FOXL2 (magenta) and FGFR3 (yellow) n=2 experiments. **f.** Quantification of FGFR3- and FGFR3+ cell proportion that are FOXL2+ (70 cells counted from week 14 ovary from 2 independent experiments).

Figure S4. FGFR3 protein is not present in PGCLCs *in vitro*



Supplemental Figure 4. FGFR3 protein is not present in PGCLCs *in vitro*.

UCLA2 hESCs differentiated into PGCLCs for 4 days. At day 4 aggregates were immunostained for OCT4 (cyan), FGFR3 (yellow) and TFAP2C (magenta), n=2 differentiation experiments. Lower panel week 14 ovary staining with VASA (cyan), FGFR3 (yellow) and TFAP2C (magenta) used as a positive control for FGFR3 staining.

Acknowledgements

We would like to thank Flow cytometry, Next Generation Sequencing and microscopy cores at the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research Center (BSCRC) for help with cell sorting, sequencing and imaging. Also, we would like to thank Technology Center for Genomics and Bioinformatics at the UCLA Johnson Comprehensive Cancer Center (JCCC) and the Translational Pathology Core Laboratory for help with histology. T.C. was supported by Boehringer Ingelheim PhD Fellowship. This work is supported by funds from the NIH to ATC (R01HD079546). K.P. was supported by the BSCRC at UCLA, the David Geffen School of Medicine at UCLA, and the UCLA JCCC, the NIH (R01HD098387, P01GM099134), and a Faculty Scholar grant from the Howard Hughes Medical Institute. Human fetal tissue research is supported by a grant to Ian Glass at the University of Washington Birth Defects laboratory, 5R24HD000836-53. Human conceptus tissue requests can be made to bdrl@u.washington.edu.

Author contributions

T.C, K.P, and A.C. designed the experiments. T.C conducted immunofluorescent experiments on tissues and PGCLC sections. F.H. contributed to hPGCLC differentiation experiments. T.C. performed scRNA-seq experiments, and the resulting data were analyzed by T.C and I.D. T.C., K.P. and A.C. interpreted all data and wrote the manuscript.

Declaration of interests

The authors declare no competing interests.

References

[1] D. Chen *et al.*, “Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors,” *Cell Rep.*, vol. 29, no. 13, pp. 4568-4582.e5, Dec. 2019.

[2] D. Chen, J. J. Gell, Y. Tao, E. Sosa, and A. T. Clark, “Modeling human infertility with pluripotent stem cells,” *Stem Cell Res.*, vol. 21, pp. 187–192, 2017.

[3] W. W. C. Tang, T. Kobayashi, N. Irie, S. Dietmann, and M. A. Surani, “Specification and epigenetic programming of the human germ line,” *Nat. Rev. Genet.*, vol. 17, no. 10, pp. 585–600, 2016.

[4] J. Guo *et al.*, “Single-cell analysis of the developing human testis reveals somatic niche cell specification and fetal germline stem cell establishment,” *Cell Stem Cell*, 2021.

[5] I. Konishi, S. Fujii, H. Okamura, T. Parmley, and T. Mori, “Development of interstitial cells and ovigerous cords in the human fetal ovary: an ultrastructural study,” *J. Anat.*, vol. 148, pp. 121–135, Oct. 1986.

[6] S. Gkountela *et al.*, “The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation,” *Nat. Cell Biol.*, vol. 15, no. 1, pp. 113–122, 2012.

[7] S. Gkountela *et al.*, “DNA demethylation dynamics in the human prenatal germline,” *Cell*, vol. 161, no. 6, pp. 1425–1436, 2015.

[8] N. Irie *et al.*, “SOX17 is a critical specifier of human primordial germ cell fate,” *Cell*, vol. 160, no. 1–2, pp. 253–268, 2015.

[9] F. Guo *et al.*, “The transcriptome and DNA methylome landscapes of human primordial germ cells,” *Cell*, vol. 161, no. 6, pp. 1437–1452, 2015.

[10] D. Chen *et al.*, “Germline competency of human embryonic stem cells depends on eomesodermin†,” *Biol. Reprod.*, vol. 97, no. 6, pp. 850–861, Oct. 2017.

[11] D. Chen *et al.*, “The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation,” *Cell Rep.*, vol. 25, no. 13, pp. 3591-3602.e5, Dec. 2018.

[12] G. R. MacGregor, B. P. Zambrowicz, and P. Soriano, “Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells,” *Development*, vol. 121, no. 5, pp. 1487–1496, May 1995.

[13] K. Sasaki *et al.*, “Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells,” *Cell Stem Cell*, vol. 17, no. 2, pp. 178–194, 2015.

[14] M. Gomes Fernandes, M. Bialecka, D. C. F. Salvatori, and S. M. Chuva de Sousa Lopes, “Characterization of migratory primordial germ cells in the aorta-gonad-mesonephros of a 4.5-week-old human embryo: a toolbox to evaluate in vitro early gametogenesis,” *Mol. Hum. Reprod.*, vol. 24, no. 5, pp. 233–243, May 2018.

[15] K. A. Ewen *et al.*, “Expression of FGFR3 during human testis development and in germ cell-derived tumours of young adults,” *Int. J. Dev. Biol.*, vol. 57, no. 2–4, pp. 141–151, 2013.

[16] T. Chitiashvili *et al.*, “Female human primordial germ cells display X-chromosome

dosage compensation despite the absence of X-inactivation," *Nat. Cell Biol.*, vol. 22, no. 12, pp. 1436–1446, 2020.

[17] L. Li *et al.*, "Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions," *Cell Stem Cell*, pp. 1–16, 2017.

[18] Á. Vértesy *et al.*, "Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells," *Nat. Commun.*, vol. 9, no. 1, p. 1873, 2018.

[19] Z. Cheng *et al.*, " Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress ," *Mol. Syst. Biol.*, vol. 12, no. 1, p. 855, 2016.

[20] J. Auger, "Spermatogenic Cells—Structure," M. K. B. T.-E. of R. (Second E. Skinner, Ed. Oxford: Academic Press, 2018, pp. 53–60.

[21] A. Amsterdam, K. Kannan, D. Givol, Y. Yoshida, K. Tajima, and A. Dantes, "Apoptosis of Granulosa Cells and Female Infertility in Achondroplastic Mice Expressing Mutant Fibroblast Growth Factor Receptor 3G374R," *Mol. Endocrinol.*, vol. 15, no. 9, pp. 1610–1623, Sep. 2001.

[22] A. Amsterdam, K. Kannan, D. Givol, Y. Yoshida, K. Tajima, and A. Dantes, "Apoptosis of Granulosa Cells and Female Infertility in Achondroplastic Mice Expressing Mutant Fibroblast Growth Factor Receptor 3G374R," *Mol. Endocrinol.*, vol. 15, no. 9, pp. 1610–1623, Sep. 2001.

[23] J. Buratini *et al.*, "Expression of fibroblast growth factor-8 and regulation of cognate receptors, fibroblast growth factor receptor-3c and -4, in bovine antral follicles," *Reproduction*, vol. 130, no. 3, pp. 343–350, 2005.

[24] K. von Kopylow, H. Staege, W. Schulze, H. Will, and C. Kirchhoff, “Fibroblast growth factor receptor 3 is highly expressed in rarely dividing human type A spermatogonia,” *Histochem. Cell Biol.*, vol. 138, no. 5, pp. 759–772, 2012.

[25] A. MAYERHOFER, L. D. RUSSELL, C. GROTHE, M. RUDOLF, and M. GRATZL, “Presence and Localization of a 30-kDa Basic Fibroblast Growth Factor-Like Protein in Rodent Testes*,” *Endocrinology*, vol. 129, no. 2, pp. 921–924, Aug. 1991.

[26] K. Steger, F. Tetens, J. Seitz, C. Grothe, and M. Bergmann, “Localization of fibroblast growth factor 2 (FGF-2) protein and the receptors FGFR 1–4 in normal human seminiferous epithelium,” *Histochem. Cell Biol.*, vol. 110, no. 1, pp. 57–62, 1998.

[27] H. Kubota, M. R. Avarbock, and R. L. Brinster, “Growth factors essential for self-renewal and expansion of mouse spermatogonial stem cells,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 47, pp. 16489 LP – 16494, Nov. 2004.

[28] J. Guo *et al.*, “Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development,” *Cell Stem Cell*, vol. 21, no. 4, pp. 533-546.e6, 2017.

[29] A. Goriely and A. O. M. Wilkie, “Paternal age effect mutations and selfish spermatogonial selection: Causes and consequences for human disease,” *Am. J. Hum. Genet.*, vol. 90, no. 2, pp. 175–200, 2012.

[30] A. Goriely *et al.*, “Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors,” *Nat. Genet.*, vol. 41, no. 11, pp. 1247–1252, 2009.

[31] F. A. Wolf, P. Angerer, and F. J. Theis, “SCANPY: large-scale single-cell gene expression data analysis,” *Genome Biol.*, vol. 19, no. 1, p. 15, 2018.

[32] X. Huang and Y. Huang, “Cellsnp-lite: an efficient tool for genotyping single cells,” *Bioinformatics*, vol. 37, no. 23, pp. 4569–4571, Dec. 2021.

[33] Y. Huang, D. J. McCarthy, and O. Stegle, “Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference,” *Genome Biol.*, vol. 20, no. 1, p. 273, 2019.

CHAPTER 5

Concluding remarks

In our studies, we demonstrate that X chromosome dampening is a type of X chromosome dosage compensation taking place during human preimplantation development and germline formation. Unlike mice, human germline specification takes days compared to hours in mice and hPGC development is a much lengthier process, it is likely that X-chromosome dosage compensation in germ cells between the two species are different. These differences include expression of lncRNA *XIST* from both active X chromosomes in human epiblast and hPGCs. Additionally, primate specific lncRNA *XACT* is expressed in human preimplantation embryos, that is only detected in high order primates such as gorillas, chimpanzees and homo sapiens[1]. By imaging techniques, we have shown that *XACT* is not only expressed in human pre-implantation embryos but also in human embryo attachment culture, hPGCLC differentiation *in vitro*, and hPGCs *in vivo*. Presence of *XACT* in these cell types might be due to the activity of an enhancer that threads *XACT* into the pluripotency network [1]. Interestingly, we describe *XACT* as a unique marker of hPGCs, and speculate that it could be used to trace hPGCs from the time of lineage specification by RNA FISH. Additional RNA FISH analysis of the X-linked genes *ATRX* and *HUWE1,* together with absence of H3K27me3 accumulation in the nucleus of most female hPGCs demonstrates that female hPGCs harbor two Xa's from at least week 4 pf onwards.

Although by microscopy we showed the presence of two active X-chromosomes in female hPGCs, the scRNA-seq data revealed that the female hPGCs reduce X-linked gene expression to similar or lower levels as somatic cells that harbor one active and one inactive X chromosomes, indicating that X-linked dosage compensation in female hPGCs

is regulated by the XCD mechanism, similar to female naïve human pluripotent stem cells[2] and female human pre-implantation embryos[3]. While in pre-implantation embryos XCD is a transient state, in case of hPGCs, we showed that XCD is not a transitional state into XCI, rather a stable state lasting at least 9 weeks until the point of meiotic initiation. Intriguingly, the loss of XCD in the female germline upon meiotic entry is linked to the silencing of *XIST*, suggesting that *XIST* could be mediating XCD. In contrast, the further decline in the X/A ratio in primordial oocytes occurs in the absence of *XIST* expression. It remains unclear whether this regulation is achieved by XCD or other mechanisms.

X-chromosome dosage regulation might be extremely important for Turner (XO) and Klinefelter (XXY) syndrome patients, who suffer from infertility associated with loss of germline cells[4], [5]. Although germline development in fetuses diagnosed with Turner syndrome is morphologically normal, oocyte loss occurs within the first few months after birth[6]. Potentially, meiosis is not occurring correctly in Turner syndrome patients due to diminished levels of critical X-linked genes in differentiating XO hPGCs compared to XX hPGCs. Thus, upregulation of X-linked gene expression with entrance into meiosis may be necessary for the formation of mature oocytes. However, how lncRNAs such as *XIST* and *XACT* contribute to X-chromosome gene regulation in the developing human embryo, will need to be studied mechanistically.

To address this question we have knocked out *XIST* in naïve hESCs that model X chromosome dampening [2]. We uncovered that *XIST* is required for XCD and

unexpectedly we also discovered that *XIST* regulates specific autosomal genes during preimplantation development, leading to differential gene expression between male and female cells.

Our findings reveal that *XIST* can be engaged in two different types of gene expression regulation during human early embryonic development: first dampening and later silencing. Interestingly, mouse pluripotent stem cells do not express *Xist*, however the ectopic induction of *Xist* in these cells enables XCI and not XCD, suggesting that it is not simply the developmental state that determines the gene regulatory output of *XIST*. Therefore, future studies to dissect functional role of the differences in *XIST* transcripts between mouse and human could help us understand how *XIST* can be engaged in two different types of X chromosome regulation during human embryonic development [7], [8].

Mechanistically, we showed that *XIST* executes XCD and XCI based on similar principles. Specifically, on both the Xd and Xi, the same isoform of the RNA is expressed, *XIST* localizes over the entire chromosome, recruits similar protein effectors (CIZ1 and SPEN) and requires SPEN for transcriptional repression. These findings support our conclusion that XCD and XCI are regulated by *XIST* and suggest that a large number of effector proteins interacting with the RNA are conserved between XCD and XCI.

Interestingly, *XIST* binds to chromatin at different places during XCI and XCD. (i) *XIST* enriches in different regions on the inactive and dampened X chromosomes that could be

explained by different spatial organization of X chromosome in those two states. This conclusion is consistent with prior observations that Xist exploits the spatial organization of the X chromosome for the initial spread across the X chromosome during XCI initiation [9]. (ii) We observed reduced enrichment of *XIST* on the Xd compared to the Xi. This model suggests that the concentration of *XIST* within the X-chromosome territory is the major determinant of *XIST*-mediated gene repression by reduced recruitment of the effector proteins necessary for silencing genes. Alternatively, absence or reduced levels of proteins (PTBP1, MATR3, CIZ1, CELF1 etc.) involved in regulating mouse *Xist* localization on X[10], [11] , could affect spreading of XIST outside X chromosome territory.

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

Unexpectedly, our genomic approach uncovered specific autosomal regions to which *XIST* localizes in cells with XCD. This finding was particularly surprising since mouse *Xist* has always been reported to be confined to the X chromosome from which it is transcribed [12]. Although multiple studies have reported that the ectopic expression of *Xist/XIST* from an autosome is sufficient to induce autosomal gene silencing, to our knowledge, autosomal gene regulation by endogenously expressed *XIST* has never been reported. By comparing gene expression levels between females and male naïve hPSCs and

human pre-implantation embryos, we discovered that 87 autosomal genes are more lowly expressed in female cells compared to males and also regulated by *XIST*. These results support the possibility that *XIST* association with chromatin beyond the X-chromosome territory results in a reduced local concentration on the Xd in naïve pluripotent stem cells. It is possible that unbound *XIST* is from the X chromosome associates with chromatin regions on nearby chromosomes. Regardless of the mechanism, the repression of autosomal genes by *XIST* requires SPEN. These observations indicate that autosomal gene repression by *XIST* follows similar principles as on the X chromosome. Further studies are necessary to define the consequences of the female-specific downregulation of autosomal genes. Overall, the small but consistent downregulation of autosomal genes by *XIST* provides an unexpected and exciting new role for *XIST* during early human development.

Overall, our discovery that *XIST* mediates XCD could serve as a foundation for future studies of the role of dosage compensation by XCD in human pre-implantation embryos and the human germ line. Moreover, whether the autosomal gene regulation by *XIST* is critical for early human development and the differentiation of the germ line remains an exciting open question for future studies.

Many biochemical assays such as RNA antisense purification, to identify *XIST* binding site on chromatin require large number of cells. To address these questions specifically during human germline development, would require advancing hPGCs differentiation systems *in vitro*. So far, isolation of hPGCs from fetal tissues using fluorescent activated

cell sorting (FACS) has served as a critical tool to dissect hPGCs *in vivo.* In most cases, combination of different cell surface markers are used as FACS strategies to isolate hPGCs. Therefore, identification of additional surface molecules will help the field to study hPGCs.

In our work, we demonstrated that FGFR3 is expressed by PGCs in prenatal ovaries and testis. Additional characterization of FGFR3 by scRNA-seq and immunofluorescence revealed that female hPGCs maintain FGFR3 expression into the initial stages of meiotic progression and silence it upon formation of ZP3+ primordial oocytes.

We showed that sorting for FGFR3 positive cells, is a viable strategy that will enrich for hPGCs and meiotic germ cells from the embryonic and fetal ovary. In addition to enriching the germ cells, FGFR3+ FACS can also isolate somatic cells that could be due to our wider FACS gates selection, but anti-FGFR3 antibodies could be a powerful tool in combination with other markers such as cKIT [13] to isolate PGCs and meiotic germ cells from prenatal ovaries. In summary, identification of FGFR3 as a surface marker for female germ cell isolation is could potentially serve as a diagnostic surface marker for further improvements of PGCLC differentiation protocols.

Taken together, we have strengthened the notion that X chromosome dampening is not a short, transitory phase towards X chromosome inactivation rather, in human development, XCD is a different type of dosage compensation that is mediated by lncRNA XIST. Our discoveries in X chromosome regulation during germline development extends

our knowledge for future studies aiming towards *in vitro* gametogenesis and safe cell therapies.

[1] M. Casanova *et al.*, “A primate-specific retroviral enhancer wires the XACT lncRNA into the core pluripotency network in humans,” *Nat. Commun.*, vol. 10, no. 1, p. 5652, 2019.

[2] A. Sahakyan *et al.*, “Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation,” *Cell Stem Cell*, vol. 20, no. 1, pp. 87–101, 2016.

[3] S. Petropoulos *et al.*, “Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos,” *Cell*, Apr. 2016.

[4] L. J. Folsom and J. S. Fuqua, “Reproductive Issues in Women with Turner Syndrome,” *Endocrinol. Metab. Clin. North Am.*, vol. 44, no. 4, pp. 723–737, Dec. 2015.

[5] S. Franik *et al.*, “Klinefelter syndrome and fertility—Impact of X-chromosomal inheritance on spermatogenesis,” *Andrologia*, vol. 50, no. 5, pp. 1–5, 2018.

[6] K. Reynaud, R. Cortvrindt, F. Verlinde, J. De Schepper, C. Bourgain, and J. Smitz, “Number of ovarian follicles in human fetuses with the 45,X karyotype,” *Fertil. Steril.*, vol. 81, no. 4, pp. 1112–1119, 2004.

[7] N. Brockdorff *et al.*, “Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome,” *Nature*, vol. 351, no. 6324, pp. 329–331, 1991.

[8] T. B. Nesterova *et al.*, “Characterization of the genomic Xist locus in rodents

reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence," *Genome Res.*, vol. 11, no. 5, pp. 833–849, May 2001.

[9] J. M. Engreitz *et al.*, "The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome," *Science (80-.).*, vol. 341, no. 6147, pp. 1237973–1237973, 2013.

[10] C. A. McHugh *et al.*, "The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3," *Nature*, vol. 521, no. 7551, pp. 232–236, Apr. 2015.

[11] C. Chu *et al.*, "Systematic discovery of Xist RNA binding proteins," *Cell*, vol. 161, no. 2, pp. 404–416, 2015.

[12] N. Brockdorff, "Localized accumulation of Xist RNA in X chromosome inactivation," *Open Biol.*, vol. 9, no. 12, p. 190213, Jun. 2020.

[13] S. Gkountela *et al.*, "The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation," *Nat. Cell Biol.*, vol. 15, no. 1, pp. 113–122, 2012.