X-chromosome upregulation is dynamically linked to the X-inactivation state

Antonio Lentini[1], Christos Coucoravas[1], Nathanael Andrews[2], Martin Enge[2], Qiaolin Deng[3] and Björn Reinius[1]*

[1]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.
[2]Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden.
[3]Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden.
*Correspondence to: bjorn.reinius@ki.se

One sentence summary: Expression of mammalian X chromosomes is dynamically balanced between both alleles to compensate dosage imbalances.

Keywords: X-chromosome upregulation, X-inactivation, embryo development, dosage compensation, Ohno's hypothesis.

Abstract

Mammalian X-chromosome dosage balance is regulated by X-chromosome inactivation (XCI) and X-chromosome upregulation (XCU), but the dynamics of XCU as well as the interplay between the two mechanisms remain poorly understood. Here, we mapped XCU throughout early mouse embryonic development at cellular and allelic resolution, revealing sex- and lineage-specific dynamics along key events in X-chromosome regulation. Our data show that XCU is linearly proportional to the degree of XCI, indicating that dosage compensation ensues based on mRNA levels rather than number of active X chromosomes. In line with this, we reveal that the two active X chromosomes in female naïve embryonic stem cells are not hyperactive as previously thought. In all lineages, XCU was underlain by increased transcriptional burst frequencies, providing a mechanistic basis *in vivo*. Together, our results demonstrate unappreciated flexibility of XCU in balancing X-chromosome expression, and we propose a general model for allelic dosage balance, applicable for wider mechanisms of transcriptional regulation.

Main

In therian mammals, the X chromosome is present as two copies in females and one in males. Correct balance of gene dosage is vital for homeostasis and normal cell function and two major, X-chromosome-wide, mechanisms ensure balanced genomic expression. XCI silences one X allele in females, equalizing expression between the sexes, and XCU increases X-linked expression to resolve dosage imbalance against autosomal gene networks [reviewed in ref. (*1*)]. Although the existence of XCU is well supported by relative measurements between X and autosomes (*2–6*), the dynamics of its establishment and maintenance remain elusive (*6–8*). Such unraveling would require quantitative gene expression measurements at cellular and allelic resolution, only recently made possible by allele-sensitive single-cell RNA-seq (scRNA-seq), and analytical disentanglement of other allelic processes at cellular level such as stochastic allelic expression and random XCI (rXCI) (*9*). Here, we have mapped the XCU process in mouse at allelic and cellular resolution throughout its establishment *in vitro* and *in vivo*, revealing fundamental features and a surprising plasticity in this process, and we provide a unified model of this chromosome-wide mode of gene regulation.

Results

Establishment of X-upregulation during mESC differentiation

We started by culturing naïve male and female mouse embryonic stem cells (mESCs) derived from F1 hybrids of strains C57BL/6J and CAST/EiJ (hereafter C57 and CAST, respectively), and subsequently differentiated female mESCs into epiblast stem cells (EpiSCs) to induce rXCI (**Fig. 1a, Methods**). We harvested naïve mESCs and EpiSCs throughout differentiation (day 1, 2, 4 and 7) and subjected cells to scRNA-seq using the recently published Smart-seq3 method (*10*) (**Fig. 1a, Methods**), allowing sensitive allele- and molecule-level analysis at cellular resolution (**Supplementary Fig. 1a**). Differentiation induced distinct expression changes accompanied by loss of pluripotency factors (*e.g. Sox2* & *Nanog*) and induction of lineage-specific factors (*e.g. Fgf5* & *Krt18*) (**Fig. 1b-c**, **Supplementary Fig. 1b and Supplementary Table 1**) as well as enrichment for associated pathways (**Supplementary Fig. 1c and Supplementary Table 1**). *Xist* induction was detected already by differentiation day 1 (**Supplementary Fig. 1d**) and accompanied by the transition from biallelic to monoallelic X-chromosome expression (**Fig. 1d**), signifying successful rXCI. To investigate XCU, we stratified female cells based on rXCI status into active (XaXa) and inactive (XaXi) states inferred from allelic expression (**Fig. 1d, Methods**) and calculated X-chromosome expression per cell and allele. Intriguingly, we detected chromosome-wide upregulation of the active X allele in female cells with established rXCI and in male mESCs, but lack thereof

in female cells of XaXa state (**Fig. 1e-f**). We calculated total expression of the different X states revealing that the combined output of two moderately active XaXa alleles was higher than that of a single hyperactive allele ($FDR_{TPM} \leq 4.35 \times 10^{-34}$, $FDR_{UMI} \leq 6.86 \times 10^{-5}$, FDR corrected Mann-Whitney *U* test; **Fig. 1g-h and Supplementary Fig. 1e**). Furthermore, gene-wise fold-changes of the active allele transitioning from XaXa into XaXi state (**Fig. 1i**, shown for both C57 and CAST alleles) demonstrated distinct XCU, with fold-changes below two (median $X_{C57} = 1.62$, $P_{C57} = 1.41 \times 10^{-40}$ and median $X_{CAST} = 1.67$, $P_{CAST} = 5.72 \times 10^{-48}$, Wilcoxon signed-rank test). These findings are important as they alter the notion that XCI acts on an hyperactive XaXa state (*3*, *5*, *11*). It furthermore directly explains why measured female-to-male X-chromosome expression fold-change in naïve ESC tend to be less than two-fold (*5*, *11*), and is consistent with the notion that expression dosage of two active X alleles blocks or delays differentiation of mESCs (*12*).

Previous studies have linked XCU to increased transcriptional initiation (*3*, *13*, *14*) and bursting (*15*), suggesting that XCU is driven by transcriptional events on the active allele. We inferred parameters of transcriptional burst kinetics at the resolution of single alleles using both molecule- (UMI) and read-count (TPM) data (**Methods**), revealing moderate and balanced burst frequencies for the two XaXa alleles whereas all states of a singly active X allele displayed markedly increased burst frequency ($FDR_{TPM} \leq 3.97 \times 10^{-10}$, $FDR_{UMI} \leq 5.68 \times 10^{-03}$, FDR corrected Mann-Whitney *U* test; **Fig. 1j**). Burst sizes were not increased but remained largely constant (FDR ≥ 0.08, **Fig. 1k**), consistent with our previous findings that XCU is driven by increased transcriptional burst frequencies (*15*).

Two waves of X-upregulation during early mouse development

Naïve mESCs are derived from, and mimic, the inner cell mass (ICM) of preimplantation stage embryos. To obtain an allelic map of XCU dynamics throughout the *in vivo* pre- and peri-implantation development, we leveraged the large allele-level scRNA-seq datasets we recently generated for the C57BL/6J *x* CAST/EiJ F1 breed across early murine embryogenesis (*16–18*), the developmental window encompassing the main events of XCI (*1*); imprinted XCI (iXCI), X-chromosome reactivation (XCR) and rXCI (**Fig. 2a**). We calculated X-chromosome expression at cell and allelic resolution and observed initiation of XCU at the 4-cell stage in males, *i.e.* around completion of zygotic genome activation (ZGA) (**Fig. 2b**), whereas female cells displayed balanced allelic expression until the 8-to-16-cell stage (**Fig. 2b**) coinciding with iXCI. This implies that XCU is initiated in concert with unbalanced chromosomal dosage in a sex-specific manner. We observed maintained XCU throughout preimplantation blastocyst development in both sexes (**Fig. 2b**). To investigate XCU in the emerging lineages in this window of development, we first performed graph-based clustering of blastocyst cells into ICM, trophectoderm (TE) and two immature clusters, followed by hierarchical clustering of ICM into epiblast (EPI) and primitive endoderm (PrE) based on known lineage markers (**Supplementary Fig. 2a-b, Methods**). This showed XCU to be present in all lineages of the late blastocyst, including EPI cells prior to XCR (**Supplementary Fig. 2c**), again suggesting that XCU is regulated based on active X alleles. Indeed, during post-implantation development (E5.5-6.5), male cells (independent of lineage) and female cells of extraembryonic lineages (visceral endoderm, VE; extraembryonic ectoderm, ExE) retained XCU along with iXCI across all timepoints (**Fig. 2c**). As expected, female E5.5 EPI cells, residing in reactivated XaXa state, showed X-linked expression close to autosomal levels for each allele (**Fig. 2c**), importantly demonstrating erasure of XCU *in vivo*. This was followed by a second wave of XCU in cells where rXCI was either underway or completed (E6.0–6.5) (**Fig. 2c**). As rXCI is an asynchronous process (*16*, *17*), we stratified and analyzed EPI cells based on rXCI status which verified that cells of XaXa state expressed each X allele at moderate levels, similar to autosomes (**Fig. 2d**), consistent with our *in vitro* findings (**Fig. 1**). Intriguingly, XCU displayed gradual chromosome-wide re-establishment along with rXCI establishment (**Fig. 2d-e**), showing that XCU is not a trigger-like mechanism but a dynamic 'tuning-like' process. To control for effects related to differentiation, we inferred pseudotime differentiation trajectories for each lineage (**Supplementary Fig. 2d-e**, **Methods**), demonstrating minimal association with X-linked expression compared to allele usage (**Supplementary Fig. 2f**).

A linear relation of X-chromosome upregulation with X-inactivation

If allelic X-chromosome expression dosage is compensated based on output rather than the number of active chromosome copies, such a mechanism would ensure similar mRNA levels regardless of degree of XCI. Indeed, allelic X-linked expression was highly correlated with maternal fractions in EPI cells (adjusted

$R^2_{maternal} = 0.80$, $P_{maternal} = 2.92 \times 10^{-69}$, adjusted $R^2_{paternal} = 0.91$, $P_{paternal} = 1.19 \times 10^{-103}$, linear regression; **Fig. 3a and Supplementary Fig. 3a**). Surprisingly, genes escaping XCI showed similar trends to other X-linked genes ($P_{maternal} = 0.62$, $P_{paternal} = 0.71$, Kolmogorov–Smirnov test; **Supplementary Fig. 3a**) which would be unlikely if XCU is independently regulated per gene and allele. Furthermore, variability of iXCI completeness in extraembryonic lineages and pre-implantation stages also followed the trend projected from EPI cells (**Fig. 3a and Supplementary Fig. 3b**), suggesting that the first, iXCI-associated, wave of XCU is achieved by the same mechanism.

Notably, previous studies lacking allelic resolution approximated XCU using X:autosome (X:A) expression ratios (*3–6*) for which the tuning-like compensation (XaXi > XaXs > XaXa, considering expression of one active allele in each state) would be undetectable. Indeed, X:A ratios were indistinguishable in non-XaXa cells, regardless of sex, lineage and rXCI status (median EPI = 0.85–0.86 (XaXa = 0.98); VE = 0.81–0.93; ExE = 0.86–0.90; **Fig. 3c and Supplementary Fig. 3c-d, Methods**). Interestingly, ancestral X-Y gene pairs were not upregulated in males whereas corresponding X-linked gametologs were subject to XCU in females (**Supplementary Fig. 3e-g and Supplementary Table 2**).

Due to caveats with indirect comparisons across chromosomes (*i.e.* X:A ratios), we sought to directly assess the dosage balancing dynamics of XCU. As male cells persistently exhibited XCU past ZGA, we compared male and female Xa alleles along rXCI which indeed revealed that expression of the female Xa allele gradually increased towards male X levels as rXCI progressed (**Fig. 3c & Supplementary Fig. 3h**). We extended this comparison by calculating gene-wise female:male expression ratios at allelic- (Xa) and global level (Xm+Xp; sum of maternal and paternal allele) per lineage, embryonic day and rXCI state (**Methods**). This showed that global gene-level expression was balanced in all except XaXa cells whereas female Xa alleles showed progressive upregulation along rXCI (**Fig. 3d**), consistent with balanced mRNA levels between both alleles. The increased relative expression in XaXa cells is in agreement with our *in vitro* results (**Fig. 1h-i**) and was further supported by reanalysis of bulk RNA-seq data (**Methods**) where mESCs of male sex (XY) and female cells lacking one X-chromosome copy (XO) displayed lower X-linked expression compared to XX cells (**Fig. 3e**), confirming our *in vitro* observation that XCU of one allele does not fully reach the combined levels of both alleles in cells of XaXa state.

If XCU tunes transcription based on mRNA levels, we expect transcription from the Xa allele to be merely slightly increased in cells with partially established XCI due to residual mRNAs from the other allele. We inferred transcriptional kinetics of XCU *in vivo* and indeed found that burst frequencies on the Xa allele were progressively increased during rXCI establishment (**Fig. 3f**) whereas burst sizes remained similar to autosomal levels upon XCU (**Supplementary Fig. 3i**). We furthermore inferred transcriptional kinetics in extraembryonic cells subject to iXCI, also revealing increased burst frequencies for the Xa allele (**Fig. 3f**), again pointing to a unified XCU mechanism in all lineages.

Discussion

By allele-level mRNA measurements in single cells throughout embryonic stem cell differentiation *in vitro* and early development *in vivo*, we have provided a unified model of XCU and its interplay with the XCI dynamics, schematically summarized in **Fig. 4**. We show distinct sex-specific initiation of XCU where X:autosome dosage would otherwise be unbalanced; after completion of ZGA in males and together with iXCI in females (**Fig. 2b**). This is in contrast to previous studies suggesting XCU to be established already in the zygote (*6*) or progressively after the 4-cell stage in both sexes (*7*, *8*) using X:A ratios. These discrepancies may be explained by the inability of non-allelic measurements to correctly distinguish XCU in the presence of other allelic processes such as ZGA and XCI, whereas we in this study directly attribute XCU to the active X allele. Furthermore, we reveal that naïve mESCs cells lack XCU whereas previous studies suggested XCU to be present in these cells (*3*, *5*, *19*), which could be explained by the fact that total XaXa expression levels are higher than what is achieved through XCU (**Fig. 1g–i and Fig. 3b, d–e**).

What is the mechanism underlying the XCU dynamics? Our findings are in line with a recently proposed model (*20*) in which expression balance is achieved by local transcription factor concentrations. As XCI is gradually established (*21*), transcription factors would be available and shift to the active X allele, increasing transcriptional burst frequencies (*22*) resulting in increased mRNA levels. This would explain why active histone modifications do not scale proportionally with mRNA levels on the hyperactive X allele (*13*, *14*) and why genes that escape XCI are subject to similar dosage balancing as other X-linked genes (**Supplementary**

Fig. 3e–g). Enhancer elements likely play key roles in mediating XCU as they may not only increase local transcription factor concentrations through DNA contacts (*23*) but also regulate transcriptional burst frequencies (*24*) which is a driver of XCU (*15*).

Our findings that XaXa cells lack XCU (**Fig. 1g-h and Fig. 2c-d**) is in line with observations of balanced X:A dosage in haploid cells (including MII oocytes; **Fig. 2b**) and cells with two active X chromosomes, such as primary oocytes and primordial germ cells (*6*, *25–27*). Two active X chromosomes is considered a hallmark of the naïve female stem cell state and a gold standard in reprogramming studies (*28*). Our model of two moderately active alleles in the XaXa state in contrast to hyperactive female XaXi and male XaY alters the interpretation of X-chromosome expression level measurements for assessment of reprogramming success and naïveness. Furthermore, in light of our current results in mouse, it is plausible that the "dampened" biallelic (XaXa) X-chromosome state reported in female human preimplantation development (*29–31*) and human naïve pluripotent stem cells (*32*) reflect lack or erasure (*33*) of XCU.

Together, our study provides a unified model of XCU and its link to XCI, and evidence for flexible balancing of allelic expression levels to compensate dosage of the X chromosome.

References

1. X. Deng, J. B. Berletch, D. K. Nguyen, C. M. Disteche, X chromosome regulation: Diverse patterns in development, tissues and disease. Nat. Rev. Genet. 15, 367–378 (2014).
2. D. A. Adler, E. I. Rugarli, P. A. Lingenfelter, K. Tsuchiya, D. Poslinski, H. D. Liggitt, V. M. Chapman, R. W. Elliott, A. Ballabio, C. M. Disteche, Evidence of evolutionary up-regulation of the single active X chromosome in mammals based on Clc4 expression levels in Mus spretus and Mus musculus. Proc. Natl. Acad. Sci. U. S. A. 94, 9244–9248 (1997).
3. X. Deng, J. B. Hiatt, D. K. Nguyen, S. Ercan, D. Sturgill, L. W. Hillier, F. Schlesinger, C. A. Davis, V. J. Reinke, T. R. Gingeras, J. Shendure, R. H. Waterston, B. Oliver, J. D. Lieb, C. M. Disteche, Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. Nat. Genet. 43, 1179–1185 (2011).
4. V. Gupta, M. Parisi, D. Sturgill, R. Nuttall, M. Doctolero, O. K. Dudko, J. D. Malley, P. S. Eastman, B. Oliver, Global analysis of X-chromosome dosage compensation. J. Biol. 5 (2006), doi:10.1186/jbiol30.
5. H. Lin, V. Gupta, M. D. Vermilyea, F. Falciani, J. T. Lee, L. P. O'Neill, B. M. Turner, Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. PLoS Biol. 5, 2809–2820 (2007).
6. D. K. Nguyen, C. M. Disteche, Dosage compensation of the active X chromosome in mammals. Nat. Genet. 38, 47–53 (2006).
7. M. Borensztein, L. Syx, K. Ancelin, P. Diabangouaya, C. Picard, T. Liu, J. B. Liang, I. Vassilev, R. Galupa, N. Servant, E. Barillot, A. Surani, C. J. Chen, E. Heard, Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. Nat. Struct. Mol. Biol. 24, 226–233 (2017).
8. F. Wang, J. D. Shin, J. M. Shea, J. Yu, A. Bošković, M. Byron, X. Zhu, A. K. Shalek, A. Regev, J. B. Lawrence, E. M. Torres, L. J. Zhu, O. J. Rando, I. Bach, Regulation of X-linked gene expression during early mouse development by Rlim. eLife. 5 (2016), doi:10.7554/eLife.19127.
9. B. Reinius, R. Sandberg, Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. Nat. Rev. Genet. 16, 653–664 (2015).
10. M. Hagemann-Jensen, C. Ziegenhain, P. Chen, D. Ramsköld, G.-J. Hendriks, A. J. M. Larsson, O. R. Faridani, R. Sandberg, Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat. Biotechnol. 38, 708–714 (2020).
11. H. Lin, J. A. Halsall, P. Antczak, L. P. O'Neill, F. Falciani, B. M. Turner, Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis. Nat. Genet. 43, 1169–1170 (2011).
12. E. G. Schulz, J. Meisig, T. Nakamura, I. Okamoto, A. Sieber, C. Picard, M. Borensztein, M. Saitou, N. Blüthgen, E. Heard, The Two Active X Chromosomes in Female ESCs Block Exit from the Pluripotent State by Modulating the ESC Signaling Network. Cell Stem Cell. 14, 203–216 (2014).
13. X. Deng, J. B. Berletch, W. Ma, D. K. Nguyen, J. B. Hiatt, W. S. Noble, J. Shendure, C. M. Disteche, Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. Dev. Cell. 25, 55–68 (2013).
14. E. Yildirim, R. I. Sadreyev, S. F. Pinter, J. T. Lee, X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. Nat. Struct. Mol. Biol. 19, 56–62 (2012).
15. A. J. M. Larsson, C. Coucoravas, R. Sandberg, B. Reinius, X-chromosome upregulation is driven by increased burst frequency. Nat. Struct. Mol. Biol. 26, 963–969 (2019).
16. G. Chen, J. P. Schell, J. A. Benitez, S. Petropoulos, M. Yilmaz, B. Reinius, Z. Alekseenko, L. Shi, E. Hedlund, F. Lanner, R. Sandberg, Q. Deng, Single-cell analyses of X Chromosome inactivation dynamics and pluripotency during differentiation. Genome Res. 26, 1342–1354 (2016).
17. S. Cheng, Y. Pei, L. He, G. Peng, B. Reinius, P. P. L. Tam, N. Jing, Q. Deng, Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition and X Chromosome Dynamics during Early Mouse Development. Cell Rep. 26, 2593–2607.e3 (2019).
18. Q. Deng, D. Ramsköld, B. Reinius, R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science. 343, 193–196 (2014).
19. M. Wang, F. Lin, K. Xing, L. Liu, Random X-chromosome inactivation dynamics in vivo by single-cell RNA sequencing. BMC Genomics. 18, 9–10 (2017).
20. X. Deng, C. M. Disteche, Rapid transcriptional bursts upregulate the X chromosome. Nat. Struct. Mol. Biol. 26, 851–853 (2019).
21. H. Marks, H. H. D. Kerstens, T. S. Barakat, E. Splinter, R. A. M. Dirks, G. van Mierlo, O. Joshi, S. Y. Wang, T. Babak, C. A. Albers, T. Kalkan, A. Smith, A. Jouneau, W. de Laat, J. Gribnau, H. G. Stunnenberg, Dynamics of gene silencing during X inactivation using allele-specific RNA-seq. Genome Biol. 16, 9–10 (2015).
22. I. Brouwer, T. L. Lenstra, Visualizing transcription: key to understanding gene expression dynamics. Curr. Opin. Chem. Biol. 51, 122–129 (2019).
23. R. Cortini, G. J. Filion, Theoretical principles of transcription factor traffic on folded chromatin. Nat. Commun. 9, 9–10 (2018).

24. A. J. M. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, \AAsa Segerstolpe, C. M. Rivera, B. Ren, R. Sandberg, Genomic encoding of transcriptional burst kinetics. Nature. 565, 251–254 (2019).
25. A. Fukuda, M. Tanino, R. Matoba, A. Umezawa, H. Akutsu, Imbalance between the expression dosages of X-chromosome and autosomal genes in mammalian oocytes. Sci. Rep. 5, 9–10 (2015).
26. X. Li, Z. Hu, X. Yu, C. Zhang, B. Ma, L. He, C. Wei, J. Wu, Dosage compensation in the process of inactivation/reactivation during both germ cell development and early embryogenesis in mouse. Sci. Rep. 7 (2017), doi:10.1038/s41598-017-03829-z.
27. M. N. Sangrithi, H. Royo, S. K. Mahadevaiah, O. Ojarikre, L. Bhaw, A. Sesay, A. H. F. M. Peters, M. Stadler, J. M. A. Turner, Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. Dev. Cell. 40, 289–301.e3 (2017).
28. S. Augui, E. P. Nora, E. Heard, Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. 12, 429–442 (2011).
29. S. Petropoulos, D. Edsgärd, B. Reinius, Q. Deng, S. P. Panula, S. Codeluppi, A. Plaza Reyes, S. Linnarsson, R. Sandberg, F. Lanner, Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell. 165, 1012–1026 (2016).
30. F. Zhou, R. Wang, P. Yuan, Y. Ren, Y. Mao, R. Li, Y. Lian, J. Li, L. Wen, L. Yan, J. Qiao, F. Tang, Reconstituting the transcriptome and DNA methylome landscapes of human implantation. Nature. 572, 660–664 (2019).
31. I. Okamoto, A. P. Otte, C. D. Allis, D. Reinberg, E. Heard, Epigenetic dynamics of imprinted X inactivation during early mouse development. Science. 303, 644–649 (2004).
32. A. Sahakyan, R. Kim, C. Chronis, S. Sabri, G. Bonora, T. W. Theunissen, E. Kuoy, J. Langerman, A. T. Clark, R. Jaenisch, K. Plath, Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. Cell Stem Cell. 20, 87–101 (2017).
33. S. Mandal, D. Chandel, H. Kaur, S. Majumdar, M. Arava, S. Gayen, Single-Cell Analysis Reveals Partial Reactivation of X Chromosome instead of Chromosome-wide Dampening in Naive Human Pluripotent Stem Cells. Stem Cell Rep. 14, 745–754 (2020).

Acknowledgements: This study was made possible by grants to B.R. from the Swedish Research Council (2017-01723), Åke Wiberg's Foundation, and the Ragnar Söderberg Foundation. We thank members of the Reinius lab for discussions and comments on the manuscript.

Contributions: A.L.: Investigation, Methodology, Formal analysis, Conceptualization, Visualization, Data Curation, Writing - Original Draft, Writing - Review & Editing. C.C.: Investigation. N.A.: Investigation. M.E.: Resources. Q.D.: Methodology, Resources. B.R.: Investigation, Methodology, Conceptualization, Supervision, Resources, Project administration, Funding acquisition, Writing - Original Draft, Writing - Review & Editing.

Competing interests: The authors declare no competing interests.

Data and materials availability: Raw and pre-processed Smart-seq3 data generated is publicly available at ArrayExpress under accession E-MTAB-9324. Previously published raw data is available at Gene Expression Omnibus under accessions GSE45719, GSE74155, GSE109071, GSE23943 and GSE90516. Data and code generated during this study are available at github: github.com/reiniuslab/Lentini_XCU_in_vivo.

Materials and Methods

Ethics statement

All animal experimental procedures were performed in accordance with Karolinska Institutet's guidelines and approved by the Swedish Board of Agriculture (permits N225/15, N343/12 and 18729-2019 Jordbruksverket).

Derivation and culturing of cell lines

Male and female cell lines were established as previously described (*16*). In brief, mESCs were derived from E4 blastocysts of F1 embryos (female C57BL/6J × male CAST/EiJ) and adapted to 2i condition by growing them in gelatin-coated flasks in N2B27 medium (50% neurobasal medium [Gibco], 50% DMEM/F12 [Gibco], 2 mM L-glutamine [Gibco], 0.1 mM β-mercaptoethanol, NDiff Neuro-2 supplement [Millipore], B27 serum-free supplement [Gibco]) supplemented with 1,000 units/mL LIF, 3 µM Gsk3 inhibitor CT-99021, 1 µM MEK inhibitor PD0325901, and passaged with accutase [Gibco]. To induce differentiation toward EpiSCs, mESCs grown in serum/LIF were plated on FBS coated tissue culture plates (coated overnight at 37°C) in N2B27 medium supplemented with 8 ng/mL Fgf2 (R&D) and 20 ng/mL Activin A (R&D) at a cell density of 1×10^4 cells/cm^2 and cultured for up to 7 days. Cells were split and re-plated in the same condition after 1, 2, 4 and 7 days of differentiation. At each split an aliquot of cells was collected for single-cell sorting into 96-well plates containing Smart-seq3 lysis buffer.

Single-cell RNA sequencing (Smart-seq3)

scRNA-seq libraries were constructed as previously described (*10*) with slight modification. Briefly, cells were single-cell sorted into 96-well low-bind PCR-plates [Eppendorf] containing 3 µl of lysis buffer (0.5 units/µl RNase inhibitor [Takara], 0.15% Triton X-100 [Sigma], 0.5 mM (each) dNTP [Thermo Scientific], 1 µM oligo-dT primer [5′-biotin-ACGAGCATCAGCAGCATACGAT30VN-3′; IDT], 5% PEG [Sigma]). Sorting was performed using an SH800 [Sony]. Plates were briefly centrifuged immediately after sorting, sealed, and stored at −80°C. For cell lysis and RNA denaturation, plates were incubated at 72°C for 10 min and immediately placed on ice. Next, 5 µl of reverse transcription mix (50 mM Tris-HCl, pH 8.3 [Sigma], 75 mM NaCl [Ambion], 1 mM GTP [Thermo Scientific], 3 mM $MgCl_2$ [Ambion], 10 mM DTT [Thermo Scientific], 1 units/µl RNase inhibitor [Takara], 2 µM of template-switch oligo [5′-biotin-AGAGACAGATTGCGCAATGNNNNNNNNrGrGrG-3′; IDT] and 2 U/µl of Maxima H-minus reverse transcriptase [Thermo Scientific]) was added to each sample. Reverse transcription was carried out at 42°C for 90 min followed by 10 cycles of 50°C for 2 min and 42°C for 2 min and the reaction was terminated at 85°C for 5 min. PCR pre-amplification was performed directly after reverse transcription by adding 17 µl of PCR mix (containing DNA polymerase, forward and reverse primer) bringing the final concentration in the 25 µl reaction to 1x KAPA HiFi ReadyMix [Roche] 0.1 µM forward primer [5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTGCGCAATG-3′; IDT] and 0.1 µM reverse primer [5′-ACGAGCATCAGCAGCATACGA-3′; IDT]). Thermocycling was performed as follows: 3 min at 98°C, 22 cycles of 20 s at 98°C, 30 s at 65°C and 6 min at 72°C, and final elongation at 6 min at 72°C. After PCR pre-amplification, samples were purified with AMpure XP beads [Beckman Coulter] at volume ratio 0.8:1. Library size distributions were monitored using high-sensitivity DNA chips (Agilent Bioanalyzer 2100) and cDNA concentrations were quantified using the Quant-iT PicoGreen dsDNA Assay Kit [Thermo Scientific]. cDNA was subsequently diluted to 100–200 pg/µl.

Tagmentation was performed using in-house produced Tn5 (*34*). 2 ng of cDNA in 5 µl water was mixed with 15 µl tagmentation mix (0.2 µl Tn5, 2 µl 10x TAPS $MgCl_2$ Tagmentation buffer; 5 µl 40% PEG-8000; 7.8 µl water, per reaction) and incubated 8 min at 55°C in a thermal cycler. Tn5 was inactivated and released from the DNA by the addition of 4 µl 0.2% SDS and 5 min incubation at room temperature. Library amplification was performed by adding 5 µl mix of 1 µM of forward and reverse custom-designed Nextera index primers [forward: 5′-CAAGCAGAAGACGGCATACGAGATNNNNNNNNNNNNGTCTCGTGGGCTCGG-3′, reverse: 5′-AATGATACGGCGACCACCGAGATCTACACNNNNNNNNNNNNTCGTCGGCAGCGTCIDT-3′, where N represents the 10-bp index bases; IDT] and 15µl PCR mix (1µl KAPA HiFi DNA polymerase [Roche]; 10µl 5x KAPA HiFi buffer; 1.5 µl 10 mM dNTPs; 3.5µl water, per reaction), and thermal cycling: 3 min 72°C, 30 s 95°C, 13 cycles of 10 s 95°C; 30 s 55°C; 30 s 72°C, followed by final elongation at 5 min 72°C; 4°C. DNA sequencing libraries were purified using 0.8:1 volume of AMPure XP beads [Beckman Coulter]. Libraries were sequenced using a NextSeq 550 and High Output kits [Illumina].

Smart-seq3 data analysis

A hybrid mouse genome index was constructed by N-masking the reference genome (GRCm38_68) for CAST/EiJ SNPs from the Mouse Genomes Project (mgp.v5.merged.snps_all.dbSNP142) (*35*) using SNPsplit (0.3.2) (*36*). Raw Smart-seq3 data was processed using zUMIs (2.8.0) (*37*). Briefly, sample barcodes were filtered and data was aligned with STAR (2.7.2a) (*38*) (options: --clip3pAdapterSeq CTGTCTCTTATACACATCT --quantMode TranscriptomeSAM) and reads were assigned to both intron and exon features (Mus_musculus.GRCm38.97.chr.gtf) using FeatureCounts (*39*). Next, barcodes were collapsed

using 1 hamming distance and gene expression was calculated for both reads and UMIs. Finally, allele-level expression was calculated from the zUMIs output as previously described (https://github.com/sandberg-lab/Smart-seq3/tree/master/allele_level_expression) (*10*). 77 cells were excluded due to low read depth (>3 MADs).

Processing of published C57BL/6J × CAST/EiJ scRNA-seq data

Pre-processed Smart-seq data for MII oocyte – 16-cell stages were obtained (*18*) and two cells were excluded (8cell_8-3 & 16cell_4-2) due to potential sample problems (high paternal X ratio in male and non-ZGA with clustering together with Zygote samples, respectively). Pre-processed blastocyst Smart-seq2 data was obtained (*18*) and 18 late blastocysts were excluded due to low read depth (>3 MADs). Pre-processed Smart-seq2 data for post-implantation embryos was obtained (*16*, *17*).

Genes escaping X inactivation

A list of known mouse escapee genes (*1810030O07Rik*, *2010000l03Rik*, *2010308F09Rik*, *2610029G23Rik*, *5530601H04Rik*, *6720401G13Rik*, *Abcd1*, *Araf*, *Atp6ap2*, *BC022960*, *Bgn*, *Car5b*, *D330035K16Rik*, *D930009K15Rik*, *Ddx3x*, *Eif1ax*, *Eif2s3x*, *Fam50a*, *Flna*, *Ftsj1*, *Fundc1*, *Gdi1*, *Gemin8*, *Gpkow*, *Huwe1*, *Idh3g*, *Igbp1*, *Ikbkg*, *Kdm5c*, *Kdm6a*, *Lamp2*, *Maged1*, *Mbtps2*, *Med14*, *Mid1*, *Mmgt1*, *Mpp1*, *Msl3*, *Ndufb11*, *Nkap*, *Ogt*, *Pbdc1*, *Pdha1*, *Prdx4*, *Rbm3*, *Renbp*, *Sh3bgrl*, *Shroom4*, *Sms*, *Suv39h1*, *Syap1*, *Tbc1d25*, *Timp1*, *Trap1a*, *Uba1*, *Usp9x*, *Utp14a*, *Uxt*, *Xist*, *Yipf6*) was compiled from previous work (*7*, *40*–*42*) and excluded from analyses of positional expression levels and expression ratios (described below).

Expression calculations

Allelic reference ratios were calculated as $\text{Counts}_{\text{C57}}/\text{Counts}_{\text{Total}}$ after exclusion of the top 10% expressed X-linked genes to avoid bias from highly expressed X genes. XCI status was determined as active (Xa/Xa), semi-inactivated (Xa/Xs) or fully inactivated (Xa/Xi) for allelic ratios in the intervals (0.4, 0.6), [0.6, 0.9) and [0.9, ∞) and the inverse, respectively. TPM was calculated for gene *i* in cell *j* as $\text{TPM}_{ij} = (\text{FPKM}_i / \sum_j \text{FPKM}_j) \times 10^6$ and allelic TPM was calculated by scaling TPM by reference ratios per gene and cell. Relative UMIs was calculated like TPM but using UMI counts and an offset of 10^2. A gene was considered expressed in a dataset/lineage if the average TPM expression was > 0. To obtain a robust estimate for whole-chromosome expression, 20% trimmed TPM means was calculated per cell.

Expression ratios

X:Autosomal ratios are well-known to be highly dependent on the thresholds used (*3*, *11*, *14*) due to X-chromosome-specific expression patterns (**fig. S3C**), therefore we tested several different thresholds using XaXa cells (**fig. S3D**). Two different threshold were selected; >10 TPM as the X expression distribution was comparable to autosomal levels in XaXa cells (see **fig. S3C**, used in **Fig. 3B**), and >1 TPM to be comparable to previous studies (used in **fig. S3E & G**) (*3*, *5*, *14*). Chromosome:autosome expression ratios were calculated as TPM relative to median of autosomes after applying expression thresholds and excluding genes that escape XCI. To account for different number of genes between chromosomes, a bootstrapping method was used (*7*). For bootstrapped ratios, random autosomal gene sets of the same size were selected as a background, repeated $n = 10^3$ times.

Female:male ratios were calculated after exclusion of X escapees as TPM relative to gene average per embryonic day and lineage, as well as for the active X allele for allelic data.

Differential expression of scRNA-seq data

For Smart-seq3 data, global count data was size-factor normalized using scater (1.12.2)/scran (1.12.1) (*43*) and genes expressed in >10% of cells were kept. Differential expression was calculated along days of differentiation as a continuous variable using likelihood ratio tests as implemented in MAST (1.10.0) (*44*). Gene set enrichment for mouse GO biological process gene sets obtained from MGI (http://www.informatics.jax.org/downloads/reports/index.html#go; accessed 2020-06-23) was performed on the differential expression model against a bootstrapped (n = 100) control model as implemented in MAST.

Dimensionality reduction, clustering and trajectory inference

For Smart-seq3 data, highly variable genes (HVGs) were identified from size-factor-normalized counts using scater/scran and ordered by biological variance and FDR. Data was visualized for top 1,000 HVGs using diffusion maps (*45*).

For *in vivo* blastocyst data, HVGs were obtained as explained above and dimensionality reduction was performed for top 1,000 HVGs using UMAP (*46*) and cells were Louvain clustered based on top 1,000 HVG ranks using scran.

For *in vivo* post-implantation data, pseudotime trajectories were inferred using Slingshot (1.2.0) (*47*) from normalized counts following Mclust (5.4.5) (*48*) clustering on diffusion map coordinates.

X-Y homology

List of X-Y homologs was obtained from (*49*) and sequence similarity of expressed X-Y homolog transcripts was calculated using the nucleotide BLAST webservice (*50*) using default settings (performed 2020-03-24).

Kinetics inference

Missing allelic data points were set to 0 if the gene was detected on the other allele. Kinetic parameters were calculated per lineage/genotype/XCI status/growth condition (depending on dataset) using txburst (https://github.com/sandberg-lab/txburst) (*24*). Genes not passing filtering steps were excluded and relative burst frequency (k_{on}) and burst size (k_{syn}/k_{off}) was calculated relative to median of expressed autosomal genes (per lineage). Only groups with > 20 cells were kept to increase the reliability of the statistical inference.

Re-analysis of published bulk RNA-seq data

Raw RNA-seq data including both male and female mESCs was obtained (*51*, *52*) and quantified to protein-coding transcripts from GENCODE vM22 (*53*) using pseudoalignment with Salmon (0.14.1) (*54*). Transcript abundance estimates were summarized to gene-level using tximport (1.12.3) (*55*) and differential expression was calculated using likelihood ratio tests in DESeq2 (1.24.0) (*56*). A full model including cellular genotype (XX, XY or XO), cell culture condition (2i or serum) and study accession was tested against a reduced model without the genotype term. Lowly expressed genes (<100 average normalized counts) were excluded from final plots. Serum growth conditions only showed a minor effect on global X expression compared to 2i (not shown), consistent with a low degree of cells exhibiting complete XCI (*16*).

Statistics and data visualization

All statistical tests were performed in R (3.6.1) as two-tailed unless otherwise stated. Heatmaps were visualized using ComplexHeatmap (2.0.0) (*57*) and all other plots were made using ggplot2 (3.2.1) (*58*). Box plots are presented as median, first and third quartiles, and 1.5x inter-quartile range (IQR). For median ± confidence interval plots, bootstrapped 95% confidence intervals (n = 1,000) were calculated using the percentile method (*59*) as implemented in the R boot package (1.3-23).

Supplementary references

34. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
35. T. M. Keane, L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N. A. Furlotte, E. Eskin, C. Nell\aaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T. G. Belgard, P. L. Oliver, R. E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. Van Der Weyden, C. A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I. J. Jackson, A. Czechanski, J. Guerra-Assunçao, L. R. Donahue, L. G. Reinholdt, B. A. Payseur, C. P. Ponting, E. Birney, J. Flint, D. J. Adams, Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*. **477**, 289–294 (2011).
36. F. Krueger, S. R. Andrews, SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Research*. **5**, 1479 (2016).
37. S. Parekh, C. Ziegenhain, B. Vieth, W. Enard, I. Hellmann, zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *GigaScience*. **7** (2018), doi:10.1093/gigascience/giy059.
38. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
39. Y. Liao, G. K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47–e47 (2019).
40. B. Reinius, J. E. Mold, D. Ramsköld, Q. Deng, P. Johnsson, J. Michaëlsson, J. Frisén, R. Sandberg, Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* **48**, 1430–1435 (2016).
41. B. Reinius, C. Shi, L. Hengshuo, K. S. Sandhu, K. J. Radomska, G. D. Rosen, L. Lu, K. Kullander, R. W. Williams, E. Jazin, Female-biased expression of long non-coding RNAs in domains that escape X-inactivation in mouse. *BMC Genomics*. **11** (2010), doi:10.1186/1471-2164-11-614.
42. F. Yang, T. Babak, J. Shendure, C. M. Disteche, Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* **20**, 614–622 (2010).
43. A. T. L. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research*. **5**, 2122 (2016).
44. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 9–10 (2015).
45. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*. **31**, 2989–2998 (2015).
46. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction (2018) (available at http://arxiv.org/abs/1802.03426).
47. K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, S. Dudoit, Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. **19** (2018), doi:10.1186/s12864-018-4772-0.
48. C. Fraley, A. E. Raftery, Bayesian regularization for normal mixture estimation and model-based clustering. *J. Classif.* **24**, 155–181 (2007).

49. Y. Q. S. Soh, J. Alföldi, T. Pyntikova, L. G. Brown, T. Graves, P. J. Minx, R. S. Fulton, C. Kremitzki, N. Koutseva, J. L. Mueller, S. Rozen, J. F. Hughes, E. Owens, J. E. Womack, W. J. Murphy, Q. Cao, P. De Jong, W. C. Warren, R. K. Wilson, H. Skaletsky, D. C. Page, Sequencing the mouse y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*. **159**, 800–813 (2014).
50. M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T. L. Madden, NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5-9 (2008).
51. H. Marks, T. Kalkan, R. Menafra, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A. F. Stewart, A. Smith, H. G. Stunnenberg, The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*. **149**, 590–604 (2012).
52. R. J. Werner, B. M. Schultz, J. M. Huhn, J. Jelinek, J. Madzo, N. Engel, Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biol. Sex Differ.* **8**, 28 (2017).
53. A. Frankish, M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin, L. Martínez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B. M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, B. Paten, A. Reymond, M. L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
54. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. **14**, 417–419 (2017).
55. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. **4**, 1521 (2015).
56. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
57. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. **32**, 2847–2849 (2016).
58. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (2016; http://link.springer.com/10.1007/978-3-319-24277-4%0Ahttp://doi.wiley.com/10.1111/j.1467-985X.2010.00676_9.x%0Ahttp://link.springer.com/10.1007/978-3-319-24277-4), vol. 174.
59. A. C. Davison, D. V. Hinkley, Bootstrap Methods and their Application. *Bootstrap Methods Their Appl.* (1997), doi:10.1017/cbo9780511802843.

Figure 1

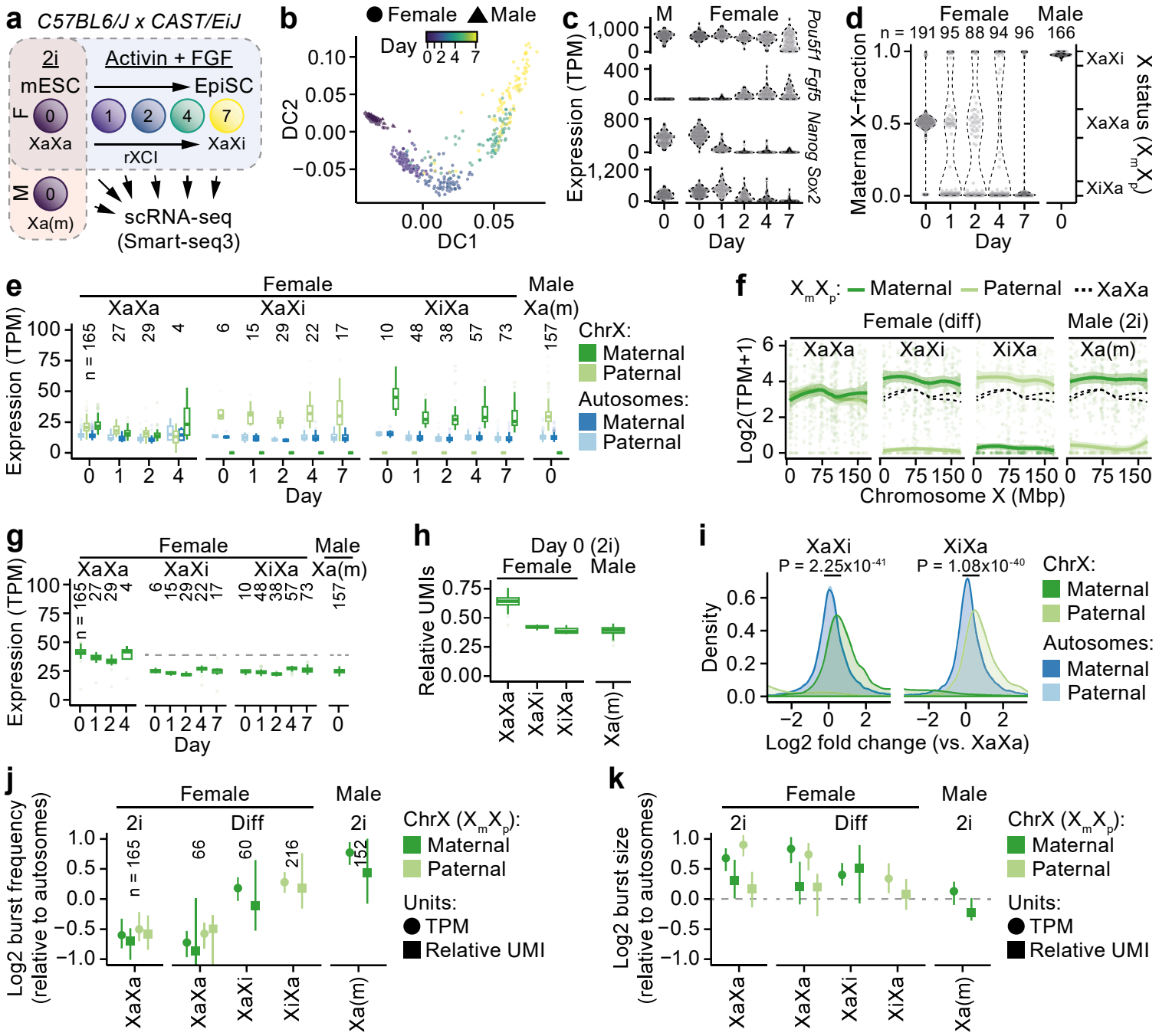


Figure 1. X-upregulation during X-inactivation establishment *in vitro*. a, Schematic overview of experimental setup. **b**, Diffusion map of mESC differentiation for 687 cells sequenced using Smart-seq3. **c-d**, Expression of pluripotency and differentiation marker genes *Pou5f1*, *Fgf5*, *Nanog* and *Sox2* (c) and maternal fraction of X-chromosome expression (d) along differentiation time. **e**, Allelic expression across differentiation days grouped by rXCI status for autosomal (n = 15,683–16,543) and X (n = 508–518) genes. **f**, Smoothed mean (LOESS fit ± 95% confidence interval) of allelic expression across the X chromosome (n = 508–524 genes). **g**, Same as (E) but for global expression for X genes (n = 1,158–1,337). Dashed line indicates XaXa median. **h**, Relative UMI expression for mESCs under 2i conditions for X genes (n = 805). **i**, Density plots of gene-wise fold changes of matching alleles between XaXa and XCI state (XaXi or XiXa; C57 and CAST allele active, respectively) in female differentiating mESCs for autosomal (n = 13,696–14,845) and X (n = 383–433) genes. Mann-Whitney *U* test P-values for active alleles. **j-k**, Relative burst- frequency (j) and size (k) for active X alleles (Xa) grouped by growth condition and rXCI status. Data shown as median ± 95% confidence interval for TPM and relative UMIs for autosomal (n = 10,836–16,714) and X (n = 348–545) genes.

Figure 2

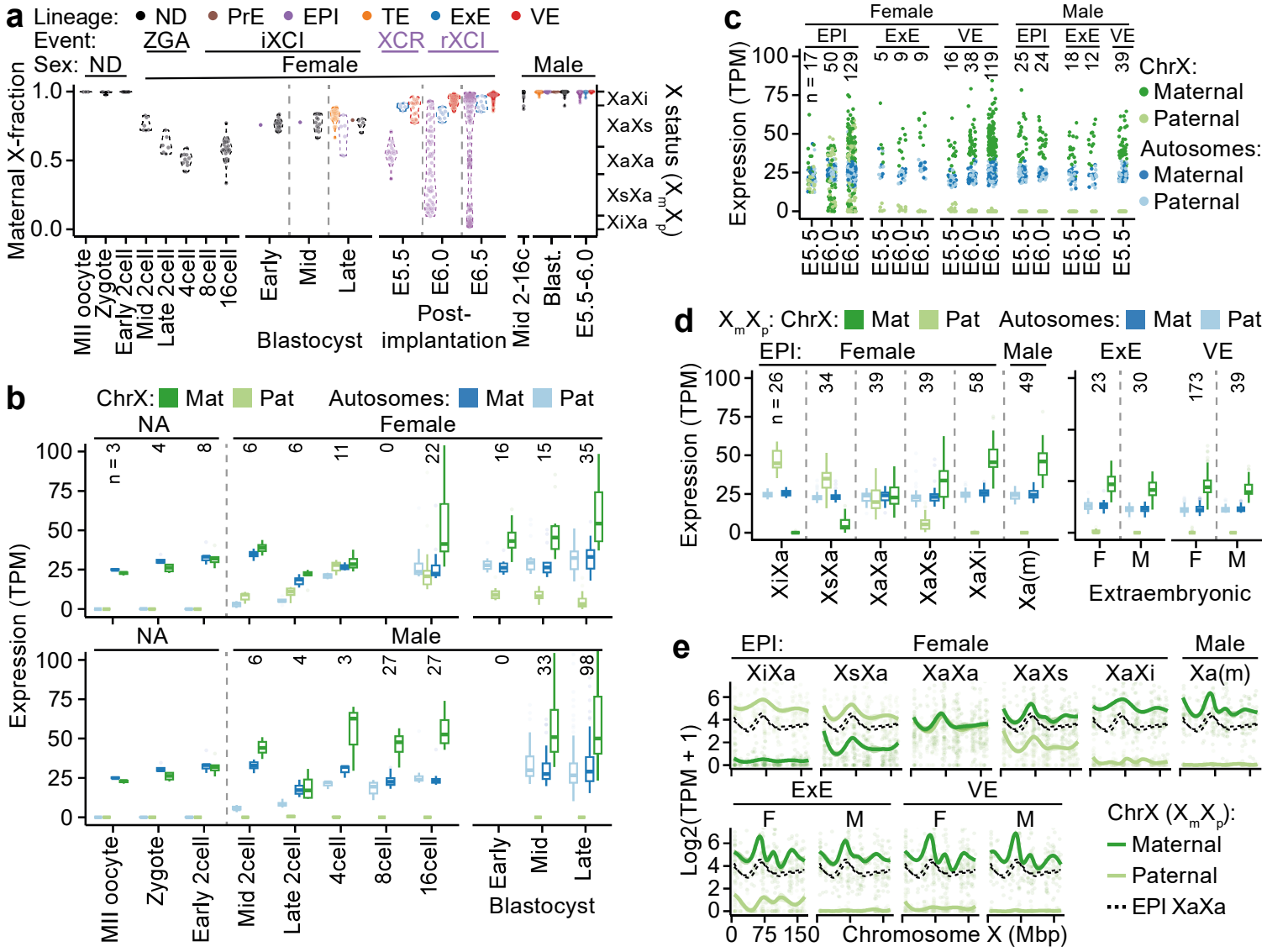


Figure 2. The dynamics of X-upregulation *in vivo*. a, Maternal fraction X-chromosome expression across female mouse early embryo development with key events of X-chromosome regulation labelled above; ZGA (mid-2- to 4-cell stage), iXCI (initiating at 8- to 16-cell stage), XCR (epiblast-specific, late blastocyst to early implantation) and rXCI (epiblast-specific, following XCR). Male data shown to the right. ND = Sex not determined. **b**, Expression levels per allele of pre-implantation development for autosomal (n = 11,164–13,347) and X (n = 323–437) genes. **c-d**, Allelic expression of post-implantation embryos grouped by embryonic day (c) or rXCI status (d) for autosomal (n = 14,941–17,643) and X (n = 597–713). **e**, Smoothed mean (GAM fit ± 95% confidence interval) of allelic expression across the X chromosome (genes n = 543–659), dashed lines represent EPI XaXa expression.

Figure 3

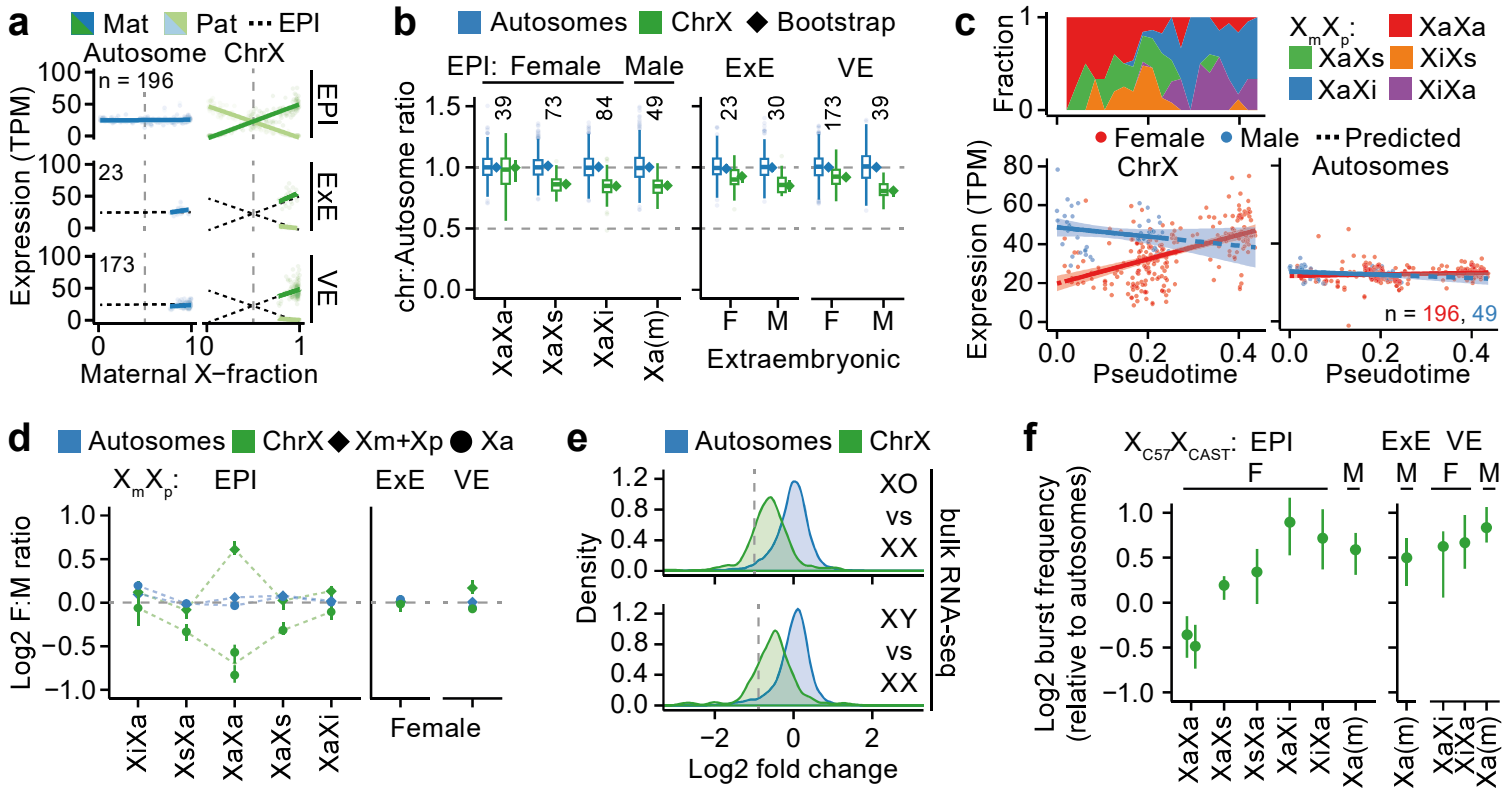


Figure 3. Linear relationship between X-upregulation and X-inactivation. a, Scatter of expression level per allele and maternal X-fraction per female cell and lines indicating linear model mean ± 95% confidence interval, shown for three lineages. Dashed black lines projects EPI and dashed gray lines indicate maternal fraction 0.5. **b**, Chromosomal expression ratios shown as box plots or bootstrapped medians (♦) ± 95% confidence interval. Expression-level threshold 10 TPM. F = female, M = male. **c**, Binned X status (top) and expression of Xa allele shown as linear model mean ± 95% confidence interval for autosomal (n = 17,643) and X (n = 713) genes (bottom) along pseudotime trajectory in EPI cells. Dashed lines represent values predicted from linear model. **d**, Female:male ratios for the active X allele (Xa; •) or global expression (Xm+Xp; ♦) shown as median and ± 95% confidence interval. **e**, Density plot of differential expression in mESC bulk RNA-seq data (XO n = 2, XX n = 12, XY n = 14) for autosomal (n = 10,961) and X (n = 405) genes. Dashed line indicates expected half of autosomal expression. **f**, Burst frequencies of the Xa allele (n = 252–316 genes) relative to autosomal alleles of same strain (n = 9,204–10,577 genes), shown as median and 95% confidence interval. **a-b, d**, Autosomal (n = 14,941–17,643) and X (n = 597–713) genes.

Figure 4

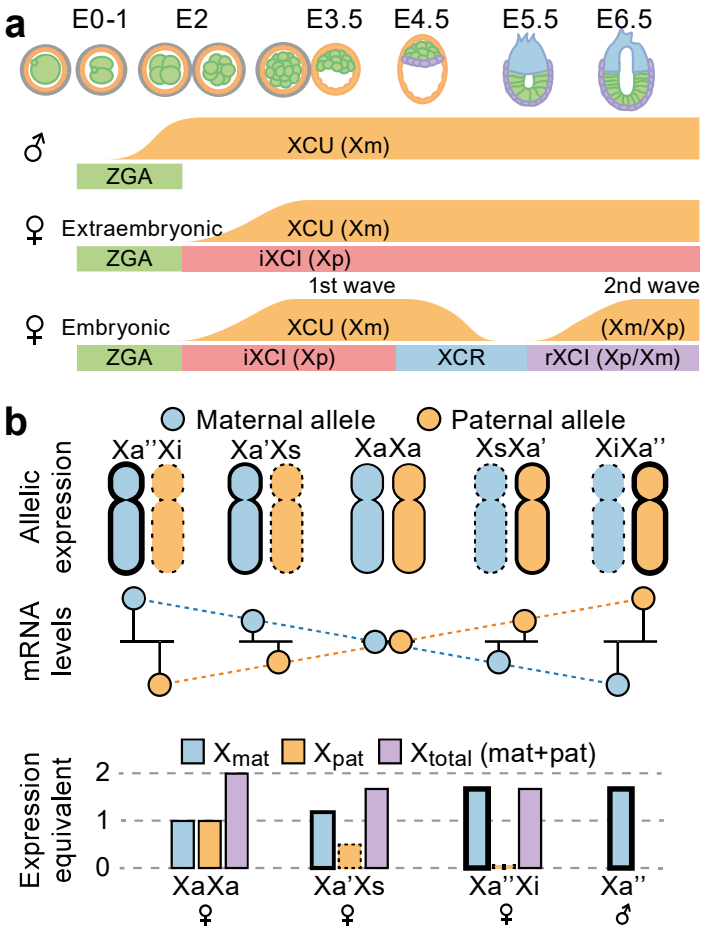
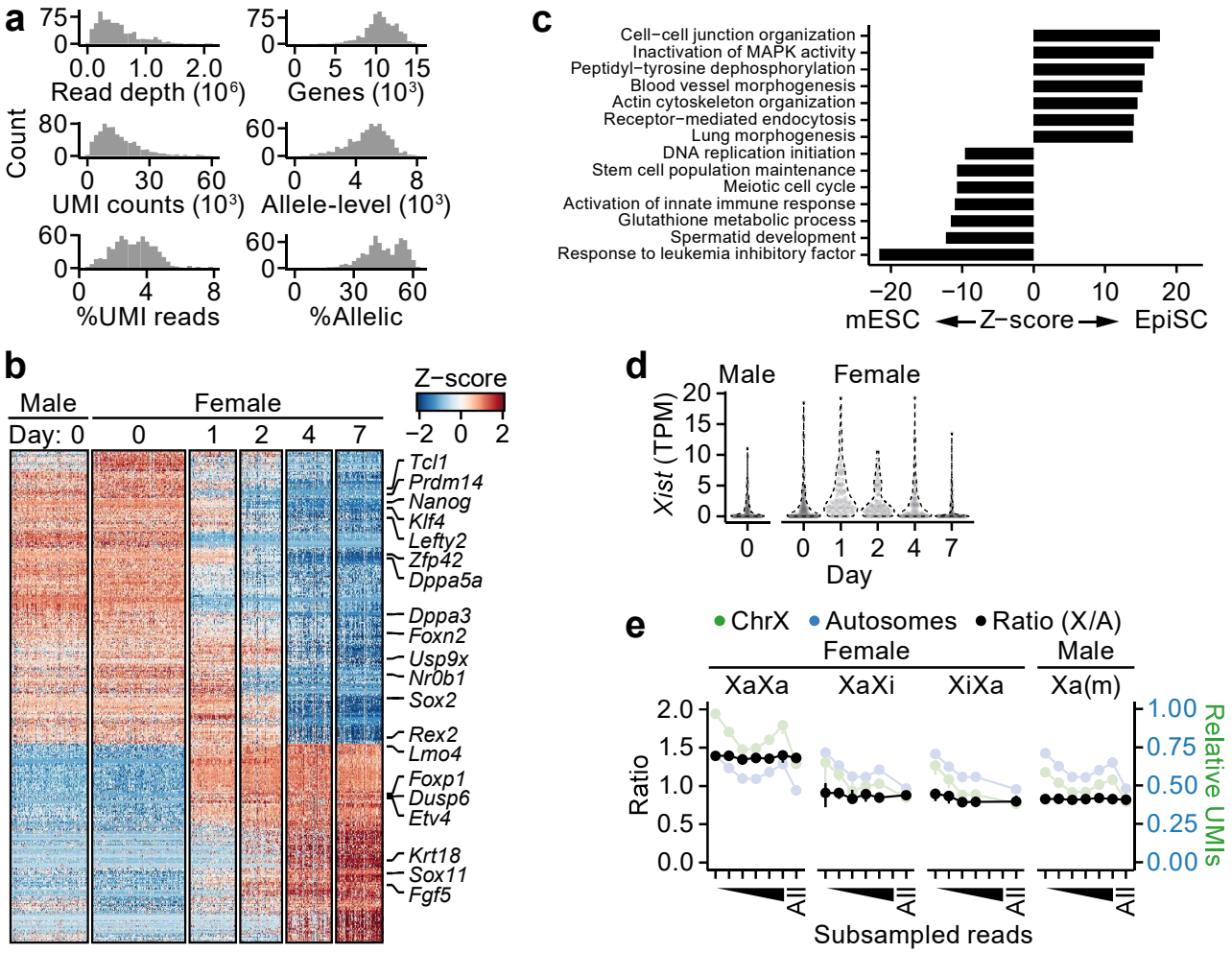


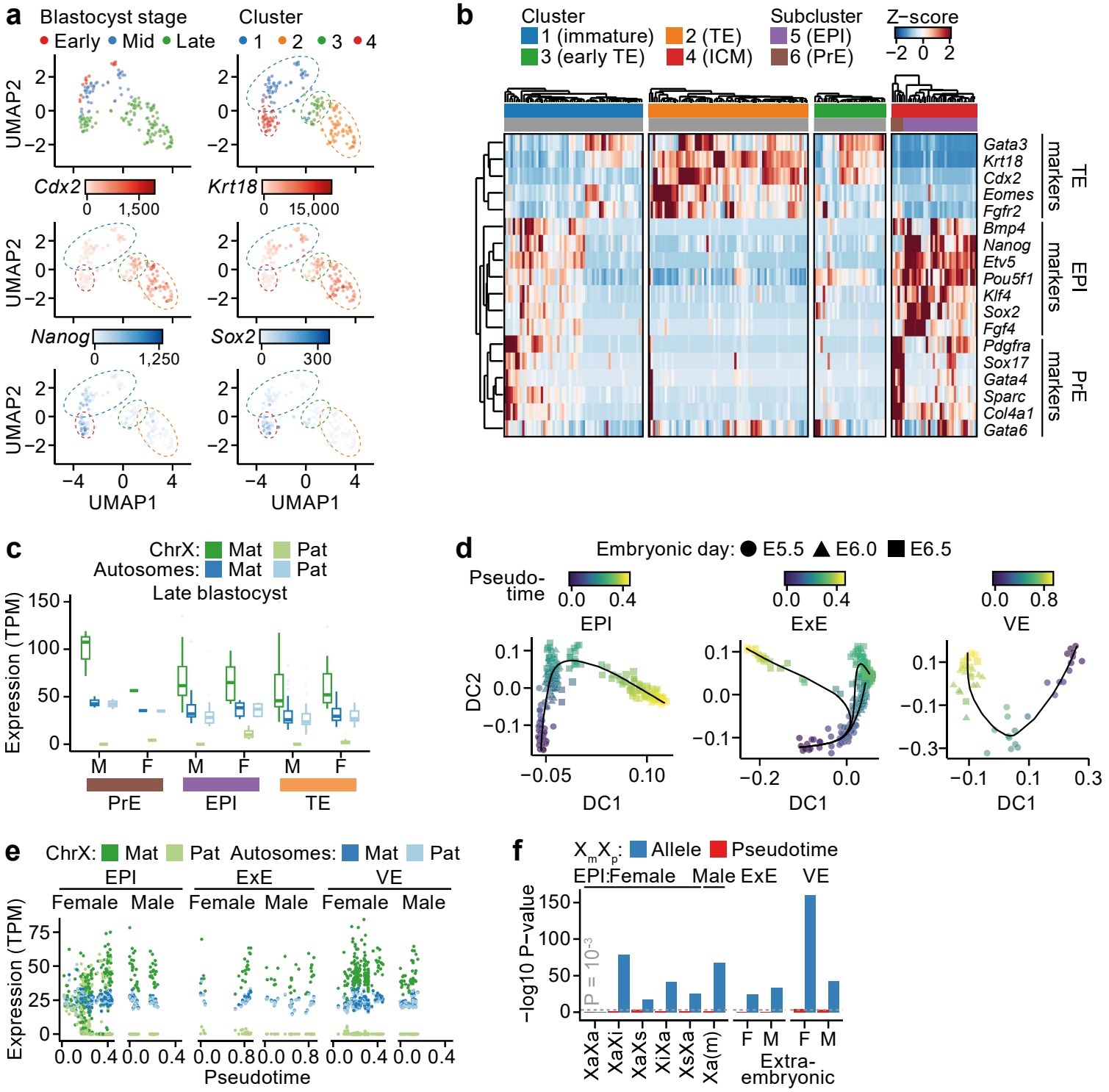
Figure 4. Model of X-upregulation. a, Schematic of sex- and lineage-specific dynamics of X-upregulation (XCU) across mouse embryonic development. **b**, Proposed model of XCU as a flexible mechanism of allelic dosage balancing. As one allele is gradually silenced, the concentration of available transcription factors is shifted towards the active allele (Xa), increasing transcriptional burst frequencies and mRNA levels in a linear fashion (top). However, the single hyperactive allele (Xa'') does not fully reach the combined expression level of two moderately active alleles in XaXa state (bottom).

Supplementary Figure 1



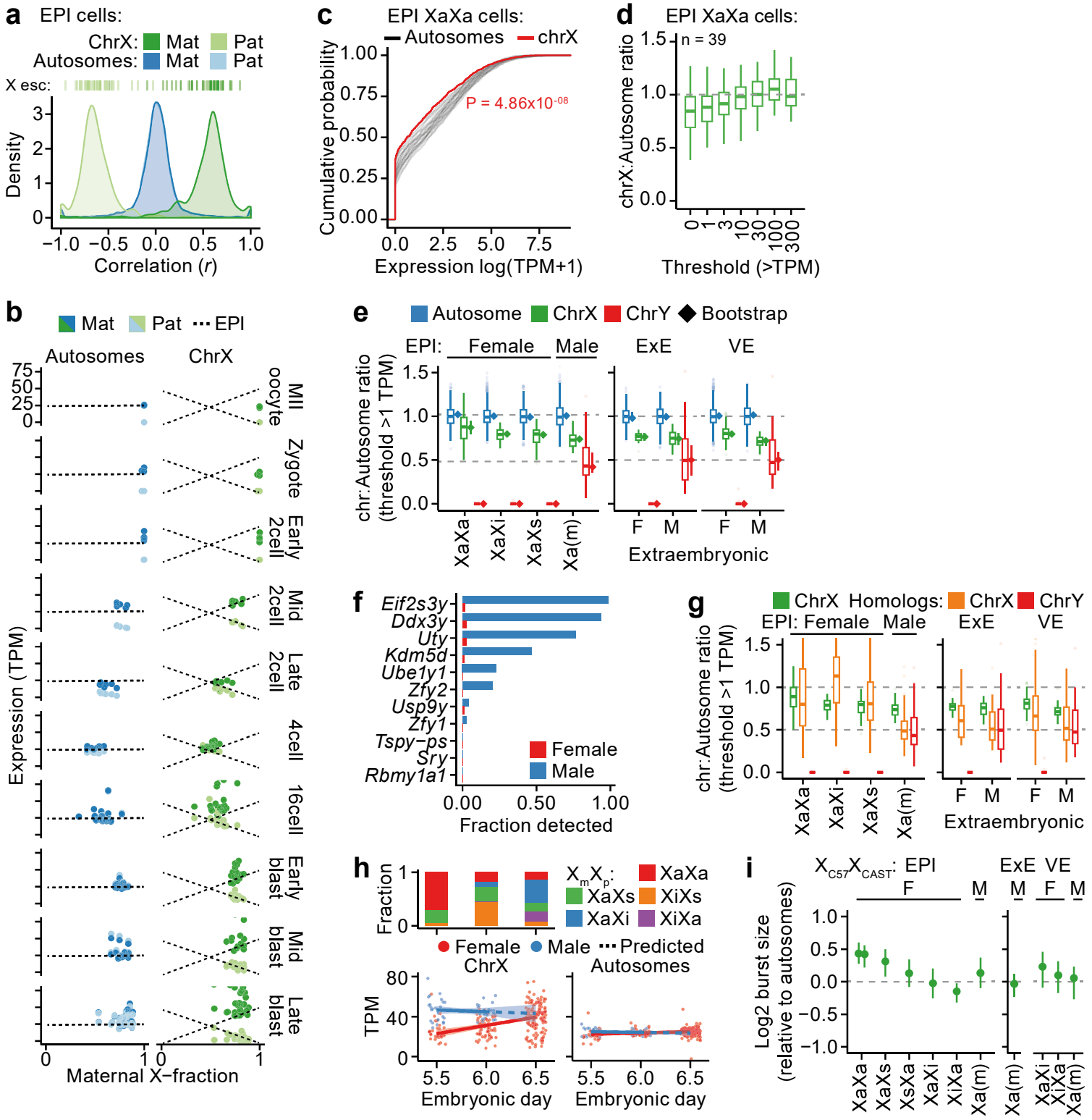
Supplementary Figure 1. Validation of mESC differentiation. a, Metrics for generated Smart-seq3 data. **b**, Heatmap of differentially expressed genes along mESC differentiation. **c**, Enriched pathways for differentially expressed genes in (b). **d**, Xist expression along mESC differentiation. **e**, UMI counts for subsampled reads (n = All, 1M, 0.75M, 0.5M, 0.25M, 0.1M and 0.05M reads).

Supplementary Figure 2



Supplementary Figure 2. Extended analysis of early embryo development. a, UMAP dimensionality reduction and graph-based clustering of blastocysts. **b**, Heatmap and hierarchical clustering of blastocysts using lineage-specific marker genes. **c**, Allelic expression of late blastocyst subclusters identified in (a-b). **d**, Diffusion map dimensionality reduction and trajectory inference per post-implantation lineage. **e**, Allelic expression along pseudotime trajectories for post-implantation embryos. **f**, Association of allele usage or pseudotime on allelic expression using linear modelling.

Supplementary Figure 3



Supplementary Figure 3. Extended analyses of linear relationship of X-upregulation. a, Density plot of gene-wise correlations (Pearson) against maternal X-fractions in EPI cells. X esc = X escapees. **b**, Relationship between X expression and maternal X-fraction, dashed lines represent EPI. **c**, Empirical cumulative distribution of expression for autosomes and the X chromosome. Kolmogorov-Smirnov test P-value. **d**, Effect of expression thresholds for calculating X:Autosome ratios in EPI XaXa cells. **e**, Autosomal expression ratios at >1 TPM threshold presented as box plots or bootstrapped ratios shown as median ± 95% confidence interval (♦). **f**, Fraction of expressed genes on the Y chromosome. **g**, Same as (e) but X-Y gene pair homologs shown separately. **h**, X status per time point (top). Expression of Xa allele along embryonic development day in EPI cells (bottom). Data shown as linear model mean ± 95% confidence interval, dashed lines representing values predicted from linear model. (**i**) Burst sizes of the active X allele (Xa) relative to autosomal genes, shown as median ± 95% confidence interval.