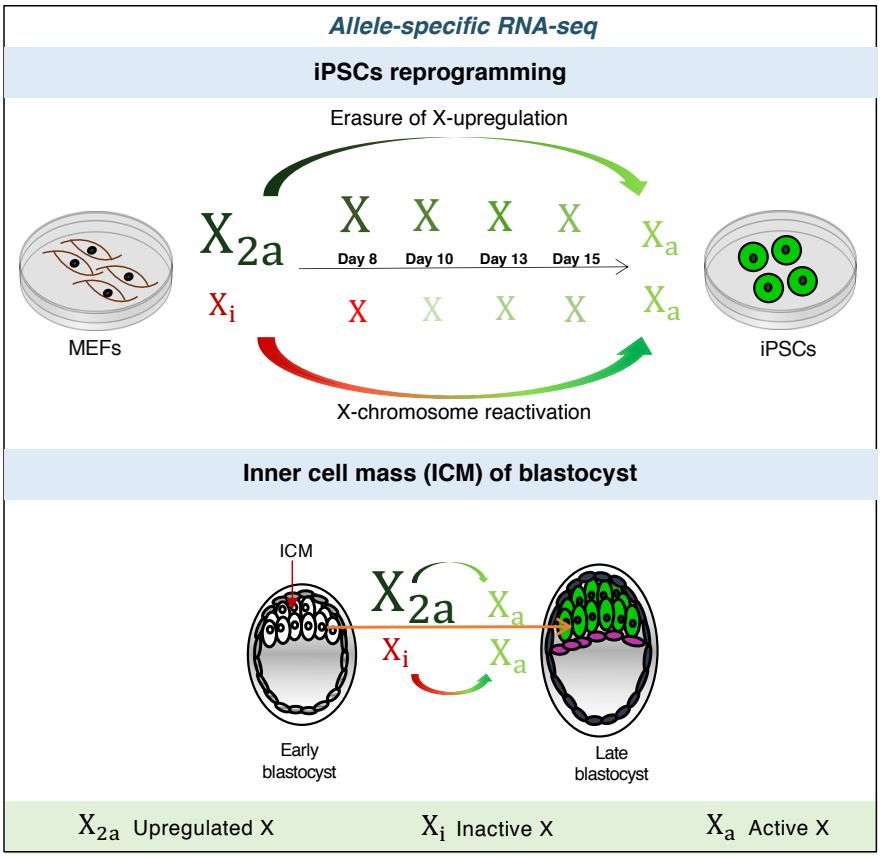
Dynamic erasure of X chromosome upregulation during iPSC reprogramming and in the inner cell mass

Chandel D[1], Naik C H[1] and Gayen S*

Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore-560012, India.

*Correspondence: srimonta@iisc.ac.in

[1] Equal contribution



Summary:

Recent studies have provided substantial evidence supporting Ohno's hypothesis that upregulation of active X chromosome genes balances the dosage relative to autosomal genes. Here, we found the evidence for erasure of active X-chromosome upregulation upon reactivation of inactive X in female cells during iPSC reprogramming and in the inner cell mass of mouse blastocyst. Moreover, our analysis revealed that the kinetics of erasure of X upregulation is tightly linked with the dynamics of X reactivation.

In therian mammals, sex is determined by the sex chromosomes: XX in female and XY in male. During the evolution of sex chromosomes from a pair of autosomes, the Y chromosome's degradation led to the dosage imbalance between X-linked and autosomal genes [1]. In 1967, Ohno hypothesized that the expression of X-linked genes is upregulated two-fold to balance the dosage relative to autosomal genes [2]. Subsequently, to counteract the extra dosage from X-

chromosome, the evolution of X-chromosome inactivation happened in female cells [3]. While X-inactivation has been extensively studied, the existence of X-chromosome upregulation process remains controversial and has been challenged [4–6]. Recent studies by us and others have provided substantial shreds of evidence for the existence of upregulated active-X chromosome in both males and females [7–15]. However, the status of X-upregulation in the context of reactivation of the inactive X chromosome has not been explored. Here, we dissect the X-upregulation dynamics upon reactivation of the inactive X chromosome in female cells during cellular reprogramming *in vivo* and *in vitro*.

We started with exploring the X-chromosome dynamics during the reprogramming of female mouse embryonic fibroblast (MEF) to induced pluripotent stem cells (iPSC). First, we profiled the X-to-autosomal (X:A) gene expression ratio in different female MEF cell lines to confirm the presence of upregulated active X. If a diploid female cell (harboring one active X (Xa) and one inactive X (Xi)) has upregulated active X, the X:A ratio should be more than 0.5 and closer to 1. Indeed, we found that X:A ratio of all MEF cell lines was greater than 0.5, signifying the presence of upregulated active X chromosome (Fig. 1A, Fig S1). Next, we determined the X:A ratio in female iPSCs. Since inactive-X gets reactivated in iPSC, we expected a higher X:A ratio in iPSC than MEF. However, we found no significant difference in X:A ratio between iPSCs and MEF cells. Moreover, X:A ratio of iPSCs was very similar to female embryonic stem cells (ESCs), which also have two active X-chromosomes (XaXa) (Fig. 1A, Fig. S1). We extended our analysis by profiling X:A ratio throughout the different iPSC reprogramming stages using published RNA-seq dataset from Janiszewski et al. [16] (Fig. 1B). Janiszewski et al. showed that reactivation of Xi triggers robustly at d10 of reprogramming and completed at d15 or iPSC, which is consistent in our analysis. Intriguingly, we found that X:A ratio remains almost unaltered throughout the different days of reprogramming and reaches ~1 at the end though there was robust Xi reactivation (Fig. 1C, Fig S1). Altogether, this fine balance of X to autosome dosage throughout the reprogramming stages, despite reactivation of Xi, suggested dynamic erasure of X-upregulation during iPSC reprogramming. To probe this further, we compared the allelic expression of autosomal and X-linked genes at different stages of reprogramming (Fig. 1D). These cells were derived from two diverge mouse strains (129S1 and CAST) and have many polymorphic sites between the paternal and maternal chromosomes, which allowed us to perform allele-specific analysis [16]. Notably, X-inactivation is skewed to the 129S1 allele in these cells, making it possible to distinguish allelic X-linked gene expression from the active vs. inactive allele [16].

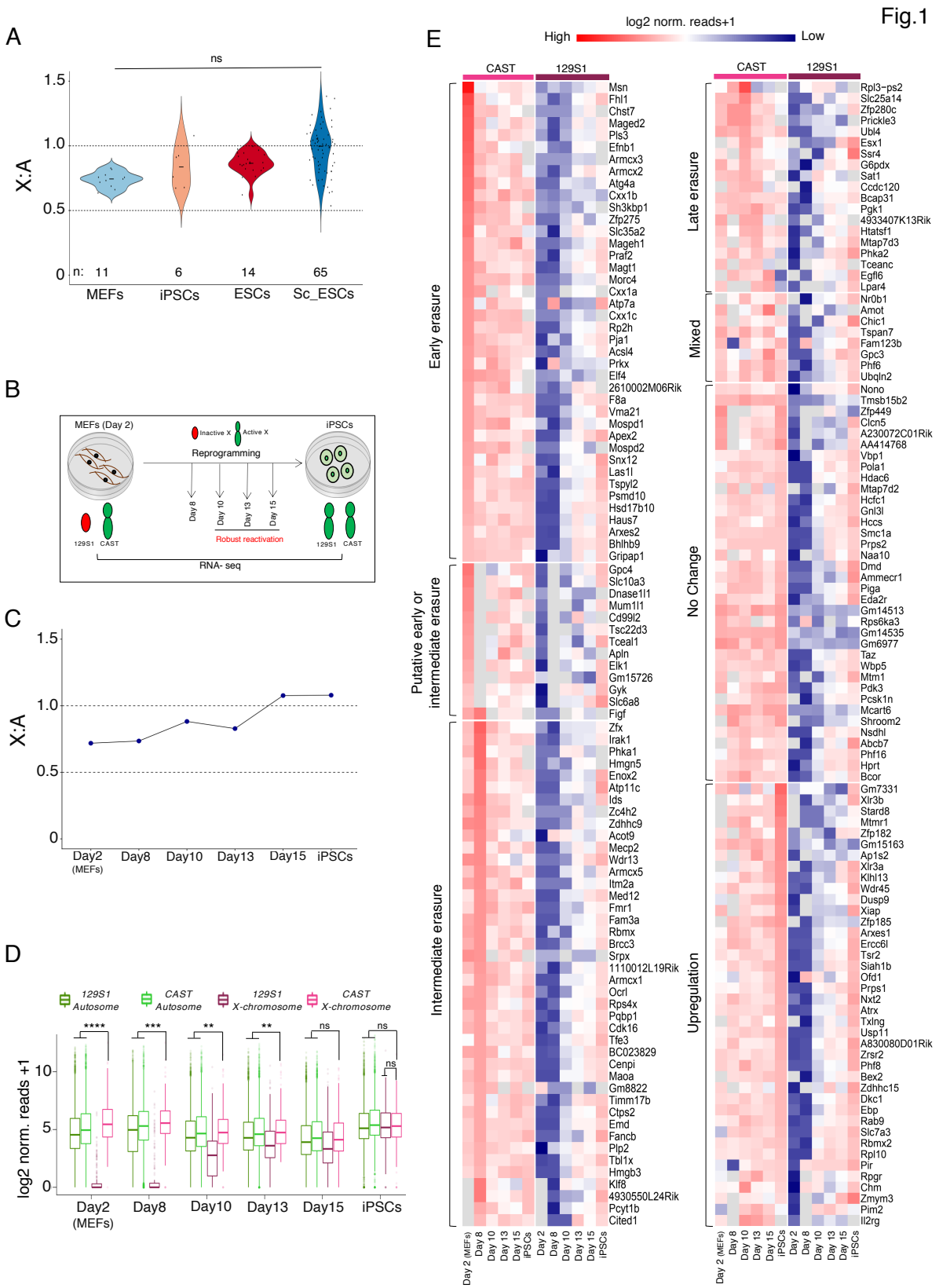


Figure 1: Erasure of active-X upregulation during iPSCs reprogramming. (A) X:A ratio in different mouse cell lines of MEFs (n=11), iPSCs (n=6), ESCs (n=14) and Sc-ESCs (Single-cells)

(n=65). Non-significant at $p < 0.05$. (B) Schematic of iPSC reprogramming stages along with X-reactivation status. (C) X:A ratio throughout the different stages of iPSC reprogramming. (D) Allelic expression of X-linked and autosomal genes (log2 normalized reads) during the reprogramming period. Statistical significance was assessed using Wilcoxon's test. **** $p < 0.0001$, *** $p < 0.001$, ns at $p < 0.05$. (E) Heatmap representing allelic expression from 129S1 and CAST allele (Log2 normalized allelic reads) of X-linked genes (n=198) throughout the reprogramming process. Genes are classified into different categories based on the erasure kinetics during the reprogramming process.

As expected, in MEF cells, we found significantly higher expression from the active X (CAST allele) compare to the autosomes, corroborating distinct upregulation of X-linked gene expression from the active X (Fig. 1D). Moreover, we found that the difference of expression between autosomes and active X (CAST) gradually decreases throughout the reprogramming and becomes non-significant at Day15 or iPSC, signifying dynamic erasure of active-X upregulation (Fig. 1D). Importantly, erasure of X-upregulation was linked dynamically to the kinetics of X-reactivation (Fig. 1D). Next, we delineated the dynamics of erasure of X upregulation of individual X-linked genes throughout the different stages of reprogramming (Fig. 1E). We found indeed many X-linked genes (n=114) showing the decrease in expression from the active X chromosome upon reprogramming. However, the kinetics of upregulation erasure was diverse; while some genes were early erasure, others showed intermediate or late erasure (Fig. 1E). Altogether, these analyses indicated dynamic erasure of X-chromosome upregulation upon reactivation of the inactive X chromosome during iPSC reprogramming.

Next, we asked if the erasure of X-upregulation is mimicked *in vivo* during the reactivation of Xi in the inner cell mass of mouse late blastocyst. In mice, X-inactivation happens in two phases: initially imprinted X-inactivation (IXi) and later random X-inactivation (RXi) in post-implantation embryos [17]. Switching of imprinted to random X-inactivation is mediated through the reactivation of the imprinted inactive-X in the inner cell mass of late blastocyst [18,19]. We profiled the dynamics of X-chromosome upregulation during early development in different lineages of mouse pre-and post-implantation (E5.5, E6.25, E6.5) embryos at the single-cell level using available single-cell RNA-seq dataset (Fig. 2; Fig S2). First, we categorized the cells into three categories based on the X-chromosome inactivation status: no X-inactivation (XaXa), partial X-inactivation (XaXp), and complete X-inactivation (XaXi) through profiling the fraction of maternal expression (Fig. 2A; Fig S2). Intriguingly, we found that X:A ratio of XaXp/XaXi cells is always >0.5 and close to 1 despite of X-inactivation and indistinguishable

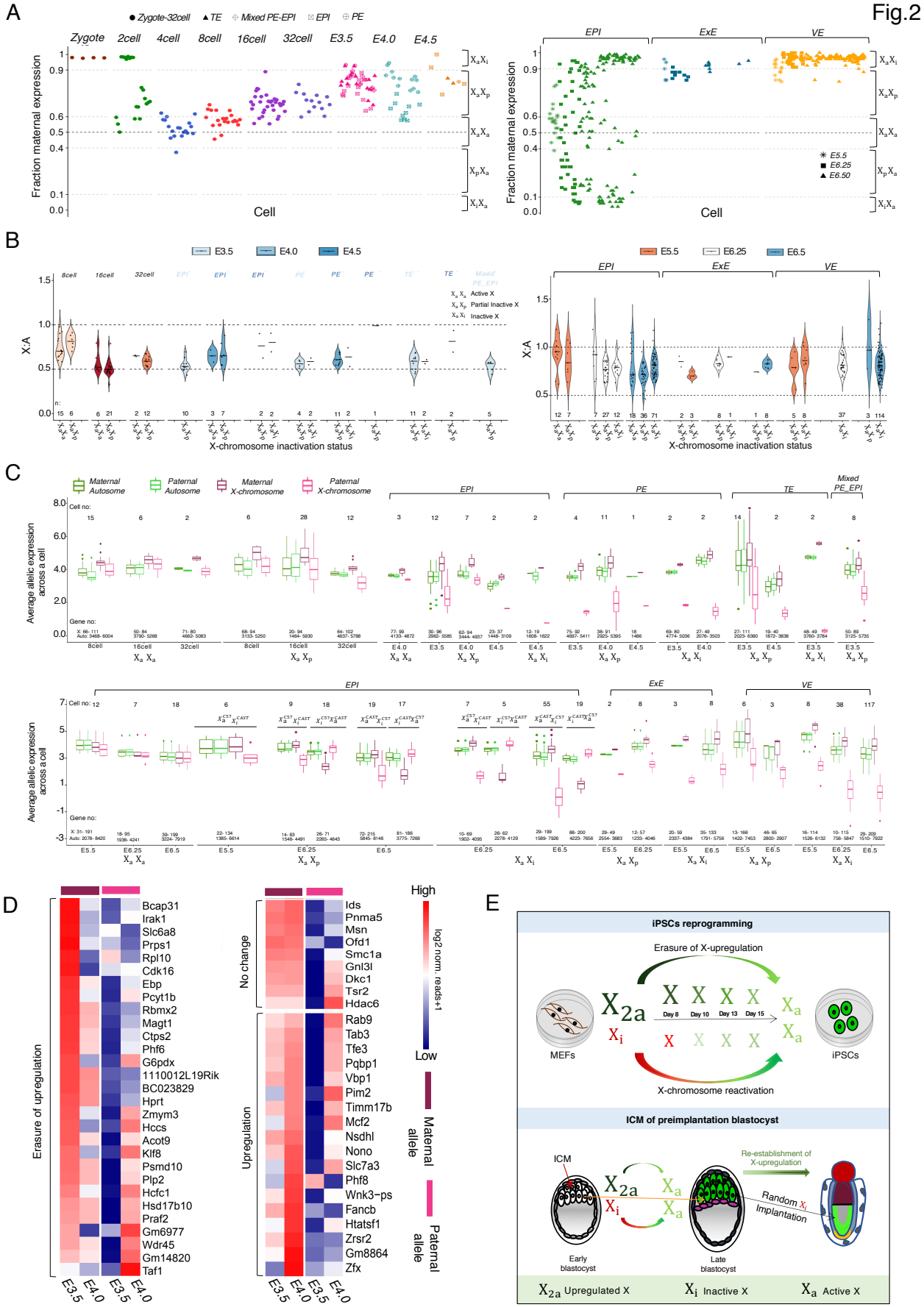


Fig.2
A
Zygote-32cell
TE
Mixed PE-EPI
EPI
PE
Zygote
2cell
4cell
8cell
16cell
32cell
E3.5
E4.0
E4.5
Fraction maternal expression
Cell
EPI
ExE
VE
E5.5
E6.25
E6.50
B
E3.5
E4.0
E4.5
E5.5
E6.25
E6.5
X:A
Active X
Partial Inactive X
Inactive X
X-chromosome inactivation status
C
Maternal Autosome
Paternal Autosome
Maternal X-chromosome
Paternal X-chromosome
Mixed PE_EPI
Average allelic expression across a cell
Cell no:
Gene no:
D
Erasure of upregulation
No change
Upregulation
Bcap31
Irak1
Slc6a8
Prps1
Rpl10
Cdk16
Ebp
Pcyt1b
Rbmx2
Magt1
Ctps2
Phf6
G6pdx
1110012L19Rik
BC023829
Hprt
Zmym3
Hccs
Acot9
Klf8
Psmd10
Plp2
Hcfc1
Hsd17b10
Praf2
Gm6977
Wdr45
Gm14820
Taf1
Ids
Pnma5
Msn
Ofd1
Smc1a
Gnl3l
Dkc1
Tsr2
Hdac6
Rab9
Tab3
Tfe3
Pqbp1
Vbp1
Pim2
Timm17b
Mcf2
Nsdhl
Nono
Slc7a3
Phf8
Wnk3-ps
Fancb
Htatsf1
Zrsr2
Gm8864
Zfx
High
Low
log2 norm. reads+1
Maternal allele
Paternal allele
E3.5
E4.0
E3.5
E4.0
E
iPSCs reprogramming
Erasure of X-upregulation
MEFs
Day 8
Day 10
Day 13
Day 15
iPSCs
X-chromosome reactivation
ICM of preimplantation blastocyst
ICM
Re-establishment of X-upregulation
Random Xi Inactivation
Early blastocyst
Late blastocyst
X2a Upregulated X
Xi Inactive X
Xa Active X

Figure 2: Erasure of active-X upregulation in the inner cell mass. (A) Classification of cells based on X-inactivation state through profiling of fraction maternal expression of X-linked genes in the single cells of different lineages of early embryonic stages beginning from Zygote to E6.5. EPI: Epiblast, PE: Primitive endoderm, TE: Trophectoderm, ExE: extraembryonic ectoderm, and VE: Visceral endoderm. (B) X:A ratios represented as violin plots for the XaXa, XaXp, XaXi cells of different embryonic stages (8cell- E6.5). (C) Allelic expression levels (log normalized reads) of X-linked and autosomal genes in XaXa, XaXp, XaXi cells of different embryonic stages (8cell- E6.5). (D) Heatmap representing allelic expression (Log2 normalised allelic reads) of X-linked genes (n=56) in E3.5 (EPI and Mixed_PE_EPI) XaXi cells vs E4.0 (EPI) XaXa cells. Genes are classified into different categories based on erasure kinetics. (E) Model representing dynamic erasure of X-chromosome upregulation during iPSCs reprogramming as well as *in vivo* in the ICM.

from XaXa cells (Fig. 2B; Fig S2). This key observation indicated the presence of dynamic X-upregulation from the active X chromosome along with X-chromosome inactivation to balance the X to autosomes dosage. To explore this further, we profiled the allelic expression pattern from autosome and as well as X-chromosome. Interestingly, we found that active X expression is always significantly higher compared to the autosomal allelic expression in XaXp/XaXi cells, corroborating upregulation of gene expression from the active X -chromosome (Fig. 2C; Fig S2). Altogether, these analyses suggested that X-chromosome upregulation is dynamically linked with X-inactivation to balance the X to autosomal dosage during early embryonic development. Notably, X:A ratio of reactivated (XaXa) late blastocyst (E4.0) epiblast precursor cells was indistinguishable from the XaXp/XaXi cells indicating the erasure of upregulation in these cells (Fig. 2B). This notion was further supported by allelic expression from maternal X in these cells, which showed no significant difference with the autosomal allelic expression (Fig. 2C). We extended our analysis by comparing the maternal X expression between XaXi E3.5 Epi/Epi-PE and XaXa E4.0 Epi cells and found that many X-linked genes showed downregulation of expression from the maternal X chromosome in XaXa E4.0 Epi cells compared to the early blastocyst XaXi E3.5 cells (Fig. 2D). Together, we found erasure of X-upregulation *in vivo* during the reactivation of Xi in the inner cell mass. Additionally, we found that the erased genes again get upregulated in RXi cells (Fig. S3B). Interestingly, we found that significant no. of genes undergoing upregulation erasure were common between iPSC reprogramming and inner cell mass (Fig. S3A).

In summary, our study for the first time shows dynamic erasure of X-chromosome upregulation during the iPSC reprogramming and in the inner cell mass (Fig. 2E). Moreover, we found that X-chromosome states (upregulation vs. inactivation) are linked together during reprogramming or early development.

Author's Contribution

SG conceptualized and supervised the study. DC and HCN did bioinformatic analyses. SG, DC, and HCN wrote, edited, and proofread the manuscript. The final manuscript was edited and approved by all the authors.

Acknowledgments

Study is supported by DBT grant (BT/PR30399/BRB/10/1746/2018), DST-SERB (CRG/2019/003067), DBT-Ramalingaswamy fellowship (BT/RLF/Re-entry/05/2016) and Infosys Young Investigator award to SG. We also thank DST-FIST [SR/FST/LS11-036/2014(C)], UGC-SAP [F.4.13/2018/DRS-III (SAP-II)] and DBT-IISc Partnership Program Phase-II (BT/PR27952-INF/22/212/2018) for infrastructure and financial support.

References:

1. Graves, J. A. M. Evolution of vertebrate sex chromosomes and dosage compensation. *Nature Reviews Genetics* **17**, 33–46 (2016).

2. Ohno S. Sex Chromosomes and Sex-linked Genes. *Springer-Verlag, Berlin, Heidelberg, New York* **68**, 1375 (1967).

3. Lyon, M. F. Gene action in the X-chromosome of the mouse (mus musculus L.). *Nature* **190**, 372–373 (1961).

4. Xiong, Y. *et al.* RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat. Genet.* **42**, 1043–1047 (2010).

5. Chen, X. & Zhang, J. The X to autosome expression ratio in haploid and diploid human embryonic stem cells. *Mol. Biol. Evol.* **33**, 3104–3107 (2016).

6. Julien, P. *et al.* Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* (2012). doi:10.1371/journal.pbio.1001328

7. Mandal, S. *et al.* Single-Cell Analysis Reveals Partial Reactivation of X Chromosome instead of Chromosome-wide Dampening in Naive Human Pluripotent Stem Cells. *Stem Cell Reports* **14**, 745–754 (2020).

8. 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).

9. 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).

10. 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.e3 (2017).

11. Li, X. *et al.* Dosage compensation in the process of inactivation/reactivation during both germ cell development and early embryogenesis in mouse. *Sci. Rep.* **7**, (2017).

12. Deng, X. *et al.* Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev. Cell* **25**, 55–68 (2013).

13. Lin, H. *et al.* Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis. *Nat. Genet.* (2011). doi:10.1038/ng.992

14. De Mello, J. C. M., 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**, 1–12 (2017).

15. Kharchenko, P. V., Xi, R. & Park, P. J. Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nature Genetics* **43**, 1167–1169 (2011).

16. Janiszewski, A. *et al.* Dynamic Reversal of Random X-Chromosome Inactivation during iPSC Reprogramming. *Genome Res.* (2019). doi:10.1101/545558

17. Saiba, R., Arava, M. & Gayen, S. Dosage compensation in human pre-implantation

embryos: X-chromosome inactivation or dampening? *EMBO Rep.* **19**, (2018).

18. Maclary, E. *et al.* Differentiation-dependent requirement of Tsix long non-coding RNA in imprinted X-chromosome inactivation. *Nat. Commun.* **5**, (2014).

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

Supplementary

Method

Data acquisition: Transcriptome datasets (related to Fig.1) used for the analysis was obtained from Gene Expression Omnibus (GEO): ESCs- GSE99491[1]; GSE94481[2]; GSE90516 [3]; GSE110215[4]; GSE126229 [5]; GSE101292 [6]; scRNA-seq dataset for ESCs- GSE55291[7] GSE74155[8]; MEFs- GSE47395 [9]; GSE118443[10] GSE67265[11]; GSE126229[5] iPSCs - GSE110215[4]; GSE67265[11] GSE126229[5]. iPSC reprogramming data with different intermediate stages was obtained from GSE126229 [5].

Single-cell transcriptomic datasets (related to Fig.2), were obtained from following datasets: pre-implantation embryos: GSE45719 [12]; GSE74155 [8]; GSE80810[13]; GSE89900[14]. Post-implantation embryos: GSE109071[15].

Lineage classification: To determine the lineages of the pre-implantation blastocyst cells (79 cells), we used the R package Seurat (version 3.2.2)[16,17] with a modified Cell Cycle Scoring function using lineage-specific (EPI, PE and TE) marker genes. A cell was assigned as mixed PE-EPI (typically the inner cell mass) if there were no significant difference of score between PE and EPI.

For the post-implantation blastocysts, single cells (510 cells) from different stages were subjected to dimension reduction algorithm using t-distributed stochastic neighbor embedding (t-SNE) to identify the different lineages as described previously[18]. t-SNE was performed taking 3000 most variable genes using the Seurat package in R. The allocation of each cell cluster to different lineages EPI, ExE, and VE was based on the expression of bona fide marker genes: Oct4 for EPI, Bmp4 for ExE, and Amn for VE.

Read alignment and counting: RNA-seq reads were aligned to the mouse reference genome GRCm38 (mm10) using STAR[19] with default parameters. The reads aligned were then counted using HTSeq-count (Anders et al., 2015). To avoid the dropout events due to low amount of mRNA sequenced within embryonic single cells, we used a statistical imputation method scImpute[20], which identifies the likely dropouts without introducing any bias in the rest of the data. Expression levels of transcripts was computed using *Transcripts per million (TPM)* method.

Allele-specific expression analysis: For the single-cell allelic expression analysis of genes, we first constructed an *in silico* CAST specific parental genome by incorporating CAST/EiJ specific SNPs into the GRCm38 (mm10) reference genome using VCF tools. CAST specific SNPs were obtained from the Mouse Genomes Project (https://www.sanger.ac.uk/science/data/mouse-genomes-project). Reads were mapped onto both C57BL/6J (mm10) reference genome and CAST/EiJ *in silico* parental genome using STAR allowing no multi-mapped reads. To exclude any false positives, we only considered those genes with at least 2 informative SNPs (minimum 3 reads per SNP site). We took an average of SNP-wise reads to have the allelic read counts. We normalized allelic read counts using the available spike-ins in the dataset. However, the data without spike-ins was normalized using DESeq2 scaling factor calculated from raw read count matrix. We considered only those genes for downstream analysis which had at least 1 average reads across the cells of each lineage from a specific developmental stage. Further, only those single-cells were considered for downstream analysis which showed at least 10 X-linked gene expression. Allelic ratio was calculated individually for each gene using formula = (Maternal/Paternal reads) ÷ (Maternal reads + Paternal reads).

For the bulk RNA-seq datasets pertaining to iPSCs reprogramming, *in silico* 129S1/SvImJ and CAST specific parental genomes were first constructed by incorporating 129S1/SvImJ and CAST/EiJ SNPs into the GRCm38 (mm10) reference genome using VCF tools. Reads were then mapped to the parental genomes using STAR allowing no multimapped reads. Genes with atleast 2 SNPs having minimum of 10 reads per SNP site were used downstream, again to avoid any false positive expression. Allelic reads thus obtained were normalized by DESeq2 scaling factor calculated from the raw read count matrix.

Sexing of the embryos: We determined the sex of the pre-implantation embryos by using the average expression of four Y-linked genes: *Eif2s3y*, *Zfy2*, *Ube1y1* and *Kdm5d.* An embryo was assigned as male if the minimum average of the sum RPKM of the four Y-linked genes across the cells of an embryo is 15, which otherwise would be a female embryo.

On the other hand, for post-implantation embryos, an embryo was classified as male if the sum of the read count of the Y-linked genes (Usp9y, Uty, Ddx3y, Eif2s3y, Kdm5d, Ube1y1, Zfy2, Zfy1) in each cell of an embryo was greater than 12, rest were considered as female cells.

X:A ratio: We calculated the X:A ratio by dividing the median expression (TPM) of X-linked genes with the median expression (TPM) of the autosomal genes. For this analysis, we

considered those X-linked and autosomal genes having ≥ 0.5 TPM. We also profiled the expression-level distribution of X-linked and autosomal genes as density plots created in R using the ggplot2 package. Moreover for the X:A analysis in the embryonic single cells, we considered an upper TPM threshold which corresponded to the lowest 90th centile of TPM expression to avoid any differences between the X-linked and autosomal gene expression distribution. Also, statistical Kolmogorov-Smirnov's test was performed which again validated the non-significant differences in the levels of gene expression distribution between X and autosomal genes (Fig S1). We excluded the escapees and genes in the Pseudoautosomal region from our analysis.

Statistical tests and plots: All statistical tests and plots were performed in R version 3.6.3.

References

1. Bao, S. *et al.* Derivation of hypermethylated pluripotent embryonic stem cells with high potency. *Cell Res.* **28**, 22–34 (2018).

2. Choi, J. *et al.* DUSP9 Modulates DNA Hypomethylation in Female Mouse Pluripotent Stem Cells. *Cell Stem Cell* **20**, 706–719.e7 (2017).

3. Werner, R. J. *et al.* Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biol. Sex Differ.* **8**, (2017).

4. Song, J. *et al.* X-Chromosome Dosage Modulates Multiple Molecular and Cellular Properties of Mouse Pluripotent Stem Cells Independently of Global DNA Methylation Levels. *Stem Cell Reports* **12**, 333–350 (2019).

5. Janiszewski, A. *et al.* Dynamic Reversal of Random X-Chromosome Inactivation during iPSC Reprogramming. *Genome Res.* (2019). doi:10.1101/545558

6. Dirks, R. A. M. *et al.* Allele-specific RNA-seq expression profiling of imprinted genes in mouse isogenic pluripotent states. *Epigenetics and Chromatin* **12**, (2019).

7. Kim, J. K., Kolodziejczyk, A. A., Illicic, T., Teichmann, S. A. & Marioni, J. C. Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.* **6**, (2015).

8. Chen, G. *et al.* Single-cell analyses of X Chromosome inactivation dynamics and

pluripotency during differentiation. *Genome Res.* **26**, 1342–1354 (2016).

9. Bhatnagar, S. *et al.* Genetic and pharmacological reactivation of the mammalian inactive X chromosome. *Proc. Natl. Acad. Sci.* **111**, 12591–12598 (2014).

10. Mizukami, H. *et al.* KDM5D-mediated H3K4 demethylation is required for sexually dimorphic gene expression in mouse embryonic fibroblasts. *J. Biochem.* **165**, 335–342 (2019).

11. Maza, I. *et al.* Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33**, 769–774 (2015).

12. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science (80-.).* **343**, 193–196 (2014).

13. Borensztein, M. *et al.* Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nat. Struct. Mol. Biol.* **24**, 226–233 (2017).

14. Borensztein, M. *et al.* Contribution of epigenetic landscapes and transcription factors to X-chromosome reactivation in the inner cell mass. *Nat. Commun.* **8**, (2017).

15. Cheng, S. *et al.* 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).

16. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

17. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).

18. Naik, C. H., Chandel, D., Mandal, S. & Gayen, S. Transcriptional bursting shape autosomal dynamic random monoallelic expression in pre-gastrulation embryos. *bioRxiv* 2020.09.18.303776 (2020).

19. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

20. Li, W. V. & Li, J. J. An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* **9**, (2018).

Fig.S1

A

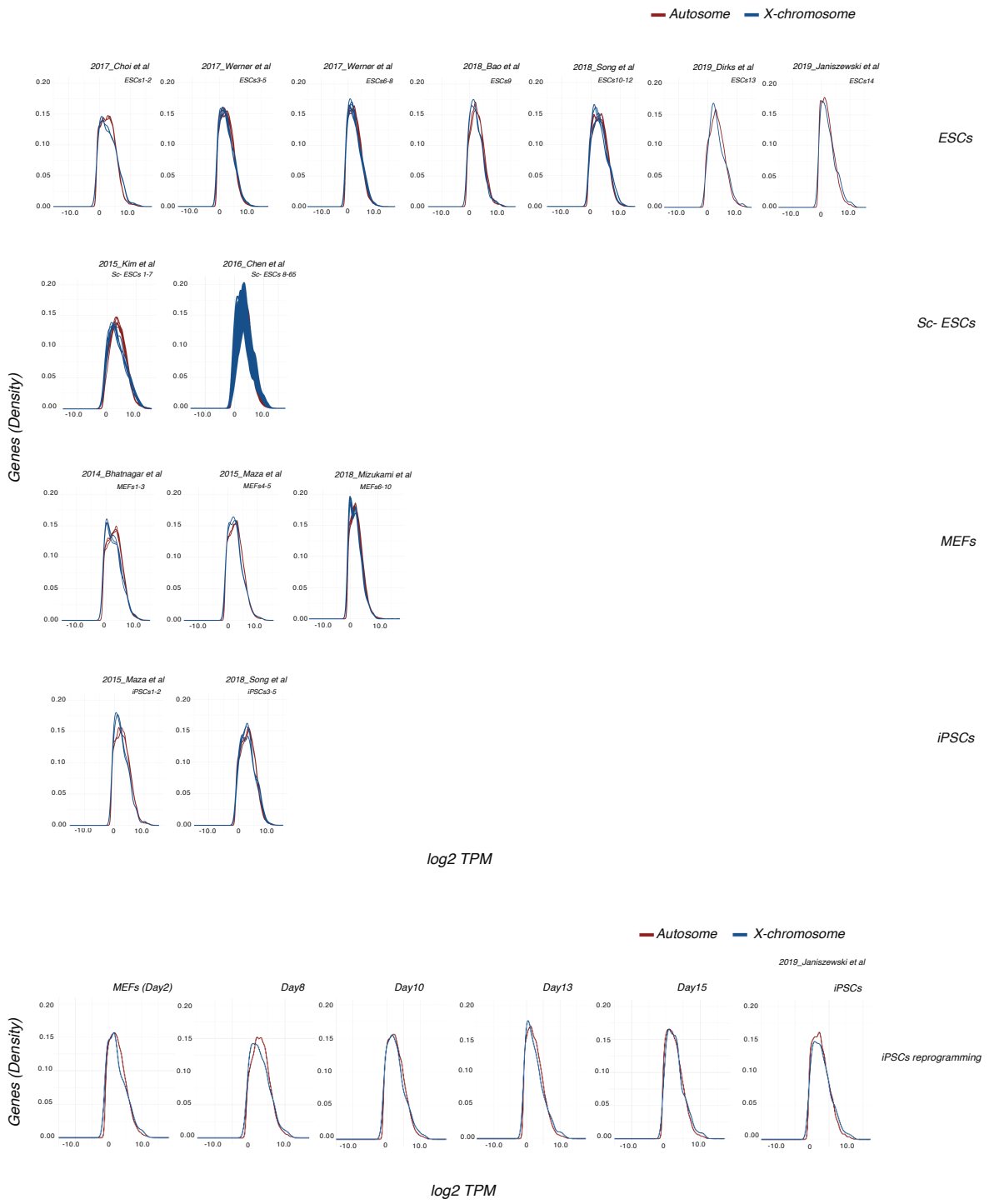
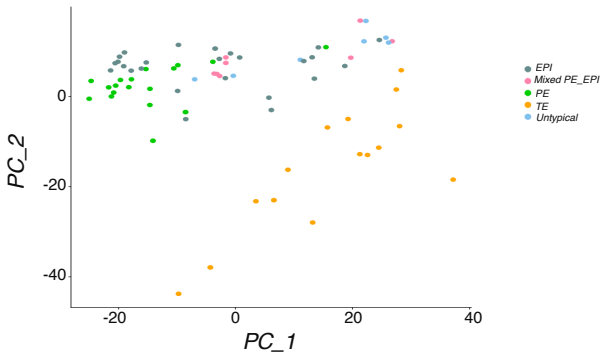


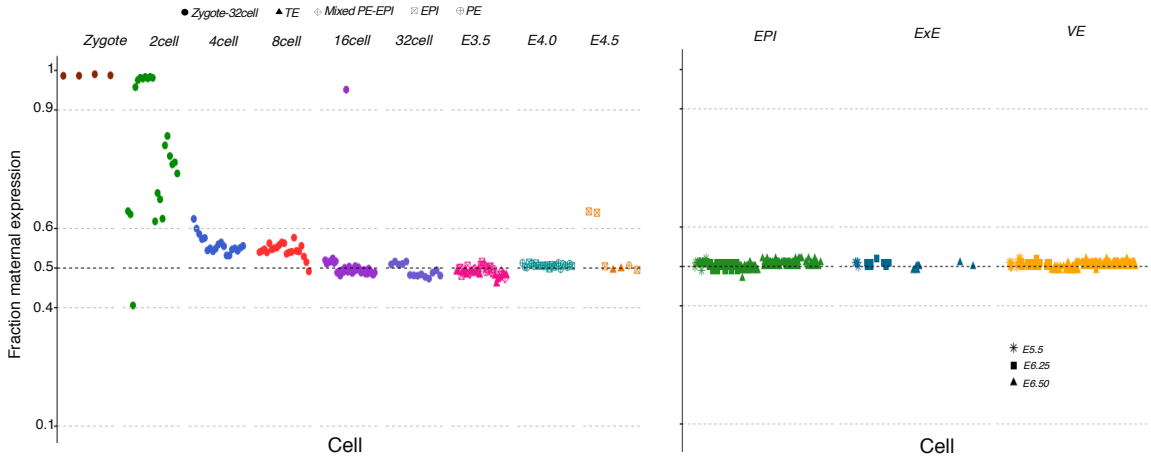
Figure S1, related to figure 1: Distribution of X-linked and autosomal gene expression (A) Histograms representing the distribution of X-linked and autosomal gene expression in ESCs, Sc-ESCs, MEFs, iPSCs (top panel) with their respective replicates and in the different stages of iPSCs reprogramming (bottom panel). ($p > 0.05$ by Kolmogorov Smirnov's test).

Fig.S2

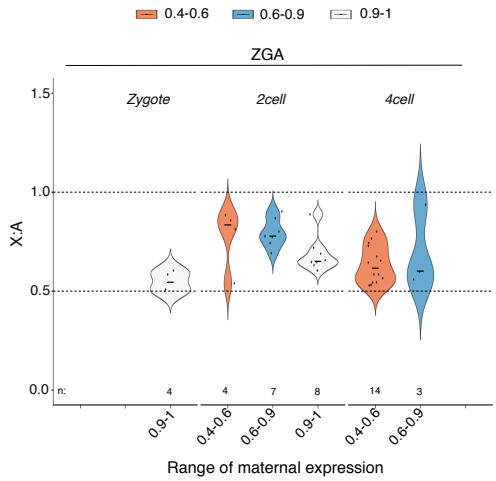
A


EPI
Mixed PE_EPI
PE
TE
Untypical
PC_2
PC_1
0
-20
-40
-20
0
20
40

B


Zygote-32cell
TE
Mixed PE-EPI
EPI
PE
Zygote
2cell
4cell
8cell
16cell
32cell
E3.5
E4.0
E4.5
EPI
ExE
VE
Fraction maternal expression
1
0.9
0.6
0.5
0.4
0.1
E5.5
E6.25
E6.50
Cell
Cell

C


0.4-0.6
0.6-0.9
0.9-1
ZGA
Zygote
2cell
4cell
X:A
1.5
1.0
0.5
0.0
n:
4
4
7
8
14
3
0.9-1
0.4-0.6
0.6-0.9
0.9-1
0.4-0.6
0.6-0.9
Range of maternal expression

D

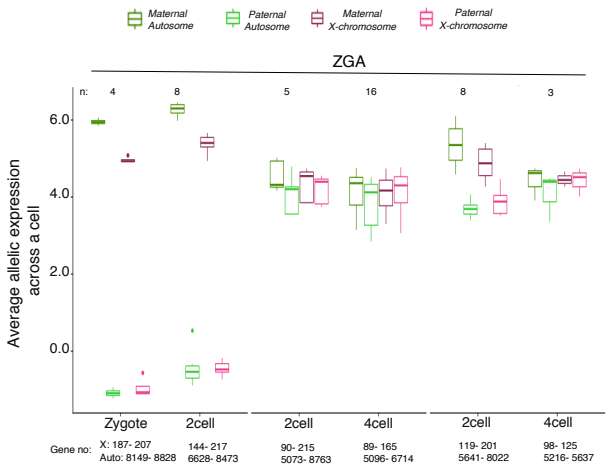

Maternal Autosome
Paternal Autosome
Maternal X-chromosome
Paternal X-chromosome
ZGA
n:
4
8
5
16
8
3
Average allelic expression across a cell
6.0
4.0
2.0
0.0
Zygote
2cell
2cell
4cell
2cell
4cell
Gene no:
X: 187- 207
Auto: 8149- 8828
144- 217
6628- 8473
90- 215
5073- 8763
89- 165
5096- 6714
119- 201
5641- 8022
98- 125
5216- 5637

Figure S2, related to figure 2: Additional analysis of pre- and post- implantation embryos (A) Clustering single cells of blastocysts based on the expression of marker genes specific to different lineages. Each dot represents a single-cell, and the color is based on its lineage. (B) Fraction maternal expression of autosomal genes in the single cells of early embryonic stages beginning from Zygote to E6.5. Cells are assigned to different lineages of pre-implantation development (EPI, PE, TE, Mixed PE-EPI; left panel) and post-implantation development (EPI, ExE and VE; right panel). (C) X:A ratio in Zygote and during the window of zygotic genome activation (2 to 4 cell stages). (D) Allelic expression levels (log2 normalised reads) of X-linked and autosomal genes in Zygote and during the window of zygotic genome activation (2 to 4 cell stages).

Fig. S3

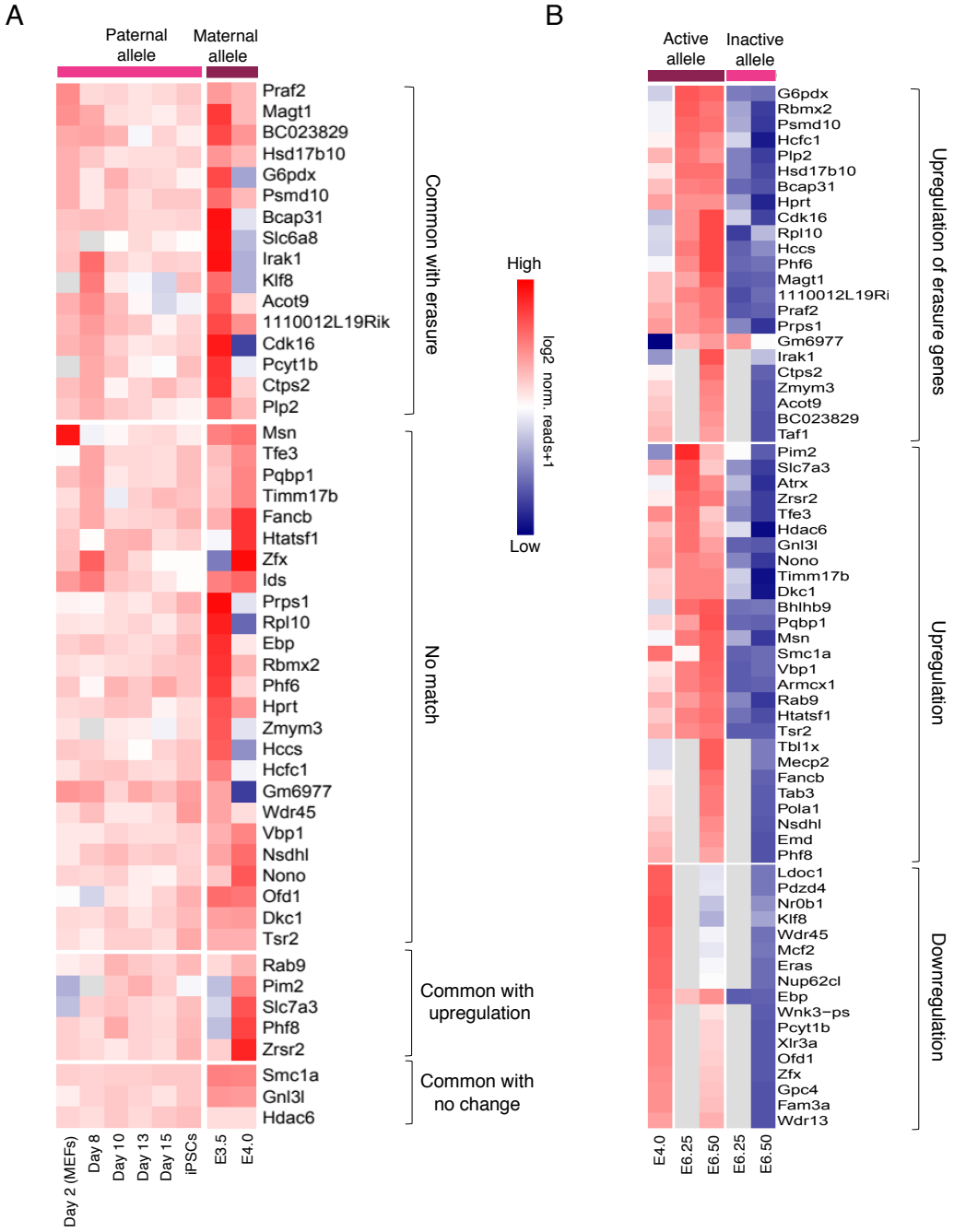


Figure S3: Extended analysis of X-chromosome upregulation dynamics of individual X-linked genes during iPSCs reprogramming and early embryonic development. (A) Comparison of the dynamics of erasure of X-upregulation of X-linked genes during iPSC and inner cell mass reprogramming. Out of 49 genes, 16 genes showed erasure of X-upregulation

in both iPSC and inner cell mass. (B) Re-establishments of X-upregulation upon random Xi in E6.25 and E6.5 embryos. Heatmap representing the dynamics of expression of X-linked genes (log2 normalized allelic reads) from active and Inactive-X in E4.0, E6.25 and E6.5 embryos.