

Supplemental Information

Lineage-specific dynamics of loss of X upregulation during inactive-X reactivation

Hemant Chandru Naik, Deepshikha Chandel, Sudeshna Majumdar, Maniteja Arava, Runumi Baro, Harshavardhan BV, Kishore Hari, Parichitran Ayyamperumal, Avinchal Manhas, Mohit Kumar Jolly, and Srimonta Gayen

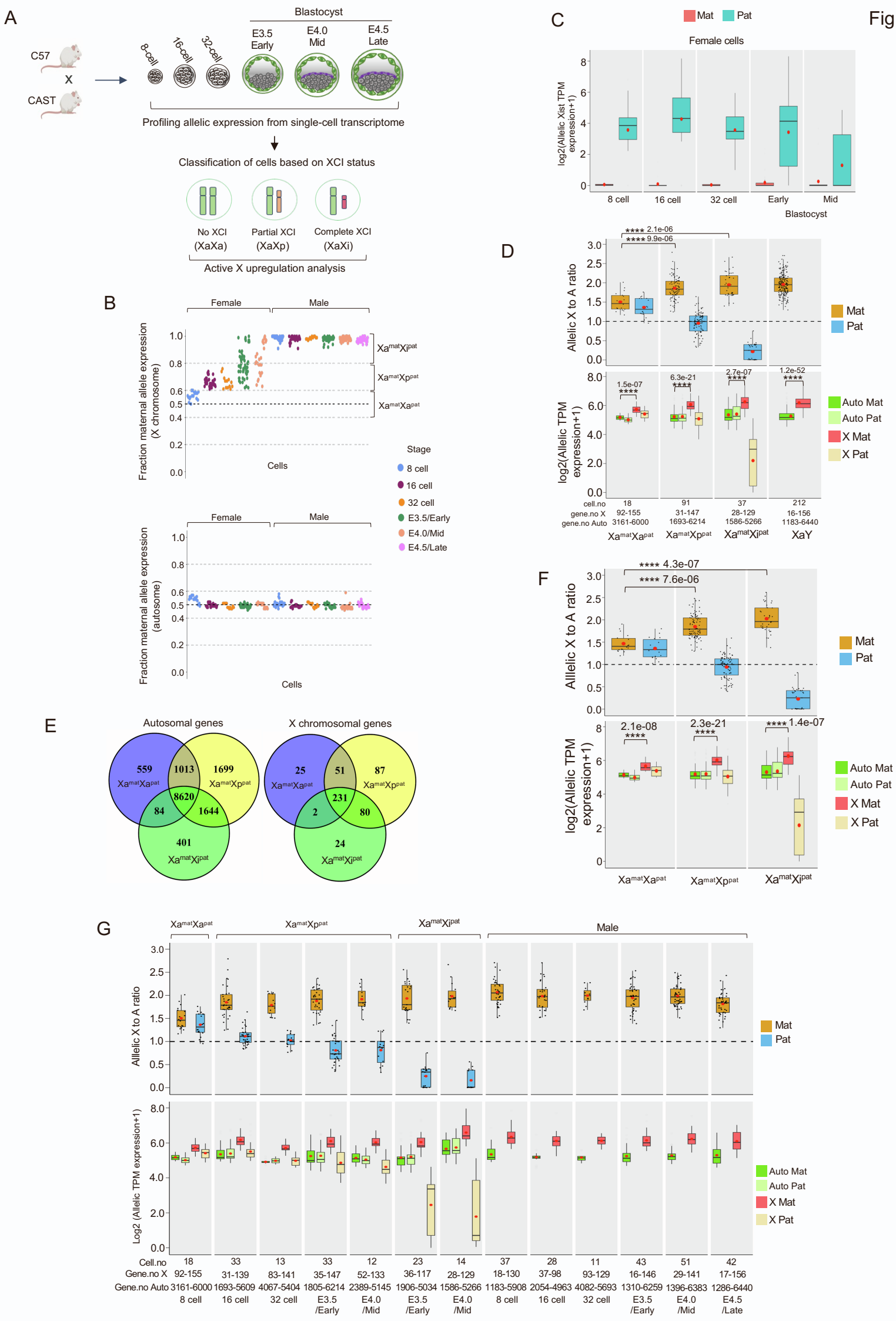


Figure S1 (related to Fig. 1). (A) Schematic representing the experimental workflow of profiling active-X upregulation upon X-inactivation in different stages of mouse pre-implantation hybrid embryos (8-cell, 16-cell, 32-cell, E3.5 early, E4.0 mid and E4.5 late blastocyst) at the single-cell level using scRNA-seq dataset. These embryos were generated from the crossing of two divergent mouse strains, C57 and cast. (B) Top: Classification of female cells of pre-implantation embryos based on X-inactivation state through profiling of fraction maternal expression of X-linked genes. Male cells showed expression from the maternal allele only. Bottom: Fraction maternal expression of autosomal genes in different stages of mouse pre-implantation embryos (8-cell, 16-cell, 32-cell, E3.5 early, E4.0 mid and E4.5 late blastocyst). (C) Box plot representing the allelic Xist expression in cells of different stages of pre-implantation embryos. (D) Top: Allelic X:A ratio plot and bottom: allelic expression (\log_2 allelic TPM+1) plots for X and autosomes in $Xa^{mat}Xa^{pat}$; $Xa^{mat}Xp^{pat}$ and $Xa^{mat}Xi^{pat}$ female cells and male cells of pre-implantation embryos. (E) Venn diagram representing the common set of genes (X and autosomal genes) among $Xa^{mat}Xa^{pat}$; $Xa^{mat}Xp^{pat}$ and $Xa^{mat}Xi^{pat}$ cells. (F) Top: Allelic X:A ratio plot and bottom: allelic expression (\log_2 allelic TPM+1) plots for a common set of genes (X and autosomal genes) among $Xa^{mat}Xa^{pat}$; $Xa^{mat}Xp^{pat}$ and $Xa^{mat}Xi^{pat}$ cells. In all boxplots, the line inside each of the boxes denotes the median value, red circle denotes the mean and the edges of each box represent 25% and 75% of dataset, respectively (Wilcoxon rank-sum test: P -value < 0.0001; ****). (G) Allelic X:A ratio and allelic expression (\log_2 allelic TPM+1) plots for X and autosomes in $Xa^{mat}Xa^{pat}$; $Xa^{mat}Xp^{pat}$ and $Xa^{mat}Xi^{pat}$ female cells and male cells throughout different stages of pre-implantation.

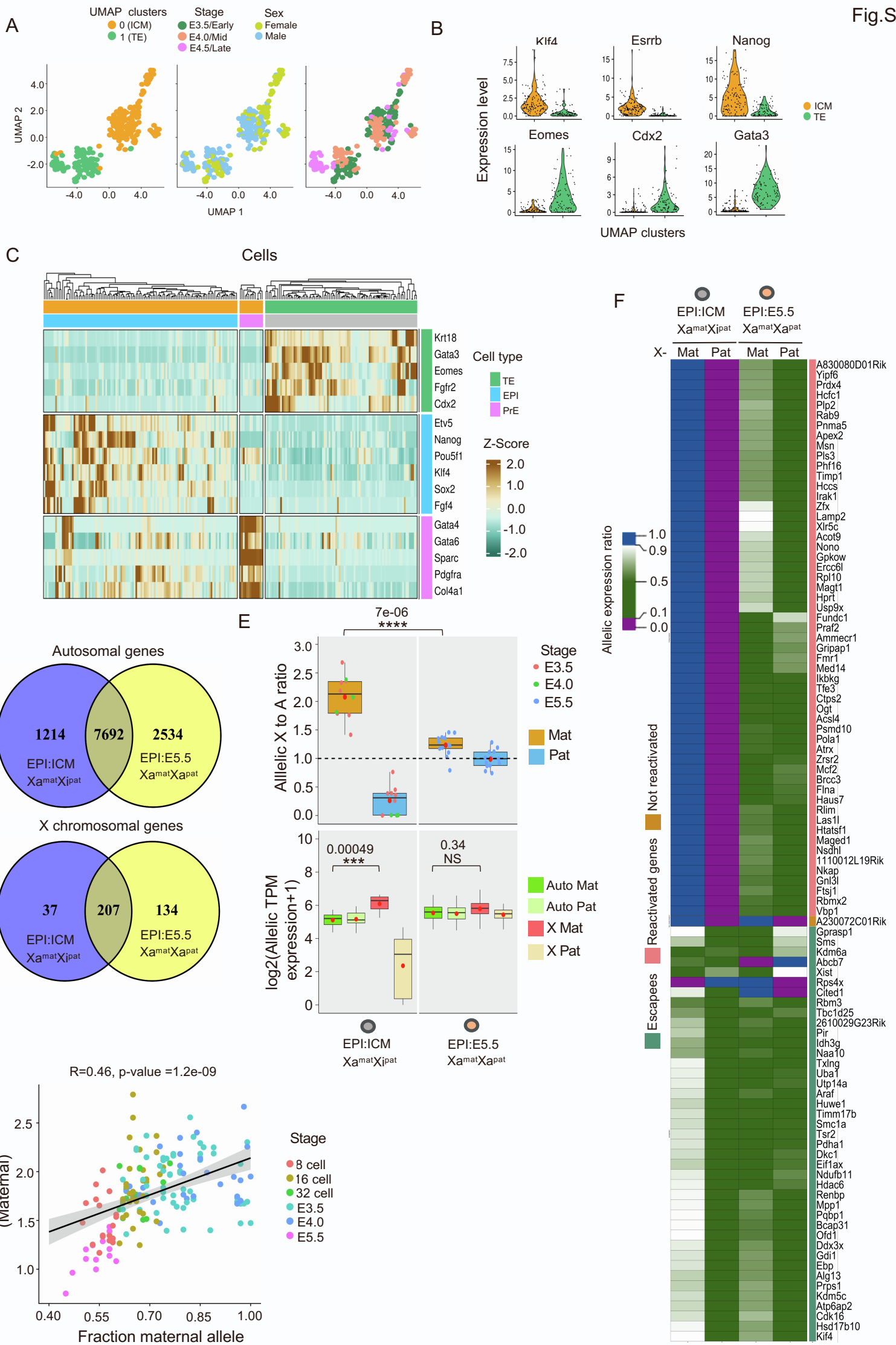


Figure S2 (related to Figure 1): (A) UMAP clustering of ICM (inner cell mass) and TE cells of pre-implantation embryos (E3.5 early, E4.0 mid and E4.5 late blastocyst). (B) Violin plots showing the expression of different markers corresponding to ICM and TE cells in UMAP-based clusters. (C) Heatmap representing the expression of markers of EPI, TE and PrE clusters. (D) Venn diagram representing the common set of genes (X and autosomal genes) among EPI:ICM ($X^{mat}X^{i\text{pat}}$) and EPI:E5.5 ($X^{mat}X^{a\text{pat}}$) cells. (E) Comparison of allelic X:A ratio (top) and allelic expression of X and autosomes (bottom) between EPI:ICM ($X^{mat}X^{i\text{pat}}$) and EPI:E5.5 ($X^{mat}X^{a\text{pat}}$) cells using a common set of genes. In boxplots, the line inside each of the boxes denotes the median value, the red circle denotes the mean and the edges of each box represent 25% and 75% of the dataset, respectively (Wilcoxon rank-sum test: P -value < 0.0001 ; **** P -value < 0.001 ; ***; NS: not significant). (F) Heatmap representing the allelic expression ratio of X-linked genes in EPI:ICM ($X^{mat}X^{i\text{pat}}$) and EPI:E5.5 ($X^{mat}X^{a\text{pat}}$) cells. Genes with fraction $X^{i\text{pat}}$ allele expression in EPI:ICM ($X^{mat}X^{i\text{pat}}$) cells < 0.10 are considered X-inactivated genes and > 0.10 as escapee genes. On the other hand, genes with a fraction of $X^{a\text{pat}}$ allele expression in EPI:E5.5 ($X^{mat}X^{a\text{pat}}$) cells > 0.10 are considered reactivated genes. Genes with fraction paternal allele expression: < 0.10 represents X-inactivated genes and > 0.10 represents escapee or reactivated genes. (G) Scatter plots of allelic X:A ratio (maternal) vs. fraction expression from maternal allele in cells of different stages (labelled with different colours) of pre-implantation embryos. R is Pearson's correlation.

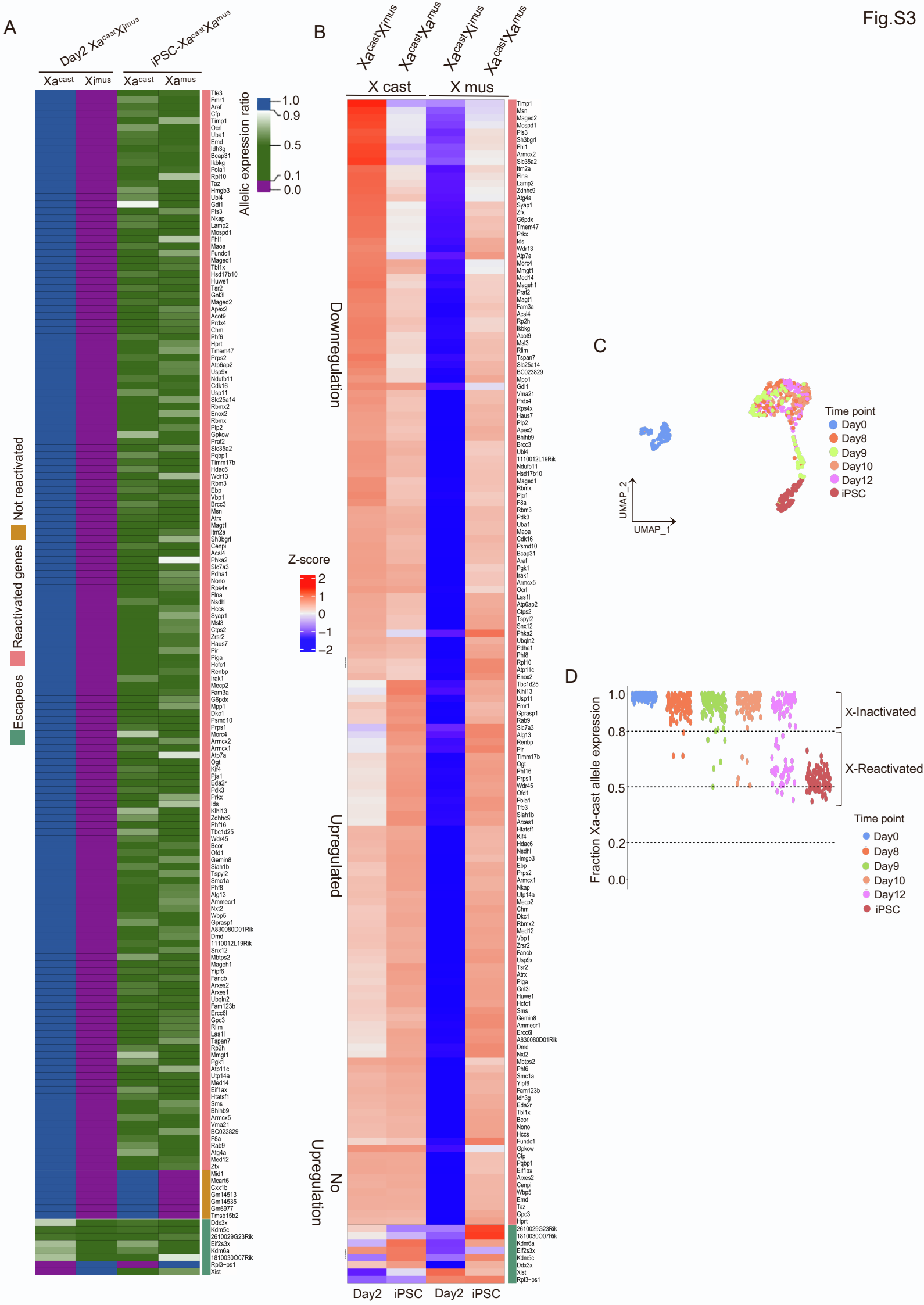


Fig S3 (related to Figure 2). (A) Heatmap representing the allelic expression ratio of X-linked genes in X-inactivated Day2 ($Xa^{cast}Xi^{mus}$) vs. X-reactivated iPSC ($Xa^{cast}Xa^{mus}$) cells. Genes with fraction Xi^{mus} allele expression in $Xa^{cast}Xi^{mus}$ cells < 0.10 are considered X-inactivated genes and >0.10 as escapee genes. On the other hand, genes with a fraction of Xa^{mus} allele expression in $Xa^{cast}Xa^{mus}$ cells >0.10 are considered reactivated genes. (B) Heatmap representing allelic expression from X^{mus} and X^{cast} allele (Log2 normalised allelic reads) of X-linked genes X-inactivated Day2 ($Xa^{cast}Xi^{mus}$) vs. X-reactivated iPSC ($Xa^{cast}Xa^{mus}$) cells. (C) UMAP clusters of different stages of reprogramming of female MEF to iPSC. (D) Identification of X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) cells through profiling fraction Xa^{cast} allele expression at different stages of reprogramming. Cells with fraction Xa^{cast} allele expression between 0.8 to 1 are categorized as X-inactivated and the rest as X-reactivated.

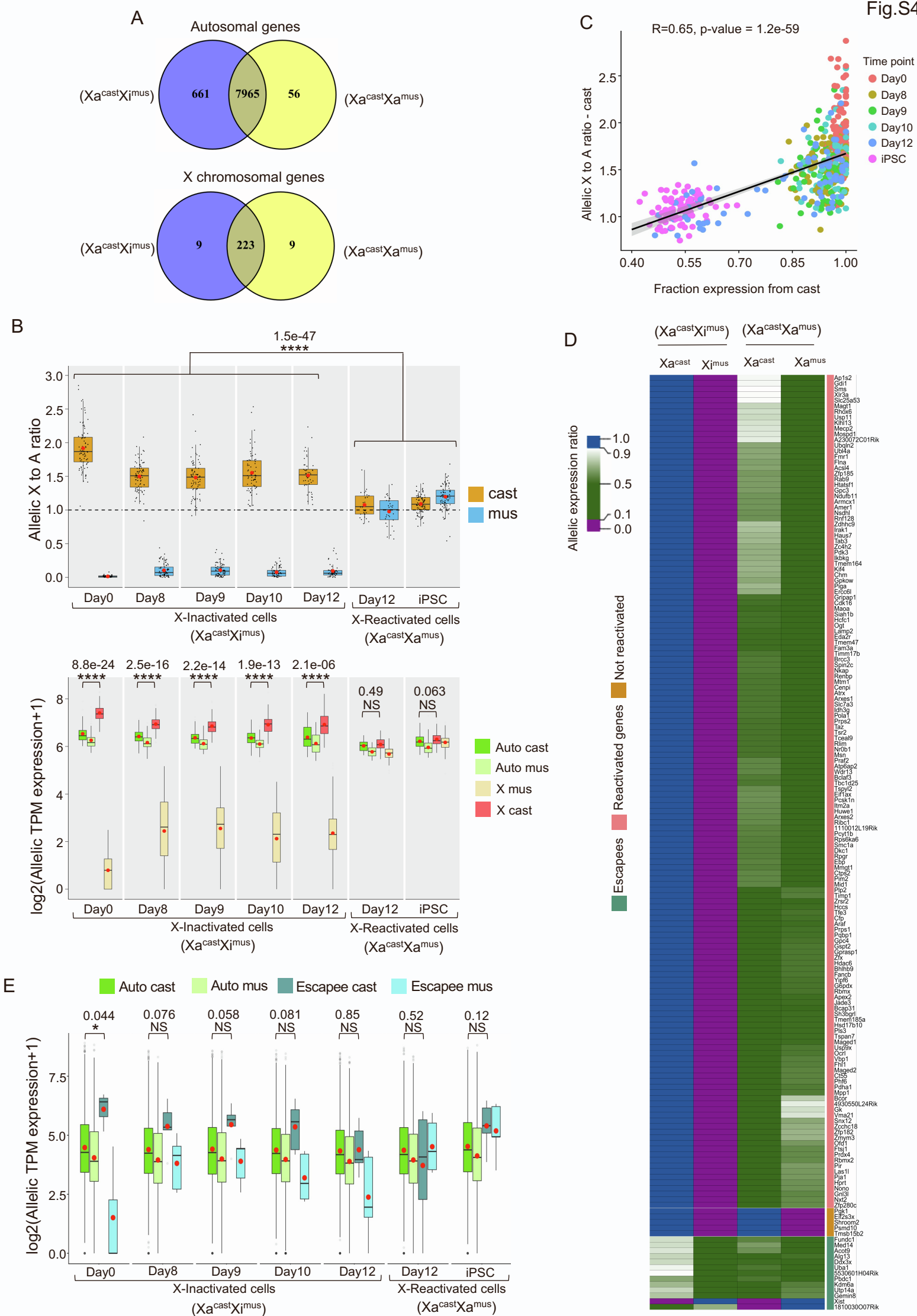
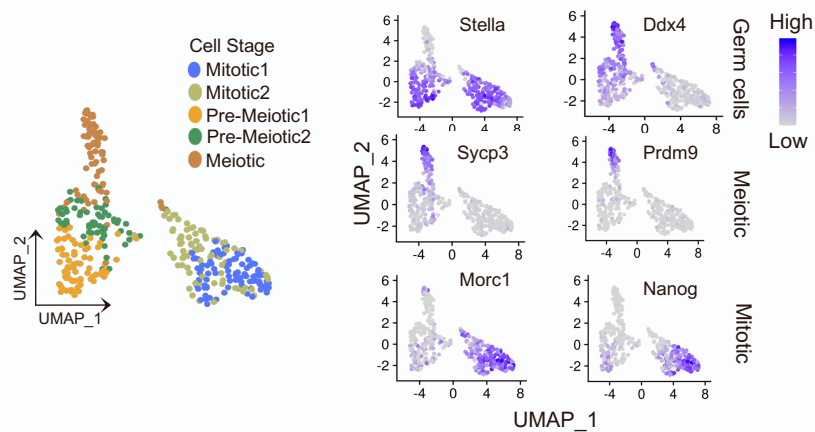
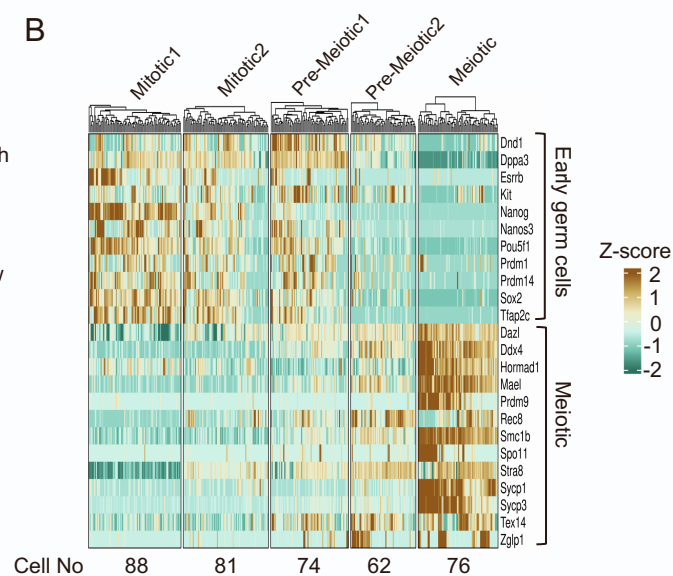


Fig S4 (related to Figure 2). (A) Venn diagram representing the common set of genes (X and autosomal genes) among X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) cells during iPSC reprogramming. (B) Comparison of allelic X:A ratio (top) and allelic expression of X and autosomes (bottom) between X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) cells using a common set of genes. (C) Scatter plots showing the correlation between Xa^{cast} upregulation ($Xa^{cast}:A^{cast}$) vs. reactivation of the Xi^{mus} (fraction expression from X^{cast} allele) in different stages of iPSC reprogramming (labelled with different colours). R is Pearson's correlation. (D) Heatmap representing the allelic expression ratio of X-linked genes in $Xa^{cast}Xi^{mus}$ vs. $Xa^{cast}Xa^{mus}$ cells of iPSC reprogramming. Genes with fraction Xi^{mus} allele expression in $Xa^{cast}Xi^{mus}$ cells < 0.10 are considered X-inactivated genes and > 0.10 as escapee genes. On the other hand, genes with a fraction of Xa^{mus} allele expression in $Xa^{cast}Xa^{mus}$ cells > 0.10 are considered reactivated genes. (E) Allelic expression (\log_2 allelic TPM+1) of escapee genes and autosomal genes in X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) cells during iPSC reprogramming. In boxplots, the line inside each of the boxes denotes the median value, the red circle denotes the mean and the edges of each box represent 25% and 75% of the dataset, respectively. (Wilcoxon rank-sum test: P -value < 0.0001 ; ****; P -value < 0.05 ; *, NS: not significant).

A



B



C

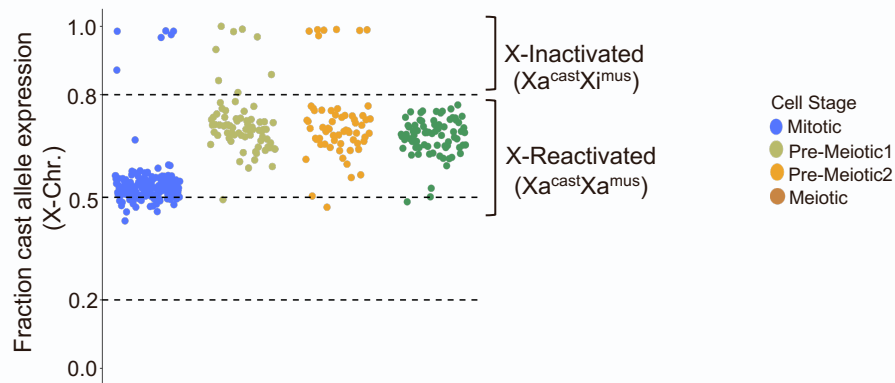


Fig S5 (related to Figure 3). (A) Plots representing UMAP-based clustering and projection of marker gene expression on UMAP plot for mitotic and meiotic germ cells. (B) Heatmap based on gene expression dynamics representing different stages of germ cell maturation: mitotic 1, mitotic 2, early meiotic (pre-meiotic 1 and 2) and late meiotic germ cells. (C) Identification of X-inactivated and X-reactivated cells through profiling fraction Xa^{cast} allele expression. Cells with fraction Xa^{cast} allele expression between 0.8 to 1 are categorized as X-inactivated and the rest as X-reactivated.

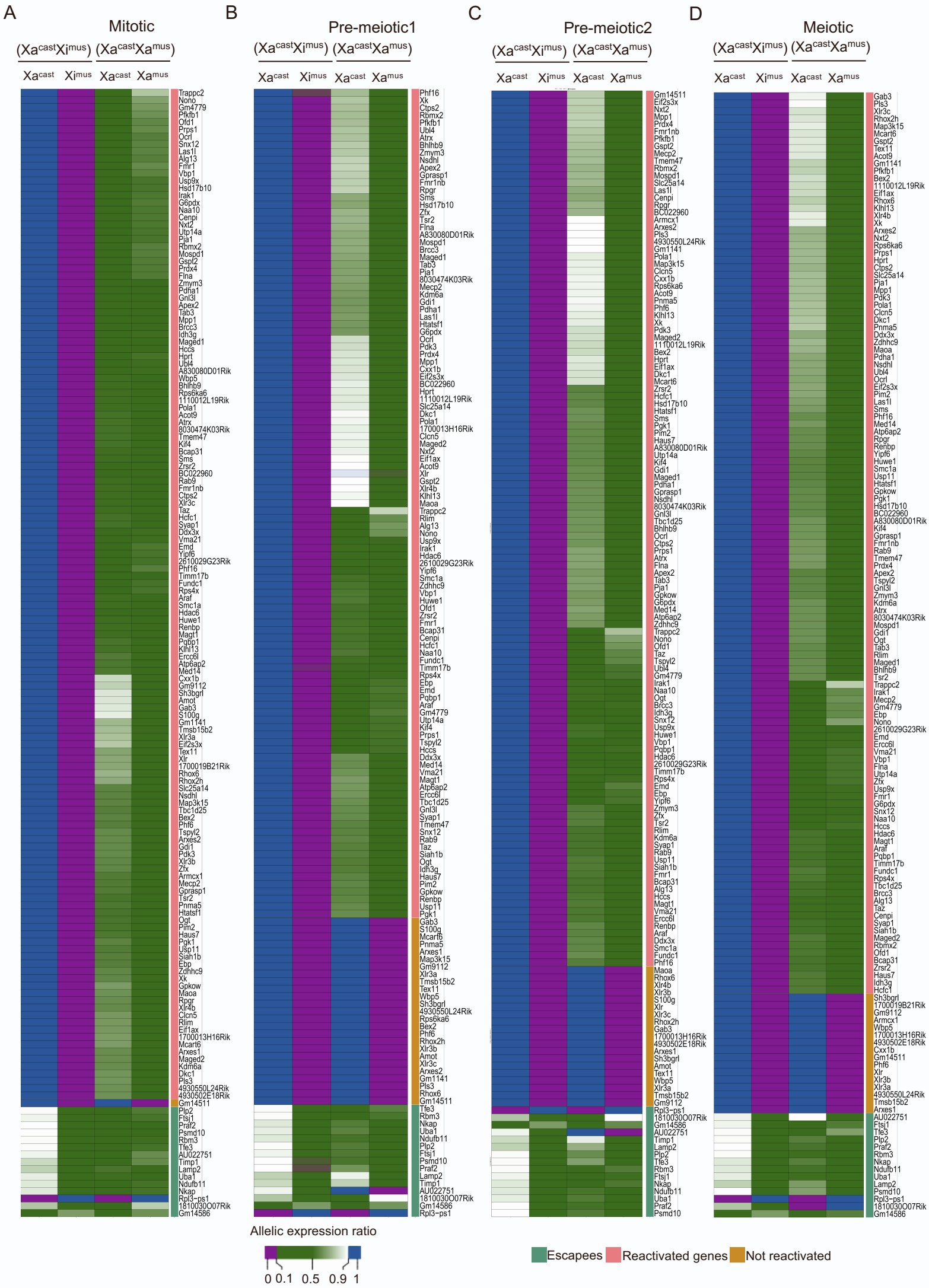


Fig S6 (related to Figure 3). Heatmap representing the allelic expression ratio of X-linked genes in $Xa^{cast}Xi^{mus}$ vs. $Xa^{cast}Xa^{mus}$ cells of (A) mitotic (B) Pre-meiotic1 (C) Pre-meiotic2 and (D) meiotic cells. Genes with fraction Xi^{mus} allele expression in $Xa^{cast}Xi^{mus}$ cells < 0.10 are considered X-inactivated genes and >0.10 as escapee genes. On the other hand, genes with a fraction of Xa^{mus} allele expression in $Xa^{cast}Xa^{mus}$ cells >0.10 are considered reactivated genes.

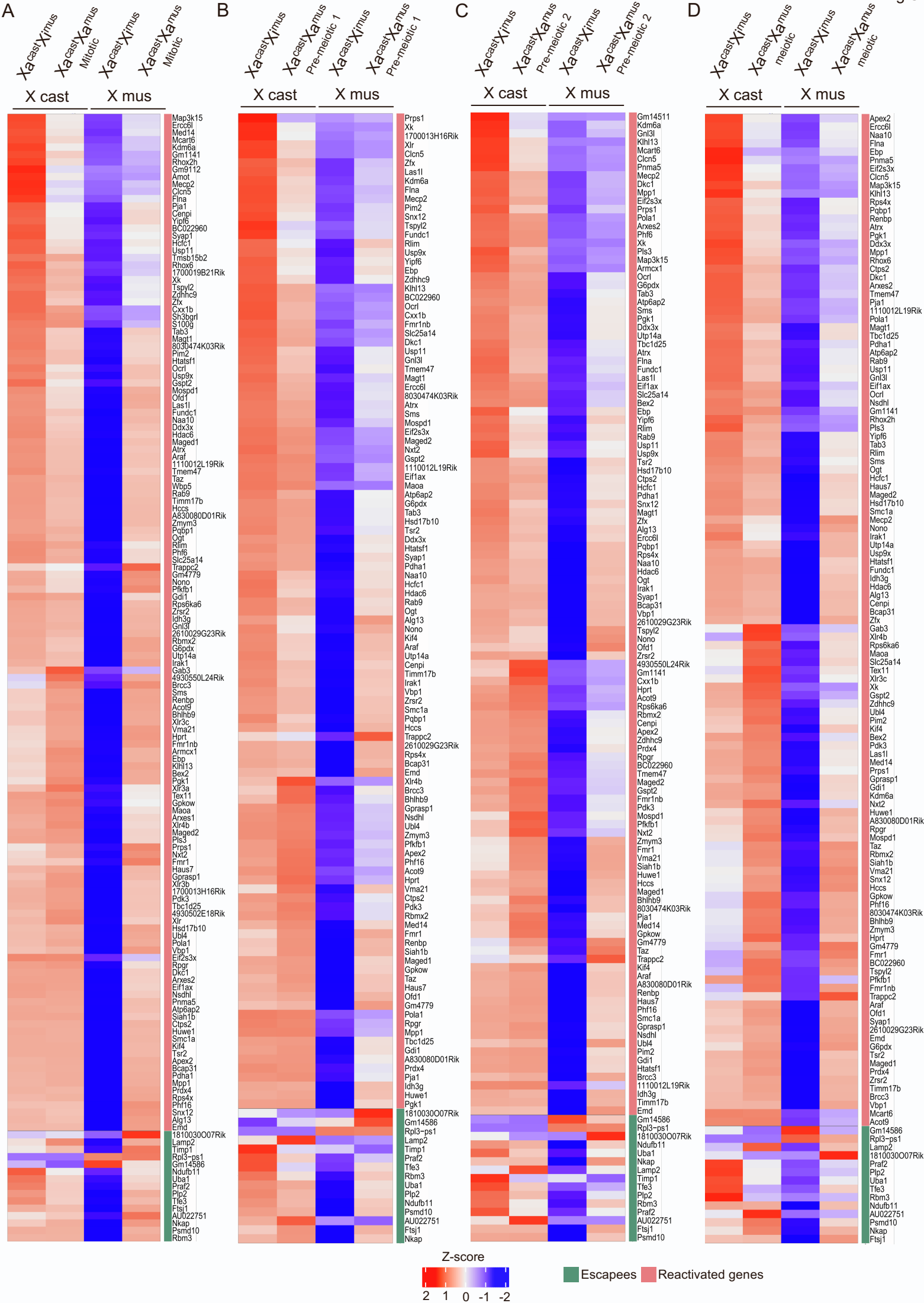


Fig S7 (related to Figure 3). Heatmap representing the allelic expression of X-linked genes from X^{cast} and X^{mus} allele in $X^{\text{cast}}X^{\text{mus}}$ vs. $X^{\text{cast}}X^{\text{mus}}$ cells of (A) mitotic (B) Pre-meiotic1 (C) Pre-meiotic2 and (D) meiotic cells.

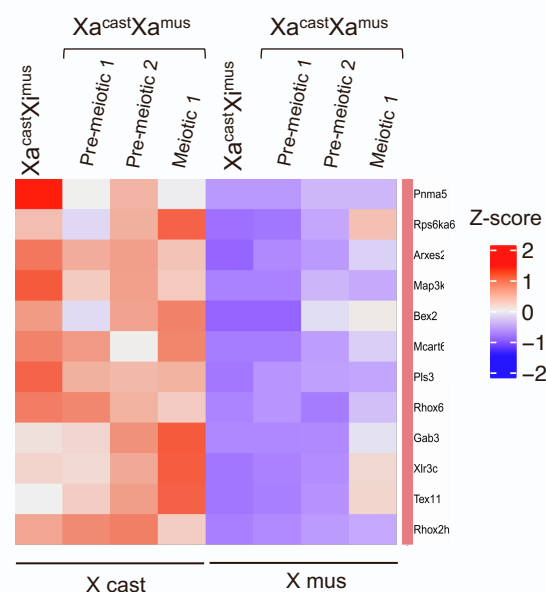
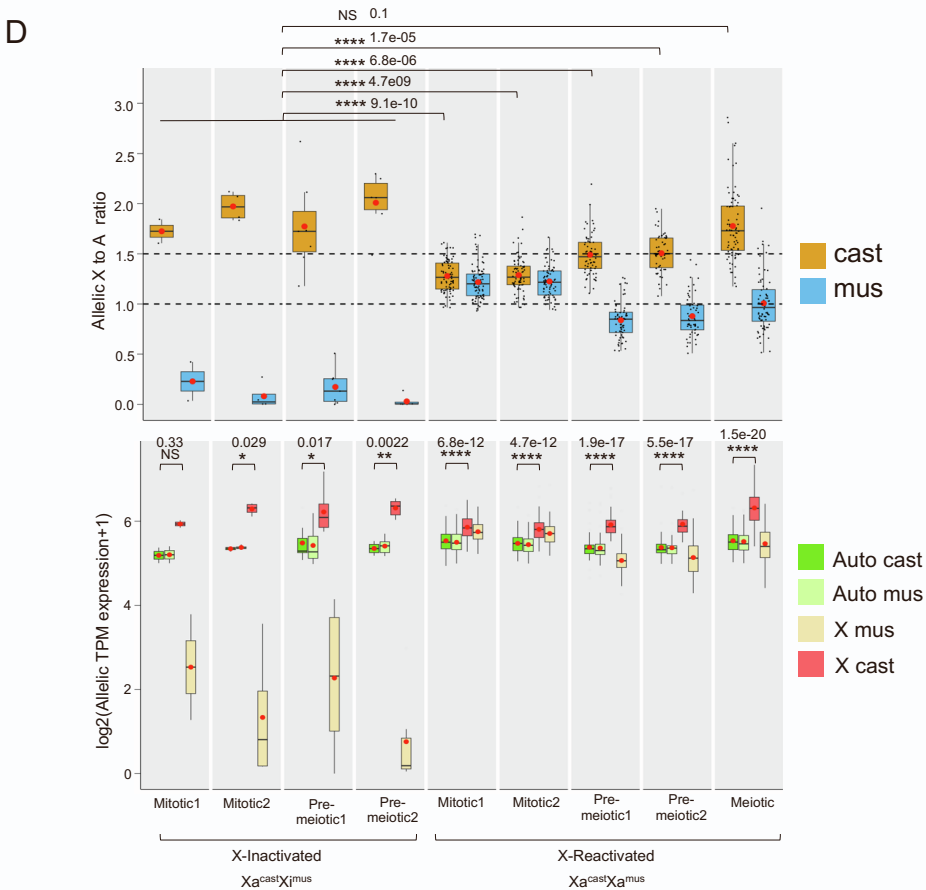
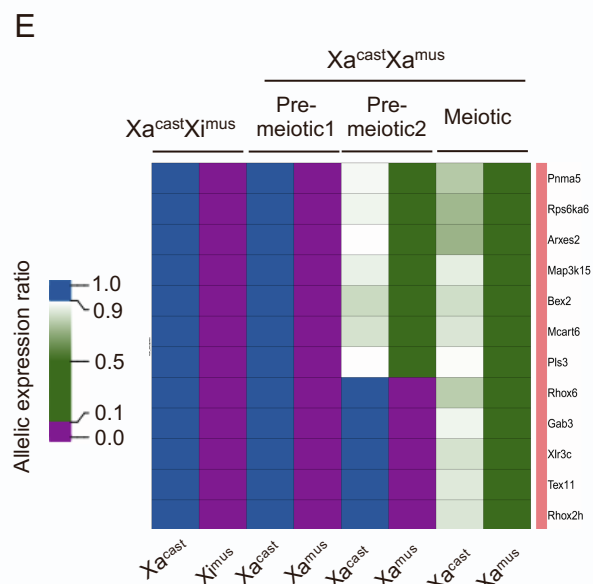
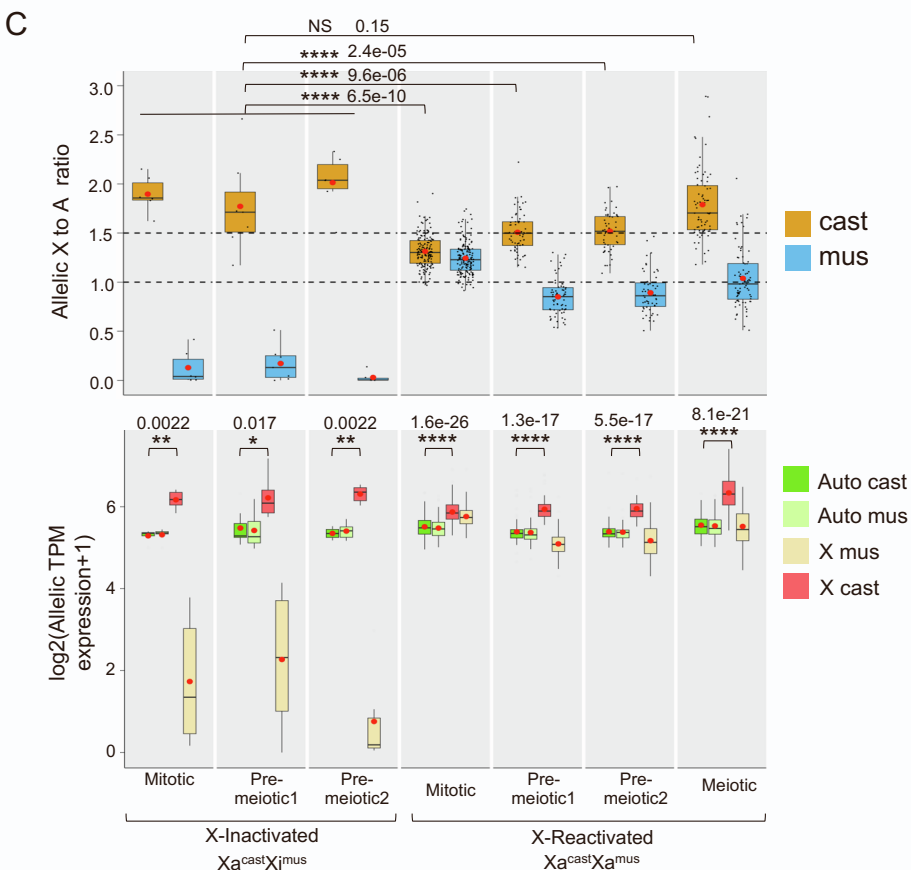
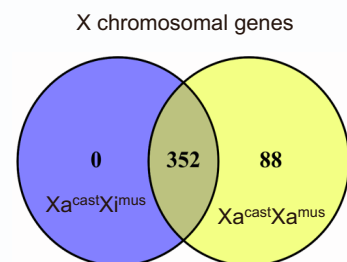
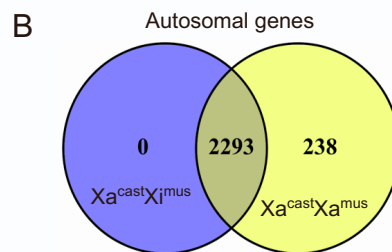
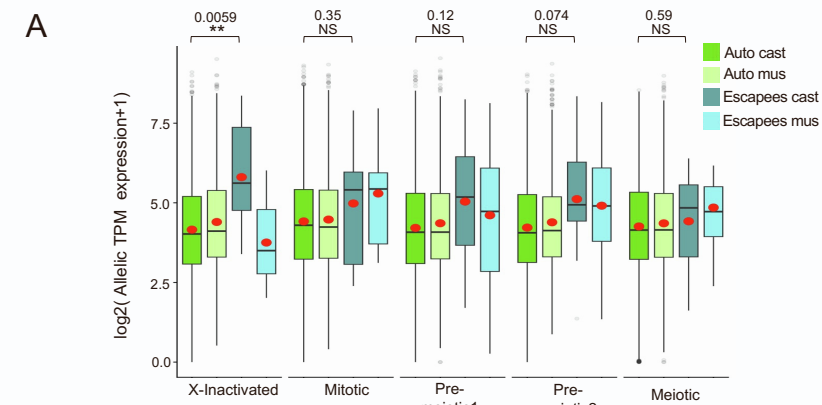
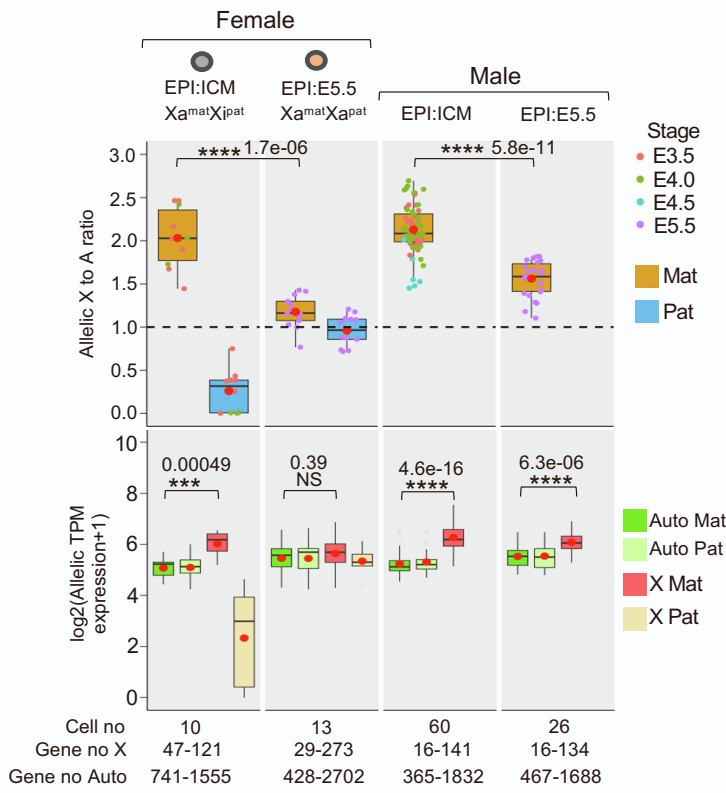
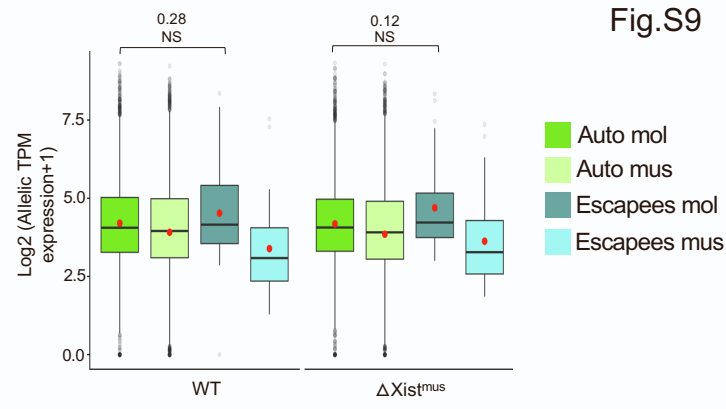


Fig S8 (related to Figure 3). (A) Allelic expression (\log_2 allelic TPM+1) of escapee genes and autosomal genes in X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) mitotic, Pre-meiotic1, Pre-meiotic2 and meiotic cells. (B) Venn diagram represents the common genes (X and autosomal genes) among $Xa^{cast}Xi^{mus}$ and $Xa^{cast}Xa^{mus}$ germ cells. (C) Comparison of allelic X:A ratio (top) and allelic expression of X and autosomes (bottom) between X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) cells using a common set of genes. In boxplots, the line inside each of the boxes denotes the median value, the red circle denotes the mean, and the edges of each box represent 25% and 75% of the dataset, respectively. (Wilcoxon rank-sum test: P -values < 0.0001 ; ****, < 0.001 ; ***, < 0.01 ; **, < 0.05 ; *). (D) Comparison of allelic X:A ratio (top) and allelic expression of X and autosomes (bottom) between X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) mitotic, Pre-meiotic1, Pre-meiotic2 and meiotic cells. In boxplots, the line inside each of the boxes denotes the median value, the red circle denotes the mean and the edges of each box represent 25% and 75% of the dataset, respectively (Wilcoxon rank-sum test: P -values < 0.0001 ; ****, < 0.01 ; **, < 0.05 ; * NS: not significant). (E) Top: heatmap representing the allelic expression ratio of X-linked genes in X-inactivated ($Xa^{cast}Xi^{mus}$) and X-reactivated ($Xa^{cast}Xa^{mus}$) pre-meiotic1, pre-meiotic2 and meiotic cells. Genes with fraction Xi^{mus} allele expression in $Xa^{cast}Xi^{mus}$ cells < 0.10 are considered X-inactivated genes and genes with a fraction of Xa^{mus} allele expression in $Xa^{cast}Xa^{mus}$ cells > 0.10 are considered reactivated genes. Bottom: Heatmap representing the allelic expression of X-linked genes from X^{cast} and X^{mus} allele in $Xa^{cast}Xi^{mus}$ vs. $Xa^{cast}Xa^{mus}$ cells of pre-meiotic1, pre-meiotic2 and meiotic cells.

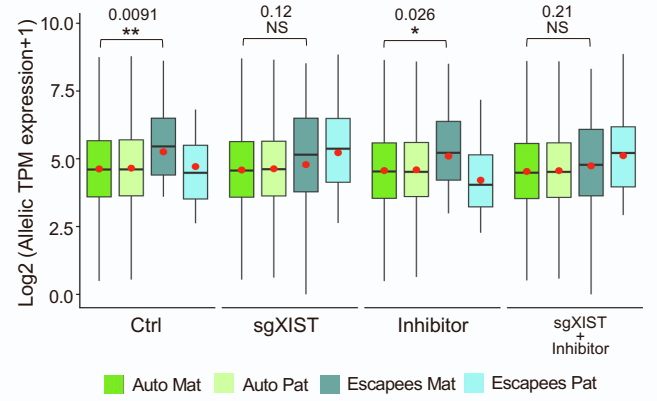
A



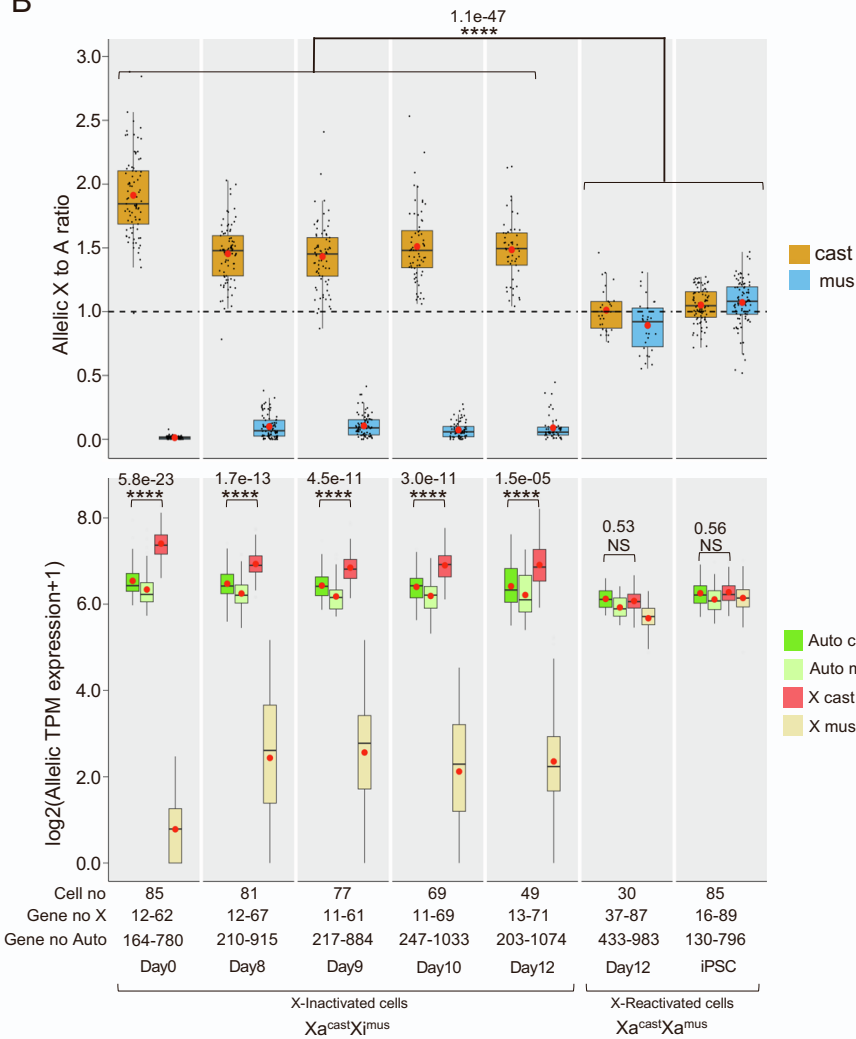
C



D



B



E

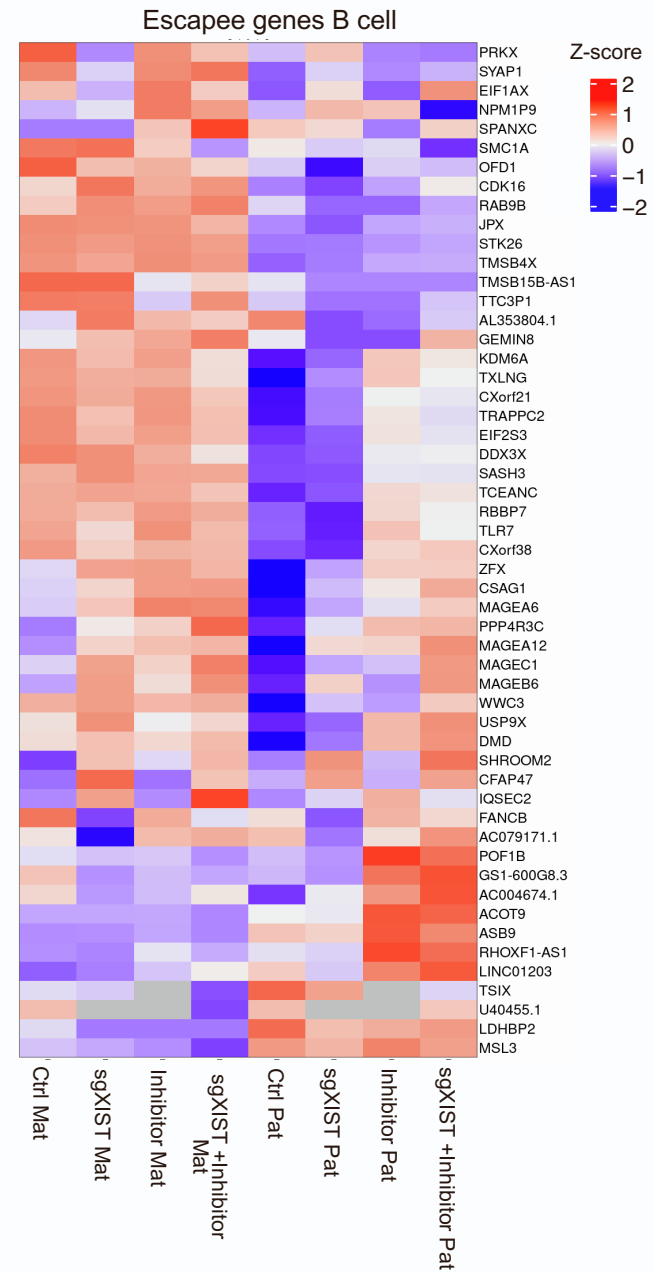


Fig S9 (related to Figure 1, 2, 3, 4 and 5). (A) Comparison of allelic X:A ratio (top) and allelic expression of X and autosomes (bottom) between EPI:ICM ($X^{mat}X^{i\text{pat}}$) and EPI:E5.5 ($X^{mat}X^{a\text{pat}}$) cells. A comparison between EPI:ICM and EPI:E5.5 of male cells is also plotted. (B) Top: Comparison of allelic X:A ratio between X-inactivated vs. X-reactivated cells during iPSC reprogramming. Bottom: Allelic expression (\log_2 allelic TPM+1) of X-linked and autosomal genes in X-inactivated vs. X-reactivated cells during iPSC reprogramming. Autosomes used for these plots: Chr13, Chr9, Chr8, Chr7 and Chr5. In all boxplots, the line inside each of the boxes denotes the median value, the red circle denotes the mean and the edges of each box represent 25% and 75% of the dataset, respectively (Wilcoxon rank-sum test: P -values < 0.0001; ****, < 0.001; *** NS: not significant). (C) Allelic expression (\log_2 allelic TPM+1) of escapee genes and autosomal genes in WT vs. $\Delta Xist^{mus}$ XEN cells. (D) Allelic expression (\log_2 allelic TPM+1) of escapee genes and autosomal genes in Ctrl and sgXIST, inhibitor and sgXIST + inhibitor treated B-cells. In boxplots, the line inside each of the boxes denotes the median value, the red circle denotes the mean, and the edges of each box represent 25% and 75% of the dataset, respectively (Wilcoxon rank-sum test: P -values < 0.01; ** and < 0.05; *). (E) Heatmap representing the allelic expression of escapee genes in Ctrl and sgXIST, inhibitor and sgXIST + inhibitor treated B-cells.

Supplemental Experimental procedure

Cell culture

XEN cells were cultured using media Dulbecco's modified eagle medium (DMEM) (Hi-media, #AL007A) supplemented with fetal bovine serum (FBS, Gibco #10270-106) L-glutamine (Gibco #25030081) Non-essential amino acids (NEAA, Gibco #11140050), penicillin-streptomycin (Gibco, # 15140122) 1mM of 2-Mercaptanol (Sigma #M6250). Cells were cultured on gelatin-coated plates and passaged through trypsinization.

RNA fluorescence in situ hybridization (RNA-FISH)

We generated double-stranded RNA-FISH probes as described previously (Gayen et al., 2015). In brief, probes were generated through random priming of BAC DNA using the Bioprime labeling kit (Invitrogen, #18094-011). Probes were labeled with Cy3-dUTP or Cy5-dUTP (Enzo Life Sciences) and purified through ProbeQuant G-50 Micro columns (Cytiva, #28903408). Probes were precipitated using 0.3M sodium acetate (Sigma, #71196), 300 µg of Yeast tRNA (Invitrogen, #15401011), 150 µg of sheared Salmon sperm DNA (Invitrogen, #15632-011) and absolute ethanol (Hayman, #F205220) at 13,000 rpm for 20 mins at 4°C. The pellet was washed with 70% and followed by 100% ethanol. After washing, probes were dried and resuspended in deionized formamide (VWR Life Sciences, #0606), followed by denatured at 95°C. Finally, probes were preserved at -20°C in a hybridization solution containing 20% Dextran sulfate (SRL, #76203), 2X SSC (SRL, #12590).

For RNA-FISH, XEN cells were seeded on the coverslip and grown to ~60-70% confluency and permeabilized with ice-cold cytoskeleton buffer (CSK, 100 mM NaCl, 300 mM sucrose, 3 mM MgCl₂, and 10 mM PIPES buffer [pH 6.8]) with 0.4% triton-X (SRL #30190). Next, cells were fixed through 3% paraformaldehyde solution (PFA Electron Microscopy Sciences #15710) for 10 min and followed by washed three times with ice-cold 70% ethanol. Next, cells were dehydrated through an ethanol series of 70%, 85%, 95% and 100% and subsequently air-dried. Cells were then hybridized with double-stranded probes for overnight at 37°C in a humid chamber. The cell samples were then washed 3× with pre-warmed 2X SSC/50% Formamide,

2X SSC, and 2 times with 1X SSC for 7 mins each at 37°C. DAPI (Invitrogen, #D1306) was added during the third 2X SSC wash. The coverslips were finally mounted using Vectashield (Vector Labs, #H1000) and visualized under the microscope.

RNA-sequencing analysis

Transcriptomic sequencing reads were mapped against mouse genome GRCm38 (mm10), and human genome GRCh37 (hg19) using STAR (2.7.9a) (Dobin et al., 2013) with default parameter and aligned reads were counted using HTSeq-count (2.0.2) (Anders et al., 2015). The expression level of transcripts was calculated using TPM (Transcripts Per Kilobase Million) counts.

Single-cell clustering and lineage identification

Seurat R package (Butler et al., 2018; Stuart et al., 2019) was used for single-cell clustering and lineage identification. In brief, highly variable features (HVG) were identified using “FindVariableFeatures” and cells were clustered using “FindClusters”. For visualization dimension reduction, UMAP was performed using the function “RunUMAP”. Clusters were annotated using a subset of lineage-specific marker gene expression. The following parameters were used for clustering; pre-implantation: HVG=1000, dims=1:35, iPSCs reprogramming: HVG=3000, dims=1:15 and germ cells: HVG=2000, dims=1:40.

X to A ratio

Considering the huge difference in the number of X-linked and autosomal genes, we calculated allelic X:A ratio using bootstrapping procedures as described previously (Naik et al., 2022; Pacini et al., 2021). In brief, we calculated the allelic X:A ratio by dividing the allelic expression of X-linked genes with the allelic expression of the same number of autosomal genes selected randomly for each cell/sample. This was repeated 1000 times and the median of 1000 values was considered. The same procedure was followed for the non-allelic X:A ratio estimation. To exclude low-expressed genes from our analysis, we used genes having >10 TPM for bulk RNA-seq data and >1 TPM for scRNA-seq data. We also excluded highly expressed genes as well using a 98-percentile threshold.

Sexing of the embryo

Sexing of the available single-cell dataset used for this study was performed if the sex was not mentioned previously. Cells were assigned as male based on Y chromosomal gene expression (Zfy2, Zfy1, Kdm5d, Uty, Usp9y, Ddx3y, Eif2s3y, Ube1y1).

Simulation

We have considered the two X-chromosomes as interacting entities, and they are modeled as differential equations given by:

$$\frac{dX_i}{dt} = g_1 \underbrace{f(K_1, X_a, n)}_{\text{cross}} \underbrace{f(K_3, X_i, n)}_{\text{self}} - \underbrace{k_1 X_i}_{\text{decay}} \quad (1)$$

$$\frac{dX_a}{dt} = g_2 \underbrace{f(K_2, X_i, n)}_{\text{cross}} \underbrace{f(K_4, X_a, n)}_{\text{self}} - \underbrace{k_2 X_a}_{\text{decay}} \quad (2)$$

Where,

$$f(K, X, n) = \begin{cases} \frac{X^n}{K^n + X^n} & \text{if activatory} \\ \frac{K^n}{K^n + X^n} & \text{if inhibitory} \\ 1 & \text{if no effect} \end{cases} \quad (3)$$

X_i is the expression level of the inactive X given as X:A ratio, X_a is the expression level of the active X given as X:A ratio, g_1 and g_2 are the production rates, k_1 and k_2 are the decay rates, n is the hill coefficient, K_1 , K_2 , K_3 and K_4 are the half-saturation constants.

For the data for full reactivation, we took the data obtained from X-reactivation in iPSC given in Fig. 2. Due to lack of data between day 0 to day 8, the starting point for X-reactivation was considered to be day 7 and with the same level corresponding to day 0 in the iPSC data. The mean value for day wise levels of X_i and X_a were considered. The levels for iPSC cells were considered as the levels from day 13 to day 15. For the data for partial reactivation, we assumed a hypothetical case of iPSC reactivation stalling at day 12. The value at day 12 was extrapolated up to day 15. These values match qualitatively with the partial reactivation state in Fig. 4 and have been done due to the lack of temporal data.

These equations are fit to the time course data for full and partial reactivation. This was done by minimizing the sum of square error using the differential evolution algorithm of scipy. The

initial population of parameters is sampled using Sobol sampling. The differential equations are solved using the explicit Runge-Kutta method of order 5(4) with these parameters. Then, the sum of square errors between the solutions evaluated at the given time points and the actual data is calculated. A new parameter set is generated by adding a weighted difference between two randomly chosen parameter sets to a third parameter set, similar to a mutation. Then, it randomly combines parameters from the old set with this new set, similar to crossover. The sum of square errors with this new set of parameters is also evaluated and compared with those of the old parameters. If the values are lower with the new set, they replace the old set in the next generation of the population. This is repeated multiple times until an optimal solution is found (Storn and Price, 1997).

For the model with noise, we added a noise term $\eta(t)$ to Equation 1 and Equation 2 where the values are sampled from a normal distribution with mean 0 and standard deviation 1. These were then solved with the fit parameters using explicit Runge-Kutta method of order 5(4).

Supplementary table legends

Table S1 (related to Fig. 1): Allelic X:A ratio in embryos.

Table S2 (related to Fig. 2): Allelic X:A ratio in cells of different stages of iPSC reprogramming.

Table S3 (related to Fig. 3): Allelic X:A ratio in germ-cells.

References:

Anders, S., Pyl, P.T., and 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>.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and 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>.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.

Gayen, S., Maclary, E., Buttigieg, E., Hinten, M., and Kalantry, S. (2015). A Primary Role for the Tsix lncRNA in Maintaining Random X-Chromosome Inactivation. *Cell Rep* 11. <https://doi.org/10.1016/j.celrep.2015.04.039>.

Naik, H.C., Hari, K., Chandel, D., Jolly, M.K., and Gayen, S. (2022). Single-cell analysis reveals X upregulation is not global in pre-gastrulation embryos. *IScience* 25. <https://doi.org/10.1016/j.isci.2022.104465>.

Pacini, G., Dunkel, I., Mages, N., Mutzel, V., Timmermann, B., Marsico, A., and Schulz, E.G. (2021). Integrated analysis of Xist upregulation and X-chromosome inactivation with single-cell and single-allele resolution. *Nat Commun* 12. <https://doi.org/10.1038/s41467-021-23643-6>.

Storn, R., and Price, K. (1997). Differential Evolution-A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* 11, 341–359. .

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