Cell Reports, Volume *35*

Supplemental information

Contiguous erosion of the inactive X in human pluripotency concludes with global DNA hypomethylation

Prakhar Bansal, Darcy T. Ahern, Yuvabharath Kondaveeti, Catherine W. Qiu, and Stefan F. Pinter

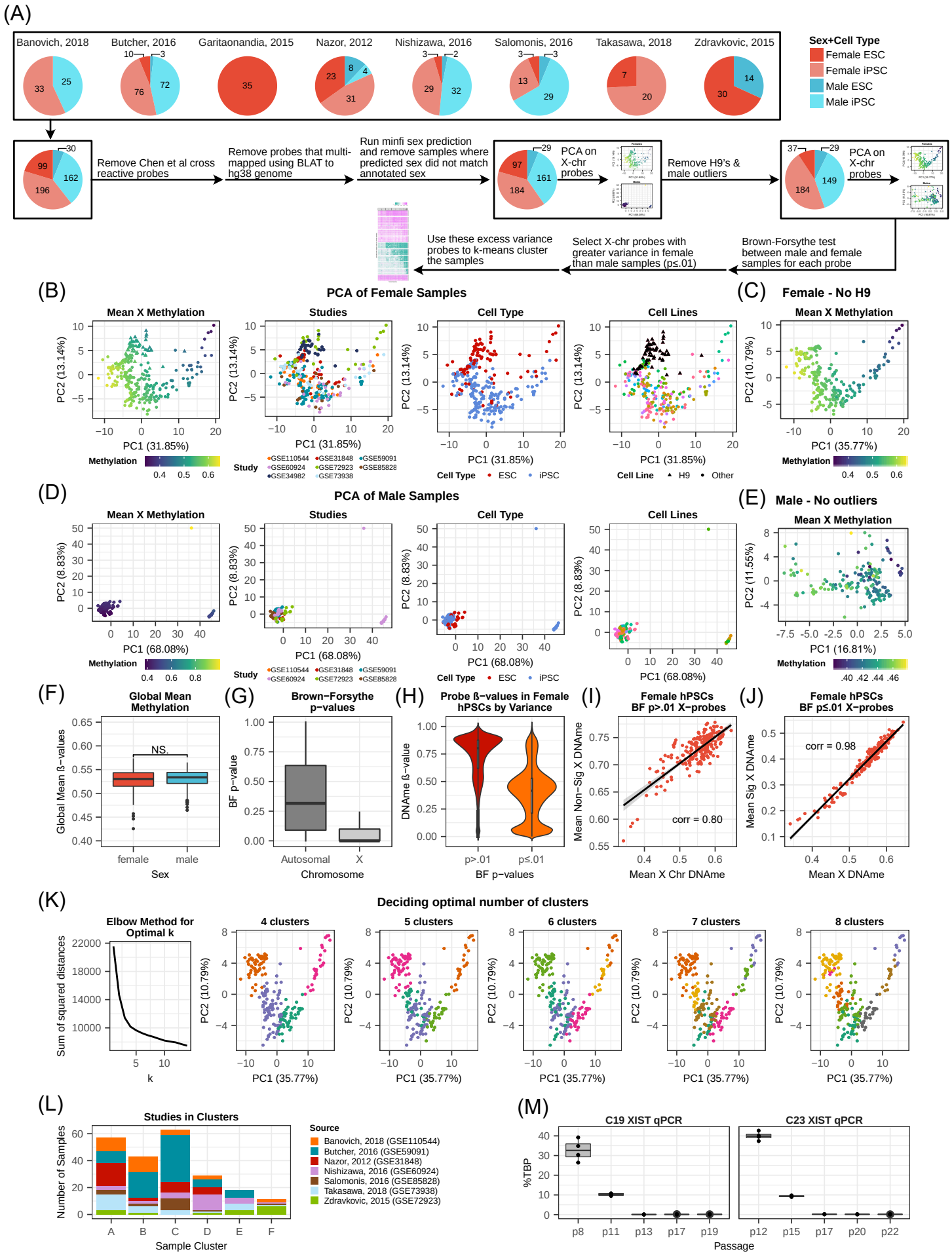
(A)
Banovich, 2018
Butcher, 2016
Garitaonandia, 2015
Nazor, 2012
Nishizawa, 2016
Salomonis, 2016
Takasawa, 2018
Zdravkovic, 2015
Sex+Cell Type
Female ESC
Female iPSC
Male ESC
Male iPSC
Remove Chen et al cross reactive probes
Remove probes that multi-mapped using BLAT to hg38 genome
Run minfi sex prediction and remove samples where predicted sex did not match annotated sex
PCA on X-chr probes
Remove H9's & male outliers
PCA on X-chr probes
Brown-Forsythe test between male and female samples for each probe
Select X-chr probes with greater variance in female than male samples (p≤.01)
Use these excess variance probes to k-means cluster the samples
(B)
PCA of Female Samples
Mean X Methylation
Studies
Cell Type
Cell Lines
PC1 (31.85%)
PC2 (13.14%)
Methylation
Study
Cell Type ESC iPSC
Cell Line H9 Other
(C)
Female - No H9
Mean X Methylation
PC1 (35.77%)
PC2 (10.79%)
(D)
PCA of Male Samples
PC1 (68.08%)
PC2 (8.83%)
(E)
Male - No outliers
PC1 (16.81%)
PC2 (11.55%)
(F)
Global Mean Methylation
Global Mean ß−values
NS.
female
male
Sex
(G)
Brown−Forsythe p−values
BF p−value
Autosomal
X
Chromosome
(H)
Probe ß−values in Female hPSCs by Variance
DNAme ß−value
p>.01
p≤.01
BF p−values
(I)
Female hPSCs BF p>.01 X−probes
Mean Non−Sig X DNAme
corr = 0.80
Mean X Chr DNAme
(J)
Female hPSCs BF p≤.01 X−probes
Mean Sig X DNAme
corr = 0.98
Mean X DNAme
(K)
Deciding optimal number of clusters
Elbow Method for Optimal k
Sum of squared distances
k
4 clusters
5 clusters
6 clusters
7 clusters
8 clusters
PC1 (35.77%)
PC2 (10.79%)
(L)
Studies in Clusters
Number of Samples
Sample Cluster
Source
Banovich, 2018 (GSE110544)
Butcher, 2016 (GSE59091)
Nazor, 2012 (GSE31848)
Nishizawa, 2016 (GSE60924)
Salomonis, 2016 (GSE85828)
Takasawa, 2018 (GSE73938)
Zdravkovic, 2015 (GSE72923)
(M)
C19 XIST qPCR
C23 XIST qPCR
%TBP
Passage
p8
p11
p13
p17
p19
p12
p15
p17
p20
p22

Fig S1. Data processing and outlier identification. Related to Figure 1 and methods about "Data Collection, Processing, and Normalization".

(A) Schematic of filtering and processing pipeline for DNA methylation data.

(B) and (C) PCAs of female samples on all X-probes before and after H9 removal, respectively. Plots are colored by different properties of the samples as indicated by the legends.

(D) and (E) PCAs of male samples on all X-probes before and after outlier removal, respectively. The plots are colored by different properties of the samples as indicated by the legends.

(F) Global mean methylation for female and male samples (N.S. = not significant).

(G) Brown-Forsythe test p-values for autosomal probes and X-probes showing that X-probes are significantly more variant between the male and female samples than autosomal probes.

(H) Distribution of β-values for X-probes that are more variant in female samples ($p \leq .01$), and those that are not ($p > .01$).

(I) and (J) Relationship between the mean X chromosome methylation and mean methylation of X-probes above or below the BF-test p-value cutoff (.01), respectively. The line shows the linear regression, and "corr" is the Pearson correlation.

(K) Plots of the elbow method to find the appropriate number of clusters for the k-means clustering of female samples.

(L) Breakdown of sample clusters by study of origin.

(M) XIST qPCR for C19 and C23 iPSCs. Each passage has one sample with 4 technical replicates each.

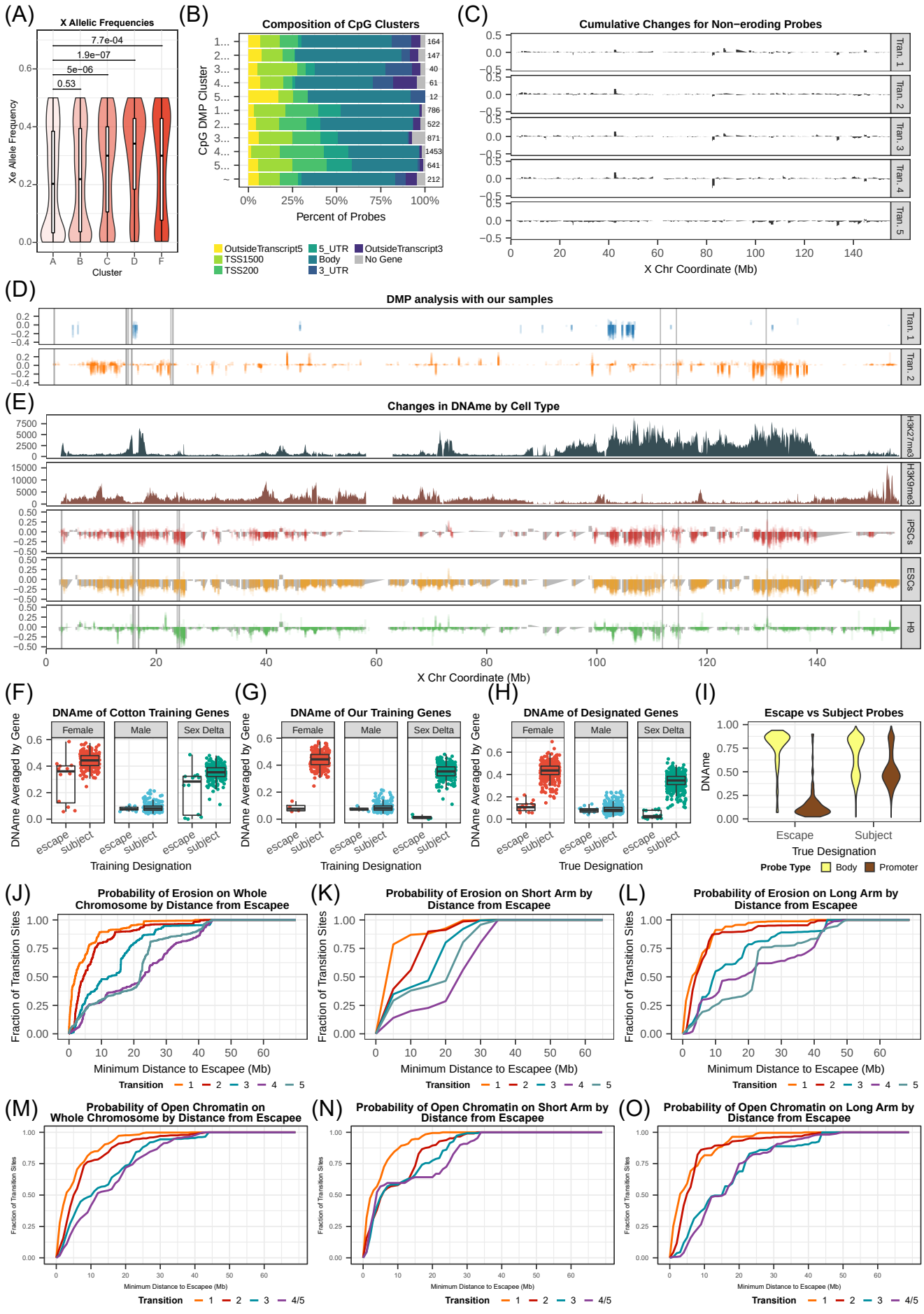
(A)
X Allelic Frequencies
7.7e-04
1.9e-07
5e-06
0.53
Xe Allele Frequency
Cluster
(B)
Composition of CpG Clusters
CpG DMP Cluster
Percent of Probes
OutsideTranscript5
TSS1500
TSS200
5_UTR
Body
3_UTR
OutsideTranscript3
No Gene
(C)
Cumulative Changes for Non−eroding Probes
Tran. 1
Tran. 2
Tran. 3
Tran. 4
Tran. 5
X Chr Coordinate (Mb)
(D)
DMP analysis with our samples
(E)
Changes in DNAme by Cell Type
H3K27me3
H3K9me3
iPSCs
ESCs
H9
X Chr Coordinate (Mb)
(F)
DNAme of Cotton Training Genes
Female
Male
Sex Delta
DNAme Averaged by Gene
escape
subject
Training Designation
(G)
DNAme of Our Training Genes
(H)
DNAme of Designated Genes
True Designation
(I)
Escape vs Subject Probes
DNAme
Escape
Subject
Probe Type
Body
Promoter
(J)
Probability of Erosion on Whole Chromosome by Distance from Escapee
Fraction of Transition Sites
Minimum Distance to Escapee (Mb)
Transition
(K)
Probability of Erosion on Short Arm by Distance from Escapee
(L)
Probability of Erosion on Long Arm by Distance from Escapee
(M)
Probability of Open Chromatin on Whole Chromosome by Distance from Escapee
(N)
Probability of Open Chromatin on Short Arm by Distance from Escapee
(O)
Probability of Open Chromatin on Long Arm by Distance from Escapee

Fig S2. DNAme dynamics during XIE distinguish escapees and reactivated genes. Related to Figure 3 and methods about "Annotating Genes as Escapees" and "Calculating distances to escapee genes."

(A) Distribution of allele frequencies of the Xe as inferred by the lesser allele fraction for each RNA-seq hPSC sample across heterozygous positions (Wilcoxon test p-values relative to cluster A).

(B) Percent composition of DMPs in each probe group from Figure 2A, as described in the methods. Probe totals for each group are listed to the right.

(C) Cumulative changes through each transition for X-linked probes with high sex-variance but without consistent direction across transitions (" ~ " symbol), plotted as in Fig. 3A.

(D) Changes in DNAme based on DMP analysis with C19 and C23 samples only. Changes for transition specific probes are shown (similar to Fig 3A) in relation to cluster A C19 and C23 samples. Correlation to meta-analysis transition-specific DMPs is shown in top-right.

(E) Changes in DNAme (similar to Fig 3A) separated by whether the sample is an iPSC, ESC, or H9. All comparisons are to cluster A samples.

(F-H) Identification of escapees using the Cotton et al method. "Sex Delta" panels show the difference in methylation for a gene between male and female samples. (E) methylation values of the initial Cotton, et al. trainee genes, (F) as G after removal of initial trainee genes that fail cutoffs (as in Cotton et al), and (G) final escapee gene set after true designation is determined and the escapees are identified.

(I) Distribution of β-values of gene body and promoter probes for escapees and genes subject to XCI.

(J) Cumulative probability of CpG demethylation in each transition as a function of distance to escapee genes on the whole X chromosome.

(K) as in J, for CpGs on Xp.

(L) as in J, for CpGs on Xq.

(M) Cumulative probability of ATAC-peak openings in each transition as a function of distance to escapee genes on the whole X chromosome.

(N) as in M, for ATAC-peaks on Xp.

(O) as in M, for ATAC-peaks on Xq.

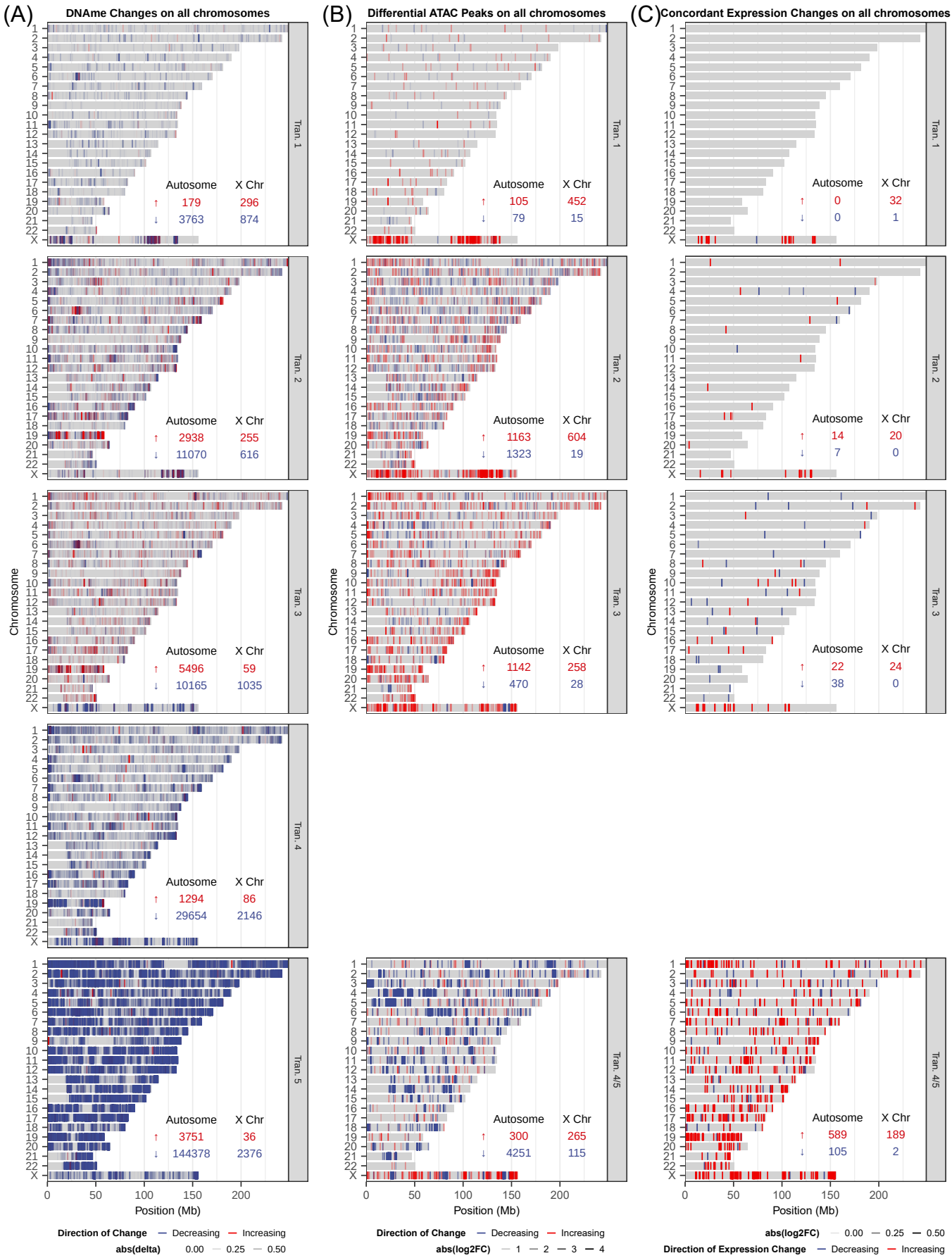
(A)
DNAme Changes on all chromosomes
(B)
Differential ATAC Peaks on all chromosomes
(C)
Concordant Expression Changes on all chromosomes
Chromosome
Tran. 1
Tran. 2
Tran. 3
Tran. 4
Tran. 5
Tran. 4/5
Autosome
X Chr
↑ 179 296
↓ 3763 874
↑ 105 452
↓ 79 15
↑ 0 32
↓ 0 1
↑ 2938 255
↓ 11070 616
↑ 1163 604
↓ 1323 19
↑ 14 20
↓ 7 0
↑ 5496 59
↓ 10165 1035
↑ 1142 258
↓ 470 28
↑ 22 24
↓ 38 0
↑ 1294 86
↓ 29654 2146
↑ 3751 36
↓ 144378 2376
↑ 300 265
↓ 4251 115
↑ 589 189
↓ 105 2
0 50 100 150 200
Position (Mb)
Direction of Change — Decreasing — Increasing
abs(delta) 0.00 0.25 0.50
Direction of Change — Decreasing — Increasing
abs(log2FC) 1 2 3 4
abs(log2FC) 0.00 0.25 0.50
Direction of Expression Change — Decreasing — Increasing

Fig S3. Global changes in DNAme and gene expression. Related to Figure 4 and methods about "Concordance and Enrichment Analysis."

(A) Chromosome-resolved DMPs assigned to specific transitions by p-value. Total counts for autosomal and X chromosome probes are provided, with direction of change indicated (blue = loss of DNAme, and red = gain of DNAme). The transparency is on a continuous scale based on the β-differential.

(B) Chromosome-resolved differential ATAC peaks assigned to specific transitions by FDR. Total counts are provided as for (A) (blue = closing ATAC peak, red = opening ATAC peak). The transparency is on a continuous scale based on the log2(fold change) of the ATAC peaks.

(C) Chromosome-resolved differentially expressed genes concordant with either DMP (A) or differential ATAC peaks (B). Total counts are provided as for (A) (blue = decreasing expression, and red = increasing expression). The transparency is on a continuous scale based on the log2(fold change) of the expression.

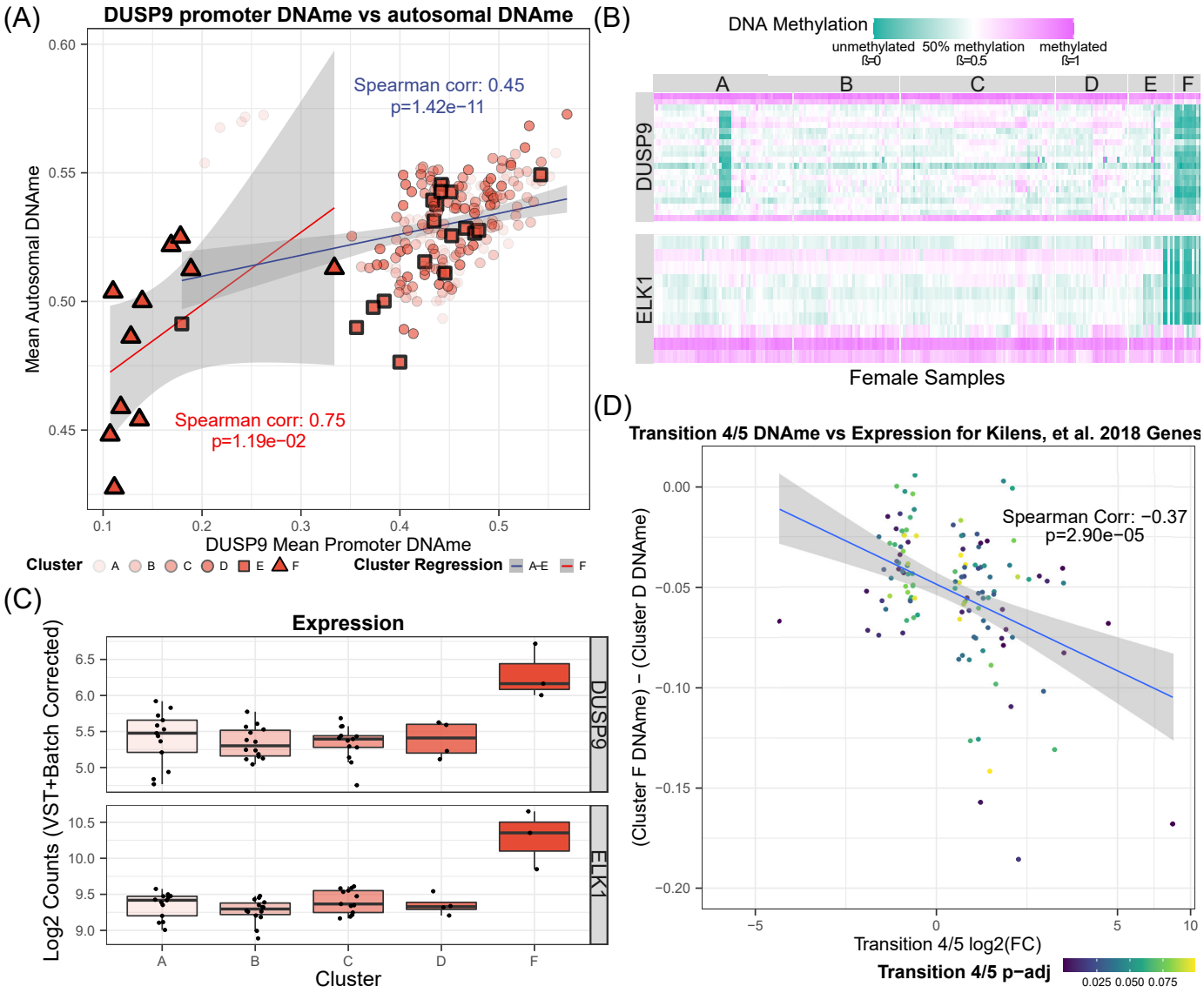

Fig S4. Autosomal DNAme changes relative to DUSP9 reactivation impact naïve markers. Related to Figure 4.

(A) Relationship of DUSP9 promoter methylation to autosomal methylation in female samples. Clusters E and F are emphasized to highlight the change in global and DUSP9 methylation in the last transition.

(B) DNAme heatmap of DUSP9 and ELK1 probes across clusters A-F.

(C) DUSP9 and ELK1 (VST+batch-corrected/log2-transformed) expression across clusters A-F.

(D) Change in methylation for genes that are differentially expressed in Transition 4/5 and are differentially expressed in naïve stem cells (Kilens et al., 2018).