Figure S1

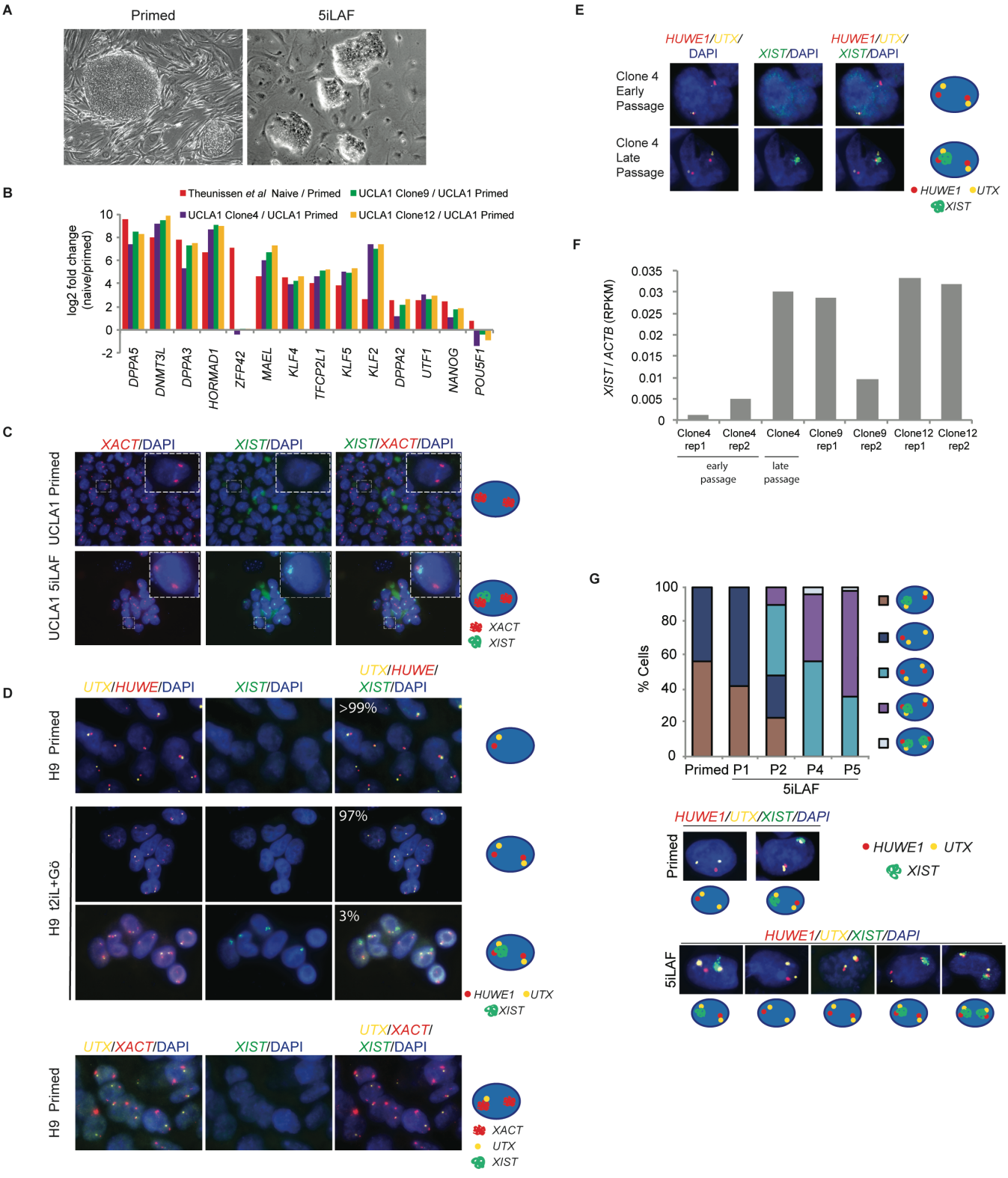

A
Primed
5iLAF
B
Theunissen et al Naive / Primed
UCLA1 Clone9 / UCLA1 Primed
UCLA1 Clone4 / UCLA1 Primed
UCLA1 Clone12 / UCLA1 Primed
log2 fold change (naive/primed)
DPPA5
DNMT3L
DPPA3
HORMAD1
ZFP42
MAEL
KLF4
TFCP2L1
KLF5
KLF2
DPPA2
UTF1
NANOG
POU5F1
C
XACT/DAPI
XIST/DAPI
XIST/XACT/DAPI
UCLA1 Primed
UCLA1 5iLAF
XACT
XIST
D
UTX/HUWE/DAPI
XIST/DAPI
UTX/HUWE/ XIST/DAPI
H9 Primed
H9 t2iL+Gö
>99%
97%
3%
HUWE1
UTX
XIST
UTX/XACT/DAPI
XIST/DAPI
UTX/XACT/ XIST/DAPI
H9 Primed
XACT
UTX
XIST
E
HUWE1/UTX/ DAPI
XIST/DAPI
HUWE1/UTX/ XIST/DAPI
Clone 4 Early Passage
Clone 4 Late Passage
HUWE1
UTX
XIST
F
XIST / ACTB (RPKM)
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
Clone4 rep1
Clone4 rep2
Clone4
Clone9 rep1
Clone9 rep2
Clone12 rep1
Clone12 rep2
early passage
late passage
G
% Cells
100
80
60
40
20
0
Primed
P1
P2
P4
P5
5iLAF
HUWE1/UTX/XIST/DAPI
Primed
HUWE1
UTX
XIST
HUWE1/UTX/XIST/DAPI
5iLAF

Figure S2

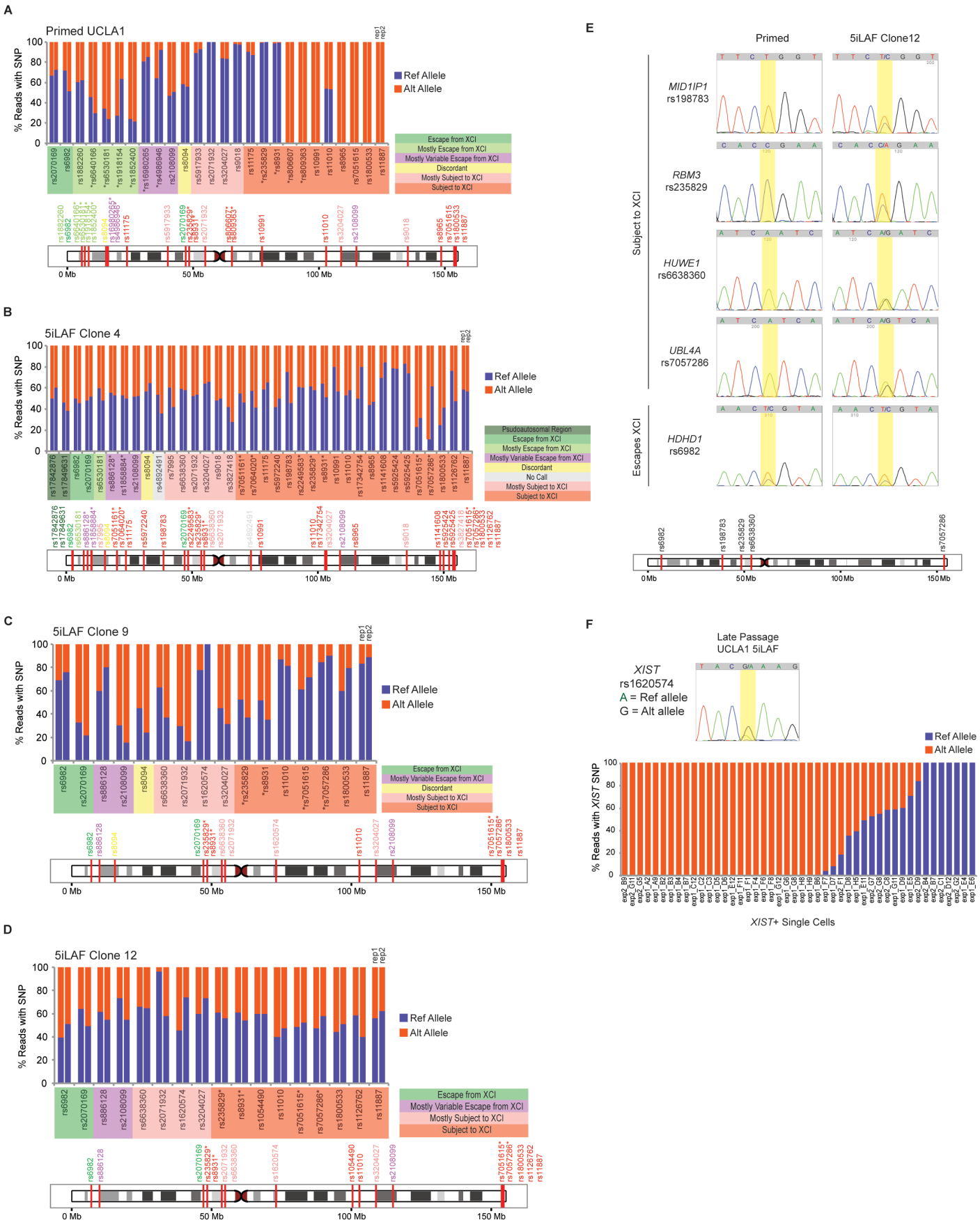

A
Primed UCLA1
% Reads with SNP
Ref Allele
Alt Allele
Escape from XCI
Mostly Escape from XCI
Mostly Variable Escape from XCI
Discordant
Mostly Subject to XCI
Subject to XCI
B
5iLAF Clone 4
Pseudoautosomal Region
No Call
C
5iLAF Clone 9
D
5iLAF Clone 12
0 Mb
50 Mb
100 Mb
150 Mb
E
Primed
5iLAF Clone12
Subject to XCI
Escapes XCI
MID1IP1
rs198783
RBM3
rs235829
HUWE1
rs6638360
UBL4A
rs7057286
HDHD1
rs6982
F
Late Passage
UCLA1 5iLAF
XIST
rs1620574
A = Ref allele
G = Alt allele
% Reads with XIST SNP
XIST+ Single Cells

Figure S3

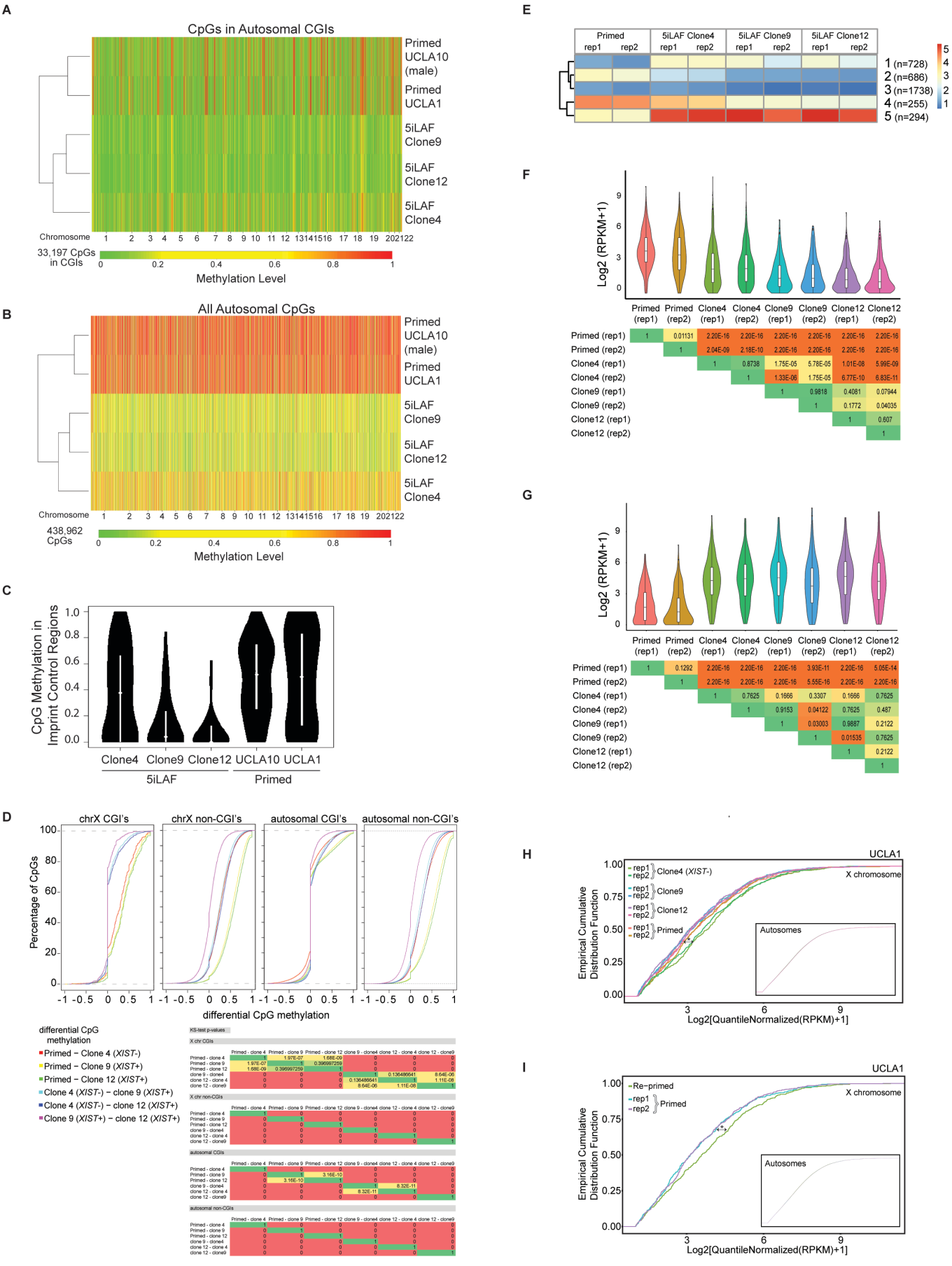
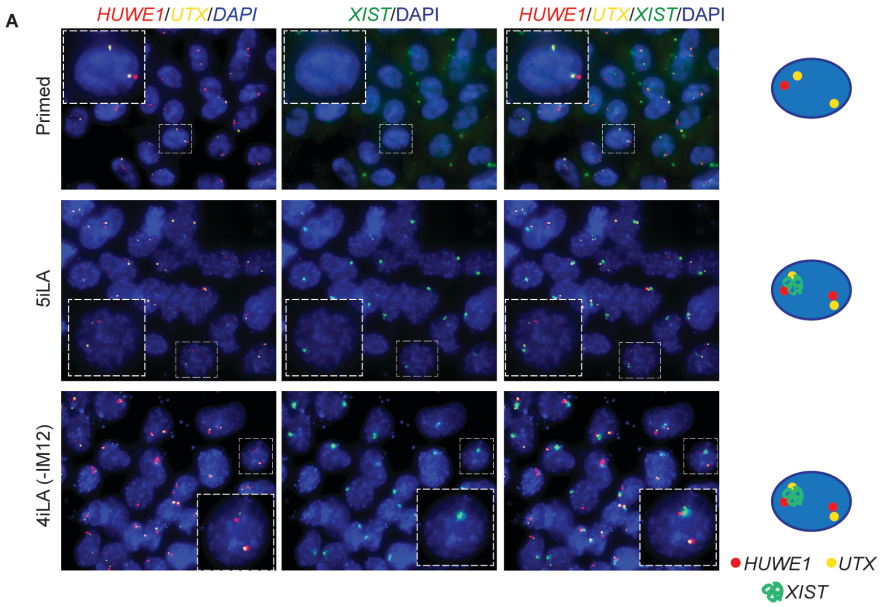
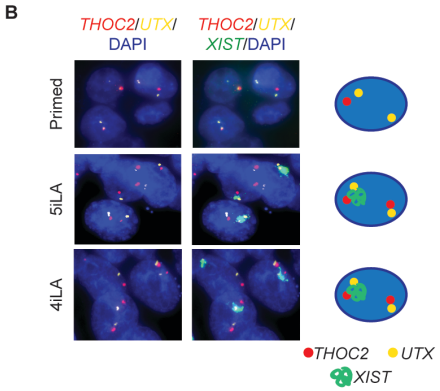

A
CpGs in Autosomal CGIs
Primed UCLA10 (male)
Primed UCLA1
5iLAF Clone9
5iLAF Clone12
5iLAF Clone4
Chromosome
33,197 CpGs in CGIs
Methylation Level
B
All Autosomal CpGs
438,962 CpGs
C
CpG Methylation in Imprint Control Regions
Clone4 Clone9 Clone12 UCLA10 UCLA1
5iLAF
Primed
D
chrX CGI's
chrX non-CGI's
autosomal CGI's
autosomal non-CGI's
Percentage of CpGs
differential CpG methylation
differential CpG methylation
Primed − Clone 4 (XIST-)
Primed − Clone 9 (XIST+)
Primed − Clone 12 (XIST+)
Clone 4 (XIST-) − clone 9 (XIST+)
Clone 4 (XIST-) − clone 12 (XIST+)
Clone 9 (XIST+) − clone 12 (XIST+)
KS-test p-values
X chr CGIs
X chr non-CGIs
autosomal CGIs
autosomal non-CGIs
E
Primed rep1 rep2
5iLAF Clone4 rep1 rep2
5iLAF Clone9 rep1 rep2
5iLAF Clone12 rep1 rep2
1 (n=728)
2 (n=686)
3 (n=1738)
4 (n=255)
5 (n=294)
F
Log2 (RPKM+1)
Primed (rep1)
Primed (rep2)
Clone4 (rep1)
Clone4 (rep2)
Clone9 (rep1)
Clone9 (rep2)
Clone12 (rep1)
Clone12 (rep2)
G
Log2 (RPKM+1)
H
UCLA1
Clone4 (XIST-)
Clone9
Clone12
Primed
X chromosome
Autosomes
Empirical Cumulative Distribution Function
Log2[QuantileNormalized(RPKM)+1]
I
UCLA1
Re-primed
Primed
X chromosome
Autosomes
Empirical Cumulative Distribution Function
Log2[QuantileNormalized(RPKM)+1]

Figure S4


A
HUWE1/UTX/DAPI
XIST/DAPI
HUWE1/UTX/XIST/DAPI
Primed
5iLA
4iLA (-IM12)
HUWE1
UTX
XIST


B
THOC2/UTX/ DAPI
THOC2/UTX/ XIST/DAPI
Primed
5iLA
4iLA
THOC2
UTX
XIST

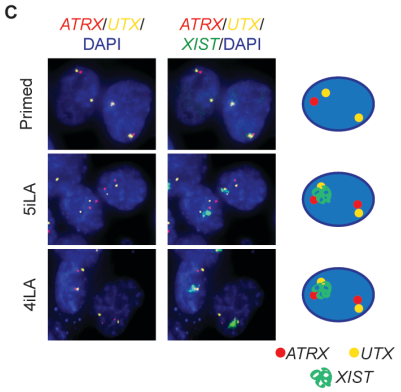

C
ATRX/UTX/ DAPI
ATRX/UTX/ XIST/DAPI
Primed
5iLA
4iLA
ATRX
UTX
XIST

Figure S5

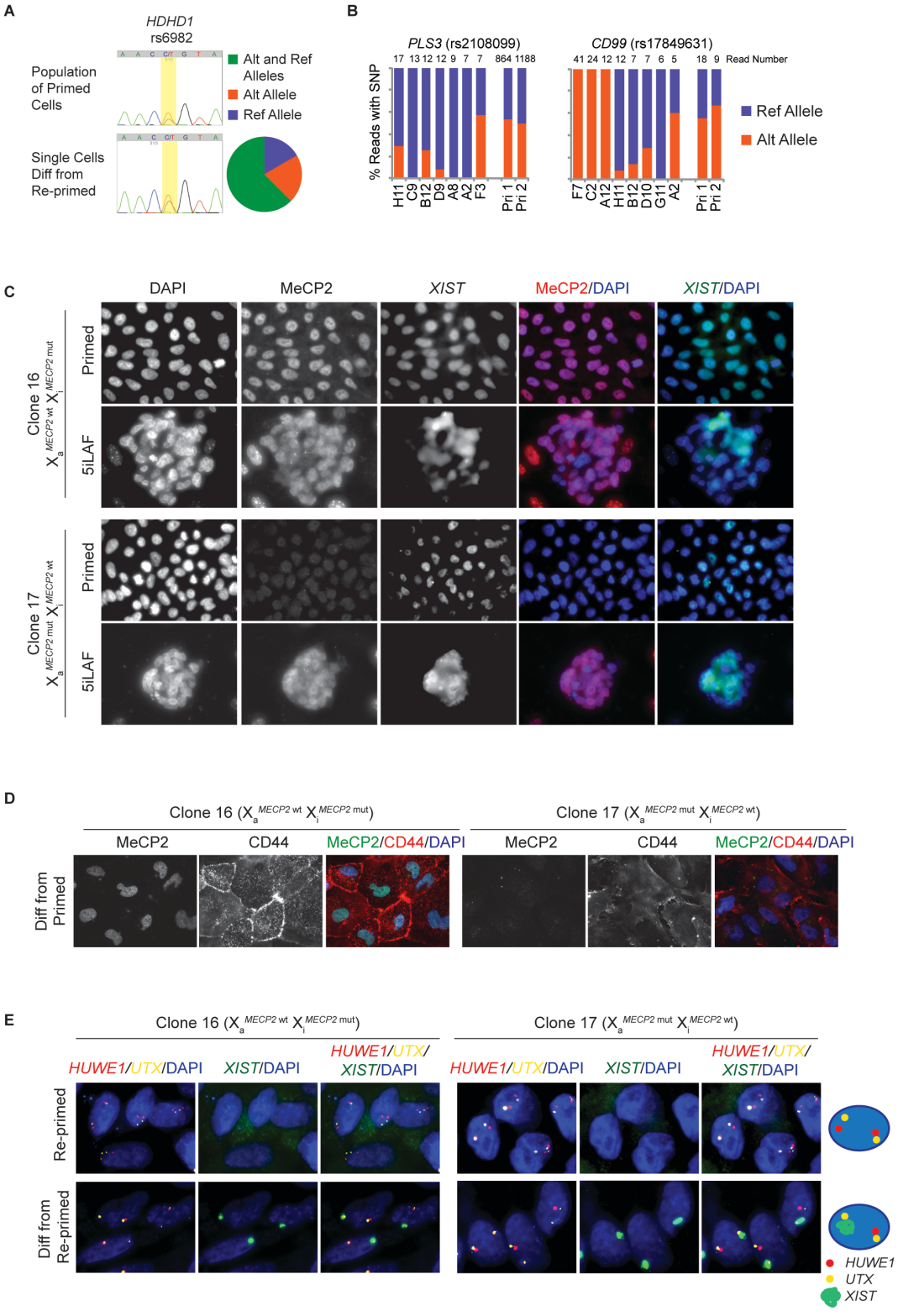

A
HDHD1
rs6982
Population of Primed Cells
Single Cells Diff from Re-primed
Alt and Ref Alleles
Alt Allele
Ref Allele
B
PLS3 (rs2108099)
CD99 (rs17849631)
Read Number
% Reads with SNP
Ref Allele
Alt Allele
C
DAPI
MeCP2
XIST
MeCP2/DAPI
XIST/DAPI
Clone 16
Clone 17
Primed
5iLAF
D
Clone 16 (Xa MECP2 wt Xi MECP2 mut)
Clone 17 (Xa MECP2 mut Xi MECP2 wt)
MeCP2
CD44
MeCP2/CD44/DAPI
Diff from Primed
E
HUWE1/UTX/DAPI
XIST/DAPI
HUWE1/UTX/ XIST/DAPI
Re-primed
Diff from Re-primed
HUWE1
UTX
XIST

Figure S6

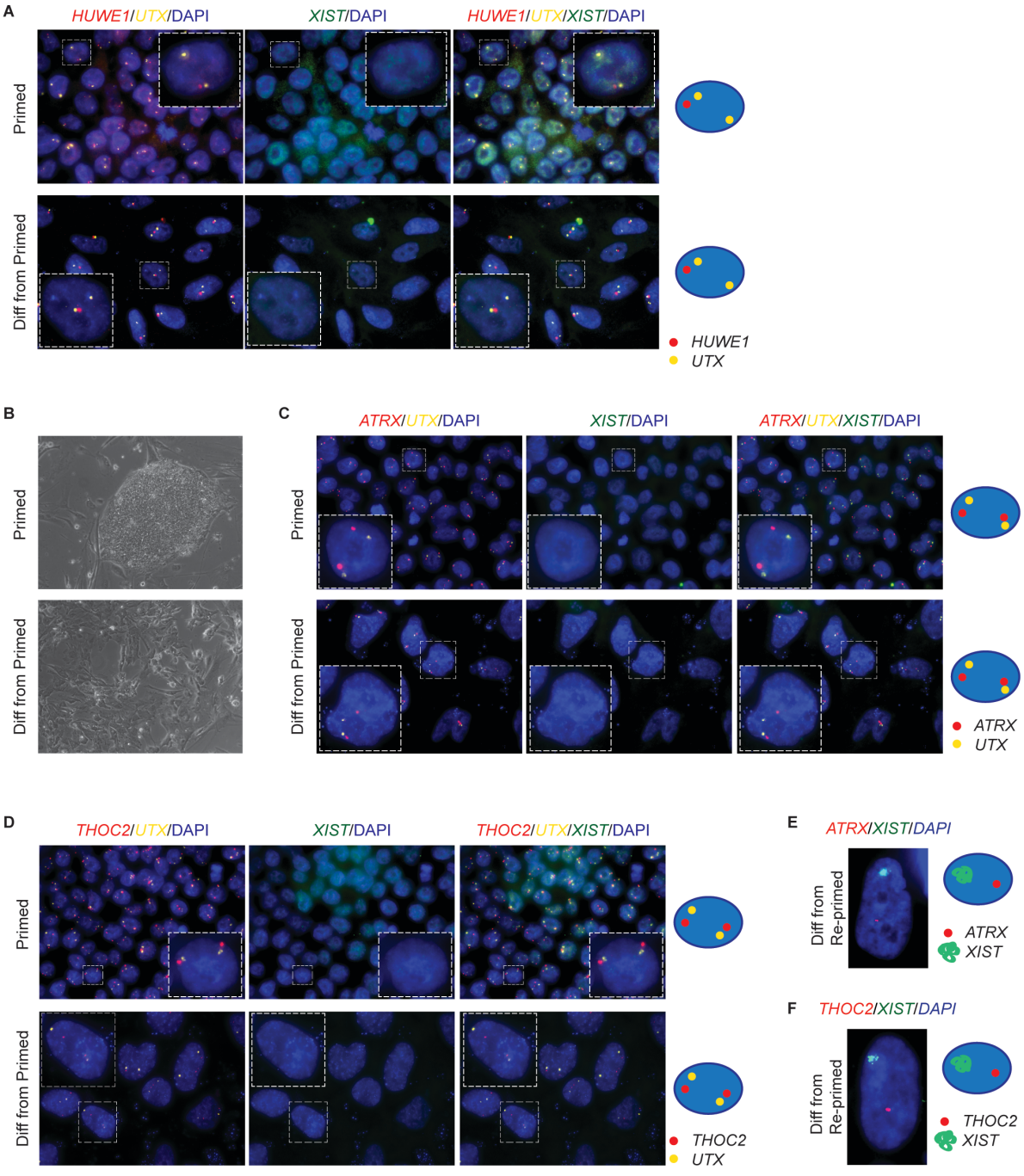
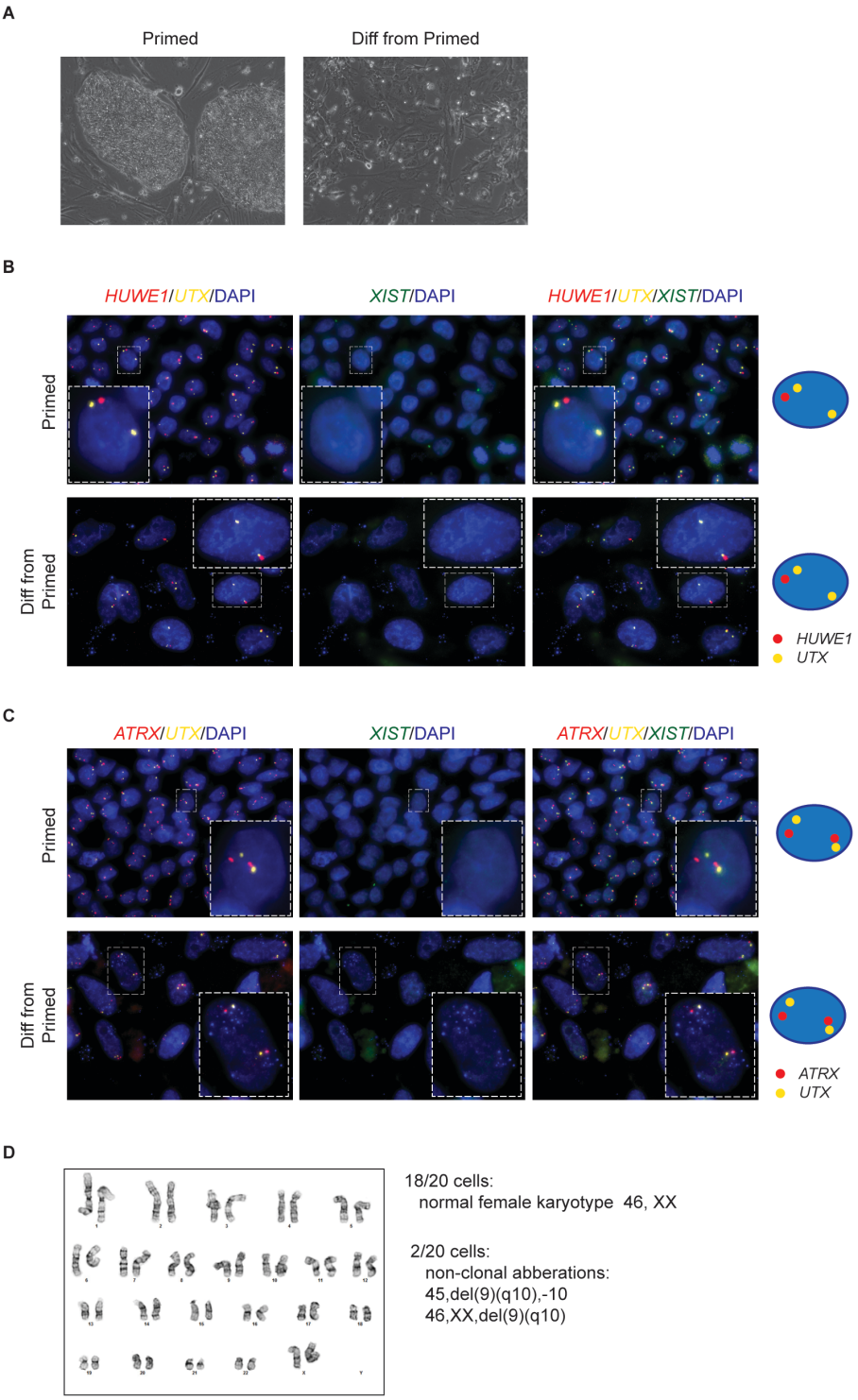

A
HUWE1/UTX/DAPI
XIST/DAPI
HUWE1/UTX/XIST/DAPI
Primed
Diff from Primed
HUWE1
UTX
B
Primed
Diff from Primed
C
ATRX/UTX/DAPI
XIST/DAPI
ATRX/UTX/XIST/DAPI
Primed
Diff from Primed
ATRX
UTX
D
THOC2/UTX/DAPI
XIST/DAPI
THOC2/UTX/XIST/DAPI
Primed
Diff from Primed
THOC2
UTX
E
ATRX/XIST/DAPI
Diff from Re-primed
ATRX
XIST
F
THOC2/XIST/DAPI
Diff from Re-primed
THOC2
XIST

Figure S7


A
Primed
Diff from Primed
B
HUWE1/UTX/DAPI
XIST/DAPI
HUWE1/UTX/XIST/DAPI
Primed
Diff from Primed
HUWE1
UTX
C
ATRX/UTX/DAPI
XIST/DAPI
ATRX/UTX/XIST/DAPI
Primed
Diff from Primed
ATRX
UTX
D
18/20 cells:
normal female karyotype 46, XX
2/20 cells:
non-clonal abberations:
45,del(9)(q10),-10
46,XX,del(9)(q10)

Supplemental Figure Legends

Figure S1. Characterization of naïve hESCs and the primed to naïve transition [Related to Figure 1]

(A) Phase-contrast images of primed and 5iLAF-adapted UCLA1 hESCs.

(B) Fold-change of transcript levels between the naïve UCLA1 clones 4 (early passage), 9, 12 and the starting primed UCLA1 population for genes known to be up-regulated in the naïve compared to the primed state (Theunissen et al., 2014). For comparison, averaged expression ratios for the same genes for originally described 5iLAF-cultured hESCs were plotted (obtained from Table S2 of Theunissen et al., 2014). *ZFP42* is an outlier in our data because primed UCLA1 expressed *ZFP42* more highly than the primed WIBR2 and WIBR3 hESC lines used by Theunissen et al., (2014), as it is among the top 3% of expressed genes in UCLA1 vs the bottom 2% in WIBR2 and WIBR3.

(C) Representative RNA FISH images detecting the lncRNAs *XIST* and *XACT* in primed and naïve UCLA1 cells at P15 and P24, respectively.

(D) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in primed (P62) and naïve t2iL+Gö (P14 in the naïve media) H9 hESCs, and for *XIST, XACT* and *UTX* in primed H9. The UTX, typically escaping XCI, displayed only mono-allelic pattern in the primed H9 cells, possibly due to it not escaping in this particular cell line at the primed state (see *UTX/HUWE1/XIST* stain in primed H9). However, the *XACT* RNA-FISH in primed H9 is expressed from two sites due to X_i-erosion, which demonstrates the presence of two X-chromosomes in these cells.

(E) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in the naive UCLA1 clone4 at early passage (P3, *XIST*-negative) and after it became predominantly *XIST*-positive (P12). The most prevalent RNA-FISH pattern in each condition is depicted by the cartoons on the right.

(F) RPKM values of *XIST* normalized to that of *ACTB* in primed UCLA1 and naïve clones 4, 9 and 12. Both early passage (replicates rep1 and rep2, P6) and late passage (P16) clone4 were included, demonstrating the transition of this clone from *XIST*-negative to the *XIST*-positive state, in agreement with data shown in (E).

(G) The early passage (P6) primed female hiPSC line #1001 (Karumbayaram et al., 2012) with about 55% of cells carrying an *XIST*-coated X_i, was converted to the naïve state with the 5iLAF approach. RNA FISH staining patterns detecting *XIST*, *HUWE1*, and *UTX* were quantified as cells progressed from primed to naïve pluripotency at the indicated passages. All counts were in cells with only two *UTX* foci. Representative RNA FISH images of single nuclei capturing the different RNA FISH patterns of X-chromosome in primed hESCs (top) and upon conversion to the naïve state (bottom) are shown, with corresponding cartoons.

Figure S2. The transition from primed to naïve pluripotency leads to X_i-reactivation [Related to Figure 2]

(A) Allelic expression proportions of X-linked genes in primed UCLA1 based on ten or more reads covering indicated SNPs in replicate RNA-seq data (rep1, rep2, closely-spaced bars). SNPs located in the same gene were placed next to each other and marked with an asterisk (*). Genes known to be subject to XCI, to escape from XCI, to be located in pseudo-autosomal regions (which is not subject to XCI), or found to display conflicting XCI-states in different studies (discordant) were color coded as by Balaton et al. (2015). Chromosomal position of the SNP-carrying genes is given along the X-chromosome on the bottom of the bar graph.

(B) Allelic expression proportions of X-linked genes in naïve UCLA1 clone4 (early passage, P6), as described in (A).

(C) Allelic expression proportions of X-linked genes in naïve UCLA1 clone9, as described in (A).

(D) Allelic expression proportions of X-linked genes in naïve UCLA1 clone12, as described in (A).

(E) Electropherograms from Sanger sequencing of cDNA obtained from the primed UCLA1 population and the naïve UCLA1 clone12, detecting the indicated SNP in the X-linked genes *MID1IP1*, *RBM3*, *HUWE1, UBL4A* (subject to XCI) and *HDHD1* (escapes XCI). Chromosomal position of the SNP-containing genes along the X-chromosome is given below. These electropherograms for the *HUWE1* SNP in primed UCLA1 was also used in main Figure 5A, and for the *HDHD1* SNP in primed UCLA1 in Figure S5A, to make data comparison easy in those figures.

(F) Electropherogram of Sanger sequencing of cDNA obtained from late passage naïve UCLA1 cells demonstrating the detection of both the reference (A) and alternate (G) SNP rs1620574 (in the last exon of *XIST*). The graph below gives allelic expression proportions of the same SNP in *XIST*-expressing single cells of naïve UCLA1 cells. Each bar represents a single cell from single cell RNA seq of early and late passage naïve UCLA1 (two different single cell RNA-seq experiments (exp1 or 2, marked along the X-axis with labels of each cell coordinate in a 96 well plate format). Experiment 1 (exp1) contained cells from late passage naïve culture, whereas experiment 2 (exp2) was from an early passage (before all cells became *XIST*-positive). Twenty-six single cells expressed solely the alternate SNP (orange) (mono-allelic *XIST*), 7 cells solely the reference SNP (blue) (mono-allelic *XIST*), and 13 cells both the reference and alternate SNPs (part orange and part blue, evidence for bi-allelic expression of *XIST* at the single-cell level).

Figure S3. Naïve *XIST*-positive hESCs resemble the human blastocyst more closely than the intermediate *XIST*-negative cells [Related to Figure 2]

(A) Heatmap of unsupervised hierarchical clustering of RRBS-based methylation levels of covered CpGs (number is given) in autosomal CGIs in indicated cell lines and states.

(B) Heatmap of DNA methylation as in (A), but of all covered autosomal CpGs.

(C) Violin plots of methylation levels of covered CpGs within primary imprinted control regions (Okae et al., 2014) in indicated cell lines and states, based on RRBS data. UCLA10 is a male primed hPSC line included for comparison.

(D) Empirical cumulative distribution functions of differential methylation values of each covered CpG in pairwise comparisons between primed UCLA1 and naïve UCLA1 clone4 (at early passage when it was largely *XIST*-negative), and the *XIST*-positive naïve UCLA1 clones 9 and 12, and among the clones, for CpGs within and outside of CGIs, separated by X-chromosome and autosomes. P-values for the difference test of these pairwise comparisons were determined using the Kolmogorov-Smirnov (KS) test and shown below (p=0 is red, p=1 is green). Note that when distributions including clone4 (*XIST*-negative) were compared to any other pair-wise distributions, the p-values were much higher only for X-linked CGIs, indicating that the methylation loss in clone4 relative to primed UCLA1 was far more striking in X-linked CGIs compared to non-CGIs in both chromosome X and autosomal contexts as well as autosomal CGIs.

(E) K-means clustering of 3701 genes differentially expressed among primed UCLA1 and naïve clones 4 (early passage), 9 and 12 (see Table S2 for a list of these genes). The average expression level for all genes in a cluster is shown and the number of genes (n) in each cluster is given on the right.

(F) Expression levels in replicate RNA-seq data (rep1, rep2) of primed UCLA1 and the naïve clones 4 (early passage), 9 and 12 for genes significantly down-regulated in epiblast cells of the human blastocyst compared to early passage primed hESC lines defined based on published single cell RNA-seq data (Yan et al., 2013), visualized with violin plots (see Table S2 for a list of these genes). KS test p-values of pair-wise comparisons are given below and color-coded based on significance.

(G) As in (F), but for genes significantly up-regulated in epiblast cells of the human blastocyst compared to early passage primed hESC lines.

(H) Empirical cumulative distribution functions of X-linked gene expression for primed UCLA1 and naïve UCLA1 clones 4 (early passage, *XIST*-negative), 9 and 12, from replicate RNA-seq data sets (rep1, rep2). The inset shows autosomal expression data for the same samples at the same scale. X-linked but not autosomal expression distributions of both replicates of early passage clone4 were the only samples that were significantly different from the distributions of any other sample (*=$p<0.006$ by Wilcoxon rank sum test with continuity correction). X-linked gene expression from primed UCLA1 cells did not differ from that of *XIST*-positive naïve clones 9 and 12 ($p>0.39$).

(I) Same as in (H), but for replicates of primed UCLA1 and re-primed cells of naïve UCLA1 after 30 passages in the naïve condition and 7 passages in the re-primed state. X-linked but not autosomal gene expressions of re-primed cells were significantly different from both replicates of primed cells ($p<0.003$ by Wilcoxon rank sum test with continuity correction).

Figure S4. Originally described naïve WIBR3 hESCs have two active X-chromosomes with predominantly mono-allelic *XIST* expression [Related to Figure 3]

(A) Representative RNA FISH images detecting *XIST, HUWE1*, and *UTX* in primed WIBR3 (P24) and WIBR3 in two modified naïve culture conditions (5iLA, without FGF2 at P15, and 4iLA, without IM12 and FGF2 at P12; Theunissen et al., 2014; Theunissen et al., 2016). A single nucleus with the most prevalent pattern is highlighted with a dotted box and enlarged for ease of viewing, and also depicted by the cartoon on the right.

(B) Representative RNA FISH images detecting *XIST*, *THOC2,* and *UTX* in WIBR3 hESCs as described in (A).

(C) Representative RNA FISH images detecting *XIST*, *ATRX,* and *UTX* in WIBR3 hESCs as described in (A).

Figure S5. Non-random XCI in differentiating naïve hESCs and hiPSCs [Related to Figure 5]

(A) Representative electropherograms from Sanger sequencing of a SNP-containing region in the X-linked gene *HDHD1*, known to escape XCI, on cDNA obtained from the primed UCLA1 population at P19 (top), and from an individual cell differentiated from the re-primed state after transition through the naïve state (bottom). The pie chart summarizes the Sanger sequencing results for this SNP in 24 *XIST*-positive, single cells differentiated from the re-primed state, considering three categories: when only the reference (Ref) allele, only the alternate (Alt) allele, or both alleles are expressed.

(B) Allelic expression proportions of X-linked genes known to escape XCI (*PLS3*, *CD99*) based on RNA-seq reads (number of reads indicated on top) from single *XIST*-expressing cells differentiated from naïve UCLA1 after transition through the re-primed state. The alphanumeric labels along the X-axis refer to single cell coordinates in a 96-well plate. For comparison, allelic proportions in primed UCLA1 were calculated from two replicates of population RNA-seq data of primed UCLA1 (Pri1/Pri2).

(C) Representative immunofluorescence images of MeCP2 in combination with RNA FISH for *XIST* in primed hiPSCs derived from Rett syndrome patient fibroblasts, bearing the wild-type (clone16) or mutant (clone17) *MECP2* allele on the X_a, and upon conversion to the naïve state (P3 in 5iLAF media). The appearance of the MeCP2 protein in almost all cells of clone17 at P3 in naïve culture media captured the reactivation of the X_i, which occurred without induction of *XIST* expression consistent with sequential order of X_i-reactivation and *XIST* induction from the X_a.

(D) Representative immunofluorescence images of cells differentiated from primed hiPSC clones 16 and 17 described in (C) detecting MeCP2 and the negative pluripotency marker CD44 (Quintanilla et al., 2014).

(E) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in re-primed cells (P3) obtained from naïve hiPSC clones 16 and 17 described in (C), and upon differentiation of these re-primed cells. Both re-primed clones 16 and 17 displayed biallelic *HUWE1* expression, indicating the X_aX_a state of re-primed cells. Differentiation was accompanied by *XIST*-mediated XCI demonstrated by mono-allelic *HUWE1* focus not overlapping with the *XIST* cloud.

Figure S6. Characterization of the XCI status in primed UCLA1 and UCLA9 and upon differentiation [Related to Figure 6]

(A) Representative RNA FISH images detecting *XIST*, *HUWE1*, and *UTX* in primed UCLA1 (P16) and after seven days of differentiation from the primed state. A single cell is enlarged in each row for ease of viewing of the prevalent RNA FISH pattern (dotted box), also depicted by the cartoon on the right.

(B) Representative phase-contrast images of UCLA9 in the primed state (P15) and after 7 days of differentiation from the primed state.

(C) Representative RNA FISH images detecting *XIST*, *ATRX* and *UTX* in primed UCLA9 and after 7 days of differentiation from the primed state. The absence of *XIST* expression and lack of silencing upon induction of differentiation of primed UCLA9 is also described and extensively discussed by Patel et al. (in review).

(D) As in (C), but for *XIST*, *THOC2* and *UTX*.

(E) Representative RNA FISH images detecting *XIST* and *ATRX* in differentiated cells derived from naïve then re-primed UCLA9. The single cell shown depicts the prevalent RNA FISH pattern in differentiated cells, which is also depicted by the cartoon on the right. Contrary to the differentiated cells originating from

the original primed UCLA9 hESC line, these cells, which are the progeny of naïve then re-primed UCLA9, demonstrated XCI with *XIST* expression.

(F) As in (E), but for *XIST* and *THOC2*.

Figure S7. UCLA4 maintains an eroded X_i upon differentiation from the primed state and is karyotypically normal in early passage naïve state [Related to Figure 7]

(A) Representative phase-contrast images of UCLA4 in the primed state (P14) and upon 7 days of differentiation from the primed state.

(B) Representative RNA FISH images detecting *XIST* RNA and the nascent transcription foci of *HUWE1* along with *UTX* in primed UCLA4 and their day seven differentiated product. A single nucleus representing the predominant X-chromosome pattern is highlighted with a dotted box and enlarged for ease of viewing, and its X-pattern is depicted by the cartoon on the right.

(C) Representative RNA FISH images detecting *XIST*, *ATRX* and *UTX* as described in (B).

(D) A representative metaphase chromosome spread and summary of the cytogenetic analysis of naïve UCLA4 at P9.

Supplemental Table Legends

Table S1. Summary of the X-chromosome state in primed, naïve, re-primed and differentiated cells described in this study. Related to Figures 1, 4, 5, 6 and 7.
Summary of the X-chromosome state for cell lines analyzed in this study in primed, naïve, re-primed, and differentiated conditions, as applicable.

Table S2. Gene lists and RPKM values of RNA-seq datasets used. Related to Figures 2 and S3.
Raw RPKM values from stranded RNA-sequencing data for primed UCLA1 and 5iLAF-cultured naïve UCLA1 clones 4 (early passage), 9 and 12 (two replicates each), with corresponding gene names and chromosomal locations, for: (i) all genes; (ii) genes differentially expressed between primed UCLA1 and naïve UCLA1 clones 4 (early passage; Figure S3E), 9 and 12; (iii) genes that are up-regulated in the blastocyst compared to primed hESCs (Figure S3F); (iv) genes that are down-regulated in the blastocyst compared to primed hESCs (Figure S3G); and (v) cell lines used in ECDF plots (Figures 2 and S3).

Supplemental Experimental Procedures

Cell Culture

Primed hPSCs were obtained from the Human Embryonic and Induced Pluripotent Stem Cell Core at UCLA and maintained in primed media consisting of 20% KnockOut Serum Replacement (KSR) (Life Technologies) in DMEM/F12 (Life Technologies) supplemented with 1x penicillin/streptomycin (Life Technologies), 1x nonessential amino acids (Life Technologies), 0.5x GlutaMAX (Life Technologies), 0.1mM β-mercaptoethanol (Sigma), and 10ng/ml FGF2 (Peprotech). Cells were passaged every 6-7 days by detaching colonies with 1mg/ml collagenase IV (Life Technologies) at 37°C for 5-15 min, followed by manually breaking up colonies via pipetting. Conversion to the naïve state with the 5iLAF culture protocol was done as described previously (Theunissen et al., 2014). Briefly, two days post passage of primed hPSCs, media was changed to modified primed media which is made exactly as primed media, but instead of 20% KSR contained 5% KSR and 15% FBS (Omega Scientific). On day 7aZ post passage, cells were dissociated into a single cell suspension with 0.05% trypsin-EDTA at 37°C for 3 min and passed through a 40µm strainer. $2x10^5$ single cells were plated in one well of a 6-well plate in modified primed media in the presence of 10µM ROCK inhibitor Y-27632 (Stem Cell Technologies). Two days post passage, 5iLAF media was applied, which consisted of a 1:1 mixture of DMEM/F12 and Neurobasal (Life Technologies), supplemented with 1x N2 (Life Technologies), 1x B27 (Life Technologies), 1x penicillin/streptomycin, 1x nonessential amino acids, 0.5x GlutaMAX, 0.5% KSR, 0.1mM β-mercaptoethanol, 50µg/ml bovine serum albumin (Sigma), 20ng/ml rhLIF (EMD Millipore), 20ng/ml Activin A (Peprotech), 8ng/ml FGF2, 1µM MEK inhibitor PD0325901 (Stemgent or Bio-Techne), 0.5µM B-Raf inhibitor SB590885 (Bio-Techne), 1µM GSK3β inhibitor IM-12 (Enzo), 1µM Src inhibitor WH-4-023 (A Chemtek), and 10µM ROCK inhibitor Y-27632. At about 11 to 12 days post plating, cells were dissociated by a 3 min treatment with StemPro Accutase (Life Technologies) at 37°C and re-plated after passing through a 40µm cell strainer in 5iLAF medium. Naïve hESCs were passaged as single cells every 5−6 days.

For sub-cloning of the 5iLAF UCLA1 hESC line at early passage, individual dome-shaped colonies that arose after single-cell plating were manually picked and expanded (into clones 4, 9, and 12). Conversion in 4iLA and 5iLA media, lacking both FGF2 and IM-12 or only FGF2, respectively, was done following the steps described for the 5iLAF protocol (Theunissen et al., 2016). In general, primed and naïve hPSCs were grown on irradiated CF-1 or DR4 mouse embryonic fibroblast feeder cells and maintained in a humidified 37°C incubator at 5% CO_2 and atmospheric oxygen levels. Conversion from primed UCLA1 to 5iLAF naïve pluripotency was also done at 5% O_2, but there was no change in cell growth rate, colony morphology, or X-chromosome state when compared to cells at atmospheric oxygen.

For the transition of naïve cells to the primed state (re-priming), 5iLAF media was switched to primed media 3−4 days after splitting the naïve cells. 2−3 days later, we observed significant cell death and naïve colonies became flat, visually resembling a primed hPSC culture. Re-primed cells were passaged with collagenase IV as described for primed hPSCs.

For differentiation, primed and re-primed hPSC lines were detached briefly with collagenase IV, transferred into fibroblast media and collected at the bottom of a conical tube by gravity to remove feeder cells. The colonies collected at the bottom were washed with Dulbecco's phosphate-buffered saline (DPBS), incubated with StemPro Accutase at 37°C for 2 min, passed through a 40µm strainer and plated at a final 1:2 split ratio in fibroblast media supplemented with 10µM ROCK inhibitor Y-27632 on plates or glass coverslips coated with matrigel (Fisher Scientific). Subsequently, the media was changed after two days and then every other day, which excluded the ROCK inhibitor. Differentiation was carried out for seven days.

The primed hESC lines WIBR3, UCLA1, UCLA4, and UCLA9 were described and characterized previously (Lengner et al., 2010; Diaz Perez et al., 2012; Patel et al., in review). Similarly, naïve WIBR3 and UCLA1 lines, the 5iLAF naïve hESC lines UCLAn19 and UCLAn20 (directly derived from blastocysts), and the naïve t2iL+Gö adapted H9 cells were reported previously (Theunissen et al., 2014; Pastor et al., 2016; Takashima et al., 2014). t2iL+Gö-adapted naïve H9 cells and their primed counterparts were grown at 5% O_2 and 7% CO_2 conditions. The low passage *XIST*-expressing X_aX_i primed iPSC line

#1001 used for the conversion to the naïve state in 5iLAF was generated and analyzed previously (Karumbayaram et al., 2012). Conversion of primed hESC line UCLA1 to 5iLAF naïve state was repeated four times and each time similar timing of epigenetic events and morphological changes were observed. Conversions of the primed hESC lines UCLA4 and UCLA9 were repeated three times each, again with similar timing of events and morphological changes. Naïve UCLA4 hESCs were sent to Cell Line Genetics (Madison, WI) for G-banding karyotype analysis at passage 9 and karyotypes for naïve UCLA1 and the blastocyst-derived naïve hESC lines UCLA19n and UCLA20n have been reported (Pastor et al., 2016) Re-priming and differentiation of all naïve hPSC lines was repeated at least twice.

Rett syndrome hiPSCs were generated from GM07982 female fibroblasts (NIGMS Human Genetic Cell Repository) heterozygous for the 705delG frame-shift mutation in *MECP2*, which leads to a premature stop codon in the mRNA and therefore absence of a full length MeCP2 protein product (Lee et al., 2001). Before reprogramming, the presence of the mutation was verified by PCR of genomic DNA. For reprogramming to hiPSCs, 100,000 fibroblasts were plated in one well of a 6-well plate, and one day later infected overnight with 20ul of concentrated (about $5x10e^8$ TU/ml) STEMCCA lentivirus, a single polycistronic lentiviral vector encoding Oct4, Klf4, Sox2, and c-Myc under control of the constitutive EF1a promoter (Sommer et al., 2009) in 1ml fibroblast media (10% FBS in DMEM supplemented with 1x penicillin/streptomycin, 1x nonessential amino acids, 1x GlutaMAX and 0.1mM β-mercaptoethanol) with 5µg/ml polybrene. On day 5 post-infection, cells were trypsinized and re-plated on feeders. The next day media was changed to primed hESC media and replaced daily. Human ESC-like colonies were picked between weeks 2−3 post infection and enzymatically passaged with collagenase IV. Several clones were analyzed for their XCI state as described (Tchieu et al., 2010) to demonstrate their $X_aX_i^{XIST+}$ state at early passage and the transition to the *XIST*-negative X_aX_i state over subsequent passages. The location of the mutant or wild-type *MECP2* allele, respectively, on the X_a was determined by Sanger sequencing of RT-PCR products, which allowed the classification of clone16 as $X_a^{MECP2\text{wt}}X_i^{MECP2\text{mut}}$ and clone17 as $X_a^{MECP2\text{mut}}X_i^{MECP2\text{wt}}$. In addition, the expression of endogenous pluripotency genes and the silencing of the ectopic reprogramming cassette were confirmed by RT-PCR, cytogenetic analysis was performed by Cell Line Genetics to demonstrate a karyotypic normal state, and teratoma formation assays were conducted by testis injection following standard procedures to demonstrate ability to differentiate into all three germ layers, as described (Tchieu et al., 2010).

RNA fluorescent in situ hybridization (FISH)

Cells were grown on gelatinized 18mm circular glass coverslips (Fisher Scientific, 12-545-100), washed with DPBS, fixed with 4% formaldehyde for 10 min, permeabilized with cold (4°C) 0.5% Triton X-100 in DPBS for 10min, and serially dehydrated with cold (4°C) 70-100% ethanol. Coverslips were air dried and hybridized with labeled DNA probes in a chamber humidified with 50% formamide in 2x SSC at 37°C for 24−36 hours, washed for three 5min intervals with 50% formamide in 2x SSC, 2x SSC, then 1x SSC at 37°C, then mounted with ProLong Gold Antifade reagent containing DAPI (ThermoFisher). Double-stranded DNA probes were generated from BACs with the BioPrime Array CGH Genomic Labeling System and fluorescently labeled ChromaTide nucleotides (ThermoFisher). The BACs used include *XIST* (RP11-13M9), *XACT* (RP11-35D3), *HUWE1* (RP11-975N19), *UTX* (RP11-256P2), *ATRX* (RP11-1145J4), and *THOC2* (RP11-121P4). The labeled DNA pellet, along with salmon sperm DNA and human Cot1 DNA (ThermoFisher), was stored at -80°C after re-suspension in deionized formamide (VWR) and 2x hybridization buffer (0.2g/ml dextran sulfate, average M_w >500,000 (Sigma) in 4x SSC and 0.1M NaH_2PO_4). Every new batch of probes was tested on normal human dermal fibroblasts before use in experiments. Every 5iLAF naïve hESC line generated was subjected to RNA FISH analysis of its X-chromosome state three or more times at different passages and the data obtained were consistent for each time point.

Immunofluorescence

Cells were grown on coverslips and fixed as for RNA FISH, but after the 10min incubation in 0.5% Triton X-100 in DPBS, coverslips were incubated in 0.2% TWEEN-20 in DPBS for 10min. Coverslips were then incubated with blocking buffer (5% donkey serum (Fisher Scientific), 0.2% gelatin from cold water fish skin (Sigma), 0.2% TWEEN 20) for 30min in a humidified chamber. This was followed by incubation with primary antibody, diluted in blocking buffer, overnight at 4°C (anti-MeCP2: Diagenode, C15410052, used

at 50ng/ml; anti-CD44: also called Hermes-1, deposited to the Developmental Studies Hybridoma Bank by E.C. Butcher, used at 1:10; Quintanilla et al., 2014). Coverslips were then washed three times (5min each) with 0.2% TWEEN-20 containing DPBS and incubated with secondary antibody diluted in blocking buffer for 1hr at room temperature. Coverslips were washed again as after the primary antibody incubation and mounted with ProLong Gold Antifade reagent containing DAPI. For MeCP2, H3K27me3 or RNA-PolII co-detection with *XIST* RNA, immunofluorescence (IF) was followed by RNA FISH. IF was done as described above, but the primary antibody incubation time was reduced to 1hr at room temperature and RNase-out (ThermoFisher) was added to the blocking buffer at 1:200. Anti-H3K27me3 antibody was used at 1:400 (Active Motif, 39155) and the anti-RNA PolII antibody at 1:1000 (Millipore 05-623, clone CTD4H8). After secondary antibody washes, the RNA FISH protocol was carried out as described above, starting from the ethanol dehydration steps.

Microscopy and Image analysis

Images were taken as Z-stacks with the Imager M1 microscope (Zeiss) at 630x magnification using the Axio Vision software. All image processing was done with the ImageJ software (NIH). Z-stack images were merged using maximum intensity on gray scale images, and the color merge function was used for overlaying merged Z-stacks of different channels. For FISH images, brightness and contrast was adjusted for each channel after merging of Z-stacks but before overlaying, to remove background signal. RNA FISH signals were quantified by eye either from images or through the eyepiece of the microscope. For IF and IF/RNA-FISH, the exposure time was kept constant across all samples during image acquisition in all channels except for DAPI and the images were processed without any changes to brightness or contrast. A minimum of 100 cells from at least five different colonies (in the case of hPSCs) or fields (in the case of differentiated cells) was analyzed.

RNA-sequencing

For strand-specific RNA sequencing of the primed hESC line UCLA1, the naïve UCLA1 clones 4 (at early (*XIST*-negative, P6) and late (*XIST*-positive, P16) passage), clones 9, and 12 (P8), re-primed UCLA1, and naïve UCLA4 at early (P5) and late passage, cells were harvested, washed with DPBS, and collected in Trizol (ThermoFisher). Where indicated, two replicates were obtained by harvesting cells from two different wells of a 6-well plate and independent processing through all downstream steps. Only for late passage naïve UCLA4 (*XIST*-positive), the replicates represent two independent conversions from the primed to the naïve state at P22 and P17, respectively. RNA was isolated after chloroform extraction using the RNease Mini Kit (Qiagen). Before elution of RNA, the column was subjected to DNase digestion (Qiagen) for 30 min at room temperature to remove any contaminating genomic DNA. RNA quality and amount were determined using a NanoDrop spectrophotometer ND-1000 (Thermo Scientific). Four µg of total RNA were then used for mRNA isolation and library preparation using the TruSeq Stranded mRNA Library Prep Kit (Illumina) following the manufacturer's Sample Preparation Guide. All libraries were amplified for 15 cycles and the final PCR product was run on a low-melt agarose gel and DNA with a median of 250bp was extracted from gel slices. The resulting libraries were sequenced as single-end 50bp reads at the UCLA Broad Stem Cell Center High-Throughput Sequencing facility. Reads were mapped using TopHat v2.0.13 and assigned to genes with HTSeq-0.6.1 (using the Illumina iGenomes UCSC hg19 assembly). Reads Per Kilobase per Million mapped reads (RPKM) values were calculated (Table S2) and differential gene expression analysis for the replicates of primed UCLA1 and naïve UCLA1 clones 4, 9 and 12 was preformed using the DESeq2 (version 1.6.2) R package for each pairwise comparison among the two replicates of primed hESC line UCLA1 and its naïve clones 4 (early passage), 9 and 12. Genes differentially expressed (p-value < 0.01 and abs(Log_2 fold-change) > 2) among all comparisons (n=3,701) (Table S2 were concatenated and used for k-means clustering (k=5) after log_2 transformation of RPKM values. The heatmap (version 1.0.8) R package was used to construct the heatmap.

For comparing the expression pattern of primed UCLA1 cells and naïve clones 4, 9 and 12 to that of epiblast cells of the human pre-implantation blastocyst, we used the 1,008 genes differentially expressed between human epiblast cells of the blastocyst and newly derived primed hESCs defined by Yan et al., (2013) that satisfied a significance threshold (p-value < 0.01 and abs(Log_2 fold-change) > 2), and divided them into ordered sets of up- and down-regulated genes according to fold-change and directionality (Table S2). In these two gene sets, after Log_2 transformation of the RPKM expression values, the geom_violin function of ggplot2 (version 2.1.0) R package was used to generate violin plots, and the Kolmogorov-

Smirnov (KS) test to identify significantly different distributions. Again, only UCLA1 samples with replicates were used in this analysis.

For globally comparing X-linked and autosomal gene expression across all samples, the complete gene set (n=26,364) was subset into autosomal (n=25,265) and X-linked (n=1,099) genes. The Empirical Cumulative Distribution Function (ECDF) (computed using the stat_ecdf() function of ggplot2 (version 2.1.0) R package; http://www.r-project.org (Dean and Nielsen, 2007) was used to plot the gene expression distribution for both subsets across all samples in each of indicated comparison groups and the Wilcoxon rank sum test with continuity correction (performed with the wilcox.test() function of the stats (version 3.3.0 R package) was used to identify significantly different distributions.

For single cell RNA-seq, we used early and late passage UCLA1 5LAF naïve hESCs (not sub-cloned) to capture *XIST*-negative (enriched at early passage) and *XIST*-positive (predominant at late passage) naïve cells, respectively, as well as UCLA1 cells differentiated for 7 days from the re-primed state (after transition through the naïve state for over 20 passages). Cells were detached with 0.05% trypsin-EDTA (differentiated) or Accutase (naïve) at 37°C for 2–3 min, passed through a 40µm strainer, and counted using a hemacytometer. Cell concentration was adjusted to 200,000 cells/ml before loading on a Fluidigm C_1 Single-Cell Auto Prep System (Fluidigm) following the manufacturer's guidelines. For differentiated cells we used the Single Cell Preamp IFC 17–25 µm, for naïve cells the Single Cell Preamp IFC 10–17 µm. cDNAs were made on-chip with the Clontech SMARTer Ultra Low RNA kit for Illumina using protocols provided by Fluidigm. Libraries were constructed in 96-well plates using the Illumina Nextera XT DNA Sample Preparation kit according to the standard protocol supplied by Fluidigm, and sequenced as paired-end 100 base pair reads at the UCLA Broad Stem Cell Center High-Throughput Sequencing facility. 1/10 of the produced cDNA was diluted 5-fold and used for experiments with direct Sanger sequencing. Microsoft Excel and BoxPlotR (http://boxplot.tyerslab.com) were used for generating figures describing the single cell data.

Determination of allelic expression of X-linked genes based on SNPs

To determine the allelic expression state of X-linked genes, heterozygous SNPs were defined genome-wide using the Affymatrix SNP6.0 array at the UCLA Clinical Microarray Core, on genomic DNA of the primed hESC line UCLA1. Taking advantage of SNPs in exons of X-linked genes, allelic expression was assessed using either population or single cell RNA-seq data, or SANGER sequencing of RT-PCR products, as indicated. For population RNA-seq data, the proportion of reads covering the reference or alternative SNP was graphed for those SNPs common to primed and naïve sub-clones with 5 or more reads (Figure 2B), or for all SNPs in primed or each naïve sub-clone with 10 or more reads (Figures S2A–S2D). For differentiated UCLA1 single cell RNA-seq data, the proportion of reads covering the reference or alternative SNP was graphed for SNPs with 5 ore more reads in each cell expressing *XIST,* which was confirmed by PCR from single cell cDNA. Only SNPs with coverage in two or more individual cells were graphed (Figures 5A–5C, S5A and S5B). For naïve early- and late- passage UCLA1 single cell RNA-seq data, cells were classified as *XIST*-negative and *XIST*-positive based on *XIST* RNA counts (if the log2-depth normalized count value was < 2, a single cell was considered *XIST*-negative). As expected, early passage naïve UCLA1 had more *XIST*-negative single cells than late passage naïve UCLA1. For uncovering the allelic expression status of *XIST*, *XIST*-positive single cells with 10 or more reads spanning the *XIST* SNP rs1620574 were graphed in Figure S2F. For Figure 2C, the allelic expression status of X-linked genes normally subject to XCI covered by 10 or more reads at informative SNPs was determined. Each circle in the figure represents the result for a particular SNP in a single cell.

In case of Sanger sequencing, we used either cDNA from single differentiated cells from re-primed UCLA1 hESCs or from populations of primed UCLA1 at P19 and naïve clone12. Only for the *XIST* expression analysis in primed UCLA1, early passage cells were employed (P4), since *XIST* was silenced at P19. cDNA was synthesized using SuperScript III First-Strand Synthesis SuperMix (ThermoFisher), and amplified with primers spanning the SNP of interest using KAPA HiFi PCR polymerase (KAPA Biosystems). The following forward and reverse primers were used: 5'CCATAGCTGACCAAGGCCAG3' and 5'CGGCAGCACCGAGATAAAAGG3' for *MID1IP1* (SNP ID rs198783), 5'TGGTTATGACCGCTACTCAGG3' and 5'CTGCCCCCACTTTTAATTTGC3' for *RBM3* (SNP ID rs235829), 5'CTACCCGTGAAGTCCTTGGC3' and 5'CGTTCCTCTGTACCAACAACC3' for *HUWE1*

(SNP ID rs6638360), 5' CAGCAGGGTCCTGGAACAG3' and 5'CAGTGCTGGGGATGAGGAC3' for *UBL4A* (SNP ID rs7057286), 5'GTTTGCTACCTCACAACAACC3' and 5' GCAGACATATATTCAGGCCATC3' for *HDHD1* (SNP ID rs6982), 5'CATTGCTAGGCATTGGGGATG3' and 5'CCAGGAAGCATGTATCTTCTGG3' for *XIST* (SNP ID rs1620574). PCR products were run on an agarose gel, expected sized bands gel-eluted using the MinElute Gel Extraction Kit (Qiagen), and sent for Sanger Sequencing (Retrogen Inc., San Diego, CA) using the forward PCR primer. The 4Peaks software was used to visualize the electropherograms of sequencing results.

DNA methylation

Genomic DNA was harvested from primed hESCs and UCLA1 naïve clones using the DNAeasy blood and tissue kit (Qiagen). For each cell population two replicates were obtained by harvesting cells from two different wells of a 6-well plate and independent processing. Libraries for Reduced Representation Bisulfite Sequencing (RRBS) were created as previously described (Meissner, 2005), and size-selected between 50 and 500bp. DNA methylation analysis was performed using BS-Seeker2 (2.0.32) (Guo et al., 2013) using Bowtie (0.12.9) (Lee et al., 2001) for read alignment to the human genome (hg19) on the UCLA Hoffman2 computer cluster. Reads with adapter contamination were trimmed. CpG island (CGI) coordinates were obtained from UCSC (http:// genome.ucsc.edu). Only CpG sites covered by at least five reads across all samples under consideration were used in an effort to obtain reliable methylation levels. Since replicate samples showed good correlation (data not shown), we merged replicates by summating counts at each CpG site.

Methylation data were hierarchically clustered using complete linkage and the Euclidean distance metric. Statistical analysis, clustering, and heat map generation were performed using custom R scripts. For the presentation of X-linked CGIs in Figure 2, CpGs were additionally filtered and required to have less then 20% methylation in the male primed hESC line UCLA10 (Patel et al., in review), to emphasize the methylation state due to XCI. For the analysis of methylation in imprint control regions, all CpGs were filtered for minimum 5-fold coverage from merged replicates of RRBS data, then selected for overlap with maternal and paternal non-placental imprinted regions (Okae et al., 2014). Methylation was visualized using the R-software package. For Figure S3D, the pairwise distributions of differential DNA methylation was plotted using CDFs for X-linked and autosomal CpGs within and outside of CGIs with at least 5x coverage and the Kolmogorov-Smirnov test was used to test the difference of pairwise comparisons between the distributions.

Supplemental References

Balaton, B.P., Cotton, A.M., and Brown, C.J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol. Sex Differ. *6*, 35.

Dean, C.B., and Nielsen, J.D. (2007). Generalized linear mixed models: a review and some extensions. Lifetime Data Anal. *13*, 497–512.

Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., and Clark, A.T. (2012). Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. Hum. Mol. Genet. *21*, 751–764.

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.-Y., and Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics *14*, 774.

Karumbayaram, S., Lee, P., Azghadi, S.F., Cooper, A.R., Patterson, M., Kohn, D.B., Pyle, A., Clark, A., Byrne, J., Zack, J.A., et al. (2012). From Skin Biopsy to Neurons Through a Pluripotent Intermediate Under Good Manufacturing Practice Protocols. Stem Cells Transl. Med. *1*, 36–43.

Lee, S.S., Wan, M., and Francke, U. (2001). Spectrum of MECP2 mutations in Rett syndrome. Brain Dev. *23 Suppl 1*, S138–S143.

Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J., et al. (2010). Derivation of Pre-X Inactivation Human Embryonic Stem Cells under Physiological Oxygen Concentrations. Cell *141*, 872–883.

Meissner, A. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

Okae, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., Kikuchi, H., Yoshida, H., Tanaka, A., Suyama, M., et al. (2014). Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet. *10*, e1004868.

Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., and Clark, A.T. (2016). Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. Cell Stem Cell *18*, 323–329.

Quintanilla, R.H., Asprer, J.S.T., Vaz, C., Tanavde, V., and Lakshmipathy, U. (2014). CD44 Is a Negative Cell Surface Marker for Pluripotent Stem Cell Identification during Human Fibroblast Reprogramming. PLoS ONE *9*, e85419.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells Dayt. Ohio *27*, 543–549.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell *158*, 1254–1269.

Tchieu, J., Kuoy, E., Chin, M.H., Trinh, H., Patterson, M., Sherman, S.P., Aimiuwu, O., Lindgren, A., Hakimian, S., Zack, J.A., et al. (2010). Female Human iPSCs Retain an Inactive X Chromosome. Cell Stem Cell *7*, 329–342.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., et al. (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. Cell Stem Cell.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.