I. SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Summary of cell lines used in this study and maintenance culture

Two male and ten female ESC lines used in this study were derived at UCLA, among which UCLA1-5 were described previously (Diaz Perez et al., 2012). The UCLA ESC lines are all listed on the NIH Human Embryonic Stem Cell Registry, and are eligible for use in NIH-funded research projects. No federal grant funding was used for work with human embryos or derivation of new ESC lines. No payment was provided to embryo donors for their generous gift of surplus embryos to stem cell research.

All UCLA ESC lines described in this study were routinely propagated in DMEM/F-12 (Life Technologies) supplemented with 20% knockout serum replacement (KSR; Life Technologies), 2mM GlutaMAX-I (Life Technologies), 1% non-essential amino acids, 100 units/ml penicillin with 100 ug/ml streptomycin (Life Technologies), 0.1 mM 2-mercaptoethanol (Sigma) and 10-20 ng/ml bFGF (Peprotech) (primed human ESC media), on γ-irradiated mouse embryonic fibroblasts (MEFs or feeders). All lines showed normal karyotype, the capacity to produce teratomas with differentiation into all three germ layers, and expressed markers of pluripotency (Table S1). To prepare UCLA ESC lines for analysis by RNA FISH, RRBS, and RNA-seq, cells were first feeder depleted by allowing feeder cells to settle in a culture tube, and then plated onto matrigel-coated plates (BD Biosciences) in mTesr1 (StemCells Inc). After 2–3 passages with Versene solution (Lonza), feeder-free ESC cultures were used for all analyses.

H7, H9, ESI02, and ESI03 ESCs were obtained from the WiCell Research Institute's WISC Bank; WIBR2 and WIBR3 ESCs from the Whitehead Institute (Lengner et al., 2010); and the iPSC lines 3S-5F-4, K-3F-1, K-3F-2 after at least 15 passages under SNL/LIF culture conditions from the Gladstone Institute (Tomoda et al., 2012), with required approval forms (Table S1). WIBR2 and WIBR3 ESCs and Gladstone iPSCs were shipped to UCLA as live cultures. WIBR2 and WIBR3 were grown to semi-confluence in hypoxia for one week and then harvested for analysis by RNA FISH and RRBS. The Gladstone iPSC lines were harvested immediately upon receipt for RRBS analysis and re-plated for RNA FISH, which was conducted 48 hours later. Thus, for WIBR2 and WIBR3 ESC and Gladstone iPSC cultures, the RRBS analysis was performed directly with human ESCs in the presence of CF1 and SNL feeders, respectively. H7, H9, ESI02, and ESI03 ESCs were cultured and prepared for analyses as UCLA ESC lines.

In addition, we used female (Lonza 17914) and male (Coriell GM0227) normal human dermal fibroblasts (NHDFs) as controls for RNA FISH and RRBS analyses. Passage numbers of ESC and iPSC lines for key analyses are summarized in Table S1. SSEA4 and TRA1-81 levels of newly established UCLA ESC lines (Table S1) were determined as previously described (Diaz Perez et al., 2012). For teratoma assays, institutional approval and guidelines by the UCLA Animal Research Committee (ARC) were followed for surgical procedures and appropriate care and use of laboratory animals. About 2 to 3 wells of a densely populated 6-well plate of ESCs were used per injection site. Cells grown on feeders were harvested with collagenase IV (Life Technologies), re-suspended in 50uL Matrigel (Corning) per injection, and

delivered into the testis of adult male SCID beige mice using a 27G needle. Six to eight weeks post injection, tumors were removed and fixed in 4% PFA in 1x PBS (diluted from 16% PFA from Electron Microscopy Sciences) overnight and then stored in 70% ethanol. Tissue pieces were sectioned, mounted onto glass slides and H&E stained by the Translational Core Pathology Laboratory (UCLA).

Pre-implantation embryo handling, ESC derivation and maintenance

Surplus embryos were donated from couples that underwent *in vitro* fertilization treatment under informed consent. Prior to initiating work, the study was approved by the UCLA Embryonic Stem Cell Research Oversight (ESCRO) Committee and the UCLA Institutional Review Board. Cryopreserved cleavage stage embryos were thawed using THAW-KIT1 (Vitrolife) according to manufacturer's instructions and cultured until blastocyst formation in Continuous Single Culture media (CSC; Irvine Scientific) supplemented with 10% Serum Substitute Supplement (SSS; Irvine Scientific) under 5% O_2. Cryopreserved blastocysts were thawed using either the Blastocyst Thaw Media Kit (Irvine Scientific) or the G-ThawKit Blast (Vitrolife) according to manufacturer's instructions and cultured overnight in Global medium (Life Global) or CSC supplemented with 10% SSS under 5% O_2. For ESC derivation, the zona pellucida of blastocyst-stage embryos was removed using acidified Tyrode's solution (Irvine Scientific), and blastocysts were plated onto γ-irradiated MEFs in propagation media as described above, except with 30ng/ml bFGF. One week after plating, the derivation media was supplemented with MEF-conditioned media (50:50% v/v) to maintain the blastocyst outgrowth on the same plate. One ESC line, UCLA17, was derived in derivation media containing 10% KSR and 10% FBS and supplemented with human ESC-conditioned media (Tachibana et al., 2013). During derivation, normoxic and hypoxic culture conditions were altered as summarized in Table S1. About three weeks after plating, small parts of the large blastocyst outgrowth were manually isolated and transferred into a new dish with irradiated feeders and from then on maintained under normoxic conditions. For established ESC lines, manual passaging was initially carried out for several passages, and subsequently switched to enzymatic dissociation using collagenase type IV 1mg/mL (Life Technologies). At this point, the bFGF concentration was reduced to 10 or 20 ng/ml in primed human ESC media.

ESC differentiation

For retinoic acid differentiation, ESCs were plated at 50% confluence on Growth Factor Reduced-matrigel in mTesr1 and after 48 hours transferred into media containing 15% FBS in DMEM/F12 with 1 uM retinoic acid (Tocris). Cultures were analyzed after 10 days, when OCT4 was lost nearly in all cells. Directed differentiation into hepatocyte-like cells, cardiomyocyte-like cells, and neurons was as described previously (Chambers et al., 2009; Lian et al., 2012; Medine et al., 2008). Briefly, hepatocyte-like cell cultures were established from feeder-depleted ESCs. Endoderm induction was initiated in RPMI/B27 (Life Technologies) with human Activin A and Wnt3a (Peprotech). Priming of definitive endoderm was monitored by morphology, and subsequent differentiation induced using DMSO and KSR-containing media. Final hepatic maturation was achieved in L-15 media (Lonza) with 8.3% heat-inactivated FBS and 10 μM hydrocortisone 21-hemisuccinate (Sigma), 1μM insulin (bovine pancreas, Sigma), 1% L-Glutamine, and 0.2% ascorbic acid (Sigma). Hepatocyte-like cells were maintained in HGF and

Oncostatin M-containing media and assayed for XCI state 30 days after induction of differentiation. Albumin concentrations in hepatocyte-differentiated cultures were measured in media conditioned for 24 hours using the Human Albumin ELISA kit (Bethyl). Cardiomyocyte monolayer differentiation was initiated after expanding ESCs to 80% confluence on matrigel-coated plates. At day 0, cells were transferred into RPMI medium supplemented with B27 without insulin (1:50) and 6uM CHIR-99021 (Tocris). At day 1, differentiation media was replaced with RPMI supplemented with B27 without insulin. Media was changed at day 3 to RPMI/B27 without insulin and with 5 μM IWP-2 (Tocris) and at day 5 to RPMI/B27 without insulin and without IWP-2. From day 7 onwards cells were maintained in RPMI/B27 medium. Functional analysis on cardiac cell populations was performed at day 21 of differentiation. For functional analysis, cardiomyocytes were treated with Fluo 4 and maintained at 37^{o}C for 30 min prior to analysis. Ca^{2+} flux was measured by the rate and pattern of fluorescence fluctuation of individual cells. Cells displaying oscillating fluorescence were considered spontaneously active (contractile). Each cell line was measured for 10 min to determine basal Ca^{2+} flux. Subsequently, cells were treated with either 1μM of isoproterenol, or an isoproterenol vehicle control and imaged for a further 10 min. Ca^{2+} transients were recorded and analyzed after a series of depolarizations that enabled each transient to fully decay to establish steady-state. Neuronal cultures were derived using the dual SMAD method as described without modifications (Chambers et al., 2009).

RNA FISH and immunostaining

RNA FISH, immunofluorescence, and the acquisition of images were performed as described previously (Tchieu et al., 2010). For RNA FISH, we used double stranded DNA probes labeled with Alexa-flour labeled nucleotides (Invitrogen) generated with exo-Klenow (Enzymatics) by random priming, from the following BACs or FOSMIDS, which were obtained from CHORI: *XIST* (RP11-13M9), *MID1* (RP11-798D20), *CASK* (RP11-154-P12, RP11-820G10, and RP11-977L20), *HUWE1* (RP11-975N19), *ATRX* (RP11-1145J4), *GPC3* (RP11-585F14, RP11-35B13, RP11-614F19), *MTMR1* (WI2-2809A11 and WI2-922F21), *WDR44* (WI2-2159C5, WI-2 2160O10, WI2-585C17, WI2-1772N18), *DIAPH2* (RP11-315I18, RP11-1031O6, RP11-1019O12, RP11-641M1, RP11-1103G2), *AMMERCR* (RP11-938C19), *THOC2* (RP11-121P4), *UTX* (RP11-256P2)*, MECP2 (*wi2-324il8, wi2-3901pc), *XACT* (RP11-35D3). We ensured that the genes were long and rich in exonic and intronic length to achieve a high signal to noise ratio and avoid stochastic fluctuations in expression between experimental conditions. The identity of the BAC was assessed by restriction digest of the source DNA. For specificity, we tested the resultant labeled probe for a solitary signal in male H1 and female XiXa H9 ESCs and fibroblasts, when appropriate. We combined BACs or fosmids that covered a gene by taking molar ratios equivalent to their length.

RNA FISH and immunostaining on human blastocysts were performed as described (Namekawa et al., 2010) with the following modifications: After dissociation of the zona pellucida with Tyrode's acid, blastocysts were washed in three consecutive drops of 6 mg/mL BSA (Sigma) in RNAse-DNAse-free PBS (Ambion). They were then individually transferred to a poly-lysine coated slide, fluid aspirated until dry, placed in a coplin jar of ice-cold 1% paraformaldehyde and 0.2% Triton X-100 and 2mM Sodium Vanadate in RNAse-DNAse-free

PBS for five minutes, and transferred into ice-cold 1% paraformaldehyde in 2mM Sodium Vanadate for another 5 minutes. Finally, the slides were transferred to ice-cold 70% ethanol in RNAse free water (Ambion) and serially dehydrated to 100% EtOH for RNA-FISH or to ice-cold PBS with 2 mM sodium vanadate for immunostaining. Embryos at day 5 and day 6 of development were thawed for these experiments and cultured for 24 hours as described above before the staining protocol was started. All embryos analyzed were at the blastocyst state.

For the analysis of blastocyst outgrowths by RNA FISH with co-immunostaining, day 5 and day 6 embryos were thawed and, 24hrs later, at the blastocyst stage, the zona pellucida was removed with Tyrode's acid and embryos transferred onto feeder-coated MatTek dishes in derivation media containing 10% FBS and 10% KSR under normoxic conditions.

The combination of immunostaining with RNA FISH was performed as described (Tchieu et al., 2010). Antibodies for immunofluorescence used include NANOG (R&D, AF1997), cTNI (Abcam ab47003), POU5F1 (Santa Cruz, C10), H3K27me3 (Millipore, 07449), AFP (Sigma, SAB3300007), ALBUMIN (Sigma, A6684), and TUJ1 (Abcam).

To quantify staining patterns, at least 200 single cells distributed over 10 fields of view were inspected under the microscope for each gene and cell line. In all cases, experiments were performed at least in triplicate for RNA FISH and immunostainings of undifferentiated ESC and retinoic acid differentiation, and in duplicate for the directed differentiations. The data were highly reproducible and are shown for one representative experiment. In retinoic acid-differentiation samples, only OCT4-negative cells were quantified for their RNA FISH staining pattern, while in cardiomyocytes and hepatocytes only cTNI- and AFP/ALBUMIN- positive cells, respectively, were analyzed. For all RNA FISH experiments, the escapee *UTX* was typically detected in addition to one or two other X-linked genes and *XIST*, and only cells with biallelic *UTX* expression were scored. Notably, in UCLA1, which is karyotypically normal, *UTX* was not escaping XCI in some cells, and therefore, in this line, the *UTX* expression state was not considered when quantifying the expression patterns of other X-linked genes.

SNP/CNV Copy Number Array Analysis and Karyotyping

Genomic DNA from ESC lines and differentiated cardiomyocyte derivatives was extracted according to standard protocols. Genomic stability was analyzed on an Affymetrix genome-wide SNP 6.0 array as per manufacturers instructions at the UCLA microarray core. The analysis was performed as described (Greenman et al., 2010). For G-band karyotyping, ESCs grown on mouse feeders were plated in a T25 flask and submitted to Cell Line Genetics according to company instructions.

Reduced Representation Bisulfite Sequencing

Table S1 summarizes all samples that were analyzed by RRBS in this study and, for ESC and iPSC lines, the passage number. Genomic DNA was isolated using the DNEasy blood and tissue Kit (Qiagen) with Riboshredder treatment (Epicentre). RRBS libraries were generated as previously described with minor modifications (Orozco et al., 2015). DNA purifications were carried out using AMPure XP beads (Beckman Coulter) and bisulfite conversion was performed

twice using the Epitect kit (Qiagen) to optimize the efficiency. Bisulfite-converted libraries were amplified using MyTaq Mix (Bioline) with the following program: [98° for 15 sec, 60° for 30 sec, 72° for 30 sec] for 12 cycles, 72° for 5 min, 4°C storage. Methylation levels per CpG were called using BS-Seeker2 (2.0.32) using Bowtie for read alignment (Langmead et al., 2009).

DNA methylation analysis

Clustering and heatmap generation were performed using custom R scripts (http://www.r-project.org). Only CpG sites covered by at least 5 reads across all samples under consideration in a given analysis/figure were used, in an effort to obtain a reliable measurement of methylation levels. Methylation levels of samples were hierarchically clustered using the complete linkage and the Euclidean distance metric. In some plots (as indicated in the legends), CpG sites with constitutive hypermethylation or hypomethylation across all samples were discarded for presentation purposes to highlight changes in Xi-methylation. UCSC genome browser-defined CGIs were used to filter for CGI-associated CpGs.

To determine the relationship between CpG methylation erosion and CpG density on the X-chromosome, the standard deviation of methylation levels at CpGs covered at least 5 times across seven female ESC lines with various XCI states (ESI03, ESI02, H9, UCLA1, UCLA4, UCLA8, and UCLA9) was plotted against the CpG density in the surrounding 200nt window of each CpG. For the comparison of XCI state and differentiation propensity, X-linked CGI methylation data for a large number of female ESC and iPSC lines were obtained from (Bock et al., 2011) and processed as described above. In addition, the median methylation levels of all CpGs (with 5X coverage across lines) falling within CGIs were plotted against the lineage score described for the same cell lines (Bock et al., 2011). The methylation state of the X-linked genes *ATRX* and *HUWE1* was based on the methylation level of CpGs with 5X coverage across all cell lines tested, which fell within the CGI that most closely neighbored the respective TSS (regions considered: *ATRX* (chrX: 77041000 –77041800) and *HUWE1* (chrX:53710300 – 53713800)). Imprinted differentially methylated regions classified as maternal and paternal control regions (Okae et al., 2014).

Expression analysis

For gene expression analysis of ESC and cardiomyocyte-derivative cells, RNA was extracted under feeder-free conditions. For expression analysis of the UCLA8 and UCLA9, and UCLA8 subclones, libraries were generated using the standard strand-specific Illumina Tru-seq v2 RNA-Seq workflow and sequenced as single end 50 bp reads. For RNA-seq of all other ESC lines and cardiomyocytes, libraries were built by integrating an unique molecular identifier tagged to the 5' end using a custom-made PAGE-purified template switching oligo (TSO, 5'-AAGCAGTGGTATCAACGCAGAGTNNNNNNNNACATrGrGrG -3'). Reverse transcription (RT) was initiated by annealing 10ng of purified RNA with a custom 5' biotinylated oligodT primer with conserved 5' sequence to the TSO oligo (Biotin-GAAGCAGTGGTATCAACGCAGAGTACT(30)VN; V is A, C or G), using SMARTScribe reverse transcriptase (Clontech) according to the manufacturers instructions. Reverse transcription products were amplified by PCR with the primer AAGCAGTGGTATCAACGCAGAGT and subsequently tagmented using the NEXTERA XT kit (Illumina). Regardless of how libraries were

built, reads were mapped using TopHat v2.0.13 and assigned to genes using HTSeq-0.6.1. X-chromosome to autosome (X:A) expression ratios were determined based on median quantile-normalized RPKM expression values. Downstream analysis, including plotting of cumulative distribution functions and significance testing, was performed using R (http://www.r-project.org). Quantile normalization was applied prior to plotting CDFs.

To assign genes as 'eroded' or 'inactive' in a given female ESC line for gene expression analysis (via CDFs), we used a procedure that involved three steps: First, we filtered CpG sites using the following criteria: (i) location within CGIs; (ii) 5X coverage across the eight ESC lines ESI03, ESI02, H9, UCLA1, UCLA4, UCLA8, UCLA10, and UCLA9 (=seven female and one male line); (iii) not being constitutively highly (≥ 0.85) and lowly (≤ 0.15) methylated; (iv) be intermediately methylated (0.2 ≥ mC ≤ 0.6) in the un-eroded reference XiXa line ESI03; (v) be unmethylated (≤ 0.2) in the male ESC line UCLA10. Second, we defined CGIs that displayed erosion or maintenance of XCI in a given cell line by determining the mean methylation level within each CGI with at least one covered CpG site. Any CGI with an average methylation level < 0.15 was considered eroded (ie. active) and with average methylation level ≥ 0.15 deemed to be silenced. Third, X-linked UCSC genome browser-based RefSeq genes were classified as eroded if their promoters (TSS +/- 2kb) overlapped eroded CGIs and as inactive if their promoters overlapped with inactive CGIs. Once eroded and inactive genes were defined for a particular cell line, the expression of these two sets of genes was analyzed in the respective cell line and other lines via CDFs.

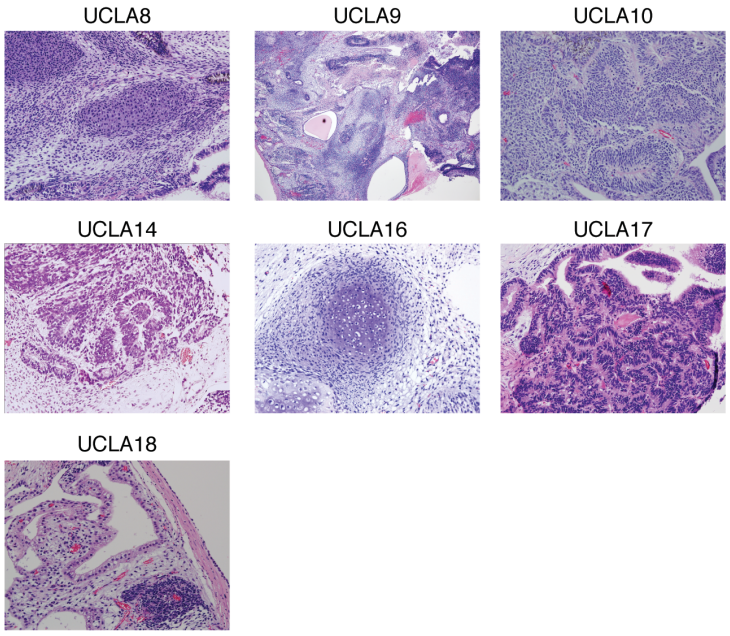
Public availability of data

All genome-wide data are available at GEO: GSE88933.

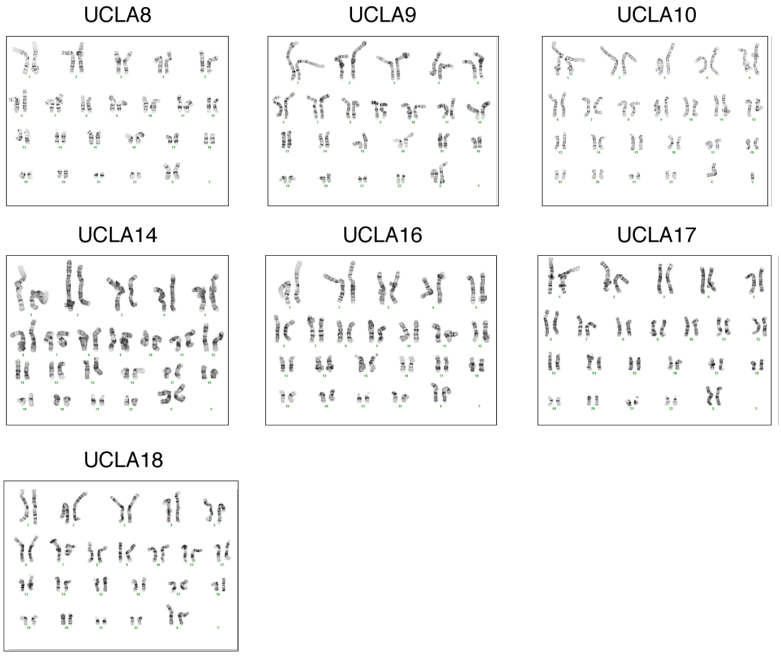
II. SUPPLEMENTAL FIGURES AND FIGURE LEGENDS

Supplemental Figure 1

A



B



Supplemental Figure 1. Characterization of newly derived UCLA ESC lines (Related to Figure 1)

A) Representative sections of hematoxylin and eosin staining from teratomas formed by indicated newly derived female UCLA ESC lines displaying differentiation into all three germ layers.

B) Representative chromosome spreads of the indicated ESC lines as in (A) showing normal karyotype.

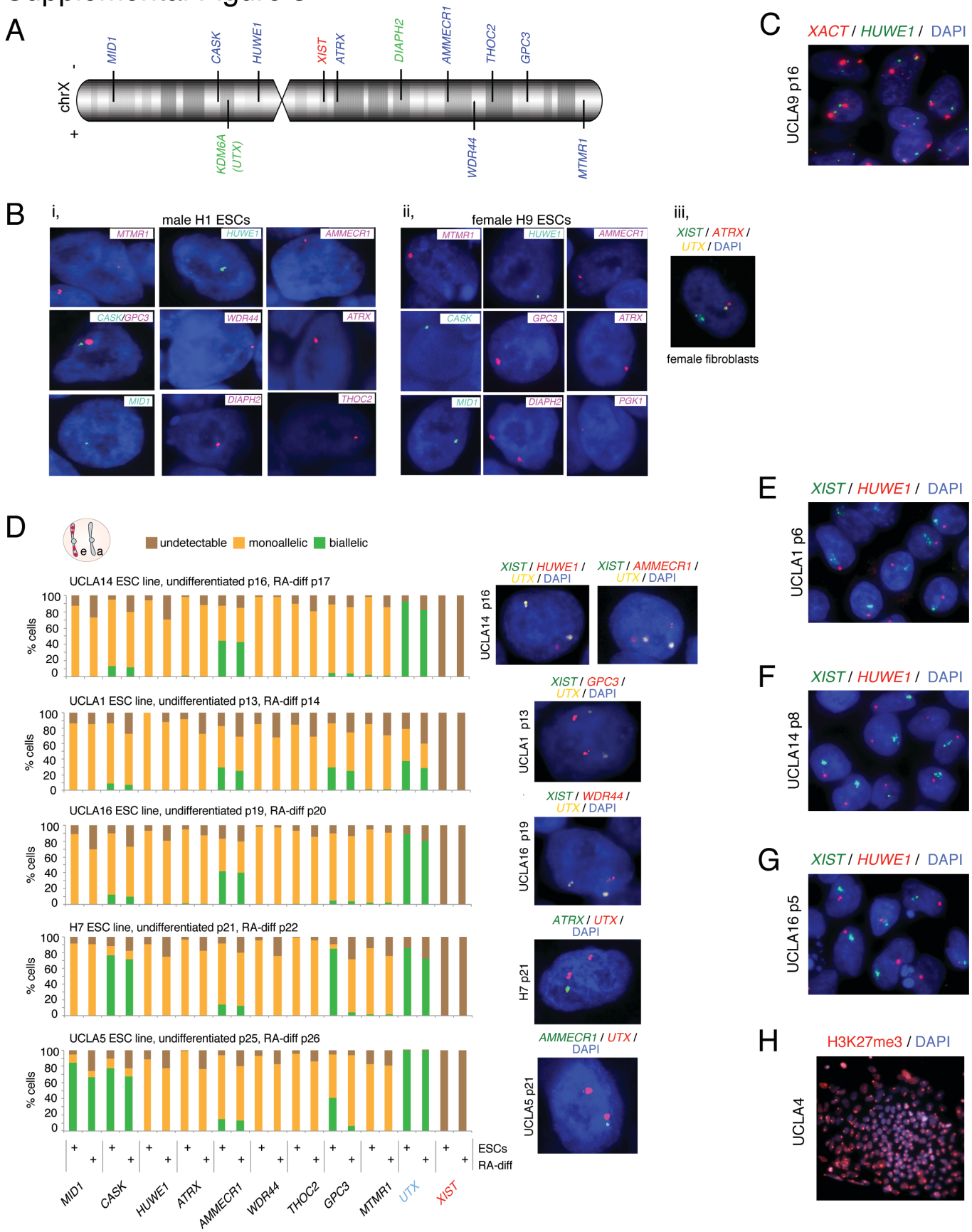
Supplemental Figure 2

Supplemental Figure 2. Confirmation of normal X-chromosome count of undifferentiated ESCs and ESC-derived cardiomyocytes (Related to Figure 1)

A) Genome-wide CNV arrays were analyzed for each chromosome in the undifferentiated XaXa ESC line UCLA9, the strongly eroded XeXa ESC line UCLA4, and in cardiomyocytes obtained from these ESC lines. In each chromosome-wide plot, presented as described in Greenman et al. (Greenman et al., 2010), the top row depicts copy number intensities with green and blue lines indicating total and minor estimated copy number. The three rows below give the genotype intensities with black and red lines representing heterozygous and homozygous segments, respectively. The graph at the bottom depicts the likelihood of state change. Horizontal scale gives the genomic position along the indicated chromosome in mega-bases. These plots demonstrate that the two X-chromosomes are present in these cell lines and do not display any major genomic abnormalities.

B) As in (A), except for only the X-chromosome in the undifferentiated ESC line UCLA1 (weakly eroded XeXa) and its cardiomyocyte derivative. All other chromosomes were normal but are not displayed here due to space restrictions.

Supplemental Figure 3



Supplemental Figure 3. Characterization of the XCI-state of ESC lines and their differentiation products at single cell resolution by multi-gene RNA FISH (Related to Figures 1 and 4)

A) Schematic diagram depicting the locations of genes on the X-chromosome analyzed by the multi-gene RNA FISH approach in our study. Genes normally subject to XCI are highlighted in blue, *XIST* in red, and the escapees *UTX* and *DIAPH2* in green (Balaton et al., 2015).

B) Representative RNA FISH images for X-linked genes shown in (A) (i) in male ESC line H1 demonstrating their expected mono-allelic expression pattern; (ii) in the female XiXa ESC line H9 demonstrating mono-allelic expression of genes known to be subject to XCI and bi-allelic expression for the gene *DIAPH2* which escapes XCI; and (iii) in female human $Xi^{XIST+}Xa$ fibroblasts demonstrating mutually exclusive mono-allelic expression of *XIST* and *ATRX* from the Xi and Xa, respectively, and bi-allelic expression of the escapee gene *UTX*.

C) Representative RNA-FISH image of the X-linked lncRNA *XACT* (red) and nascent transcription foci of the X-linked gene *HUWE1* (green) in UCLA9 ESCs (XaXa) at passage 16.

D) Analysis of additional XeXa ESC lines by multi-gene RNA FISH similar to Figure 1C. Quantification of the expression pattern of nine X-linked genes that are normally subject to XCI, of the X-linked gene *UTX* that in most cell lines escapes XCI, and of *XIST* in UCLA14, UCLA1, UCLA16, UCLA5, and H7 in the undifferentiated state at the indicated passage (p) (after at least one freeze-thaw cycle since derivation) and after 10 days of retinoic acid-induced differentiation, induced from the same ESC culture at the indicated passage. Nascent transcription foci of X-linked genes and *XIST* signals were scored as mono-allelic, bi-allelic, and undetectable. Note, for UCLA1, which is karyotypically normal, *UTX* was not bi-allelically expressed in all cells despite the fact that it usually escapes XCI and UCLA1 is karyotypically normal (Figure S2). Representative images of RNA FISH staining for various genes in undifferentiated ESCs are given.

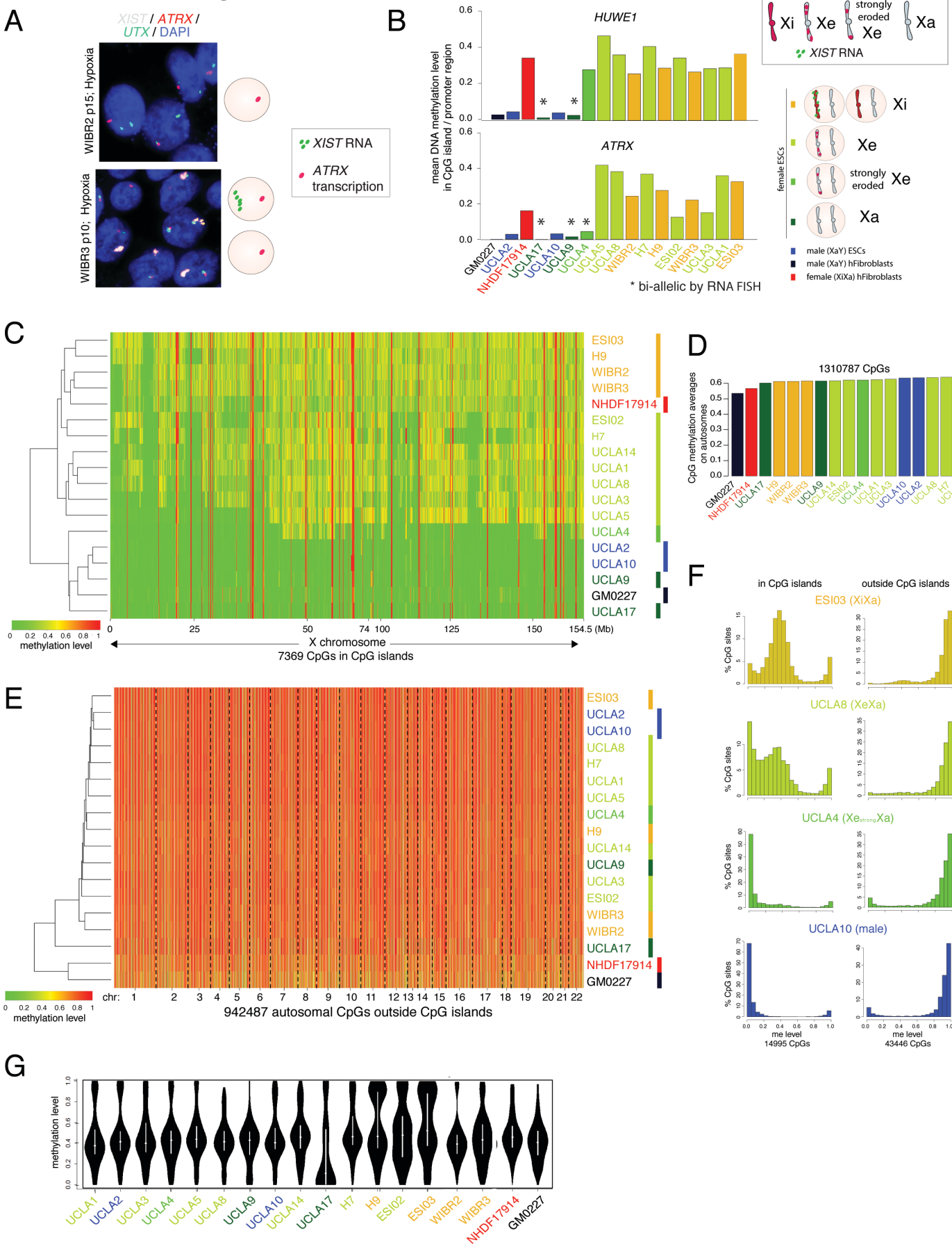
E) Representative RNA FISH image for *XIST* and nascent transcription foci of the X-linked gene *HUWE1* for UCLA1 at early passage, indicating that this line was $Xi^{XIST+}Xa$ at early passage.

F) As in (E), except for UCLA14.

G) As in (E), except for UCLA16.

H) Representative immunostaining image for H3K27me3 for the ESC line UCLA4 at early passage, revealing an Xi-like accumulation in a large number of cells, suggesting that UCLA4 was $Xi^{XIST+}Xa$ at early passage.

Supplemental Figure 4



Supplemental Figure 4. Analysis of XCI state by DNA methylation (Related to Figure 2)

A) Representative RNA FISH images for *XIST*, nascent transcription foci of the X-linked gene *ATRX*, and the X-linked escapee *UTX* for the ESC lines WIBR2 and WIBR3. The schemes to the right depict the most prominent XCI-state(s).

B) Mean methylation level in CGIs most closely associated with the promoter region of the X-linked genes *HUWE1* and *ATRX* in the indicated cell lines, showing that CGI methylation correlates well with the RNA FISH results for mono- and biallelic expression found in Figures 1 and S3D. In fact, the CGI of *HUWE1* is intermediately methylated in all female ESC lines with the exception of the XaXa lines UCLA9 and UCLA17. Conversely, the CGI of *ATRX* lacks methylation in the XaXa lines and, additionally, in the strongly eroded line UCLA4. Asterisks mark those female ESC lines that displayed biallelic expression of *HUWE1* and *ATRX*, respectively, based on the RNA FISH analysis in Figures 1 and S3D.

C) Heatmap of unsupervised hierarchical clustering of RRBS-based methylation levels for CpGs within X-linked CGIs in indicated female and male ESC lines and fibroblast controls (GM0227 and NHDF17914, respectively). This heatmap is similar to that shown in Figure 2A, except that CpGs shown in this heatmap were not filtered for constitutively lowly and highly methylated sites.

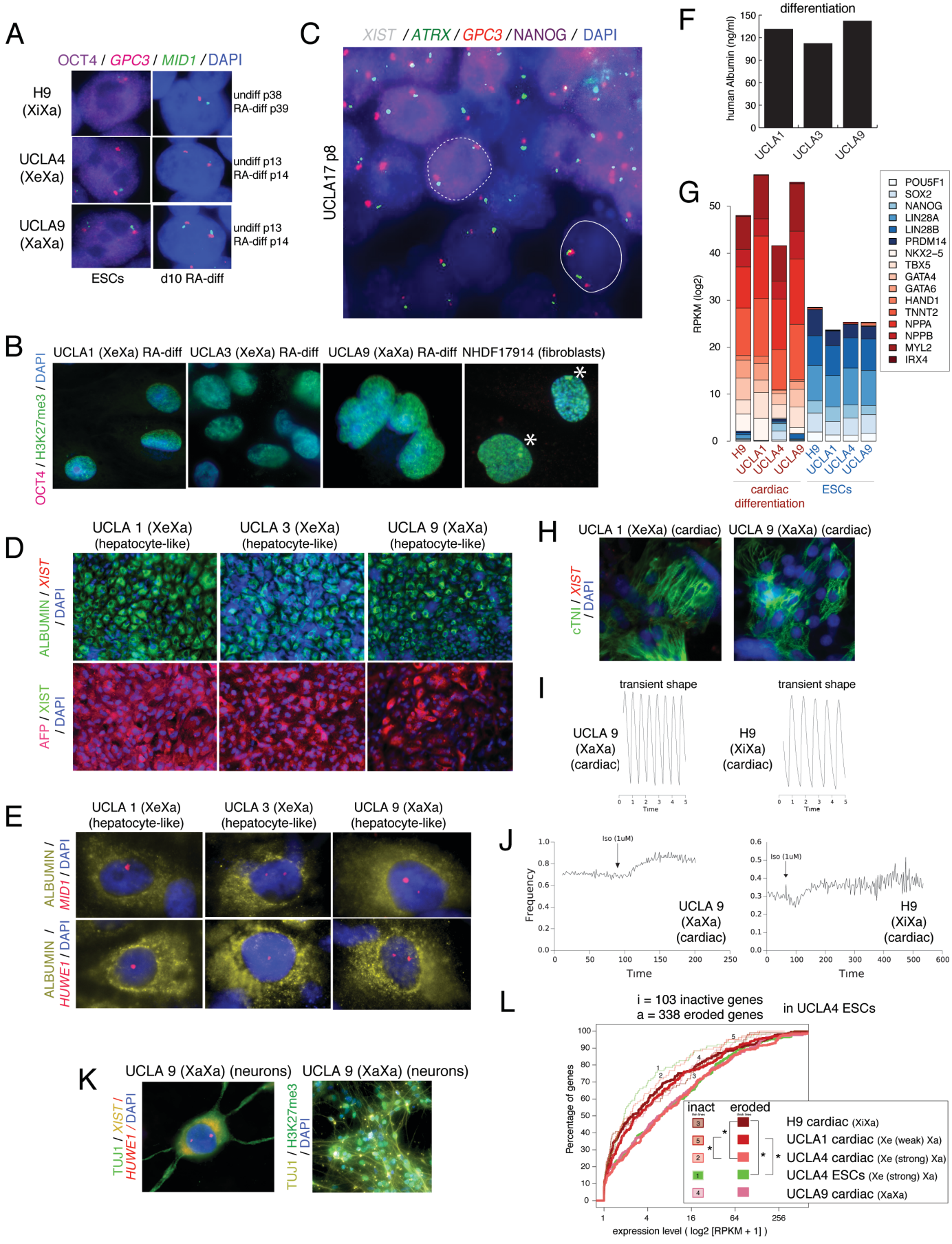
D) Mean methylation level across all autosomal CpGs in the same cell lines as shown in (B and C).

E) As in (C), except for all autosomal CpGs, excluding those in CGIs.

F) Histograms displaying methylation level distributions for X-linked CpGs inside and outside of CGIs for the XiXa line ESI03, the weakly eroded XeXa line UCLA8, and strongly eroded XeXa line UCLA4, and the male ESC line UCLA10 for comparison, demonstrating the specific erosion of CGI-linked methylation.

G) Violin plot of methylation levels for CpGs in 31 primary imprinted control regions that are known to be differentially methylated in the human blastocyst (see Supplemental Experimental Procedures), indicating variation of imprint methylation across ESC lines, but no correlation with XCI-state.

Supplemental Figure 5



Supplemental Figure 5. Maintenance of the XCI state pre-existing in ESCs upon differentiation (Related to Figures 1 and 4)

A) Representative images of RNA FISH analysis for nascent transcription foci of the X-linked genes *MID1* and *GPC3*, combined with immunostaining for the pluripotency marker OCT4 in indicated ESC lines in their undifferentiated state (OCT4-positive) and ten days after induction of retinoic acid (RA)- induced differentiation (OCT4-negative). These stainings were used to evaluate the state of the X-chromosome in OCT4-negative cells upon RA-induced differentiation shown in Figures 1 and S3D.

B) Representative images of H3K27me3 and OCT4 co-immunostainings on indicated ESC lines at day 10 of RA-differentiation, demonstrating the lack of an Xi-like H3K27me3 accumulation in OCT4-negative cells. Female fibroblasts (NHDF17914) served as positive control for the Xi-like accumulation of H3K27me3 (marked by asterisks).

C) Cells spontaneously differentiating in the XaXa ESC line UCLA17 under self-renewing culture conditions did not display XCI. Representative image of RNA FISH staining for *XIST* and the X-linked genes *ATRX* and *GPC3*, co-immunostained for the pluripotency marker NANOG. DAPI staining detected nuclei. The dashed line indicates an example of a NANOG-positive nucleus and the solid line highlights a NANOG-negative nucleus. Both NANOG-positive and -negative cells displayed biallelic expression of the X-linked genes *ATRX* and *GPC3* and lacked *XIST* expression.

D) Representative images of ALBUMIN and AFP immunostaining, respectively, in combination with RNA FISH for *XIST* on hepatocyte-like cells from indicated ESC lines. ALBUMIN and AFP staining indicated that most cells in the culture were differentiated to the hepatocyte-like state (>95%). *XIST* staining was not observed in differentiated hepatocytes. Similarly, an accumulation of the *XIST*-dependent Xi-mark H3K27me3 was not found (not shown).

E) As in (D), except that ALBUMIN immunostaining was combined with RNA FISH for the X-linked genes *MID1* and *HUWE1*, respectively. *MID1* was expressed mono-allelically in UCLA1 and biallelically in UCLA3 and UCLA9, while *HUWE1* was mono-allelic in UCLA1 and UCLA3 and biallelic in UCLA9 (Figures 1 and S3D). The same expression states are maintained in hepatocytes derived from these ESC lines.

F) ALBUMIN concentration in the culture media was measured 24 hours after adding fresh media to indicated ESC-derived hepatocyte cultures, displaying levels expected for ESC-derivatives.

G) Expression level of pluripotency and cardiomyocyte markers based on RNA-seq data for the four indicated ESC lines (H9=XiXa, UCLA1=XeXa, UCLA4=strongly eroded XeXa, and UCLA9=XaXa) and their cardiomyocyte derivatives. The data demonstrated that silencing of pluripotency genes had occurred in the differentiated culture, arguing against any significant contribution from pluripotent cells towards the results in differentiated cells.

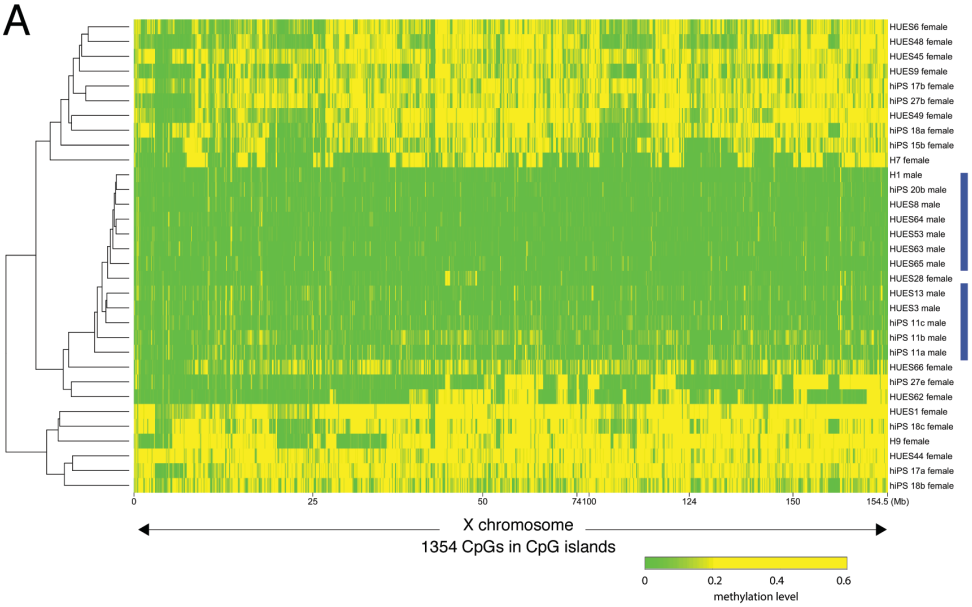
H) Representative images of cTNI immunostaining in combination with RNA FISH for *XIST* on cardiomyocyte-like cells derived from the XeXa ESC line UCLA1 and the XaXa ESC line UCLA9. Most cells in the culture were positive for cTNI, with a striated pattern. Moreover, the culture displayed spontaneous beating in large patches of cells. *XIST* staining could not be observed in differentiated cardiomyocytes.

I) Calcium transient waveforms of XaXa UCLA9- and XiXa H9-derived cardiomyocytes, supporting cardiomyocyte identity of these cells.

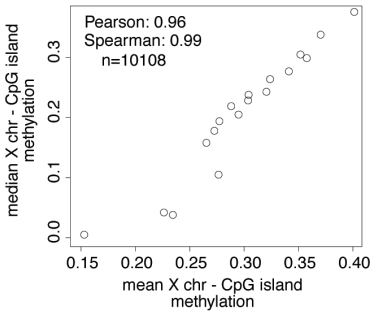
J) As in (I), except for beat frequency at baseline and after addition of 1uM isoproterenol, showing an increase in beating frequency in response to beta-adrenergic stimulation.
K) Differentiation of the XaXa line UCLA9 into neurons detected by TUJ1 immunostaining. Combination with RNA FISH for *XIST*/*HUWE1* and immunostaining for H3K27me3, respectively, indicated the maintenance of biallelic *HUWE1* expression and the absence of *XIST*/H3K27me3 in UCLA9-derived neurons.
L) Eroded and inactive genes in the strongly eroded line UCLA4 were classified based on the methylation level of promoter-associated CGIs in the undifferentiated state, and the cumulative distribution function for the expression of eroded (thick lines) and inactive (thin lines, numbered to help to distinguish the lines) X-linked gene sets plotted for undifferentiated UCLA4 and UCLA4-derived cardiomyocytes. In addition, for the same two gene sets, CDFs for gene expression values in cardiomyocytes differentiated from the XaXa line UCLA9, the XeXa line UCLA1, and the XiXa line H9 are given. * = ks-test p-value < 0.0005 for the difference in the gene expression function for the eroded gene set between UCLA4 ESCs and UCLA4 cardiomyocytes, respectively, and the various other cardiomyocyte lines, indicating significantly higher expression of these genes in UCLA4 ESCs and cardiomyocytes compared to H9- and UCLA1-derived cardiomyocytes, and similar expression levels as in the XaXa ESC line UCLA9.

Supplemental Figure 6

A



B

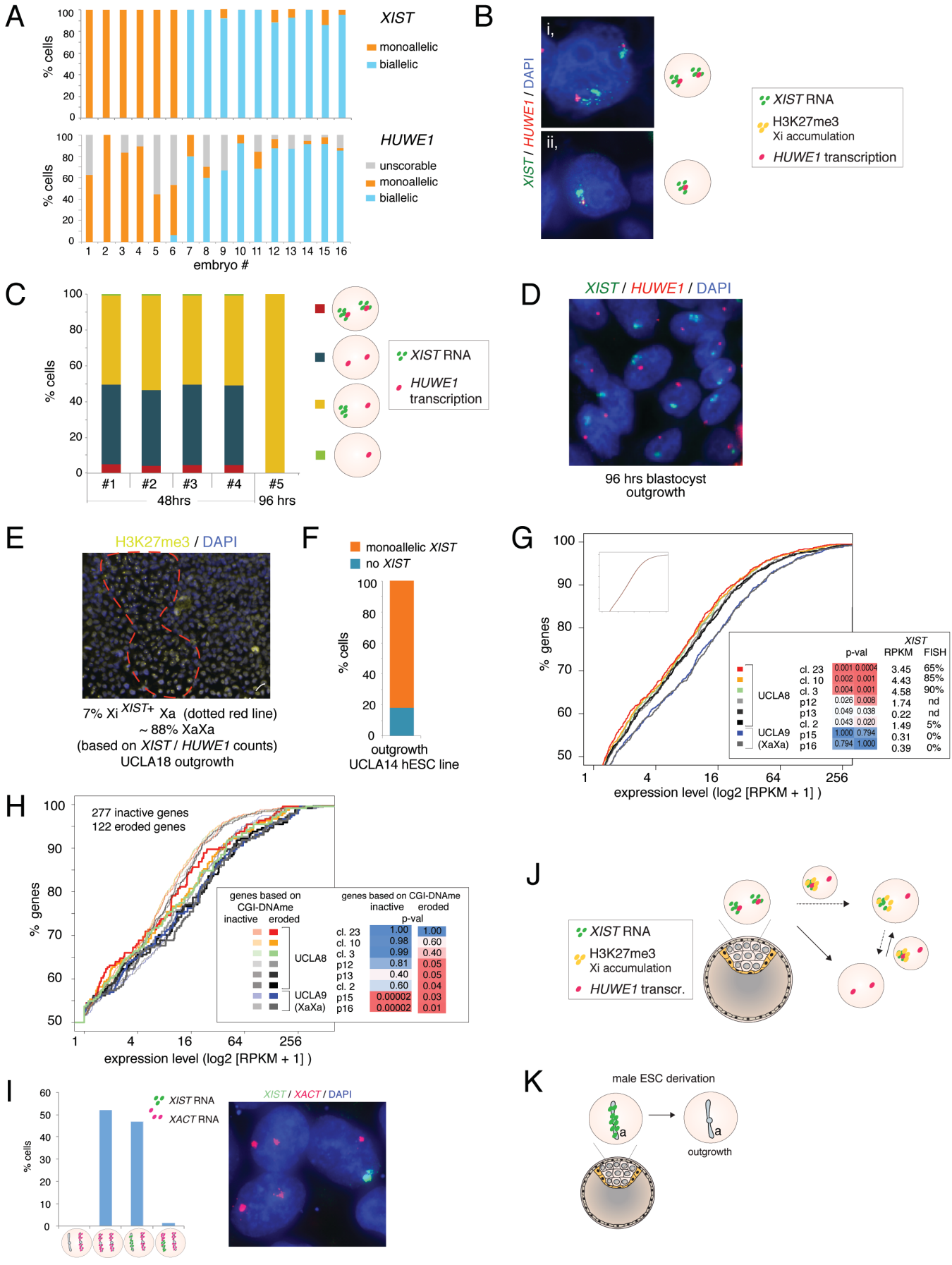


Supplemental Figure 6. Characterization of XCI status in a large set of human ESCs and iPSCs (Related to Figure 5)

A) Heatmap of unsupervised hierarchical clustering of RRBS-based methylation levels for CpGs within X-linked CGIs in female ESC and iPSC lines shown in Figure 5A and, additionally, in male ESC and iPSC lines (marked by the blue line) from the same study (Bock et al., 2011). This analysis reveals that among all cell lines only the female ESC line HUES28 carries the low methylation pattern of males, and that iPSC27e is strongly eroded. CpGs with constitutive low (<0.2) and high (>0.6) methylation were not included.

B) Comparison of mean and median CpG methylation levels within CGIs of the female ESC and iPSC lines (Bock et al., 2011) shown in Figure 4. 10108 CpG sites were considered.

Supplemental Figure 7



Supplemental Figure 7. Changes in XCI-state during ESC derivation and early passages of established ESC lines (Related to Figures 6 and 7)

A) Quantification of the *XIST* expression pattern and nascent transcription sites of the X-linked gene *HUWE1* in cells of 16 human blastocysts based on RNA FISH. Signals for *XIST* and *HUWE1* were scored independently, and, for *XIST* counts, only cells with either a mono- or bi-allelic *HUWE1* signal were considered. Blastocysts with bi-allelic *XIST* accumulation predominantly carried bi-allelic *HUWE1* signals and those with mono-allelic *XIST* displayed mono-allelic *HUWE1* transcription foci, suggesting that embryos 1-6 were male and 7-16 female. On average, 27 cells were analyzed per embryo.

B) Images of the two most representative RNA FISH patterns of *XIST*/*HUWE1* expression for male (i) and female (ii) blastocysts observed in (A) and summarized by the scheme to the right. DAPI staining (blue) detected nuclei.

C) Quantification of RNA FISH staining patterns for *XIST* RNA and nascent transcription sites of the X-linked gene *HUWE1* in cells of female blastocysts at 48 and 96 hours after plating onto a feeder layer in ESC culture media. The classification of these blastocysts as female was based on the *HUWE1*/*XIST* pattern in cells of the blastocyst/outgrowths, indicating the presence of two X-chromosomes. These results were obtained from a derivation experiment performed in addition to the one shown in Figure 6B. Compared to Figure 6B, in the experiment shown here cells had already grown out of the blastocyst more clearly at 48 hours.

D) Representative RNA FISH image of the outgrowth from blastocyst #5 from (C).

E) Immunostaining for H3K27me3 on the blastocyst outgrowth that yielded the ESC line UCLA18 at three weeks in the derivation culture, obtained immediately after cells for establishment of the UCLA18 ESC line were separated. The red encircled region highlights a patch of $X^{H3K27me3+}X$ cells, indicative of proper X-inactivation in these cells. Below the white dotted line (right bottom corner), cells were taken for the establishment of the ESC line UCLA18, away from the Xi- positive patch. Quantification of the XCI-state in the outgrowth is given below and based on *XIST*/*HUWE1* RNA FISH, which is also given in Figure 6Cii (first bar).

F) Quantification of *XIST*-positive cells in the outgrowth of the blastocyst that yielded the ESC line UCLA14, based on RNA FISH, after cells were removed from the outgrowth for the establishment of the ESC line UCLA14 (accompanying Figure 6Di).

G) Relationship between *XIST* loss and Xi-erosion. To analyze the relationship between *XIST* loss and Xi erosion, we sub-cloned the ESC line UCLA8, which predominantly displayed the $Xi^{XIST+}Xa$ state at early passage (Figure 6Dii), at p9 when *XIST*-expressing cells were still present. Four sub-clones (cl. 2, 3, 10, 23) were expanded and analyzed by RNA FISH for *XIST*. The parent line UCLA8 at p12/p13 (when it largely was *XIST*-negative) and the four sub-clones were profiled by RNA-seq together with the XaXa ESC line UCLA9 at p15 and p16. The *XIST* numbers to the right in the inset give the number of cells with an Xi-like *XIST* accumulation based on RNA FISH and on the expression level from RNA-seq data (RPKM) (nd=not determined), showing different proportions of *XIST*+ cells in the four sub-clones. These findings indicate that the loss of *XIST* occurs progressively in these cells and is not due to selection of *XIST*-negative cells. The plot depicts the cumulative distribution function (CDF) of X-linked gene expression for these lines/sub-clones and the inset shows the autosomal expression data at the same XY-scales (see Supplemental Experimental Procedures). The two p-value columns give the ks-test result for the difference of X-linked gene expression distributions between UCLA9

p15 (left) and UCLA9 p16 (right), respectively, and all other lines/clones, indicating that X-linked genes in UCLA8 sub-clones are more lowly expressed when *XIST* is still expressed in the majority of cells. Thus, UCLA8 sub-clones carrying a higher proportion of *XIST*-positive cells exhibited lower expression of X-linked genes compared to those largely lacking *XIST*. These results are most consistent with Xi-erosion occurring at around the time of *XIST* RNA loss.

H) Xi-erosion leads to higher gene expression. X-linked genes with were divided into those where XCI is eroded and those that display XCI in the ESC line UCLA8, based on CGI methylation obtained from UCLA8 at a time after *XIST* was lost completely in the culture and the Xe already formed (Fig 2A). Plotted is the cumulative distribution function for X-linked gene expression based on RNA-seq data for ESC lines and sub-clones described in (G), for the inactive (thin lines) and eroded (thick lines) gene sets of UCLA8. The results of the ks-test for the difference in distribution between the *XIST*-positive UCLA8 sub-clone 23 data and all others cell lines/sub-clones are given. These data indicated that inactive genes were significantly lower expressed in UCLA8 and its sub-clones compared to the XaXa UCLA9 line, consistent with the fact that these genes have maintained the inactive state in UCLA8 and its sub-clones, regardless of *XIST* RNA expression. Conversely, genes that displayed erosion in UCLA8 at later passage (thick lines) were expressed at a lower level in the *XIST*-positive UCLA8 sub-clones than in UCLA8 sub-clones/lines that lacked *XIST* or the XaXa UCLA9 lines. Thus, Xi-erosion occurred at around the time of *XIST* loss and leads to higher expression of the affected genes as they are now active on both X-chromosomes.

I) Quantification of the expression patterns of *XIST* and the X-linked lncRNA *XACT* (Vallot et al., 2015) in early passage human iPSCs (p6) based on RNA FISH. The image is representative of the RNA FISH result showing that *XACT* was only bi-allelically expressed in this line during propagation when *XIST* silencing had occurred.

J) Scheme summarizing the changes of the X-chromosome in female blastocyst outgrowths similar to Figure 7D, except for the addition of the $Xa^{XIST+/H3K27me3+}Xa$ state, which likely is an intermediate in the transition to the $Xi^{XIST+/H3K27me3+}Xa$ state as shown in differentiating mouse ESCs, where *Xist* RNA coating and H3K27me3 accumulation occur before silencing.

K) Scheme summarizing the changes of the X-chromosome in male blastocyst outgrowths, indicating silencing of *XIST* in derivation conditions soon after blastocyst plating.

III. SUPPLEMENTAL TABLE LEGENDS

Supplemental Table 1. Summary of characteristics of and analyses performed for various ESC and iPSC lines (Related to Figures 1, 2, 3, 4 and 7)
Summary of characteristics and analyses performed for ESC and iPSC lines in this study. Related to Figures 1-7.

Supplemental Table 2. Gene level analysis of XCI-state based on DNA methylation (Related to Figure 3)
Gene level analysis of XCI-state based on DNA methylation levels for the following female ESC lines: ESI02, ESI03, H9, UCLA1, UCLA4, UCLA8, UCLA9, and UCLA10. The table gives the mean DNA methylation level of CGIs overlapping the promoter region (TSS +/- 2kb) of X-linked RefSeq genes as a measure of XCI-state. Low methylation indicates the active state and intermediate methylation the silent state of the X. Related to Figures 2 and 3.

IV. 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.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H.*, et al.* (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439-452.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol *27*, 275-280.

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.

Greenman, C.D., Bignell, G., Butler, A., Edkins, S., Hinton, J., Beare, D., Swamy, S., Santarius, T., Chen, L., Widaa, S.*, et al.* (2010). PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. Biostatistics *11*, 164-175.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

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.

Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L.B., Azarin, S.M., Raval, K.K., Zhang, J., Kamp, T.J., and Palecek, S.P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proceedings of the National Academy of Sciences of the United States of America *109*, E1848-1857.

Namekawa, S.H., Payer, B., Huynh, K.D., Jaenisch, R., and Lee, J.T. (2010). Two-step imprinted X inactivation: repeat versus genic silencing in the mouse. Molecular and cellular biology *30*, 3187-3205.

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.

Orozco, L.D., Morselli, M., Rubbi, L., Guo, W., Go, J., Shi, H., Lopez, D., Furlotte, N.A., Bennett, B.J., Farber, C.R.*, et al.* (2015). Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. Cell Metab *21*, 905-917.

Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.S., Sritanaudomchai, H.*, et al.* (2013). Human embryonic stem cells derived by somatic cell nuclear transfer. Cell *153*, 1228-1238.

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.

Tomoda, K., Takahashi, K., Leung, K., Okada, A., Narita, M., Yamada, N.A., Eilertson, K.E., Tsang, P., Baba, S., White, M.P.*, et al.* (2012). Derivation conditions impact X-inactivation status in female human induced pluripotent stem cells. Cell Stem Cell *11*, 91-99.

Vallot, C., Ouimette, J.F., Makhlouf, M., Feraud, O., Pontis, J., Come, J., Martinat, C., Bennaceur-Griscelli, A., Lalande, M., and Rougeulle, C. (2015). Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. Cell Stem Cell *16*, 533-546.