Stem Cell Reports, Volume *10*

Supplemental Information

X Chromosome Dosage Influences DNA Methylation Dynamics during Reprogramming to Mouse iPSCs

Vincent Pasque, Rahul Karnik, Constantinos Chronis, Paula Petrella, Justin Langerman, Giancarlo Bonora, Juan Song, Lotte Vanheer, Anupama Sadhu Dimashkie, Alexander Meissner, and Kathrin Plath

Figure S1

A

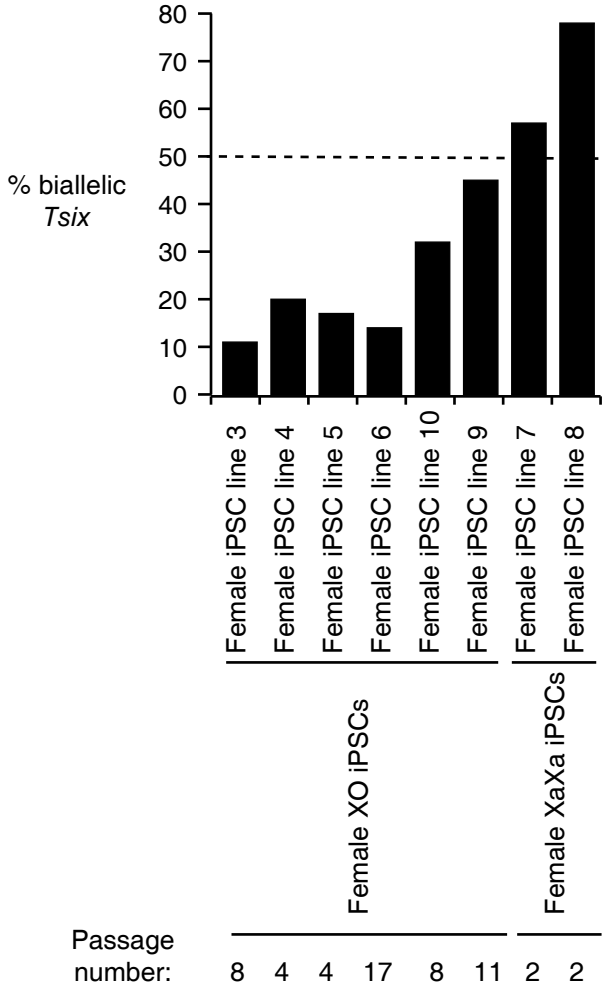
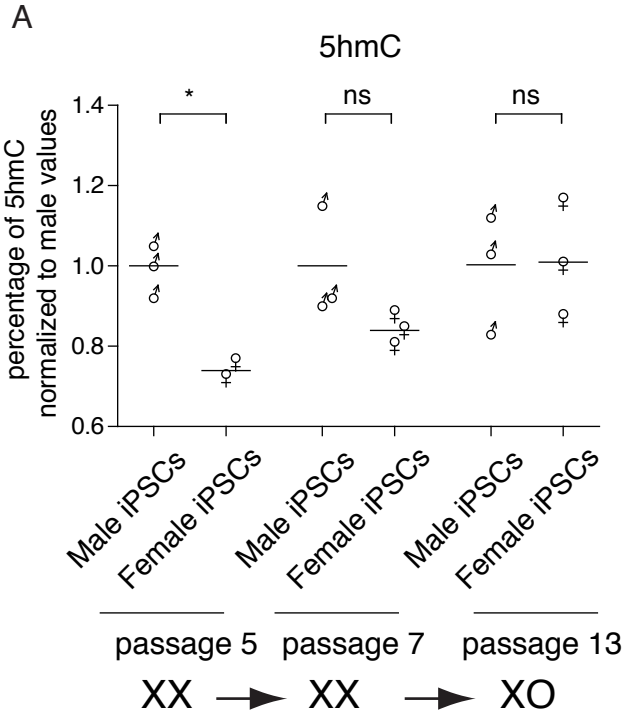
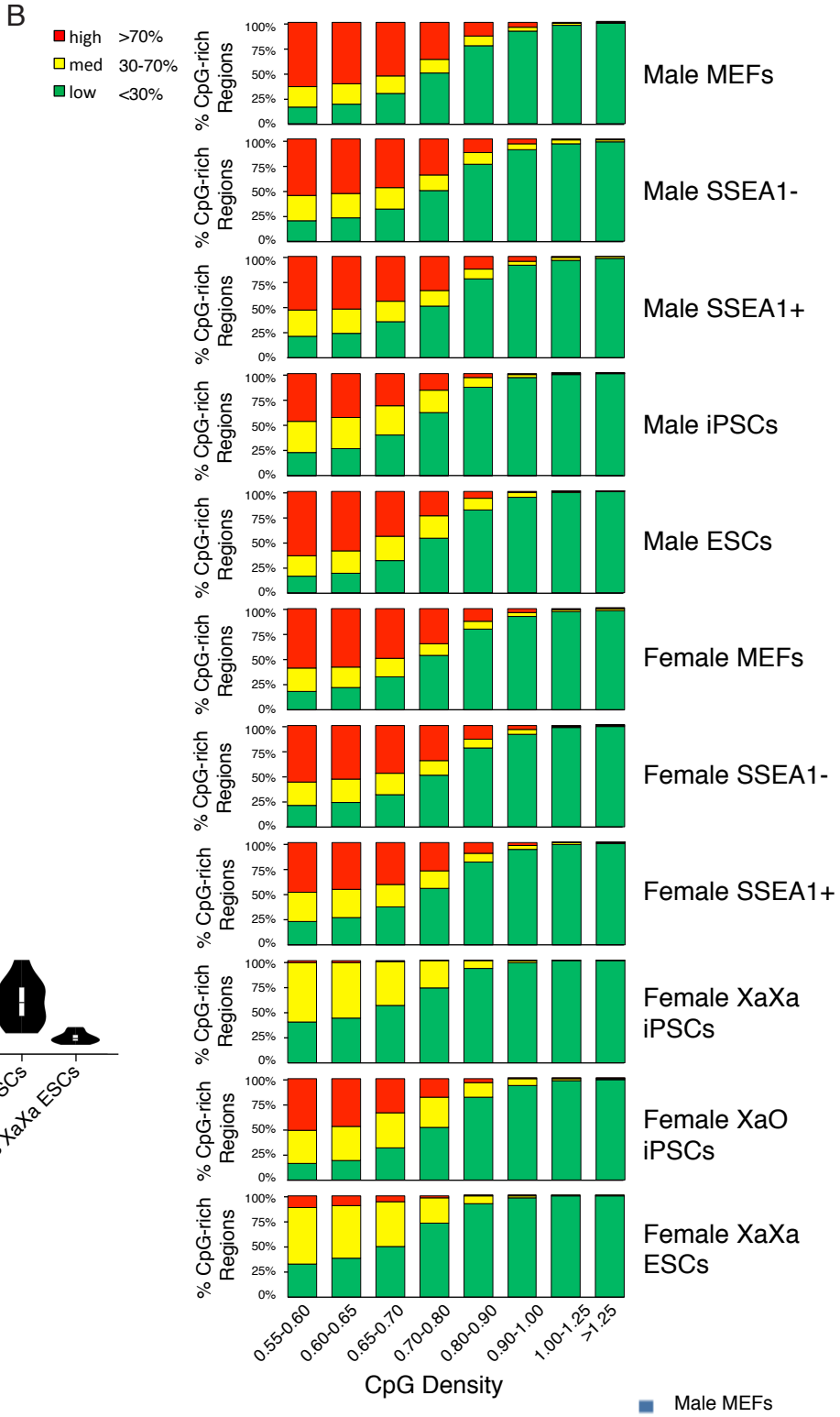
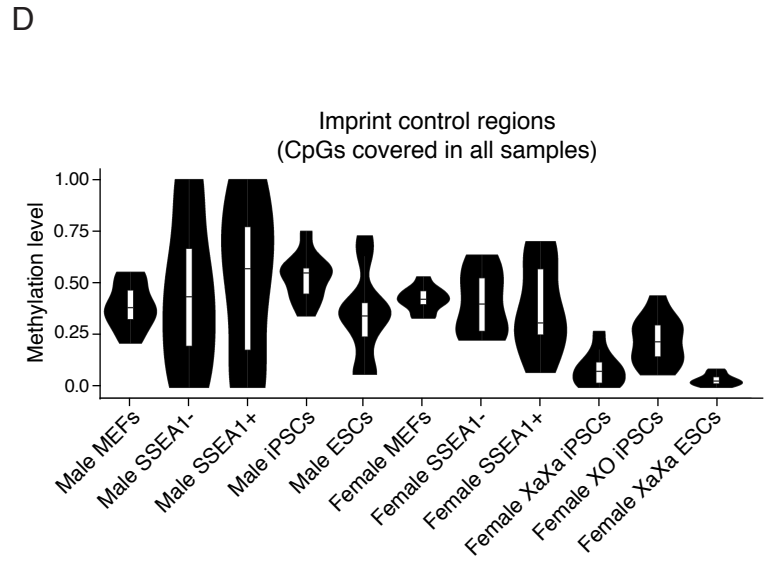

80
70
60
50
40
30
20
10
0
% biallelic
Tsix
Female iPSC line 3
Female iPSC line 4
Female iPSC line 5
Female iPSC line 6
Female iPSC line 10
Female iPSC line 9
Female iPSC line 7
Female iPSC line 8
Female XO iPSCs
Female XaXa iPSCs
Passage
number:
8 4 4 17 8 11 2 2

Figure S2


A
5hmC
percentage of 5hmC normalized to male values
1.4
1.2
1.0
0.8
0.6
*
ns
ns
Male iPSCs
Female iPSCs
Male iPSCs
Female iPSCs
Male iPSCs
Female iPSCs
passage 5
passage 7
passage 13
XX → XX → XO


B
high >70%
med 30-70%
low <30%
% CpG-rich Regions
100%
75%
50%
25%
0%
Male MEFs
Male SSEA1-
Male SSEA1+
Male iPSCs
Male ESCs
Female MEFs
Female SSEA1-
Female SSEA1+
Female XaXa iPSCs
Female XaO iPSCs
Female XaXa ESCs
0.55-0.60
0.60-0.65
0.65-0.70
0.70-0.80
0.80-0.90
0.90-1.00
1.00-1.25
>1.25
CpG Density


D
Imprint control regions
(CpGs covered in all samples)
Methylation level
1.00
0.75
0.50
0.25
0.0
Male MEFs
Male SSEA1-
Male SSEA1+
Male iPSCs
Male ESCs
Female MEFs
Female SSEA1-
Female SSEA1+
Female XaXa iPSCs
Female XO iPSCs
Female XaXa ESCs

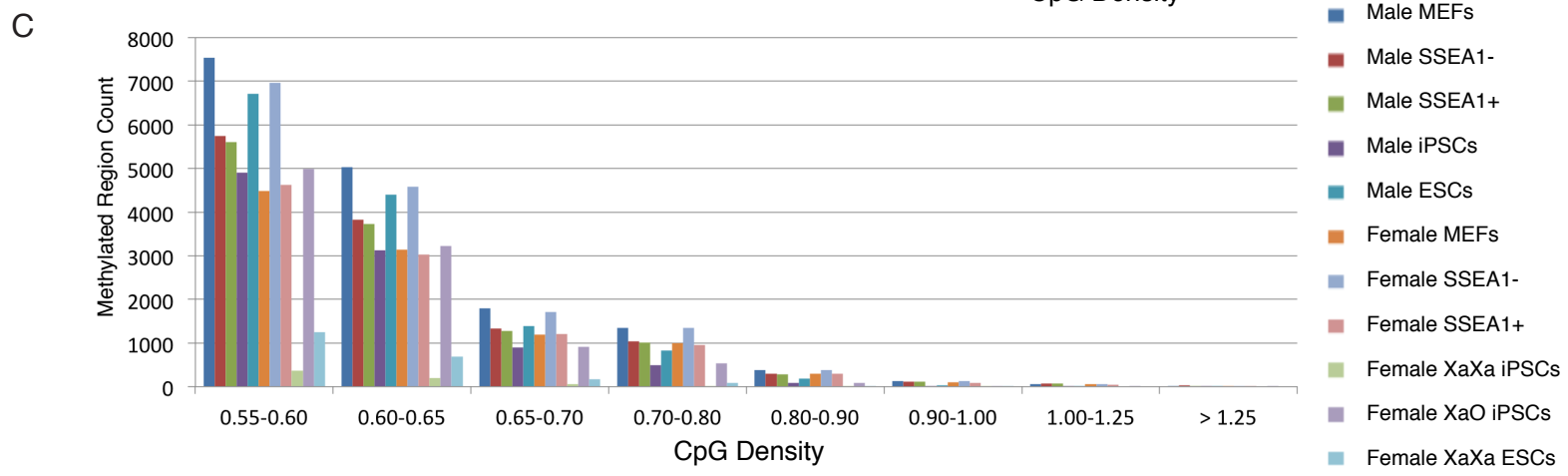

C
Methylated Region Count
8000
7000
6000
5000
4000
3000
2000
1000
0
0.55-0.60
0.60-0.65
0.65-0.70
0.70-0.80
0.80-0.90
0.90-1.00
1.00-1.25
> 1.25
CpG Density
Male MEFs
Male SSEA1-
Male SSEA1+
Male iPSCs
Male ESCs
Female MEFs
Female SSEA1-
Female SSEA1+
Female XaXa iPSCs
Female XaO iPSCs
Female XaXa ESCs

Figure S3

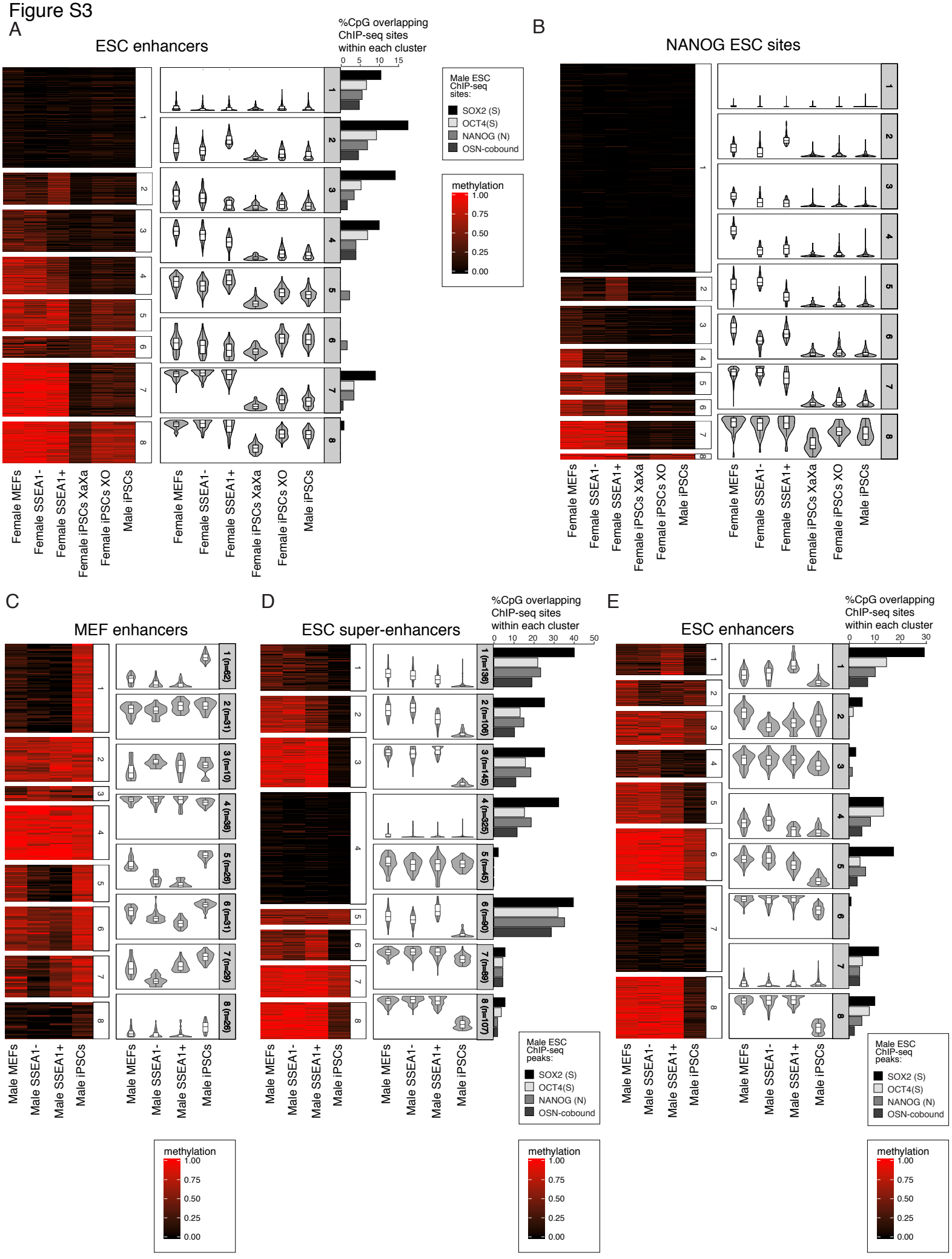
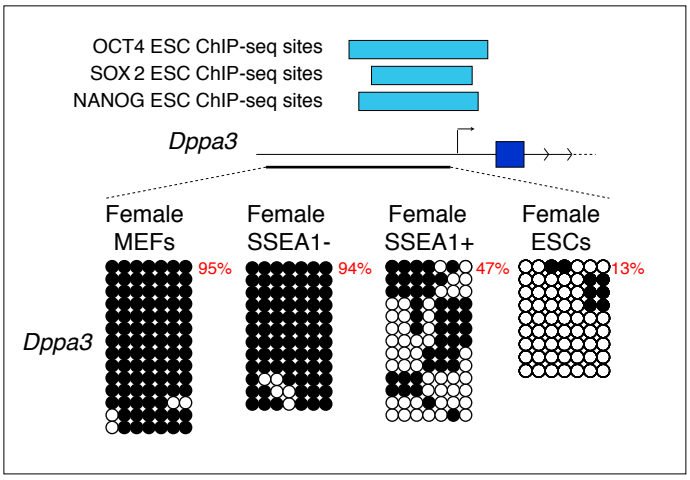
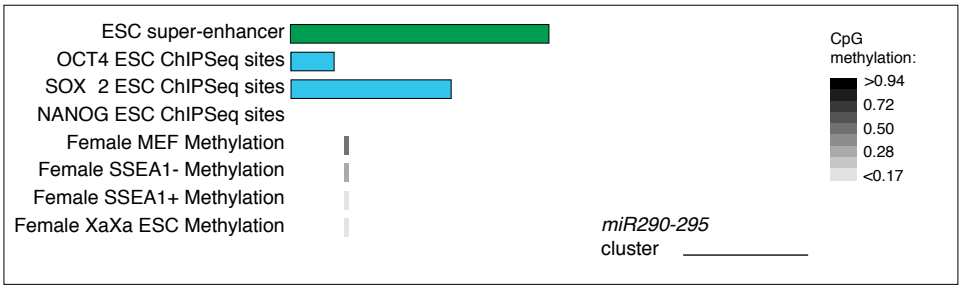

A
ESC enhancers
%CpG overlapping ChIP-seq sites within each cluster
0 5 10 15
Male ESC ChIP-seq sites:
SOX2 (S)
OCT4(S)
NANOG (N)
OSN-cobound
methylation
1.00
0.75
0.50
0.25
0.00
Female MEFs
Female SSEA1-
Female SSEA1+
Female iPSCs XaXa
Female iPSCs XO
Male iPSCs
B
NANOG ESC sites
C
MEF enhancers
1 (n=82)
2 (n=31)
3 (n=10)
4 (n=38)
5 (n=26)
6 (n=31)
7 (n=29)
8 (n=28)
Male MEFs
Male SSEA1-
Male SSEA1+
Male iPSCs
D
ESC super-enhancers
%CpG overlapping ChIP-seq sites within each cluster
0 10 20 30 40 50
1 (n=136)
2 (n=106)
3 (n=145)
4 (n=325)
5 (n=45)
6 (n=90)
7 (n=69)
8 (n=107)
Male ESC ChIP-seq peaks:
SOX2 (S)
OCT4(S)
NANOG (N)
OSN-cobound
E
ESC enhancers
%CpG overlapping ChIP-seq sites within each cluster
0 10 20 30

Figure S4

A


OCT4 ESC ChIP-seq sites
SOX 2 ESC ChIP-seq sites
NANOG ESC ChIP-seq sites
Dppa3
Female MEFs
Female SSEA1-
Female SSEA1+
Female ESCs
95%
94%
47%
13%
Dppa3

B


ESC super-enhancer
OCT4 ESC ChIPSeq sites
SOX 2 ESC ChIPSeq sites
NANOG ESC ChIPSeq sites
Female MEF Methylation
Female SSEA1- Methylation
Female SSEA1+ Methylation
Female XaXa ESC Methylation
CpG methylation:
>0.94
0.72
0.50
0.28
<0.17
miR290-295 cluster

SUPPLEMENTAL FIGURES AND LEGENDS

Figure S1, related to Figure 1. XX versus XO assessment of iPSC cells and methylation analysis of the non-repeat genome across reprogramming and iPSC/ESC lines

(A) RNA FISH analysis for *Tsix* in female iPSC lines. The proportion of nuclei with biallelic *Tsix* signal is indicated. This information was used to deduce the extent to which cell lines are XaXa and XO.

Figure S2, related to Figure 2. Further analysis of global 5hmC as well as 5mC and ICR methylation

(A) Mass spectrometry analysis of 5hmC in male and female iPSCs. 5hmC content is expressed as the percentage of 5hmC in the total pool of cytosine, normalized to male values in each time point analyzed. The average for 3 male and 2-3 female lines are shown for different passages. Data are presented as mean + SEM. Two-tailed unpaired t-test, n=2 experiments, *P< 0.05, ns= non-significant.

(B) Proportion (%) of CpG-rich regions (non-overlapping 150bp windows with >10% of CpGs covered) of various CpG densities displaying low (<30%), medium (30-70%) and high (>70%) methylation in the indicated female and male cell types and reprogramming intermediates. CpG density refers to an Observed/Expected ratio equal to (#Number of CpGs)/(Size of regions*(1/16)).

(C) Number of methylated CpG-rich regions (non-overlapping 150bp windows with >10% of CpGs covered) per CpG density for the data shown in (B).

(D) Distributions of CpG methylation levels within ICRs across indicated samples, shown as violin plots. Only CpGs covered in all samples were considered.

Figure S3, related to Figure 4. Single base pair resolution analysis of methylation during female and male cell reprogramming

(A) Heatmap showing the k-means clustering of single CpG methylation levels for typical ESC enhancers as defined in (Koche et al., 2011) in the indicated female iPSCs and female reprogramming intermediates, as well as male iPSCs. The methylation level ranges from 0 (no methylation) to 1 (100% methylation). Violin plots show the distribution of CpG methylation levels for each cluster across all samples shown in the heatmap. The proportion of

CpGs overlapping O, S, N ESCs binding sites, or those sites co-bound by OSN (OSN) in ESCs, within a given cluster is indicated by bar graphs.

(B) As in (A), except for sites bound by the transcription factor NANOG in ESCs. The methylation legend is the same as in figure S3A.

(C) As in Figure 4A but exclusively for male samples.

(D) As in Figure 4B but exclusively for male samples.

(E) As in Figure S3A but exclusively for male samples.

Figure S4, related to Figure 4. Methylation state of selected regions

(A) Bisulfite sequencing analysis of the *Dppa3* locus in female SSEA1- and SSEA1+ reprogramming intermediates, female MEFs and female ESCs. The region just upstream *Dppa3* (promoter) and the first exon (blue) are shown. Partial loss of DNA methylation is seen in SSEA1+ cells, but not in the SSEA1- cells, and female ESCs display complete demethylation. O, S and N ESC binding sites are indicated.

(C) RRBS data depicting a CpG within an ESC super-enhancer and overlapping with a Oct4 and Sox2 ESC binding site upstream the miR290-295 cluster. The data show a focal loss of DNA methylation in the SSEA1- and SSEA1+ intermediates but not in starting MEFs. The demethylated region is even further expanded in female ESCs.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell culture

Male ESCs (V6.5) and female ESCs (F1-2-1, passage 15) were grown in ESC culture media: KO DMEM (Invitrogen 10829-018) containing 15% FBS, leukemia inhibiting factor (LIF), penicillin/streptomycin, L-glutamine, b-mercaptoethanol, and nonessential amino acids. For methylation analysis, male and female iPSCs were grown in ESC medium after isolation from reprogramming cultures (Pasque et al., 2014) (see below). Male and female MEFs were derived from d14.5 embryos and cultured in DMEM (Invitrogen 11995-065) with the same components as for ESC media except for LIF and with 10% FBS instead of 15% FBS.

Reprogramming stages

Reprogramming and the isolation of reprogramming intermediates was previously described (Pasque et al., 2014).

Briefly, we used female and male MEFs carrying M2rtTA in the Rosa26 locus and a single polycistronic cassette encoding *Oct4*, *Sox2*, *Klf4*, *c-Myc* and produced by two different laboratories (OSKM, (Carey et al. 2010); OKSM (Sridharan et al., 2013) as indicated in Table S1). To initiate reprogramming, 2ug/ul dox were added to MEFs at passage one, and replaced every two days. On day 5, reprogramming cultures were switched from FBS-media to Knock-out Serum Replacement (KSR)-containing media, which was extensively described before (Pasque et al., 2014). One exception was the isolation of male and female iPSCs from the tetO-OSKM mice (Carey et al., 2010), which was done in ESC media supplemented with ascorbic acid (50 ug/ml). Female SSEA1 reprogramming intermediates were isolated by FACS on day 9 of reprogramming (Pasque et al., 2014). Briefly, reprogramming cultures were dissociated, stained with a SSEA1-PE antibody (R&D FAB2155P, Clone MC-480, lot LOY0410071), and sorted on a FACS Aria Cell Sorter (female SSEA1 cells). Purity checks indicated high purity of the sorted cells, and replating experiments demonstrated that SSEA1+ reprogramming intermediates are enriched for cells poised to become iPSCs (Pasque et al., 2014). Male SSEA1 reprogramming intermediates were isolated by MACS at day 9 of reprogramming. iPSCs were picked from day 21 reprograming cultures and grown in ESC media on feeders. Table S1 lists all the cell lines, replicates, reprogramming method and passage number of the primary and established cell lines used in this study.

RRBS and data analysis

iPSCs and ESCs were feeder-depleted before isolation for RRBS. The generation of RRBS data used in this study was described previously (Pasque et al., 2014). At least two RRBS biological replicate libraries were constructed for each cell type analyzed, and up to 4 biological replicates for female MEFs and female ESCs were utilized (Table S1). The GEO accession number for the RRBS data used in this study is GSE58109. Some of these data sets (GSE58109) were used previously to determine the methylation state of CpG-islands on the X chromosome in different reprogramming stages (Pasque et al., 2014). The genomic features analyzed for DNA methylation include the following: MEF- and ESC-specific enhancers were obtained from (Koche et al., 2011), ESC super-enhancers from (Whyte et al., 2013), and NANOG, OCT4 and SOX2 male ESC ChIP-Seq sites from (Chronis et al., 2017). Repetitive LINE, SINE, and LTR elements were downloaded from UCSC and are based on RepeatMasker annotation. When specified, repeats were filtered with RepeatMasker (http://www.repeatmasker.org). Mouse imprinted DMR loci were obtained from (Meredith et al., 2015) and used to calculate the distribution of DNA

methylation levels for CpGs covered >5x in any of the cell type and plotted as violin plots using R. These ICRs were defined by (Meredith et al., 2015) as GSK3-dependant. Apart from the PCA analysis, biological replicate samples of given cell types were merged. Average methylation levels for a given genomic feature were defined as the mean (weighted by coverage) methylation of the CpGs within that region, with coverage capped at 25x per CpG. False discovery rates were calculated using the Benjamini-Hochberg method (Storey and Tibshirani, 2003) and an FDR threshold of 0.05 was defined as significant in all analyses. An FDR threshold of 0.05 was defined as significant in all other analyses. CpG-rich regions were defined by scanning and integrating 150bp windows across the genome for CpG frequency (Figures S2B/C), and then CpG methylation was averaged across regions with >10% of CpGs covered. CpG-regions were retained when the ratio of observed/expected CpG frequency was greater than 0.55. All other statistical analyses (including PCA and k-means clustering) used standard packages and methods in R. PCA was done only on 100 bp tiles on the autosomes to avoid influence by sex chromosome methylation states.

Dotblot

Genomic DNA was extracted with the Purelink Genomic DNA kit (Invitrogen, k182001), prepared with 2-fold serial dilutions in TE buffer and denatured in 0.2M NaOH/5mM EDTA (final concentration) at 95C for 10 min and followed by adding an equal volume of ice-cold 2M ammonium acetate (pH 7.0). Denatured DNA samples were spotted on nitrocellulose membranes (VWR,10600002) and washed with 0,4M NaOH once in an assembled Bio-Dot apparatus (Bio-Rad) according to manufacturer's instruction. The membrane was air-dried for 5 min and UV-cross linked twice at 1200 uJ/cm2. The membrane was blocked with 5% skimmed milk in TBST for 1 hr, followed by incubation with a 1:1000 dilution of the anti-5mC antibody (Millipore, MABE146) overnight at 4° C. After 3 washes with TBST, membranes were incubated with secondary antibody HRP conjugated goat anti-mouse IgG (Bio-Rad: 1706516, 1/5000). The membranes were then washed with TBST and visualized using the ECL chemiluminescence reagent (Perkin-Elmer: NEL103001EA). The quantification of dot-blot was performed using ImageJ.

Mass Spectrometry

For mass spectrometry analysis of DNA methylation, 1 ug of genomic DNA was analyzed using liquid chromatography triple-quadrupole mass spectrometry (LC-QQQ) (Le et al., 2011). The concentration (uM) of

Cytosine (unmodified), 5mC and 5hmC were obtained using standard curves of known C, 5mC and 5hmC amounts. The percentage of 5mC or 5hmC in the total pool of C was obtained by calculating the ratio of the concentration of 5mC or 5hmC to the concentration of total C.

SUPPLEMENTAL REFERENCES

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., Jaenisch, R., 2010. Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat Meth 7, 56–59.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., Plath, K., 2017. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell 168, 442–459.e20.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., Meissner, A., 2011. Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. Cell Stem Cell 8, 96–105.

Le, T., Kim, K.-P., Fan, G., Faull, K.F., 2011. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. Analytical biochemistry 412, 203–209.

Meredith, G.D., D'Ippolito, A., Dudas, M., Zeidner, L.C., Hostetter, L., Faulds, K., Arnold, T.H., Popkie, A.P., Doble, B.W., Marnellos, G., Adams, C., Wang, Y., Phiel, C.J., 2015. Glycogen synthase kinase-3 (Gsk-3) plays a fundamental role in maintaining DNA methylation at imprinted loci in mouse embryonic stem cells. Molecular Biology of the Cell 26, 2139–2150.

Pasque, V., Tchieu, J., Karnik, R., Uyeda, M., Sadhu Dimashkie, A., Case, D., Papp, B., Bonora, G., Patel, S., Ho, R., Schmidt, R., McKee, R., Sado, T., Tada, T., Meissner, A., Plath, K., 2014. X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency. Cell 159, 1681–1697.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., Carey, M., Garcia, B.A., Plath, K., 2013. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat Cell Biol 15, 872–882.

Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100, 9440–9445.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., Young, R.A., 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319.