Stem Cell Reports, Volume *12*

Supplemental Information

Enhancer Chromatin and 3D Genome Architecture Changes from Naive to Primed Human Embryonic Stem Cell States

Stephanie L. Battle, Naresh Doni Jayavelu, Robert N. Azad, Jennifer Hesson, Faria N. Ahmed, Eliah G. Overbey, Joseph A. Zoller, Julie Mathieu, Hannele Ruohola-Baker, Carol B. Ware, and R. David Hawkins

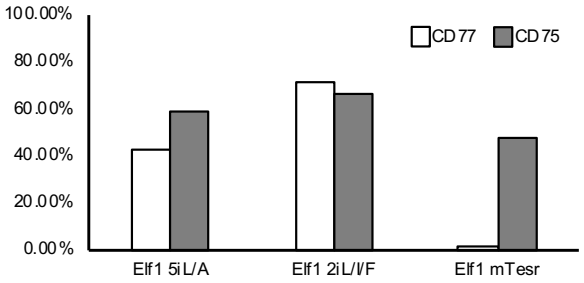
SUPPLEMENTAL INFORMATION

Enhancer chromatin and 3D genome architecture changes from naïve to primed human embryonic stem cell states

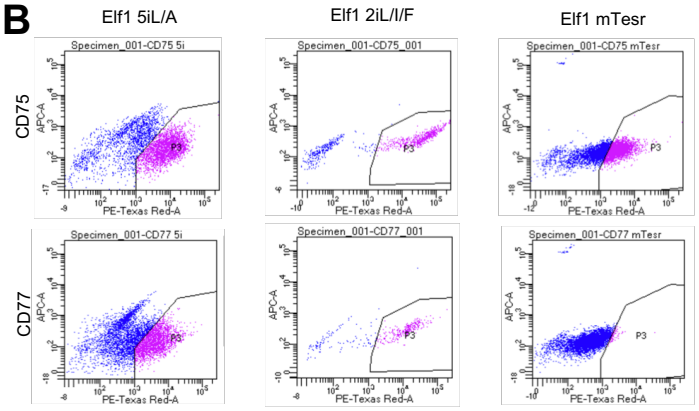
Stephanie L. Battle[1,2], Naresh Doni Jayavelu[1,2,3,6], Robert N. Azad[1,2,6], Jennifer Hesson[2,4], Faria N. Ahmed[1,2], Eliah G. Overbey[1,2], Joseph A. Zoller[1], Julie Mathieu[2,5], Hannele Ruohola-Baker[2,5], Carol B. Ware[2,4], R. David Hawkins[1,2,3,*]

Figure S1

A



B

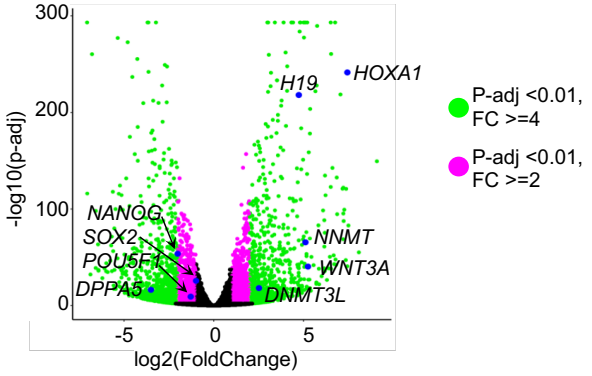


C

Average CpG Methylation Levels

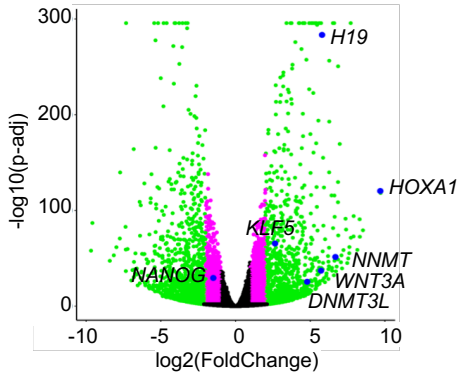
Growth Condition	# CpGs	Coverage cutoff	Avg Meth level
Elf1 5iL/A	2,678,704	10x	14%
Elf1 2iL/I/F	2,847,041	10x	30%
Elf1 AF	2,533,524	10x	27%
H9	1,552,396	10x	53%

G



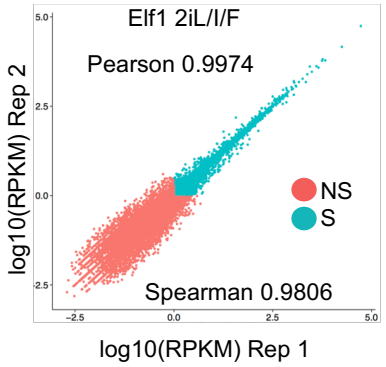
TR Up	2iL/I/F Up
995	883
2,037	1,920

H

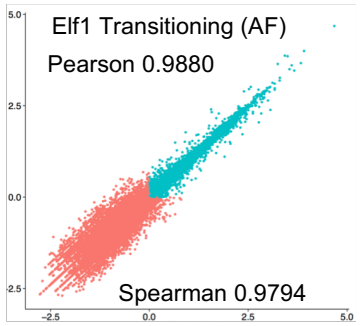


Primed Up	2iL/I/F Up
1,505	1,184
2,993	2,431

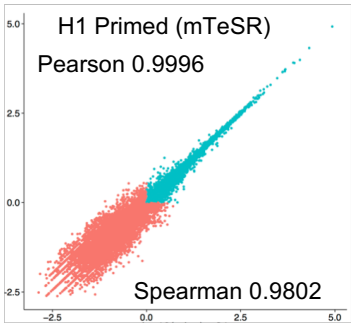
D



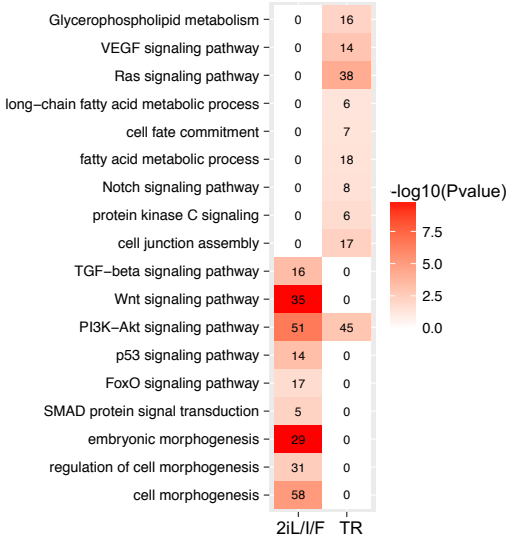
E



F



I DEGs Pathways 2iL/I/F : TR



J DEGs Pathways 2iL/I/F : Primed

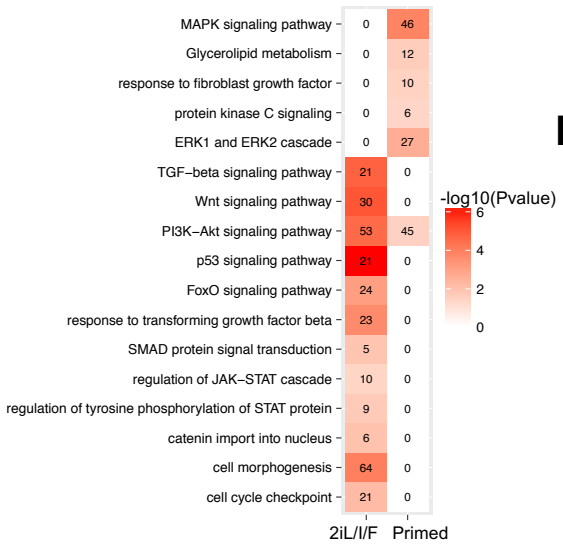
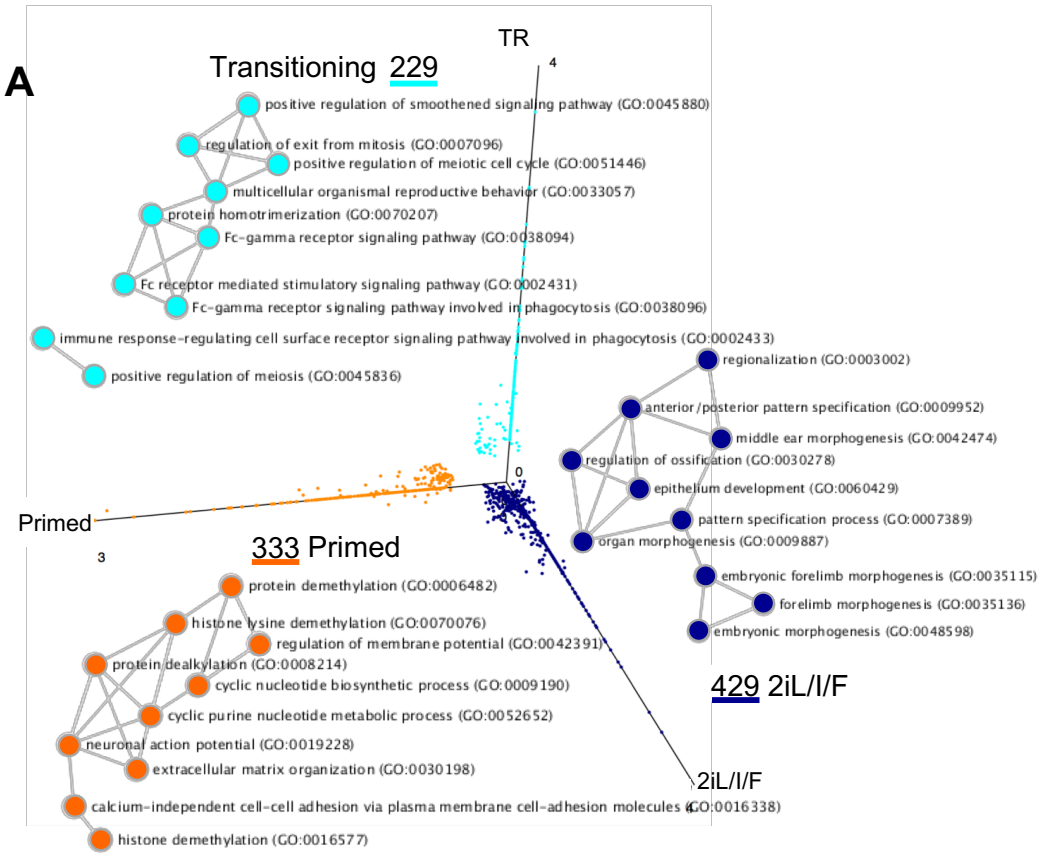
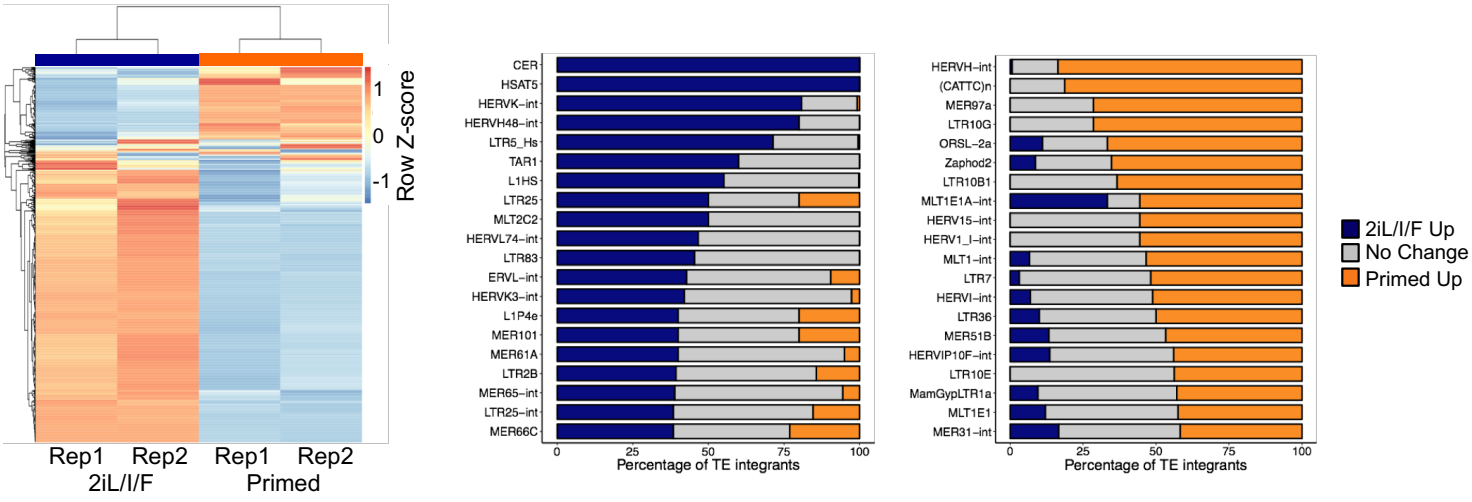


Figure S1: RNA-seq replicates and DEGs in Key Pathways
(A) Graph of percent of Elf1 5iL/A, 2iL/I/F, and mTesr (primed) hESCs positive for naïve cell surface markers as determined by Collier et al., 2017. (B) FACS scatterplots after gating for single cells, the P3 population was counted as positive for either CD75 or CD77. (C) Table shows average CpG DNA methylation levels as determined by RRBS. H9 data were taken from Kyttala et al., 2016. (D-F) Correlation between RNA-seq replicates, log10(RPKM) is plotted. On figure "NS" means not significant and "S" means significant. (G-H) Volcano plot of differentially expressed genes (DEGs) in 2iL/I/F versus transitioning (G) and 2iL/I/F versus primed (H) pairwise comparison. Genes in magenta have P-adj < 0.01 and fold change ≥ 2 while genes in green have P-adj < 0.01 and fold change ≥ 4. (I-J) Heatmap showing significantly overrepresented GO terms and KEGG pathways based on DEGs in 2iL/I/F versus transitioning and 2iL/I/F versus primed pairwise comparison.

Figure S2


A
TR
Transitioning 229
positive regulation of smoothened signaling pathway (GO:0045880)
regulation of exit from mitosis (GO:0007096)
positive regulation of meiotic cell cycle (GO:0051446)
multicellular organismal reproductive behavior (GO:0033057)
protein homotrimerization (GO:0070207)
Fc-gamma receptor signaling pathway (GO:0038094)
Fc receptor mediated stimulatory signaling pathway (GO:0002431)
Fc-gamma receptor signaling pathway involved in phagocytosis (GO:0038096)
immune response-regulating cell surface receptor signaling pathway involved in phagocytosis (GO:0002433)
positive regulation of meiosis (GO:0045836)
regionalization (GO:0003002)
anterior/posterior pattern specification (GO:0009952)
middle ear morphogenesis (GO:0042474)
regulation of ossification (GO:0030278)
epithelium development (GO:0060429)
pattern specification process (GO:0007389)
organ morphogenesis (GO:0009887)
embryonic forelimb morphogenesis (GO:0035115)
forelimb morphogenesis (GO:0035136)
embryonic morphogenesis (GO:0048598)
Primed
333 Primed
protein demethylation (GO:0006482)
histone lysine demethylation (GO:0070076)
regulation of membrane potential (GO:0042391)
protein dealkylation (GO:0008214)
cyclic nucleotide biosynthetic process (GO:0009190)
cyclic purine nucleotide metabolic process (GO:0052652)
neuronal action potential (GO:0019228)
extracellular matrix organization (GO:0030198)
calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules (GO:0016338)
histone demethylation (GO:0016577)
429 2iL/I/F
2iL/I/F
4
0
3


B
TE Signature in 2iL/I/F and Primed
Row Z-score
1
0
-1
Rep1
Rep2
Rep1
Rep2
2iL/I/F
Primed
CER
HSAT5
HERVK-int
HERVH48-int
LTR5_Hs
TAR1
L1HS
LTR25
MLT2C2
HERVL74-int
LTR83
ERVL-int
HERVK3-int
L1P4e
MER101
MER61A
LTR2B
MER65-int
LTR25-int
MER66C
0
25
50
75
100
Percentage of TE integrants
HERVH-int
(CATTC)n
MER97a
LTR10G
ORSL-2a
Zaphod2
LTR10B1
MLT1E1A-int
HERV15-int
HERV1_I-int
MLT1-int
LTR7
HERVI-int
LTR36
MER51B
HERVIP10F-int
LTR10E
MamGypLTR1a
MLT1E1
MER31-int
0
25
50
75
100
Percentage of TE integrants
2iL/I/F Up
No Change
Primed Up

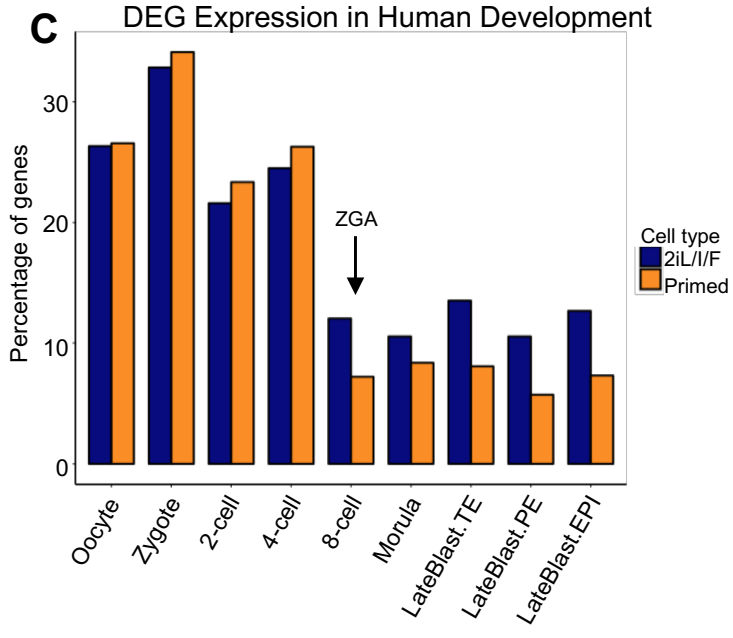

C
DEG Expression in Human Development
Percentage of genes
30
20
10
0
ZGA
Cell type
2iL/I/F
Primed
Oocyte
Zygote
2-cell
4-cell
8-cell
Morula
LateBlast.TE
LateBlast.PE
LateBlast.EPI

Figure S2: Cell-type Specific Genes and Transposable Elements
(A) Cell type-specific genes in the different hESC stages by applying a cutoff of a RPKM value greater than or equal to two in one cell type and less than one in the other two cell types. GO terms shown in network, connected by shared genes between terms. (B) Hierarchical clustering of transposable elements expression separates 2iL/I/F from primed hESCs in first panel. Second and third panel show the percentage of integrants that are upregulated in each cell type. (C) Percentage of genes upregulated in pairwise comparison of 2iL/I/F or primed hESCs that are also found to be upregulated in human embryo developmental stages.

Figure S3

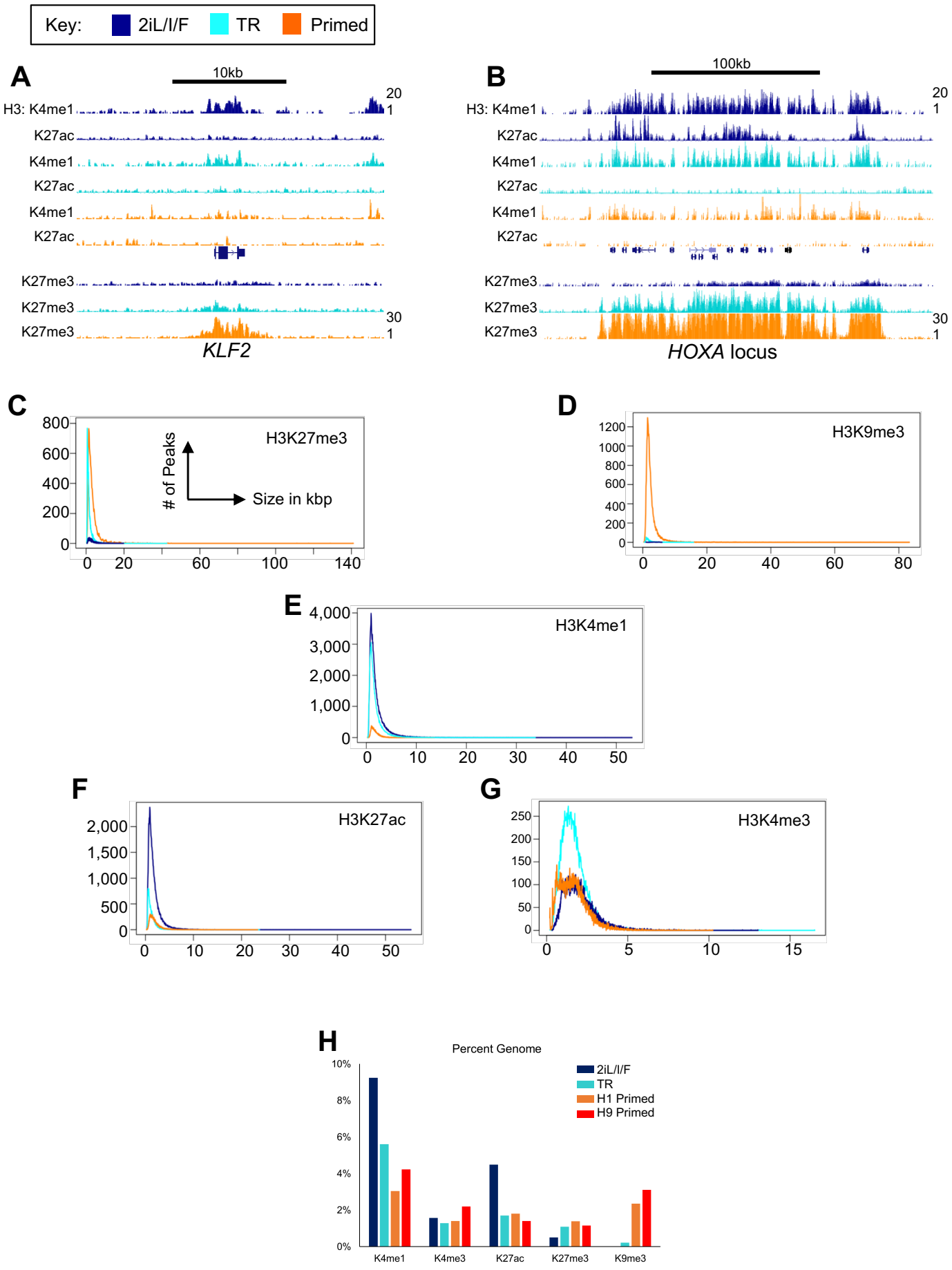
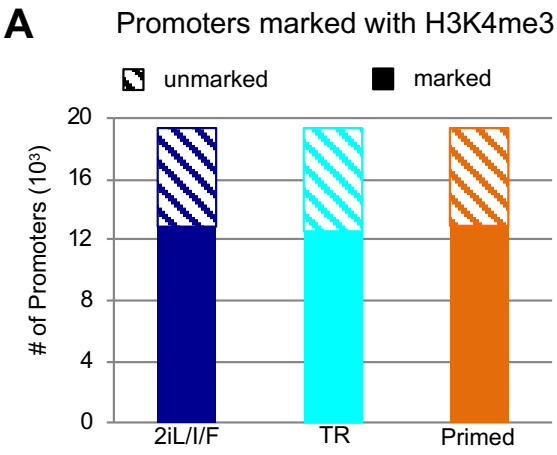
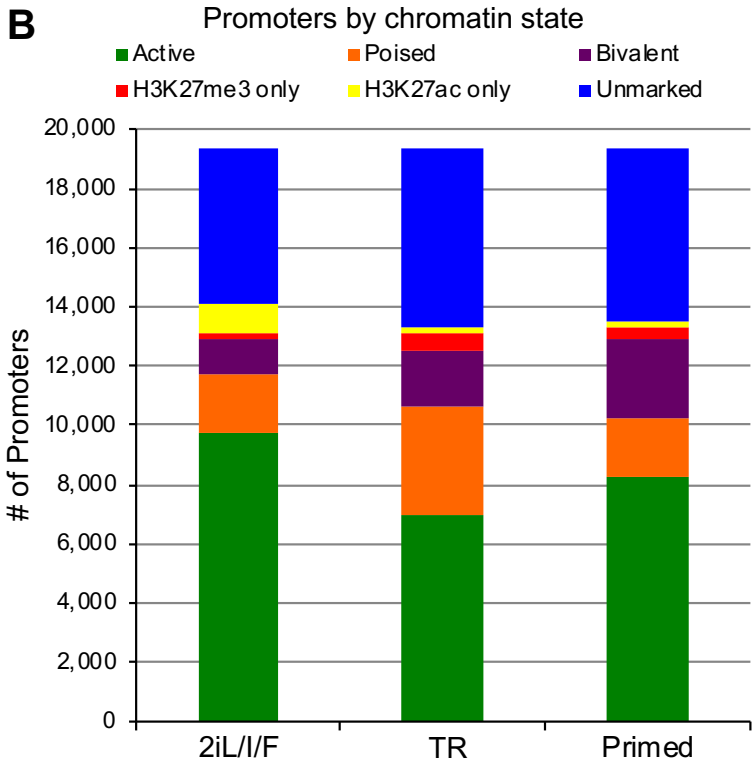

Key: 2iL/I/F TR Primed
A
10kb
H3: K4me1
K27ac
K4me1
K27ac
K4me1
K27ac
K27me3
K27me3
K27me3
20
1
30
1
KLF2
B
100kb
H3: K4me1
K27ac
K4me1
K27ac
K4me1
K27ac
K27me3
K27me3
K27me3
20
1
30
1
HOXA locus
C
H3K27me3
# of Peaks
Size in kbp
800
600
400
200
0
0 20 60 100 140
D
H3K9me3
1200
1000
800
600
400
200
0
0 20 40 60 80
E
H3K4me1
4,000
3,000
2,000
1,000
0
0 10 20 30 40 50
F
H3K27ac
2,000
1,500
1,000
500
0
0 10 20 30 40 50
G
H3K4me3
250
200
150
100
50
0
0 5 10 15
H
Percent Genome
2iL/I/F
TR
H1 Primed
H9 Primed
10%
8%
6%
4%
2%
0%
K4me1
K4me3
K27ac
K27me3
K9me3

Figure S3: Histone Modifications globally and Chromosome X
(A-B) UCSC Browser image of the *KLF2* (A) and *HOXA* (B) loci. H3K4me1 and H3K27ac (top 6 tracks, RPKM scale 1-20) and H3K27me3 (bottom 3 tracks, RPKM scale 1-30) are shown for 2iL/I/F, transitioning and primed ChIP-seq data. (C-G) Histograms showing distribution of peak lengths for (C) H3K27me3, (D) H3K9me3, (E) H3K4me1, (F) H3K27ac and (G) H3K4me3. (H) Percent of human genome bases covered by histone modifications in primed cell line H9. H9 ChIP-seq data taken from the Roadmap Project.

Figure S4


A
Promoters marked with H3K4me3
unmarked
marked
# of Promoters (10³)
20
16
12
8
4
0
2iL/I/F
TR
Primed


B
Promoters by chromatin state
Active
Poised
Bivalent
H3K27me3 only
H3K27ac only
Unmarked
# of Promoters
20,000
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
2iL/I/F
TR
Primed

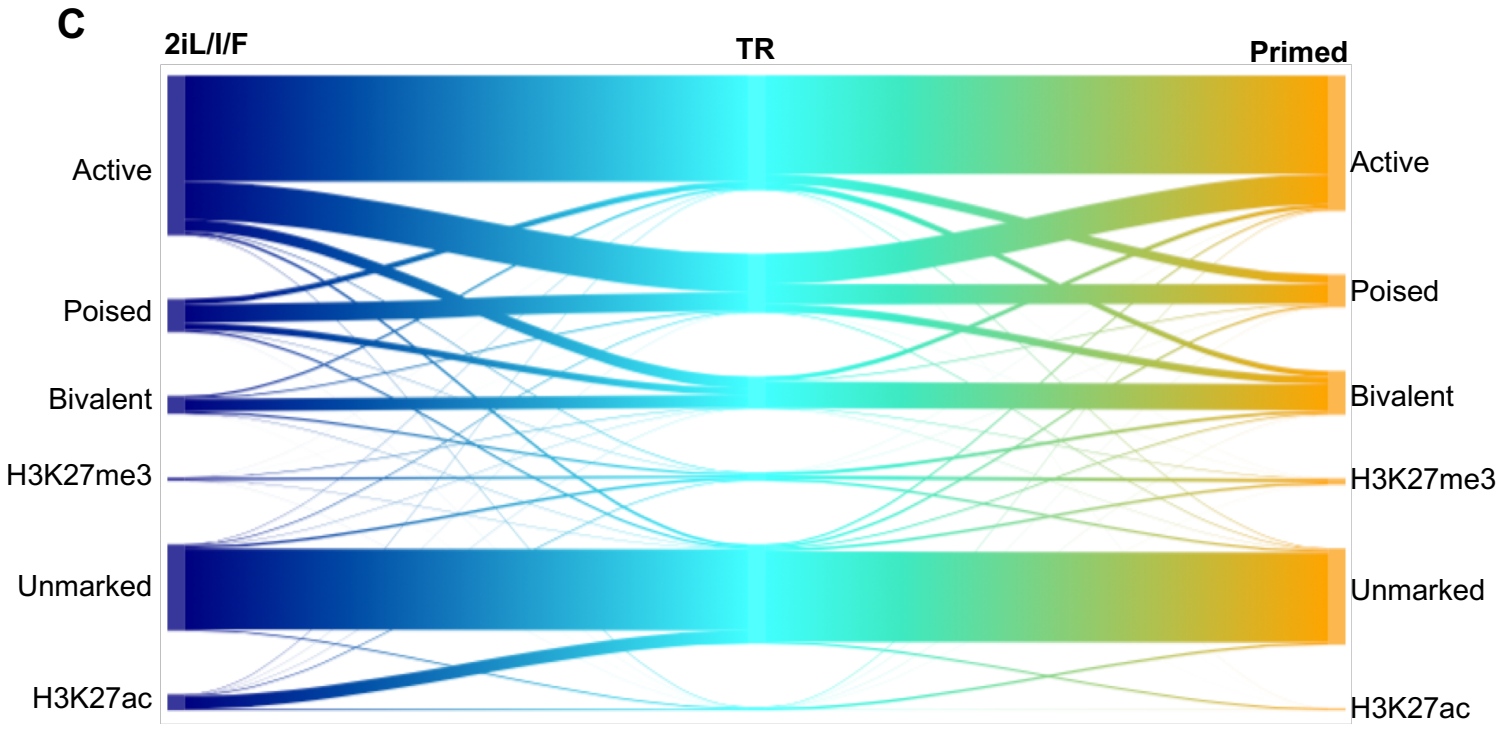
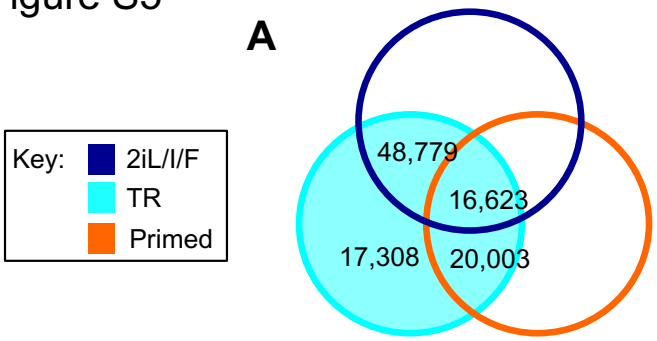
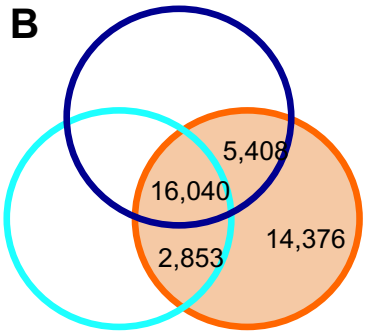
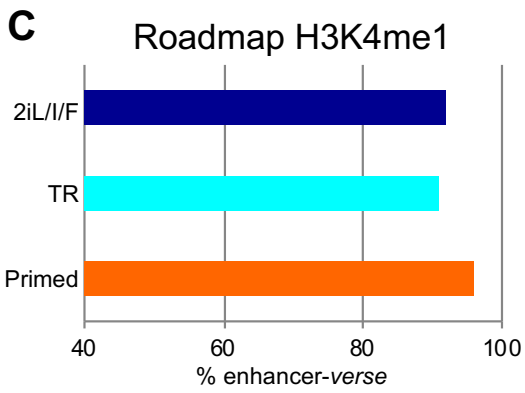

C
2iL/I/F
TR
Primed
Active
Poised
Bivalent
H3K27me3
Unmarked
H3K27ac
Active
Poised
Bivalent
H3K27me3
Unmarked
H3K27ac

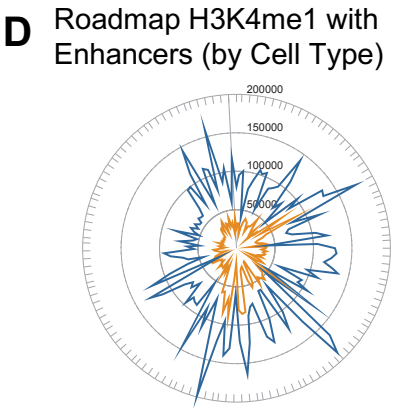
Figure S4: Histone Modifications at Promoters and Bivalent Gene Ontology
(A) Number of GENCODE coding gene promoters marked with H3K4me3 or unmarked in each hESC stage. (B) Breakdown of promoter chromatin state categories. Graph shows how many gene promoters are found in each category. (C) Sankey plot of all promoter state transitions between three hES cell types.

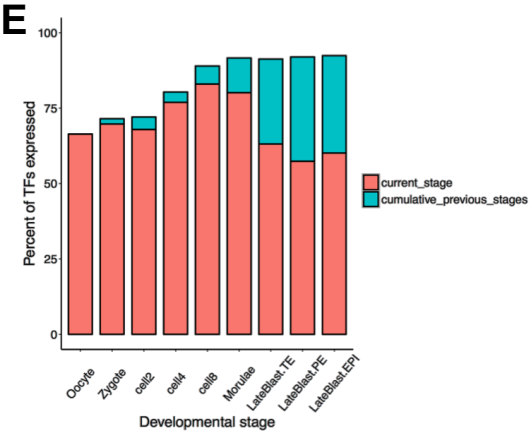
Figure S5











F

Enhancer Class	2iL/I/F	TR	Primed
Active H3K4me1 + H3K27ac	31,673	7,990	8,946
Poised H3K4me1 only	65,334	58,497	28,067
Poised H3K4me1 + H3K27me3	1,013	2,980	1,664

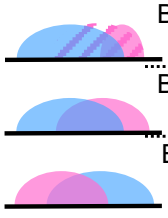
G

	H3K4me1				H3K27ac				H3K4me3
	< 5 kb	≥ 5 kb	≥ 10 kb	≥ 20 kb	< 5 kb	≥ 5 kb	≥ 10 kb	≥ 20 kb	≥ 5 kb
	Short	**Broad**	**Grande**	**Venti**	**Short**	**Broad**	**Grande**	**Venti**	**Broad**
2iL/I/F	109,587	7,412	1,413	157	65,170	2,330	294	20	283
TR	80,818	3,122	391	18	36,460	241	10	1	74
Primed	46,472	371	10	0	23,865	803	53	3	144

H

2iL/I/F	8,140	6,610	7,428
TR	7,383	5,153	6,559
Primed	6,123	5,249	6,700

I



	2iL/I/F	TR	Primed
Broad K4me1 w/ K27ac	87.72%	49.01%	71.16%
Broad K4me1 w/ Broad K27ac	23.46%	4.13%	22.64%
Broad K27ac w/ Broad K4me1	78.58%	52.70%	10.46%

Figure S5: Characteristics of Enhancers in hESCs
(A-B) Venn diagrams of enhancer overlaps for transitioning (A) and primed (B) cells with other cell types. Venns are separate because one peak in one cell type can overlap multiple peaks in another cell type. (C) Percent of hESC H3K4me1 genomic space (% bases or enhancer-verse) occupied by Roadmap H3K4me1 from 127 cell types. (D) Number of Roadmap H3K4me1 peaks from 127 cell types overlapping with 2iL/I/F and primed H3K4me1 enhancers. (E) Percent of transcription factors, annotated from Animal TFDB, expressed at each stage of embryogenesis and cumulatively across stages using single cell RNA-seq data from Yan et al. (Yan et al., 2013). (F) Number of enhancers peaks that fall into the category of active, poised HK4me1 only or poised H3K27me3 enhancer classes. (G) Number of H3K4me1, H3K27ac and H3K4me3 peaks, broken down by size. (H) Average length of broad (≥ 5 kb) H3K4me1 and H3K27ac domains and the average number of bases overlapped in shared regions. (I) Percent of broad peaks that overlap in these comparisons: broad H3K4me1 with any H3K27ac overlap, broad H3K4me1 with only broad H3K27ac overlap, or broad H3K27ac with broad H3K4me1 overlap.

Figure S6

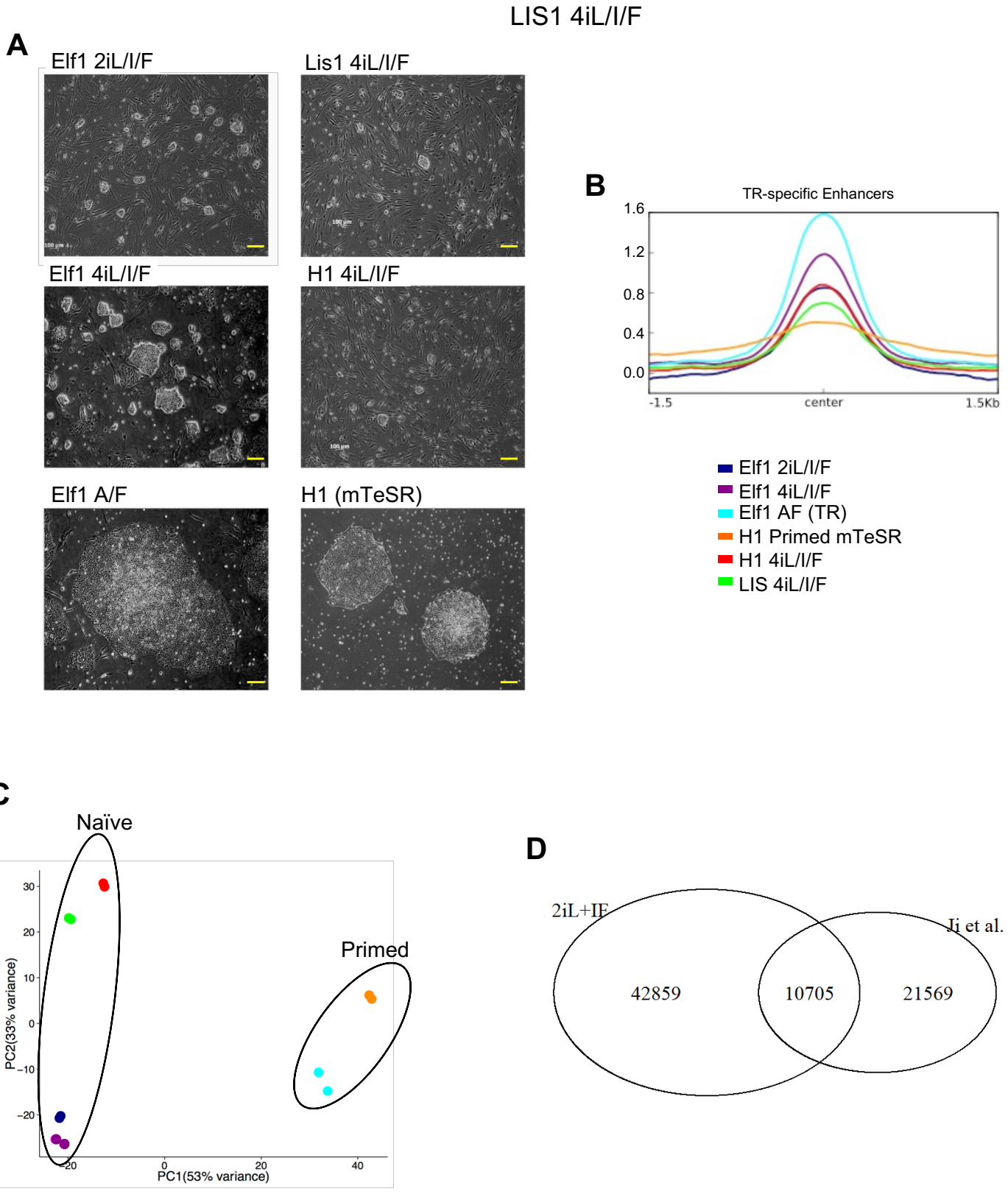

LIS1 4iL/I/F
A
Elf1 2iL/I/F
Lis1 4iL/I/F
Elf1 4iL/I/F
H1 4iL/I/F
Elf1 A/F
H1 (mTeSR)
B
TR-specific Enhancers
1.6
1.2
0.8
0.4
0.0
-1.5
center
1.5Kb
Elf1 2iL/I/F
Elf1 4iL/I/F
Elf1 AF (TR)
H1 Primed mTeSR
H1 4iL/I/F
LIS 4iL/I/F
C
Naïve
Primed
PC2(33% variance)
PC1(53% variance)
D
2iL+IF
Ji et al.
42859
10705
21569

Figure S6: RNA-seq and ChIP-seq of hESCs in Different Growth Conditions
(A) Representative images of cell lines in each growth condition used for this study. Yellow bar is 200um. (B) Average normalized ChIP-seq signal in all naïve and primed hESCs at transitioning-specific enhancers. (C) PCA of autosomal gene expression, naïve cells cluster separately from primed cells. (D) Venn diagram showing the number of Elf1 2iL/I/F H3K27ac peaks that overlap 5iL/A H3K27ac peak calls from Ji et al., 2016.

Figure S7

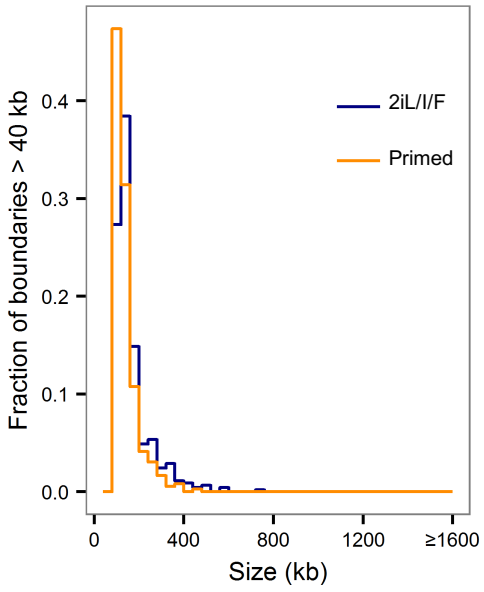
A

Hi-C library	Raw read pairs	Processed read pairs	Cis interactions	Cis interactions within 20 kb
Elf1 Replicate 1	798,771,830	318,226,786	224,980,721	65,753,953
Elf1 Replicate 2	770,256,147	324,524,006	221,137,452	62,491,765
H1 Replicate 1	237,662,270	35,822,152	19,713,347	3,369,816
H1 Replicate 2	496,522,946	160,344,628	124,240,997	20,935,372

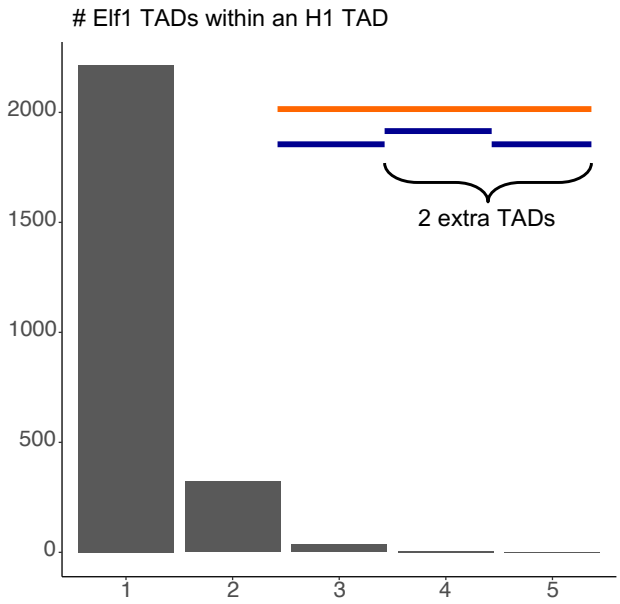
B

Feature	2iL/I/F	Primed
TAD	6,119	5,822
Boundary	5,487	5,279
Boundary > 40 kb	433	351

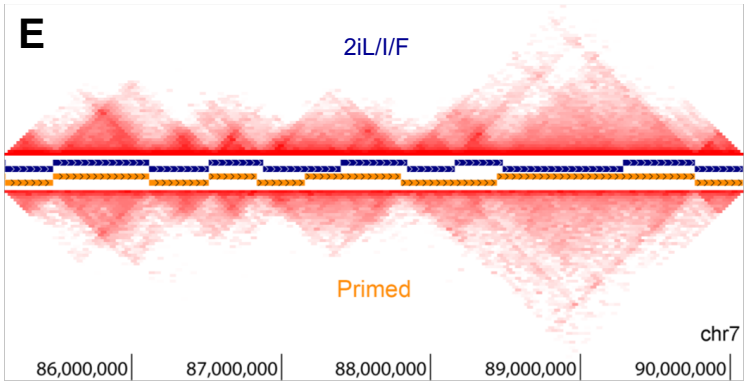
C



D



E



F

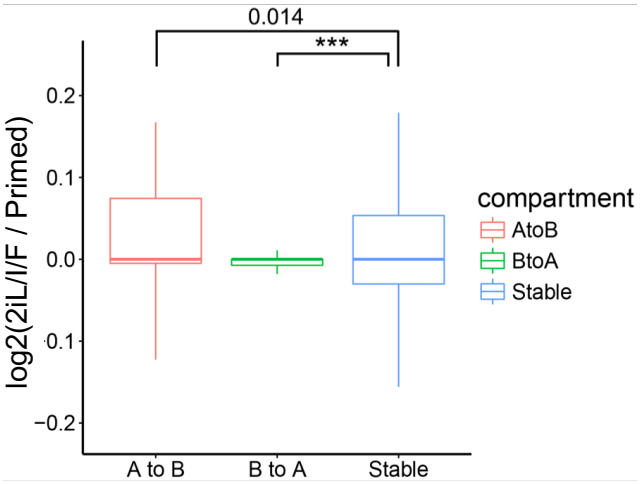


Figure S7: Hi-C libraries and TAD structure
(A) Summary of Hi-C sequencing read pairs and interactions in 2iL/I/F and primed hESCs. (B) Total counts of TADs and boundaries identified in 2iL/I/F and primed hESCs after discarding X and Y chromosome interactions. (C) Global size distributions of boundaries > 40 kb. (D) Bar chart of the number of Elf1 TADs within an H1 TAD and depiction of how "extra" TADs are defined. (E) Contact heatmaps of a region in chromosome 7. Navy and dark orange tracks denote TADs in 2iL/I/F and primed hESCs, respectively. (F) Boxplot of gene expression (RPKM) overlapping A to B compartment switches. "A to B" and "B to A" are 2iL/I/F to primed directions. Stable are compartments that do not switch. P-values are computed using two-sample t-test with one sided alternative. *** P-value < 1.34×10^{-10}.

Table S1: DEGs and DEG Pathways
List of differentially expressed genes and Gene Ontology pathways

Table S2

Elf1 2iL/I/F Sample	Number of Reads Sequenced	Mapped reads	Mapping Efficiency
ENK27acRep1	10,289,987	9,885,369	96.1%
ENK27acRep2	34,868,010	31,613,836	90.7%
ENK27me3Rep1	11,131,576	10,256,807	92.1%
ENK27me3Rep2	28,435,049	25,806,462	90.8%
ENK4me1Rep1	29,070,786	26,338,406	90.6%
ENK4me1Rep2	23,139,610	22,376,410	96.7%
ENK4me3Rep1	15,960,227	14,359,079	90.0%
ENK4me3Rep2	25,423,585	23,909,154	94.0%
ENK9me3Rep1	28,486,276	25,936,935	91.1%
ENK9me3Rep2	42,803,514	38,115,104	89.0%
ENInputRep1	31,149,021	23,377,765	75.1%
ENInputRep2	20,116,865	18,585,106	92.4%
Total 2iL/I/F (1.35E10 bases)		270,560,433	
Elf1 Transitioning Sample ID	Number of Reads Sequenced	Mapped reads	Mapping Efficiency
EPK27acRep1	16,542,791	15,466,596	93.5%
EPK27acRep2	19,045,610	14,495,788	76.1%
EPK27me3Rep1	14,661,174	12,484,843	85.2%
EPK27me3Rep2	17,360,465	10,303,499	59.4%
EPK4me1Rep1	13,466,280	12,643,726	93.9%
EPK4me1Rep2	14,628,581	13,524,908	92.5%
EPK4me3Rep1	14,142,256	12,502,355	88.4%
EPK4me3Rep2	12,112,327	9,124,628	75.3%
EPK9me3Rep1	37,350,962	32,818,042	87.9%
EPK9me3Rep2	42,244,794	36,811,767	87.1%
EPInputRep1	8,581,717	7,920,395	92.3%
EPInputRep2	37,772,483	35,291,325	93.4%
Total EP genome (1.07E10 bases)		213,387,872	

Table S2: Summary of Mapping Statistics for Elf1 2iL/I/F and Transitioning (AF) Cells
ChIP-seq DNA libraries were sequenced on Illumina platform, SE75 and mapped using Bowtie2.

Table S3

	H3K4me1		H3K4me3		H3K27ac		H3K27me3		H3K9me3		# enhancers after removing H3K4me3 overlapping peaks
	# peaks	% genome	# peaks	% genome	# peaks	% genome	# peaks	% genome	# peaks	% genome	
Elf1 2iL/I/F	116,999	9.25%	21,071	1.58%	67,500	4.49%	4,512	0.51%	12	0.00%	98,020
Elf1 TR	83,940	5.61%	20,541	1.29%	36,701	1.71%	17,379	1.09%	2,831	0.22%	69,467
H1 Primed	46,843	3.04%	22,866	1.41%	24,668	1.81%	10,454	1.39%	22,911	2.35%	38,677

Table S3: Number of ChIP-seq Peaks and Percent Genome covered by Histone Modification
Numbers in bar chart for Figure 2B. Number of ChIP-seq peaks called by MACS that pass FDR 5% for each cell type. Numbers in bar chart for Figure 2C. Percent of genome covered by each histone modification. Includes number of K4me1 peaks after excluding peaks that overlap K4me3.

Table S4: Enhancer overlaps with ENCODE DHS and Roadmap H3K4me1
Charts of the number or fraction of a cell type's peaks were found in 2iL/I/F, transitioning or primed hESCs from 177 ENCODE DHS cell types or 127 Roadmap Epigenome Project cell types.

Supplemental Experimental Procedures

Human Embryonic Stem Cell Culture

All human ESC culture conditions were as previously described (Sperber et al., 2015), with the following modifications. Base hESC media consisted of DMEM/F12, 20% knock-out serum replacer (KSR), 1mM sodium pyruvate, 0.1 mM nonessential amino acids, 50 U/mL penicillin, 50 ug/mL streptomycin, and 0.1mM β-mercaptoethanol. Growth conditions: 2iL/I/F - 1uM Mek inhibitor (PD0325901) [catalog #S1036, Selleck Chemicals, Houston, TX, USA], 1uM GSK3 inhibitor (CHIR-99021) [catalog #S2924, Selleck Chemicals, Houston, TX, USA], 10 ng/mL human Leukemia inhibitory factor [catalog #YSP1249, Speed Biosystems, Gaithersburg, MD, USA], 5ng/mL IGF-1 [catalog #100-11 Peprotech, Rocky Hill, NJ], 10ng/mL FGF [catalog #PHG0263, Thermo Fisher Scientific, Waltham, MA, USA]; 4iL/I/F - 1uM Mek inhibitor (PD0325901), 1uM GSK3 inhibitor (CHIR-99021), 5uM JNK inhibitor (SP600125) [catalog #S1460, Selleck Chemicals, Houston, TX, USA], 2uM p38 inhibitor (BIRB796) [catalog #S1574, Selleck Chemicals, Houston, TX, USA], 10 ng/mL Leukemia inhibitory factor, 5ng/mL IGF-1, 10ng/mL FGF.

Flow Cytometry for CD75 and CD77 Markers

2iL/I/F Elf1 cells were harvested from one 10cm^2 plate. The plate was rinsed with PBS, then cells were separated from the plate with trypsin. After the incubation, cells were unicellularized through pipetting of the cell suspension. Trypsin was neutralized using 5% FBS/PBS solution. Cells were rinsed with PBS twice and resuspended in 500 uL of 5% FBS/PBS. 5mL of a 70% EtOH solution was slowly added while cells were vortexed. Cells were stored at -20°C for two days.

To perform antibody staining, cells were removed from -20°C and washed twice with PBS. Cells were resuspended in 300uL of 5% FBS/PBS and then split equally into three samples: an unstained control, CD75 stained, and CD77 stained. 1ug of Anti-CD75 antibody (Abcam, cat #: ab77676, lot #: GR3181182-5) was added to the CD75 stained sample. 1ug of Anti-CD77 antibody (BD Biosciences, cat #: 551352, lot #: 7214648) was added to the CD77 stained sample. The control sample was stored at 4°C in 400uL 5% FBS/PBS. CD75 and CD77 samples were stained with the primary antibody for 30 minutes at room temperature, washed once with 5% FBS/PBS. They were then resuspended in 100uL of 5% FBS/PBS containing the secondary antibody, Alexa Fluor 586 donkey anti-mouse (Invitrogen, cat #: A10037, lot #: 1917938), at a dilution of 1:1000 and incubated for 30 minutes at room temperature in the dark. CD75 and CD77 samples were rinsed once with 5% FBS/PBS and resuspended in 400uL of 5% FBS/PBS. The control, CD75, and CD77 samples were fixed by adding 100uL 4% PFA and were mixed by pipetting. Flow cytometry was performed with BD FACSAria II. The program BD FACSDiva was used for gating and analysis.

Reduced Representation Bisulfite Sequencing

Genomic DNA was extracted with Qiagen All Prep Kit (Qiagen, cat #: 80204). Unmethylated lambda DNA was sheared to 300 bp using Covaris M-series and spiked in to each sample at 0.5% of DNA quantity. About 100-500 ng of genomic DNA was digested in a total volume of 20 uL with 1 uL MspI enzyme (NEB, cat #: R0106S) with 2 uL NEB buffer 2 and incubated at 37C for 2-4 hours, followed by heat inactivation at 65C for 10min. Next, samples underwent A-tailing reaction by adding 1.2 uL Klenow exo- (NEB, cat #: M0212S), 1.2 uL of dNTP mix (10mM dATP, 1mM dGTP, 1mM dCTP (NEB, cat #: N0446S)) to each sample and incubating at 30C for 30 min, then 37C for 30 min, followed by heat inactivation at 65C for 10min. Methylated Illumina compatible adapters with ligated to the DNA samples by adding to each sample 1.4 uL of NEB buffer 2, 1 uL of 10uM methylated adapters, 1.2 uL of T4 DNA ligase (NEB, cat #: M0202S), 3.4 uL of 10mM ATP, and sterile water to a final volume of 34 uL. The ligation reaction sat at RT for 1 hour. Ligated DNA samples were then gel size selected for 150-350 bp fragment size and then samples were proceeded to bisulfite conversion (MethylCode Bisulfite Conversion Kit, ThermoFisher, cat #: MECOV50). Converted DNA samples underwent indexing PCR using KAPA

HIFI Hotstart URACIL+ Readymix (Fisher Scientific, cat #: NC0682281) for 12-15 cycles. Final libraries were cleaned with AMPure beads at 1.1x ratio and sequenced on NextSeq 500. Samples were mapped using Bismarck(Krueger and Andrews, 2011) and methylation levels were calculated using Bismark Methylation Extractor. H9 data were taken from Kyttala et al., 2016 (Kyttala et al., 2016).

ChIP-seq Analysis and Visualization

ChIP-seq was performed and analyzed as previsously described (Hawkins et al., 2013; Valensisi et al., 2017). Sequencing raw reads were trimmed for low quality and adapters using TrimGalore!, aligned to genome (version hg19) using Bowtie2 (Langmead and Salzberg, 2012). For the UCSC genome browser tracks, ChIP-seq signals were normalized by RPKM followed by subtraction of input from ChIP using deepTools suite (Ramirez et al., 2014). Heatmaps and histograms are of normalized ChIP-seq signal: samples are normalized by read count and log2(chip reads/input reads) per 10kb bin is plotted using deepTools suite (Ramirez et al., 2014).

Peak Calling

ChIP-seq peaks were called on merged replicates and normalized to input using MACS v1.4 (Zhang et al., 2008). Peak calls with a FDR of 5% or less were used for downstream analysis. Percent of genome covered, also referred to as genomic space in text, was defined as total number of bases under the peak divided by 2.7e9, the effective genome size. This was found it to be a better representation of global chromatin structure (e.g. a 10kb region can be covered by one or many ChIP-seq peaks due to peak size; the number of peaks may vary more than the total number of bases under the peaks). Peak comparisons and overlaps were done using the BedTools suite (Quinlan and Hall, 2010) for autosomal chromosomes only.

In order to compare the histone marks (H3K4me1 and H3K27ac) across cell types, we divided the genome into 10 kb bins and counted the reads across these 10 kb genomic regions using *featurecounts* in Rsubread package (Liao et al., 2014). Then, PCA was performed on regularized log transformed read count data obtained using DESeq2 (Love et al., 2014).

RNA-seq and Gene Expression

Embryonic stem cells were counted and 200,000 cells were pelleted for RNA extraction using the Qiagen All Prep Kit (cat # 80204). RNA-seq libraries were constructed using the Scriptseq RNA-seq Library Preparation Kit on ¾ of total RNA. Libraries were sequenced single-end 75 on Illumina NextSeq. The quality of the reads and contamination of adapter sequences were checked with FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to human hg19 genome (UCSC) using TopHat2 (Kim et al., 2013). Transcript quantification was performed by Cufflinks (Trapnell et al., 2010) using GENCODE's comprehensive gene annotation release 19 as reference annotation.

Differential Gene Expression Analysis

The raw read counts were calculated using *featurecounts* in Rsubread package (Liao et al., 2014) and GENCODE's release 19 as reference annotation. Differential gene expression analysis was performed with DESeq2 (Love et al., 2014) using read counts matrix. Two sets of differentially expressed genes (DEGs) are identified with P-value < 0.01, log2FC > |1| and P-value < 0.01, log2FC > |2|. The P-values were adjusted for multiple hypothesis correction. DEGs in all pairwise sample comparisons were identified. PCA was performed on regularized log transformed read count data from autosomes of top 500 highly variant genes obtained using DESeq2(Love et al., 2014) and plot was generated using ggplot2 in R(Wickham, 2009).

For transposable elements (TE) analysis, transcripts were quantified using hg19 UCSC RepeatMasker TE annotation. We considered unique reads as well as multi mapped reads during quantification of TE

transcripts. PCA was performed on regularized log transformed read count data of top 500 highly variant TE transcripts obtained using DESeq2.

Identification of Overrepresented GO Terms and Enriched Pathways

ClueGO (Bindea et al., 2009) was used to identify the overrepresented GO terms and enriched pathways with the data from gene ontology consortium and KEGG pathways database. The input gene lists to the ClueGO were DEGs with P-value < 0.01, log2FC > |1|. We used all genes in the genome as background. The statistically significant GO terms and pathways were filtered with P-value < 0.05 and GO term/pathway should contain at least 5 DEGs. P-values were adjusted with Benjamini Hochberg method for multiple hypothesis correction.

Sankey Plot

We looked at promoter chromatin state transitions from 2iL/I/F to primed to gain insight into the establishment of bivalency and other chromatin state changes occurring at gene promoters. In order to accomplish this goal we focused on the over 19,000 autosomal protein-coding gene TSS annotated by GENCODE. We assigned a promoter to a gene if the H3K4me3 peak was within -2kb to +500bp of the TSS. Sankey plot is limited by the presence of multiple promoters within overlapping regions. Sankey plot were created using Google Charts (https://developers.google.com/chart/interactive/docs/gallery/sankey)

In situ DNase Hi-C

Samples were prepared in a manner similar to Deng et al. (Deng et al., 2015). Briefly, nuclei from ~5 x 106 cross-linked Elf1 cells were isolated and permeabilized, and chromatin was digested with 4 U DNase I at room temperature for 4 min. Following end-repair and dA-tailing reactions, chromatin ends were ligated to biotinylated bridge adapters, and nuclei were purified with two volumes of AMPure XP beads (Beckman Coulter). Chromatin ends were phosphorylated and ligated in situ, and protein-DNA cross-links were reversed by proteinase K digestion and incubation at 60C overnight. Following purification, DNA was sonicated to an average size of 400 bp, and chimeric species were enriched via pull-down with streptavidin-coated magnetic beads (Active Motif). Preparation of Hi-C libraries was accomplished by ligating sequencing adapters to the ends of bead-bound DNA fragments and PCR-amplifying the products in the presence of forward and barcoded reverse primers. Libraries were purified with AMPure XP beads, DNA concentrations were determined using a Qubit 2.0 (Thermo Fisher), and size distributions were quantified using a Bioanalyzer with a high sensitivity kit (Agilent). A 10 ng aliquot from each library was digested with BamHI, run on the Bioanalyzer, and compared to an undigested control in order to confirm the presence of a reconstituted BamHI site at the junctions of ligated bridge adapters.

Hi-C Sequencing and Data Processing

Raw Hi-C sequencing reads from H1 hESCs were downloaded from GEO (GSE35156). Reads were aligned using Bowtie2 (Langmead and Salzberg, 2012) to the hg19 reference genome and filtered for MAPQ ≥ 10, uninformative ligation products, and PCR duplicates using HiC-Pro (Servant et al., 2015).

Valid Hi-C read pairs from biological replicates of Elf1 and H1 hESCs were combined, respectively, and used to generate raw chromosome-wide interaction matrices binned at a resolution of 40kb. Raw matrices were ICE-normalized using the HiTC Bioconductor package (Servant et al., 2012) for R, and TADs and boundaries were identified using TopDom (Shin et al., 2016) with a window size of 5. X and Y chromosomes were removed for the datasets for all Hi-C analyses.

Insulation scores were calculated from the ICE-normalized matrices of both the Elf1 and H1 hESCs, separately. Insulation vectors were detected via cworld (Giorgetti et al., 2016) using the script matrix2insulation.pl, and using the following options: (--is 240000 --nt 0.1 --ids 160000 --im median --

bmoe 0). Differential insulation scores computing Elf1 score minus H1 score for the all autosomes via cworld using the script compareInsulation.pl, with inputs being the two insulation scores above.

High-Confidence SMC1 ChIA-PET interactions for naïve and primed hESCs were downloaded as a supplemental table (Ji et al., 2016). A ChIA-PET was considered to span a TAD if both PET termini were located within 40 kb of a TAD boundary.

Spatial compartments and activity status were identified via principal component analysis (PCA) using Homer Tools(Heinz et al., 2010). Processed Hi-C reads were imported into Homer. For each chromosome, a contact matrix was constructed at 40 kb resolution and normalized using a sliding window of 400 kb as background. Next, the correlation between intra-chromosomal contact profiles was computed and the first principal component (PC1) vector was extracted and saved as a bedGraph file. H3K27ac ChIP-seq peaks served as a seed for determining which regions are active (PC1 > 0). A genomic region was considered cell type-specific if it met the following three criteria: 1) the average PC1 value was positive in one cell type and negative in the other, 2) the difference in the average PC1 value was > 50 and 3) the correlation between contact profiles was < 0.4. Randomization was achieved by selecting coordinates from a pool of 40 kb regions that had associated PC1 values and were not located within any cell type-specific sub-compartments.

Supplemental References

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091-1093.

Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E.*, et al.* (2016). Structural organization of the inactive X chromosome in the mouse. Nature *535*, 575-579.

Hawkins, R.D., Larjo, A., Tripathi, S.K., Wagner, U., Luu, Y., Lonnberg, T., Raghav, S.K., Lee, L.K., Lund, R., Ren, B.*, et al.* (2013). Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. Immunity *38*, 1271-1284.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Ji, X., Dadon, D.B., Powell, B.E., Fan, Z.P., Borges-Rivera, D., Shachar, S., Weintraub, A.S., Hnisz, D., Pegoraro, G., Lee, T.I.*, et al.* (2016). 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell *18*, 262-275.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Kyttala, A., Moraghebi, R., Valensisi, C., Kettunen, J., Andrus, C., Pasumarthy, K.K., Nakanishi, M., Nishimura, K., Ohtaka, M., Weltner, J.*, et al.* (2016). Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. Stem Cell Reports *6*, 200-212.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. In Bioinformatics, pp. 841-842.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187-191.

Servant, N., Lajoie, B.R., Nora, E.P., Giorgetti, L., Chen, C.J., Heard, E., Dekker, J., and Barillot, E. (2012). HiTC: exploration of high-throughput 'C' experiments. Bioinformatics *28*, 2843-2844.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol *16*, 259.

Shin, H., Shi, Y., Dai, C., Tjong, H., Gong, K., Alber, F., and Zhou, X.J. (2016). TopDom: an efficient and deterministic method for identifying topological domains in genomes. Nucleic Acids Res *44*, e70.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Valensisi, C., Andrus, C., Buckberry, S., Jayavelu, N.D., Lund, R.J., Lister, R., and Hawkins, R.D. (2017). Epigenomic landscapes of hESC-derived neural rosettes: modeling neural tube formation and diseases. Cell Rep *20*, 1448-1462.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. In (Springer-Verlag New York).

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.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.