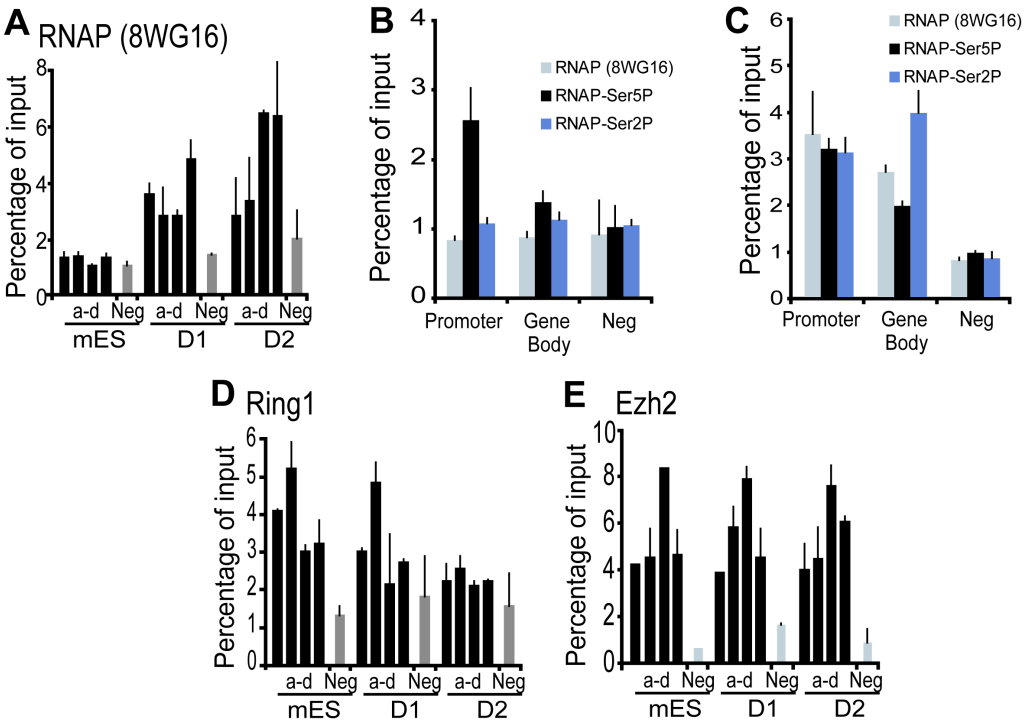
The histone demethylase Jmjd3 sequentially associates with the transcription factors Tbx3 and Eomes to drive endoderm differentiation

Apriliana E. R. Kartikasari, Josie X. Zhou, Murtaza S. Kanji, David N. Chan, Arjun Sinha, Anne Grapin-Botton, Mark A. Magnuson, William E. Lowry and Anil Bhushan

SUPPLEMENTARY INFORMATION

Inventory of Supplementary Information:



Supplementary Figure 1. Release of paused RNAP is accompanied by eviction of PRC1.

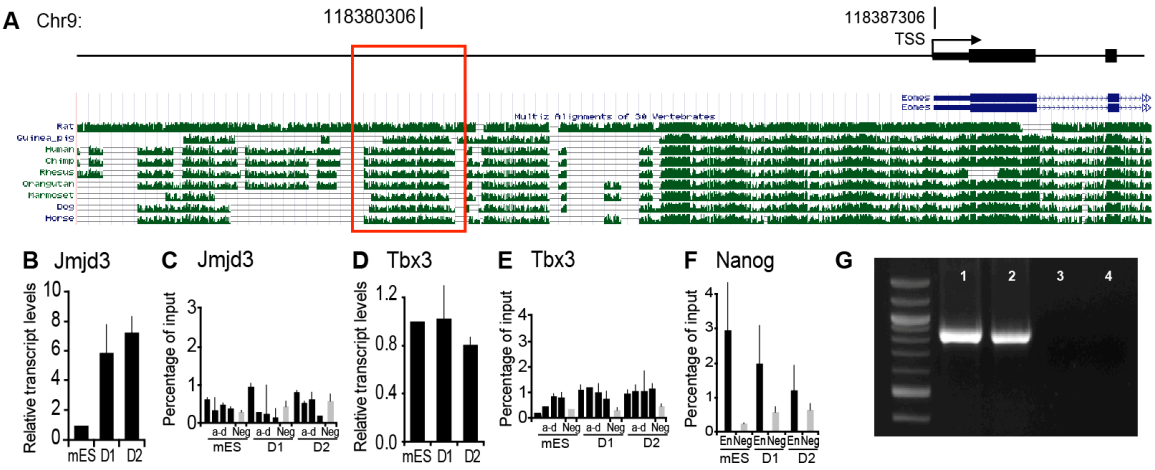
(A) RNAP (8WG16) enrichment on the *Eomes* promoter is only detected in D1-D2 differentiation.

(B) In D0, a paused conformation of RNAP was detected in the *Eomes* locus. The binding of RNAP (8WG16), RNAP-Ser5P and RNAP-Ser2P in the *Eomes* promoter and the gene body region was analyzed using the ChIP analyses. Only RNAP-Ser5P was detected primarily in the promoter region, indicating the presence of the paused RNAP in this region.

(C) In D2, the elongation form of RNAP (RNAP-Ser2P) was highly abundant both in the promoter and the gene body of *Eomes*, analyzed by ChIP.

(D-E) The release of paused RNAP is accompanied by eviction of Ring1, the core component of PRC1, but not Ezh2, the core component of PRC2, during D1-D2 differentiation, shown by ChIP analyses.

All values are mean±SEM (n=2-3).



Supplementary Figure 2. *Eomes* enhancer in mammals and recruitment of Tbx3-Jmjd3 complex during early differentiation.

(A) *Eomes* enhancer region (red box) is highly conserved in mammals as shown by sequence alignment of the *Eomes* locus proximal region by the UCSC genome browser.

(B) *Jmjd3* expression is upregulated in D1-D2-differentiated cells.

(C) Jmjd3 is not recruited to the *Eomes* promoter proximal region during early differentiation. Jmjd3 recruitment to *Eomes* promoter (a-d) and negative region (Neg) were analyzed in ES cells and D1-D2-differentiated cells.

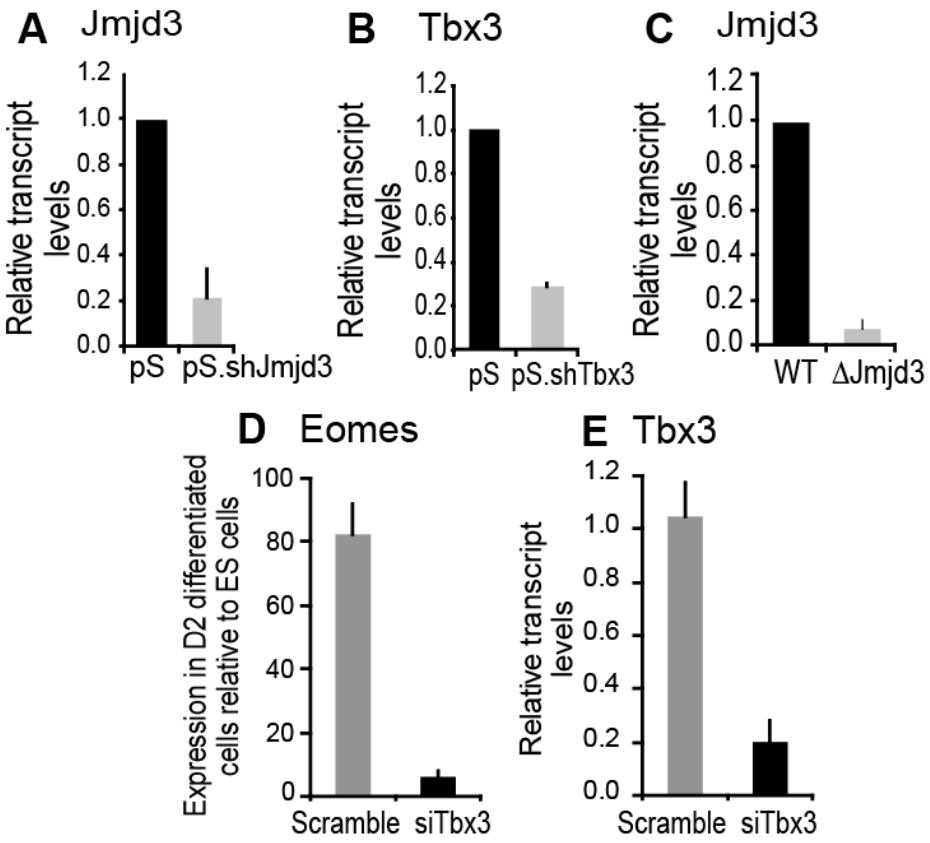
(D) *Tbx3* expression is sustained in D1-D2-differentiated cells.

(E) Tbx3 is minimally recruited to the *Eomes* promoter proximal region during early differentiation.

(F) Nanog binding to *Eomes* enhancer declines as ES cells are differentiated.

(G) *Eomes* enhancer region of 850bp was amplified to test the presence of genomic DNA in the cell lysate following DNAse1 treatment. Template DNA was isolated from cell lysate without (lane2) or with (lane3) nuclease treatment using DNase1. Lane1 was PCR from genomic DNA control. Lane4 was water control.

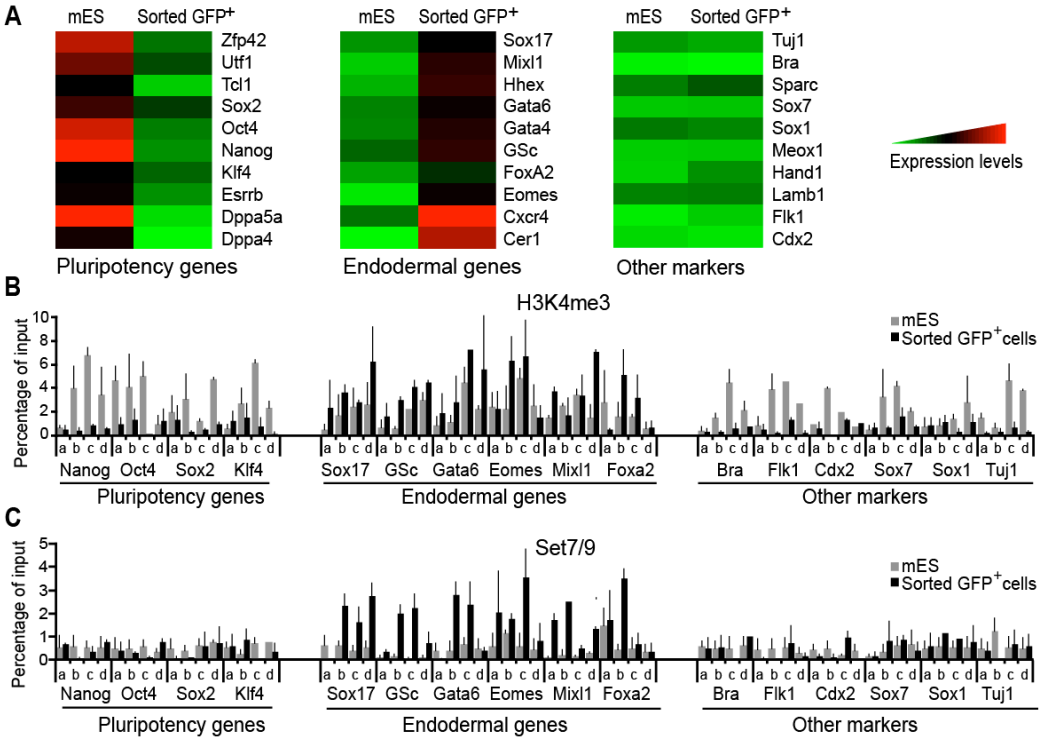
Values are mean±SEM (n=2-3).



Supplementary Figure 3. The T-box protein, Tbx3, is required for *Eomes* activation

Mean±SEM (n=2-3) of *Jmjd3* and *Tbx3* in D2-differentiated *Jmjd3*-knockdown (pS.shJmjd3) (A), *Tbx3*-knockdown (pS.shTbx3) (B) and *Jmjd3*-null (Δ*Jmjd3*) (C) cells are reduced compared to their respective controls.

Mean±SEM (n=2-3) of *Eomes* (D) and *Tbx3* (E) expression is lower in D2-differentiated cells transfected with si*Tbx3* SMARTpool compared to the si*Scramble* control as determined by qPCR.



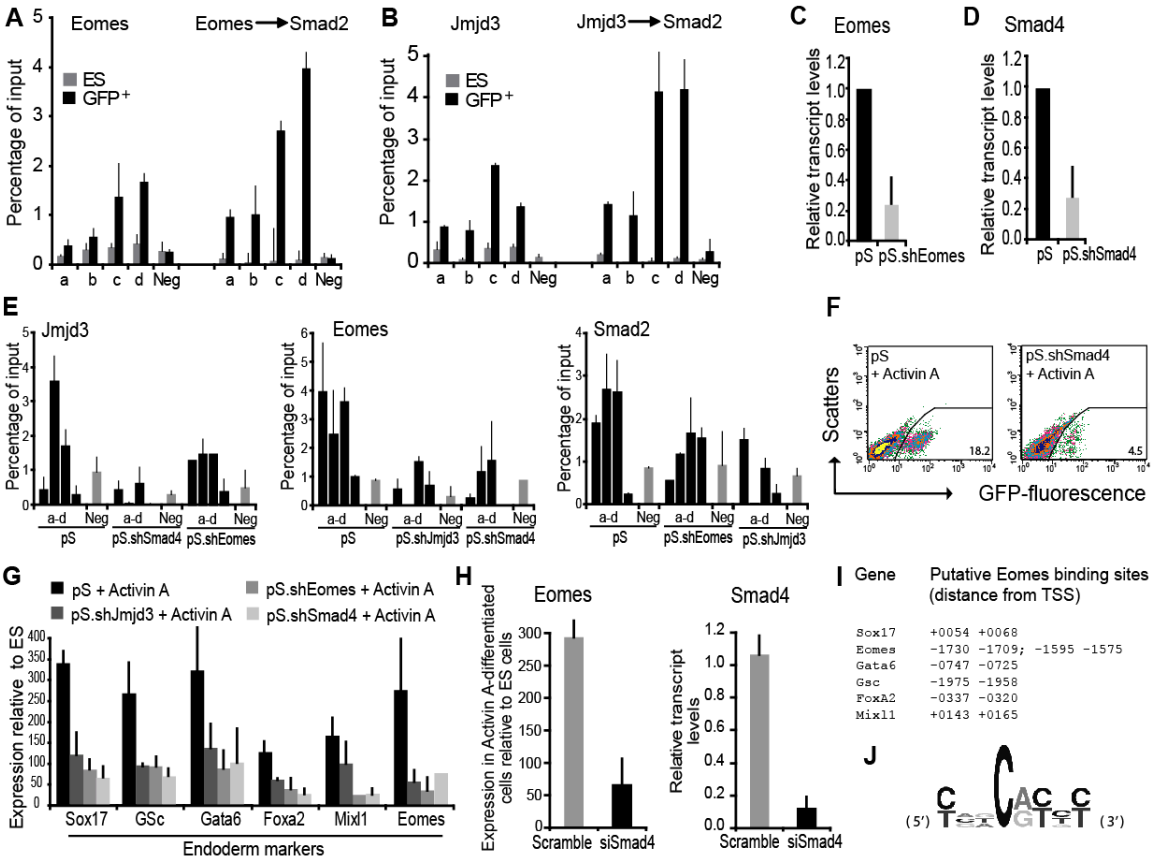
Supplementary Figure 4. The core definitive endoderm genes are selectively expressed in Sox17.GFP$^+$ cells.

(A) Microarray expression analyses of ES and the sorted Sox17.GFP$^+$ cells. The levels of gene expression are shown as green (low expression) to red (high expression).

(B) Promoters of definitive endoderm genes are marked by H3K4me3. ChIP analyses of H3K4me3 were performed using chromatin from ES and the sorted Sox17.GFP$^+$ cells. Four pairs of primers at the TSS-promoter region (a,b,c,d) were used for each promoter.

(C) Set7/9 selectively occupies the promoters of endodermal genes in the sorted Sox17.GFP$^+$ cells.

All values are mean±SEM (n=2-3).



Supplementary Figure 5. Eomes directs Jmjd3 and Smad2 to the promoters of core definitive endoderm regulators.

(A, B) Jmjd3, Eomes and Smad2 coocupy the *Eomes* promoter. Cooccupancy of Jmjd3, Eomes and Smad2 in the *Eomes* promoter region (a, b, c, d) and the negative control region (Neg) was analyzed using the sequential ChIPs. Eomes (A, left panel) and Jmjd3 (B, left panel) enrichment, as well as the sequential Smad2 enrichment (A and B, right panels) at the *Eomes* promoter were observed in sorted GFP+ cells but not in undifferentiated mES cells.

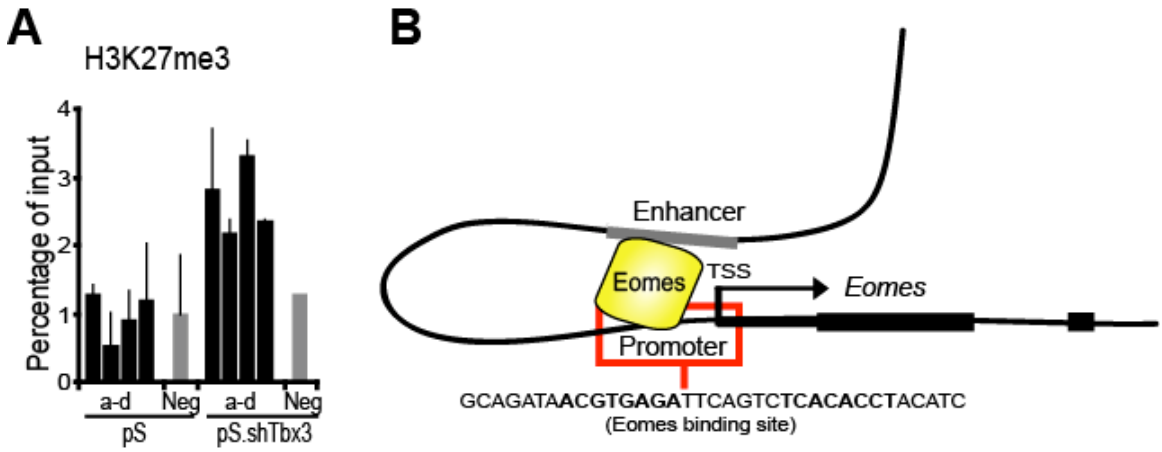
(C, D) *Eomes* and *Smad4* are efficiently knocked-down using shRNAs. Expression of *Eomes* (C) and *Smad4* (D) in D2-differentiated knockdown cells were determined by qPCR.

(E) Jmjd3, Eomes and Smad2 occupancy on the *Eomes* promoter are dependent on each other. ChIP analyses were performed using Activin A-differentiated cells on the promoter (a-d) and the negative region (Neg).

(F) Flow cytometry analyses show lower number of Sox17.GFP^{+} cells in Smad4 knockdown (pS.sh*Smad4*) cells (right panel) compared to the control (pS) cells differentiated with Activin A (left panel). (G) Mean of definitive endoderm gene expression±SEM in the Activin A-differentiated knockdowns are reduced in the knockdowns compared to the controls. (H) Expression of *Eomes* was also reduced in the Activin A-differentiated cells transfected with si*Smad4* SMARTpool compared to the si*Scramble* control.

(I, J) Putative T-box binding sites are present in the promoter regions of definitive endoderm genes. (I) The position of putative T-box binding sites on the endodermal promoters relative to the transcription start site (TSS). (J) The core motif of Eomes binding site formulated using our exhaustive motif search algorithm.

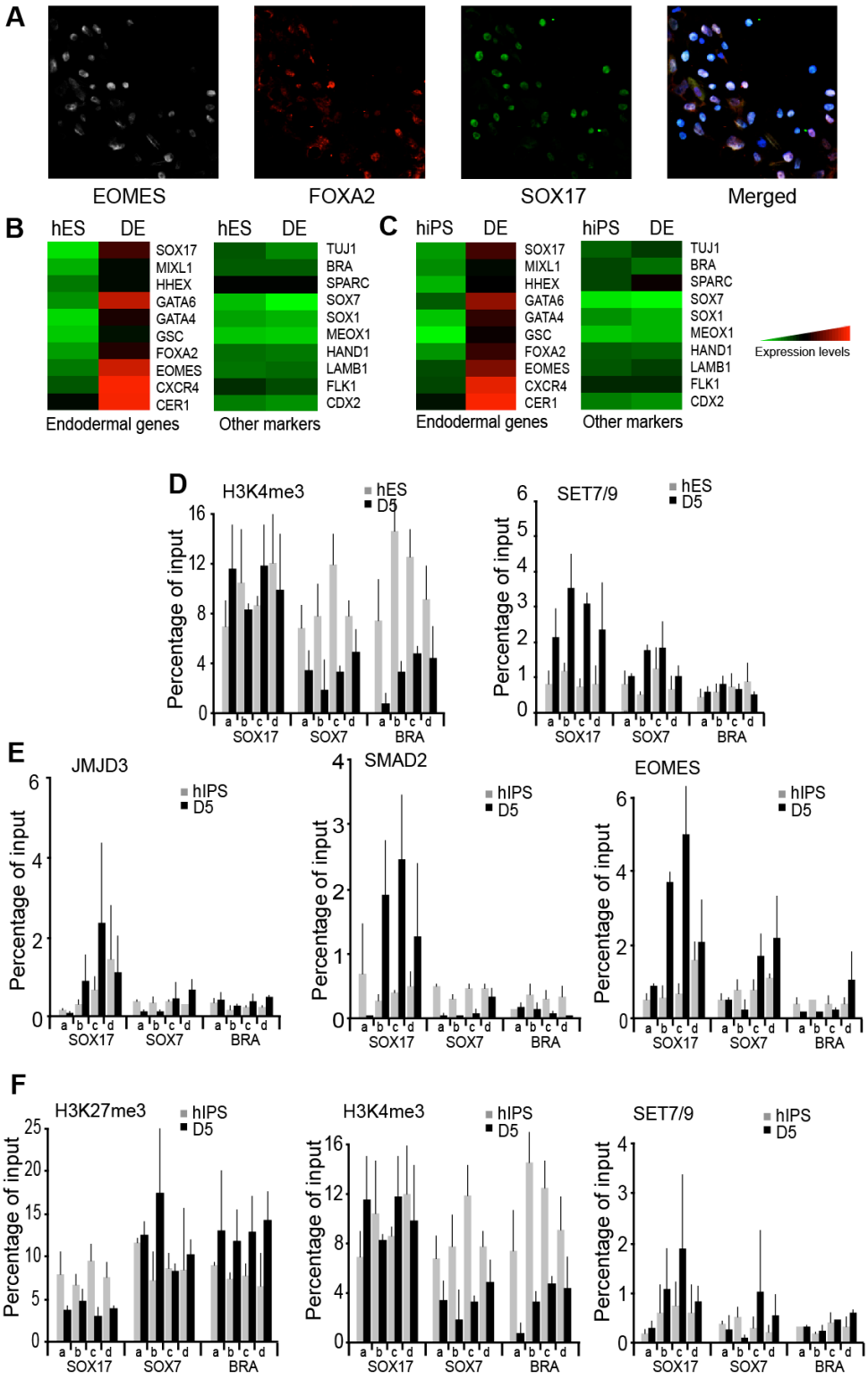
All values are mean±SEM (n=2-3).



Supplementary Figure 6. Tbx3 is required for *Eomes* gene activation and chromosomal interaction between enhancer and promoter of the *Eomes* locus.

(A) H3K27me3 is enriched at the *Eomes* promoter (a-d) and not at the negative region (Neg) in Activin A-differentiated *Tbx3*-knockdown cells but not the control cells as determined by ChIP analyses. Values are mean±SEM (n=2).

(B) Model of DNA looping in the *Eomes* locus with a putative Eomes binding site on the promoter region of the *Eomes* locus identified using our exhaustive search algorithm.



Supplementary Figure 7. The sequential two-step activation of *EOMES* locus is conserved in hiPS derived from fibroblast.

(A) Colocalization of EOMES, FOXA2 and SOX17 in the Activin A-differentiated hES.

(B, C) Microarray expression analyses of hES, hiPS and their respective Activin A-induced definitive endoderm. The levels of gene expression are shown as green (low expression) to red (high expression).

(D) H3K4me3 and SET7/9 are enriched on the promoter of *SOX17* but not *SOX7* and *BRA* in hES-derived definitive endoderm.

(E) JMJD3, SMAD2, EOMES and (F) H3K27me3, H3K4me3 and SET7/9 are enriched on the promoter of *SOX17* but not *SOX7* and *BRA* in hiPS-derived Activin A-induced definitive endoderm. Four pairs of primers at the TSS-promoter region (a,b,c,d) were used to perform ChIP analyses for each promoter.

Values are mean±SEM (n=2-4).

SUPPLEMENTARY MATERIALS AND METHODS

Flow cytometry and cell sorting

Prior to flow cytometry and cell sorting, single cell suspension was prepared by dissociation using Trypsin (Invitrogen) and gentle pipetting. Cells were then resuspended at approximately $1x10^5$ cells/mL in PBS supplemented with 3% FBS. Live cells, 20.000 events per analysis, were analyzed for GFP or mCherry fluorescence using four-laser digital LSRII analyzer (BD Bioscience) or sorted for GFP^+ using FACSAriaII benchtop cell sorter (BD Bioscience)**.**

Coimmunostaining

Coverslips were fixed in 4% PFA, blocked for 1 h in 10% serum + 0.1% Tween 20 (or 0.1% Triton-X-100), then incubated overnight at 4 °C with primary antibodies. Following primary antibody incubation, the coverslips were incubated with Alexa Fluor secondary antibodies (Invitrogen) at room temperature for 1 h and mounted in Prolong Gold (Invitrogen). All imaging was performed on Zeiss Axio Imager A1.

RNA extraction, reverse-transcription and real-time quantitative PCR

Total RNAs were extracted using Absolute RNA Microprep Kit (Agilent Technology) or RNeasy Plus Mini Kit (Qiagen) and digested with RNase-free DNase (Fermentas) to eliminate DNA. The quality and quantity of the extracted RNA were assessed using Nanodrop (Spectrophotometers) and Bioanalyzer (Agilent Technologies). 0.5 µg of total RNA from each sample was denatured at 65^0C and then reverse transcribed using SuperscriptIII reverse transcriptase (Invitrogen) at 50^0C for 1 hour. Real-time quantitative PCR (qPCR) was performed using ABI7900HT (Applied Biosystems) with initial denaturation at 95^0C for 20 seconds, followed by 50 cycles of 94^0C for 1 second and 60^0C for 20 seconds, then continued with a dissociation stage. Each qPCR reaction contained 1x FastSybrGreen Mix (Applied Biosystems), 500 nM of each primer and 100 ng cDNA. Relative expression of a certain target gene was determined using the comparative cycle threshold (Ct) method, where the amount of target cDNA was normalized to the internal control, cylophilin cDNA, and expressed as relative to the baseline normal control cDNA ($2^{-\Delta\Delta Ct}$). Data are representative of two to three independent experiments, and error bars indicate standard error of the replicates. Data are

statistically analyzed using Student's t-test followed by Bonferroni correction for multiple comparisons. Primer sequences are included in Supplementary Table 2.

Microarray profiling

Two to three biological replicates for undifferentiated and Activin A-differentiated stem cells were hybridized to GeneChip Human Gene 1.0ST Array or GeneChip Mouse Gene 1.0ST Array (Affymetrix) depending on the cell source. Hybridization was done at UCLA Microarray Core Facility. Raw data obtained by Affymetrix scanner passed data quality control steps prior to RMA normalization using the Affymetrix Suite Analysis software. The normalized data were analyzed for differential gene expression with statistical analysis using Z-test FDR-adjusted *P*-values. Heatmaps of the fold-changes were generated using Mayday microarray data analysis as described in (Battke et al., 2010).

Eomes binding motif analysis

Mathematica7.0 was used to build an exhaustive motif search algorithm to map the Eomes binding motif on mouse promoters between (-2500 to +500) based on our ChIP-Eomes binding profile and the previously established Eomes binding specificity (Conlon et al., 2001). The extracted consensus sequences were graphically represented using Weblogo (UC Berkeley). Mouse *Sox17* promoter (-683 to +82) was cloned into pmCherry-1 vector (Clontech) containing the identified Eomes binding motif (+54 to +68). A mutated version of the Eomes element was also cloned into pmCherry-1 vector. The resulting constructs were transfected to Hek293T using Lipofectamine 2000, followed by 100 ng/mL Activin A treatment for 3 days in serum free DMEM as indicated. Promoter activity reflected by mCherry fluorescence was analyzed using flow cytometry. Protein binding to the promoter constructs was characterized by ChIP, using a pair of primers (see Supplementary Table 2) to amplify a region between the 3' end of the promoter and the mCherry coding sequence.

Chromatin conformation capture

Chromatin conformation capture (3C) was performed according to (Gavrilov et al., 2009) with minor modifications. Briefly, single cell suspension of $1x10^6$ cells was cross-linked with 1% formaldehyde for 15 minutes at room temperature. The reaction was quenched with 125 mM glycine. The cells were washed once with ice cold PBS and then lysed with

5 mL of ice cold lysis buffer containing 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% NP-40 and 1x complete proteinase inhibitors. Following 20 minutes incubation on ice, nuclei were pelleted and resuspended in 0.5 mL 1.2x MspI restriction buffer (Fermentas). SDS was added to 0.3% final concentration and nuclei were incubated for 1 hour at 37^0C with vigorous shaking. Then, TritonX-100 was added at a final concentration of 1.8% and nuclei were incubated for 1 hour at 37^0C with shaking. The sample was incubated with 1000 Units of MspI, overnight at 37^0C with shaking. Similar digestion efficiency was obtained from each sample preparation, assessed by qPCR analysis with primers, 3C-1 and 3C-2, spanning the MspI site. MspI was then inactivated with addition of SDS (1.3% final concentration) and shaking for 20 minutes at 65^0C. Nuclei were incubated with 7 mL of 1x ligation buffer and 1% final concentration of TritonX-100, for 1 hour at 37^0C. Samples were then incubated with 100 Units of T4 DNA ligase (Fermentas) at 16^0C overnight with slow agitation. Protein digestion and reverse cross-linking was performed overnight at 65^0C in the presence of Proteinase K with agitation. RNA was degraded with RNase for 30 minutes at 37^0C. DNA was then phenol/chloroform extracted and ethanol-precipitated. BAC clone RPCI.23C-448K1 (Invitrogen) containing mouse *Eomes* locus and BAC clone RP11-1107O23 (Invitrogen) containing human *EOMES* locus were used as positive control. Ten μg of the BAC clone were digested with 500 Units of MspI overnight at 37^0C with agitation. The digested DNA was phenol chloroform extracted and ethanol precipitated. These DNA fragments were ligated with 50 Units of T4 DNA ligase at 16^0C overnight with slow agitation. The ligated BAC DNA was then extracted with phenol/chloroform and precipitated with ethanol. Taqman probes flanking the *Eomes/EOMES* enhancer region and primers were designed using Primer Express 3.0 (Applied Biosystems). All primers were designed to anneal at and face the same ends of the MspI sites, have melting temperatures of 58-60^0C and yield amplicons of 100-150 bp. Each PCR efficiency was determined based on the calibration curve generated using qPCR of series of successive dilutions of the BAC preparation. Primer pairs with similar PCR efficiencies were chosen to analyze 3C. qPCR was performed using ABI7900HT with initial denaturation at 95^0C for 3 minutes, followed by 50 cycles of 94^0C for 5 seconds and 60^0C for 20 seconds. The quality and quantity of each 3C samples were analyzed using Nanodrop (Spectrophotometers). qPCR was done in triplicate for each primer pairs and 20 ng of 3C DNA was used for each qPCR reaction. The interaction frequencies were calculated based on the amount of PCR products obtained with the 3C primer pairs, followed by normalization with the log ratio of the interaction frequencies

between fragments in the control region, to correct for the crosslinking and ligation efficiencies and the amount of the templates. The control region was the anchor probe primer (3C-1) with the primer of the adjacent restriction fragment (3C-2). Data are representative of two independent experiments, and error bars indicate standard error of the replicates. Data are statistically analyzed using the Student t-test, followed by Bonferroni correction for multiple comparisons. Genomic coordinates are based on build mm9 for mice and hg19 for human. Probe and primer sequences are included in Supplementary Table 2.

Supplementary Table 1. shRNA sequences used in this study

Targeted gene	shRNA sequence (5' to 3')
Jmjd3	**CCTCTGTTCTTGAGGGACAA**TTCAAGAGA**TTGTCCCTCAAGAACAGAGG**
Eomes	**ACTCACCAATAACAAAGGTG**TTCAAGAGA**CACCTTTGTTATTGGTGAGT**
Smad4 (Deckers et al., 2006)	**CCAGCTACTTACCATCATA**TTCAAGAGA**TATGATGGTAAGTAGCTGG**
Tbx3 (Han et al., 2010)	**GAATTGTACATAGCTGGAT**TTCAAGAGA**ATCCAGCTATGTACAATTC**

Supplementary Table 2. Probe and primers used in this study

Species	Name	Approximate distance from TSS	Sequence (5' to 3')	Amplicon size (bp)
Mouse	Eomes -Enhancer	-7000	GCATTGCCTTCCGTCTACTCA TTAAGGCCCCATTTCCTGAA	106
Mouse	Eomes-A	-1500 to -1000	CTGATCGCGCAGGATGTCA GTAGTGGGTCTTGGGCGCT	101
Mouse	Eomes-B	-1000 to -500	CGGAATCCTTTTGCCATCTC GGAACCGCTTTTTGTCCTGTC	101

Mouse	Eomes-C	-500 to 0	GCCGCACATATATAAGCAGCC	101
			AGCAACCAGCCATTTCCTCTC	
Mouse	Eomes-D	0 to +500	GCGCCCACGTCTACCTGTG	101
			TCCACCCTGCACTCACCTG	
Mouse	Negative	Chr13	GCCACAGTTAGCCATCTGCA	101
			GCCCTGACCTCTGGACCCTA	
Mouse	Eomes-body	+7000 to +7500	TTCGGGAAGAGAAGGGTTGG	101
			TTTCCCCGCTTGACTCTCAG	
Mouse	Sox17-A	-1500 to -1000	TGTGAGCAAGGCTTCATTCCT	142
			ACCGGTCGGGAATGTTTCTT	
Mouse	Sox17-B	-1000 to -500	AGGGTCTGGCCTGAATGTTCT	101
			GGAGGTGGAGATGGAAATTCG	
Mouse	Sox17-C	-500 to 0	CTTGGCCATCAGAACACCAA	114
			GTGATCAACTGGGACTCCGAA	
Mouse	Sox17-D	0 to +500	TCTGAGAGGCAGGAGAGCTACA	101
			ATCGATTCTCTAGCCAGGTGCT	
Mouse	GSc-A	-1500 to -1000	TGGAAGGACAGCTAGAGGCC	101
			CGTGGGATCTACCGGGACT	
Mouse	GSc-B	-1000 to -500	ACAACTTGTACGAAGGCGGAC	101
			CCCTCCGTCTCTTTCCCAA	
Mouse	GSc-C	-500 to 0	CCAGGAACTAAATCCAGGATGA	101
			CGCCTAGCTGAACTCAACCC	
Mouse	GSc-D	0 to +500	TCTCTTTCGGTTTGCTCGC	101
			GGACGCGCCGAAGAAG	
Mouse	Gata6-A	-1500 to -1000	AAAGAACCCCTTCGCAACG	101
			TTGGAGAAAGGAGAGGGATCC	
Mouse	Gata6-B	-1000 to -500	ACGCGACAACCAAACTAGCG	101
			TTCGGTTGTGGTGTATGGGA	
Mouse	Gata6-C	-500 to 0	CCCCACCCGCTTTTTTTCTC	96
			CAGGCTGTGGGTCGGAACT	
Mouse	Gata6-D	0 to +500	CTCGGTGAGTCCAATCAGGAG	101
			GGGTCTGCTAGTCGCTGCTG	
Mouse	FoxA2-A	-1500 to -1000	CTTTGCTCCGCGTACCCACT	101

			CTGCTCTTTGATTGTGCTCCC	
Mouse	FoxA2-B	-1000 to -500	CACCTGGATTTGGGCCAACTA	101
			TTCTTGGTTCTGAAAATGTGACTAGG	
Mouse	FoxA2-C	-500 to 0	CTGCCCTACCCCCACCTACTG	101
			ATGAGGGCCTCGGTGTTTCA	
Mouse	FoxA2-D	0 to +500	CCCTTCGCAACGAGGTCTCT	81
			AGGGATCCAAGCAGATTGAAAC	
Mouse	Mixl1-A	-1500 to -1000	AGCTGGTGGACAGGTCAAGC	101
			TGGAAGAGCAGCAAGTGGC	
Mouse	Mixl1-B	-1000 to -500	CCTCAAATGAACCCCAGGG	109
			CAGTGCTCAGAACCCGAGAAG	
Mouse	Mixl1-C	-500 to 0	TTGAGAGGTCTTCTGGCCCA	99
			CAATCCACATACCTCGAACCG	
Mouse	Mixl1-D	0 to +500	AGTTGCTGGAGCTCGTCTTCC	102
			TCACCTGGATCCTGGACTCG	
Mouse	Oct4-A	-1500 to -1000	CAGACGGCAGATGCATAACAA	101
			CAGATTAAGGAAGGGCTAGGACG	
Mouse	Oct4-B	-1000 to -500	GCACCTAGAGCCACTGACCC	102
			CAAAAAGCCCAGAGCACCC	
Mouse	Oct4-C	-500 to 0	GAAAATGAAGGCCTCCTGGG	101
			AAACTGAGGCGAGCGCTATCT	
Mouse	Oct4-D	0 to +500	TAGTCCCCCAAGTTGGCGTG	101
			GTCGGCACAGGGCTCAGAG	
Mouse	Nanog-A	-1500 to -1000	GAGAGGACGGCCCTTCCCT	101
			TGCCGATCAGTCCTTGTGCT	
Mouse	Nanog-B	-1000 to -500	GAGGCCAGTCTCGGGTACGT	101
			GCCTCAGCCGTCTAAGCAATG	
Mouse	Nanog-C	-500 to 0	CAATGTCCATGGTGGACCCT	101
			TACCCACCCCCTATTCTCCC	
Mouse	Nanog-D	0 to +500	TCTGGGAACGCCTCATCAAT	101
			TCCTTACCAGCCTCTGTGCAG	
Mouse	Sox2-A	-1500 to -1000	GGGCTTTGTTTGACTCCGTGT	101
			ATCGGCAGGCGCTCAAAA	

Mouse	Sox2-B	-1000 to -500	TGGGAGTTAGAAAGAGCGGGA GCACGTGTCCTCAACCCCTA	101
Mouse	Sox2-C	-500 to 0	GCCAATATTCCGTAGCATGGG CCCCGTTTTCAGCAACAGGT	101
Mouse	Sox2-C	0 to +500	GAGAGAAGTTTGGAGCCCGA CGAAGTGCAATTGGGATGAAA	106
Mouse	Klf4-A	-1500 to -1000	CCTTCCTCAAAGGCTCCCA CCGACCACAGAAGCGCTCTT	103
Mouse	Klf4-B	-1000 to -500	CCTCCCAGTGAAGTCCCTTTG ATAAAGAGGTGAAGCGGCGAG	101
Mouse	Klf4-C	-500 to 0	GCCTTCCTTGTTCCTTCCATT GGCGCTAGGAATCTCAGAAGC	109
Mouse	Klf4-D	0 to +500	TCTGCCTTGCTGATTGTCTATTTT TTCAATATAAACCGGCGATGTCTT	101
Mouse	T-A	-1500 to -1000	GCCGAACACCCCAGTATTCA CCCTACTTCTGGGAGAGACGG	103
Mouse	T-B	-1000 to -500	ATGCGTTCCAACAATCCCC GCGTCCCACATAAAAGCTGG	101
Mouse	T-C	-500 to 0	GTGCAGGGACCCAGGTGTAA TGACTCCCTACCCAACAGCC	101
Mouse	T-D	0 to +500	TTTCTTGGAAAAGCGGTGGC CCGGCGAGCTCATCCTC	101
Mouse	Flk1-A	-1500 to -1000	TCCAGGCAGGAGAGAGCATT TGACGTTCCAAACTACCCAGG	102
Mouse	Flk1-B	-1000 to -500	CCTACCTCCTGGGTTGGTCA CGCAGCCTTTCGTACATTCTC	103
Mouse	Flk1-C	-500 to 0	CTGTGATCTGGTCGCTGCC GCTACGCACACAGTCCCCAC	102
Mouse	Flk1-D	0 to +500	TCCCCAGACAGCATGGTGAG GCGCACGGAGCTAGGATACAT	101
Mouse	Cdx2-A	-1500 to -1000	TATGCGCAGCCGCTCTATC CATTCATTTCACAAGAGCCGC	101
Mouse	Cdx2-B	-1000 to -500	GGAGCTATAATCCGCCACCG	101

			CGAGTCGCTTGAGGTGGGA	
Mouse	Cdx2-C	-500 to 0	GTCAGCATGAGGGAGATCGTG	101
			GGGTCTATGCCGTGCTCATC	
Mouse	Cdx2-D	0 to +500	ACAGTGTTCGGTCTGGCGC	101
			CGTATAGTGCCCTTTGGCCC	
Mouse	Sox7-A	-1500 to -1000	TTACCTCATCACCGCCTGTG	101
			TGAGTTACTGAATGCGTGACAGC	
Mouse	Sox7-B	-1000 to -500	TCCCACATCCACCTAAGACCC	101
			ATCGGCCAAGTATGCTGTTGT	
Mouse	Sox7-C	-500 to 0	CACAACAGCATACTTGGCCG	101
			AACTCCCGCAAAGGAAAACC	
Mouse	Sox7-D	0 to +500	ATGAGAGGAAACGTCTGGCAG	101
			CATCCTCGCAACAGCTTCGT	
Mouse	Sox1-A	-1500 to -1000	TTGATCATCCCCAGAACTGCC	101
			CGGAATTGGAAGGTCCAGTG	
Mouse	Sox1-B	-1000 to -500	GGAGCGTAGCACCACACAAA	101
			GGTGCATTCTCAGAGCCTGC	
Mouse	Sox1-C	-500 to 0	CCTCCTTCCTGCCCTTGTTC	101
			CGCTGGCGAGTGACG TG	
Mouse	Sox1-D	0 to +500	CCTTGCTAGAAGTTGCGGTCC	101
			CGACGCAGGGTTCGCTG	
Mouse	Tuj1-A	-1500 to -1000	GCATTTGGTTTTTGCCAGGAG	102
			CAGCCTTCACCCCTCGTTTT	
Mouse	Tuj1-B	-1000 to -500	TCTTCTCTGTGCATTTGCCCT	101
			TCGGTTGTGAACTCCTGCCT	
Mouse	Tuj1-C	-500 to 0	AGAAACCGAGCTGCGTCCA	101
			TTTGGGTTTCCCGGTTTCC	
Mouse	Tuj1-D	0 to +500	GGCGCTCTCTTTCTGCCCT	102
			CAGTCTGGGCTTTCAGTCCTG	
Mouse	3C-1	-7500 to -7000	CTCACGGGTCTAAGCAAGCATGATGGCTCAT	
Mouse	3C-2	-7000 to -6500	CTTAAGGCCCCATTTCCTGAA	
Mouse	3C-3	-6000 to -5500	TCTCTCTCTCTTCCTTGGTTATCTGA	
Mouse	3C-4	-1000 to -1500	GTGTGATCGCATTGGGTTGGG	

Mouse	3C-5	-1000 to -500	AAAACCGAGCAGAGGGAATTCT	
Mouse	3C-6	-1000 to -500	GAGGGAATTCTGAGTAATGAAAGTG	
Mouse	3C-7	-500 to 0	TCTCCAGCCCTCCCCTAGAT	
Mouse	3C-8	-500 to 0	AGACCATGTTCGCAGACTTCAA	
Mouse	3C-9	0 to +500	GTCCTCTTCTTCCTTAGTGTGAG	
Mouse	3C-10	0 to +500	GAAGGAAGGAGAGAGACACCTACAA	
Mouse	3C-11	0 to +500	TCCCCCAACTGGCCTTTATAA	
Mouse	3C-12	+1000 to +1500	CCCACGTTACTCTCCTCTCTAGTCA	
Mouse	3C-13	+1500 to +2000	CAAAGAGGGCTCGTTGAGAGA	
Mouse	3C-14	+2000 to +2500	GCCATTTTTGTTTGTTTGCTGTT	
Mouse	3C FAM-TAMRA	-7500 to -7000 (Probe)	TCCCTGGTGGTCACTGACTGCATTTCATA	
Human	EOMES -Enhancer	-7000	GGCCTTGGGAAAAACAGAAT GGAGAGGTGCACTTCAAAGC	101
Human	EOMES-A	-1500 to -1000	CCAAAAGCTTAGCGCGAGTT AACTCCCTGGCCTTCAACG	101
Human	EOMES-B	-1000 to -500	CAAGCTGATAGCTTGCAGGGA ACAGCTTCCGACCAGGAAAG	101
Human	EOMES-C	-500 to 0	AACTTGACCGATGCTTTGGC GGCTCACACGCTGGAAGAAG	101
Human	EOMES-D	0 to +500	GGCCCCCTCTCCTCAGAAG GGAGAGACTGCCGGAAAACTT	101
Human	Negative	Chr13	CTCCCAAATTGCTGGGATTA ATTCCAGGCACCACAAAAAG	156
Human	SOX17-A	-1500 to -1000	ACTTGTCCCCTAGCCAAGGC TCTGGCATTGCTTTTTCTCAAA	101
Human	SOX17-B	-1000 to -500	AGAGCTGCAAGTCTGAAGGCG GAAAGGAAACGGAGGCAGAGA	101
Human	SOX17-C	-500 to 0	TGAAACACCCTCGGCTTCTG TGGAATCGACGTACTTGCCC	101
Human	SOX17-D	0 to +500	AAGATGCTGGGTGAGTCCGA CAGGCTCGCAAAGAACAGTTT	102
Human	SOX7-A	-1500 to -1000	TTCACCTTACACTTGCCGGG	101

			AGCCTGAGTCCGGATTTCAAT	
Human	SOX7-B	-1000 to -500	TGGGTGGCGCTCAGTAAATT	101
			GCCTGCAGAATGAGTGAGCA	
Human	SOX7-C	-500 to 0	GGCGGCCCAAGTTGACTG	101
			CCATGAACCCCTCAGAGCAG	
Human	SOX7-D	0 to +500	GCGGTGCCCAGGATTCAG	101
			GCCGATCTCCGCTCCTGTC	
Human	T-A	-1500 to -1000	CTGGGAACCGAGGAACAAAG	102
			TTGGCCAGCACTTGTGTAGTG	
Human	T-B	-1000 to -500	TAACACGCACCACGCTCTCC	102
			GGCACCAAAAGGACCTCATTT	
Human	T-C	-500 to 0	GCGTCAGGAGCAAATGAGGT	101
			TTGATGTCTGGGATGCATGG	
Human	T-D	0 to +500	TGACCTCGGCAACAAGTCCT	101
			GCAGAGAGGGAAATGGACGG	
Human	3C-1	-7500 to -7000	GATGACCGATTCTACCGCAAA	
Human	3C-2	-7000 to -6500	GATACGCCTTTACTATGGCTCCTT	
Human	3C-3	-6500 to -6000	AGGTGGAGAACAAAAGCTGAGCTA	
Human	3C-4	-6500 to -6000	GAAGGCACAGCCTGGTTAGG	
Human	3C-5	-6500 to -6000	CCTGGTTCACGTTGCTTTTGA	
Human	3C-6	-4500 to -4000	CAACTCACCACAAAAAACCCATT	
Human	3C-7	-2000 to -1500	CGAAGAGTTTCCCGTGTGATC	
Human	3C-8	-1500 to -1000	GGAAGCTTTGTTTCTTGGTGTCA	
Human	3C-9	-1500 to -1000	TTCTGATTTAGAAGCTGAGAGCTTAGG	
Human	3C-10	-1000 to -500	GCTTCCGACGAGGAAAGAGA	
Human	3C-11	-1000 to -500	GCGACTTGATCCAATTAGTAAATGC	
Human	3C-12	-500 to 0	GGAGAGGGCTTCTGGGTTAGTT	
Human	3C-13	-500 to 0	TCCCCCATCCTGCCTTTATAAT	
Human	3C-14	0 to +500	GCCCCCTCTCCTCAGAAGTT	
Human	3C-15	0 to +500	GCGGCCATGCTTAGTGACA	
Human	3C-16	0 to +500	GGCCACTGCGCGCTACT	
Human	3C-17	+500 to +1000	CTGCTCACTCTTCCCGTACCA	
Human	3C-18	+1000 to +1500	CCACAAGCTGGATTCGTTAGATC	

Human	3C-19	+1000 to +1500	TAGATGGGCTCCCGCAGTT	
Human	3C FAM-VIC	-7500 to -7000 (Probe)	TTGGGATTTGCAACCGAGTCATGGA	
Promoter -mCherry	mCherry_F		ACTCACACTGCTGGCGGGT	125
	mCherry_R		GCGCATGAACTCCTTGATGAT	
Mouse	Cyclophilin	cDNA	GTCTCCTTCGAGCTGTTTGC	85
			AGCCAAATCCTTTCTCTCCA	
Mouse	Sox17	cDNA	CACAACGCAGAGCTAAGCAA	65
			CGCTTCTCTGCCAAGGTC	
Mouse	Eomes	cDNA	ACCGGCACCAAACTGAGA	64
			AAGCTCAAGAAAGGAAACATGC	
Mouse	GSc	cDNA	GAGACGAAGTACCCAGACGTG	93
			GCGGTTCTTAAACCAGACCTC	
Mouse	Gata6	cDNA	GGTCTCTACAGCAAGATGAATGG	94
			TGGCACAGGACAGTCCAAG	
Mouse	Foxa2	cDNA	GAGCAGCAACATCACCACAG	64
			CGTAGGCCTTGAGGTCCAT	
Mouse	Mixl1	cDNA	CCATGTACCCAGACATCCACT	88
			CGGTTCTGGAACCACACCT	
Mouse	Jmjd3	cDNA	GTACAGACCCCCGGAACC	81
			TGGTGGAGAAAAGGCCTAAG	
Mouse	Smad4	cDNA	CCGTGGGTGGAATAGCTC	63
			GGTCATCCACACCGATGC	
Mouse	Oct4	cDNA	GTTGGAGAAGGTGGAACCAA	61
			CTCCTTCTGCAGGGCTTTC	
Mouse	Tbx3	cDNA	TTGCAAAGGGTTTTCGAGAC	64
			TGCAGTGTGAGCTGCTTTCT	
Mouse	Nanog	cDNA	TTCTTGCTTACAAGGGTCTGC	67
			AGAGGAAGGGCGAGGAGA	
Human	Cyclophilin	cDNA	TGCTGTCTTTGGGACCTTGTC	91
			ACCCCACCGTGTTCTTCGAC	
Human	SOX17	cDNA	ACGCCGAGTTGAGCAAGA	82
			TCTGCCTCCTCCACGAAG	

Supplementary Table 3. Antibodies used in this study

Protein recognized	Raised in (isotype)	Company/Origin
H3K4me1	Rabbit (IgG)	Abcam
H3K4me3	Rabbit (IgG)	Millipore
H3K27me3	Rabbit (IgG)	Millipore
H3K27Ac	Rabbit (IgG)	Abcam
H2A.Ub	Mouse (IgM)	Millipore
H3	Rabbit (IgG)	Cell Signaling
Pol2 (8WG16)	Rabbit (IgG)	Abcam
CTD-Pol2-Ser5P	Rabbit (IgG)	Abcam
CTD-Pol2-Ser2P	Mouse (IgM)	Covance
Med12	Rabbit (IgG)	Bethyl Laboratories
Cdk9	Rabbit (IgG)	Santa Cruz
Ezh2	Mouse (IgG)	Millipore
Ring1	Rabbit (IgG)	Santa Cruz
Jmjd3	Rabbit (IgG)	Millipore
Set7/9	Rabbit (IgG)	Bethyl Laboratories
Smad2	Rabbit (IgG)	Cell Signaling
Smad2/3	Rabbit (IgG)	BD Biosciences
Nanog	Rabbit (IgG)	Bethyl Laboratories
Tbx3	Goat (IgG)	Santa Cruz
Oct4	Goat (IgG)	Santa Cruz
Sox2	Goat (IgG)	Santa Cruz
Eomes	Rabbit (IgG)	Abcam
Mouse IgM	Rabbit (IgG)	Millipore
Mouse IgG	Rabbit (IgG)-HRP	Jackson Laboratories
Rabbit IgG	Donkey (IgG)-HRP	Jackson Laboratories
Goat IgG	Donkey (IgG)-HRP	Jackson Laboratories

Supplementary References

Battke, F., Symons, S., and Nieselt, K. (2010). Mayday--integrative analytics for expression data. BMC Bioinformatics *11*, 121.

Conlon, F. L., Fairclough, L., Price, B. M., Casey, E. S., and Smith, J. C. (2001). Determinants of T box protein specificity. Development *128*, 3749-3758.

Deckers, M., van Dinther, M., Buijs, J., Que, I., Lowik, C., van der Pluijm, G., and ten Dijke, P. (2006). The tumor suppressor Smad4 is required for transforming growth factor beta-induced epithelial to mesenchymal transition and bone metastasis of breast cancer cells. Cancer Res *66*, 2202-2209.

Gavrilov, A., Eivazova, E., Priozhkova, I., Lipinski, M., Razin, S., and Vassetzky, Y. (2009). Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. Methods Mol Biol *567*, 171-188.

Han, J., Yuan, P., Yang, H., Zhang, J., Soh, B. S., Li, P., Lim, S. L., Cao, S., Tay, J., Orlov, Y. L.*, et al.* (2010). Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature *463*, 1096-1100.