Cell Reports, Volume *37*

Supplemental information

Modulating mesendoderm competence during human germ layer differentiation

James R. Valcourt, Roya E. Huang, Sharmistha Kundu, Divya Venkatasubramanian, Robert E. Kingston, and Sharad Ramanathan

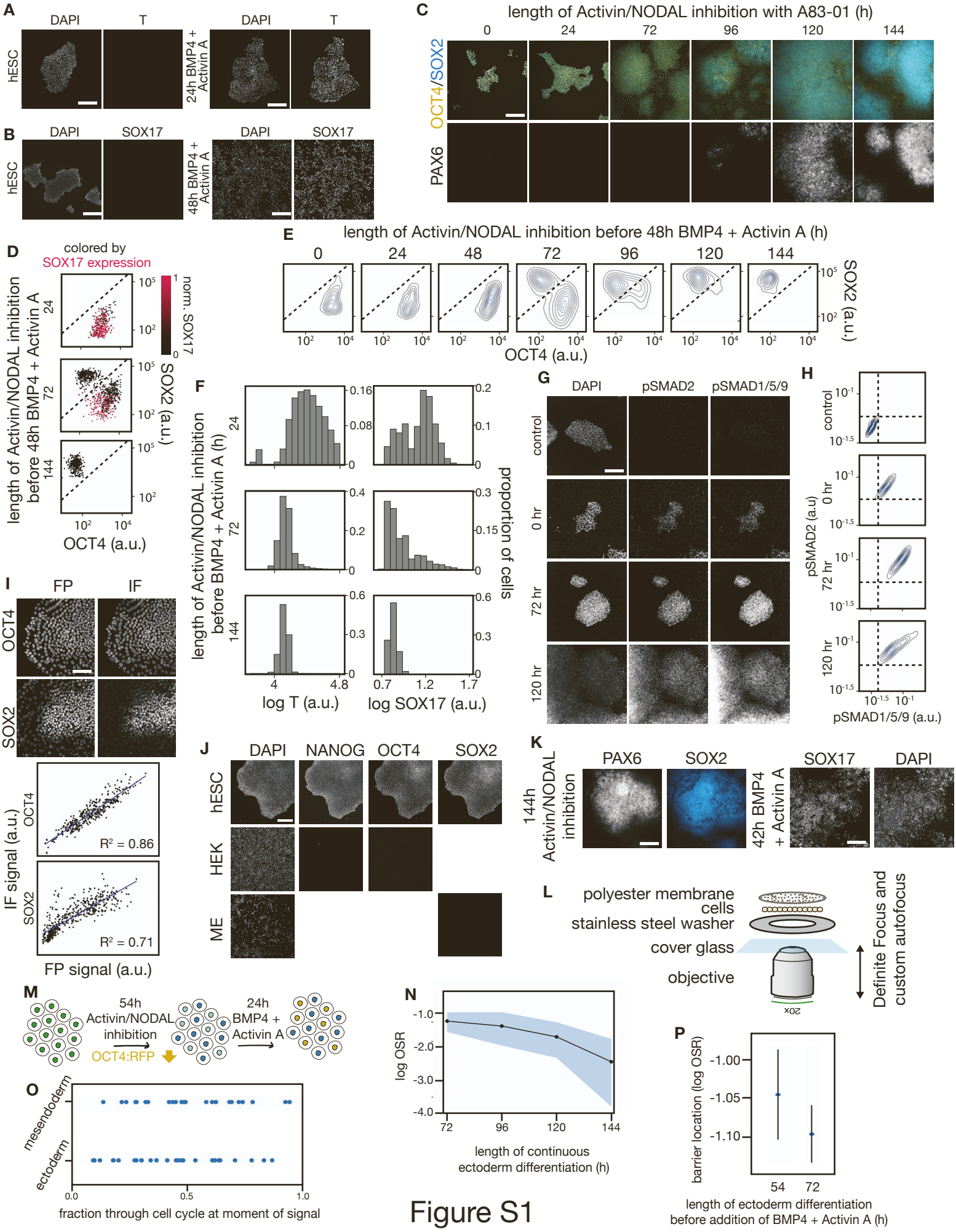

A
DAPI
T
hESC
24h BMP4 + Activin A
B
SOX17
48h BMP4 + Activin A
C
length of Activin/NODAL inhibition with A83-01 (h)
0
24
72
96
120
144
OCT4/SOX2
PAX6
D
colored by SOX17 expression
norm. SOX17
length of Activin/NODAL inhibition before 48h BMP4 + Activin A
SOX2 (a.u.)
OCT4 (a.u.)
E
length of Activin/NODAL inhibition before 48h BMP4 + Activin A (h)
48
SOX2 (a.u.)
OCT4 (a.u.)
F
length of Activin/NODAL inhibition before BMP4 + Activin A (h)
proportion of cells
log T (a.u.)
log SOX17 (a.u.)
G
DAPI
pSMAD2
pSMAD1/5/9
control
0 hr
72 hr
120 hr
H
pSMAD2 (a.u.)
pSMAD1/5/9 (a.u.)
I
FP
IF
OCT4
SOX2
IF signal (a.u.)
FP signal (a.u.)
R² = 0.86
R² = 0.71
J
DAPI
NANOG
OCT4
SOX2
hESC
HEK
ME
K
144h Activin/NODAL inhibition
PAX6
SOX2
42h BMP4 + Activin A
SOX17
DAPI
L
polyester membrane
cells
stainless steel washer
cover glass
objective
20x
Definite Focus and custom autofocus
M
54h Activin/NODAL inhibition
OCT4:RFP
24h BMP4 + Activin A
N
log OSR
length of continuous ectoderm differentiation (h)
O
mesendoderm
ectoderm
fraction through cell cycle at moment of signal
P
barrier location (log OSR)
length of ectoderm differentiation before addition of BMP4 + Activin A (h)

Figure S1

Figure S1: Differentiating stem cells produce mesendoderm-derived and ectoderm-derived cell types while maintaining the ability to receive BMP4 and Activin A signals. Related to Figures 1 and 2.

(A) Immunostained epifluorescence microscopy images of hESC colonies in pluripotency-supporting media (left) and after 24h of BMP4 + Activin A stimulation (right);. These cells largely adopt T^+ mesendodermal fates. Scale bar = 300 μm.

(B) Immunostained epifluorescence microscopy images of hESC colonies in pluripotency-supporting media (left) and after 48h of BMP4 + Activin A stimulation (right). These cells largely adopt $SOX17^+$ endoderm fates. Scale bar = 300 μm.

(C) Epifluorescence microscopy images of hESC colonies immunostained for the neurectoderm marker PAX6 (gray), OCT4 (yellow), and SOX2 (blue) after 0 to 144h of Activin/NODAL inhibition. After 120 to 144h, these cells largely adopt $PAX6^+$ $OCT4^-$ $SOX2^+$ neurectoderm fates. Scale bar = 300 μm.

(D) Scatter plots of cells stained for SOX17 from selected timepoints in Fig. 1D as function of their OCT4 and SOX2 levels. Cells were treated with Activin/NODAL inhibition for 24, 72, or 144h and then switched to BMP4 + Activin A for 48h. Cells in scatter plot colored by level of SOX17. Black dotted line separates mesendoderm- (bottom left) and ectoderm- (top left) derived cells.

(E) Density contour plots of cells as a function of their OCT4 and SOX2 levels from one sample for all timepoints in Fig. 1D. Cells were treated with Activin/NODAL inhibition for 0 to 144h and then switched to BMP4 + Activin A for 48h. Dotted line as in (D).

(F) Distribution of BRACHYURY (T) (left) and SOX17 (right) signal intensity of cells immunostained after 0, 72, or 144h of Activin/NODAL inhibition followed by BMP4 + Activin A. T was assayed after 24h of BMP4 + Activin A stimulation. SOX17 was assayed at 48h of BMP4 + Activin A because of the delayed appearance of SOX17 compared to T.

(G) Immunofluorescence images acquired via epifluorescence microscopy showing DAPI, phosphorylated SMAD1/5/9 (pSMAD1/5/9), and phosphorylated SMAD2 (pSMAD2) levels in hESC colonies upon stimulation with BMP4 + Activin for 45 mins after 0, 72, or 120h of Activin/NODAL inhibition. Unstimulated control shown in top row. All stimulated cells at all time points displayed nuclear-localized pSMAD1/5/9 and pSMAD2 upon BMP4 + Activin A exposure despite the loss of mesendoderm competence that occurs during this window. Scale bar = 300 μm.

(H) Density contour plots of cells in (G), as a function of pSMAD2 and pSMAD1/5/9 levels determined by antibody staining. Dotted lines drawn for reference.

(I) OCT4 and SOX2 levels as determined by immunofluorescence (IF) and by OCT4:RFP and SOX2:YFP fluorescence (FP) in H1 OCT4:RFP SOX2:YFP double reporter line. Left, epifluorence microscopy images. Right, Quantification of images. Scale bar = 100 μm. Blue line, best-fit linear regression.

(J) Immunofluorescence images of H1 OCT4:RFP SOX2:YFP double reporter cells in pluripotency conditions (hESC) stained for NANOG, OCT4, and SOX2. Nuclei (DAPI) are shown for reference. These cells can maintain pluripotency indefinitely. Fluorescence levels of immunostained pluripotency factors are compared to levels of NANOG and OCT4 in HEK cells

(HEK) and levels of SOX2 in mesendoderm cells (ME, H1 cells differentiated with 48h of BMP4. +Activin A). Scale bar = 300 μm.

(K) Immunofluorescence images of H1 OCT4:RFP SOX2:YFP double reporter cells that were (top) stained for PAX6 and SOX2 after 144h of Activin/NODAL inhibition by A83-01 or (bottom) stained for SOX17 after 42h of BMP4 and Activin A with nuclei (DAPI) shown for reference. This cell line's ability to form the germ layers is unaffected. Scale bar = 300 μm.

(L) Schematic of the live-cell imaging apparatus constructed for this study. Cells were grown on the bottom of a permeable membrane glued to a thin stainless-steel washer and imaged from below with an epifluorescence microscope. Focus was maintained with a combination of Zeiss Definite Focus and periodic software autofocus adjustments.

(M) Schematic of treatment time course in Fig. 2B. hESCs expressing both OCT4 (blue) and SOX2 (yellow) were treated with 54h of Activin/Nodal inhibitor A83-01, in response to which cells downregulated OCT4 at different rates. A83-01 was then removed and BMP4 + Activin A were added. In response to signal, some cells fully downregulated OCT4 to become blue, while other cells fully downregulated SOX2 to become yellow.

(N) Distribution of OSR during an ectoderm differentiation time course in which cells were exposed continuously for to up to 144h of A83-01. Quantification of pixel intensity of epifluorescence microscopy images of cells immunostained for OCT4 and SOX2. Black points = mean, shaded = std.

(O) Final cell fates from Fig. 2B as a function of each cell's position in the cell cycle at the moment of signal addition, measured as proportion of time between mitotic events during time lapse imaging. The cell's cell cycle state when signal is added does not correlate with the final fate ($p=0.66$ from chi square test).

(P) Point of mesendoderm competence loss along OSR during extended ectoderm differentiation. An independent time-lapse experiment was performed as in Fig. 2B except that ectoderm treatment was extended to 72h. $p(\text{mesendoderm}|\text{OSR})$ for this time-course was measured directly from live imaging data as in Fig. 2D. The point of competence loss, defined as the OSR value at the center of the sigmoid, $p(\text{mesendoderm}|\text{OSR}) = 0.5$, does not shift substantially with time, and the OSR distributions of the competent cells were statistically indistinguishable ($p = 0.11$).

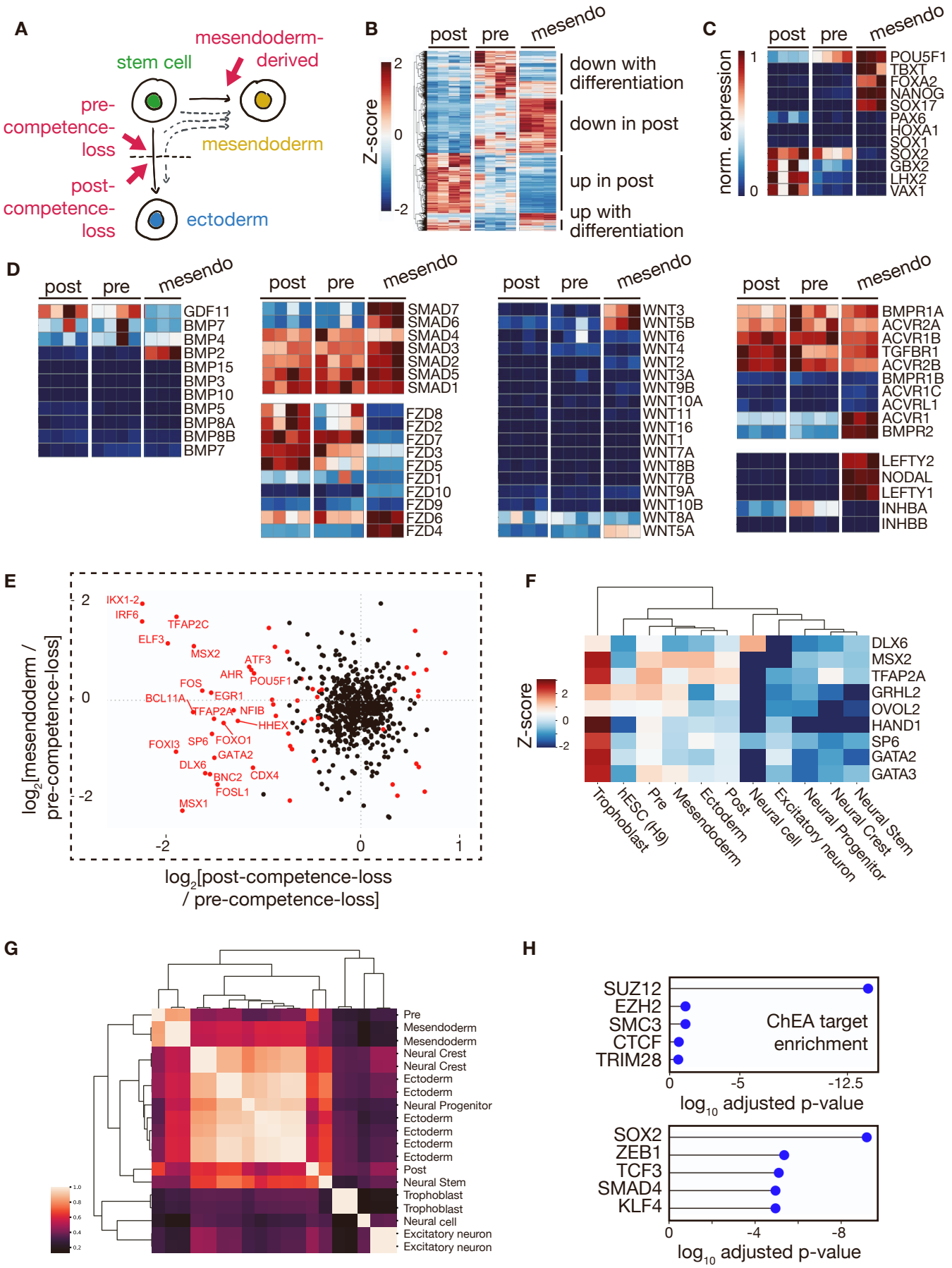

A
stem cell
mesendoderm-derived
pre-competence-loss
mesendoderm
post-competence-loss
ectoderm
B
post
pre
mesendo
Z-score
down with differentiation
down in post
up in post
up with differentiation
C
norm. expression
POU5F1
TBXT
FOXA2
NANOG
SOX17
PAX6
HOXA1
SOX1
SOX2
GBX2
LHX2
VAX1
D
GDF11
BMP7
BMP4
BMP2
BMP15
BMP3
BMP10
BMP5
BMP8A
BMP8B
BMP7
SMAD7
SMAD6
SMAD4
SMAD3
SMAD2
SMAD5
SMAD1
FZD8
FZD2
FZD7
FZD3
FZD5
FZD1
FZD10
FZD9
FZD6
FZD4
WNT3
WNT5B
WNT6
WNT4
WNT2
WNT3A
WNT9B
WNT10A
WNT11
WNT16
WNT1
WNT7A
WNT8B
WNT7B
WNT9A
WNT10B
WNT8A
WNT5A
BMPR1A
ACVR2A
ACVR1B
TGFBR1
ACVR2B
BMPR1B
ACVR1C
ACVRL1
ACVR1
BMPR2
LEFTY2
NODAL
LEFTY1
INHBA
INHBB
E
log2[mesendoderm / pre-competence-loss]
log2[post-competence-loss / pre-competence-loss]
IKX1-2
IRF6
TFAP2C
ELF3
MSX2
ATF3
AHR
POU5F1
FOS
BCL11A
EGR1
TFAP2A
NFIB
HHEX
FOXI3
SP6
FOXO1
GATA2
DLX6
BNC2
CDX4
FOSL1
MSX1
F
Z-score
DLX6
MSX2
TFAP2A
GRHL2
OVOL2
HAND1
SP6
GATA2
GATA3
Trophoblast
hESC (H9)
Pre
Mesendoderm
Ectoderm
Post
Neural cell
Excitatory neuron
Neural Progenitor
Neural Crest
Neural Stem
G
Pre
Mesendoderm
Mesendoderm
Neural Crest
Neural Crest
Ectoderm
Ectoderm
Neural Progenitor
Ectoderm
Ectoderm
Ectoderm
Ectoderm
Post
Neural Stem
Trophoblast
Trophoblast
Neural cell
Excitatory neuron
Excitatory neuron
H
SUZ12
EZH2
SMC3
CTCF
TRIM28
ChEA target enrichment
log10 adjusted p-value
SOX2
ZEB1
TCF3
SMAD4
KLF4
log10 adjusted p-value

Figure S2

Figure S2: Analysis of gene expression changes that coincide with mesendoderm competence loss. Related to Figure 4.

(A) Pre- and post-competence-loss cells were isolated based on OSR using FACS as in Fig. 3C for RNA-seq and ATAC-seq. Mesendoderm-derived cells (hESCs stimulated with 42 h of BMP4 + Activin A) were also assayed as an outgroup to allow for the identification of ectoderm lineage-specific changes.

(B) Heatmap showing normalized gene expression changes for genes that display significant differential expression between at least two of the three assayed populations in (A).

(C) Heatmap showing normalized expression of selected marker genes whose expression is associated with the early germ layer lineages. Gene expression was normalized by the maximal expression observed in each row or 10 transcripts per kilobase million (TPM), whichever was greater.

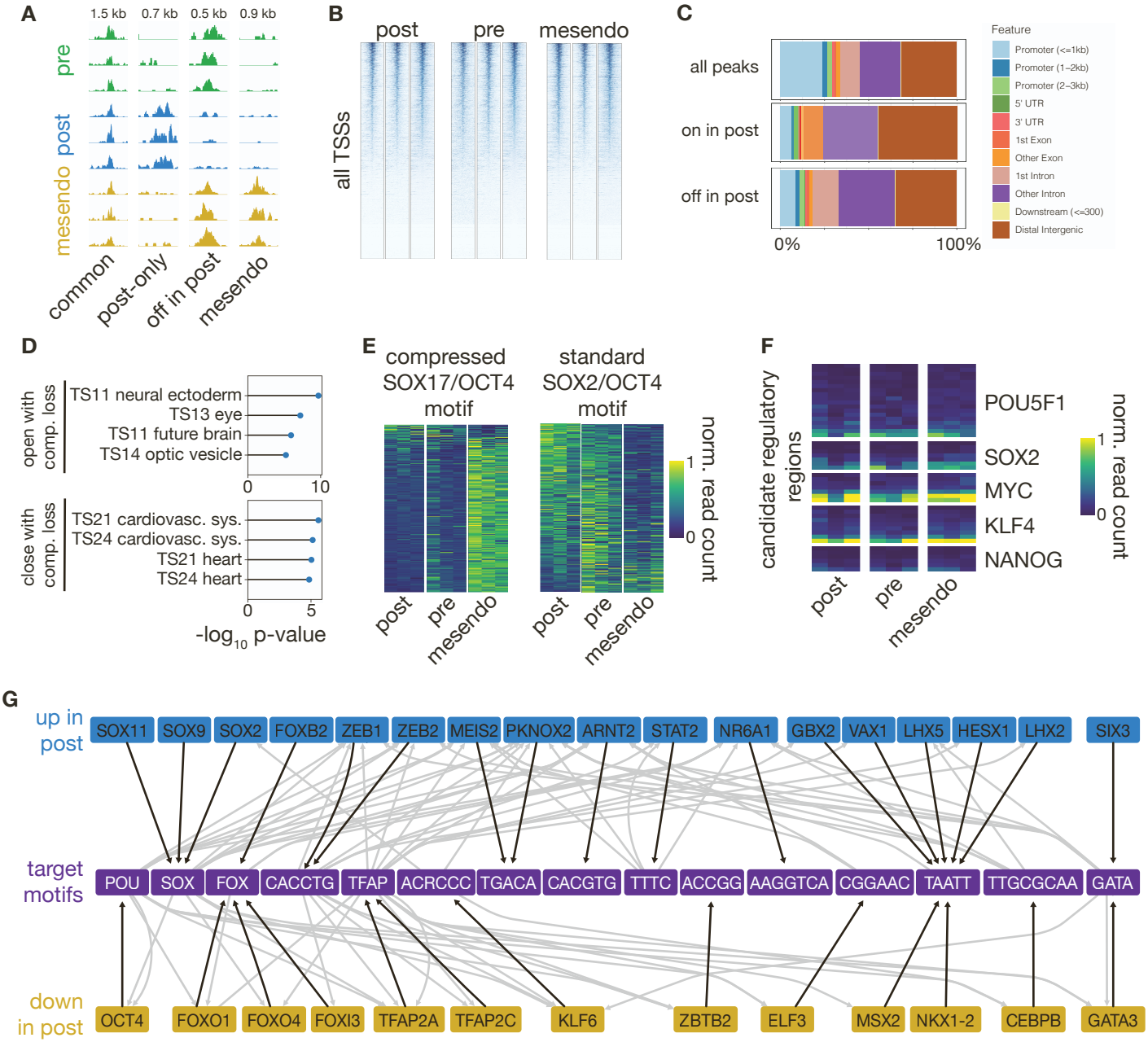
(D) Heatmap showing normalized expression of selected signaling molecules, transducers, and receptors. Gene expression was normalized by the maximal expression observed in each row or 10 TPM, whichever was greater. No significant expression changes between the pre- and post-competence-loss populations were observed in these groups.

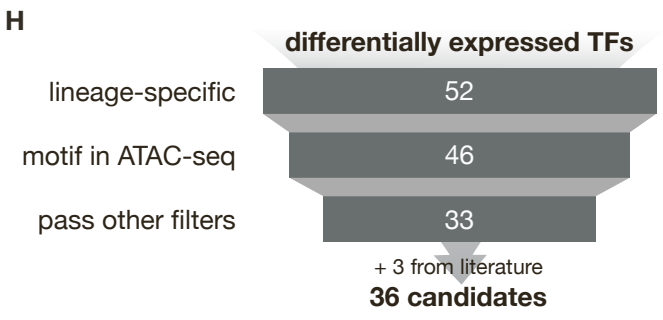
(E) Inset from Fig. 4C. Scatterplot showing log2 fold change in expression for all transcription factors between the (x-axis) post- and pre-competence-loss populations and the (y-axis) mesendoderm and pre-competence-loss populations. TFs that are significantly differentially expressed between the pre- and post-competence-loss populations are shown in red. TFs in the second and fourth quadrants of this plot have lineage-specific expression patterns.

(F) Bulk RNA-seq characterization of extra-embryonic-associated genes and comparison to ENCODE reference samples. Both pre- and post-competence-loss populations showed expression of these genes comparable to embryonic cell types. Values shown are the Z-score by gene across samples after normalization to TPM.

(G) Fate analysis by bulk RNA-seq confirms competence of FACS sorted populations. Samples shown are those from Fig. 3D with additional ENCODE reference samples. Pre- and post-competence-loss FACS-sorted populations were treated with 24h of BMP4 + Activin A and analyzed with bulk RNA-seq and compared with reference ENCODE samples representing various cell fates. The matrix shows the Pearson correlation between each pair of samples in the expression space of all human transcription factors. Hierarchical clustering reveals that the pre-competence-loss cells adopt cell fates that are most similar to mesendoderm, and the post-competence-loss cells adopt fates that are most similar to ectoderm types.

(H) Left, most significant ChIP-seq target set enrichments as reported by Enrichr for genes whose expression was upregulated from pre-competence-loss to post-competence-loss cells but not significantly upregulated from the pre-competence-loss to mesendoderm-derived populations. Right, same for genes that were downregulated from pre- to post-competence-loss but not downregulated from pre-competence loss to mesendoderm populations.





I **full list of candidates**

AHR	GRHL1	NFIB	SOX2DBD
ARNT2	GSC	NRF2	SOX9
ATF3	HESX1	OCT4	TFAP2A
FEZF1	ID2	OCT4DBD	TFAP2C
FOXB2	JUNB	OTX2	VAX1
FOXJ2	KLF6	PAX3	ZSCAN23
FOXO1	LHX2	PKNOX2	
FOXO4	LHX5	POU2F3	
GATA3	MBNL2	POU6F1	
GBX2	NANOG	SOX2	

Figure S3

Figure S3: Analysis of chromatin accessibility changes that coincide with mesendoderm competence loss. Related to Figure 4.

(A) Normalized ATAC-seq read depth at example loci with the indicated accessibility patterns across pre-competence-loss, post-competence-loss, and mesendoderm-derived populations. Text at top indicates the length of the region shown.

(B) ATAC-seq read depth at all transcription start sites. Each heatmap block represents all transcription start sites (TSSs) in a single biological replicate for the indicated condition. Each row shows a heatmap representation of read depth in 2kb windows centered on each TSS. Rows are sorted for each replicate individually by mean read depth across the window.

(C) Genomic feature type of the indicated ATAC-seq peaks as annotated by ChIPSeeker. Annotations are shown for all detected peaks across all samples, for peaks that are significantly more accessible in post-competence-loss cells, and for peaks that are significantly less accessible in post-competence-loss cells. Most differentially accessible sites are located in distal intergenic or intronic regions.

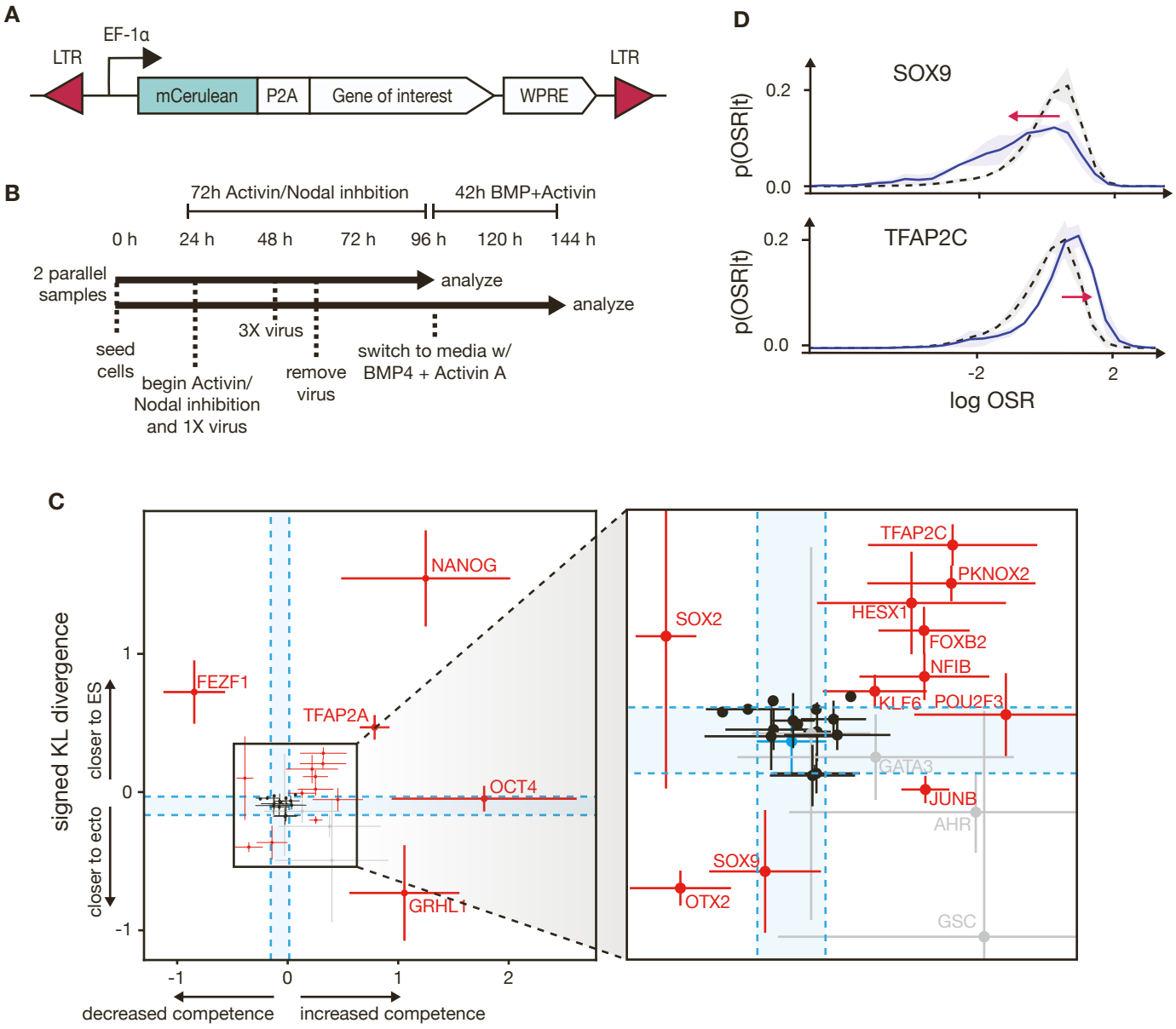
(D) Overlap enrichment of mouse homologs of the differentially expressed gene sets from this study with developmental stage gene expression sets from the Mouse Genome Informatics database as calculated by the software GREAT. Top, the four most enriched gene expression set matches for genes that are upregulated in a lineage-specific manner upon competence loss. Bottom, the same for genes that are downregulated.

(E) Heatmap showing row-normalized mean ATAC-seq depth in 500 bp regions centered on each peak that shows significant changes in accessibility and which contains the indicated motif. Left, the regions that contain a compressed SOX17/OCT4 motif associated with endoderm occupancy. Right, the same for regions that contain a standard SOX2/OCT4 motif associated with pluripotency occupancy.

(F) Heatmap showing normalized ATAC-seq read depth in all ENCODE-annotated candidate regulatory regions for the indicated pluripotency-associated genes. None of these regions show significant differences between pre- and post-competence-loss cells

(G) Candidate gene regulatory network that governs mesendoderm competence. TFs that are upregulated (top row, blue) or downregulated (bottom row, yellow) specifically between pre- and post-competence-loss populations are connected to their DNA binding motifs by black arrows. Each motif is connected to possible regulatory targets by gray arrows according to the following criteria (1) there is a significant match to that motif in an ATAC-seq peak that is (2) predicted to be associated with the target gene via the CisMapper software.
(H) Schematic depicting the candidate selection process (see Methods).

(I) Full list of the 36 TFs identified as candidates for perturbation to affect mesendoderm competence. OCT4DBD = OCT4 DNA binding domain; SOX2DBD = SOX2 DNA binding domain.



shift in location of p(mesendoderm|OSR) = 0.5 from non-transduced to transduced populations (log_2 OSR)

Figure S4

Figure S4: Analysis of lineage bias and differentiation dynamics in cells perturbed by candidate TFs. Related to Figure 5.

(A) Schematic of the cassette transduced via lentiviral vector. Under the control of the EF1alpha promoter, we placed the cyan fluorescent protein mCerulean fused in frame to a P2A ribosomal skip sequence and the gene of interest. A Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE) was included to improve transcript stability. The entire sequence was flanked by the Long Terminal Repeat regions (LTR) that the viral transposases recognize to effect integration into the genome.

(B) Schematic of flow cytometry protocol. 24h after seeding in pluripotency maintenance conditions, parallel samples of hESCs were subjected to simultaneous Activin/Nodal inhibition and, for the first 36h, lentiviral transduction of the gene of interest. After 72h of Activin/Nodal inhibition, one membrane was analyzed by flow cytometry while the other was stimulated with 42h of BMP4 + Activin A before analysis by flow cytometry.

(C) Scatterplot of the shift in p(mesendoderm|OSR) location versus a metric of disruption of p(OSR|t) for each TF tested. Disruption of initial differentiation dynamics was measured as the Kullback-Leibler divergence of the distribution of the log OSR of CFP^+ cells using the same distribution from CFP^- cells as a reference; this KL-divergence was then reported with a positive sign if the mean OSR of the transduced cells was higher than in the non-transduced cells and a negative sign if not. Candidates that were significantly different from the CFP negative control (FDR < 0.1) along either dimension are shown in red. Candidates whose effects were not consistent across replicates are shown in gray. CFP:P2A:CFP is shown in blue. Blue dotted lines and shaded region indicate CFP:P2A:CFP overexpression outcomes, shown as the mean ± one standard deviation of n=4 biological replicates, for each axis. OCT1, POU6F1, PAX3 were tested but are excluded from this plot because their overexpression caused premature downregulation of both OCT4:RFP and SOX2:YFP, indicating possible differentiation to an alternative fate. Interestingly, exogenous OCT4 overexpression does not impact the dynamics of endogenous OCT4:RFP and SOX2:YFP during ectodermal differentiation, as evidenced by the location of the OCT4 point near zero on the y-axis.

(D) Examples of two candidates, SOX9 and TFAP2C, whose overexpression altered the OSR distribution of cells before the signal, p(OSR | t). OCT4:RFP and SOX2:YFP levels were normalized to an hESC sample measured in the same batch. Black dotted line, wildtype. Blue, transduced. Shaded area represents one standard deviation across biological replicates.

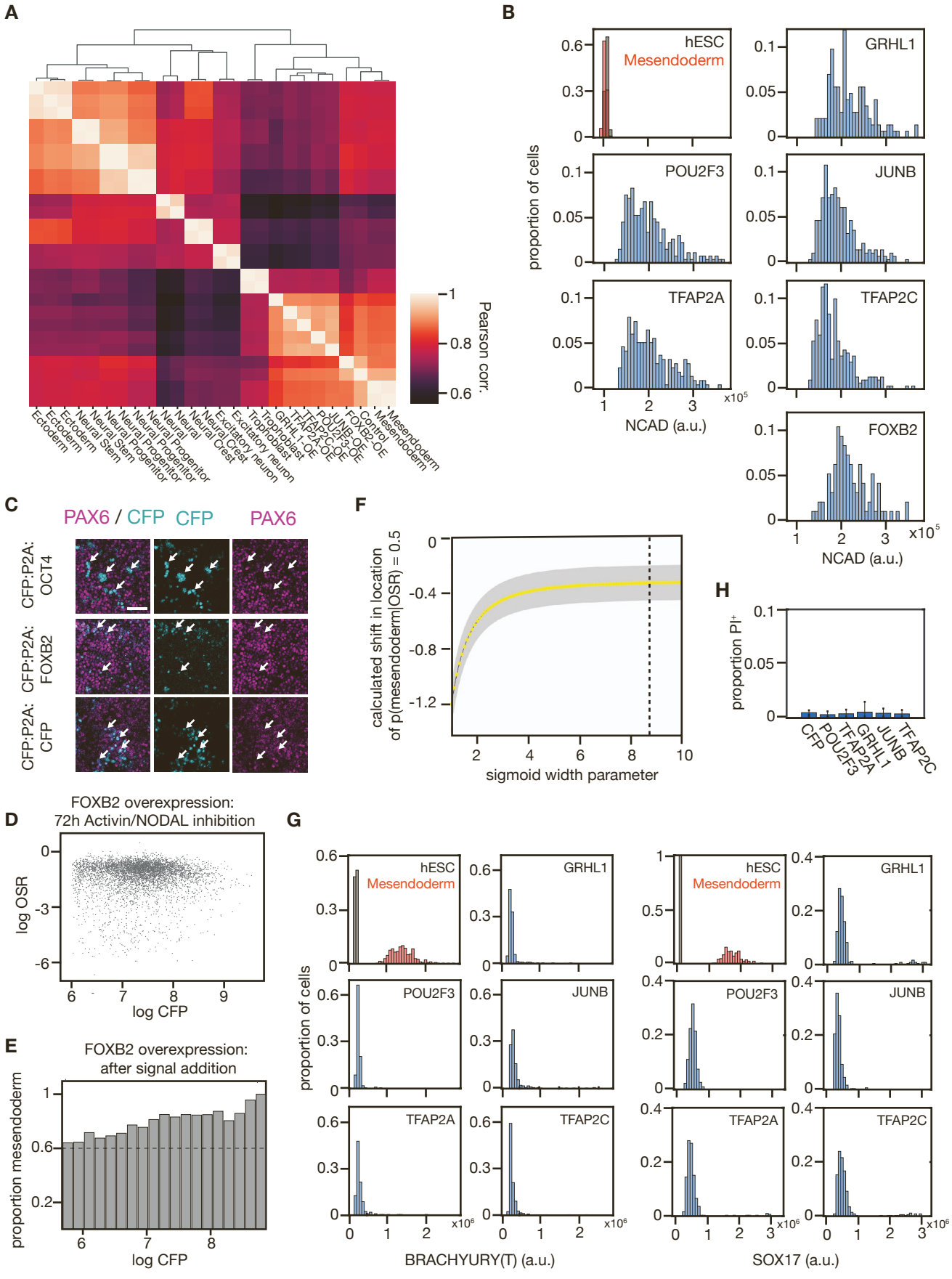

A
Pearson corr.
B
hESC
Mesendoderm
GRHL1
POU2F3
JUNB
TFAP2A
TFAP2C
FOXB2
proportion of cells
NCAD (a.u.)
C
PAX6 / CFP
CFP
PAX6
CFP:P2A: OCT4
CFP:P2A: FOXB2
CFP:P2A: CFP
D
FOXB2 overexpression:
72h Activin/NODAL inhibition
log OSR
log CFP
E
FOXB2 overexpression:
after signal addition
proportion mesendoderm
log CFP
F
calculated shift in location of p(mesendoderm|OSR) = 0.5
sigmoid width parameter
G
BRACHYURY(T) (a.u.)
SOX17 (a.u.)
H
proportion PI+
CFP
POU2F3
TFAP2A
GRHL1
JUNB
TFAP2C

Figure S5

Figure S5: FOXB2 overexpression analysis. Related to Figure 5.

(A) Fate analysis by bulk RNA-seq confirms mesendoderm identity of cells overexpressing key candidate TFs after BMP and Activin A signal. Shown are six samples overexpressing candidate TFs after BMP and Activin A signal, one unperturbed control after BMP and Activin A signal, and 19 reference ENCODE samples representing various cell fates. The matrix shows the Pearson correlation between each pair of samples in the expression space of all human transcription factors. Hierarchical clustering reveals that the TF-overexpressing cells are most similar to the two mesendoderm ENCODE samples and to the unperturbed control sample from this study. Overexpression of these TFs does not impair the cell's ability to adopt mesendodermal fates.

(B) Post-competent cells that were perturbed with genetic candidates were exposed to 48 hours of BMP4 + Activin A, immunostained for N-cadherin, which is highly expressed in neuroectoderm (Pijuan-Sala, et al., 2019) and imaged with epifluorescence microscopy. Quantification of N-cadherin immunofluorescence per cell are compared to those of hESC cells (black) and mesendodermal cells (red, derived from hESCs exposed to 48 hours of BMP4 + Activin A).

(C) Immunofluorescence images showing PAX6 (magenta) and CFP (cyan) after 144h of neurectoderm-directed differentiation using inhibitors of both Activin/NODAL and BMP. Top row, exogenous expression of CFP:P2A:OCT4; Middle row, CFP:P2A:FOXB2; Bottom row, CFP:P2A:CFP. White arrows indicate select CFP-expressing cells. Look-up tables for the channels in each condition are scaled individually because variable membrane positioning in the well created spurious changes in overall apparent signal intensity. Scale bar = 50 μm. OCT4 overexpression prevents PAX6 induction, while FOXB2 or CFP overexpression does not disrupt normal PAX6 induction.

(D) Scatterplot of log OSR versus CFP:P2A:FOXB2 expression level in CFP^+ cells (see Fig. 5D, bottom) after 72h of Activin/Nodal inhibition. FOXB2 expression level does not influence distribution of OSR during ectoderm differentiation.

(E) Effect of CFP:P2A:FOXB2 expression level on final fate distribution. CFP^+ cells (see Fig. 5E, bottom) after 42h of signal addition were binned according to CFP expression level. Shown is proportion of mesendoderm fate for each bin, as measured by OSR, with proportion of CFP^- control population (black dotted line) shown as reference. As expected, cells with low CFP expression level match the fate distribution of non-transduced controls, while cells with higher levels of genetic perturbation show stronger population phenotypes.

(F) Insensitivity of the shift in the calculated location of the point at which p(mesendoderm|OSR) = 0.5 relative to the non-transduced internal control cells (y-axis) to the value of the sigmoid width parameter (x-axis) for CFP:P2A:FOXB2 overexpression. Gray region represents one standard deviation. The calculated shift is statistically significant at all points shown, and the effect of the overexpression is broadly insensitive to the sigmoid width parameter used. Parameter value used when fitting p(mesendoderm|OSR) in Figure 5 was 8.38.

(G) Post-competent cells that were perturbed with genetic candidates were exposed to 48 hours of BMP4 + Activin A, reseeded on glass, immunostained for SOX17 and BRACHYURY(T), and imaged with epifluorescence microscopy. Quantification of BRACHYURY(T) (left two columns) and SOX17 (right two columns) immunofluorescence per nucleus are compared to

those of hESC cells (black) and mesodermal cells (hESCs exposed to 48 hours of BMP4 + Activin A, red).

(H) Proportion of cells that showed positive propidium iodide (PI) stain upon overexpression of each candidate. Candidates did not show significant cell death above CFP control.