Changing the Waddington landscape to control mesendoderm competence

James R. Valcourt[1,2,3,†], Roya Huang[2,3], Sharmistha Kundu[4,5,6], Divya Venkatasubramanian[2,3], Robert E. Kingston[4,5], Sharad Ramanathan[2,3,7,8,†]

[1] Systems, Synthetic, and Quantitative Biology Program, Harvard University, Cambridge, MA 02138

[2] Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

[3] Division of Applied Physics, Harvard University, Cambridge, MA 02138

[4] Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA

[5] Department of Genetics, Harvard Medical School, Boston, MA, USA

[6] Present address: Intellia Therapeutics, Cambridge, MA 02139

[7] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

[8] Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

[†] To whom correspondence should be addressed. Email: valcourt@g.harvard.edu, sharad@cgr.harvard.edu

Abstract

As pluripotent human embryonic stem cells progress towards one germ layer fate, they lose the ability to adopt alternative fates. It is unknown how the cells' competence for these alternative fates changes along their developmental trajectory or if this competence can be modulated. Here, we show that a differentiating stem cell's probability of adopting a mesendodermal fate when given the appropriate signal falls sharply at a specific point along the ectodermal trajectory, and we further demonstrate that this point can be moved using genetic perturbations. Using a low-dimensional reaction coordinate to monitor progression towards ectoderm, we can determine the probability that individual cells at different points along this path can transition to the

mesendodermal fate upon BMP4 and Activin A signal exposure. Knowing this probability allows us to prospectively isolate and profile differentiating cells based on their mesendoderm competence. Analysis and validation of these RNA-seq and ATAC-seq profiles identified transcription factors that can independently control the cell's mesendoderm competence and its progression along the ectodermal developmental trajectory. In the classical picture of a Waddington landscape, these effects correspond to altering the barrier between fates and changing the cell's location on the landscape, respectively. The ability of the underlying gene regulatory network to modulate these two aspects of the developmental landscape could allow separate control of the dynamics of differentiation and tissue size proportions.

Main

Pluripotent cells have the ability to produce any of the myriad cell types seen in the adult body[1], but they lose this potential as they differentiate. During initial lineage specification, cells can change their fate choice upon exposure to signals that induce an alternative selection[1–4], such as by transplantation to a different location in the embryo. In time, however, the cell's fate becomes determined, and it is no longer competent to choose a different lineage in response to the same external signals[5–7]. While lineage specification is relatively well-understood[8–17], whether and how competence for adopting alternative lineages can be tuned during differentiation is not. Determining how this competence is set and modulated is essential for understanding developmental patterning and plasticity.

There is a fundamental challenge in understanding the competence of a cell to choose a specific fate in response to a signal because a cell's fate choice is evident only after the expression of fate-

49 specific marker genes. In mouse and humans, the fate markers for the germ layers are not expressed
50 until at least twelve hours after exposure to the appropriate signals[20–22]. Since the gene expression
51 and epigenetic state of the cell changes substantially during this period, determining how the
52 molecular state of the cell at the time of signal exposure governs its competence to adopt alternative
53 fates has been difficult.

54 To overcome this challenge, we use a framework motivated by the Waddington landscape[23]. In
55 this picture, the pluripotent cell moves down into the valley corresponding to its chosen fate, and
56 it is prevented from adopting the alternative fate by a barrier that rises between the valleys (Figure
57 1a). Thus, a cell's ability to transition to the alternative fate depends both on its location along the
58 developmental trajectory and the size of the barrier at that point. Using the germ layer selection of
59 human embryonic stem cells (hESCs) as a model system, we measured the probability that a given
60 cell can transition to the mesendoderm fate as a function of its position along the ectodermal
61 developmental trajectory, thereby probing the location of the barrier between fates. Using this
62 probability distribution, we could prospectively isolate and characterize cells based on their
63 mesendoderm competence. Computational analysis of both the gene expression and epigenetic
64 profiles of these populations allowed us to identify candidate genes that we predicted to control
65 mesendoderm competence. By perturbing the levels of these genes, we were able to independently
66 tune the barrier between fates and the cell's position on the Waddington landscape. These dual
67 mechanisms could play a role in regulating the proportions of the different germ layers in the
68 developing embryo.

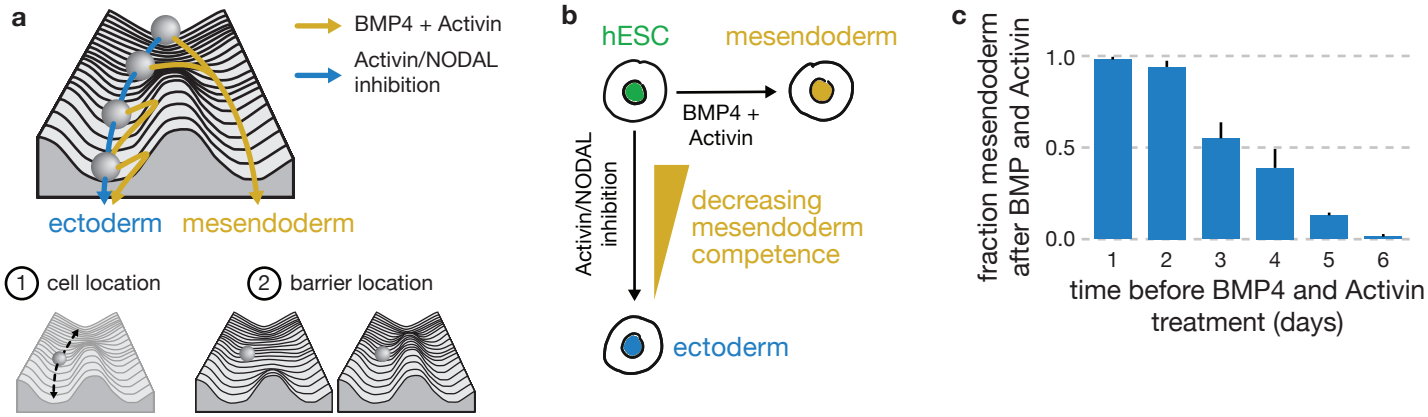

a
BMP4 + Activin
Activin/NODAL inhibition
ectoderm
mesendoderm
1 cell location
2 barrier location
b
hESC
mesendoderm
BMP4 + Activin
Activin/NODAL inhibition
decreasing mesendoderm competence
ectoderm
c
fraction mesendoderm after BMP and Activin
1.0
0.5
0.0
1
2
3
4
5
6
time before BMP4 and Activin treatment (days)

Figure 1

Figure 1: Stem cells lose competence to adopt mesendodermal fates upon BMP4 and Activin A signal exposure with increasing duration of Activin/NODAL inhibition.

a, Schematic of a Waddington landscape illustrating the ectoderm (blue) / mesendoderm (yellow) fate choice. In this picture, the competence of a cell to produce mesendoderm depends both on the cell's location along the ectodermal developmental trajectory and on the position of the barrier between the ectoderm and mesendoderm fates. **b,** hESCs choose the ectodermal lineage in response to Activin/NODAL inhibition and mesendodermal lineages in response to BMP4 + Activin A. **c,** Bar graph of the fraction of cells adopting mesendodermal fate as a function of the duration of ectodermal differentiation after which BMP4/Activin A was added (see Methods). At the population level, with an increasing duration of prior ectodermal differentiation, the competence to choose mesendodermal fates in response to signal decreases. Error bars: std (n=3).

69 ## Results

70 **hESCs lose mesendoderm competence along the ectoderm developmental trajectory**

71 Given the appropriate signals, hESCs *in vitro* can adopt either mesendodermal or ectodermal
72 progenitor fates: BMP4 and Activin/NODAL signals induce mesendodermal fate[24], while
73 Activin/NODAL inhibition promotes ectodermal fate[21] (Figure 1b). When cultured in conditions
74 that allow the signaling ligands to access their corresponding receptors on the cells' basolateral
75 side (see Methods), hESCs exposed to BMP4 and Activin A uniformly produce cells expressing
76 BRACHYURY, SOX17, and other markers associated with mesendoderm-derived cell types
77 (Figure 1b, S1a, S1b). In contrast, inhibiting Activin/NODAL signaling promoted ectoderm-
78 derived fates[21], ultimately producing PAX6^{+} neurectoderm after 4-5 days (Figure 1b, S1c, S1d).

79 Consistent with previous data from mouse[22,25], we demonstrated that competence to adopt
80 mesendoderm-derived fates decreases at the population level as cells differentiate towards the
81 ectodermal fate. Increasing the duration of differentiation towards ectoderm reduced the fraction
82 of cells that adopted mesendodermal fates in response to BMP4 and Activin A signal (Figure 1c,
83 S1b, S1e, S1f). The temporal decrease in the mesendoderm fraction occurred despite the cells'
84 continued ability to respond to BMP and Activin signals throughout this period (Figure S1g, S1h).
85 These findings suggested that the probability of cells transitioning to the mesendodermal fate in
86 response to signal decreases over time within the differentiating population.

87 **Dynamics of OCT4 and SOX2 predict the mesendoderm competence of individual cells**

88 We next sought to directly measure and predict the probability of individual cells adopting a
89 mesendodermal fate in response to BMP and Activin signals along the ectodermal differentiation

90 trajectory. To do so, we first developed a measure of each live cell's position along the
91 developmental trajectory by choosing a low-dimensional coordinate—in this case, the expression
92 levels of key genes—whose dynamics accurately report on the choice of the two germ layer
93 lineages. Our recent computational work allows us to identify these key genes for a given lineage
94 decision from single cell gene expression data[26,27], and we have demonstrated the accuracy of this
95 method for germ layer, neural, and hematopoietic development. For the decision between the two
96 germ layer lineages, the genes that allow us to continuously monitor the progression of fate choice
97 are the transcription factors OCT4 (POU5F1) and SOX2. Our previous work in mouse showed
98 that the protein levels of Oct4 and Sox2 reflect the transitions of pluripotent cells to the
99 mesendoderm or ectoderm fates[26–28]. We validated that, in humans as in mouse, both OCT4 and
100 SOX2 are symmetrically highly expressed in the pluripotent stem cell, but they are asymmetrically
101 downregulated in the two lineages. OCT4 expression is maintained in the mesendoderm while
102 SOX2 is downregulated; in contrast, ectoderm differentiation involves SOX2 maintenance and
103 OCT4 downregulation. Both TFs are also functionally important for these state transitions: OCT4
104 downregulation is necessary for neurectoderm induction[28,29], while SOX2 downregulation is
105 required for mesendoderm fate selection[28]. Furthermore, direct conversion to a neural fate silences
106 OCT4[28], underscoring the fundamental nature of this reaction coordinate to the fate decision in
107 question.

108 To monitor developmental trajectories in real time, we employed our validated hESC line in which
109 one allele each of OCT4 and SOX2 had been replaced with OCT4:RFP and SOX2:YFP,
110 respectively, at the endogenous locus[30] (Figure S2a-c). Using flow cytometry, we followed the
111 developmental trajectories of hESCs to ectodermal progenitors as they downregulated OCT4 over

112 the course of six days. When BMP4 and Activin A signals were added at an intermediate stage of
113 differentiation (Figure 2a), we could visualize a bifurcation of developmental trajectories: the
114 mesendoderm-competent cells adopted $OCT4^{+}$ $SOX2^{-}$ mesendoderm-derived fates, while the cells
115 that were not mesendoderm-competent proceeded towards $OCT4^{-}$ $SOX2^{+}$ ectoderm-derived fates.

116 We next sought to measure the probability of an individual cell adopting a mesendoderm-derived
117 fate given the cell's location along the ectodermal trajectory. To do so, we performed a time lapse
118 experiment using the OCT4:RFP SOX2:YFP hESC line (Figure 2b). To monitor the cells
119 throughout this process, we developed a custom live-cell microscopy setup that was capable of
120 imaging cells on the flexible membrane every 15 minutes for over five days (Figure S2d). Based
121 on the timing of mesendoderm competence loss in our flow cytometry experiments, we first
122 differentiated the pluripotent stem cell population in this apparatus for 2.25 days in ectodermal
123 differentiation conditions to obtain a heterogeneous population in which some cells had already
124 lost mesendoderm competence and some had not. We then added BMP4 and Activin A signals for
125 25 hours, prompting mesendoderm-competent cells to adopt mesendoderm fates and non-
126 mesendoderm-competent cells to adopt ectodermal fates.

127 Using our time lapse data, we demonstrated that each cell's location along its developmental
128 trajectory as measured by OCT4:RFP and SOX2:YFP levels was predictive of the cell's
129 mesendoderm competence, and that this measure allowed such prediction days before the
130 expression of classical master regulators. We tracked individual cells from pluripotency through
131 ectoderm-directed differentiation and subsequent BMP4 + Activin A signal (Figure 2b).
132 Pluripotent cells were tightly localized in OCT4:RFP/SOX2:YFP space but downregulated

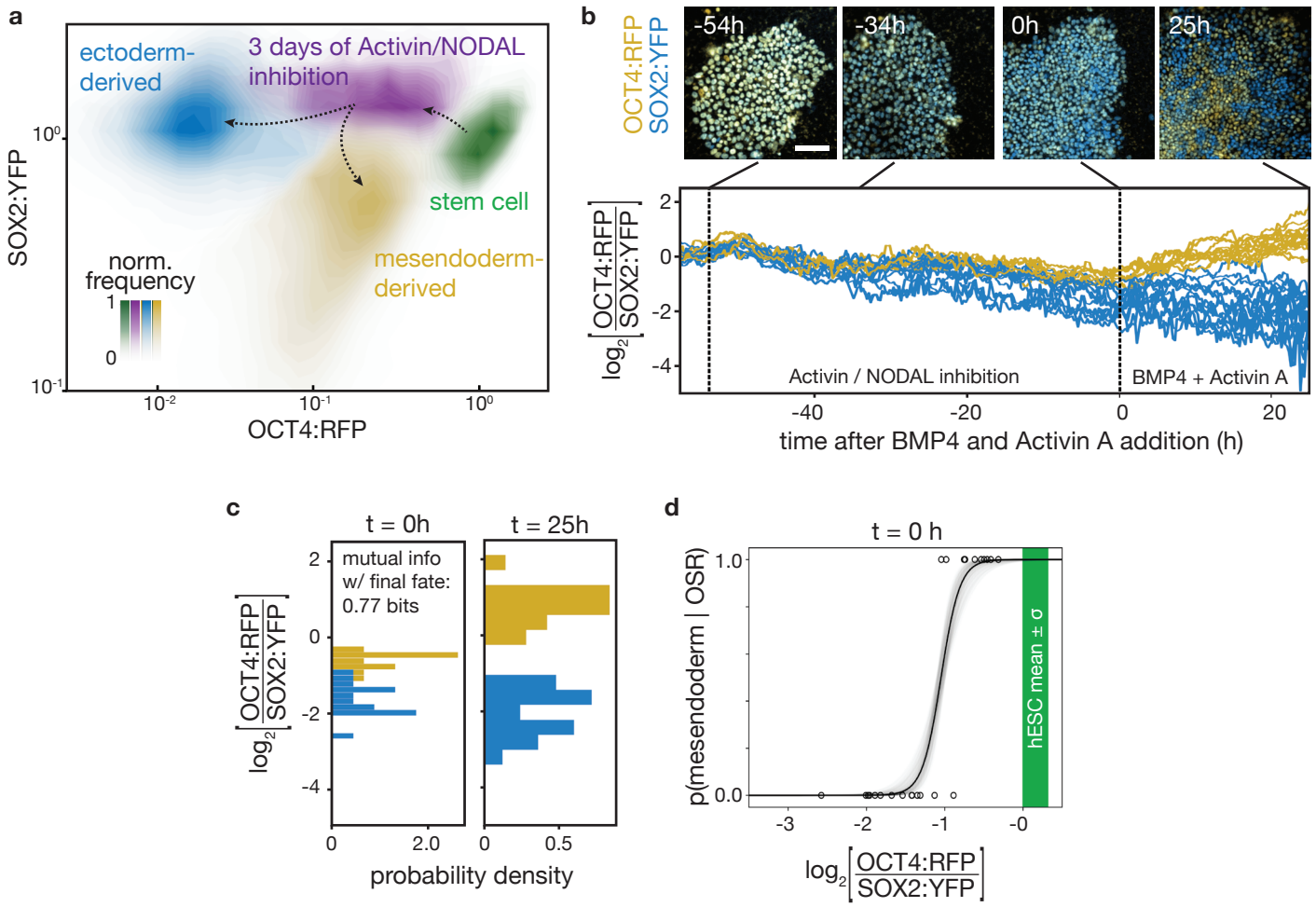

a
ectoderm-derived
3 days of Activin/NODAL inhibition
stem cell
mesendoderm-derived
norm. frequency
1
0
SOX2:YFP
OCT4:RFP
b
OCT4:RFP
SOX2:YFP
-54h
-34h
0h
25h
Activin / NODAL inhibition
BMP4 + Activin A
time after BMP4 and Activin A addition (h)
c
t = 0h
t = 25h
mutual info w/ final fate: 0.77 bits
probability density
d
t = 0 h
p(mesendoderm | OSR)
hESC mean ± σ

Figure 2

Figure 2: Dynamics of OCT4 and SOX2 accurately predict mesendoderm competence.

a, Contour plots of flow cytometry data showing levels of OCT4:RFP and SOX2:YFP normalized to the mean level in the hESC population. hESCs (green) downregulate OCT4:RFP after 3 days of ectodermal differentiation (purple). After a subsequent two days of BMP4 + Activin A stimulation, the cells in this purple population bifurcate: the competent fraction chooses a mesendodermal fate (yellow), while the remaining continue to become ectoderm (blue). **b,** Top, snapshots from a time-lapse microscopy experiment of a field of hESCs showing endogenous OCT4:RFP (yellow) and SOX2:YFP (blue). Cells started in pluripotency conditions; at t=-54h, ectodermal differentiation began; at t=0h, BMP4 + Activin A stimulation began; at t=25h, experiment ended. Scale bar = 100 μm. Bottom, plot of the log ratio of the OCT4:RFP to SOX2:YFP signal (OSR) in individual cells through the time course above. Time traces of cells adopting a mesendodermal (ectodermal) fate at the end of the time course are shown in yellow (blue). **c,** Histogram of the log OSR from (b) shown at the time of signal addition (t=0h) and at the end of the experiment (t=25h). Histograms corresponding to cells adopting an eventual mesendodermal (ectodermal) fate at the end of the time course are in yellow (blue). The mutual information between the OSR at the moment of signal addition and the final fate choice is 0.77 bits. **d,** Plot of the probability of a single cell adopting a mesendoderm-derived fate given OSR, p(mesendoderm | OSR), calculated from time course in (b), mean black, 1 std in grey (see Methods). Green bar: mean value +/- std of the ratio in pluripotent stem cells.

133 OCT4:RFP at different rates upon ectoderm differentiation. Importantly, we found that, rather than
134 the levels of the individual proteins, the OCT4:RFP to SOX2:YFP fluorescence ratio at the
135 moment of BMP4 and Activin A signal addition was predictive of the ultimate fate of the cells
136 with high accuracy (Figure 2c). Each cell carried 0.77 bits of mutual information (the maximum
137 possible value being 1) about its mesendoderm competence state in its ratio of OCT4:RFP to
138 SOX2:YFP at the moment of signal addition. Cells with a high ratio of OCT4:RFP to SOX2:YFP
139 were able to become mesendoderm in response to the BMP4 and Activin A signal, while cells with
140 a low ratio of OCT4:RFP to SOX2:YFP were not.

141 We then computed the probability of a cell adopting a mesendodermal fate given its OCT4:RFP
142 to SOX2:YFP ratio (OSR) after normalizing OCT4:RFP and SOX2:YFP intensities to the mean
143 levels measured in hESCs in pluripotency conditions (Figure 2d). The probability of adopting a
144 mesendodermal fate given OSR, $p(\mathrm{mesendoderm}|\mathrm{OSR})$, had a sharp transition from 1 to 0,
145 indicating that there was a defined point along the developmental trajectory at which cells lose
146 their ability to become mesendoderm even when exposed to the relevant signals. On the
147 Waddington landscape, the OSR corresponds to a cell's position along the ectoderm trajectory,
148 and the point at which $p(\mathrm{mesendoderm}|\mathrm{OSR})$ transitions from 1 to 0 corresponds to the location
149 of the barrier that arises between the two fates.

150 **Prospective isolation of cell populations based on mesendoderm competence**

151 Having computed $p(\mathrm{mesendoderm}|\mathrm{OSR})$ for single cells in our time lapse, we sought to predict
152 the mesendoderm competence of cells in a heterogeneous differentiating population. Since cells in
153 a population move along the developmental trajectory at differing rates, at any given time, t, the

154 cells have a distribution of OSRs, $p(\mathrm{OSR}|t)$. In the Waddington landscape, this corresponds to
155 the distribution of cell locations along the ectoderm trajectory at that given time. The fraction of
156 the overall cell population that adopts a mesendoderm fate after BMP4 and Activin stimulation
157 should be determined by the fraction of the cells that have not yet reached the point along the
158 ectodermal trajectory where the barrier to mesendoderm arises. Thus, the total fraction of cells that
159 become mesendoderm can be calculated by multiplying the fraction of cells with a given OSR by
160 the probability that cells with that ratio will become mesendoderm, then summing over all OSRs
161 (Figure 3a):

162 Equation 1: $$p(\mathrm{mesendoderm}|\mathrm{t}) = \sum_{\mathrm{OSR}} p(\mathrm{mesendoderm}|\mathrm{OSR}) \times p(\mathrm{OSR}|\mathrm{t})$$

163 We next reasoned that we should be able to use our ability to predict mesendoderm competence to
164 prospectively isolate competent from non-competent cells from a single differentiating population.
165 To this end, we measured $p(\mathrm{OSR}|t)$ of a population of differentiating hESCs using flow
166 cytometry. Our analysis suggested that the cells in the high OSR region where our computed
167 $p(\mathrm{mesendoderm}|\mathrm{OSR}) \approx 1$ would be competent to differentiate into mesendoderm, while those
168 in the low OSR region where $p(\mathrm{mesendoderm}|\mathrm{OSR}) \approx 0$ would not (Figure 3b). We will
169 hereafter refer to these populations as "pre-competence-loss" and "post-competence-loss,"
170 respectively. To validate our predictions, we sorted cells based on their OSR using fluorescence-
171 activated cell sorting (FACS) from a population that had been subjected to 3 days of ectodermal-
172 directed differentiation (Figure 3c). We then added BMP4 and Activin A to the sorted populations
173 to compare our predicted mesendodermal competence with the observed fate choice of these cells.

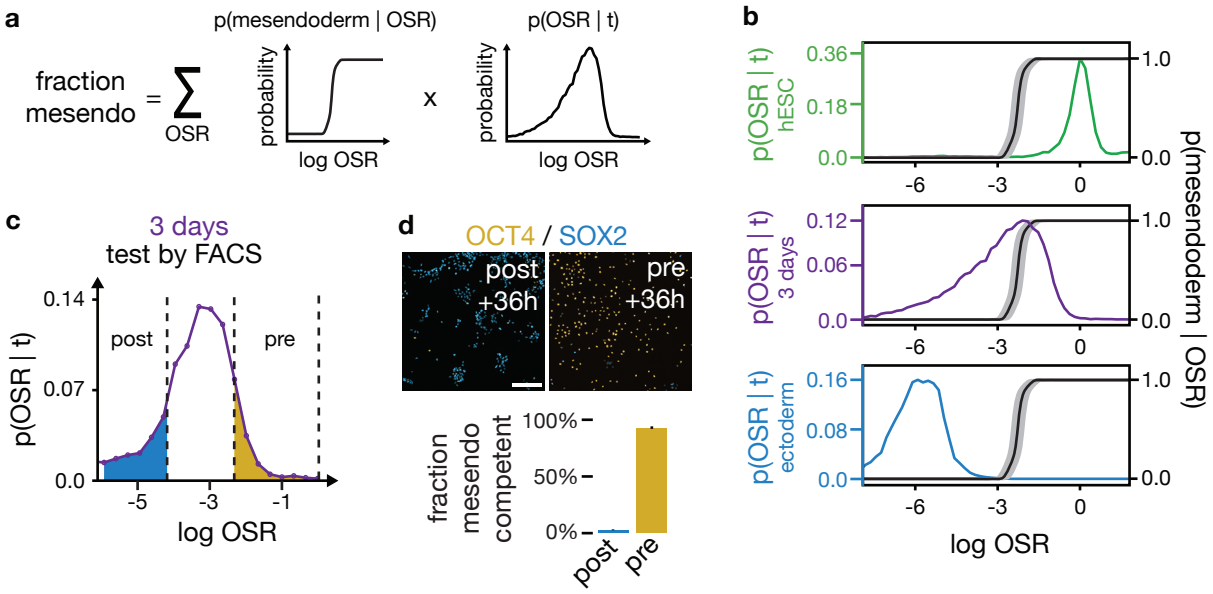

a
fraction mesendo = ∑ OSR
p(mesendoderm | OSR)
probability
log OSR
x
p(OSR | t)
probability
log OSR
b
p(OSR | t) hESC
0.36
0.18
0.0
1.0
0.0
-6
-3
0
p(OSR | t) 3 days
0.12
0.06
0.0
1.0
0.0
p(OSR | t) ectoderm
0.16
0.08
0.0
1.0
0.0
p(mesendoderm | OSR)
log OSR
c
3 days
test by FACS
p(OSR | t)
0.14
0.07
0.0
post
pre
-5
-3
-1
log OSR
d
OCT4 / SOX2
post +36h
pre +36h
fraction mesendo competent
100%
50%
0%
post
pre

Figure 3

Figure 3: Prospective isolation of cell populations based on mesendoderm competence.

a, The predicted fraction of cells that adopt a mesendodermal fate upon BMP4 and Activin stimulation at a given time t, p(mesendoderm | t), equals the sum over OSR of the probability of adopting a mesendodermal fate given OSR, p(mesendoderm | OSR), times the probability distribution of OSR in the population at that given time, p(OSR | t). **b,** Plot of p(OSR | t) as a function of OSR for the populations in (a) shown using the same color: pluripotent stem cells (green), 3 days of ectodermal differentiation (purple), ectodermal population (blue). p(mesendoderm | OSR) (black curve) inferred based on the final mesendoderm fraction from (a). p(OSR | t) moves leftward as cells differentiate. The cells in the purple population corresponding to the region with p(mesendoderm | OSR)=1 are predicted to be mesendoderm competent while the cells where p(mesendoderm | OSR)=0 are predicted not to be. **c,** Histogram showing p(OSR|t) for a cell population at t=3 days of Activin/NODAL inhibition. Overlaid on the histogram are the FACS gates based on (f) used to sort subpopulation "post," shown in blue, predicted to have lost mesendoderm competence; and subpopulation "pre," shown in yellow, predicted to retain mesendoderm competence. **d,** Top, images of pre- and post-competence-loss subpopulations after cell sorting by FACS followed by 40h of BMP4 + Activin A stimulation and immunostaining for OCT4 (yellow) and SOX2 (blue). Scale bar = 300 μm. Bottom, fraction of cells in the pre and post sorted population adopting mesendodermal fate. Error bars: std (n=3). The sorted populations essentially uniformly adopt the predicted fate.

174 As predicted, we obtained essentially pure populations of OCT4$^-$ SOX2$^+$ ectoderm-derived fates
175 from the post-competence-loss population and OCT4$^+$ SOX2$^-$ mesendoderm-derived fates from
176 the pre-competence-loss population (Figure 3d). Thus, we were indeed able to prospectively
177 isolate individual cells at points before and after loss of mesendodermal competence.

178 Having achieved our first goal of predicting mesendoderm competence, we turned to our second
179 goal: to change this competence by modulating the Waddington landscape. We hypothesized that
180 we could change the fraction of mesendoderm competent cells in one of two ways: (1) by altering
181 the position of cells along the ectoderm-directed developmental trajectory, $p(\mathrm{OSR}|t)$, so that more
182 or fewer cells would be past the point of competence loss at the time of BMP4 + Activin A signal
183 addition, or (2) by modulating mesendoderm competence along the trajectory,
184 $p(\mathrm{mesendoderm}|\mathrm{OSR})$, such that competence loss (the sharp drop-off in probability from 1 to 0)
185 happens earlier or later along the trajectory. We further hypothesized that important factors
186 controlling these two aspects would be DNA-binding factors whose expression patterns or access
187 to binding sites changed along the developmental trajectory. Thus, to identify candidate factors
188 that control these two probabilities, we characterized pre- and post-competence-loss populations
189 using RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin using
190 sequencing (ATAC-seq).

191 **Key TF families are remodeled upon loss of mesendoderm competence**

192 For our RNA-seq and ATAC-seq experiments, we isolated populations of pre- and post-
193 competence cells using FACS from a single heterogeneous population of stem cells that had been
194 subjected to 3 days of ectodermal differentiation. As a control, we reserved a small fraction of

195 sorted cells from each sample that we then treated with BMP4 and Activin A to confirm the
196 competence of that sorted population. We also included a third, mesendoderm-derived population
197 produced by subjecting pluripotent stem cells to BMP4 + Activin A for 40 hours, which allowed
198 us to distinguish lineage-specific changes in expression and chromatin accessibility from changes
199 that are shared by cells entering either lineage (Figure S3a). These populations displayed
200 significant and concerted changes in both their RNA-seq and ATAC-seq signatures upon the loss
201 of mesendoderm competence (Figure 4a).

202 Differential analysis of our RNA-seq data between the pre- and post-competence-loss populations
203 using mesendoderm as an outgroup (see methods) showed that 544 genes were upregulated
204 specifically in the post-competence-loss cells, 23 of which were TFs (Figure 4b, S3b). We also
205 found 673 genes (32 TFs) that were specifically downregulated. In particular, we observed the
206 differential expression of TFs such as SOX9, HESX1, LHX2, FOXB2, TFAP2A, TFAP2C,
207 PKNOX2, ZEB1, ZEB2, and GBX2, along with the expected expression pattern of OCT4 (Figure
208 S3c). Consistent with our earlier observation (Figure S1g-h), the differentially expressed genes did
209 not include signaling pathway components (Figure S3d). Our data further showed that competence
210 loss occurred prior to the expression of master regulators such as PAX6 and SOX1 (Figure S3c).
211 The expression pattern of all TFs is plotted in Figure S3e.

212 We validated our findings with complementary analysis of GO term enrichment (Figure S3f) and
213 ChIP-seq target enrichment (Figure S3g). Again, we observed that genes that are specifically
214 downregulated upon competence loss are most enriched for putative targets of factors previously
215 implicated in pluripotency and germ layer selection, such as SOX2, ZEB1, SMAD4, TCF3, and

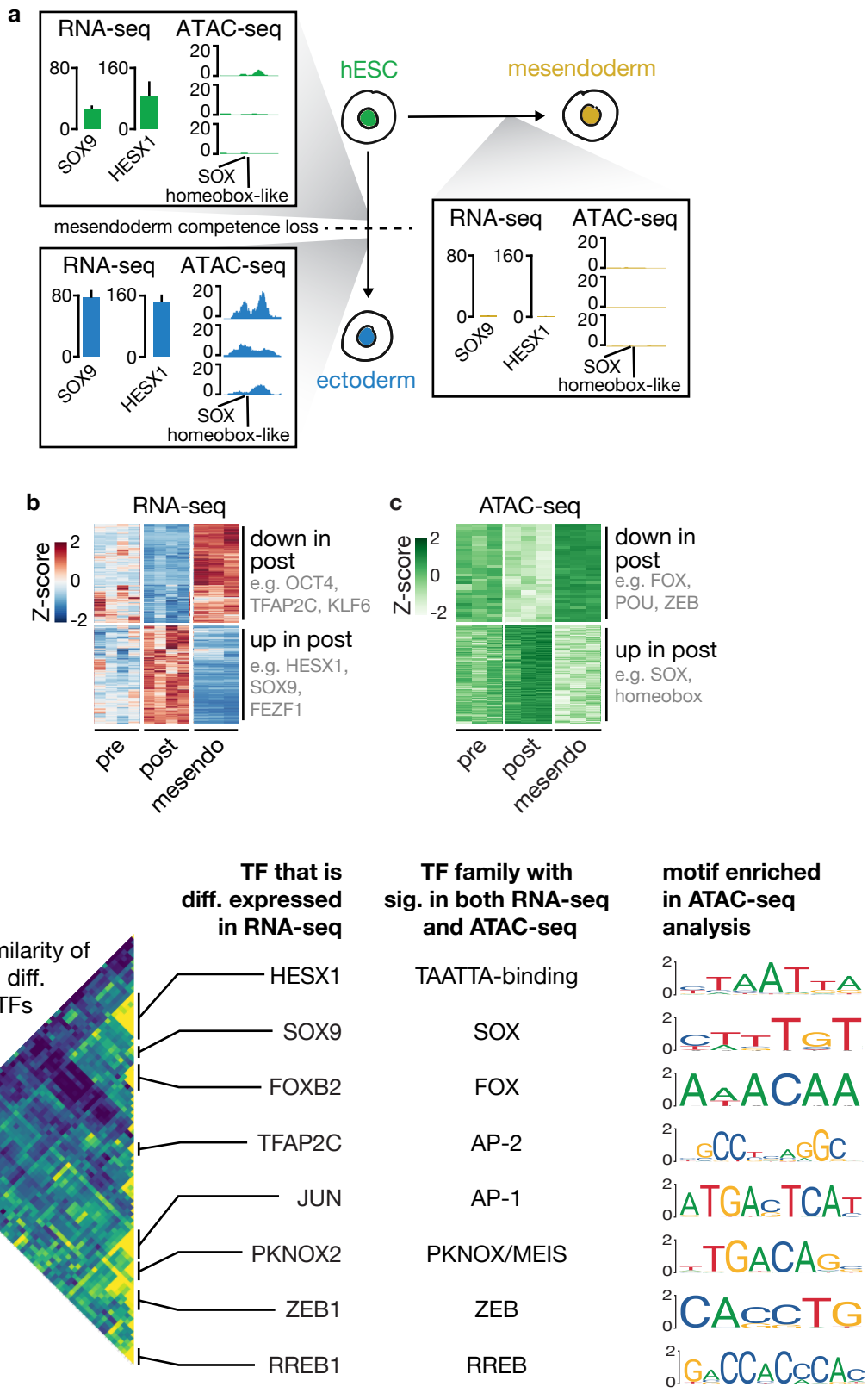

a
RNA-seq
ATAC-seq
SOX9
HESX1
SOX homeobox-like
hESC
mesendoderm
ectoderm
mesendoderm competence loss
b
RNA-seq
Z-score
down in post
e.g. OCT4, TFAP2C, KLF6
up in post
e.g. HESX1, SOX9, FEZF1
pre
post
mesendo
c
ATAC-seq
down in post
e.g. FOX, POU, ZEB
up in post
e.g. SOX, homeobox
d
pairwise similarity of motifs of all diff. expressed TFs
similarity (bits)
TF that is diff. expressed in RNA-seq
TF family with sig. in both RNA-seq and ATAC-seq
motif enriched in ATAC-seq analysis
HESX1
TAATTA-binding
SOX9
SOX
FOXB2
FOX
TFAP2C
AP-2
JUN
AP-1
PKNOX2
PKNOX/MEIS
ZEB1
ZEB
RREB1
RREB

Figure 4

Figure 4: Key TF families show concordant changes in expression and motif accessibility upon mesendoderm competence loss, suggesting perturbation candidates

a, Illustration showing the three assayed populations: pre- and post-competence loss (Left) and mesendoderm (Right). Inset boxes illustrate, with one example, combined analysis of gene expression and chromatin accessibility data from the three populations. Within each inset: Left, gene expression of SOX9 and HESX1 measured in transcripts per million, error bars: sd (n=3). Right, ATAC-seq read depth for three biological replicates at a genomic locus that contains SOX and TAATTA homeobox-like motifs, to which SOX9 and HESX1, respectively, are known to bind. SOX9 and HESX1 are upregulated in the post-committed population along with increased chromatin access to their binding site. **b,** Shown are all genes with significant lineage-specific differential expression level changes (Heat map with Z-scores) between the pre-competence-loss and post-competence loss populations (n=4) as well as mesendoderm-derived outgroup (n=3 biological replicates). Key TFs downregulated (upregulated) upon loss of mesendoderm competence include OCT4, TFAP2C, and KLF6 (HESX1, SOX9, and FEZF1). **c,** Heatmap showing row-normalized ATAC-seq read depth in all 250 bp peaks with a significant change in read depth between competent and non-competent populations (n=3). Like gene expression, these regions display clear, commitment and lineage-specific accessibility changes. **d,** Left, heatmap showing an information-based measure of similarity (see Methods) between the known DNA binding motifs of all pairs of differentially expressed TFs. Each row/column corresponds to the motif of one TF, and the matrix is arranged using hierarchical clustering. Only one half of the symmetric matrix is shown. TFs with similar preferences for DNA primary sequence cluster together, and notable families are labeled with the name of one cluster member in the column at Center Left. Center Right, the name of the TF family whose signatures are seen in both the RNA-seq analysis and ATAC-seq analysis. Right, the corresponding motif identified as significant in the ATAC-seq analysis for each labeled TF family. These key TF families create concordant signatures in gene expression and chromatin accessibility data during mesendoderm competence loss, and are good candidates for perturbing mesendoderm competence.

216 KLF4; in contrast, genes that are specifically upregulated upon mesendoderm competence loss are
217 most enriched for putative targets of Polycomb repressive complex 2 component SUZ12, which is
218 known to repress ectoderm target genes until their expression is appropriate.

219 From our ATAC-seq analysis, we observed DNA accessibility peaks that showed reproducible,
220 clear changes between groups, alongside many peaks that were present in all samples (Figure S4a).
221 Accessibility, as assayed by read depth, showed clear peaks at transcription start sites (Figure S4b).
222 Differential analysis of our ATAC-seq data between the pre- and post-competence-loss
223 populations using mesendoderm as an outgroup (see Methods) showed thousands of regions that
224 change accessibility between pre- and post-competence-loss populations. We found 2071 regions
225 that were more accessible after competence loss, and 233 that were less accessible (FDR < 0.05;
226 Figure 4c). Most of these differentially accessible regions were located in distal intergenic or
227 intronic regions (Figure S4c).

228 We next confirmed that we could reproduce expected patterns in our ATAC-seq data. Using the
229 software GREAT[31], we found that genes near loci with increased accessibility in the post-
230 competence-loss samples were most significantly enriched for orthologs of mouse genes expressed
231 in the Theiler Stage 11 neurectoderm (Figure S4d), increasing our confidence that we were indeed
232 sampling the transition we hoped to capture. We further noted that we could recover previously
233 hypothesized motif signals from our data. For example, our ATAC-seq peaks that changed
234 significantly and which also contained a compressed OCT4/SOX17 motif were almost exclusively
235 open in the mesendoderm-derived outgroup (Figure S4e), consistent with what would be predicted
236 from previous work[32]; in contrast, such regions that contained the classical OCT4/SOX2 motif

237 were generally closed in the mesendoderm-derived cells. Interestingly, we did not observe
238 significant changes in accessibility at any of the ENCODE-annotated candidate regulatory
239 elements of pluripotency genes such as OCT4, SOX2, NANOG, KLF4, and MYC (Figure S4f).

240 To find TFs that potentially bind to the differentially accessible regions, we searched for sequence
241 motifs that were enriched at these loci. We found more than 20 such motifs, many of which
242 matched the known DNA-binding motifs of the differentially expressed TF families we had
243 identified, including motifs that were similar to those bound by SOX, forkhead box (FOX), AP-2,
244 AP-1, TAATTA-binding homeobox-like, PKNOX/MEIS, Zinc Finger E-Box Binding Homeobox
245 (ZEB), and POU family TFs (Figure S4g), the latter of which includes OCT4 as a member. As a
246 complementary analysis, we determined which known sequence motifs could best explain the
247 changes in chromatin accessibility that we observed across the point of competence loss (Figure
248 S4h).

249 We clustered the known binding motifs of the differentially expressed TFs by calculating a
250 measure of similarity between each pair of motifs, and we found that many such TFs shared similar
251 binding motifs (Figure 4d). These clusters correspond to major TF families—including the FOX,
252 SOX, AP-2, PKNOX/MEIS, ZEB, and homeobox-like TAATTA-binding TF families—and
253 multiple members of each family are differentially expressed. Each of these major TF family DNA
254 binding motifs was also enriched in the ATAC-seq analysis, indicating that the expression changes
255 of these TFs have clear signatures in the chromatin accessibility data. Taken together, the gene
256 expression and chromatin accessibility data revealed that a small, core set of TFs from key families
257 are remodeled upon loss of mesendodermal competence (Figure S5a).

258 **Key TFs independently modulate barrier between fates and location along developmental**
259 **trajectory**

260 We hypothesized that perturbation of these factors could modulate the position of cells along the
261 ectoderm-directed developmental trajectory, $p(\mathrm{OSR}|t)$, or the cell's mesendoderm competence
262 along the trajectory, $p(\mathrm{mesendoderm}|\mathrm{OSR})$. To test these predictions, we selected 36 genes,
263 composed largely of the differentially expressed TFs in our core network plus their paralogs
264 (Figure S5b and S5c; see Methods), for overexpression studies. We seeded two parallel samples
265 of pluripotent stem cell populations for each candidate gene. Using a lentiviral delivery system,
266 we transduced cells in both samples with a payload of a single gene of interest separated from the
267 C-terminal end of a mCerulean cyan fluorescent protein (CFP) by a P2A self-cleaving peptide
268 sequence, all under the control of an EF-1α promoter (Figure S6a). Viral transduction was initiated
269 at the onset of ectoderm-directed differentiation (Figure S6b). After three days of differentiation,
270 we analyzed one sample by flow cytometry to measure $p(\mathrm{OSR}|t)$. We switched the other sample
271 to media containing BMP4 + Activin A for 42 hours and assayed the final fate fractions by flow
272 cytometry. We confirmed that our non-transduced controls had produced about 50% OCT4:RFP^{+}
273 SOX2:YFP^{-} mesendoderm-derived cells and 50% OCT4:RFP^{-} SOX2:YFP^{+} ectoderm-derived
274 cells. The transduced cells, in contrast, expressed CFP (Figure S6c) and displayed varying final
275 lineage proportions. The co-expressed CFP marker allowed us to monitor transduction efficiency
276 at the single-cell level. We titrated viral concentration to achieve <50% transduction so that each
277 sample also contained many non-transduced (CFP^{-}) cells to serve as an internal control population.

278 In principle, a change in the fraction of lentiviral-perturbed cells that form mesendoderm-derived
279 cell types after BMP4 and Activin A treatment could result from either of the hypothesized effects:

280 first, the perturbation could facilitate or impede the initial ectoderm-directed differentiation,
281 thereby changing $p(\mathrm{OSR}|t)$; second, the perturbation could change $p(\mathrm{mesendoderm}|\mathrm{OSR})$, the
282 probability of transitioning to mesendoderm given the location along the developmental trajectory.
283 For each candidate TF, we computed $p(\mathrm{mesendoderm}|\mathrm{OSR})$ based on the $p(\mathrm{OSR}|t)$ and the
284 observed final fraction of mesendoderm of lentiviral-perturbed cells (Equation 1).

285 For certain candidate TFs, their overexpression principally affected mesendoderm competence by
286 shifting the barrier between fates, captured by $p(\mathrm{mesendoderm}|\mathrm{OSR})$. For example,
287 overexpression of JUNB, POU2F3, or FOXB2 increased mesendoderm competence by shifting
288 $p(\mathrm{mesendoderm}|\mathrm{OSR})$ (Figure 5a). We concluded that these candidates tuned mesendoderm
289 competence by moving the barrier between fates on the developmental landscape.

290 In contrast, other candidates affected progression along the developmental trajectory, captured by
291 $p(\mathrm{OSR}|t)$. For example, overexpression of SOX9 facilitated movement along the developmental
292 trajectory, thereby shifting the distribution of $p(\mathrm{OSR}|t)$ towards the ectoderm fates (Figure S6d).
293 In contrast, overexpression of TFAP2C hindered movement along the developmental trajectory,
294 thereby shifting $p(\mathrm{OSR}|t)$ towards the pluripotent state. We concluded that these candidates tuned
295 fate competence by changing cellular position along the developmental trajectory.

296 Notably, our candidates impacted competence by tuning $p(\mathrm{mesendoderm}|\mathrm{OSR})$ and $p(\mathrm{OSR}|t)$
297 independently of one another (Figure 5b, S6e). For example, while overexpression of FEZF1 and
298 TFAP2A both shifted $p(\mathrm{OSR}|t)$ towards the pluripotent state, they shifted $p(\mathrm{mesendoderm}|\mathrm{OSR})$
299 in opposite directions. The outcomes seen upon overexpression of FEZF1, TFAP2A, OTX2 and

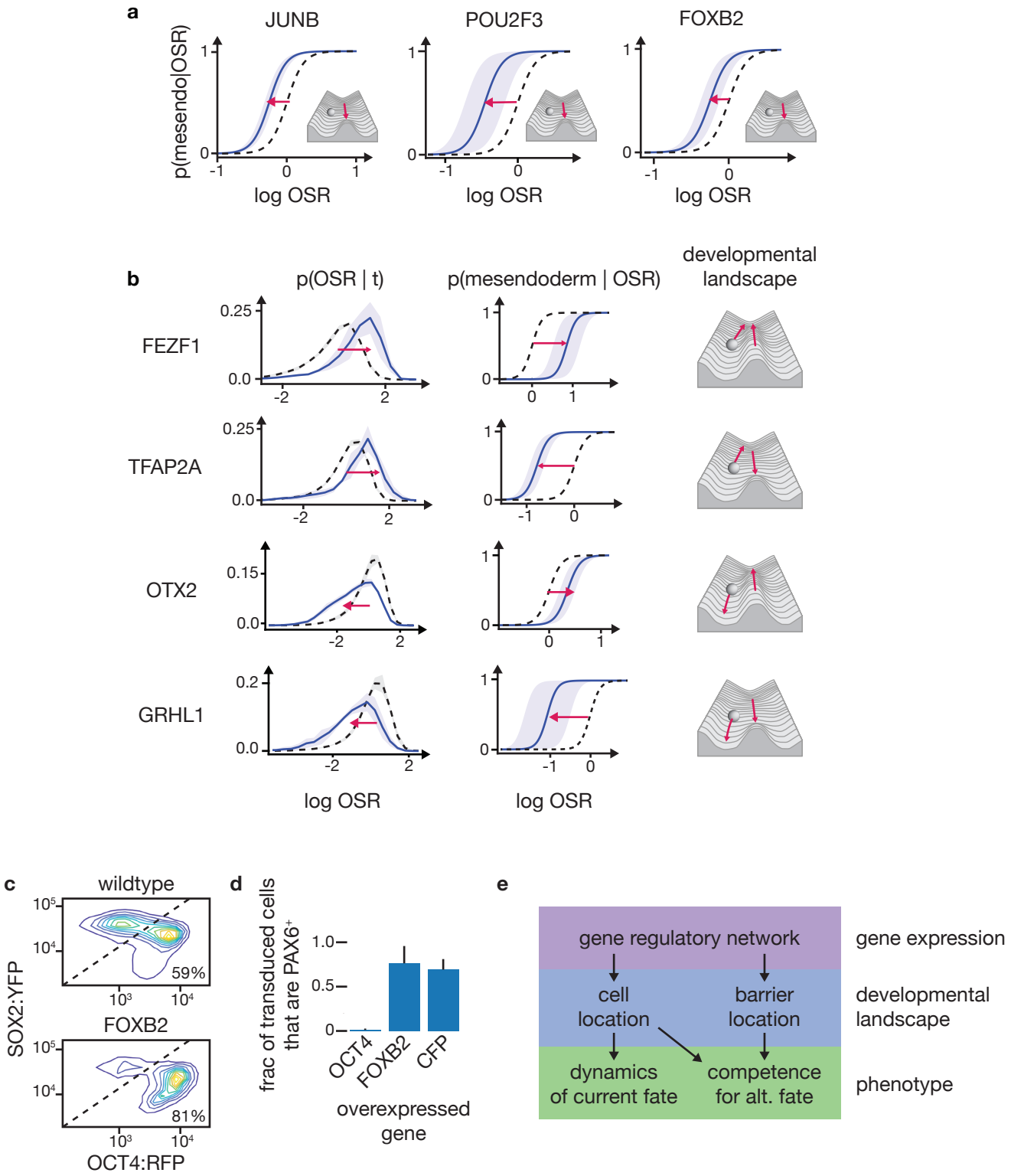

a
JUNB
POU2F3
FOXB2
p(mesendo|OSR)
log OSR
b
p(OSR | t)
p(mesendoderm | OSR)
developmental landscape
FEZF1
TFAP2A
OTX2
GRHL1
log OSR
log OSR
c
wildtype
FOXB2
SOX2:YFP
OCT4:RFP
59%
81%
d
frac of transduced cells that are PAX6+
OCT4
FOXB2
CFP
overexpressed gene
e
gene regulatory network
cell location
barrier location
dynamics of current fate
competence for alt. fate
gene expression
developmental landscape
phenotype

Figure 5

Figure 5: Perturbation of TFs candidates from RNA-seq and ATAC-seq analysis can independently tune progression along the developmental trajectory and inter-fate barrier.

a, Over expression of three candidates, JUNB, POU2F3, and FOXB2, shifted p(mesendoderm | OSR) to keep cells mesendoderm competent for longer along the ectodermal differentiation trajectory. Black dotted: wildtype; blue: transduced cells; shaded: std (n=3, FOXB2 n=4). Inset, schematic depicting the effect of overexpression on the Waddington landscape. **b,** Over expression of FEZF1, TFAP2A, OTX2, and GRHL1 (top to bottom) altered both p(OSR | t) (left column) and p(mesendoderm | OSR) (middle column). Black dotted: wildtype control; blue: transduced cells, shaded: std (n=3). Left, p(OSR | t). Center, p(mesendoderm | OSR). Right, schematic as in (a). Overexpression of candidates can tune p(OSR | t) and p(mesendoderm | OSR) independently. **c,** Top, FACS density plot (OCT4 vs SOX2) of wild type cells (un-transduced CFP-) showing two peaks corresponding to ectodermal lineage above diagonal (41% of cells) and mesendodermal lineage, below diagonal (59% of cells). Bottom, transduced cells (CFP+) from the same population with CFP:P2A:FOXB2 cassette show 81% of cells adopting mesendodermal fate. Overexpression of FOXB2 increases the fraction of cells that adopt a mesendodermal fate. **d,** Fraction of cells adopting ectodermal fate (PAX6+) under ectodermal differentiation conditions. Cells with CFP:P2A:FOXB2 or the control CFP:P2A:CFP show normal ectodermal differentiation. OCT4 over expression precludes PAX6 expression and ectodermal differentiation. Error bars: std. **e,** The gene regulatory network governs both the progression of the cell along the developmental trajectory and shape of barrier between fates. In turn, both the cell's location on the developmental landscape and the location of the barrier determine the cell's fate in response to alternative signals. Only perturbations of the location of the barrier alters competence without altering the dynamics of movement along the original trajectory in the absence of a new signal.

300 GRHL1 together represent all four possible combinations of directional shifts in $p(\mathrm{OSR}|t)$ and
301 $p(\mathrm{mesendoderm}|\mathrm{OSR})$. Thus, our candidates independently tuned $p(\mathrm{OSR}|t)$ and
302 $p(\mathrm{mesendoderm}|\mathrm{OSR})$, thereby separately altering cellular location on the Waddington landscape
303 and the barrier between fates.

304 Lastly, we explored the downstream developmental consequences of mesendoderm competence
305 modulation via these two mechanisms. Overexpression of FOXB2, a candidate that increased
306 mesendoderm competence primarily via a shift in $p(\mathrm{mesendoderm}|\mathrm{OSR})$, increased the fraction
307 of cells that adopted mesendoderm by 32±5% (Figure 5c). Despite this change in mesendoderm
308 competence, FOXB2 overexpression did not prevent normal neurectodermal differentiation as
309 assayed by PAX6 expression in the absence of the BMP and Activin signal (Figure 5d, S6f). Thus,
310 alternative fate competence can be modulated without preventing normal lineage progression in
311 the absence of alternative-fate-inducing signals.

312 Discussion

313 Our finding that competence for an alternative fate can be controlled either by changing the
314 location of the cell along the developmental trajectory or moving the barrier that prevents cells
315 from crossing over to the mesendoderm valley suggests possible evolutionary and developmental
316 consequences. While moving the cell along the trajectory interferes with normal development in
317 the absence of alternative signal, moving the barrier does not (Figure 5e). These two effects
318 represent fundamentally different mechanisms for tuning competence, and both could be at play
319 in the developing embryo. During the patterning of the mammalian epiblast, for example, the
320 mesendodermal progenitors are generated along the primitive streak as it extends anteriorly from

321 the posterior end of the epiblast. We speculate that changing the dynamics of epiblast competence
322 loss anterior to the primitive streak could be a possible mechanism for tuning the length and extent
323 of the streak. Further investigation of the role of competence modulation during mammalian
324 gastrulation could be an important element in a full description of this important process, and the
325 same mechanism could be acting to tune relative tissue sizes during any given cellular decision.

326 We expect that discovering reaction coordinates for different lineage decisions and monitoring the
327 dynamics of differentiation along these coordinates will be fruitful for dissecting many cell fate
328 decisions and state transitions. Our recent computational work has demonstrated how to find
329 candidate genes that accurately reflect any given lineage choice[26,27], which greatly simplifies the
330 prospect of following this approach in another context. Indeed, several other choices and
331 competence restrictions occur immediately adjacent to the loss of mesendoderm competence in
332 the early germ layer lineage tree, such as the presumptive loss of non-neural potential during
333 neurectoderm fate determination or the loss of ectoderm competence during mesendoderm
334 differentiation. In sum, these findings reveal two independently tunable mechanisms for adopting
335 alternative fates along a developmental trajectory. Understanding how competence is tuned along
336 distinct mechanistic axes will be crucial for elucidating the dynamics of mammalian embryonic
337 patterning during development.

338

339 **References**

340 1. Gilbert, S. The Developmental Mechanics of Cell Specification. in *Developmental Biology*
341 (Sinauer Associates, 2000).
342 2. Handyside, A. H. Time of commitment of inside cells isolated from preimplantation mouse
343 embryos. **45**, 37–53 (1978).
344 3. Pedersen, R. A., Wu, K. & BaŁakier, H. Origin of the inner cell mass in mouse embryos: Cell
345 lineage analysis by microinjection. *Dev. Biol.* **117**, 581–595 (1986).
346 4. Berg, D. K. *et al.* Trophectoderm Lineage Determination in Cattle. *Dev. Cell* **20**, 244–255
347 (2011).
348 5. Rossant, J. & Lis, W. T. Potential of isolated mouse inner cell masses to form trophectoderm
349 derivatives in vivo. *Dev. Biol.* **70**, 255–261 (1979).
350 6. Rossant, J. & Vijh, K. M. Ability of outside cells from preimplantation mouse embryos to form
351 inner cell mass derivatives. *Dev. Biol.* **76**, 475–482 (1980).
352 7. Oron, E. & Ivanova, N. Cell fate regulation in early mammalian development. *Phys. Biol.* **9**,
353 045002 (2012).
354 8. Kiecker, C., Bates, T. & Bell, E. Molecular specification of germ layers in vertebrate embryos.
355 *Cell. Mol. Life Sci.* **73**, 923–947 (2016).
356 9. Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle
357 gene transcription. *Development* **132**, 2685–2695 (2005).
358 10. Patthey, C. & Gunhaga, L. Signaling pathways regulating ectodermal cell fate choices. *Exp.*
359 *Cell Res.* **321**, 11–16 (2014).

360 11. Trompouki, E. *et al.* Lineage regulators direct BMP and Wnt pathways to cell-specific
361 programs during differentiation and regeneration. *Cell* **147**, 577–589 (2011).
362 12. Mullen, A. C. *et al.* Master transcription factors determine cell-type-specific responses to TGF-
363 β signaling. *Cell* **147**, 565–576 (2011).
364 13. Takaoka, K. & Hamada, H. Cell fate decisions and axis determination in the early mouse
365 embryo. *Development* **139**, 3–14 (2012).
366 14. Chng, Z., Teo, A., Pedersen, R. A. & Vallier, L. SIP1 Mediates Cell-Fate Decisions between
367 Neuroectoderm and Mesendoderm in Human Pluripotent Stem Cells. *Cell Stem Cell* **6**, 59–70
368 (2010).
369 15. Jang, J. *et al.* Primary Cilium-Autophagy-Nrf2 (PAN) Axis Activation Commits Human
370 Embryonic Stem Cells to a Neuroectoderm Fate. *Cell* **165**, 410–420 (2016).
371 16. Sheng, G., Dos Reis, M. & Stern, C. D. Churchill, a Zinc Finger Transcriptional Activator,
372 Regulates the Transition between Gastrulation and Neurulation. *Cell* **115**, 603–613 (2003).
373 17. Yang, Y. *et al.* Pax6 Is a Human Neuroectoderm Cell Fate Determinant. *Cell Stem Cell* **7**, 90–
374 100 (2010).
375 18. Hou, P.-S. *et al.* LHX2 regulates the neural differentiation of human embryonic stem cells via
376 transcriptional modulation of PAX6 and CER1. *Nucleic Acids Res.* **41**, 7753–7770 (2013).
377 19. Mizuseki, K., Kishi, M., Matsui, M., Nakanishi, S. & Sasai, Y. Xenopus Zic-related-1 and
378 Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural
379 induction. *Dev. Camb. Engl.* **125**, 579–87 (1998).
380 20. Loh, K. M. *et al.* Mapping the Pairwise Choices Leading from Pluripotency to Human Bone,
381 Heart, and Other Mesoderm Cell Types. *Cell* **166**, 451–467 (2016).

382 21. Lupo, G. *et al.* Inhibition of Activin/Nodal signaling promotes specification of human
383 embryonic stem cells into neuroectoderm. *Dev. Biol.* **313**, 107–117 (2007).
384 22. Li, L. *et al.* Ectodermal progenitors derived from epiblast stem cells by inhibition of Nodal
385 signaling. *J. Mol. Cell Biol.* **7**, 455–465 (2015).
386 23. Waddington, C. H. *The Strategy of the Genes*. (Routledge, 1957).
387 24. Xu, X., Browning, V. L. & Odorico, J. S. Activin, BMP and FGF pathways cooperate to
388 promote endoderm and pancreatic lineage cell differentiation from human embryonic stem
389 cells. *Mech. Dev.* **128**, 412–427 (2011).
390 25. Li, L. *et al.* Location of transient ectodermal progenitor potential in mouse development.
391 *Development* **140**, 4533–4543 (2013).
392 26. Jang, S. *et al.* Dynamics of embryonic stem cell differentiation inferred from single-cell
393 transcriptomics show a series of transitions through discrete cell states. *Elife* **6**, (2017).
394 27. Furchtgott, L. A., Melton, S., Menon, V. & Ramanathan, S. Discovering sparse transcription
395 factor codes for cell states and state transitions during development. *eLife* **6**, 1–33 (2017).
396 28. Thomson, M. *et al.* Pluripotency factors in embryonic stem cells regulate differentiation into
397 germ layers. *Cell* **145**, 875–889 (2011).
398 29. Greber, B. *et al.* FGF signalling inhibits neural induction in human embryonic stem cells: FGF
399 inhibits neural induction in hESCs. *EMBO J.* **30**, 4874–4884 (2011).
400 30. Zhang, Z., Zwick, S., Loew, E., Grimley, J. S. & Ramanathan, S. Mouse embryo geometry
401 drives formation of robust signaling gradients through receptor localization. *Nat. Commun.* **10**,
402 (2019).

403 31. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat.*
404 *Biotechnol.* **28**, 495–501 (2010).
405 32. Merino, F. *et al.* Structural basis for the SOX-dependent genomic redistribution of OCT4 in
406 stem cell differentiation. *Structure* **22**, 1274–1286 (2014).
407 33. Etoc, F. *et al.* A Balance between Secreted Inhibitors and Edge Sensing Controls Gastruloid
408 Self-Organization. *Dev. Cell* **39**, 302–315 (2016).
409 34. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A method for assaying
410 chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).
411 35. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful
412 Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
413 36. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.*
414 **16**, e2005970 (2018).
415 37. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
416 38. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016
417 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
418 39. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more
419 genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic*
420 *Acids Res.* **47**, D419–D426 (2019).
421 40. Edelstein, A. D. *et al.* Advanced methods of microscope control using μManager software. *J.*
422 *Biol. Methods* **1**, 10 (2014).
423

424 # Methods

425 **Cell lines**

426 We conducted our experiments using WA01 (H1, WiCell) human embryonic stem cells. We also
427 used an H1 cell line in which both OCT4 and SOX2 were tagged with fluorescent proteins as
428 previously described[30]. In these cells, one endogenous copy of OCT4 was replaced with
429 OCT4:tdTomato followed by an internal ribosomal entry site and a neomycin resistance gene to
430 allow for selection, and one endogenous copy of SOX2 was replaced with
431 SOX2:FLAG:Citrine:P2A:PuroR.

432 **Cell culture**

433 hESCs were cultured in 6-well tissue culture dishes treated with Matrigel (Corning) and supplied
434 with mTeSR media (STEMCELL Technologies) according to the manufacturer's specifications.
435 For routine culture, we passaged by washing with phosphate buffered saline (PBS) followed by
436 ReLeSR (Stem Cell) treatment according to the manufacturer's instructions. Cells were passaged
437 in clumps of 8-10 cells and seeded in mTeSR supplemented with the Rho-associated protein kinase
438 inhibitor γ-27632 (STEMCELL Technologies) at 10 μM for the first day to improve survival. All
439 cell lines used were routinely tested for mycoplasma contamination.

440 For all experiments, we seeded cells on polyester membrane filters (Sterlitech) with 3 μm pores
441 that had been treated with Matrigel. We chose this substrate to allow all cells to receive the BMP
442 and Activin/NODAL signals we added to the media. TGF-β superfamily receptors, such as those
443 for BMP4 and Activin A, are localized basolaterally in epithelial stem cell colonies and *in vivo* in
444 the epiblast, so they are insulated from ligands in the apical media or luminal fluid[30,33]. Typical

445 tissue culture conditions allow for only the cells on the edge of the colony to receive signals, but
446 growing cells on a membrane allows all cells in a colony access to the BMP and Activin ligands.

447 For live cell imaging, membranes were first glued to a custom 300 µm thick stainless-steel washer
448 with Cytoseal 60 (Thermo Fisher), allowed to dry, sterilized with washes in 70% ethanol and with
449 UV treatment, then treated with Matrigel for cell seeding.

450 **Differentiation conditions**

451 Differentiation towards the ectoderm lineage was effected using mTeSR supplemented with 0.5
452 µM A83-01 (R&D Systems), a small molecule inhibitor of Activin and Nodal signaling. BMP4 +
453 Activin A treatment was accomplished by treating cells with mTeSR supplemented with 3 ng/mL
454 recombinant human BMP4 protein (R&D Systems) and 100 ng/mL recombinant human Activin
455 A protein (R&D Systems). For neurectoderm-directed differentiation, we inhibited BMP signaling
456 with 0.5 µM LDN-193189 in addition to Activin/Nodal inhibition with 0.5 µM A83-01 for 6 days.

457 **Flow cytometry**

458 Cells were washed with PBS (Lonza) and removed from membranes by treatment with Accutase
459 (Innovative Cell Technologies) until the cells were dissociated, about 20 minutes. Cells were
460 analyzed on an LSRFortessa (BD Biosciences).

461 **Fluorescence activated cell sorting**

462 Accutase-dissociated cells were sorted using a BD Aria III (BD Biosciences) using a 100 µm
463 nozzle. Cells were gated such that the pre-competence-loss population was taken as the cells with
464 the top 10-15% OCT4:RFP to SOX2:YFP ratio, and the post-competence-loss population was the

465 bottom 10-15% OCT4:RFP to SOX2:YFP ratio. We sorted around 250,000 cells per subpopulation
466 in a typical experiment. Populations were sorted into 1.5 mL centrifuge tubes (Eppendorf) filled
467 with 500 μL of mTeSR supplemented with 10 μM γ-27632; by the end of the sort, ~800 μL of
468 sheath and sorted cells had been added to each tube. After the sort had completed, we pelleted the
469 cells in a microcentrifuge at 250 xg for 3 minutes, then resuspended in PBS.

470 For each sorted sample, about 10% of the sorted cells were reserved for competence testing to
471 confirm the pre-/post-competence-loss status of the sorted population. These cells were seeded
472 back into glass-bottom 24-well plates (Ibidi) treated with Matrigel and filled with 1 mL of mTeSR
473 supplemented with γ-27632 and allowed to recover for 3 hours. The media was then changed to
474 mTeSR supplemented with BMP4 and Activin A for 36 hours. Cells were fixed and stained for
475 OCT4 and SOX2 according to the protocols described under "Immunofluorescence."

476 **RNA-seq**

477 Total RNA was prepared from sorted or dissociated cells with an RNeasy Mini Kit (Qiagen)
478 according to the manufacturer's instructions. For the mesendoderm-derived outgroup samples, the
479 input to the RNA extraction kit was a cell population directly after dissociation with Accutase; for
480 FACS sorted populations, the input was sorted cells suspended in PBS. RNA integrity was
481 quantified with a TapeStation 4200 (Agilent). All RINe scores were ≥ 9.9. Sequencing libraries
482 were prepared by the Bauer Core at Harvard University using a Kapa mRNA Hyper Prep kit with
483 Poly-A selection. Sequencing was performed on a NextSeq High output flow cell that generated
484 paired-end 38bp reads. We obtained ≥42M reads per sample.

485 **ATAC-seq**

486 ATAC-seq was performed as previously described[34]. Briefly, live cells were lysed and incubated
487 with Tn5 transposase for 30 min at 37°C. After DNA purification, samples were amplified for the
488 appropriate number of cycles as determined by qPCR to minimize PCR bias. Sequencing was
489 performed by the sequencing core at Massachusetts General Hospital. We obtained ~100M
490 mapped paired-end reads per sample.

491 **Plasmid construction**

492 Overexpression targets were subcloned from plasmids available through the Harvard PlasmID
493 database, where available. Other targets were cloned from complementary DNA (cDNA) libraries.
494 To prepare cDNA libraries for cloning, we differentiated human stem cells for 3 days in mTeSR
495 + 0.5 μM A8301 and extracted RNA with RNeasy Mini Kit (Qiagen) according to the
496 manufacturer's instructions. We then performed first strand cDNA synthesis using SuperScript II
497 Reverse Transcriptase (Thermo Fisher). We amplified the relevant cDNAs using Phusion
498 polymerase (NEB) or Kapa HiFi (Kapa Biosystems). The OCT4 DNA binding domain and the
499 SOX2 DNA binding domain (OCT4DBD and SOX2DBD) were amplified from cDNA. The
500 OCT4DBD consisted of amino acids 131-296 of OCT4A (NCBI reference sequence:
501 NM_002701.5). The SOX2DBD consisted of amino acids 37-117 of SOX2 (NCBI reference
502 sequence: NM_003106.3). All cDNA-amplified clones were fully sequence confirmed by Sanger
503 sequencing (Genewiz).

504 We constructed a vector from the second-generation lentiviral transfer backbone pWPXL with an
505 EF-1α promoter. pWPXL was a gift from Didier Trono (Addgene plasmid #12257;
506 http://n2t.net/addgene:12257; RRID:Addgene_12257). We first joined sequences for fluorescent

507 protein mCerulean (CFP) and 2A peptide P2A (a ribosomal skip sequence) with Q5 (NEB) fusion
508 PCR and added them to the pWPXL vector with Gibson Assembly Master Mix (NEB). We then
509 constructed final transfer vectors by inserting target cDNA after the P2A using Gibson assembly.
510 All constructed vectors were sequence confirmed at Gibson assembly junctions by Sanger
511 sequencing (Genewiz) prior to use. Plasmids were grown and stored in NEB Stable *E. coli* (NEB).

512 **Lentiviral overexpression and flow cytometry analysis**

513 To produce virus, we used jetPrime (Polyplus) according to the manufacturer's instructions to
514 transfect Lenti-X 293 T HEK cells (Takara) with lentiviral production plasmids pMD2.G and
515 psPAX2 along with our individual transfer plasmids. pMD2.G and psPAX2 were gifts from Didier
516 Trono (Addgene plasmid #12259; http://n2t.net/addgene:12259; RRID:Addgene_12259; Addgene
517 plasmid #12260; http://n2t.net/addgene:12260; RRID:Addgene_12260) We collected viral media
518 at 24 and 48 hours and concentrated using Lenti-X Concentrator (Clontech) according to the
519 manufacturer's instructions.

520 We seeded human stem cells in mTeSR medium containing γ-27632 on Matrigel-treated
521 membrane filters as described above. We treated cells with 1x and 3x viral titer at 24 hours and 48
522 hours post-seeding, respectively, in order to obtain transduction efficiency of ~10%. 1x viral
523 treatment was performed simultaneously with the beginning of A83-01 treatment.

524 Two samples of each overexpression condition were performed in parallel. We harvested cells of
525 one sample after 3 days of treatment with A83-01 and the cells of the second sample after 3 days
526 of A83-01 followed by 42 hours of BMP4 + Activin A treatment. We analyzed each sample
527 immediately after harvest using an LSRFortessa (BD Biosciences).

528 We analyzed differential OCT4:RFP to SOX2:YFP ratio distributions between CFP-positive and
529 CFP-negative populations of each 3-day sample by calculating the Kullback-Leibler divergence in
530 MATLAB (MathWorks). To determine differences in proportions of end fates
531 (OCT4:RFP+/SOX2:YFP- and OCT4:RFP-/SOX2:YFP+), we manually gated ectoderm and
532 mesendoderm populations using a custom MATLAB script and used identical gates for both CFP-
533 positive and CFP-negative populations. We performed at least 3 biological replicates for all
534 candidates that showed an initial phenotype except for SOX2, AHR, ARNT2, and GBX2, each of
535 which had two replicates. Significance compared to the CFP-overexpressing negative control was
536 determined using a two-sided *t*-test, and we controlled the FDR at 10% across the set of all
537 candidates using the method of Benjamini and Hochberg[35].

538 **Immunofluorescence**

539 Cells were fixed with 4% formaldehyde for 15 min at room temperature. Fixed cells were treated
540 with blocking buffer (PBS + 5% normal donkey serum + 0.3% Triton X-100) for 1 h, then
541 overnight at 4°C with primary antibody diluted in staining buffer (PBS + 1% BSA + 0.3% Triton
542 X-100). The following primary antibodies were used: OCT4 (1:400, Cell Signaling C30A3);
543 SOX2 (1:400, Thermo Fisher BTJCE); SOX17 (1:100, R&D Systems AF1924); phosphorylated
544 SMAD1/5/9 (1:200, Cell Signaling D5B10); phosphorylated-SMAD2 (1:200, Cell Signaling
545 E8F3R); PAX6 (1:200, DSHB AB_528427); and NANOG (1:500, R&D Systems AF1997). The
546 PAX6 antibody was deposited to the DSHB by Kawakami, A. (DSHB Hybridoma Product PAX6).
547 After overnight incubation, samples were washed three times with PBS, then secondary antibodies
548 diluted in staining buffer were added. We used the following secondary antibodies all at a dilution
549 of 1:1000: donkey anti-rabbit Alexa 568 (Thermo Fisher), donkey anti-rat Alexa 488 (Thermo

550 Fisher), and donkey anti-mouse Alexa 647 (Thermo Fisher). We incubated with a 300 nM DAPI
551 (Thermo Fisher) solution in PBS for 5 minutes to visualize DNA. For analysis of the resulting
552 images, we used CellProfiler 3.1.8[36] to segment well-separated nuclei for samples where
553 automated segmentation performed well (those shown in Figure 2), and we used manual
554 segmentation for more challenging images (those shown in Figures 1, S1, and S6).

555 **ATAC-seq analysis**

556 Reads were trimmed using NGmerge 0.2_dev in adapter removal mode with minimum overlap (-
557 e flag) set to 20 to remove any remaining adapter sequence. Reads were aligned to the hg38 build
558 of the human genome using bowtie2 2.2.9 using the --very-sensitive preset and with a maximum
559 fragment size of 2000, then collated with samtools 1.9. Duplicate fragments were removed with
560 picard 2.8.0. Peaks were called with MACS2 2.1.1 in callpeak -f BAMPE mode. Differentially
561 accessible peaks were identified using the Bioconductor package DiffBind 2.12.0 in R 3.6.1. Peaks
562 were annotated by genomic region type using ChIPSeeker 1.20.0.

563 For differential accessibility analysis with DiffBind and DESeq2, we used a design matrix with an
564 "sample" column, which indicated the well from which the cells had been sorted (since each pair
565 of pre- and post-competence-loss samples was derived from a single population sorted by FACS),
566 and a "competenceloss" column, which was 1 for the post- competence-loss population and 0 for
567 the pre-competence-loss and mesendoderm-derived populations. Thus, we identified regions that
568 changed specifically with competence loss while controlling for original sample identity.

569 The primary DNA sequences of differentially accessible peaks were retrieved from Ensembl and
570 examined for motifs using MEME-ChIP 5.0.3. ATAC-seq read depth was modeled as a function

571 of known motif presence using chromVAR 1.4.1. Significant motif matches were identified with
572 FIMO 5.0.3. For the gene regulatory network, possible associations between genomic regions and
573 target genes were identified using CisMapper 5.0.5. The full list of human TFs and the motifs for
574 each TF were extracted from the list in Lambert *et al.*[37]. Mutual information between pairs of
575 motifs was calculated with a custom python script.

576 Enrichment of functions of genes near differentially accessible genomic regions was performed
577 using the web interface of GREAT[31] in June 2019.

578 **RNA-seq analysis**

579 Reads were pseudoaligned using kallisto 0.45.1 to transcripts from the human genome build hg38.
580 Differentially expressed genes were identified using DESeq2 1.24.0 on R 3.6.1. For analysis with
581 DESeq2 when comparing pre- and post-competence-loss populations, we used a design matrix
582 with an "sample" column, which indicated the well from which the cells had been sorted and a
583 "population" column, which indicated the pre- or post-competence-loss state. For comparison of
584 pre-competence-loss and mesendoderm-derived populations, our design matrix contained
585 sequencing batch and pre-competence-loss or mesendoderm-derived population identity.

586 For clustering the motifs of differentially expressed transcription factors, similarity between each
587 pair of motifs was quantified as the Kullback-Leibler divergence of the product of the two motifs
588 from a reference distribution, which was the product of two uniform motifs (0.25 probability for
589 each base at each position). Motif alignment was performed by calculating the aforementioned
590 divergence at each possible offset and using the maximum value obtained at any offset. This
591 calculation was performed using a custom python script. The linkage was computed using the

592 scipy.cluster.hierarchy.linkage function from scipy 1.3.0 with the "average" clustering method and
593 the "braycurtis" distance.

594 ChIP-seq target set enrichments were calculated using the Enrichr[38] web interface in July 2019.
595 GO term enrichment was calculated using the PANTHER[39] web interface in July 2019.

596 **Overexpression candidate selection**

597 We selected TFs by incorporating information from both RNA-seq and ATAC-seq analyses. We
598 began with all TFs that were differentially expressed between the pre- and post-competence-loss
599 populations ($q < 0.05$ with DESeq2). A gene was considered to be a TFs if it was so annotated in
600 Lambert *et al.*[37]. We then limited this list only to those that were expressed in a lineage-specific
601 pattern and had above-background expression levels in at least one of the three populations. We
602 defined genes with a lineage-specific expression pattern as those genes that (1) were differentially
603 expressed between pre- and post-competence-loss populations and (2) either were not
604 differentially expressed from the pre-competence-loss to mesendoderm populations or were
605 differentially expressed in the opposite direction (upregulated from pre- to post- and
606 downregulated from pre- to mesendoderm, or vice versa). By these criteria, 23 TFs were
607 specifically upregulated with competence loss and 32 were specifically downregulated with
608 competence loss. We also added select paralogs of the TFs that passed our expression pattern
609 cutoffs: POU6F1, GRHL1, POU2F3, and FOXJ2, along with the OCT4 and SOX2 DNA binding
610 domains. We further restricted the list to those candidates that had a known, high-quality DNA
611 binding motif that appeared in either the DiffBind/MEME-ChIP or chromVAR analyses of our
612 data. We also added four TFs (OTX2, JUNB, ZSCAN23, and GSC) whose motifs appeared in our

613 ATAC-seq analyses but did not pass our differential expression cutoffs. We excluded ZEB1 and
614 ZEB2 as candidates because their size precluded delivery by lentivirus. We also eliminated 10 TFs
615 for which a clone was not readily accessible to us, either from the Harvard PlasmID database,
616 Addgene, or genes that had previously been cloned from cDNA in our lab. We note that one
617 candidate that was tested before all RNA-seq analysis was complete, MBNL2, missed significance
618 cutoffs in the final analysis but is nevertheless included for completeness. After adding three
619 candidates based on the literature (NRF2, ZNF521, and ID2), we were left with 36 candidates in
620 total.

621 **Epifluorescence imaging of fixed samples**

622 Samples were imaged on a Zeiss AxioObserver Z1 inverted microscope using Zeiss 10x and 20x
623 plan apo objectives (NA 1.3) using the appropriate filter sets. Images were acquired using an Orca-
624 Flash 4.0 CMOS camera (Hamamatsu). The 43 HE DsRed/46 HE YFP/47 HE CFP/49 DAPI/50
625 Cy5 filter sets from Zeiss were used. The microscope was controlled using the ZEN software.

626 **Live cell time lapse imaging**

627 Samples were imaged on a Zeiss AxioObserver Z1 inverted microscope using a Zeiss 20x plan
628 apo objective (NA 0.8) using the appropriate filter sets and a Hamamatsu ImagEM EMCCD
629 camera. Cells were maintained in a 37 degree incubation chamber at 5% CO_2. Cells were imaged
630 every 15 minutes. Focus was maintained using a combination of Zeiss Definite Focus and, using
631 a custom script in MicroManager 2.0 beta[40], software autofocus adjustments every hour to
632 compensate for slight movement of the membrane. For maximum accuracy, cells in this time lapse

633 were tracked manually, and the tracks were analyzed with a custom python script that performed
634 illumination profile correction and background subtraction.

635 **Confocal imaging**

636 For Figure 1 and S1, cells were imaged on a Leica inverted microscope with a Zeiss 20x objective
637 (NA 0.8) with the appropriate filter sets. Detection was performed with photomultiplier tubes (for
638 detection of Alexa 488 and Alexa 647) and a Leica HyD Photon Counter (for Alexa 568). For
639 Figure S6, cells were imaged on a Zeiss LSM 880 with Airyscan using a Zeiss 20x objective (NA
640 0.8). Detection was performed with photomultiplier tubes (Alexa 568 and Alexa 647) and a GaAsP
641 detector (CFP and Alexa 488).

642 **p(mesendoderm|OSR) curve fitting and location inference**

643 For the initial $p(\mathrm{mesendoderm}|\mathrm{OSR})$ curve fitting to the single cell data extracted from the time
644 lapse, we fit a two-parameter sigmoid function $\frac{1}{1+\exp(-a(x-b))}$ to the data using
645 scipy.optimize.curve_fit to minimize the squared difference between data and prediction. We used
646 the learned sigmoid shape parameter, a, for all subsequent $p(\mathrm{mesendoderm}|\mathrm{OSR})$ inference. To
647 infer $p(\mathrm{mesendoderm}|\mathrm{OSR})$ for a given population, we fit the location parameter, b, by
648 minimizing the squared difference between the observed final mesendoderm fate proportion and
649 the mesendoderm proportion predicted by $p(\mathrm{mesendoderm}|\mathrm{OSR})$ at varying locations, b, given
650 the observed $p(\mathrm{OSR}|t)$.

651 **Ethical compliance**

652 We used hESCs in accordance with approvals by Harvard University IRB (protocol #IRB18- 0665)
653 and Harvard University ESCRO (protocol E00065).

654 Data availability

655 The RNA-seq and ATAC-seq datasets generated in this study will be made available in the NCBI
656 GEO repository and the NCBI SRA by the time of publication, or in advance of that date upon
657 reasonable request. All other data are available upon reasonable request.

658 Code availability

659 All custom code used in this study is available from the authors upon reasonable request.

660 Author contributions

661 J.R.V. and S.R. conceived the project. J.R.V., R.H., S.K., and S.R. designed the experiments.
662 J.R.V., R.H., S.K., and D.V. conducted the experiments. J.R.V. and R.H. analyzed the data. J.R.V.,
663 R.H., S.K., D.V., R.E.K., and S.R. wrote the manuscript.

664 Acknowledgments

665 We thank the staff of the Bauer Core at Harvard University for their work on the RNA sequencing
666 used in this manuscript as well as for their expertise and assistance with flow cytometry and FACS.
667 We thank the Massachusetts General Hospital sequencing core for their work on the ATAC-seq
668 data used in this manuscript, Manashree Damle for bioinformatics help, the Harvard Physics/SEAS
669 Instructional Machine Shop for making the stainless-steel washers used in the live cell imaging in

670 this study, and the Harvard Center for Biological Imaging for the use of their equipment. We thank
671 Andrew Murray, Sean Eddy, and all of the members of the Ramanathan Lab for their helpful
672 comments. J.R.V. was funded by The Fannie and John Hertz Foundation, the National Science
673 Foundation Graduate Research Fellowship Program, and the NSF-Simons Center for
674 Mathematical and Statistical Analysis of Biology at Harvard, award #1764269. Some confocal
675 imaging was conducted on an instrument provided by the Harvard MRSEC (DMR-1420570). This
676 work was supported in part by NIH R01GM131105-01.

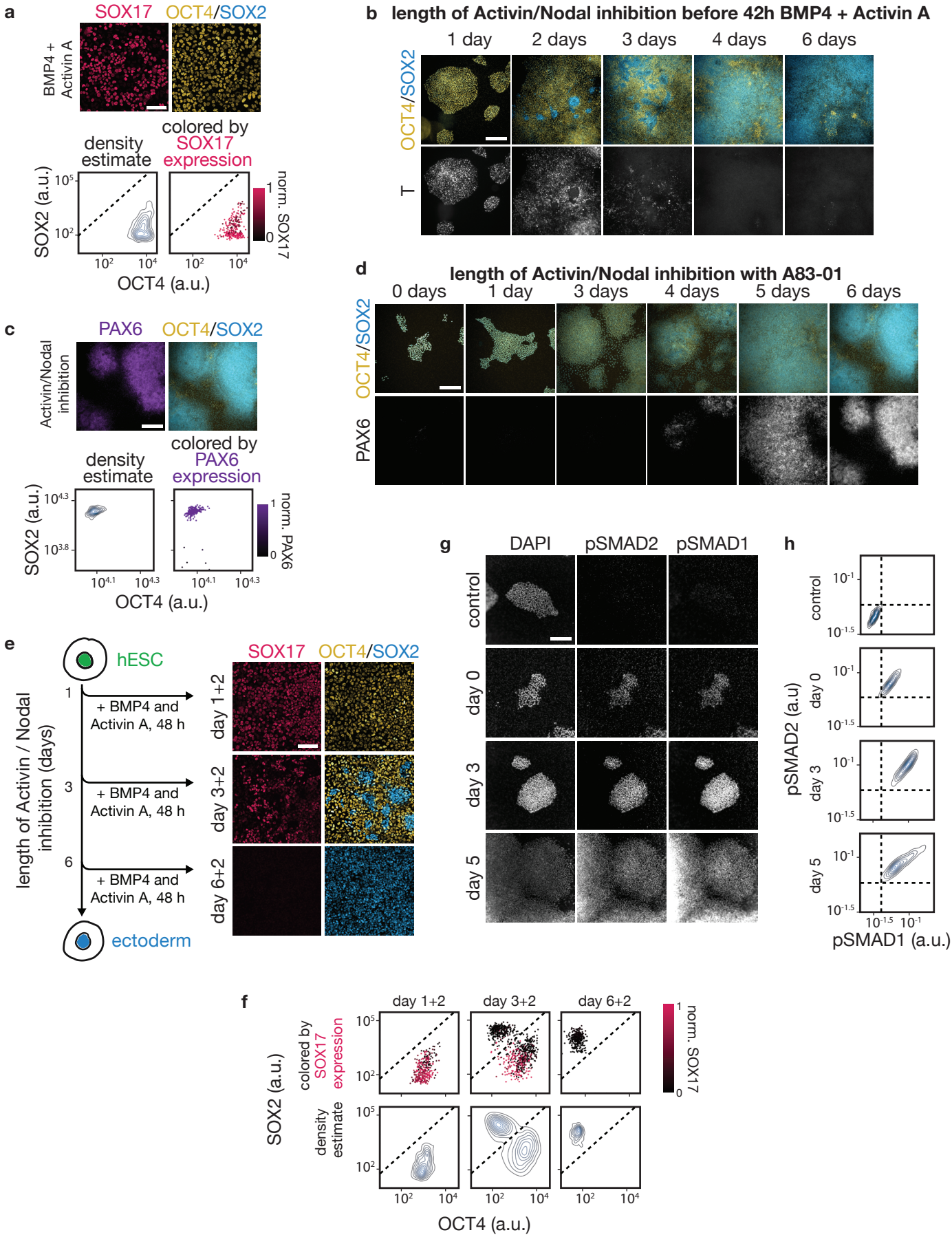

a
SOX17
OCT4/SOX2
BMP4 + Activin A
density estimate
colored by SOX17 expression
SOX2 (a.u.)
OCT4 (a.u.)
norm. SOX17
b
length of Activin/Nodal inhibition before 42h BMP4 + Activin A
1 day
2 days
3 days
4 days
6 days
OCT4/SOX2
T
c
PAX6
OCT4/SOX2
Activin/Nodal inhibition
density estimate
colored by PAX6 expression
SOX2 (a.u.)
OCT4 (a.u.)
norm. PAX6
d
length of Activin/Nodal inhibition with A83-01
0 days
1 day
3 days
4 days
5 days
6 days
OCT4/SOX2
PAX6
e
hESC
length of Activin / Nodal inhibition (days)
+ BMP4 and Activin A, 48 h
ectoderm
SOX17
OCT4/SOX2
day 1+2
day 3+2
day 6+2
f
day 1+2
day 3+2
day 6+2
colored by SOX17 expression
density estimate
SOX2 (a.u.)
OCT4 (a.u.)
norm. SOX17
g
DAPI
pSMAD2
pSMAD1
control
day 0
day 3
day 5
h
control
day 0
day 3
day 5
pSMAD2 (a.u)
pSMAD1 (a.u.)

Figure S1

Figure S1: Differentiating stem cells produce mesendoderm-derived and ectoderm-derived cell types while maintaining the ability to receive BMP4 and Activin signals

a, Top, confocal microscopy image of immunostained hESC colony after 48 hours of BMP4 + Activin A; SOX17 (pink), OCT4 (yellow), SOX2 (blue). These cells largely adopt $SOX17^+$ endoderm fates and are uniformly $OCT4^+$ $SOX2^-$. Scale bar = 100 µm. Bottom Left, contour plot of cell density as a function of OCT4 and SOX2 levels in individual cells; Bottom Right, same plot as left displayed as scatter plot with individual cells colored based on the level of endodermal marker SOX17. **b,** Images of hESC colonies immunostained for OCT4, SOX2, and mesendoderm marker T after 1-6 days of Activin/NODAL inhibition followed by 42h of BMP4 + Activin A stimulation. T^- $OCT4^-$ $SOX2^+$ ectoderm cells first appear at 2 days and are widespread by 4 days of Activin/NODAL inhibition. The spatial structure seen here largely appears after BMP4 and Activin A signal induction and is likely due to a combination of local density impacts on ectoderm-directed differentiation rate and homophilic interactions between cells. Scale bar = 300 µm. **c,** Top, epifluorescence microscopy image of an hESC colony immunostained for the neurectoderm marker PAX6 (purple), OCT4 (yellow), and SOX2 (blue) after 6 days of Activin/NODAL inhibition. These cells largely adopt $PAX6^+$ $OCT4^-$ $SOX2^+$ neurectoderm fates. Below, contour plot of cell density as a function of OCT4 and SOX2 levels in individual cells (left), same plot as left displayed as scatter plot with individual cells colored based on the level of PAX6 (right). Scale bar = 300 µm. **d,** Images of hESC colonies immunostained for OCT4, SOX2, and the neurectoderm marker PAX6 after 0-6 days of Activin/NODAL inhibition. PAX6 expression is first seen after 4 days and is widespread by 5 days. Scale bar = 300 µm. **e,** hESCs were exposed to a pretreatment of Activin/NODAL inhibition for 1, 3, or 6 days before BMP4 + Activin A signal exposure (left). Confocal microscopy images of immunostained hESCs at the end of the experiment after 48h of BMP4 + Activin A treatment (right). Increasing duration of Activin/NODAL inhibition reduced the population's competence to produce $SOX17^+$ endoderm and, more broadly, $OCT4^+$ (yellow) $SOX2^-$ (blue) mesendoderm-derived cell types. Scale bar = 100 µm. **f,** Scatter plot (top) and density contour plot (bottom) of cells from E, as a function of their OCT4 and SOX2 levels. Cells in scatter plot colored by level of SOX17. **g,** Immunofluorescence images acquired via epifluorescence microscopy showing DAPI, phosphorylated SMAD1, and phosphorylated SMAD2 levels in hESC colonies upon stimulation with BMP4 + Activin for 45 mins after 0, 3, or 5 days of Activin/NODAL inhibition. Unstimulated control shown in top row. All stimulated cells at all time points displayed nuclear-localized pSMAD1 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 treated as in G, as a function of pSMAD2/3 and pSMAD1/5 levels determined by antibody staining. Dotted lines drawn for reference.

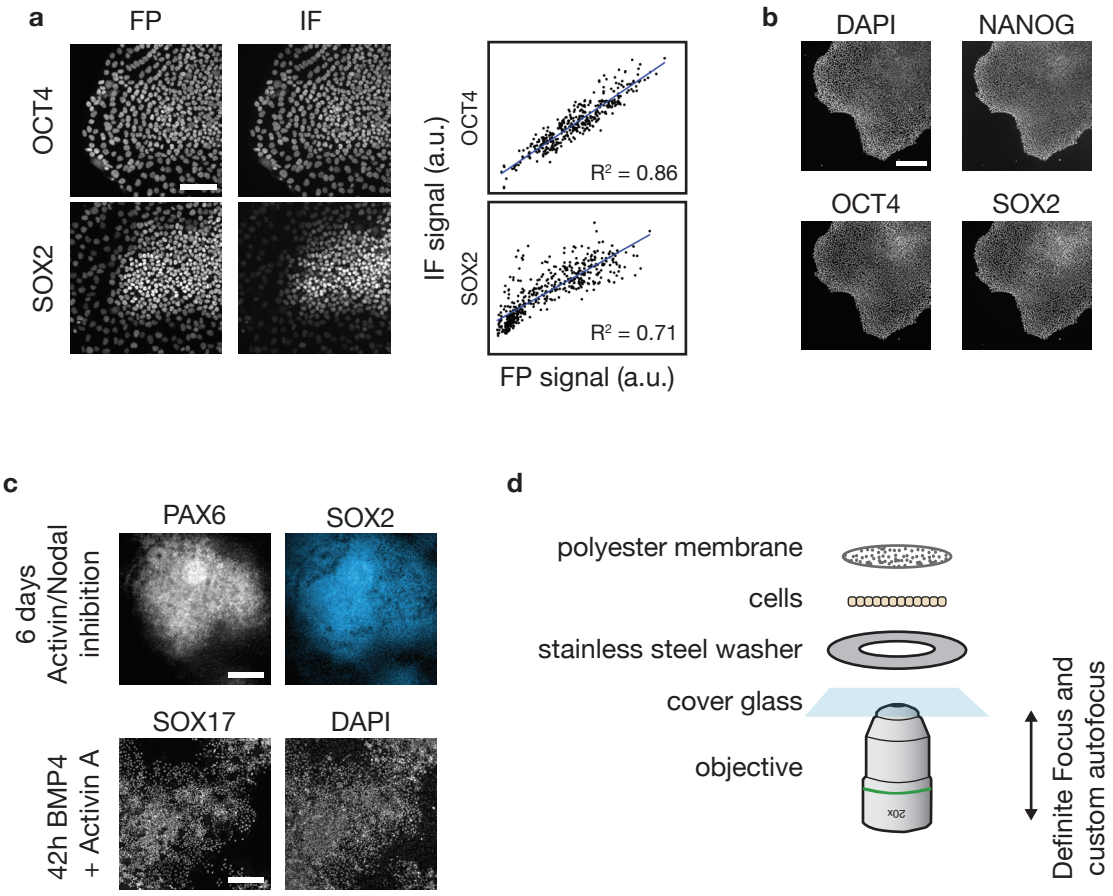

a
FP
IF
OCT4
SOX2
IF signal (a.u.)
OCT4
SOX2
R² = 0.86
R² = 0.71
FP signal (a.u.)
b
DAPI
NANOG
OCT4
SOX2
c
PAX6
SOX2
6 days
Activin/Nodal
inhibition
SOX17
DAPI
42h BMP4
+ Activin A
d
polyester membrane
cells
stainless steel washer
cover glass
objective
20x
Definite Focus and
custom autofocus

Figure S2

Figure S2: Monitoring OCT4:RFP and SOX2:YFP accurately reflects OCT4 and SOX2 levels without perturbing cell potential or behavior

a, OCT4 and SOX2 levels as determined by immunofluorescence (IF) and by OCT4:RFP and SOX2:YFP fluorescence (FP). Scale bar = 100 µm. Blue line, best-fit linear regression. **b,** Immunofluoresence images of H1 OCT4:RFP SOX2:YFP double reporter cells stained for NANOG, OCT4, and SOX2. Nuclei (DAPI) are shown for reference. These cells can maintain pluripotency indefinitely. Scale bar = 300 µm. **c,** Immunofluorescence images of H1 OCT4:RFP SOX2:YFP double reporter cells that were (top) stained for PAX6 and SOX2 after 6 days 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 line's ability to form the germ layers is unaffected. Scale bar = 300 µm. **d,** 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.

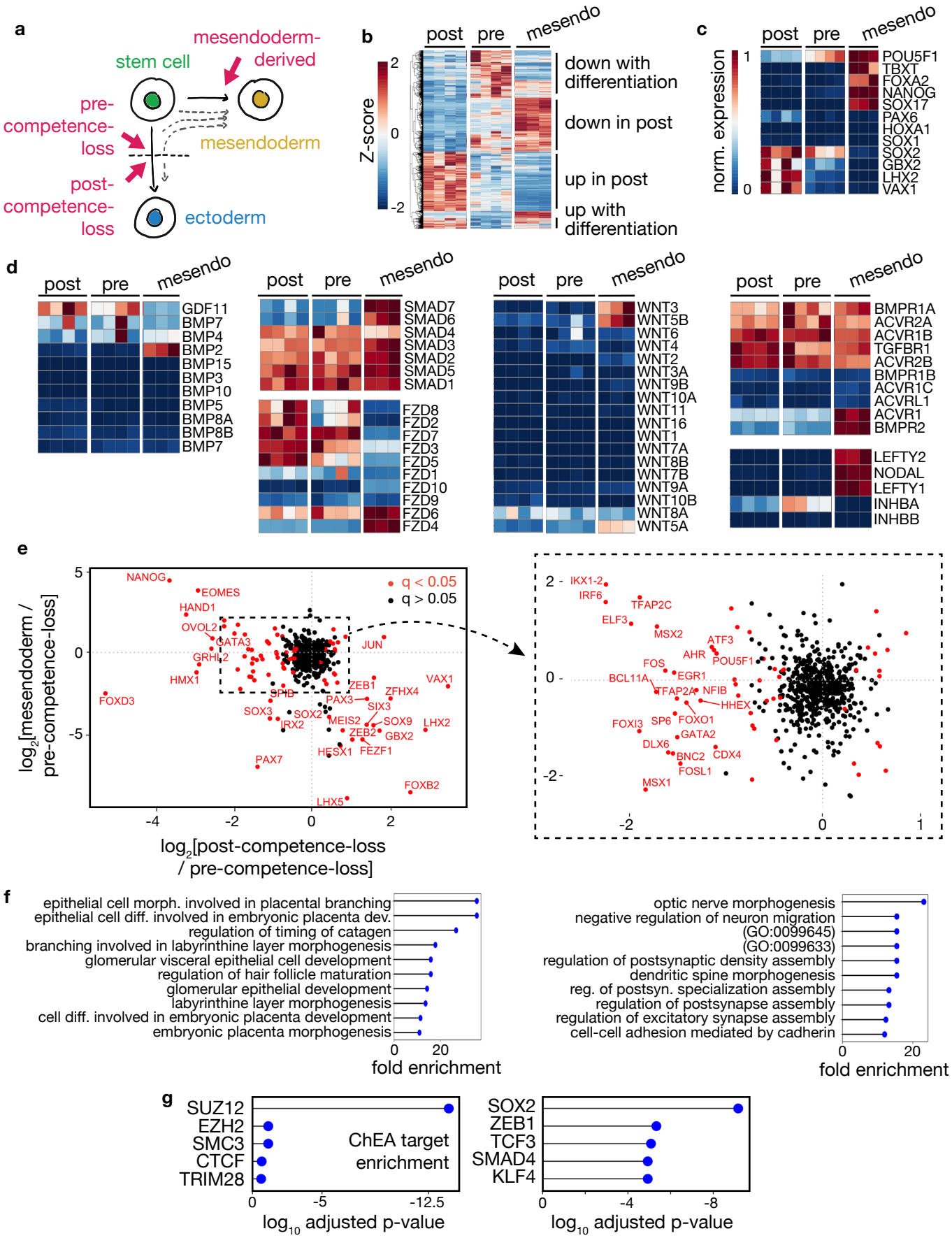

a
mesendoderm-derived
stem cell
pre-competence-loss
mesendoderm
post-competence-loss
ectoderm
b
post
pre
mesendo
Z-score
down with differentiation
down in post
up in post
up with differentiation
c
norm. expression
POU5F1
TBXT
FOXA2
NANOG
SOX17
PAX6
HOXA1
SOX1
SOX2
GBX2
LHX2
VAX1
d
GDF11
BMP7
BMP4
BMP2
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
q < 0.05
q > 0.05
log2[mesendoderm / pre-competence-loss]
log2[post-competence-loss / pre-competence-loss]
NANOG
EOMES
HAND1
OVOL2
GATA3
GRHL2
HMX1
FOXD3
JUN
VAX1
ZEB1
ZFHX4
SPIB
PAX3
SIX3
SOX3
SOX2
IRX2
MEIS2
SOX9
LHX2
ZEB2
GBX2
HESX1
FEZF1
PAX7
FOXB2
LHX5
IKX1-2
IRF6
TFAP2C
ELF3
MSX2
ATF3
AHR
POU5F1
FOS
EGR1
BCL11A
TFAP2A
NFIB
HHEX
FOXI3
SP6
FOXO1
GATA2
DLX6
BNC2
CDX4
FOSL1
MSX1
f
epithelial cell morph. involved in placental branching
epithelial cell diff. involved in embryonic placenta dev.
regulation of timing of catagen
branching involved in labyrinthine layer morphogenesis
glomerular visceral epithelial cell development
regulation of hair follicle maturation
glomerular epithelial development
labyrinthine layer morphogenesis
cell diff. involved in embryonic placenta development
embryonic placenta morphogenesis
fold enrichment
optic nerve morphogenesis
negative regulation of neuron migration
(GO:0099645)
(GO:0099633)
regulation of postsynaptic density assembly
dendritic spine morphogenesis
reg. of postsyn. specialization assembly
regulation of postsynapse assembly
regulation of excitatory synapse assembly
cell-cell adhesion mediated by cadherin
fold enrichment
g
SUZ12
EZH2
SMC3
CTCF
TRIM28
ChEA target enrichment
log10 adjusted p-value
SOX2
ZEB1
TCF3
SMAD4
KLF4
log10 adjusted p-value

Figure S3

Figure S3: Analysis of gene expression changes that coincide with mesendoderm competence loss

a, Based on the OCT4:RFP to SOX2:YFP ratio, pre- and post-competence-loss cells were isolated using FACS for RNA-seq and ATAC-seq. Mesendoderm-derived cells (hESCs stimulated with 40 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 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,** Scatterplot showing log2 fold change 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,** Ten most significantly enriched GO terms for gene expression sets from this study as calculated by PANTHER. Left, enriched terms for genes that are downregulated upon competence loss in a lineage-specific pattern. Right, same for genes that are upregulated. **g,** 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.

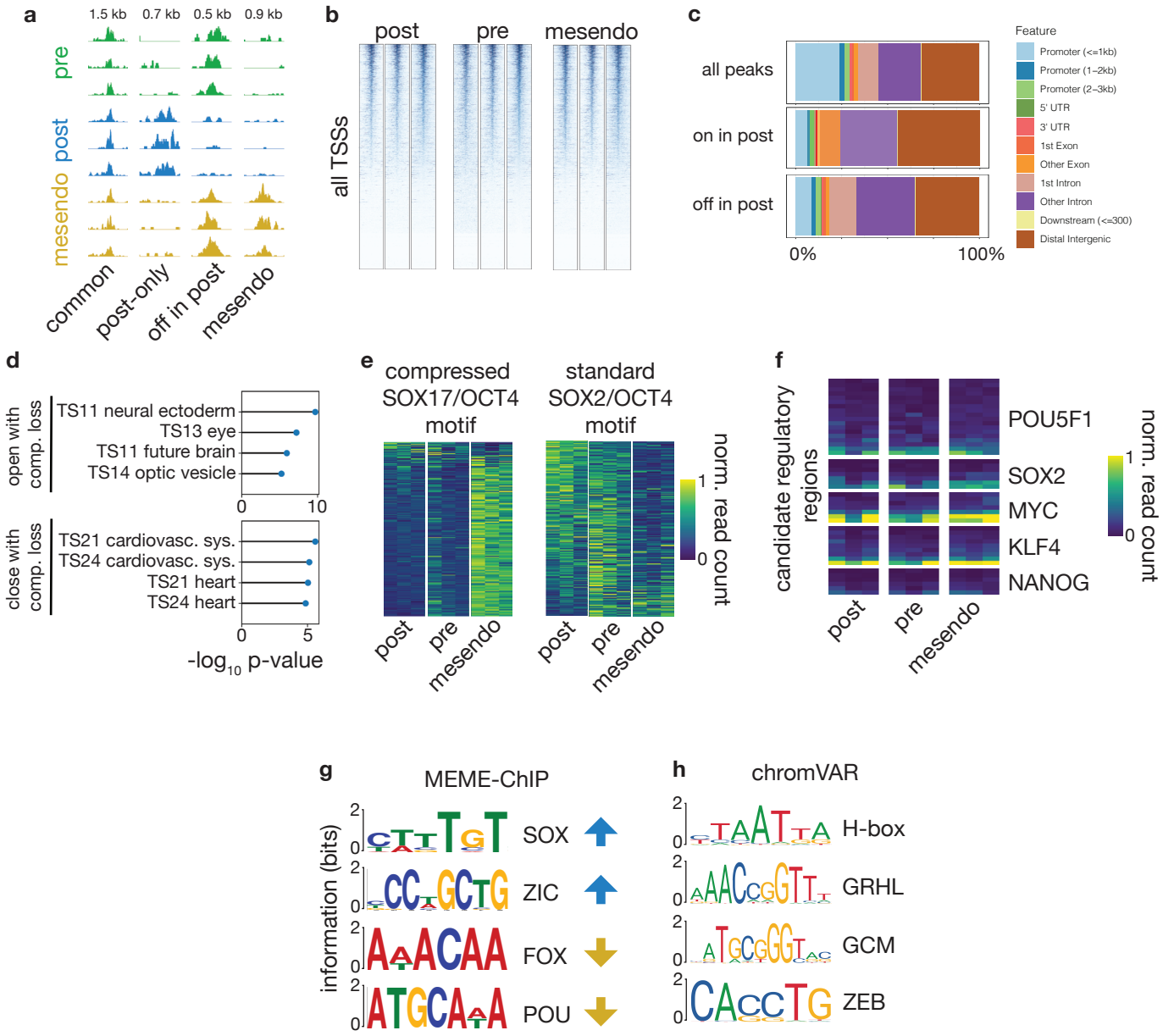
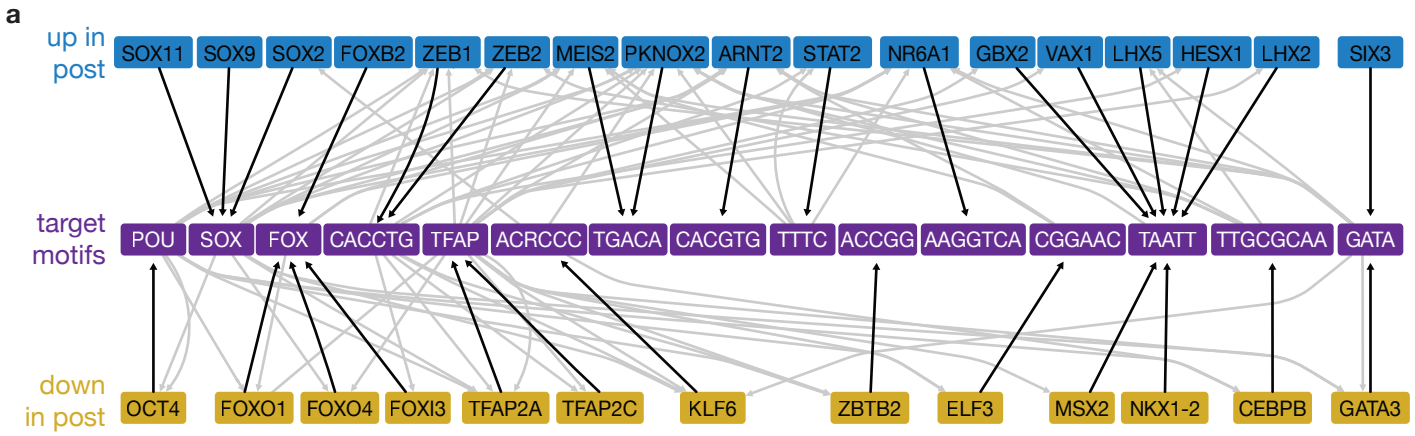
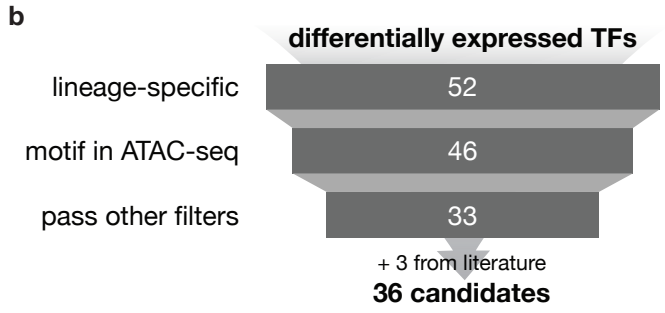

a
1.5 kb
0.7 kb
0.5 kb
0.9 kb
pre
post
mesendo
common
post-only
off in post
mesendo
b
post
pre
mesendo
all TSSs
c
all peaks
on in post
off in post
0%
100%
Feature
Promoter (<=1kb)
Promoter (1–2kb)
Promoter (2–3kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=300)
Distal Intergenic
d
open with comp. loss
TS11 neural ectoderm
TS13 eye
TS11 future brain
TS14 optic vesicle
0
10
close with comp. loss
TS21 cardiovasc. sys.
TS24 cardiovasc. sys.
TS21 heart
TS24 heart
0
5
-log10 p-value
e
compressed SOX17/OCT4 motif
standard SOX2/OCT4 motif
norm. read count
1
0
post
pre
mesendo
post
pre
mesendo
f
candidate regulatory regions
POU5F1
SOX2
MYC
KLF4
NANOG
norm. read count
1
0
post
pre
mesendo
g
MEME-ChIP
information (bits)
SOX
ZIC
FOX
POU
h
chromVAR
H-box
GRHL
GCM
ZEB

Figure S4

Figure S4: Analysis of chromatin accessibility changes that coincide with mesendoderm competence loss

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,** Selected motifs with significant enrichment in genomic regions that had a significant increase (blue arrows) or decrease (yellow arrows) from the pre-competence-loss to the post-competence-loss populations as discovered using MEME-ChIP. Motifs are labeled by the TF family/group to which they correspond. E-values: SOX, 3.7e-128; ZIC, 8.3e-033; FOX, 1.8e-02; POU, 2.9e-003. **h,** Top 4 nonredundant motifs that best explain the observed changes in ATAC-seq read depth between pre-competence-loss and post-competence-loss populations, as calculated using chromVAR. p-values: homeobox (VAX2), 9.8E-05; GCM (GCM1), 3.2E-4; GRHL (GRHL1), 6.6E-4; ZEB (ZEB2), 6.9E-4.





c **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 S5

Figure S5: Gene expression and chromatin accessibility signatures reveal a core set of TFs that, when perturbed, could modulate competence

a, 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. **b,** Schematic depicting the candidate selection process (see methods for full details). **c,** Full list of the 36 TFs identified as candidates for perturbation to affect mesendoderm competence

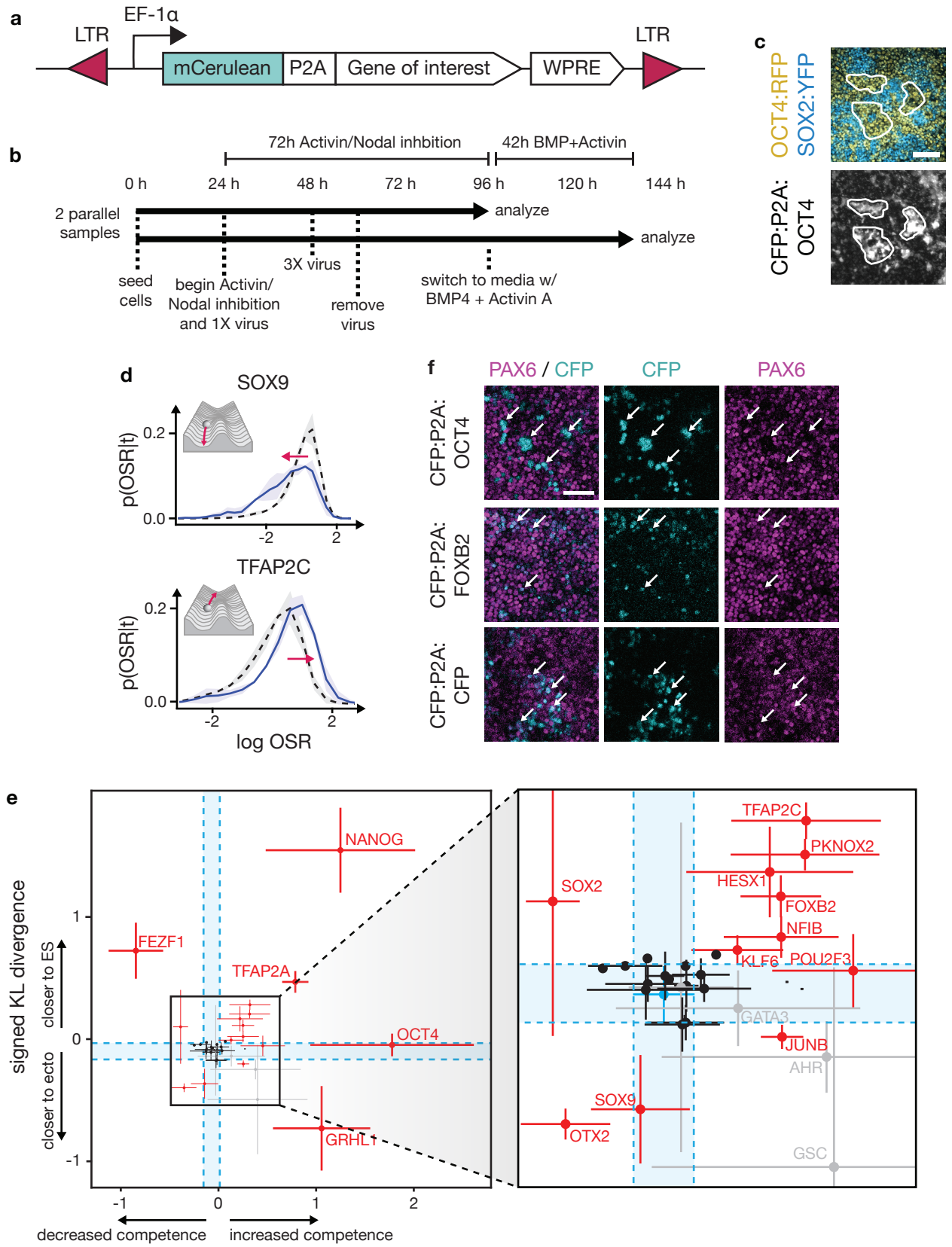

a
LTR
EF-1α
mCerulean
P2A
Gene of interest
WPRE
LTR
b
72h Activin/Nodal inhbition
42h BMP+Activin
0 h
24 h
48 h
72 h
96 h
120 h
144 h
2 parallel samples
analyze
analyze
seed cells
begin Activin/ Nodal inhibition and 1X virus
3X virus
remove virus
switch to media w/ BMP4 + Activin A
c
OCT4:RFP
SOX2:YFP
CFP:P2A: OCT4
d
SOX9
p(OSR|t)
0.2
0.0
-2
2
TFAP2C
log OSR
f
PAX6 / CFP
CFP
PAX6
CFP:P2A: OCT4
CFP:P2A: FOXB2
CFP:P2A: CFP
e
signed KL divergence
closer to ES
closer to ecto
NANOG
FEZF1
TFAP2A
OCT4
GRHL1
SOX2
TFAP2C
PKNOX2
HESX1
FOXB2
NFIB
KLF6
POU2F3
GATA3
JUNB
AHR
SOX9
OTX2
GSC
decreased competence
increased competence
shift in location of p(mesendoderm|OSR) = 0.5 from non-transduced to transduced populations (log2 OSR)

Figure S6

Figure S6: Analysis of lineage bias and differentiation dynamics in cells perturbed by candidate TFs

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 experimental 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,** Epifluorescent microscopy images of a cell colony after Activin/Nodal inhibition and subsequent BMP4 + Activin A stimulation. Cells that were transduced with a CFP:P2A:OCT4 expression cassette expressed CFP and preferentially adopted a OCT4:RFP^{+} SOX2:YFP^{-} mesendoderm fate. White regions indicate selected clusters of cells expressing CFP, indicating transduction. Scale bar = 100 µm. **d,** Examples of two candidates, SOX9 and TFAP2C, whose overexpression altered the distribution of OCT4:RFP/SOX2:YFP ratios 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. Inset, a schematic depicting the corresponding effects on the Waddington developmental landscape. **e,** 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 ratio of OCT4:RFP to SOX2:YFP 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 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, as evidenced by the location of the OCT4 point near zero on the y-axis. **f,** Immunofluorescence images showing PAX6 (magenta) and CFP (cyan) after 6 days 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.