# 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

**As pluripotent human embryonic stem cells progress along the developmental trajectory towards one differentiated fate, they lose competence to adopt other fates. Here, we show that the loss of competence for alternative fates occurs at a specific point along the developmental trajectory and that this point can be moved using genetic perturbations. We first show that competence to adopt mesendoderm-derived fates in response to BMP4 and Activin A signal exposure is lost during ectoderm-directed differentiation. By monitoring each cell's progression along its developmental trajectory, we can prospectively predict that cell's mesendoderm competence. We then exploit this predictive ability using RNA-seq and ATAC-seq to identify and validate candidate transcription factors that can modulate mesendoderm competence. These factors exert their effects by controlling the cell's**

**progression along the developmental trajectory, by tuning its competence to form mesendoderm at any given point along that trajectory, or by altering both of these aspects. In the classical picture of a Waddington landscape, these effects correspond to changing the cell's location on the landscape and altering the location of the barrier between fates, 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.**

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 external signaling milieu[5–7]. If we understood the underlying dynamical system that governs this changing response to a morphogenetic signal, we should be able to predict a cell's competence to adopt alternative fates and, further, to modulate this competence. We asked whether we can achieve these two goals using the *in vitro* germ layer differentiation of pluripotent human embryonic stem cells (hESCs) as a model system.

Given the appropriate signals, hESCs adopt mesendodermal or ectodermal progenitor fates that can each, in turn, produce a multitude of specialized cell types. *In vivo* and *in vitro*, the mesendoderm fate is induced by BMP and Activin/NODAL signals. The recent discovery that the receptors for these signals are basolaterally localized in epithelial stem cell populations[10,11] led us

to grow hESCs on permeable polyester membranes so that the ligands were able to access the receptors of all cells (see methods). Under these conditions, all hESCs in the colony uniformly produced cells expressing BRACHYURY, SOX17, and other markers associated with mesendoderm-derived cell types upon stimulation with BMP4 and Activin A (Figure 1A, S1A, S1B). In contrast, inhibiting Activin/NODAL signaling promoted ectoderm-derived fates[12], ultimately producing PAX6$^{+}$ neurectoderm after 4-5 days (Figure 1A, S1C, S1D).

To characterize the cells' competence to adopt the alternative fate, we measured how the fraction of cells that become mesendoderm changed as a function of the time for which they were first differentiated towards the ectoderm lineage. Consistent with previous data from mouse[8,9], we demonstrated that competence to adopt mesendoderm-derived fates decreases at the population level as cells differentiate towards the ectodermal fate. We found that increasing the duration of differentiation towards ectoderm before exposing the cells to BMP4 and Activin A signal lowered the fraction of cells that adopt mesendodermal fates (Figure 1B, S1B, S1E, S1F). The temporal decrease in the mesendoderm fraction occurred despite the cells' continued ability to respond to BMP and Activin signals throughout this period (Figure S1G, S1H). This decreasing fraction suggested a decreasing competence to adopt mesendodermal fates with time and, further, that this competence is heterogeneous within the differentiating population. We therefore sought to measure and predict the competence of individual cells.

Our approach to prospectively predicting an individual cell's mesendoderm competence was motivated by the notion of a Waddington-inspired developmental landscape (Figure 1C). In this qualitative picture, a barrier between the ectodermal and mesendodermal developmental

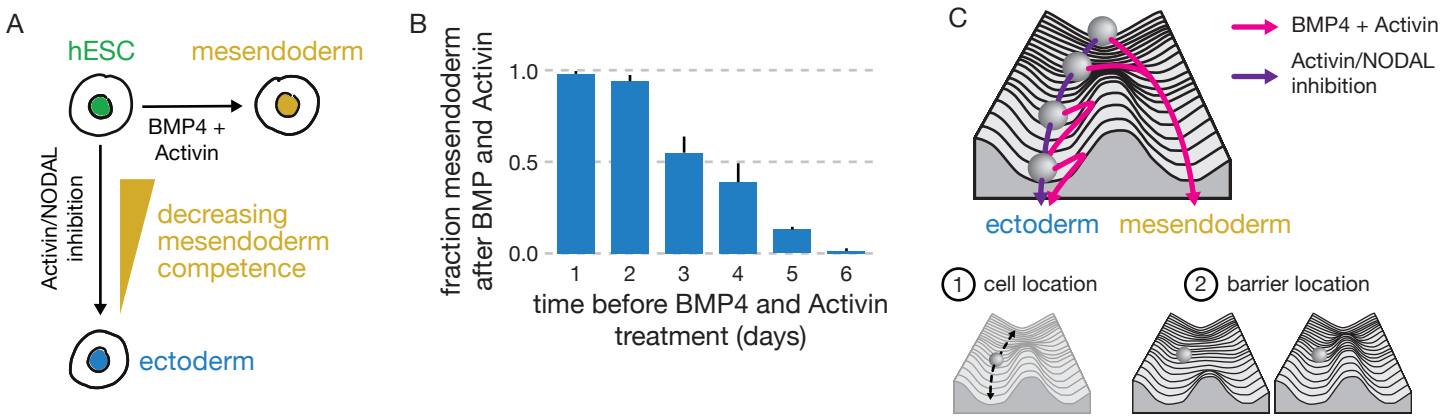
A
hESC
mesendoderm
BMP4 +
Activin
Activin/NODAL
inhibition
decreasing
mesendoderm
competence
ectoderm
B
fraction mesendoderm
after BMP and Activin
1.0
0.5
0.0
1
2
3
4
5
6
time before BMP4 and Activin
treatment (days)
C
BMP4 + Activin
Activin/NODAL
inhibition
ectoderm
mesendoderm
1 cell location
2 barrier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

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) hESCs produce mesendoderm-derived lineages in response to BMP4 + Activin A, and they produce ectoderm-derived lineages in response to Activin/NODAL inhibition. At the population level, mesendoderm competence decreases gradually with increasing duration of ectoderm-directed differentiation.

(B) Quantification of confocal microscopy images of immunostained hESC colonies after the indicated duration of Activin/NODAL inhibition using 0.5 μM A83-01 followed by two days of stimulation with 3 ng/mL BMP4 and 100 ng/mL Activin A. Cells were stained for OCT4 and SOX2. Each bar shows the mean fraction of cells that adopt an $OCT4^{+}$ $SOX2^{-}$ mesendoderm-derived fate given the time at which BMP4 and Activin A was added. Error bars represent the standard deviation across multiple fields of view. Increasing duration of Activin/NODAL inhibition reduced the probability of a cell producing $OCT4^{+}$ $SOX2^{-}$ mesendoderm-derived cell types after subsequent BMP4 and Activin A stimulation.

(C) Schematic of a Waddington landscape that represents the ectoderm/mesendoderm fate choice. In this picture, the competence of a cell to produce mesendoderm depends both on the cell's location along the developmental trajectory and on the position of the barrier between the ectoderm and mesendoderm fates.

trajectories increases in height as the cells progress along the ectodermal trajectory. A quantitative implication of such a picture is that the probability that a cell can transition to a mesendodermal fate decreases as the cell proceeds down the ectodermal trajectory. To achieve the goal of predicting the competence of a cell, we sought to quantify this probability of the cell adopting the mesendodermal fate in the presence of the appropriate signal as a function of the cell's location along the ectodermal trajectory.

To predict and test the competence of individual cells to become mesendoderm, we needed to measure each live cell's location along the developmental trajectory in time. Spectacular success in understanding complex physical systems has been achieved through identifying relevant order parameters[13–16]: low-dimensional variables whose dynamics reflect the broken symmetry associated with a state transition. Our previous results in mouse showed that the protein levels of the pluripotency factors Oct4 and Sox2 together constitute such an order parameter, and their dynamics reflect the transitions of pluripotent cells to the mesendoderm or ectoderm fates[17–19]. In humans as in mouse, we validated that both OCT4 and SOX2 are symmetrically highly expressed in the pluripotent stem cell, but they are asymmetrically downregulated in the two lineages. OCT4 expression is maintained in the mesendoderm while SOX2 is downregulated; in contrast, ectoderm differentiation involves SOX2 maintenance and OCT4 downregulation. Both TFs are also functionally important for these state transitions: OCT4 downregulation is necessary for neurectoderm induction[17,20], while SOX2 downregulation is required for mesendoderm fate selection[17]. Furthermore, direct conversion to a neural fate silences OCT4[17], underscoring the fundamental nature of this order parameter to the fate decision in question. We therefore used

OCT4 and SOX2 to monitor progression along the developmental trajectory from pluripotency to mesendoderm- or ectoderm-derived fates.

We demonstrated that the developmental trajectories of hESCs in OCT4 and SOX2 space were predictive of the germ layer fate choice of cells, and they allowed such prediction days before the expression of classical master regulators of these fates. To show this, we employed our validated hESC line in which one allele each of OCT4 and SOX2 had been replaced with OCT4:RFP and SOX2:YFP, respectively, at the endogenous locus[11] (Figure S2A-C). Using flow cytometry, we could follow the developmental trajectories of a population of hESCs (with high levels of OCT4 and SOX2) into ectodermal progenitors over the course of six days as they downregulated OCT4. When BMP4 and Activin A signals were added at an intermediate stage of differentiation (Figure 2A), we could visualize a bifurcation of developmental trajectories: the mesendoderm-competent cells adopted $OCT4^{+}$ $SOX2^{-}$ mesendoderm-derived fates, while the cells that were not mesendoderm-competent proceeded towards $OCT4^{-}$ $SOX2^{+}$ ectoderm-derived fates.

We next sought to measure the probability of an individual cell adopting a mesendoderm-derived fate given the cell's location along the ectodermal trajectory. To do so, we performed a time lapse experiment using the OCT4:RFP SOX2:YFP hESC line (Figure 2B). To monitor the cells throughout this process, we developed and deployed a custom live-cell microscopy setup that was capable of imaging cells on the flexible membrane every 15 minutes for over five days (Figure S2D). Based on the timing of mesendoderm competence loss in our flow cytometry experiments, we first differentiated the pluripotent stem cell population in this apparatus for 2.25 days in ectodermal differentiation conditions to obtain a heterogeneous population in which some cells

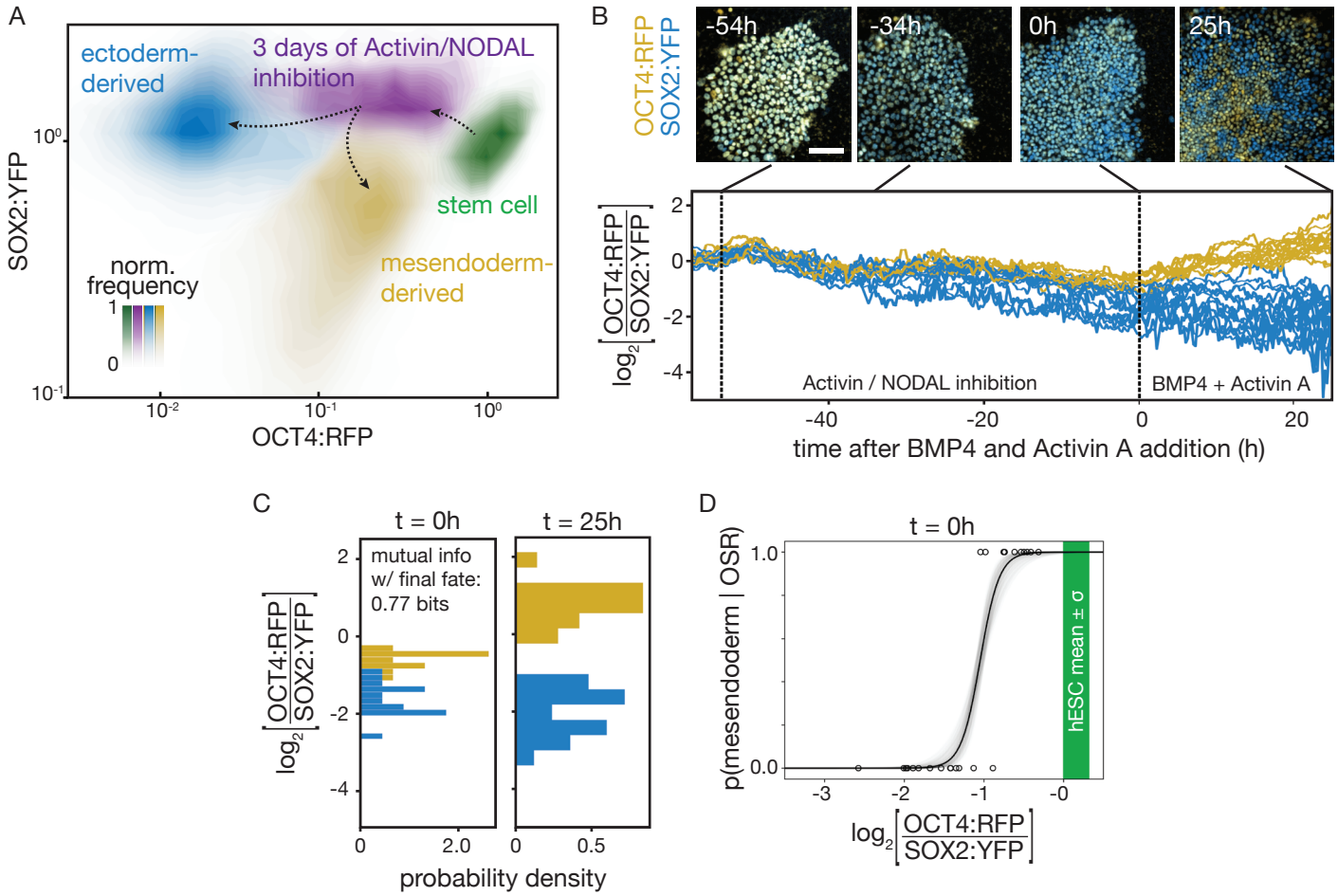
A
ectoderm-derived
3 days of Activin/NODAL inhibition
stem cell
mesendoderm-derived
norm. frequency
1
0
SOX2:YFP
OCT4:RFP
B
OCT4:RFP
SOX2:YFP
-54h
-34h
0h
25h
log2 [OCT4:RFP / SOX2:YFP]
Activin / NODAL inhibition
BMP4 + Activin A
time after BMP4 and Activin A addition (h)
C
t = 0h
t = 25h
mutual info w/ final fate: 0.77 bits
log2 [OCT4:RFP / SOX2:YFP]
probability density
D
t = 0h
p(mesendoderm | OSR)
hESC mean ± σ
log2 [OCT4:RFP / SOX2:YFP]

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 seen in the hESC population. hESCs (green) downregulate OCT4:RFP after 3 days of Activin/NODAL inhibition (purple). After a subsequent two days of BMP4 + Activin A stimulation, the cells in this purple population bifurcate: those that are still competent to choose a mesendodermal fate (yellow) do so, while cells that are no longer mesendoderm-competent continue to the ectodermal (blue) fates.

(B) Top, snapshots from a time-lapse microscopy experiment of a field of hESCs showing endogenous OCT4:RFP (yellow) and SOX2:YFP (blue). Cells were initially maintained in pluripotency conditions, Activin/NODAL inhibition was started at t = -54 h, and BMP4 + Activin A stimulation began at t = 0 h. Scale bar = 100 µm. Bottom, the ratio of the OCT4:RFP signal to SOX2:YFP signal for individual cells through the 80 hour time course is plotted as a function of time with t = 0 h marking the time of BMP4 + Activin stimulation. Cells that adopt an $OCT4^+$ $SOX2^-$ mesendodermal fate at the end of the time course are shown in yellow, and those that adopt an $OCT4^-$ $SOX2^+$ ectodermal fate are colored blue. At the moment of signal stimulation (t = 0 h), the ratio of OCT4:RFP to SOX2:YFP is predictive of the eventual fate.

(C) The distribution of the logarithm of the ratio of OCT4:RFP to SOX2:YFP signal is shown at the end of the experiment when cells have chosen their fate (right) and 25 hours earlier at the moment of BMP4 + Activin A exposure (t = 0h, left). Cells adopting an eventual $OCT4^+$ $SOX2^-$ mesendodermal fate at the end of the time course are in yellow, and those that adopt an $OCT4^-$ $SOX2^+$ ectodermal fate are in blue. The mutual information between the OCT4:RFP to SOX2:YFP ratio at the moment of signal addition and the final fate is 0.77 bits.

(D) Estimation of the probability of adopting a mesendoderm-derived fate given the OCT4:RFP/SOX2:YFP ratio, p(mesendoderm | OSR). The best-fit sigmoid function is drawn in black. Transparent gray curves represent 1000 sigmoid fits sampled randomly according to the covariance matrix of the fit parameters in which both parameters fall within one standard deviation of the best-fit values. The green region represents the mean value of pluripotent stem cells plus or minus one standard deviation. Measuring the OCT4:RFP/SOX2:YFP ratio reveals a sharp transition from a high probability of mesendoderm competence to a low probability as the ratio decreases.

had already lost mesendoderm competence and some had not. We then added BMP4 and Activin A signals for 25 hours, prompting mesendoderm-competent cells to adopt mesendoderm fates and non-mesendoderm-competent cells to adopt ectodermal fates.

Using our time lapse data, we next showed that measuring each cell's OCT4:RFP and SOX2:YFP levels allowed us to predict that cell's mesendoderm competence. We tracked individual cells from pluripotency through ectoderm-directed differentiation and subsequent BMP4 + Activin A signal (Figure 2B). In this OCT4:RFP/SOX2:YFP space, the trajectories of cells in pluripotency conditions were tightly localized. In ectoderm-promoting conditions, some cells differentiated towards ectoderm faster than others as measured by the rate of OCT4:RFP downregulation. Importantly, we found that, rather than the levels of the individual proteins, the OCT4:RFP to SOX2:YFP fluorescence ratio at the moment of Activin/Nodal signal addition was predictive of the ultimate fate of the cells with high accuracy (Figure 2C). Each cell carried 0.77 bits of mutual information about its mesendoderm competence state in its ratio of OCT4:RFP to SOX2:YFP at the moment of signal addition; we note that 1 bit would represent perfect information about the final state and is thus the maximum possible value. Cells with a high ratio of OCT4:RFP to SOX2:YFP were able to become mesendoderm in response to the BMP4 and Activin A signal, while cells with a low ratio of OCT4:RFP to SOX2:YFP were not. We then computed the probability of a cell adopting a mesendodermal fate given OSR, $p(\mathrm{mesendoderm}|\mathrm{OSR})$, where OSR is that cell's ratio of OCT4:RFP to SOX2:YFP after normalizing OCT4:RFP and SOX2:YFP intensities to the mean levels measured in hESCs in pluripotency conditions (Figure 2D). This probability had a sharp transition from 1 to 0, demonstrating that there was a defined point along

the developmental trajectory defined by the ratio of OCT4:RFP to SOX2:YFP at which cells lose their ability to become mesendoderm even when exposed to the relevant signals.

Having computed $p(\text{mesendoderm}|\text{OSR})$ for single cells in our time lapse, we sought a way to use this knowledge to predict the mesendoderm competence of cells in a heterogeneous differentiating population. Since cells in a population move along the developmental trajectory at differing rates, at any given time, $t$, the cells have a distribution of OCT4:RFP to SOX2:YFP ratios, $p(\text{OSR}|t)$. The fraction of the overall cell population that adopts a mesendoderm fate after BMP4 and Activin stimulation should be determined by the fraction of the cells with a given OCT4:RFP to SOX2:YFP ratio multiplied by the probability that cells with that ratio will become mesendoderm; that is, $p(\text{mesendoderm}|\text{t}) = \sum_{\text{OSR}} p(\text{mesendoderm}|\text{OSR}) \times p(\text{OSR}|\text{t})$ (Figure 3A). We note that this statement provides a mathematical formalization of the intuition that each cell's response to signal depends on its location on the developmental trajectory, which corresponds to $p(\text{OSR}|t)$, and the probability that the cell adopts a mesendodermal fate when exposed to signal at that location, $p(\text{mesendoderm}|\text{OSR})$. This changing probability of mesendoderm adoption can be pictured as a changing barrier between fates on a developmental landscape.

We next reasoned that we should be able to use our ability to predict mesendoderm competence to prospectively isolate competent from non-competent cells from a single differentiating population. To this end, we measured $p(\text{OSR}|t)$ of a population of differentiating hESCs using flow cytometry. Our analysis suggested that, as cells moved towards a lower OCT4:RFP/SOX2:YFP ratio as they differentiated to become ectoderm, the cells in the region where our computed

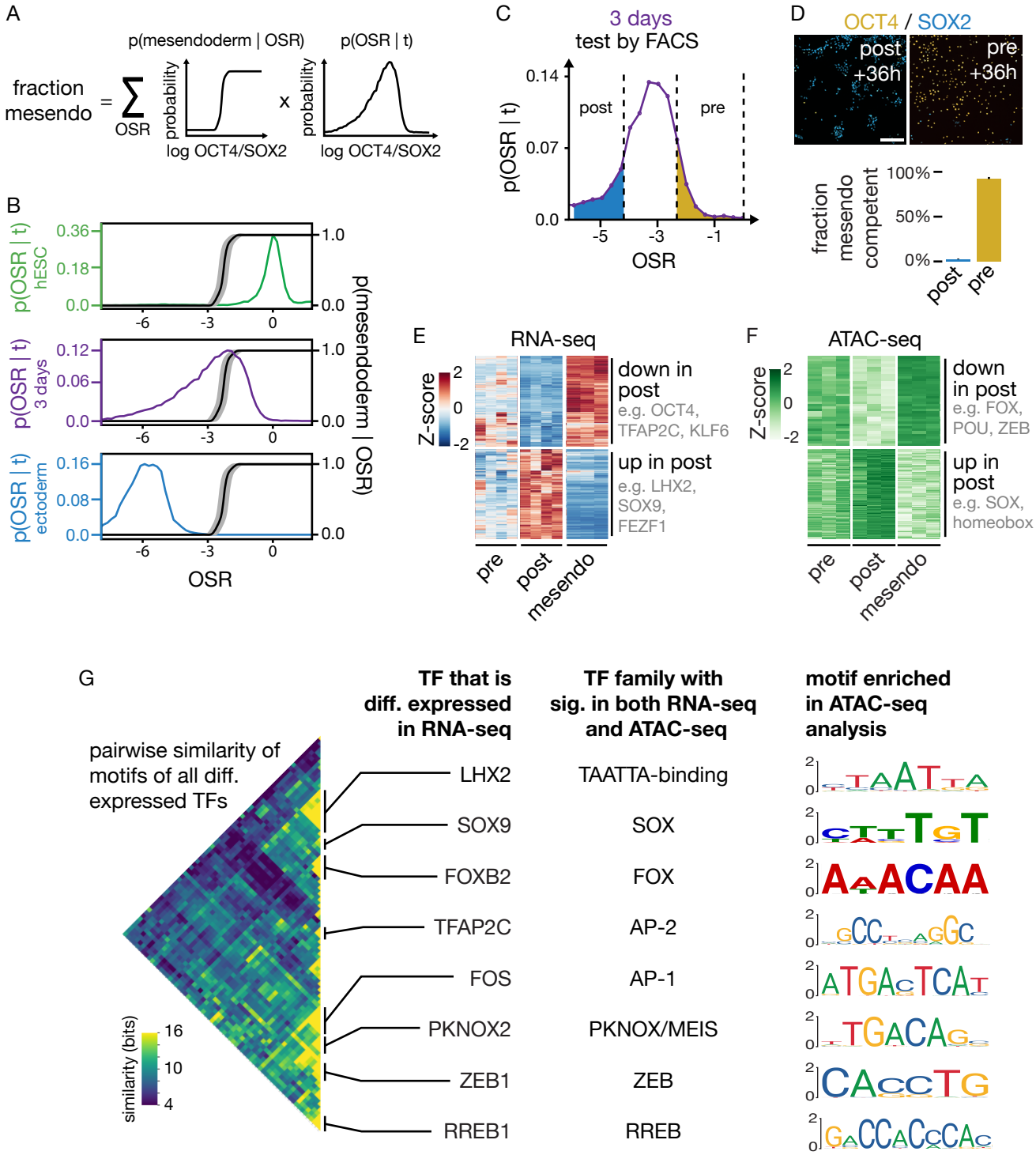
A
p(mesendoderm | OSR)
p(OSR | t)
fraction mesendo
OSR
probability
log OCT4/SOX2
probability
log OCT4/SOX2
B
p(OSR | t) hESC
p(OSR | t) 3 days
p(OSR | t) ectoderm
p(mesendoderm | OSR)
OSR
C
3 days
test by FACS
p(OSR | t)
post
pre
OSR
D
OCT4 / SOX2
post +36h
pre +36h
fraction mesendo competent
post
pre
E
RNA-seq
Z-score
down in post
e.g. OCT4, TFAP2C, KLF6
up in post
e.g. LHX2, SOX9, FEZF1
pre
post
mesendo
F
ATAC-seq
Z-score
down in post
e.g. FOX, POU, ZEB
up in post
e.g. SOX, homeobox
pre
post
mesendo
G
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
LHX2
TAATTA-binding
SOX9
SOX
FOXB2
FOX
TFAP2C
AP-2
FOS
AP-1
PKNOX2
PKNOX/MEIS
ZEB1
ZEB
RREB1
RREB

Figure 3

**Figure 3: Key TF families show concordant changes in expression and motif accessibility upon mesendoderm competence loss, suggesting perturbation candidates**

(A) The fraction of cells that adopt a mesendoderm-derived fate after BMP4 and Activin stimulation at a given time, p(mesendoderm | t), is equal to the sum over OSR of the probability of adopting a mesendoderm-derived fate given the OCT4:RFP/SOX2:YFP ratio, p(mesendoderm | OSR), times the probability distribution of OCT4:RFP/SOX2:YFP ratios at the given time, p(OSR | t).

(B) Distribution of OCT4:RFP/SOX2:YFP ratios for cell populations at three times during the transition to ectoderm as measured by flow cytometry. Shown are a population of pluripotent stem cells (green), a cell population after 3 days of Activin/NODAL inhibition (purple), and the $OCT4^-$ $SOX2^+$ ectoderm population produced after 3 days of Activin/NODAL inhibition and subsequent 48 h of BMP4 and Activin A stimulation (blue). The black sigmoid curve represents the inferred p(mesendoderm | OSR) based on the observed p(OSR | t) for the 3 day population, the observed final mesendoderm fraction after BMP4 and Activin A stimulation, and the sigmoid shape parameter learned from the time lapse data in Figure 2. The grey region represents the standard deviation of the location estimate of the sigmoid obtained from multiple biological replicates run in the same batch. In this coordinate, it is apparent that the 3 day population spans a region in OCT4:RFP/SOX2:YFP space from cells that are near-certain to adopt mesendodermal fates to cells that are near-certain to become ectoderm.

(C) Histogram showing the distribution of OCT4:RFP/SOX2:YFP ratios for a cell population after 3 days of Activin/NODAL inhibition. Overlaid on the histogram are the FACS gates used to sort subpopulation "post," shown in blue and predicted to have lost mesendoderm competence, and subpopulation "pre," shown in yellow and 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). The post-competence-loss subpopulation, which appears blue, is not able to respond to signals to downregulate SOX2 and maintain high OCT4 and is therefore not competent to become mesendoderm. The pre-competence-loss subpopulation downregulates SOX2 and maintains high OCT4 to become mesendoderm, thereby appearing yellow. Thus, the sorted populations essentially uniformly adopt the predicted fate. Scale bar = 300 μm. Bottom, fraction of cells in the populations represented above that became $OCT4^+$ $SOX2^-$ mesendoderm as determined by immunofluorescence after 40h of treatment with BMP4 and Activin A. Error bars represent the standard deviation from three biological replicates. Only "pre" is competent to become mesendoderm while "post" continues to the ectodermal fates.

(E) Heatmap showing normalized gene expression changes for genes that display lineage-specific significant differential expression patterns between the pre-competence-loss and post-competence loss populations (n=4 biological replicates each). The mesendoderm-derived outgroup is labeled "mesendo," and it likely is comprised of mostly endoderm cell types due to the high expression of endoderm markers such as SOX17 and FOXA2 at the population level (n=3 biological replicates). Key TFs are downregulated upon loss of mesendoderm competence, including OCT4, TFAP2C, and KLF6. LHX2, SOX9, and FEZF1 are among the TFs that are upregulated upon mesendoderm competence loss.

(F) 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. Each condition consisted of n=3 biological replicates. Much like the gene expression data, these regions display clear, lineage-specific accessibility changes.

(G) 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 matrix is shown because it is symmetric. 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 they are therefore good candidates for mesendoderm competence perturbation.

$p(\text{mesendoderm}|\text{OSR}) \approx 1$ would be competent to differentiate into mesendoderm, while those that were on the other side where $p(\text{mesendoderm}|\text{OSR}) \approx 0$ would not (Figure 3B). To validate this prediction, we sorted cells using FACS from a population that had been subjected to 3 days of ectodermal-directed differentiation. The sorting was performed based on each cell's ratio of OCT4:RFP to SOX2:YFP signal intensities to obtain populations that we predicted to be before and after the loss of mesendoderm competence (Figure 3C). We will hereafter refer to these sorted populations as "pre-competence-loss" and "post-competence-loss" for brevity. We then added BMP4 and Activin A to the sorted populations to compare our predicted competence with the observed fate choice of these cells. As predicted, we obtained essentially pure populations of OCT4$^-$ SOX2$^+$ ectoderm-derived fates from the post-competence-loss population and OCT4$^+$ SOX2$^-$ mesendoderm-derived fates from the pre-competence-loss population (Figure 3D). Thus, we were indeed able to predict and prospectively isolate pre- and post-competence-loss cells.

Having achieved our first goal of predicting mesendoderm competence, we turned to our second goal: to change this competence. Such a change could occur either by altering the position of cells along the ectoderm-directed developmental trajectory, $p(\text{OSR}|t)$, or by modulating how competence changes along the trajectory, $p(\text{mesendoderm}|\text{OSR})$. We hypothesized that important factors controlling mesendoderm competence would be DNA-binding factors whose expression patterns or access to binding sites changed along the developmental trajectory. Thus, to identify candidate factors that control these two probabilities, we characterized pre- and post-competence-loss populations using RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq). We obtained populations of pre- and post-competence cells using FACS from a single heterogeneous population of stem cells that had been

subjected to 3 days of ectodermal differentiation. As a control, we reserved a small fraction of sorted cells from each sample that we then treated with BMP4 and Activin A to confirm the competence of that sorted population. We also included a third, mesendoderm-derived population produced by subjecting pluripotent stem cells to BMP4 + Activin A for 40 hours, which allowed us to distinguish lineage-specific changes in expression and chromatin accessibility from changes that are shared by cells entering either lineage (Figure S3A).

Differential analysis of our RNA-seq data between the pre- and post-competence-loss populations using mesendoderm as an outgroup (see methods) showed that 544 genes were upregulated specifically in the post-competence-loss cells, 23 of which were TFs (Figure 3E, S3B). We also found 673 genes (32 TFs) that were specifically downregulated. In particular, we observed the differential expression of TFs such as SOX9, LHX2, FOXB2, TFAP2A, TFAP2C, PKNOX2, ZEB1, ZEB2, and GBX2, along with the expected expression pattern of OCT4 (Figure S3C). Consistent with our earlier observation (Figure S1G-H), the differentially expressed genes did not include signaling pathway components (Figure S3D). Our data further showed that competence loss occurred prior to the expression of master regulators such as PAX6 and SOX1 (Figure S3C). The expression pattern of all TFs is plotted in Figure S3E.

Analysis of GO term enrichment (Figure S3F) and ChIP-seq target enrichment (Figure S3G) further bolstered our confidence in our data; these analyses included, for example, the observation that genes that are specifically downregulated upon competence loss are most enriched for putative targets of factors previously implicated in pluripotency and germ layer selection, such as SOX2, ZEB1, SMAD4, TCF3, and KLF4; in contrast, genes that are specifically upregulated upon

mesendoderm competence loss are most enriched for putative targets of Polycomb repressive complex 2 component SUZ12, which is known to repress ectoderm target genes until their expression is appropriate.

We next analyzed our DNA accessibility data from ATAC-seq. We observed accessibility peaks that showed reproducible, clear changes between groups, alongside many peaks that were present in all samples (Figure S4A). Accessibility, as assayed by read depth, showed clear peaks at transcription start sites (Figure S4B). Differential analysis of our ATAC-seq data between the pre- and post-competence-loss populations using mesendoderm as an outgroup (see methods) showed thousands of regions that change accessibility between pre- and post-competence-loss populations. We found 2071 regions that were more accessible after competence loss, and 233 that were less accessible (FDR < 0.05; Figure 3F). Most of these differentially accessible regions were located in distal intergenic or intronic regions (Figure S4C).

We next confirmed that we could reproduce expected patterns in our ATAC-seq data. Using the software GREAT[21], we found that genes near loci with increased accessibility in the post-competence-loss samples were most significantly enriched for orthologs of mouse genes expressed in the Theiler Stage 11 neurectoderm (Figure S4D), increasing our confidence that we were indeed sampling the transition we hoped to capture. We further noted that we could recover previously hypothesized motif signals from our data. For example, our ATAC-seq peaks that changed significantly and which also contained a compressed OCT4/SOX17 motif were almost exclusively open in the mesendoderm-derived outgroup (Figure S4E), consistent with what would be predicted from previous work[22]; in contrast, such regions that contained the classical OCT4/SOX2 motif

were generally closed in the mesendoderm-derived cells. Interestingly, we did not observe significant changes in accessibility at any of the ENCODE-annotated candidate regulatory elements of pluripotency genes such as OCT4, SOX2, NANOG, KLF4, and MYC (Figure S4F).

To search for TFs that potentially bind to the differentially accessible regions, we found sequence motifs that were enriched at these loci. We found more than 20 such motifs, many of which matched the known DNA binding motifs of the differentially expressed TF families we had identified, including motifs that were similar to those bound by SOX, forkhead box (FOX), AP-2, AP-1, TAATTA-binding homeobox-like, PKNOX/MEIS, Zinc Finger E-Box Binding Homeobox (ZEB), and POU family TFs (Figure S4G), the latter of which includes OCT4 as a member. As a complementary analysis, we determined which known sequence motifs could best explain the changes in chromatin accessibility that we observed across the point of competence loss (Figure S4H).

Taken together, the gene expression and chromatin accessibility data revealed a core set of TFs that is remodeled upon competence loss. When we clustered the known binding motifs of the differentially expressed TFs by calculating a measure of similarity between each pair of motifs, we found that many such TFs shared similar binding motifs (Figure 3G). These clusters correspond to major TF families, and multiple members of each family are differentially expressed. Thus, a small number of key TF families change expression upon competence loss, and competition between TFs for similar binding sites may play a role in modulating competence. The core TF families we uncovered included the SOX, FOX, AP-2, AP-1, PKNOX/MEIS, ZEB, TAATTA-binding homeobox-like, and POU TF families. Furthermore, each of these major TF family DNA

binding motifs was also enriched in the ATAC-seq analysis, indicating that the expression changes of these TFs have clear signatures in the chromatin accessibility data. The concordance of our RNA-seq and ATAC-seq results shows that the gene regulatory network changes that correspond with competence loss produce consistent signatures across data modalities. These key TFs compose a core gene regulatory network that is remodeled upon competence loss (Figure S5A). We hypothesized that perturbation of these factors could modulate the position of cells along the ectoderm-directed developmental trajectory, $p(\mathrm{OSR}|t)$, or the change in mesendoderm competence along the trajectory, $p(\mathrm{mesendoderm}|\mathrm{OSR})$

Using the core set of TFs we identified as potential regulators of mesendoderm competence, we selected 40 candidate genes whose exogenous expression might perturb cellular competence (Figure S5B and S5C; see methods). These candidates were composed largely of the differentially expressed TFs in our core network plus paralogs of those TFs. We overexpressed each candidate TF to test its effects on mesendoderm competence and the dynamics of ectoderm-directed differentiation. Using a lentiviral delivery system, we transduced cells with a payload of the gene of interest separated from the C-terminal end of a mCerulean cyan fluorescent protein (CFP) by a P2A self-cleaving peptide sequence, all under the control of an EF-1α promoter (Figure S6A). The co-expressed CFP marker allowed us to monitor transduction efficiency at the single-cell level. We titrated viral concentration to achieve <50% transduction so that each sample also contained many non-transduced ($CFP^-$) cells to serve as an internal control population.

In principle, a change in the fraction of lentiviral-perturbed cells that form mesendoderm-derived cell types after BMP4 and Activin A treatment could result from one of two broad effects: first,

the perturbation could facilitate or impede the initial ectoderm-directed differentiation, thereby changing $p(\mathrm{OSR}|t)$, or second, the perturbation could change $p(\mathrm{mesendoderm}|\mathrm{OSR})$, the probability of transitioning to mesendoderm given the location along the developmental trajectory. We therefore designed our experiment to monitor the initial differentiation as well as the final fate mixture achieved, both using our OCT4:RFP and SOX2:YFP line.

To monitor the cells both before and after signal addition, we seeded two parallel samples for each candidate (Figure S6B). Both samples began with a pluripotent stem cell population, and we started viral transduction at the same time as we initiated ectoderm-directed differentiation. After three days of differentiation, we analyzed one sample by flow cytometry and switched the other to media containing BMP4 + Activin A. After 42h of this treatment, we assayed our second sample by flow cytometry and confirmed that our non-transduced controls had produced about 50% OCT4:RFP$^{+}$ SOX2:YFP$^{-}$ mesendoderm-derived cells and 50% OCT4:RFP$^{-}$ SOX2:YFP$^{+}$ ectoderm-derived cells. The transduced cells, in contrast, expressed CFP (Figure S6C) and displayed varying final lineage proportions.

For each candidate TF, we measured $p(\mathrm{OSR}|t)$ using flow cytometry before signal addition and the final fraction of mesendoderm produced after BMP4 and Activin stimulation. We then computed $p(\mathrm{mesendoderm}|\mathrm{OSR})$ based on that $p(\mathrm{OSR}|t)$ and the observed final fraction of mesendoderm. For some candidate TFs, their overexpression principally affected progression along the developmental trajectory as assayed by the measured $p(\mathrm{OSR}|t)$. For example, overexpression of SOX9 facilitated movement along the developmental trajectory, thereby shifting the distribution of $p(\mathrm{OSR}|t)$ towards the ectoderm fates, but did not affect $p(\mathrm{mesendoderm}|\mathrm{OSR})$

(Figure 4A). In contrast, overexpression of TFAP2C hindered movement along the developmental trajectory, thereby shifting $p(\mathrm{OSR}|t)$ towards the pluripotent state without affecting $p(\mathrm{mesendoderm}|\mathrm{OSR})$. Thus, these candidates tuned competence by changing cellular location along the developmental trajectory without altering the barrier between fates.

Other candidates affected mesendoderm competence by shifting the barrier between fates as assayed by $p(\mathrm{mesendoderm}|\mathrm{OSR})$. For example, overexpression of LHX2 did not impact $p(\mathrm{OSR}|t)$, but it did restrict mesendoderm competence by shifting $p(\mathrm{mesendoderm}|\mathrm{OSR})$ (Figure 4B). Similarly, FOXB2 overexpression extended mesendoderm competence by altering $p(\mathrm{mesendoderm}|\mathrm{OSR})$ despite minimal impacts on $p(\mathrm{OSR}|t)$. These candidates tuned mesendoderm competence by moving the barrier between fates on the developmental landscape, and this change occurred independently of alterations to $p(\mathrm{OSR}|t)$.

While some of the candidates principally affected either $p(\mathrm{mesendoderm}|\mathrm{OSR})$ or $p(\mathrm{OSR}|t)$, many candidates that impacted competence did so by altering both $p(\mathrm{mesendoderm}|\mathrm{OSR})$ and $p(\mathrm{OSR}|t)$ (Figure 4C, S6D). For example, while overexpression of FEZF1 and TFAP2A both shifted $p(\mathrm{OSR}|t)$ towards the pluripotent state, they shifted $p(\mathrm{mesendoderm}|\mathrm{OSR})$ in opposite directions. Such candidates further underscored that these two mesendoderm competence-modulating mechanisms could be tuned independently.

Finally, we explored the developmental consequences of mesendoderm competence modulation via these two mechanisms. Overexpression of FOXB2, a candidate that increased mesendoderm competence primarily via a shift in $p(\mathrm{mesendoderm}|\mathrm{OSR})$, increased the fraction of cells that

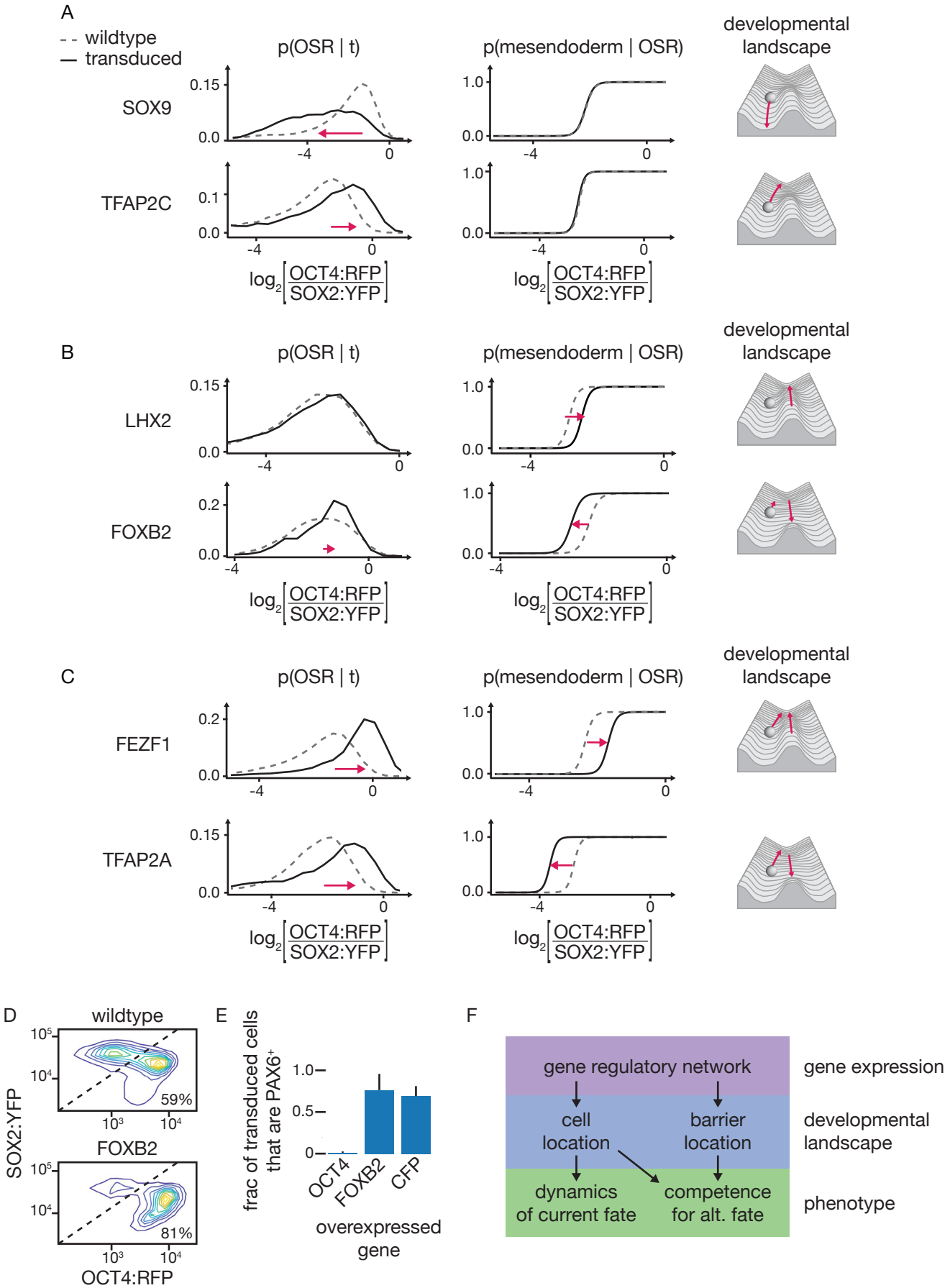
A
wildtype
transduced
p(OSR | t)
p(mesendoderm | OSR)
developmental landscape
SOX9
TFAP2C
B
LHX2
FOXB2
C
FEZF1
TFAP2A
log2[OCT4:RFP / SOX2:YFP]
D
wildtype
FOXB2
SOX2:YFP
OCT4:RFP
59%
81%
E
frac of transduced cells that are PAX6+
OCT4
FOXB2
CFP
overexpressed gene
F
gene regulatory network
cell location
barrier location
dynamics of current fate
competence for alt. fate
gene expression
developmental landscape
phenotype

Figure 4

**Figure 4: Perturbation of TFs candidates from RNA-seq and ATAC-seq can independently tune progression along the developmental landscape and inter-fate barrier.**

(A) 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), but did not alter the probability of adopting a mesendoderm fate given the pre-signal OCT4:RFP/SOX2:YFP ratio, p(mesendoderm | OSR). OCT4:RFP and SOX2:YFP levels were normalized to an hESC sample measured in the same batch. Left, the measured p(OSR | t). Center, the inferred p(mesendoderm | OSR) based on the measured p(OSR | t) and the final fraction of the paired replicate that adopted a mesendoderm-derived fate. Gray dashed lines represent the distribution in the non-transduced wildtype internal control cells, black solid lines represent the transduced cells that overexpress the gene of interest, pink arrows show the direction of probability movement. Right, a schematic depicting the corresponding effects on the Waddington developmental landscape.

(B) Examples of two candidates, LHX2 and FOXB2, whose overexpression principally shifted p(mesendoderm | OSR). Panels as in A.

(C) Examples of two candidates, FEZF1 and TFAP2A, whose overexpression altered both p(OSR | t) and p(mesendoderm | OSR). These two candidates shifted p(OSR | t) in the same direction, towards the distribution seen in pluripotent cells, but shifted p(mesendoderm | OSR) in opposite directions. Panels as in A.

(D) Contour plots showing the final fates adopted by wildtype or FOXB2-overexpressing cells after 48 h of BMP4 and Activin stimulation. Top, wildtype, non-transduced cells as determined by absence of CFP expression. Bottom, cells from the same population that were transduced with the CFP:P2A:FOXB2 cassette, as assayed by their measured CFP fluorescence. Percentage in lower right indicates the fraction of each population that adopted mesendoderm-derived fates. Dashed line drawn for reference. Overexpression of FOXB2 increases the number of cells that adopt a mesendodermal fate.

(E) Quantification of confocal microscopy images (see Fig S6E). Cells were transduced with the indicated gene expression cassette, maintained in Activin/NODAL inhibition conditions for 6 days, and immunostained for PAX6. Each bar represents the mean fraction of transduced cells that were positive for PAX6 after 6 days of Activin/NODAL inhibition across multiple fields of view. Error bars represent the standard deviation. OCT4 overexpression precludes PAX6 induction, while FOXB2 overexpression does not prevent normal PAX6 induction as cells adopt a neurectodermal fate.

(F) Schematic showing the model suggested by these results. The gene regulatory network governs both the progression of the cell along the developmental trajectory and the shape of the 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, but only perturbations of the location along the developmental landscape alter the dynamics of movement along the current trajectory in the absence of a new signal.

adopted mesendoderm by 32±5% (Figure 4D). Despite this change in mesendoderm competence, FOXB2 overexpression did not prevent normal neurectodermal differentiation and concomitant PAX6 expression in the absence of the BMP and Activin signal (Figure 4E, S6E). Thus, alternative lineage competence can be modulated without preventing normal lineage progression in the absence of alternative-lineage-inducing signals.

Our finding that competence for an alternative lineage can be controlled either by changing the location of the cell along the developmental trajectory or moving the barrier that prevents cells from crossing over to the mesendoderm canal suggests possible evolutionary and developmental consequences. While moving the cell along the trajectory interferes with normal development in the absence of alternative signal, moving the barrier does not (Figure 4F). These two effects represent fundamentally different mechanisms for tuning competence, and both could be at play in the developing embryo. During the patterning of the mammalian epiblast, for example, the mesendodermal progenitors are generated along the primitive streak as it extends anteriorly from the posterior end of the epiblast. We speculate that changing the dynamics of epiblast competence loss anterior to the primitive streak could be a possible mechanism for tuning the length and extent of the streak. Further investigation of the role of competence modulation during mammalian gastrulation could be an important element in a full description of this important process, and the same mechanism could be acting to tune relative tissue sizes during any given cellular decision.

We expect that discovering order parameters for different lineage decisions and monitoring the dynamics of differentiation along these coordinates will be fruitful for dissecting many cell fate decisions and state transitions. Our recent computational work has demonstrated how to find

candidate order parameters for any given lineage choice[18,19], which greatly simplifies the prospect of following this approach in another context. Indeed, several other choices and competence restrictions occur immediately adjacent to the loss of mesendoderm competence in the early germ layer lineage tree, such as the presumptive loss of non-neural potential during neurectoderm fate determination or the loss of ectoderm competence during mesendoderm differentiation.

In sum, these findings identify modulation of the probability of adopting alternative lineages in response to signal as a regulatory mechanism that is separable from the cell's progression along its developmental trajectory, and we anticipate that understanding how competence is tuned will be crucial for the study of patterning of the mammalian embryo during development.

## References


1 Gilbert S. The Developmental Mechanics of Cell Specification. *Dev. Biol.* 6th ed. Sunderland (MA): Sinauer Associates; 2000.

2 Handyside AH. Time of commitment of inside cells isolated from preimplantation mouse embryos 1978;**45**:37–53.

3 Pedersen RA, Wu K, BaŁakier H. Origin of the inner cell mass in mouse embryos: Cell lineage analysis by microinjection. *Dev Biol* 1986;**117**:581–95. https://doi.org/10.1016/0012-1606(86)90327-1.

4 Berg DK, Smith CS, Pearton DJ, Wells DN, Broadhurst R, Donnison M, *et al.* Trophectoderm Lineage Determination in Cattle. *Dev Cell* 2011;**20**:244–55. https://doi.org/10.1016/j.devcel.2011.01.003.

5 Rossant J, Lis WT. Potential of isolated mouse inner cell masses to form trophectoderm derivatives in vivo. *Dev Biol* 1979;**70**:255–61. https://doi.org/10.1016/0012-1606(79)90022-8.

6 Rossant J, Vijh KM. Ability of outside cells from preimplantation mouse embryos to form inner cell mass derivatives. *Dev Biol* 1980;**76**:475–82. https://doi.org/10.1016/0012-1606(80)90395-4.

7 Oron E, Ivanova N. Cell fate regulation in early mammalian development. *Phys Biol* 2012;**9**:045002. https://doi.org/10.1088/1478-3975/9/4/045002.

8 Li L, Liu C, Biechele S, Zhu Q, Song L, Lanner F, *et al.* Location of transient ectodermal progenitor potential in mouse development. *Development* 2013;**140**:4533–4543. https://doi.org/10.1242/dev.092866.

9 Li L, Song L, Liu C, Chen J, Peng G, Wang R, *et al.* Ectodermal progenitors derived from epiblast stem cells by inhibition of Nodal signaling. *J Mol Cell Biol* 2015;**7**:455–465. https://doi.org/10.1093/jmcb/mjv030.

10 Etoc F, Metzger J, Ruzo A, Kirst C, Yoney A, Ozair MZ, *et al.* A Balance between Secreted Inhibitors and Edge Sensing Controls Gastruloid Self-Organization. *Dev Cell* 2016;**39**:302–15. https://doi.org/10.1016/j.devcel.2016.09.016.

11 Zhang Z, Zwick S, Loew E, Grimley JS, Ramanathan S. Mouse embryo geometry drives formation of robust signaling gradients through receptor localization. *Nat Commun* 2019;**10**:. https://doi.org/10.1038/s41467-019-12533-7.

12 Lupo G, Harris WA, Pedersen RA, Alexander M, Smith JR, Vallier L. Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Dev Biol* 2007;**313**:107–117. https://doi.org/10.1016/j.ydbio.2007.10.003.

13 Landau LD. On the Theory of Phase Transitions. *Collect. Pap.*, vol. 1. Moscow: Nauka; 1969. p. 234–252.

14 Anderson PW. Local moments and localized states (Nobel Prize Lecture) 1977.

15 De Gennes P-G. Soft Matter (Nobel Lecture) 1991.

16 Kosterlit JM, Thouless DJ. Ordering, metastability and phase transitions in two-dimensional systems. *J Phys C Solid State Phys* 1973;**6**:1181–203.

17 Thomson M, Liu SJ, Zou L-N, Smith Z, Meissner A, Ramanathan S. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* 2011;**145**:875–889. https://doi.org/10.1016/j.cell.2011.05.017.

18 Jang S, Choubey S, Furchtgott L, Zou L-N, Doyle A, Menon V, *et al.* Dynamics of embryonic stem cell differentiation inferred from single-cell transcriptomics show a series of transitions through discrete cell states. *Elife* 2017;**6**:. https://doi.org/10.7554/eLife.20487.

19 Furchtgott LA, Melton S, Menon V, Ramanathan S. Discovering sparse transcription factor codes for cell states and state transitions during development. *ELife* 2017;**6**:1–33. https://doi.org/10.7554/eLife.20488.

20 Greber B, Coulon P, Zhang M, Moritz S, Frank S, Müller-Molina AJ, *et al.* FGF signalling inhibits neural induction in human embryonic stem cells: FGF inhibits neural induction in hESCs. *EMBO J* 2011;**30**:4874–84. https://doi.org/10.1038/emboj.2011.407.

21 McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;**28**:495–501. https://doi.org/10.1038/nbt.1630.

22 Merino F, Ng CKL, Veerapandian V, Schöler HR, Jauch R, Cojocaru V. Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. *Structure* 2014;**22**:1274–1286. https://doi.org/10.1016/j.str.2014.06.014.

23 Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 2015;**109**:21.29.1–21.29.9. https://doi.org/10.1002/0471142727.mb2129s109.

24 McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol* 2018;**16**:e2005970. https://doi.org/10.1371/journal.pbio.2005970.

25 Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, *et al.* The Human Transcription Factors. *Cell* 2018;**172**:650–665. https://doi.org/10.1016/j.cell.2018.01.029.

26 Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N. Advanced methods of microscope control using μManager software. *J Biol Methods* 2014;**1**:10. https://doi.org/10.14440/jbm.2014.36.

## Methods

### Cell lines

We conducted our experiments using WA01 (H1) human embryonic stem cells. We also used an H1 cell line in which both OCT4 and SOX2 were tagged with fluorescent proteins as previously described[11]. In these cells, one endogenous copy of OCT4 was replaced with OCT4:tdTomato followed by an internal ribosomal entry site and a neomycin resistance gene to allow for selection, and one endogenous copy of SOX2 was replaced with SOX2:FLAG:Citrine:P2A:PuroR.

**Cell culture**

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

For all experiments, we seeded cells on polyester membrane filters (Sterlitech) with 3 μm pores that had been treated with Matrigel. We chose this substrate to allow all cells to receive the BMP and Activin/NODAL signals we added to the media. TGF-β superfamily receptors, such as those for BMP4 and Activin A, are localized basolaterally in epithelial stem cell colonies and *in vivo* in the epiblast, so they are insulated from ligands in the apical media or luminal fluid [10,11]. Typical tissue culture conditions allow for only the cells on the edge of the colony to receive signals, but growing cells on a membrane allows all cells in a colony access to the BMP and Activin ligands.

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

**Differentiation conditions**

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

**Flow cytometry**

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

**Fluorescence activated cell sorting**

Accutase-dissociated cells were sorted using a BD Aria III (BD Biosciences) using a 100 µm nozzle. Cells were gated such that the pre-competence-loss population was taken as the cells with the top 10-15% OCT4:RFP to SOX2:YFP ratio, and the post-competence-loss population was the bottom 10-15% OCT4:RFP to SOX2:YFP ratio. We sorted around 250,000 cells per subpopulation

in a typical experiment. Populations were sorted into 1.5 mL centrifuge tubes (Eppendorf) filled with 500 μL of mTeSR supplemented with 10 μM γ-27632; by the end of the sort, ~800 μL of sheath and sorted cells had been added to each tube. After the sort had completed, we pelleted the cells in a microcentrifuge at 250 xg for 3 minutes, then resuspended in PBS.

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

**RNA-seq**

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

**ATAC-seq**

ATAC-seq was performed as previously described[23]. Briefly, live cells were lysed and incubated with Tn5 transposase for 30 min at 37°C. After DNA purification, samples were amplified for the appropriate number of cycles as determined by qPCR to minimize PCR bias. Sequencing was performed by the sequencing core at Massachusetts General Hospital. We obtained ~100M mapped paired-end reads per sample.

**Plasmid construction**

Overexpression targets were subcloned from plasmids available through the Harvard PlasmID database, where available. Other targets were cloned from complementary DNA (cDNA) libraries.

To prepare cDNA libraries for cloning, we differentiated human stem cells for 3 days in mTeSR + 0.5 μM A8301 and extracted RNA with RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. We then performed first strand cDNA synthesis using SuperScript II Reverse Transcriptase (Thermo Fisher). We amplified the relevant cDNAs using Phusion polymerase (NEB) or Kapa HiFi (Kapa Biosystems). The OCT4 DNA binding domain and the SOX2 DNA binding domain (OCT4DBD and SOX2DBD) were amplified from cDNA. The OCT4DBD consisted of amino acids 131-296 of OCT4A (NCBI reference sequence: NM_002701.5). The SOX2DBD consisted of amino acids 37-117 of SOX2 (NCBI reference sequence: NM_003106.3). All cDNA-amplified clones were fully sequence confirmed by Sanger sequencing (Genewiz).

We constructed a vector from the second-generation lentiviral transfer backbone pWPXL with an EF-1α promoter. pWPXL was a gift from Didier Trono (Addgene plasmid #12257;

http://n2t.net/addgene:12257; RRID:Addgene_12257). We first joined sequences for fluorescent protein mCerulean (CFP) and 2A peptide P2A (a ribosomal skip sequence) with Q5 (NEB) fusion PCR and added them to the pWPXL vector with Gibson Assembly Master Mix (NEB). We then constructed final transfer vectors by inserting target cDNA after the P2A using Gibson assembly. All constructed vectors were sequence confirmed at Gibson assembly junctions by Sanger sequencing (Genewiz) prior to use. Plasmids were grown and stored in NEB Stable *E. coli* (NEB).

**Lentiviral overexpression and flow cytometry analysis**

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

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

Two samples of each overexpression condition were performed in parallel. We harvested cells of one sample after 3 days of treatment with A83-01 and the cells of the second sample after 3 days

of A83-01 followed by 41 hours of BMP4 + Activin A treatment. We analyzed each sample immediately after harvest using an LSRFortessa (BD Biosciences).

We analyzed differential OCT4:RFP to SOX2:YFP ratio distributions between CFP-positive and CFP-negative populations of each 3-day sample by calculating the Kullback-Leibler divergence in MATLAB (MathWorks). To determine differences in proportions of end fates (OCT4:RFP+/SOX2:YFP- and OCT4:RFP-/SOX2:YFP+), we manually gated ectoderm and mesendoderm populations using a custom MATLAB script and used identical gates for both CFP-positive and CFP-negative populations.

**Immunofluorescence**

Cells were fixed with 4% formaldehyde for 15 min at room temperature. Fixed cells were treated with blocking buffer (PBS + 5% normal donkey serum + 0.3% Triton X-100) for 1 h, then overnight at 4°C with primary antibody diluted in staining buffer (PBS + 1% BSA + 0.3% Triton X-100). The following primary antibodies were used: OCT4 (1:400, Cell Signaling C30A3); SOX2 (1:400, Thermo Fisher BTJCE); SOX17 (1:100, R&D Systems AF1924); phosphorylated SMAD1/5/9 (1:200, Cell Signaling D5B10); phosphorylated-SMAD2 (1:200, Cell Signaling E8F3R); PAX6 (1:200, DSHB AB_528427). After overnight incubation, samples were washed three times with PBS, then secondary antibodies diluted in staining buffer were added. We used the following secondary antibodies all at a dilution of 1:1000: donkey anti-rabbit Alexa 568 (Thermo Fisher), donkey anti-rat Alexa 488 (Thermo Fisher), and donkey anti-mouse Alexa 647 (Thermo Fisher). We incubated with a 300 nM DAPI (Thermo Fisher) solution in PBS for 5 minutes to visualize DNA. For analysis of the resulting images, we used CellProfiler 3.1.8[24] to

segment well-separated nuclei for samples where automated segmentation performed well (those shown in Figure 3), and we used manual segmentation for more challenging images (those shown in Figures 1 and S6).

**ATAC-seq analysis**

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

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

The primary DNA sequences of differentially accessible peaks were retrieved from Ensembl and examined for motifs using MEME-ChIP 5.0.3. ATAC-seq read depth was modeled as a function of known motif presence using chromVAR 1.4.1. Significant motif matches were identified with

FIMO 5.0.3. For the gene regulatory network, possible associations between genomic regions and target genes were identified using CisMapper 5.0.5. The full list of human TFs and the motifs for each TF were extracted from the list in Lambert *et al.*[25]. Mutual information between pairs of motifs was calculated with a custom python script.

**RNA-seq analysis**

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

For clustering the motifs of differentially expressed transcription factors, similarity between each pair of motifs was quantified as the Kullback-Leibler divergence of the product of the two motifs from a reference distribution, which was the product of two uniform motifs (0.25 probability for each base at each position). Motif alignment was performed by calculating the aforementioned divergence at each possible offset and using the maximum value obtained at any offset. This calculation was performed using a custom python script. The linkage was computed using the scipy.cluster.hierarchy.linkage function from scipy 1.3.0 with the "average" clustering method and the "braycurtis" distance.

**Overexpression candidate selection**

We selected TFs by incorporating information from both RNA-seq and ATAC-seq analyses. We began with all TFs that were differentially expressed between the pre- and post-competence-loss populations ($q < 0.05$ with DESeq2). A gene was considered to be a TFs if it was so annotated in Lambert *et al.*[25]. We then limited this list only to those that were expressed in a lineage-specific pattern and had above-background expression levels in at least one of the three populations. We defined genes with a lineage-specific expression pattern as those genes that (1) were differentially expressed between pre- and post-competence-loss populations and (2) either were not differentially expressed from the pre-competence-loss to mesendoderm populations or were differentially expressed in the opposite direction (upregulated from pre- to post- and downregulated from pre- to mesendoderm, or vice versa. By these criteria, 23 TFs were specifically upregulated with competence loss and 32 were specifically downregulated with competence loss. We also added select paralogs of the TFs that passed our expression pattern cutoffs: POU6F1, GRHL1, POU2F3, FOXJ2, and POU2F1, along with the OCT4 and SOX2 DNA binding domains. We further restricted the list to those candidates that had a known, high-quality DNA binding motif that appeared in either the DiffBind/MEME-ChIP or chromVAR analyses of our data. We also added four TFs (OTX2, JUNB, ZSCAN23, and GSC) whose motifs appeared in our ATAC-seq analyses but did not pass our differential expression cutoffs. We also eliminated 10 TFs for which a clone was not readily accessible to us, either from the Harvard PlasmID database, Addgene, or genes that had previously been cloned from cDNA in our lab. We note that one candidate that was tested before all RNA-seq analysis was complete, MBNL2, missed significance cutoffs in the final analysis but is nevertheless included for completeness. After

adding three candidates based on the literature (NRF2, ZNF521, and ID2), we were left with 40 candidates in total.

**Epifluorescence imaging of fixed samples**

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

**Live cell time lapse imaging**

Samples were imaged on a Zeiss AxioObserver Z1 inverted microscope using a Zeiss 20x plan apo objective (NA 0.8) using the appropriate filter sets and a Hamamatsu ImagEM EMCCD camera. Cells were maintained in a 37 degree incubation chamber at 5% $CO_2$. Cells were imaged every 15 minutes. Focus was maintained using a combination of Zeiss Definite Focus and, using a custom script in MicroManager 2.0 beta[26], software autofocus adjustments every hour to compensate for slight movement of the membrane. For maximum accuracy, cells in this time lapse were tracked manually, and the tracks were analyzed with a custom python script that performed illumination profile correction and background subtraction.

**Confocal imaging**

For Figure 1, cells were imaged on a Leica inverted microscope with a Zeiss 20x objective (NA 0.8) with the appropriate filter sets. Detection was performed with photomultiplier tubes (for

detection of Alexa 488 and Alexa 647) and a Leica HyD Photon Counter (for Alexa 568). For Figure S6, cells were imaged on a Zeiss LSM 880 with Airyscan using a Zeiss 20x objective (NA 0.8). Detection was performed with photomultiplier tubes (Alexa 568 and Alexa 647) and a GaAsP detector (CFP and Alexa 488).

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

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

**Ethical compliance**

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

## Acknowledgments

We thank the staff of the Bauer Core at Harvard University for their work on the RNA sequencing used in this manuscript as well as for their expertise and assistance with flow cytometry and FACS. We thank the Massachusetts General Hospital sequencing core for their work on the ATAC-seq data used in this manuscript, the Harvard Physics/SEAS Instructional Machine Shop for making the stainless-steel washers used in the live cell imaging in this study, and the Harvard Center for Biological Imaging for the use of their equipment. We thank Andrew Murray, Sean Eddy, and all of the members of the Ramanathan Lab for their helpful comments. JV was funded by The Fannie and John Hertz Foundation, the National Science Foundation Graduate Research Fellowship Program, and the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard, award #1764269. Some confocal imaging was conducted on an instrument provided by the Harvard MRSEC (DMR-1420570). This work was supported in part by NIH R01GM131105-01.