Stem Cell Reports

Resource

Virtual cells in a virtual microenvironment recapitulate early development-like patterns in human pluripotent stem cell colonies

Himanshu Kaul,[1,2] Nicolas Werschler,[3] Ross D. Jones,[3,4,7] M. Mona Siu,[4,5,7] Mukul Tewary,[6] Andrew Hagner,[3] Joel Ostblom,[6] Daniel Aguilar-Hidalgo,[3,4] and Peter W. Zandstra[3,4,*]

[1]School of Engineering, University of Leicester, Leicester, UK
[2]Department of Respiratory Sciences, University of Leicester, Leicester, UK
[3]School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada
[4]Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada
[5]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada
[6]Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada
[7]These authors contributed equally
*Correspondence: peter.zandstra@ubc.ca
https://doi.org/10.1016/j.stemcr.2022.10.004

SUMMARY

The mechanism by which morphogenetic signals engage the regulatory networks responsible for early embryonic tissue patterning is incompletely understood. Here, we developed a minimal gene regulatory network (GRN) model of human pluripotent stem cell (hPSC) lineage commitment and embedded it into "cellular" agents that respond to a dynamic morphogenetic signaling microenvironment. Simulations demonstrated that GRN wiring had significant non-intuitive effects on tissue pattern order, composition, and dynamics. Experimental perturbation of GRN connectivities supported model predictions and demonstrated the role of OCT4 as a master regulator of peri-gastrulation fates. Our so-called GARMEN strategy provides a multiscale computational platform to understand how single-cell-based regulatory interactions scale to tissue domains. This foundation provides new opportunities to simulate the impact of network motifs on normal and aberrant tissue development.

INTRODUCTION

Gastrulation organizes the seemingly equivalent epiblast cells into the three germ layers: ectoderm, mesoderm, and endoderm. These germ layers contribute to the formation of all tissues and organs. How the germ layers emerge in humans remains elusive due to technical and ethical challenges. Seeding human pluripotent stem cells (hPSCs) on extracellular matrix micropatterns of defined shape and size provides a model to study the elements of this phenomenon (Tewary et al., 2017; Warmflash et al., 2014). Upon stimulation by bone morphogenetic protein-4 (BMP4), micropatterned hPSC colonies self-organize with radially segregated SOX2+ (ectoderm-associated), Brachyury+ (mesoderm-associated), SOX17+ (endoderm-associated), and CDX2+ (extraembryonic-associated) cells. This pattern has been explained using reaction-diffusion (RD) models based on BMP4 activation and autoregulation (Etoc et al., 2016; Tewary et al., 2017). The signaling cascade has been extended to include a BMP → WNT → NODAL pathway hierarchy in hPSCs (Martyn et al., 2018). Evidence exists that this cascade evolves dynamically, yielding neighboring domains in a sequential manner (Chhabra et al., 2019; Tewary et al., 2019). However, several questions remain unanswered: how do gene regulatory network (GRN) interactions mediate germ-layer commitment; how do the germ layers form their distinct identities; and how does the interplay between gene regulation and exogenous cues enable cells to decode their position during patterning?

Computational models are optimal tools to address these systems-level questions. They offer tightly controlled, interlinked variables that reflect the manipulable parameters of the physical system, thus acting as virtual laboratories for hypothesis generation, testing, and refinement. Computational methods across multiple scales have been used to study cell fates. Intracellular GRNs formalize interactions between gene products, including transcription factors (TFs), signal molecules, and epigenetic modifiers, to predict context-specific gene expression patterns (Dunn et al., 2014; Parfitt and Shen, 2014; Yachie-Kinoshita et al., 2018). Cellular models simulate intercellular interactions (Kaul and Ventikos, 2015). These can be lattice based or off lattice and mechanistic or phenomenological (Glen et al., 2019). Microenvironment gradients can be captured using transport phenomena (Kaul et al., 2016) or RD (Chhabra et al., 2019; Etoc et al., 2016; Tewary et al., 2017).

How lower-level interactions shape the macroscopic properties of a system and are, in turn, influenced by the very same macroscopic properties remains elusive. There are two key aspects to this problem: (1) how do individual GRNs coordinate interactions with thousands of GRNs operating within locally and globally neighboring cells (i.e., individual shaping the continuum), and (2) how do these interactions drive the microenvironmental gradients

that ultimately mediate the activity of the same GRNs (i.e., continuum shaping the individual). This is unknown in the context of peri-gastrulation patterns. Existing RD models do not explain how the underlying GRNs shape spatial gradients, and existing GRN models simulate the attractor states for only a single cell.

Herein, we bridge this gap by discretizing thousands of GRNs in space using agents (virtual cells) linked to morphogen gradients based on RD (virtual environment). This helped both to link thousands of GRNs and to couple spatial gradients with individual GRNs. The so-called GARMEN (GRN-agent-RD modeling environment) thus captured how GRNs mediate the microenvironment (individual shaping the continuum) and how that dynamic microenvironment regulates the GRN (continuum shaping the individual). It showed how the same master GRN can yield significantly different tissue structures depending on the signaling context, perturbation, and perturbation timing. GARMEN successfully identified OCT4 as a master regulator of germ-layer patterning in hPSCs. We also show that cells decode their position by considering the biochemical and gene-expression gradients cumulatively. Overall, our framework links a minimal GRN to tissue-like patterns in a systematic manner, providing mechanistic insights into cell-fate control.

RESULTS

Embedding GRN into agents connects regulatory decision-making to tissue patterns

GRN development and validation

We considered interactions between key signals, pluripotency factors, and lineage-specification markers. Briefly, the core pluripotency factors OCT4, NANOG, and SOX2 are closely linked via autoregulatory interactions (Boyer et al., 2005). Systematic knockout studies have revealed that, while OCT4 and NANOG are co-regulated, SOX2 can be manipulated independently (Wang et al., 2012). In addition, OCT4 expression is supported by β-catenin, likely dependent on WNT signaling duration and dosage (Davidson et al., 2012). OCT4 expression also helps mediate differentiation to mesendoderm (when "high") and extraembryonic fates (when "low") upon exposure to BMP4 (Wang et al., 2012). The β-catenin and OCT4 interplay further drives primitive streak induction (Blauwkamp et al., 2012; Funa et al., 2015). Subsequently, Brachyury (TBXT) interacts with SMAD1 and SMAD2/3 signaling to regulate mesoderm and endoderm expression, respectively (Faial et al., 2015).

We first created a dynamic GRN model consisting of the pluripotency-associated genes OCT4 and SOX2, primitive streak-associated factors TBXT and SOX17, and extraem-bryonic lineage specifier CDX2. We added BMP4 and WNT3 signaling pathways (and their inhibitors NOGGIN and DKK, respectively) by linking gene states to signaling activities (Figure S1A and Table S1). We employed a multi-level logic formalism (Collombet et al., 2017) whereby GRN node states were abstracted as a ternary variable (OFF, LOW, HIGH). This meant that the nodes could be OFF (gene not expressed), LOW (intermediate expression), or HIGH (expression at its highest level). For example, during pluripotency, OCT4 expression was set to HIGH, but for mesodermal cells OCT4 expression turned LOW, with TBXT expression turning HIGH. The GRN logic is defined in Table 1.

We next validated the GRN by comparing the model output against existing data (Tables S2 and S3, Figures S1B–S1D, and Note S1). Importantly, our model captured the duality associated with CDX2 expression as being both extraembryonic mesodermal and trophectodermal. When stimulated with BMP4, the model yielded CDX2+ attractors that were (1) extraembryonic mesoderm associated (i.e., TBXT not OFF) and also had CDX2 as HIGH or LOW, and (2) trophectoderm associated and had CDX2 = HIGH or LOW but TBXT set to OFF (Data S1).

Integrating the GRN, agent, and RD modules

We next discretized thousands of these GRNs explicitly in space by embedding them in cellular "agents." Agents are computational systems that are situated within an environment and capable of engaging in autonomous action within that environment to meet their design objectives (Kaul and Ventikos, 2015). To connect GRN activity to exogenous cues, we coupled the virtual cell with the GRN paradigm to a four-component RD model that simulated the BMP and WNT pathways. We employed the classical RD paradigm (Murray, 2002) and used Gierer and Meinhardt (1972) reaction kinetics to model activator (BMP4 and WNT3) and inhibitor (NOGGIN, DKK) interactions. We used zero-flux boundary conditions.

We validated the RD module by comparing the effective diffusivity ratios of BMP4:NOGGIN and WNT3:DKK used in the RD model (0.4 and 0.36, respectively), which were consistent with experimental estimates (0.38–0.77 and 0.34–0.36, respectively, see Note S2). The architecture of GARMEN is represented in Figure 1A and its implementation shown in Figures 1B and S2A–S2D. Parameters used in this three-layered model and the rationale for their selection are listed in Table 2. Importantly, we did not consider edge sensing (Etoc et al., 2016). Yet, our model recapitulated pSMAD1 and β-catenin profiles accurately (Figures S2E–S2I and Note S3). Agent phenotypes throughout this study were based on the combination of gene states shown in Table S4.

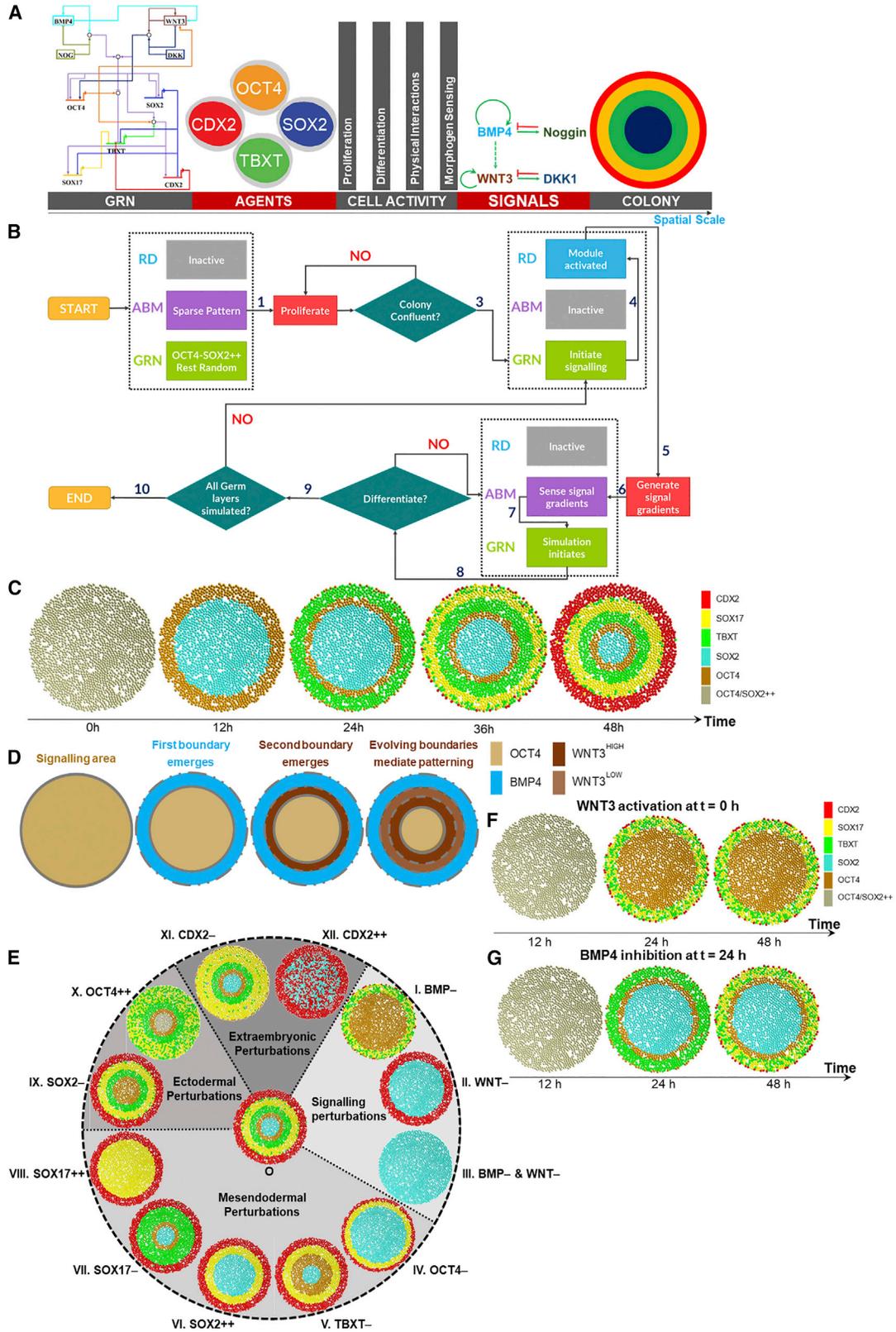To orient the reader, we consistently use descriptors such as computational or state/agent (e.g., computational

Table 1. Gene regulatory network rules regulating differentiation-related cell-fate decision-making

	Expression level	Condition
Morphogen		
BMP4	HIGH	[BMP4] $\geq$ 0.3 ($\sim$25% of peak BMP4 value)
	LOW	0.0 < [BMP4] < 0.3
	OFF	[BMP4] = 0.0
WNT3	HIGH	[WNT3] $\geq$ 0.3 ($\sim$25% of peak BMP4 value)
	LOW	0.0 < [WNT3] < 0.3
	OFF	[WNT3] = 0.0
NOGGIN	HIGH	[NOGGIN] $\geq$ 0.3
	LOW	0.0 < [NOGGIN] < 0.3
	OFF	[NOGGIN] = 0.0
DKK	HIGH	[DKK] $\geq$ 0.3
	LOW	0.0 < [DKK] < 0.3
	OFF	[DKK] = 0.0
Gene (gene category)		
OCT4 (pluripotency TF)	HIGH	(signal BMP4 > 0.0 and signal WNT3 $\leq$ signal DKK) or (signal BMP4 = OFF and signal WNT3 < signal DKK)
	LOW	(TBXT not OFF) or (signal WNT3 > signal DKK) or (signal WNT3 $\leq$ signal DKK and signal BMP4 > signal NOGGIN)
	OFF	signal BMP4 = OFF and signal WNT3 = OFF (conditions that lead to CDX2 = HIGH or SOX17 = HIGH)
SOX2 (lineage TF)	HIGH	signal WNT3 < signal DKK and signal BMP4 < signal NOGGIN
	LOW	signal BMP4 = OFF and signal WNT3 $\leq$ signal DKK
	OFF	TBXT not OFF (conditions leading to TBXT emergence)
TBXT (lineage TF)	HIGH	signal WNT3 > signal DKK and OCT4 not OFF
	LOW	(signal WNT3 > signal DKK and signal BMP4 = OFF and OCT4 not OFF) or (signal BMP4 > signal NOGGIN and OCT4 not OFF and signal WNT3 < signal DKK)
	OFF	SOX2 = HIGH or SOX17 = HIGH or CDX2 = HIGH (or conditions leading to these states)
SOX17 (lineage TF)	HIGH	signal BMP4 = HIGH and signal WNT3 = LOW and signal WNT3 > signal DKK and OCT4 not HIGH
	LOW	(OCT4 = OFF and SOX2 = OFF and signal BMP4 > signal NOGGIN) or (signal BMP4 = OFF and signal WNT3 > signal DKK and OCT4 not HIGH and TBXT not OFF)
	OFF	SOX2 = HIGH or CDX2 = HIGH (or conditions leading to these states)
CDX2 (lineage TF)	HIGH	(signal BMP4 > signal NOGGIN and signal WNT3 < signal DKK and signal WNT3 = LOW and OCT4 not HIGH) or (signal BMP4 = OFF and signal WNT3 > signal DKK and OCT4 = LOW and TBXT not OFF)
	LOW	(signal WNT3 = OFF and signal BMP4 > signal NOGGIN and OCT4 not OFF) or (OCT4 = OFF and signal BMP4 > signal NOGGIN)
	OFF	OCT4 = HIGH or SOX2 = HIGH (or conditions leading to these states)

[BMP4] or OCT4 state or TBXT agent) when sharing *in silico* results, and experimental or *in vitro* or marker expression profile (e.g., *in vitro* pSMAD1 profile, CDX2 expression profile) for *in vitro* results.

GARMEN decision-making

The simulation (Figure 1C and Note S4) was initiated with 1,200 virtual PSCs randomly located on a circular micropattern (step 1). As initial conditions, all GRN nodes were set

(legend on next page)

to OFF, except OCT4 and SOX2, which were HIGH. The virtual cells first proliferated to confluence. Adding proliferation rate in the agent-based modeling (ABM) module allowed us to synchronize the ABM and RD modules in time. Following confluence (step 2), we mimicked the *in vitro* induction of BMP4 by triggering BMP4 activity in the RD module (step 3), which yielded steady-state [BMP4] profiles at ~12 h of physical time (step 4, see Note S3 for the time step). The [BMP4] profile from the RD module was then binned (step 5) and fed to the agents/GRNs (steps 6/7) that led to the activation of WNT activity and emergence of agents transitioning to the mesendoderm state via loss of the SOX2 state (step 8). As BMP4 triggers WNT3 activity, we next simulated the WNT3 profile. The steady state reflected [WNT3] at ~24 h of physical time, which was next fed to the agents, resulting in the emergence of TBXT agents at the colony edge. The non-mesodermal agents at 24 h stay engaged with WNT activity, which we simulated again with the RD module. This yielded a [WNT3] profile at 36 h, which, when fed to the agents, resulted in radial internalization of SOX2 and TBXT agents (step 9, Figure 1B) and the emergence of SOX17 agents at the colony edge. Finally, the WNT activity, by remaining non-mesendodermal, was again simulated in the RD module, yielding the WNT profile at 48 h of physical time. When fed to the agents, this led to a further internalization of the germ-layer states and the emergence of a CDX2 agent population at the edge (step 10, Figure 1B). At the systems level, this radial internalization of germ-layer markers was aided by a progressive OCT4^{high} expression front that translocates radially inward (Figures 1C and 1D and Note S5). Importantly, the CDX2+ agents at the edge were trophectodermal (i.e., their TBXT and SOX2 states were both OFF), corroborating an earlier report (Chhabra et al., 2019).

Exploring the peri-gastrulation phase space

We simulated germ-layer patterns due to biochemical and genomic perturbations in the GRN. This entailed setting the activity levels of the relevant nodes continuously to OFF (shown as "−" in Figure 1E, reflecting *in vitro* gene knockout) or HIGH (shown as "++" reflecting *in vitro* ectopic expression). The pattern phase space is represented in Figure 1E, which shows the impact of perturbations on the patterns, and summarized in Table S5 and Note S6.

Our approach revealed the systems-level impact of genomic and signal perturbations. Specifically, in the BMP4− simulation, the TBXT agents were confined to the colony edge and co-localized with SOX17 agents, possibly suggesting this condition does not commit mesendodermal agents to either mesodermal or endodermal state (Figures 1E and 1F). We also noted a wide spread of agents expressing OCT4 at the colony center. We also noted both observations for the OCT4++ perturbation (pattern X). Together, these suggested that the peri-gastrulation patterns evolve due to interplay between OCT4 and the BMP and WNT pathways.

To investigate this interplay, we tested the consequence of halting the translocating differentiation front. We initiated the model by activating the BMP module, but turning the BMP activity OFF after 24 h in the simulation. Consequently, the TBXT state shifted inward until 24 h, but there was no shift after the BMP module was switched off (Figure 1G). The SOX17- and CDX2-associated agents were both restricted to the colony edge.

In summary, we used a validated hPSC lineage-commitment GRN structure to explore the relationship between GRN wiring and the peri-gastrulation pattern phase space. We found that an inward-traveling differentiation front can be predictively interrupted by altering the input signals or "gene" node activity in the virtual colonies, which we validated next.

Validating GARMEN phase-space predictions via signal perturbations

Validating model predictions when patterning triggered with CHIR

Our model predicted that triggering patterns with WNT3 will limit the mesendoderm-associated (TBXT/SOX17) agents toward the edge (Figure 1F). The simulation also yielded a wide core of SOX2+ agents (Figure 2A). To test these predictions, we initiated differentiation *in vitro* via CHIR-99021 (CHIR). Figure S3A shows all *in vitro* conditions used in this study to test model predictions. Further,

Figure 1. GRNs embedded in agents predicted the systems-level impact of regulatory decision-making
(A) Schematic representing GARMEN's architecture.
(B) The GARMEN decision-making flowchart (see Note S4).
(C) Simulation output capturing peri-gastrulation, including the radial internalization of the OCT4, TBXT, and SOX17 expression (see Note S5).
(D) Visualizing the "stepwise" mechanism of radial internalization of germ-layer markers (see Note S5).
(E) Results of the parametric analysis conducted to link the network structure to the germ-layer pattern phase space (see Table S5 and Note S6).
(F) Simulation output when the model bypasses BMP activity by initiating with WNT activity.
(G) Simulation output when BMP activity is inhibited midway (=24 h of real time) during patterning.

Table 2. Parameters used in the GARMEN

Layer	Parameter	Value	Comments
Agent-based modeling	agent radius	10 μm	measured in this study via internuclear distances (22 ± 5 μm, n = 15)
	colony diameter	930 μm	measured in this study at various time points, n = 10
	agent proliferation rate	12 h	this parameter allowed us to incorporate time in the ABM and thus help synchronize it with the RD module in time
	location of daughter agents	randomly generated based on parent agent coordinates	see "the virtual cell - agent-based modeling" in supplemental experimental procedures
	differentiation probability	15% per iteration (data-constrained free parameter)	based on *in vitro* observations; 1% differentiation probability yielded patterns too quickly, and 30% probability led to delayed patterning; 15% probability, however, led to emergence of ring-like patterns equivalent to 12 h of physical time, which agreed with empirical observations
	initial agent number in colony	1,200	arithmetic calculation based on the number of cells seeded per microtiter plate well: 60,000/well
	initial agent state	pluripotent (OCT4 and SOX2 set to HIGH)	based on experimental condition
Reaction diffusion	α/κ	0.1 (free parameter)	see "linearisation" in supplemental experimental procedures
	β/ω	1 (free parameter)	see "linearisation" in supplemental experimental procedures
	BMP4:NOGGIN effective diffusivity	0.4 (free parameter)	see "linearisation" in supplemental experimental procedures
	WNT3:DKK effective diffusivity	0.36 (free parameter)	see "linearisation" in supplemental experimental procedures
	initial [BMP4]	random distribution	–
	initial [Noggin]	random distribution	–
	initial [WNT3]	[BMP4] at steady state	this linked [BMP4] to the WNT3 RD module
	initial [DKK]	random distribution	–
	boundary condition	Neumann (zero flux)	based on experimental conditions that we do not add BMP4 in the microtiter plate wells after initiating differentiation

to test the statistical validity of model predictions, we calculated the median radial locus of the peak *in vitro* marker expression (peak locus, henceforth) for all conditions. Colonies exposed to both CHIR ± NODAL *in vitro* showed the TBXT and SOX17 peak loci significantly closer to the edge (Figure 2A, all p values indicated in the figure henceforth) vs. BMP4 + NODAL (control). SOX2 expression for these colonies extended from the center to the edge vs. control. The peak TBXT and SOX17 loci for CHIR + NODAL were significantly closer to the edge vs. CHIR *in vitro*. This is consistent with our proposed mechanism that, due to NODAL's BMP inhibitory effect (Tewary et al., 2019), the TBXT and SOX17 peak locus will be confined closer to the edge compared with a condition in which a BMP inhibitor is not used (CHIR in this case). The computational model had also predicted fewer CDX2+ agents for the WNT3 condition that will be

restricted to the edge vs. control. While *in vitro* colonies treated with CHIR ± NODAL yielded CDX2 cells that were radially internalized vs. control, they yielded significantly less CDX2 expression vs. control. This can be observed in Figure S3B, which compares the area under the CDX2 radial profile curves (CDX2 area under the curve [AUC]) for the different conditions.

Validating model predictions when BMP4 patterning is inhibited midway

The predictions (Figure 1G) were tested *in vitro* by initially stimulating the hPSC colonies with BMP4 + NODAL but switching to medium containing the BMP inhibitor LDN193189 (LDN), which also lacked BMP4 and NODAL, for 24 h. Consistent with our prediction, we noticed the peak TBXT, SOX2, and SOX17 expressions each confined significantly closer to the edge vs. control (Figures 2A, S3C, and S3D). We also exposed micropatterns to medium
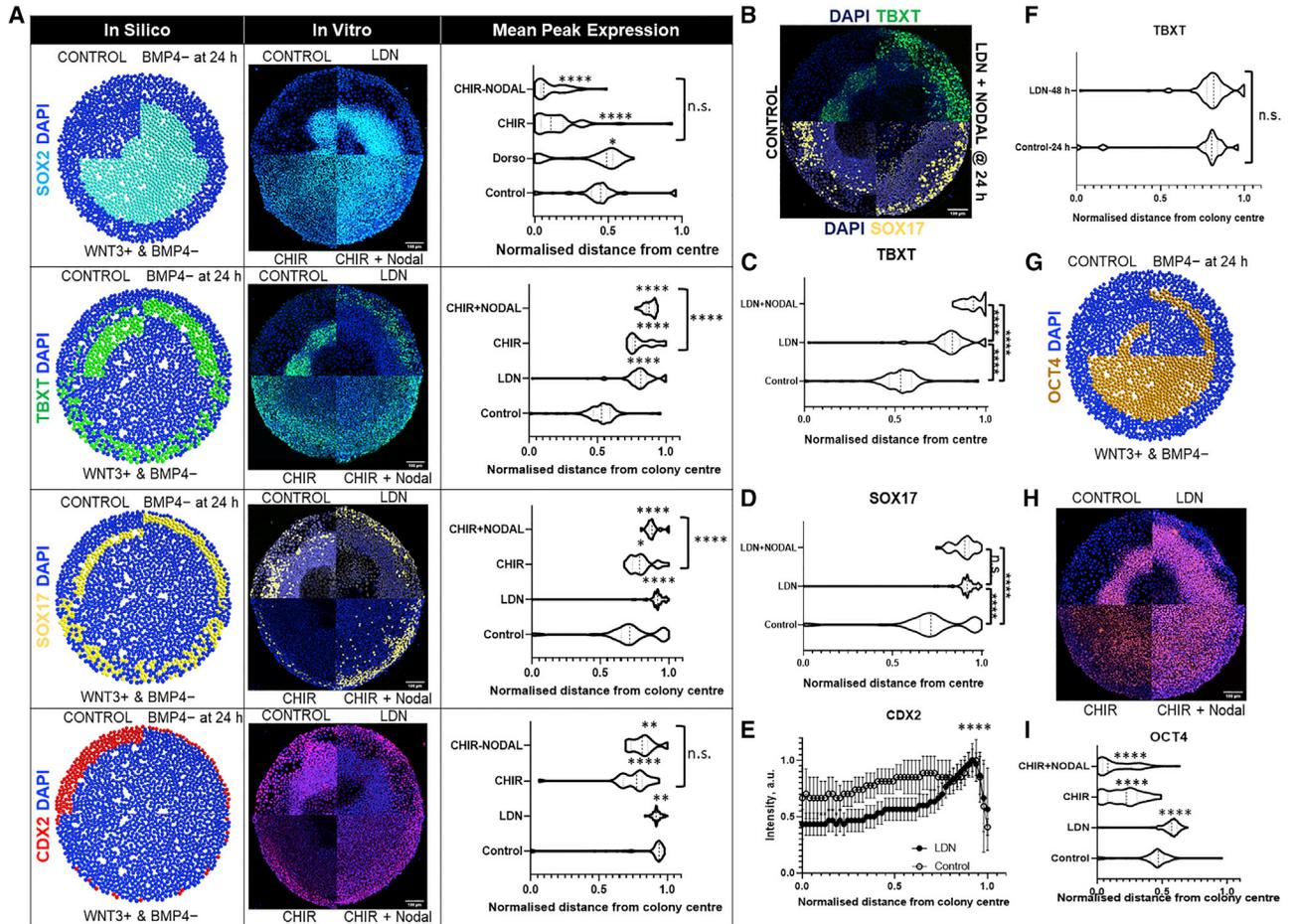
Figure 2. Signal perturbations halt the internalizing OCT4 front to alter peri-gastrulation patterning

(A) Comparison of *in silico* (left column) predictions against *in vitro* data (middle and right columns). The images under the middle column represent a composite of hPSC colonies maintained under different conditions. The plots under the right column compare the peak loci of indicated markers for the four conditions. For TBXT and SOX2: n = 118 (control), 127 (LDN), 35 (CHIR), and 63 (CHIR + NODAL). For SOX17 and CDX2: n = 75 (control), 73 (LDN), 28 (CHIR), and 19 (CHIR + NODAL).

(B) Composite showing TBXT and SOX17 expression for colonies exposed to LDN + NODAL at 24 h vs. control.

(C) The graph compares the peak TBXT loci for the control, LDN only-treated colonies, and LDN + NODAL-treated colonies; n = 118 (control), 127 (LDN), 42 (LDN + NODAL).

(D) The graph compares the peak SOX17 loci for the control, LDN only-treated colonies, and LDN + NODAL-treated colonies; n = 75 (control), 73 (LDN), 32 (LDN + NODAL).

(E) Radial CDX2 expression profile (mean values represented by circles and standard deviation by error bars) comparison between the control (n = 75) and LDN (n = 63)-treated hPSC colonies.

(F) Comparison of the peak loci of TBXT expression in control at 24 h (n = 16) vs. LDN (n = 127)-treated cells after 48 h; p = 0.68.
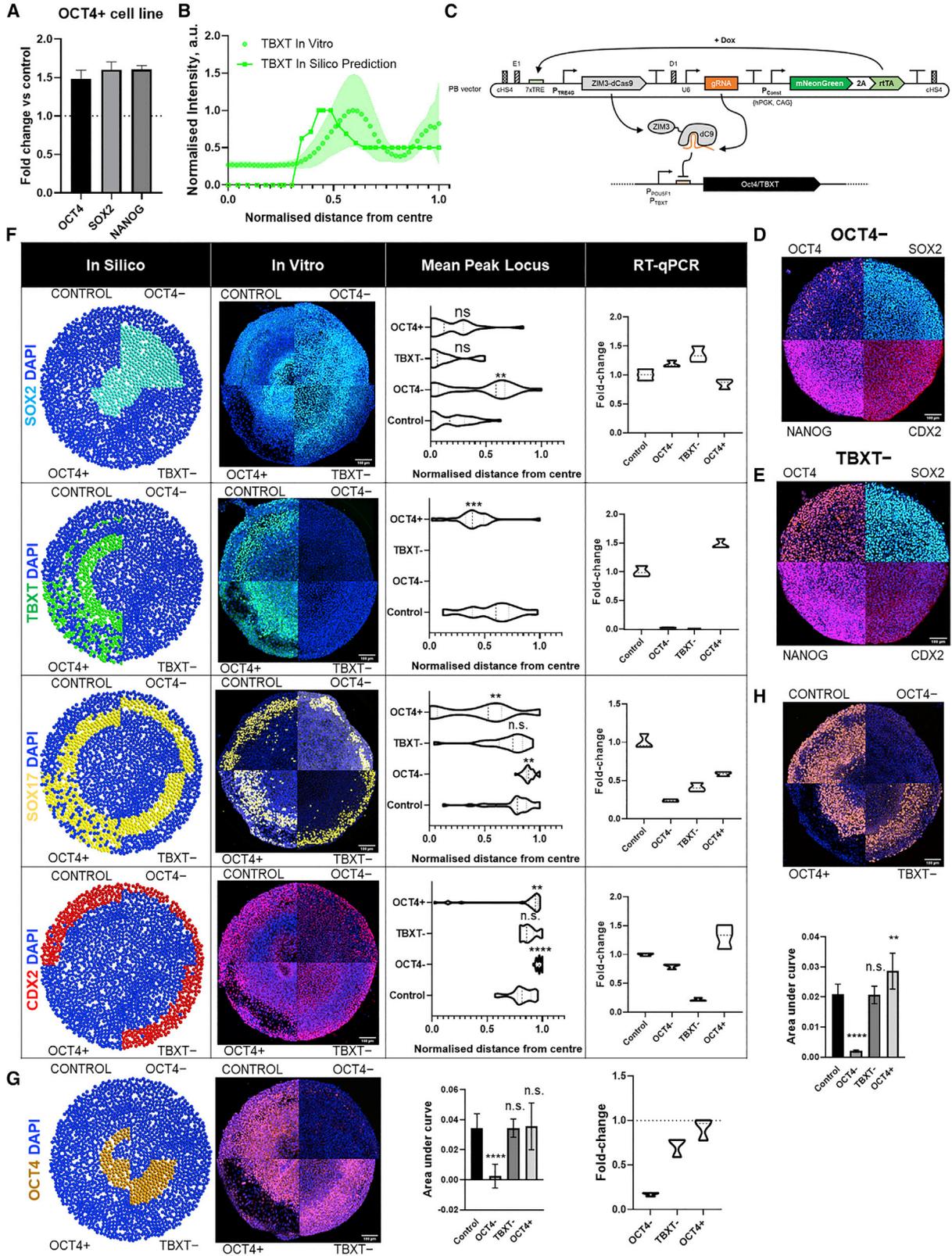
(G) Composite showing the *in silico* prediction of OCT4 expression due to the different signal perturbations.

(H) Composite showing the actual OCT4 expression *in vitro* resulting from exposure to the different conditions.

(I) Comparison of the peak OCT4 loci for the various conditions; n = 118 (control), 127 (LDN), 35 (CHIR), and 63 (CHIR + NODAL); *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; n.s., not significant; n, number of colonies, represents technical replicates. The p values represent comparison with the control condition, except where direct comparisons are specified.

containing LDN + NODAL at 24 h post-differentiation, which similarly yielded peak TBXT and SOX17 loci that were confined to the edge vs. control (p < 0.0001 for both; Figures 2B–2D). The computational model predicted that inhibiting BMP4 at 24 h will lead to fewer CDX2-asso-

ciated agents (55 vs. 618 control agents) that will be confined to the edge (Figure 2A). *In vitro*, we found that the peak CDX2 locus for both the control and the LDN conditions was located close to the edge (Figure 2A). However, the two conditions yielded significantly different data. To

(legend on next page)

understand this, we examined the *in vitro* radial CDX2 expression profiles for both conditions (Figure 2E) and noted that CDX2 expression remained significantly high toward the colony center for control hPSC colonies vs. the LDN treatment, consistent with our model (Figure 2A).

To confirm that we were able to halt marker internalization, we compared the peak TBXT locus in hPSC colonies kept in the control medium for 24 h vs. those exposed to LDN for 24 h (after 24 h of BMP4 exposure). TBXT loci for the two conditions were not significant (p = 0.68, Figure 2F), confirming the computational prediction.

Impact of these perturbations on OCT4 expression
The *in silico*/*in vitro* OCT4 state/expression (Figures 2G–2I) was consistent with the *in silico*/*in vitro* (TBXT, SOX17, CDX2) state/expression patterns (Figure 2A). Exposing the colonies to LDN led to exteriorization of the peak OCT4 locus vs. control (p < 0.0001, Figures 2H and 2I), consistent with the computational prediction. Further, exposing *in vitro* colonies to CHIR led to a spatially wide OCT4 expression profile (Figures 2G–2I, S3E, and S3F), as predicted by GARMEN (Figure 1F).

In summary, these data validate GARMEN's ability to predict changes in patterning based on altered signaling dynamics. Changes in the OCT4 profile due to exposure to CHIR ± NODAL and LDN led us to hypothesize that we can predictively control the mesendoderm expression by manipulating OCT4.

Validating GARMEN phase-space predictions via transcriptional perturbations
Given our hypothesis, we focused on testing the predictions emerging from OCT4-relevant perturbations, i.e.,

OCT4− (pattern IV) and OCT4++ (pattern X). To ensure our model was not biased toward OCT4-oriented perturbations, we also tested the TBXT− perturbation (pattern V).

Cell lines used to test the OCT4++, OCT4−, and TBXT− perturbations
We used a human induced PSC (hiPSC) line to test the OCT4++ perturbation. qRT-PCR analysis revealed that the pluripotency gene (OCT4, SOX2, NANOG) expression was upregulated in this line by 50% vs. the control hPSC line (Figure 3A). We, thus, refer to it as the OCT4+ line. GARMEN predicted the OCT4++ perturbation will yield the same pattern order as control, except that there will be a long tail of TBXT+ agents extending toward the colony edge. This was confirmed by superimposing the mean *in silico* and *in vitro* TBXT radial profiles (Figure 3B).

To test the computational predictions for the OCT4− and TBXT− perturbations, we engineered hPSC lines with transposons that encode genetic circuits (Figures 3C, S4A, and S4B) for doxycycline (Dox)-inducible silencing of OCT4 and TBXT (OCT4− and TBXT− lines, henceforth). The expression of the engineered circuits can be observed via the GFP tag shown in Figure S4C. We chose Dox-inducible OCT4 silencing (instead of knocking out OCT4) to prevent spontaneous differentiation (and total loss of pluripotency) of cells before the induction of patterning.

We first checked that the OCT4− and TBXT− colonies had not spontaneously differentiated at confluence by assessing the pluripotency markers. As expected, OCT4 expression was drastically reduced in the OCT4− line vs. NANOG and SOX2 (Figures 3D, S4D, and S4E showing qPCR data). We also noted slightly elevated SOX2 gene expression vs. control, but similar NANOG levels vs. control

Figure 3. *In silico* patterning phase space successfully recapitulated *in vitro*
(A) The histogram shows fold change in the expression of the pluripotency genes for OCT4+/hiPSC micropatterns. The dotted line represents the normalized median pluripotency gene expression for the control micropatterns generated with the hPSC line (i.e., relative OCT4, SOX2, and NANOG expression for the control cell line = 1).
(B) The graph compares the radial profiles of the computational TBXT+ agents in the virtual colony vs. *in vitro* TBXT expression (n = 71).
(C) The schematic of the gene circuit used to engineer the OCT4− and TBXT− cell lines.
(D and E) The composites show the radial profiles of the pluripotency genes and CDX2 for colonies generated using the (D) OCT4− and (E) TBXT− cell lines at t = 0 h.
(F) Comparison of *in silico* predictions vs. *in vitro* data. The images under the second column represent a composite of colonies with different perturbations. The plots under the third column compare the peak loci of indicated markers for the four conditions. The fourth column shows fold change in gene expression of the respective markers vs. control using qPCR. For SOX2 and TBXT: n = 38 (control), 54 (OCT4−), 12 (TBXT−), and 25 (OCT4+). For SOX17: n = 30 (control), 35 (OCT4−), 22 (TBXT−), and 25 (OCT4+). For CDX2: n = 40 (control), 22 (OCT4−), 11 (TBXT−), and 31 (OCT4+).
(G) Composite showing the GARMEN-predicted vs. *in vitro* OCT4 expression for the different colonies. The histograms show the AUC values calculated based on the radial OCT4 expression profile for the different conditions; n = 40 (control), 22 (OCT4−), 11 (TBXT−), and 31 (OCT4+). The violin plots show the fold change in OCT4 expression vs. control at 48 h. The horizontal line indicates the median OCT4 expression for the control colonies.
(H) Composite showing the EOMES expression for the different colonies. The histograms show the AUC values calculated based on the radial EOMES expression profile for the different conditions; n = 8 (control), 7 (OCT4−), 8 (TBXT−), and 7 (OCT4+). **p < 0.01; ***p < 0.001; ****p < 0.0001; n.s., not significant; n, number of colonies, represents technical replicates. The p values represent comparison with the control, except where a direct comparison is specified.

(Figure S4E). For the TBXT− line, all pluripotency markers showed uniform spatial distribution (Figures 3E and S4F). qPCR data (Figure S4G) confirmed similar OCT4 and NANOG levels vs. control, but a 2-fold increase in SOX2 levels. Further, we measured CDX2 expression levels via qPCR, which was significantly elevated in both cell lines. However, we observed no nuclear CDX2 expression for either via immunofluorescence microscopy (Figures 3D and 3E) vs. control (Figure S4H, control: CDX2 expression in parental colonies at 48 h post-differentiation).

Validating the impact of OCT4 perturbations on TBXT and SOX2 expression

GARMEN predicted (Figure 1E) that silencing either OCT4 or TBXT would lead to loss of TBXT+ agents. This was indeed the case *in vitro* (Figure 3F). AUC calculations confirmed significant loss of TBXT for both perturbations vs. control (Figure S4I), which was also reinforced by the qPCR analysis. The long TBXT expression tail can be observed in the OCT4+ colony composite. Further, GARMEN predicted that silencing OCT4 would lead to a wider spread of SOX2+ agents vs. control, but silencing TBXT or overexpressing OCT4 would not have an impact on SOX2+ agents vs. control. This was also confirmed *in vitro*.

Validating the impact of OCT4 perturbations on SOX17 expression

GARMEN predicted (Figure 1E) that silencing OCT4 would yield SOX17+ agents, despite the loss of TBXT agents, although the mean locus of SOX17+ agents was closer to the colony edge. This was confirmed *in vitro* (Figure 3F). In addition, the mean peak SOX17 locus for the OCT4− line was significantly toward the edge vs. control. GARMEN further predicted that silencing TBXT alone would not have an impact on SOX17 agents computationally vs. control. This was also confirmed *in vitro* (peak SOX17 locus for TBXT− colonies vs. control p > 0.99). Finally, as predicted, the OCT4+ micropatterns yielded a wide spread of SOX17+ cells vs. control.

Validating the impact of OCT4 perturbations on CDX2 expression

GARMEN predicted (Figure 1E) that the OCT4− line would yield CDX2+ agents that would be restricted to the edge computationally. This was confirmed *in vitro* (p < 0.0001 vs. control). GARMEN further predicted that silencing TBXT would yield no significant change in the location of CDX2+ agents computationally. *In vitro*, the peak CDX2 locus for TBXT− micropatterns was not significant vs. control (p = 0.95). The qPCR analysis, however, showed a significant loss in CDX2 gene expression for the TBXT− line.

Finally, GARMEN predicted there would be no CDX2+ agents in the OCT4++ perturbation computationally, although CDX2 expression was observed *in vitro*. This discrepancy needs to be contextualized. First, the predic-tions are relative to control, i.e., the OCT4++ condition would yield less CDX2 expression vs. control, confirmed by the AUC analysis (Figure S4J) that shows significantly less CDX2 expression vs. control. Second, for the OCT4++ perturbation, GARMEN assumed the OCT4 levels in the agents to always be 100% higher vs. control. But the hiPSC line we used had only 50% higher OCT4 levels pre-differentiation (Figure 3A). However, engineering a cell line with continuously HIGH OCT4 will yield the same pattern by triggering WNT following exposure to BMP4. Consequently, mesendoderm will emerge but will not diverge, given HIGH OCT4 expression, which will likely inhibit CDX2 expression.

Impact of transcriptional perturbations on post-differentiation OCT4 and EOMES expression

To confirm that these results were driven by OCT4 activity, we tested the OCT4 expression for each of these experi-mental conditions (Figure 3G). OCT4 expression was significantly lost in the OCT4− colonies (p < 0.001 calcu-lated for AUC), although the rest showed similar OCT4 expression. The qPCR analysis further confirmed that this was the case. The concomitant loss of TBXT expression in OCT4− colonies suggested that OCT4 is central to the emergence of mesendoderm. To test this, we hypothesized that silencing OCT4 will lead to the loss of eomesodermin (EOMES), a key mesendoderm regulator. Indeed, OCT4− micropatterns showed no EOMES (p < 0.0001 vs. control), compared with other cell lines that showed EOMES in the same annulus as TBXT (Figure 3H, p = 0.256 and p = 0.221 for peak EOMES vs. TBXT loci for control and OCT4+ cell line, respectively). EOMES expression was also found in colonies created using the TBXT− cell line, suggesting the need for OCT4 to trigger mesendoderm expression.

In summary, these data highlight that we can predic-tively disrupt mesendoderm emergence by disrupting OCT4 expression. Further, successful predictions for TBXT− condition highlight the unbiased and gene-agnostic nature of this model/GARMEN.

Validating novel insights gained via GARMEN

The formation of three distinct germ layers is one of the key signatures of gastrulation. However, the "distinctness" of germ layers is yet to be quantified in any meaningful way in the context of hPSC-based gastruloids. We propose that significantly separated SOX2, TBXT, and SOX17 peak loci *in vitro* represent a matured gastrula state where the three germ layers have diverged and are distinct. Figure 4A shows the significantly separated peak loci for the three markers for the control condition (BMP4 + NODAL).

Impact of varying OCT4 expression on mesendoderm divergence

GARMEN suggested that this divergence of germ layers is regulated by OCT4 expression (Figures 1E–1G). For
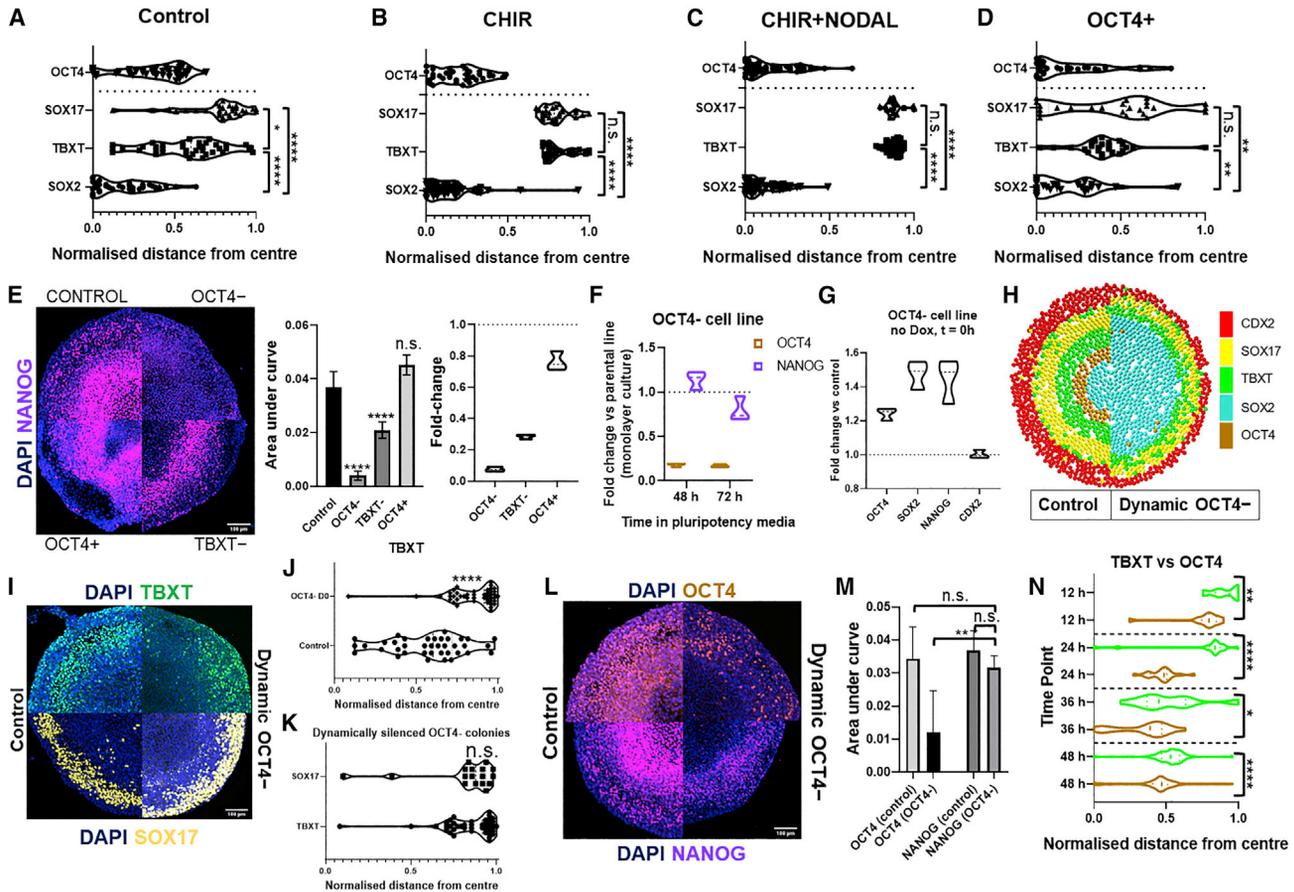
Figure 4. OCT4 serves as a master regulator of peri-gastrulation patterns in hPSCs

(A–D) The graphs compare peak SOX2, TBXT, and SOX17 loci for colonies generated using the parental line and exposed to (A) BMP4 + NODAL, (B) CHIR, (C) CHIR + NODAL, and (D) OCT4+ colonies exposed to BMP4 + NODAL. The peak OCT4 locus is shown for reference.

(E) Composite showing the NANOG expression for the different colonies. The histograms show the AUC values calculated based on the radial NANOG expression profile; n = 8 (control), 5 (OCT4−), 5 (TBXT−), and 3 (OCT4+). The violin plots show the fold change in NANOG expression vs. control at 48 h. The horizontal line indicates the median NANOG expression for the control colonies.

(F) The graph compares the fold change in OCT4 and NANOG gene expression with control (median OCT4 and NANOG gene expression value highlighted by the dotted line) at two different time points.

(G) Fold change in gene expression in colonies generated using the OCT4− cell line that was not exposed to Dox at t = 0 h vs. control (median gene expression values represented by the dotted line).

(H) Composite showing patterning predicted by GARMEN for control vs. dynamically inhibited OCT4 perturbation.

(I) TBXT and SOX17 expression for the OCT4− colonies where OCT4 expression was dynamically inhibited vs. control.

(J) Graph comparing the peak TBXT loci for the dynamically inhibited colonies (n = 45) with control (n = 38).

(K) Graph comparing the peak TBXT (n = 45) with SOX17 (n = 17).

(L) Composite showing OCT4 and NANOG expression for the dynamically inhibited OCT4− colonies vs. control.

(M) The histograms compare the AUC values calculated based on the OCT4 and NANOG radial expression profiles for the dynamically inhibited OCT4− colonies with control. For OCT4: n = 52 (OCT4−) and n = 40 (control). For NANOG: n = 5 (OCT4−) and n = 8 (control).

(N) The graph compares peak TBXT with peak OCT4 loci for hPSC colonies exposed to BMP4 + NODAL for 48 h across time. For 12 h: n = 12 (TBXT and OCT4). For 24 h: n = 19 (TBXT) and n = 16 (OCT4). For 36 h: n = 19 (TBXT and OCT4). For 48 h: n = 118 (TBXT and OCT4). *p < 0.05; **p < 0.01; ****p < 0.0001; n.s., not significant; n, number of colonies, represents technical replicates. The p values represent comparison with the control, except where a direct comparison is specified.

example, in GARMEN, the OCT4− state led to loss of TBXT agents, the OCT4++ state yielded co-localized TBXT and SOX17 agents, and exposure to WNT3 also resulted in broad OCT4 expression radius and colocalized TBXT and

SOX17 agents. To test this *in vitro*, we compared the peak loci of all three markers for the following conditions: (1) CHIR- and (2) CHIR + NODAL-induced patterning of hPSC colonies and (3) patterning of the OCT4+ colonies

via BMP4 + NODAL. The peak OCT4 locus for all conditions is shown for reference (Figures 4B–4D). Compared with control, all three conditions yielded patterns where peak TBXT and SOX17 loci are not significantly apart. The only commonality in all three conditions is a wider OCT4 spatial expression covering more than half of the colony radius. Taken together, our data suggest that OCT4 is central to the divergence of mesendoderm to meso- and endoderm.

Contribution of OCT4 vs. NANOG expression in mediating peri-gastrulation

We next explored whether OCT4 regulates patterning directly or indirectly via NANOG, which has been implicated in triggering mesoderm. We noted uniform radial NANOG expression in the OCT4− colonies at t = 0 h (Figure 3D), which was lost by 48 h post-differentiation (Figure 4E, p < 0.0001 vs. control based on AUC). All other cell lines, including TBXT− (albeit significantly less so vs. control, p < 0.0001 based on AUC), retained NANOG+ cells and showed spatially differential NANOG expression. These data suggested that OCT4 potentially regulates NANOG during the emergence and divergence of mesendoderm.

To disentangle the contributions of OCT4 and NANOG, we inhibited OCT4 expression dynamically during patterning to limit the OCT4-driven inhibition of NANOG expression. Our qPCR data showed that the gene circuit used to engineer the OCT4− cell line requires 24 h to reduce OCT4 expression in cells maintained under pluripotent conditions by >50%, during which time NANOG expression remains comparable to that of the control cell line (Figure 4F). While in previous experiments we had added Dox to OCT4− colonies shortly after cell seeding, for this experiment we added Dox to colonies along with the differentiation medium (Figure S3A). This meant that the colonies would not lose significant OCT4 expression due to the synthetic circuit for the first 24 h and, in turn, NANOG for that duration, which we confirmed via qPCR (Figure 4G). The CDX2 expression in these cells was also comparable to control.

We predicted (Figure 4H) that dynamically silencing OCT4 would lead to the emergence of TBXT+ agents, but these agents would not internalize radially. This prediction was successfully validated in vitro (Figure 4I). The peak TBXT locus for the dynamically silenced OCT4− colonies was significantly toward the edge vs. control (Figure 4J). Importantly, the peak TBXT vs. SOX17 loci were not significantly apart (p = 0.984, Figure 4K). To confirm that this was due to the silencing of OCT4 and not the potential downstream loss of NANOG, we compared the AUC for the OCT4 and NANOG profiles generated from these colonies with control (Figure 4L). Figure 4M shows that the NANOG AUC for the dynamically silenced OCT4− col-

onies was not significantly different from control (p = 0.8353), although the OCT4 AUC was significantly reduced vs. control. Taken together, our data support the role of OCT4 as a master regulator of peri-gastrulation patterning in the human context.

Spatial OCT4 vs. TBXT profiles in vitro vs. in vivo

Our radially internalizing OCT4 front hypothesis (Figure 1D) suggests spatial delineation of the germ layers, particularly mesendoderm, as a potential role for OCT4. To test this, we investigated how the OCT4 vs. TBXT spatial expression varies over time in micropatterned colonies, and whether this arrangement is recapitulated in vivo. We found that, following the emergence of TBXT expression, the peak OCT4 loci lie significantly inside of the peak TBXT loci across all time points (Figure 4N). This suggests that OCT4 mediates the TBXT expression boundary in these micropatterns. Given that the ectoderm-like SOX2+ cells in hPSC micropatterns are located at the colony center, this region represents the anterior part of the colony (vs. the TBXT annulus that represents the posterior). As OCT4 in the hPSC colonies is significantly close to the ectoderm-like region (Figure 4N), it is anterior to TBXT, consistent with a mouse gastrula.

To summarize, we validated that OCT4 potentially acts as a master regulator of germ-layer patterning in hPSCs. The successful in vitro validation of computational insights highlights the power of our multiscale approach to discovering novel mechanisms of cell-fate control.

DISCUSSION

To build GARMEN, we integrated and validated three hierarchical levels of cell-fate control: regulatory, cellular, and microenvironmental. A key validation of our GRN was its ability to reflect the duality of CDX2 as both an extraembryonic mesodermal and a trophectodermal marker. The ABM is validated by the numerous in silico composite figures that were compared with their in vitro counterparts. Finally, our RD model successfully predicted the radial switch between peak OCT4 and β-catenin loci between 24 and 48 h (Note S7 and Figures S5A–S5D). This three-layered multiscale model, which is greater than the sum of its components, successfully predicted patterning for 19 signal and "transcriptomic" perturbations (Table S6).

But, how should the GRN nodes be interpreted? This is context dependent. We took these nodes to specifically represent the genes. However, in the case of CDX2, with its dual characteristic, what would a CDX2+ state imply: an extraembryonic mesodermal or a trophectodermal state? This is where our approach to embed these GRNs into agents proves optimal. If the TBXT state of an agent is OFF, but CDX2 set to HIGH, then that agent is
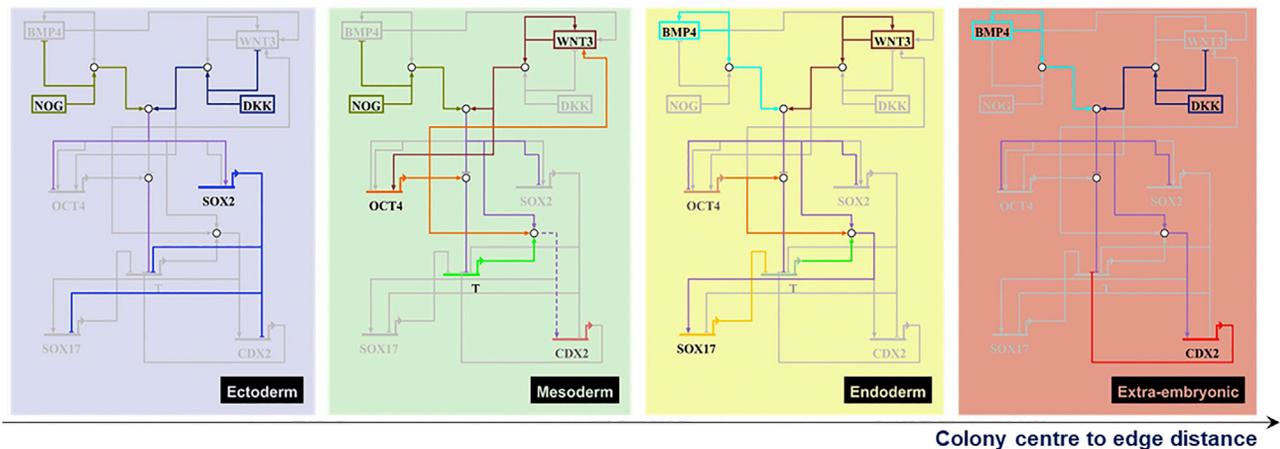
Figure 5. Peri-gastrulation patterning is mediated by spatial signals that engage the GRN in a spatially differential manner
The figure contextualizes the effects of exogenous signals on single-cell GRNs during peri-gastrulation. Different germ-layer fates are symbolized with colored boxes. Active genes and interactions that lead to a particular germ-layer fate are shown in color, whereas genes and interactions that are inactivated due to the signaling context are shown in gray. This schematic highlights how different signal inputs activate and silence different topological paths within a master GRN to reveal spatially distinct peri-gastrulation patterns.

trophectodermal. However, if the TBXT state of an agent is not OFF, and the CDX2 state of the same agent is not HIGH, then that agent is mesodermal. These rules are summarized in Table S4.

Computationally, we overcame the significant impediment of uniting these paradigms in time. The GRN and ABM models are inherently devoid of time vs. RD models that have implicit time. We achieved this by incorporating agent proliferation (i.e., *rate* of change of agent population). Consequently, we captured patterns that are dynamically relevant, i.e., they contain the information that predicts future evolution of the system: a signature of emergent phenomena (Rosas et al., 2020). For example, the status of WNT active agents (e.g., at 24 or 36 h) or OCT4 expression (whether OFF or dynamically diminishing) is key to predicting patterns at subsequent time steps.

We diverged from previous studies in three ways. First, NANOG was not included, as we did not anticipate it having a direct role in lineage commitment (Wang et al., 2012), confirmed via the dynamic OCT4 inhibition experiment. Second, we ignored NODAL, because it emerges and becomes independent after 30 h post-differentiation (Chhabra et al., 2019). Its activity, thus, falls outside the range of the time domain we aimed to capture: the first 36 h when mesendoderm emerges and starts diverging. Third, our model revealed that edge sensing is not critical to peri-gastrulation because donut-shaped colonies always expressed TBXT at their outer edges irrespective of the donut inner diameter (Figure S6 and Note S8).

Our work reveals how the environmental signals are processed by the GRN, in each cell in a spatially differential manner to yield germ-layer patterns (Figure 5). This dem-

onstrates how developmental waves (of OCT4 and WNT3) emerge in hPSC colonies during gastrulation. While Chhabra et al. (2019) proposed an RD model of WNT internalization, we demonstrate how GRNs are responsible for the emergence and maturation of these waves, and how they can be predictively manipulated. This has implications in diverse developmental contexts.

Our model further showed the role of OCT4 as a master regulator of the emergence of the mesendoderm and its divergence to meso- and endoderm in the human context. We validated this experimentally by generating a quantitative index of germ-layer "distinctness" in the context of peri-gastrulation patterns: statistically significant separation of the peak SOX2, TBXT, and SOX17 loci. We anticipate this index will be used to quantitatively identify the accuracy of 3D embryoids in recapitulating gastrulation.

Moreover, our model suggests a biphasic interplay between WNT signaling and OCT4 expression whereby loci with peak WNT activity yield mesendoderm but maintain OCT4 expression at loci where WNT is not peaking. Davidson et al. (2012) had reported the impact of WNT3 in triggering mesoderm, although their data also revealed that if WNT is inhibited or if its concentration is low, it leads to more OCT4+ cells vs. control in the short term (<2 weeks). WNT emerges in our system ~24 h post-differentiation and is fully active for only 24 h, which is extremely short term. Thus, our data and assumption are consistent with Davidson et al. (2012).

Fundamentally, our model shows that cells decode position despite the modest signal gradients available to them by considering local gradients cumulatively (Figure S7 and Note S9). This robustly supports the view that, while

genes act as the cell's computational machinery and the microenvironment offers the appropriate context, it is the cell that acts as the functional executive unit of organogenesis. Our approach, thus, offers a formal framework to understand how positional information (Wolpert, 1969) is implemented in cellular systems. Last, we show how an integrated approach is a practically tractable way to understand and predict *a priori* how lower-level interactions lead to higher-level structure and function.

We acknowledge limitations to our model. First, our GRN is minimalistic and can be extended by adding other genes and signals. Inclusion of NODAL, for example, will help capture observed differences between the CHIR and the CHIR + NODAL conditions and parallel developmental milestones, e.g., pre-neurulation. This can be done by adding an RD model of NODAL dynamics and GRN rules that link NODAL gradients to existing nodes. Second, as GARMEN simulates the activity of thousands of interacting GRNs, the number of nodes should be kept low. This, however, is not a barrier, as we showed (Heydari et al., 2022) that even for multistep differentiation typically a limited set of genes can play a significant role in differentiation. Third, our RD model is isotropic, yielding centro-symmetric solutions, which means that spontaneous symmetry breaking in 3D is likely to be challenging. This could potentially be fixed by allowing boundary conditions to self-organize. Last, GRN creation is currently manual and can be automated via GRN inference tools, e.g., IQCell (Heydari et al., 2022) or GENIE3 (Huynh-Thu et al., 2010). We recommend rigorous pruning of automated networks by considering only those genes that significantly contribute to differentiation to ensure tractability over multiple scales.

Looking ahead, GARMEN is optimally positioned to quantitatively understand two fundamental questions: how do multiscale interactions lead to biological function, and how do they lead to pathology? This will inevitably allow us to generate novel mechanistic insights into multiscale biology (as we demonstrated), engineer cell differentiation strategies to enable development of optimal cell therapies (Prochazka et al., 2017), and conduct patient-specific modeling to tailor interventions to the patient's genomic profile (Kaul, 2019).

EXPERIMENTAL PROCEDURES

Resource availability

Corresponding author
Requests for code should be directed to Himanshu Kaul (himanshu.kaul@leicester.ac.uk).

Materials availability
Plasmid sequences are listed in Supplemental Cloning Information. Plasmids and cell lines are available upon request.

Data and code availability
The accession number for GARMEN-related files reported in this paper is [Figshare]: 10.25392/leicester.data.21375642. Files are available for research purposes only.

Computational
Refer to the supplemental experimental procedures for details.

Experimental

hPSC culture
The CA1 hPSC (gift from Dr Andras Nagy) and the monoallelic mEGFP-tagged β-actin WTC hiPSC (AICS-0016 at Coriell Institute) lines, generated using the WTC-11 hiPSC line (GM25256, Coriell Institute), were cultured using Geltrex (1:50 dilution) coated on six-well tissue culture plates with mTeSR medium (StemCell Technologies). CA1 cells were passaged with ReLeSR (StemCell Technologies) and AICS-16 with StemPro Accutase (Thermo Fisher Scientific). To increase cell viability, ROCK inhibitor Y-27632-supplemented mTeSR was used for 24 h post-passaging. We employed daily medium replacement until 80% confluence was reached.

Plasmid construction and validation
The sequences and descriptions of all plasmids are provided in the Data S2 shows PiggyBac transposons used for Dox-inducible silencing of OCT4 or TBXT. Plasmid vectors encoding the transposase, transposon backbone, and related vectors for mammalian modular cloning (MoClo) (Duportet et al., 2014; Jones et al., 2020) were a gift from Ron Weiss. Gene silencing was actuated by Tet-regulated ZIM3-dCas9, a gift from Mikko Taipale (Addgene plasmid 154472; Alerasool et al., 2020). This sequence was amplified by PCR and cloned into a level 0 (pL0) backbone with Golden Gate using BpiI (Thermo Fisher). Validated gRNA sequences targeting OCT4 and TBXT were obtained from Mandegar et al. (2016) and Libby et al. (2021), respectively. The gRNA sequences were ordered as overlapping ssDNA oligos (IDT) and then annealed and phosphorylated by T4 polynucleotide kinase (NEB) to generate dsRNA with 4-bp sticky ends. The annealed oligos were ligated with T4 ligase (NEB) into a BsmBI (NEB)-linearized scaffold for U6-driven gRNA expression (pRJ 698), based on existing designs (Kiani et al., 2014). All oligo sequences used for cloning are provided in Table S7.

Tet-activated promoters for ZIM3-dCas9 were either TRE3G (TaKaRa Bio) or a variant thereof (TRE4G, Ron Weiss' group) that uses a tighter minimal promoter (Hansen et al., 2014). Constitutive promoters for reversed tetracycline transactivator (rtTA) were either hPGK (modest strength) or CAG (strong). Our rtTA was previously modified from Tet3G (TaKaRa Bio) to use the VP64 activation domain (Jones et al., 2020). Level 1 (pL1) plasmids were assembled from pL0s with Golden Gate using BsaI HF v.2 (NEB). Level 2 (pL2) PiggyBac plasmids were assembled from pL1s with Golden Gate using SapI (NEB). The mappings of parts in each cloning reaction are provided in the supplemental information.

PiggyBac cell line engineering
Approximately 90×10^3 CA-1 cells were seeded into a 12-well plate pre-coated with 0.5 mL Geltrex for 1 h at 37°C. The medium was changed the next day and the cells were transfected with

Lipofectamine Stem (Thermo Fisher) completed with plasmids encoding a puro selection marker, hyperactive PiggyBac transposase (Yusa et al. 2011), and a transposon encoding the Dox-inducible gene circuit to silence either OCT4 or TBXT (Table S8). Twenty-four hours post-transfection and 2 days thereafter, the medium was replaced with fresh medium containing 1 μg/mL Puromycin (Sigma). Ninety-six hours post-transfection, the cells were passaged into fresh six-well plates in puro-supplemented medium. Cells were fed daily in medium without puro until mature colonies formed.

RNA extraction and qRT-PCR

Total RNAs were extracted using the RNeasy Mini Kit (Qiagene) per the manufacturer's protocol. Genomic DNAs were cleared using the TURBO DNA-free Kit (Invitrogen). Total RNA was reverse transcribed into cDNAs with random hexamers by SuperScript III reverse transcriptase (Invitrogen). qPCR was performed using the PowerUp SYBR Green Master Mix (Applied Biosystems) on a QuantStudio 6 Flex System (Applied Biosystems). Selected target genes were amplified by qPCR, and GAPDH was used for normalization. Relative gene expression was calculated by the $2^{-\Delta\Delta Ct}$ method with triplicates. The primers (5′-3′) used are listed in Table S9.

Micropatterned plate preparation

We employed two ways of preparing the micropatterning plates using: (1) polyethylene glycol (Tewary et al., 2017, 2019) and (2) Lipidure. We cleaned glass coverslips with isopropyl alcohol and spin-coated them with Lipidure-CM5206 (Lipidure) (NOF America) in 100% ethanol at 0.2% (mass/volume). We covered the coverslips with 2 mL Lipidure solution and spun them at 2,500 rpm for 30 s. Lipidure-coated coverslips were left in an oven at 54°C for at least 1 h. Micropatterns were created on the Lipidure-coated side of the glass coverslip by 20 min deep UV exposure through a quartz photomask. The glass slides were subsequently attached to the 96-well bottomless plates using epoxy. Carboxyl groups on the photoactivated regions were subsequently activated and coated with Geltrex as described previously (Tewary et al., 2019).

Cell seeding and induction of peri-gastrulation patterning

hPSC micropatterns were formed using N2B27 medium with 50 ng/mL BMP4 and 100 ng/mL NODAL at a density of 60,000 cells/well (Tewary et al., 2017). We used the same approach to create micropatterns using the hiPSCs and the two engineered (OCT4− and TBXT−) cell lines. For the experiment in which we micropatterned the engineered cell lines, following the removal of ROCKi, we left the colonies in the seeding medium for 24 h to account for any cell death due to the activation of the gene circuit.

LDN- and CHIR-induced modulation of peri-gastrulation

LDN (Sigma-Aldrich), reconstituted at 2.5 μM in N2B27 (without BMP4 and NODAL), was used to inhibit the BMP pathway. CHIR (StemCell Technologies) was added to N2B27 medium (without BMP4 and NODAL) and used at 2.5 μM. Table S10 lists the reagents used in this study.

Data acquisition and immunofluorescence imaging analysis

Samples were fixed with 4% paraformaldehyde and permeabilized using 100% methanol. They were exposed to primary antibody at 4°C for 16 h and later immersed in the secondary antibody solution for 2 h at room temperature. Images were acquired using the Cellomics Arrayscan VTI platform and Zeiss LSM 800 confocal microscope at 20×. To extract single-cell data from the images, TargetActivation.V4 bioassay algorithms from the Cellomics Arrayscan VTI platform and CellProfiler (McQuin et al., 2018) were used to identify nuclear regions based on the intensities in the DAPI images. Protein intensities were measured within the nuclear regions for each respective channel. Single-cell data of nuclear location and protein intensity was analyzed via Context Explorer (Ostblom et al., 2019).

Statistics

We used the Mann-Whitney U test and Kruskal-Wallis test with Dunn's *post hoc* tests to compare the peak loci. We used the two-sample Kolmogorov-Smirnov test to compare the TBXT, SOX17, and CDX2 radial profiles for the BMP4 + NODAL and LDN conditions. Given the radial profile of a marker, we calculated the AUC and associated standard error by considering the lowest intensity of the marker as baseline. We next used the AUC values, error, and number of replicates to conduct unpaired, ordinary one-way ANOVAs. $p < 0.05$ served as the threshold of significance for all tests. All analyses were conducted using GraphPad Prism 8.1.

REFERENCES

Alerasool, N., Segal, D., Lee, H., and Taipale, M. (2020). An efficient KRAB domain for CRISPRi applications in human cells. Nat. Methods *17*, 1093–1096.

Blauwkamp, T.A., Nigam, S., Ardehali, R., Weissman, I.L., and Nusse, R. (2012). Endogenous Wnt signalling in human embryonic stem cells generates an equilibrium of distinct lineage-specified progenitors. Nat. Commun. *3*, 1070.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Chhabra, S., Liu, L., Goh, R., Kong, X., and Warmflash, A. (2019). Dissecting the dynamics of signaling events in the BMP, WNT, and NODAL cascade during self-organized fate patterning in human gastruloids. PLoS Biol. *17*, e3000498.

Collombet, S., van Oevelen, C., Sardina Ortega, J.L., Abou-Jaoudé, W., Di Stefano, B., Thomas-Chollier, M., Graf, T., and Thieffry, D. (2017). Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. Proc. Natl. Acad. Sci. USA *114*, 5792–5799.

Davidson, K.C., Adams, A.M., Goodson, J.M., McDonald, C.E., Potter, J.C., Berndt, J.D., Biechele, T.L., Taylor, R.J., and Moon, R.T. (2012). Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. Proc. Natl. Acad. Sci. USA *109*, 4485–4490.

Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naive pluripotency. Science *344*, 1156–1160.

Duportet, X., Wroblewska, L., Guye, P., Li, Y., Eyquem, J., Rieders, J., Rimchala, T., Batt, G., and Weiss, R. (2014). A platform for rapid prototyping of synthetic gene networks in mammalian cells. Nucleic Acids Res. *42*, 13440–13451.

Etoc, F., Metzger, J., Ruzo, A., Kirst, C., Yoney, A., Ozair, M.Z., Brivanlou, A.H., and Siggia, E.D. (2016). A balance between secreted inhibitors and edge sensing controls gastruloid self-organization. Dev. Cell *39*, 302–315.

Faial, T., Bernardo, A.S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G.E., Mascetti, V.L., Trotter, M.W.B., Smith, J.C., and Pedersen, R.A. (2015). Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. Development *142*, 2121–2135.

Funa, N.S., Schachter, K.A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honoré, C., Hansen, K., and Semb, H. (2015). Beta-catenin regulates primitive streak induction through collaborative interactions with SMAD2/SMAD3 and OCT4. Cell Stem Cell *16*, 639–652.

Gierer, A., and Meinhardt, H. (1972). A theory of biological pattern formation. Kybernetik *12*, 30–39.

Glen, C.M., Kemp, M.L., and Voit, E.O. (2019). Agent-based modeling of morphogenetic systems: advantages and challenges. PLoS Comput. Biol. *15*, e1006577.

Hansen, J., Mailand, E., Swaminathan, K.K., Schreiber, J., Angelici, B., and Benenson, Y. (2014). Transplantation of prokaryotic two-component signaling pathways into mammalian cells. Proc. Natl. Acad. Sci. USA *111*, 15705–15710.

Heydari, T., A Langley, M., Fisher, C.L., Aguilar-Hidalgo, D., Shukla, S., Yachie-Kinoshita, A., Hughes, M., M McNagny, K., and Zand-

stra, P.W. (2022). IQCELL: a platform for predicting the effect of gene perturbations on developmental trajectories using single-cell RNA-seq data. PLoS Comput. Biol. *18*, e1009907.

Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS One *5*, e12776.

Jones, R.D., Qian, Y., Siciliano, V., DiAndreth, B., Huh, J., Weiss, R., and Del Vecchio, D. (2020). An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. Nat. Commun. *11*, 5690.

Kaul, H. (2019). Respiratory healthcare by design: computational approaches bringing respiratory precision and personalised medicine closer to bedside. Morphologie *103*, 194–202.

Kaul, H., and Ventikos, Y. (2015). Investigating biocomplexity through the agent-based paradigm. Brief. Bioinform. *16*, 137–152.

Kaul, H., Ventikos, Y., and Cui, Z. (2016). A computational analysis of the impact of mass transport and shear on three-dimensional stem cell cultures in perfused micro-bioreactors. Chin. J. Chem. Eng. *24*, 163–174.

Kiani, S., Beal, J., Ebrahimkhani, M.R., Huh, J., Hall, R.N., Xie, Z., Li, Y., and Weiss, R. (2014). CRISPR transcriptional repression devices and layered circuits in mammalian cells. Nat. Methods *11*, 723–726.

Libby, A.R.G., Joy, D.A., Elder, N.H., Bulger, E.A., Krakora, M.Z., Gaylord, E.A., Mendoza-Camacho, F., Butts, J.C., and McDevitt, T.C. (2021). Axial elongation of caudalized human organoids mimics aspects of neural tube development. Development *148*, dev198275.

Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. Cell Stem Cell *18*, 541–553.

Martyn, I., Kanno, T.Y., Ruzo, A., Siggia, E.D., and Brivanlou, A.H. (2018). Self-organization of a human organizer by combined Wnt and Nodal signalling. Nature *558*, 132–135.

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). CellProfiler 3.0: next-generation image processing for biology. PLoS Biol. *16*, e2005970.

Murray, J. (2002). Spatial pattern formation with reaction diffusion systems. In Mathematical Biology II: Spatial Models and Biomedical Applications, L. Glass and J. Murray, eds. (Springer), pp. 71–140.

Ostblom, J., Nazareth, E.J.P., Tewary, M., and Zandstra, P.W. (2019). Context-explorer: analysis of spatially organized protein expression in high-throughput screens. PLoS Comput. Biol. *15*, e1006384.

Parfitt, D.E., and Shen, M.M. (2014). From blastocyst to gastrula: gene regulatory networks of embryonic stem cells and early mouse embryogenesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. *369*, 20130542.

Prochazka, L., Benenson, Y., and Zandstra, P.W. (2017). Synthetic gene circuits and cellular decision-making in human pluripotent stem cells. Curr. Opin. Struct. Biol. *5*, 93–103.

Rosas, F.E., Mediano, P.A.M., Jensen, H.J., Seth, A.K., Barrett, A.B., Carhart-Harris, R.L., and Bor, D. (2020). Reconciling emergences: an information-theoretic approach to identify causal emergence in multivariate data. PLoS Comput. Biol. 16, e1008289.

Tewary, M., Dziedzicka, D., Ostblom, J., Prochazka, L., Shakiba, N., Heydari, T., Aguilar-Hidalgo, D., Woodford, C., Piccinini, E., Becerra-Alonso, D., et al. (2019). High-throughput micropatterning platform reveals Nodal-dependent bisection of peri-gastrulation-associated versus preneurulation-associated fate patterning. PLoS Biol. 17, e3000081.

Tewary, M., Ostblom, J., Prochazka, L., Zulueta-Coarasa, T., Shakiba, N., Fernandez-Gonzalez, R., and Zandstra, P.W. (2017). A stepwise model of Reaction-Diffusion and Positional-Information governs self-organized human peri-gastrulation-like patterning. Development 144, 4298–4312.

Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. Cell Stem Cell 10, 440–454.

Warmflash, A., Sorre, B., Etoc, F., Siggia, E.D., and Brivanlou, A.H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nat. Methods 11, 847–854.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. 25, 1–47.

Yachie-Kinoshita, A., Onishi, K., Ostblom, J., Langley, M.A., Posfai, E., Rossant, J., and Zandstra, P.W. (2018). Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. Mol. Syst. Biol. 14, e7952.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A 108, 1531–1536. https://doi.org/10.1073/pnas.1008322108.