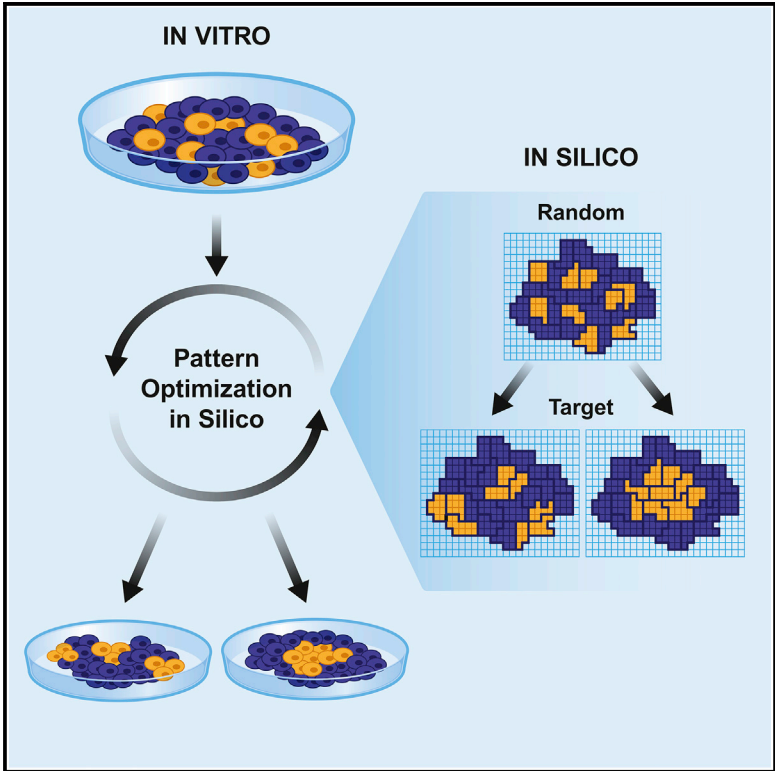


Cell Systems

Automated Design of Pluripotent Stem Cell Self-Organization

Graphical Abstract



Authors

Ashley R.G. Libby, Demarcus Briers, Iman Haghighi, David A. Joy, Bruce R. Conklin, Calin Belta, Todd C. McDevitt

Correspondence

todd.mcdevitt@gladstone.ucsf.edu (T.C.M.), cbelta@bu.edu (C.B.)

In Brief

Libby et al. created a computational model of pluripotent stem cell (PSC) dynamics, enabling a machine learning optimization approach to predict experimental conditions that yield targeted multicellular patterns. These results demonstrate that cellular dynamics can be predicted through model-driven exploration of behaviors, thereby facilitating spatial control of multicellular organization.

Highlights

- Extended cellular Potts model captures pluripotent stem cell organization dynamics
- Machine learning optimization yields conditions for multicellular patterns
- *In silico* predicted experimental parameters generate desired patterns *in vitro*



Automated Design of Pluripotent Stem Cell Self-Organization

Ashley R.G. Libby,^{1,2,8} Demarcus Briers,^{3,8} Iman Haghghi,⁴ David A. Joy,^{2,5} Bruce R. Conklin,^{2,6} Calin Belta,^{4,*} and Todd C. McDevitt^{2,7,9,*}

¹Developmental and Stem Cell Biology PhD Program, University of California, San Francisco, San Francisco, CA, USA

²Gladstone Institute of Cardiovascular Disease, Gladstone Institutes, San Francisco, CA, USA

³Boston University Bioinformatics Program, Boston, MA, USA

⁴Systems Engineering Department at Boston University, Boston, MA, USA

⁵UC Berkeley-UC San Francisco Bioengineering Graduate Program, San Francisco, CA, USA

⁶Departments of Medicine, Pharmacology, and Ophthalmology, University of California, San Francisco, San Francisco, CA, USA

⁷Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA

⁸These authors contributed equally

⁹Lead Contact

*Correspondence: todd.mcdevitt@gladstone.ucsf.edu (T.C.M.), cbelta@bu.edu (C.B.)

<https://doi.org/10.1016/j.cels.2019.10.008>

SUMMARY

Human pluripotent stem cells (hPSCs) have the intrinsic ability to self-organize into complex multicellular organoids that recapitulate many aspects of tissue development. However, robustly directing morphogenesis of hPSC-derived organoids requires novel approaches to accurately control self-directed pattern formation. Here, we combined genetic engineering with computational modeling, machine learning, and mathematical pattern optimization to create a data-driven approach to control hPSC self-organization by knock down of genes previously shown to affect stem cell colony organization, CDH1 and ROCK1. Computational replication of the *in vitro* system *in silico* using an extended cellular Potts model enabled machine learning-driven optimization of parameters that yielded emergence of desired patterns. Furthermore, *in vitro* the predicted experimental parameters quantitatively recapitulated the *in silico* patterns. These results demonstrate that morphogenic dynamics can be accurately predicted through model-driven exploration of hPSC behaviors via machine learning, thereby enabling spatial control of multicellular patterning to engineer human organoids and tissues. A record of this paper's Transparent Peer Review process is included in the Supplemental Information.

INTRODUCTION

During the early stages of embryonic development, patterned self-assembly of cells is essential for the organization of primitive germ layers, multicellular tissues, and complex organ systems (Montero and Heisenberg, 2004). Similarly, human pluripotent stem cells (hPSCs) maintain the ability to self-organize, differen-

tiate to all three germ layers, and generate 3D organoids that replicate primitive tissue structure and function (Bredenoord et al., 2017; Sasai, 2013; Warmflash et al., 2014); hence, hPSCs provide a robust and tractable system to observe, quantify, predict, and ultimately control collective cellular behaviors (Pir and Le Novère, 2016). The ability to direct heterotypic cell self-organization and concurrently specify cell fate can enable the possibility of directing organogenesis via cell-intrinsic routes.

Although several *in vitro* and *in silico* frameworks for multicellular patterning have been independently developed, the ability to predict and direct *de novo* multicellular organization has yet to be demonstrated. Previously, several groups (Molitoris et al., 2016; Tewary et al., 2017; Warmflash et al., 2014) have induced radial organization of differentiated germ layers by restricting hPSC colonies to micropatterned islands or have used molecular engineering of cell surface and/or substrate properties to extrinsically control cell location and subsequent multicellular patterning *in vitro* (Chandra et al., 2005; Hsiao et al., 2009; MacKay et al., 2014; Molitoris et al., 2016; Toda et al., 2018). However, the resulting patterns that arise spontaneously afford limited control of precise multicellular organization or circumvent the intrinsic mechanisms that regulate cell-mediated morphogenic assembly. Theoretical *in silico* frameworks have been developed to computationally model multicellular organization (Bartocci et al., 2016; Briers et al., 2016; Sharpe, 2017) and automate the design of non-spatial cellular logic (Nielsen et al., 2016). However, although computational approaches can test general principles of biology *in silico*, it is often difficult to directly map these models to specific *in vitro* mechanisms and perturbations, making it challenging to systematically synthesize experimentally tractable perturbations *in silico* that can be accurately reproduced *in vitro*.

In this proof-of-principle study, we paired CRISPR interference (CRISPRi)-driven genetic perturbations of human-induced pluripotent stem cells (hiPSCs) with computational modeling, machine learning, and mathematical optimization to facilitate a “closed-loop” cycle of *in silico* hypothesis generation that could be experimentally validated *in vitro*. To predict multicellular pattern formation, we combined a multi-scale cellular Potts model (Graner and Glazier, 1992; Krieg et al., 2008; Magno et al., 2015; Marée et al., 2007; Ouchi et al., 2003; Pir and Le



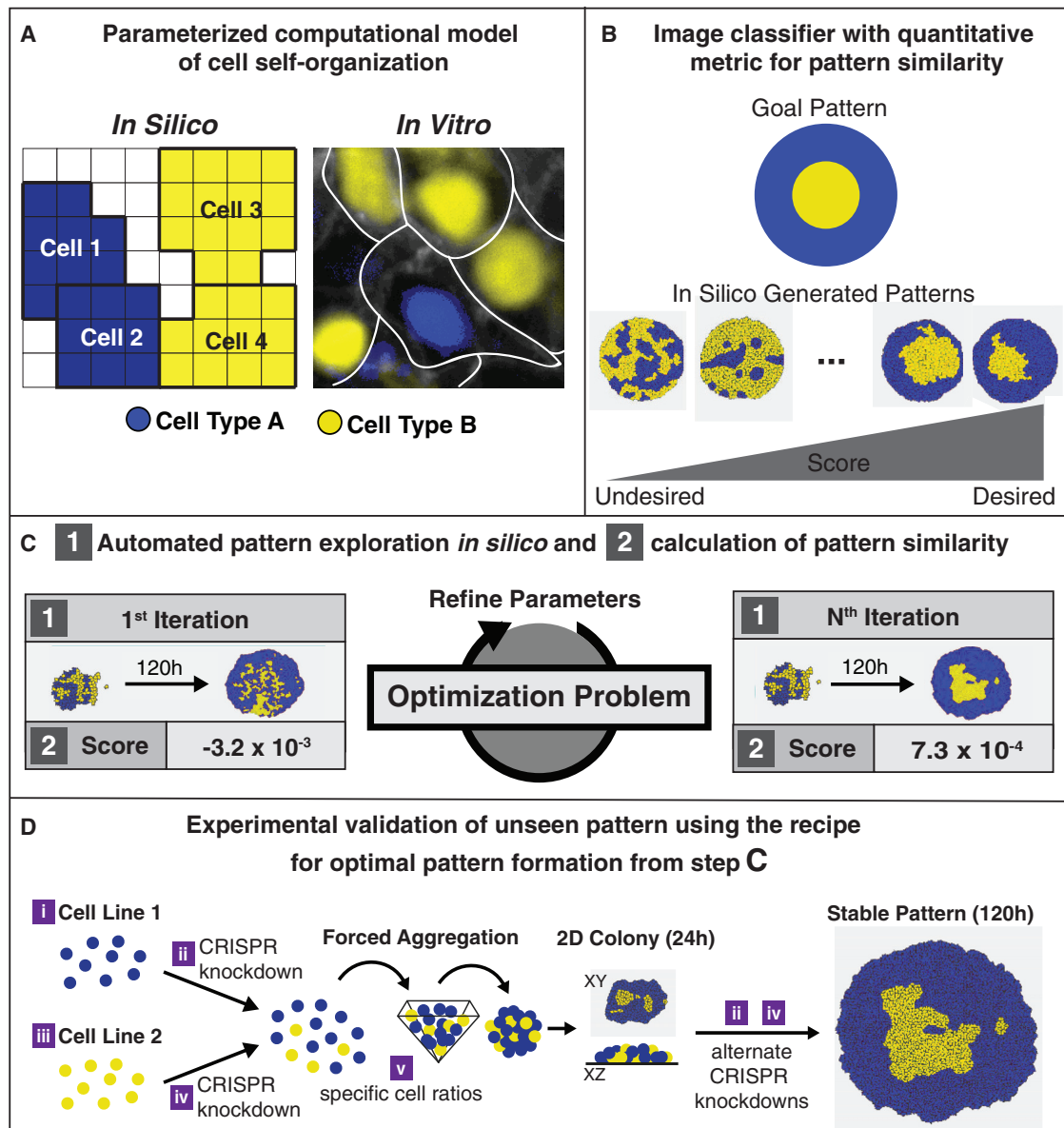


Figure 1. Overview of Synthesis of Spatial Patterns

Pattern Synthesis is a computational procedure used to predict the *in vitro* conditions needed to produce target spatial patterns.

(A) The first input to Pattern Synthesis is a parameterized computational model of mechanically driven spatial patterning in iPSC colonies. Five parameters of the computational model map directly to perturbations that can be made by experimentalists, and the output of the model was a series of images.

(B) The second input to Pattern Synthesis is a trained image classifier that produces a numerical score indicating the similarity of an image to the desired pattern class. In this scenario our desired pattern was a “bullseye” pattern.

(C) Given the parameterized model and pattern classifier, particle swarm optimization was used to explore parameter combinations, which map directly to *in vitro* perturbations, in order to identify the optimal conditions to produce the desired pattern *in silico*.

(D) Given the “recipe” of perturbations suggested by parameter optimization, we validate the control strategy is able to produce the desired pattern *in vitro*.

Novère, 2016) of behavior-driven cell sorting with an automated machine learning and optimization procedure, referred to as “Multicellular Pattern Synthesis” (Bartocci et al., 2016; Briers et al., 2016), which consisted of four steps (Figure 1). First, we created a computational model of observed hiPSC self-organization that quantified collective stem cell dynamics and captured how targeted changes in the mechanical profiles of subpopulations of cells affected stem cell colony patterning. Second, a su-

pervised machine learning classifier was trained to quantify pattern similarity to the desired pattern using images from our computational model. Third, we employed mathematical optimization, specifically particle swarm optimization (PSO), to simulate thousands of potential designs and identify specific experimental conditions that yielded unique patterns in *in silico* simulations. Finally, we tested the *in silico* predicted conditions with hiPSCs *in vitro* and obtained the desired multicellular patterns

with similar frequency and quantitative characteristics, thereby validating the predictive *in silico* system. As an initial exploration of the impact of patterning, we examined regional changes to cell fate commitment in patterned colonies of hiPSCs upon differentiating in response to morphogen treatment (BMP4).

RESULTS

Pattern Synthesis: *In Silico* Prediction and Automated Discovery of Spatial Behaviors

To observe multicellular pattern formation, we used a previously established hiPSC line with a doxycycline (DOX)-inducible CRISPRi system, allowing for temporal gene knockdown (KD) wherein mixed populations establish KD in only a portion of the colony, creating a symmetry breaking event and subsequent pattern formation (Libby et al., 2018; Mandegar et al., 2016). However, the generation of new patterns in a predictable manner requires the ability to test large numbers of experimental conditions that would require a massive amount of time and manual effort to comprehensively test the vast number of experimental parameters possible. For example, to experimentally explore the parameter space of a single gene KD where the following parameters are varied; KD timing (3 timing schemes tested), duration of experiment (5 durations tested), degree of gene KD (5 KD levels tested), and proportion of population that is knocked down (9 percentages tested), one would need to perform 675 total experiments. Given the biological variability we observe within our *in vitro* experiments (Figure 4), a power analysis suggests that a minimum of approximately 13 biological replicates would be necessary to detect significant differences between individual experiments (12.85 observations required, with significance assessed at $p < 0.05$, 80% probability of accepting the alternate hypothesis, corrected for multiple comparisons), yielding approximately 9,000 total conditions or roughly ninety four 96-well plates, where one well represents a single condition. Alternatively, a machine learning and optimization algorithm, such as “Pattern Synthesis” (Bartocci et al., 2016; Briers et al., 2016), can automatically and efficiently discover experimental conditions and robustly predict the *de novo* self-organization of hiPSCs into desired target patterns.

Pattern Synthesis required two inputs: a model of hiPSC behavior and images of the desired pattern (i.e., “goal”) outcomes. First, we developed a computational model of hiPSC colony organization as a result of a single gene KD (Figure 1A). Next, we generated images of desired and undesired spatial patterns to train a machine learning algorithm that establishes a pattern classifier with a quantitative metric of pattern similarity (Figure 1B) (Bartocci et al., 2016; Haghghi et al., 2015). Given these inputs, we formalized pattern discovery as an optimization problem where the objective was to maximize the similarity score of images from our computational model to our desired spatial pattern (Figure 1C). The variation between different simulations was based upon five categories of *in vitro* perturbations that could be readily created in hiPSC colonies (Figure 1D).

Data-Driven Computational Model of Human iPSC Self-Organization

Several different experimental and computational studies support the vital role of local cell-cell mechanical interactions in

the form of interfacial tension in spatial patterning (Heisenberg, 2017; Maître et al., 2012). Differences in cell-cell adhesion (Foty and Steinberg, 2005; Jia et al., 2007; Maître and Heisenberg, 2013; Steinberg, 1975), cell-cell repulsion (Taylor et al., 2017), and cortical tension (Brodland, 2002; Heisenberg and Bellaïche, 2013; Krieg et al., 2008) have been shown to collectively orchestrate spatial organization in diverse organisms such as *Dictyostelium discoideum* (slime mold) (Kay and Thompson, 2009; Palsson, 2008) and *Danio rerio* (zebrafish) (Merks et al., 2008) and in mammalian cells (Bentley et al., 2014; Toda et al., 2018). However, it is challenging to both predict and control spatial patterning in human iPSCs since the design of multicellular systems rapidly increases in complexity when considering the dynamics of single-cell mechanics and cell-cell interactions. These dynamics include, but are not limited to, temporal changes in interfacial-tension-associated proteins, cell type abundance, cell division, and cell migration velocities.

To capture the complex dynamic interactions involved in multicellular patterning, we developed a data-driven cellular Potts model (CPM) to predict spatial patterning in hiPSCs due to the time-dependent modulation of cell-cell adhesion and cortical tension (Supplemental Information). The CPM represents the spatial environment of stem cells grown in a monolayer using a 2D square lattice grid (Figure 1A). Each square region in the grid (i.e., a lattice site) is equal to 1 square micrometer, hence each lattice site represents a partial region of a cell’s membrane or the medium surrounding a cell. A cell ID is assigned to each lattice site to identify the region of a cell that occupies a lattice site. For example, 100 lattice sites each having a cell ID equal to 1 represent a single stem cell with an area of 100 square micrometers. Complex behaviors such as preferential cell-cell adhesions, cortical tension, and cell migration are achieved by copying lattice sites to adjacent regions (i.e., moving a region of the cell to a new lattice site), which in the CPM is referred to as a copy attempt. Each copy attempt is accepted with a probability related to a Hamiltonian function (SI Equations 3–5). The Hamiltonian function is the sum of four competing forces influencing intracellular behaviors and cell interactions with the environment: (1) conservation of cell area, (2) locally polarized cell migration, (3) cell-cell adhesion, and (4) and cortical tension (SI Equations 5–10). Every competing force is represented by a score and a weight, where the score represents a reward or penalty depending on the divergence of a cell from its target behavior, while the weight represents the relative importance of the respective cell’s behavior.

Briefly, using the CPM, we modeled an *in vitro* system consisting of two populations of iPSCs co-cultured for up to 120 h. To connect the *in silico* model to potential genetic targets for *in vitro* experimental manipulation, we focused on using CRISPRi KD, which provides precise temporal control over protein expression, of two molecules associated with regulating cellular mechanics and cell-cell interactions: E-cadherin (CDH1) and Rho-associated coiled-coil containing protein kinase (ROCK1). CDH1 is a classical cadherin cell-cell adhesion molecule, whose modulation allows for changes in the adhesive interactions between neighboring cells, and ROCK1 is a protein kinase that regulates non-muscle myosin activity and indirectly modulates the actinomyosin cytoskeletal tension within and between cells. These two molecules contribute to feedback loops that regulate

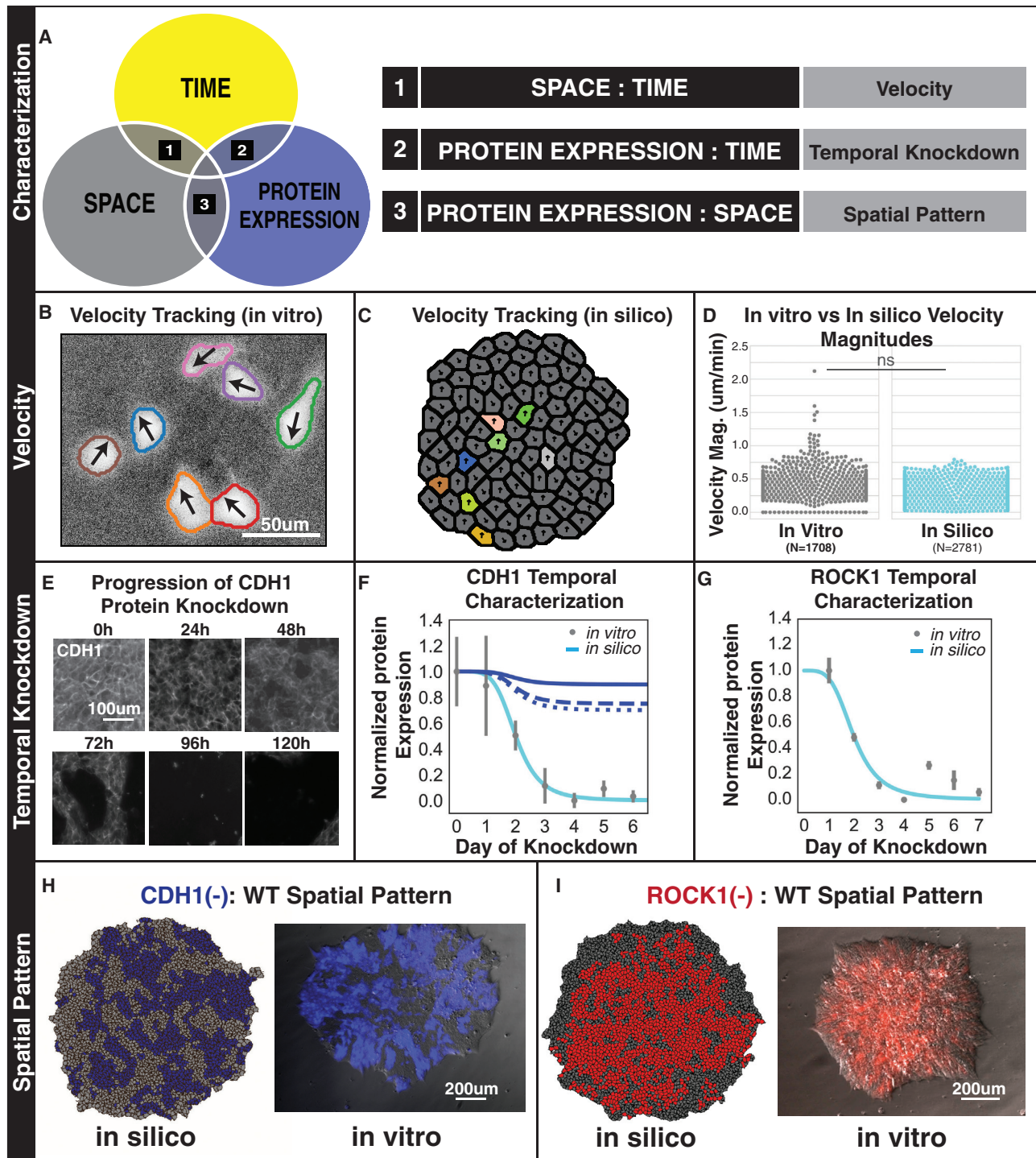


Figure 2. Pairwise Experiments to Characterize Dynamic Changes in Spatiotemporal Behaviors

(A) We characterize cellular behavior in a pairwise manner to reduce the complexity of possible interactions. Space, time, and protein expression are the minimal necessary properties to characterize and model spatiotemporal behavior. Space-time relationships are captured with velocity characterizations, time-protein expression is captured characterizing the relative protein expression for several days after knockdown, and protein-space relationships are characterized by confocal microscopy imaging of spatial behavior due to cell mechanical perturbations.

(B and C) We performed paired *in vitro* (B) and *in silico* (C) experiments to match the velocity distributions of iPSCs.

(D) The gray swarm plot represents the distribution of *in vitro* velocity magnitudes ($n = 1,708$), while the cyan swarm plot represents the distribution of *in silico* velocity magnitudes ($n = 2,781$). Using the Mann-Whitney U test, there was no statically significant difference in cell velocity (p value = 0.29).

(E) Representative images of DOX-inducible modulation of protein expression.

(legend continued on next page)

interfacial tension between cells within a tissue and facilitate the physical organization of multiple cell types, making them ideal candidates that when knocked down alter the cellular organization within a iPSC colony (Libby et al., 2018).

To fit the *in silico* model to an *in vitro* experimental training set, pairwise *in vitro* characterization experiments were performed to determine the relationship between space, time, and protein expression (Figure 2A) in wild-type (WT), CDH1 KD, and ROCK1 KD hiPSCs. These relations were established by *in vitro* measurements of single-cell morphological changes (Figure S1), migration velocity magnitudes (Figures 2B–2D), protein expression changes (Figures 2E–2G), and colony organization (Figures 2H and 2I) before and after mosaic KD of CDH1 and ROCK1 in hiPSC colonies. The purpose of these characterization experiments was 2-fold: (1) to reduce the complex interactions into quantifiable relationships and (2) create a closed-loop mapping between *in vitro* perturbations and *in silico* simulation parameters.

To characterize cell morphology, brightfield images of WT, CDH1(–), and ROCK1(–) cells were collected 6 days (144 h) after gene KD. Single-cell *in vitro* cell area and membrane length measurements (Figure S1) were acquired to set the target cell area and target cortical tension in the simulations, respectively ($n = 110$ per cell type). In the CPM, the weight associated with cortical tension constraint regulates how readily a cell can change its cell membrane length and relates to cell membrane stiffness. Because of differences in cell crowding in the center versus the edge of colonies, cell morphology measurements were fixed given a cell's mechanical modulation and its radial position in the colony ($n = 55$ central and 55 edge) (Table S2; Figure S1). Cell division was assumed to be asynchronous among the population, and cell division times specific to each type of KD were incorporated into the model to provide an accurate depiction of population growth kinetics.

The relationship between cells in space with respect to time was characterized by measuring the *in vitro* distribution of individual cell velocities, resulting in an empirical median velocity magnitude of $0.29 \mu\text{m}/\text{min}$ and median absolute deviation (MAD) of $0.10 \mu\text{m}/\text{min}$ (Figures 2B and 2D). The distribution of velocity magnitude values was then used to model collective cell migration as locally polarized motility where the direction of cell migration is influenced by the relative cell adhesion strength of neighboring cells (Czirók et al., 2013). A grid search was performed for cell-cell adhesion and cortical tension parameters for WT. We then selected the parameter combination that mostly closely mimics the *in vitro* velocity measurements, producing a comparable distribution with a median *in silico* velocity magnitude of $0.31 \mu\text{m}/\text{min}$ and MAD of $0.15 \mu\text{m}/\text{min}$ (Figures 2C and 2D). Importantly, the *in silico*-generated velocity distributions were not significantly different from the *in vitro* measured velocities (Mann-Whitney U test, $p = 0.29$, $N = 2,781$). To further test how robust these results were to random variation in the initial

population size of the colony, we generated 24 additional simulations of the optimal parameter combination that had a median velocity of $0.34 \mu\text{m}/\text{min}$ and MAD of $0.15 \mu\text{m}/\text{min}$ ($N = 78,747$ cells). We then performed a Mann-Whitney U test ($p = 0.51$, $N = 78,747$ *in silico* and $N = 1,708$ *in vitro*), showing our simulations robustly represent the velocity magnitudes observed experimentally. When selecting optimal parameters, we also manually inspected images and only evaluated parameter combinations where individual cells remained part of a dense epithelial colony without migrating from the exterior borders to match the hiPSC phenotype observed *in vitro*. After fitting the model to empirical data of cell morphology, velocity, and single-cell morphology, collective cell migration of human iPSC colonies was accurately recapitulated (Videos S1 and S2).

To characterize how protein expression changes in relation to time, CDH1 and ROCK1 were knocked down using CRISPRi, and the relative mRNA and protein expression was assessed for 6 consecutive days via qPCR, fluorescence microscopy, (Figures 2E and S2), and western blot analyses (Libby et al., 2018). Because of our previous observation of the phenotypic robustness of CDH1 KD in promoting cell self-organization (Libby et al., 2018), we designed several CRISPRi guide RNAs to target CDH1 producing different levels of transcriptional KD at 10%, 25%, 30%, and 98% compared to WT expression. A single guide RNA for ROCK1 KD was used to achieve an 80% KD of WT expression levels (Figure S3). The data were normalized (min-max [0,1]) and Hill functions were fit to the normalized median expression (per day) using least-squares optimization to create a time-dependent response function for CDH1 knocked down to 90%, 75%, 70%, and 2% of the original mRNA expression (Figure 2F). This range of KD efficiency allowed us to computationally model how differing levels of CDH1 expression could impact spatial patterning. Using the same approach as the CDH1 KD, we created a Hill response function for ROCK1 knocked down to 20% of the original mRNA expression. Because of a delay in protein KD compared to mRNA levels, the Hill functions were shifted by 24 h to account for the delay in protein loss (Figure 2G), allowing us to model the average change in ROCK1 protein expression for individual cells over time.

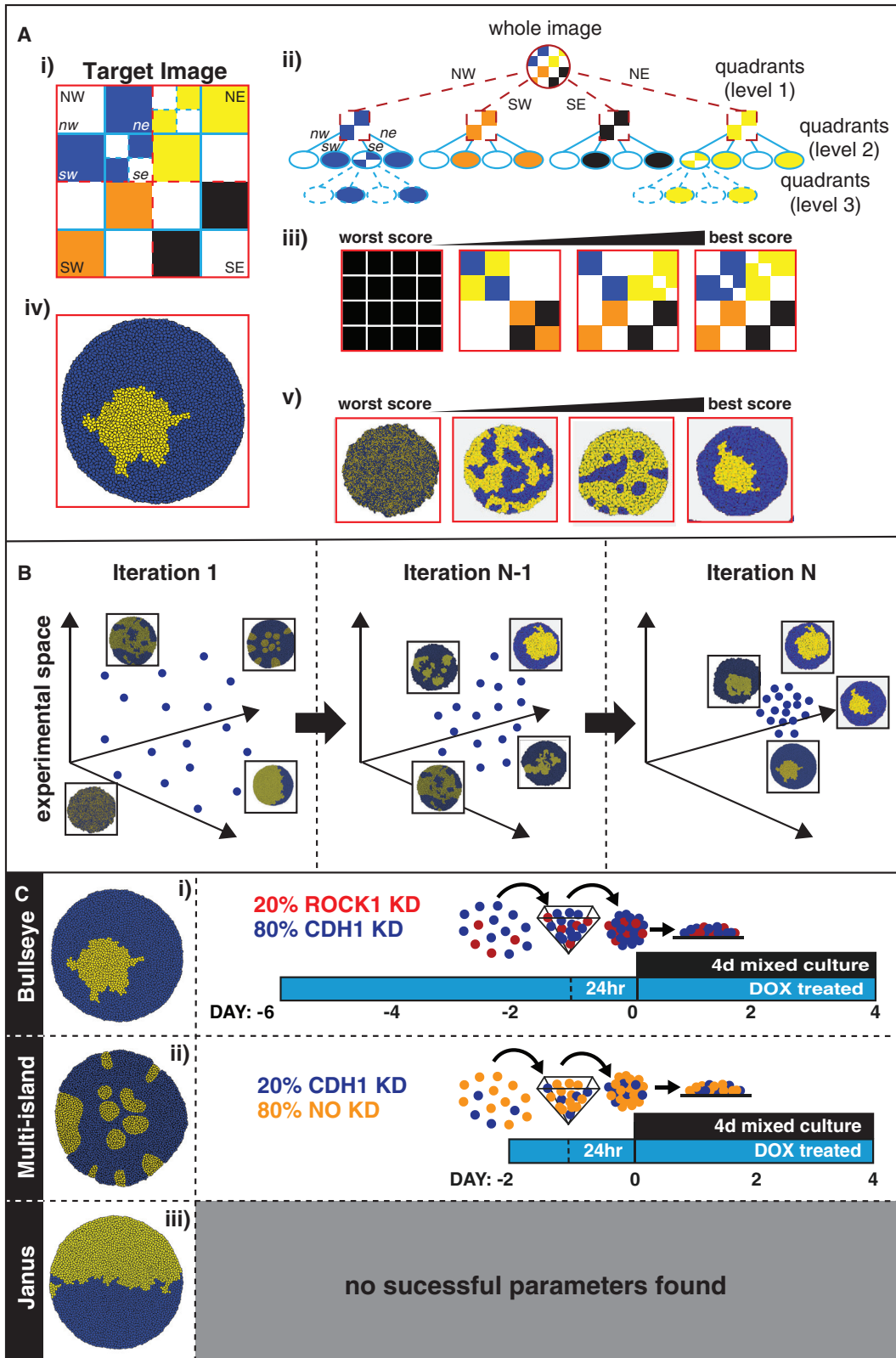
Given the previous characterization experiments, we were able to model collective cell migration and temporal changes in cell mechanics. To model the spatial patterning due to the temporal modulation of cell-cell adhesion via CDH1 or cortical tension via ROCK1, either inducible ROCK1 KD or inducible CDH1 KD iPSCs were co-cultured with WT iPSCs, where KD of gene expression was induced upon mixing the two cell types. Then, images of the mixed populations were collected 96 h after gene KD and co-culture. As previously reported (Libby et al., 2018), mixed colonies with a subpopulation of cells that had reduced CDH1 or ROCK1 expression produced distinct mosaic patterns due to reduced cell-cell adhesion and increased

(F) We used Hill functions to mathematically model CDH1 knockdown over time from quantification of mRNA by qPCR ($n = 3$) and then adding a 24 h delay to account for protein production depicted by light blue lines. Gray circles represent the normalized median expression 0–6 days after CDH1 knockdown. Error bars represent 1 standard deviation from the mean. The dark blue line depicts Hill function models of partial KD of CDH1.

(G) We use a Hill function to model ROCK1 knocked down over time as previously described for CDH1 knockdown ($n = 3$).

(H) Paired *in vitro* and *in silico* images of spatial patterning 96 h after CDH1 knockdown in a subpopulation of cells (blue).

(I) Paired *in vitro* and *in silico* images of spatial patterning 96 h after ROCK1 knockdown in a subpopulation of cells (red). Given the previous characterizations, the relative strength of cell-cell adhesion and cortical tension can be tuned in the *in silico* simulations to recapitulate the spatiotemporal patterns observed *in vitro*.



(legend on next page)

membrane stiffness properties, respectively (Figured 2H and 2I, left). *In silico*, parameter sweeps were run over a range of adhesion strength and membrane length values to explore the phenotypic space resulting from decreases in cell-cell adhesion and increases in membrane stiffness. Computationally varying the adhesion strength produced a variety of spatiotemporal patterns due to progressively weaker cell-cell adhesion or progressively stiffer cell membrane parameter values. Double-blind analysis of *in silico*- and *in vitro*-generated data sets was then conducted to identify parameters that yielded closely matching multicellular patterns (Figures 2H and 2I, right). Given the characterization experiments of cell morphology, cell migration speed, time-dependent modulation of cell mechanics, and the resulting spatial organization, the computational model was able to recapitulate the spatial patterning due to the CDH1 and ROCK1 KDs (Videos S3, S4, S5, and S6).

Overall, after incorporating *in vitro* measurements into our computational model, we accurately recapitulated hiPSC spatial patterns with the initial experimentally derived parameters in mixed colonies of WT and CDH1 KD cells or WT and ROCK1 KD cells (Videos S2, S3, S4, S5, and S6) (Libby et al., 2018).

Formulating Parameters for Design Automation

Given the success in matching the computational model to the *in vitro* experimental training set, we then introduced five new design parameters to simulate *in vitro* experimental perturbations, allowing us to model exponentially more permutations of experimental design than would be feasible *in vitro*. The five design parameters were; (1) the gene KD target of cell population 1, (2) the KD time for cell population 1, (3) the gene KD target of cell population 2, (4) the KD time for cell population 2, and (5) the ratio of the distinct cell populations (Figure 1D; Table S1). These additional design parameters allowed us to convert trial-and-error-based design into a mathematical optimization problem that could be computationally solved *in silico* without time-consuming and costly additional experiments. Although computational design frameworks for multicellular spatiotemporal patterning have been used in several previous studies (Krieg et al., 2008; Marcon et al., 2016; Tewary et al., 2017), they often propose underlying morphogenic mechanisms with limited perturbation potential *in vitro*. Thus, an *in silico* optimization framework that directly informs subsequent experimental design is critical to survey the high-dimensional landscape of morphogenesis.

Quantitative Pattern Classification

The second input to the Pattern Synthesis procedure was a supervised image classifier known as tree spatial superposition

logic (TSSL) (Bartocci et al., 2016). TSSL uses a quadtree data structure to represent spatial relationships in an image at multiple levels of detail, where the highest level captures global aspects of an image, while the lower levels capture local spatial relationships. For example, examining a checkerboard image with some variation (Figure 3Ai), the TSSL would generate a unique quadtree (Figure 3Aii) representing the levels of complexity within the image (Bartocci et al., 2016; Finkel and Bentley, 1974; Jackins and Tanimoto, 1983). A rule-based machine learning algorithm (RIPPER) (Cohen, 1995) was employed to automatically learn a set of rules over the values of quadtree vertices specific to an *in silico* training set of 3,000 positive and 13,000 negative manually rendered images of cells precisely organized into target patterns, such that a quantitative score of pattern similarity could be assigned to any image from the associated quadtrees (STAR Methods) (Figure 3Aiii). The magnitude of the similarity score, which can range from -1 to $+1$, indicates how strongly a simulation image matches (positive scores) or violates (negative scores) the target spatial behavior. Use of a TSSL robustness score replaces qualitative manual observation of simulation images with a quantitative score of pattern similarity.

Analogous to the checkerboard example, this algorithm can be applied to more complex images such as a target organizational pattern within the CPM (Figures 3Aiv and 3Av) where the generated quadtree from the TSSL of each desired pattern is used to recognize and rank pattern similarity (Figure 3Av). As a proof of principle, we first attempted a concentric ring (i.e., “bullseye”) pattern, defined as one population of 50 or more connected cells completely surrounded by a second population (Figure 3Ci). The annular bullseye pattern was chosen because similar asymmetric cell organization occurs multiple times in human development, such as during early embryo compaction leading to the formation of the inner cell mass in the human blastocyst (Deglincerti et al., 2016; Ducibella and Anderson, 1975; Ziemek and Johnson, 1980). The second target was a multi-island pattern, consisting of at least three distinct clusters of 25 or more cells completely surrounded by a separate larger population (Figure 3Cii). The island pattern was chosen to demonstrate the reproducibility of previously observed segregation of cell populations *in vitro* (Libby et al., 2018) and to test whether this can be predictably controlled. To first demonstrate that the automated classifiers could reliably detect and distinguish between desired and undesired spatial patterns, the classifiers were tested using an *in silico* set of 1,000 positive and 5,000 negative images. The TSSL classifiers achieved a 98.2% classification accuracy for the bullseye pattern and 96.9%

Figure 3. Quantitative Pattern Classifier with TSSL

A quadtree is used to represent an image at multiple levels of detail.

(A) A representative quad tree for an example checkerboard image (Ai and Aii). An image (Ai) is subdivided into sequential quadrants until each quadrant is one singular color. This is then depicted as a tree (Aii) where both the values and branches of the tree are specific to each image. (Aiii) Given a quadtree representation of a target image, TSSL produces a numerical score corresponding to the similarity of an image to the desired target image. This score can then be used to rank images by similarity to the desired image. (Aiv) An example image of a desired pattern generated in the CPM. (Av) A pictorial example of how the TSSL would be able to distinguish different CPM images and score them against the desired pattern.

(B) Schematic representation of a particle swarm algorithm depicted in a 3D search space where each particle represents an *in silico* simulation. With each iteration of the algorithm, PSO reduces the breadth of exploration in the experimental space and travels toward increasing TSSL scores, indicating that the optimization procedure has located *in silico* experiments that are generating patterns of increasing similarity to the goal pattern.

(C) Schematics of example target patterns given as classifiers in the machine learning Pattern Synthesis process and parameters produced by Pattern Synthesis that predict the creation of the desired patterns: (Ci) bullseye, (Cii) island, and (Ciii) Janus.

classification accuracy for the multi-islands pattern, meaning that the TSSL algorithm can properly recognize and score bullseye or island patterns nearly 100% of the time.

By quantifying how well images from an *in silico* multicellular arrangement matched images of our target organization, we enabled the optimization algorithm (described in the next section) to incrementally improve and learn a unique combination of design parameters that could give rise to a desired goal pattern.

Automated Discovery of Pattern-Producing Conditions

The CPM allows simulation of more than 40,000 distinct parametric conditions and facilitating the study of emerging behaviors of hiPSCs much faster than *in vitro* experiments. Distributing the computation over 12 processors at 2.1 GHz on a server cluster, it only took approximately 5 min to simulate the evolution of one cell population over 120 h. To recapitulate this same experiment *in vitro*, 13 96-well plates would need to be cultured in parallel for 120 h, demonstrating that *in silico* experimentation can accelerate parameter exploration more than 1,000-fold. The simulation speed permitted examination of a wide range of different experimental conditions in a rapid and inexpensive manner, taking both the labor and reagent costs into account. However, it quickly became impractical to enumerate every possible set of conditions to identify parameter combinations that yielded the highest robustness score(s) because of the tens of thousands of experimental conditions to consider and the resulting months of computation for such a large number of simulations. Thus, to automate the discovery of conditions that yielded goal spatial patterns, we formulated the selection of experimental conditions as an optimization problem.

Using the TSSL-provided metric of image similarity, a PSO (Eberhart and Kennedy, 1995) was employed to identify regions of the 5-dimensional (5D) design space, created by the available design parameters, with the highest probability of producing a target pattern (bullseye or multi-island) (Figure 3B). In brief, the PSO first explores the extremes of the 5D experimental space, where every extreme represents a set of experimental parameters that are run as an *in silico* experiment using the previously described CPM. Then, the resulting patterns from this first set of *in silico* experiments are given scores. The algorithm then narrows its focus to the experimental space that produced experiments resulting in higher scores, doing this iteratively further selecting for the experimental space that is most likely to produce the highest TSSL score and therefore the patterns that most closely resemble the goal pattern. A full explanation of the particle swarm algorithm can be found in the STAR Methods.

For any *in silico* simulation, where the previously described design parameters are varied to represent a different experimental condition, the patterning synthesis algorithm determined whether the generated pattern was successful (Figure S4A), and whether the similarity score improved over the simulated period of 120 h by at least one order of magnitude, eventually reaching a steady state (Figure S4B). Analyzing the temporal dynamics of robustness scores provided insight into the exact time a pattern emerged *in silico*, and optimized design parameters for target patterns that closely resembled, but still resembled the desired spatial behavior. The final output of the particle swarm algorithm is a list of experimental parameters that are predicted to

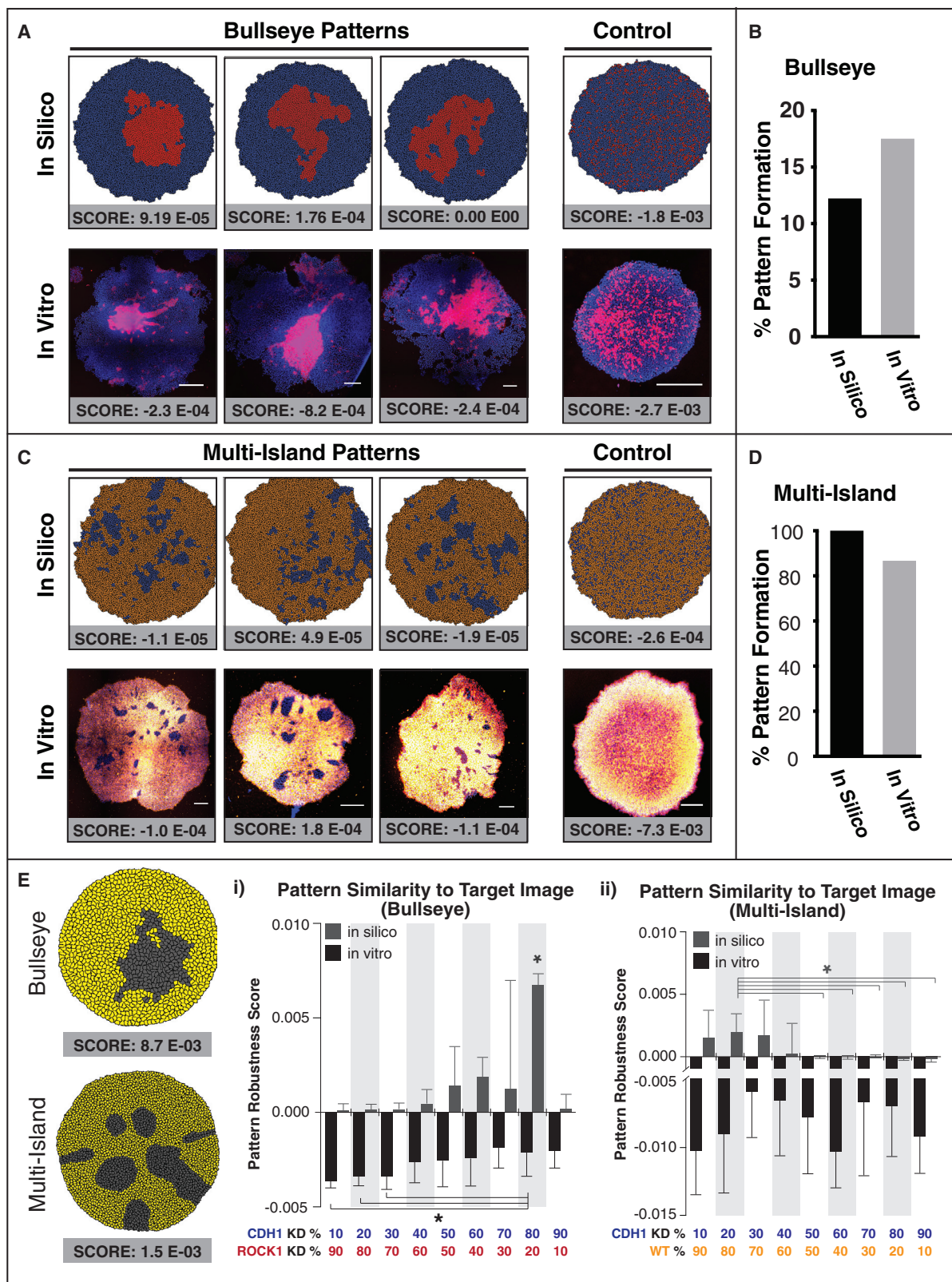
generate the desired pattern both in the *in silico* CPM and the *in vitro* stem cell culture system after 120 h of mixed culture (Figure 3C).

In addition to automating the design of *de novo* spatial patterns, we could also determine the feasibility of any spatial pattern given the tunable conditions of the system. Although it is impossible to exclude experimentally that a particular pattern can never be generated *in vitro* (it would require testing all possible conditions), *in silico* certain *de novo* patterns resulted in negative robustness scores (violating the pattern specification), indicating that the cell population under the current perturbations available (KDs, mixing ratios, etc.) was unable to perfectly recapitulate the desired spatial behavior. For example, the algorithm was able to determine that a perfectly symmetrical “Janus” pattern (left-right) (Figure 3Cii) was not achievable with the primary experimental variables (i.e., timing of CDH1/ROCK1 KDs and the ratio of cell types co-cultured in an approximately 2D monolayer), indicating that additional parameters such as silencing of other genes may be necessary to yield such a pattern.

In Silico Model Accurately Predicts In Vitro Experimental Validation

The Patterning Synthesis algorithm yielded different sets of instructions to produce either a bullseye pattern or a multi-island pattern of hiPSCs. The Pattern Synthesis predicted that a mixture of 1:4 ROCK1 KD iPSCs to CDH1 KD iPSCs that were independently pretreated with DOX for 6 days prior to mixing and cultured together for 4 days would yield a bullseye pattern (Figure 3Ci) and that a mixture of WT cells with CDH1 KD at a ratio of 1:4 with DOX pretreatment of iPSCs for 48 h prior to mixing would create the multi-island pattern (Figure 3Cii).

Based on these predictions, *in vitro* experiments were performed using the specified conditions, and the incidence of pattern formation was independently analyzed for *in silico* and *in vitro* results (Figure 4A–4D). The experiments were performed with unrestricted colony growth (i.e., no patterned matrix restriction) (Tewary et al., 2017; Warmflash et al., 2014) to ensure that cellular organization within the colony was not driven by imposed boundary conditions. To account for colony size differences affecting the resulting patterns, only colony sizes within two standard deviations of the mean number of cells per colony were examined for pattern formation. We characterized the morphology of *in silico*- and *in vitro*-generated patterns by interrogating subpopulation cluster circularity, number of clusters, and cells per cluster within the colony (Figure S5). However, the *in vitro* experimental results were more variable and yielded a wider range of results, which may be due to biological variability in wet lab experimentation or subtle variations in cellular behavior that the *in silico* model does not take into account. Comparing the robustness scores generated for both the parallel *in silico* and *in vitro* experiments indicated that the optimal *in vitro* bullseye and multi-island patterns had larger robustness than their respective control images (at least an order of magnitude difference). The robustness scores are highly comparable only when they are calculated in the same setting also known as a domain; thus, a simulation versus a simulation control is quite comparable whereas an *in silico* simulation versus an *in vitro* experimental image will inherently differ to some extent. The *in*



(legend on next page)

silico model and experimental optimization predicted that a bullseye pattern would be achieved ~12% of the time, which closely matched the *in vitro* frequency (~16%; Figure 4B). Similarly, a multi-island pattern was predicted to occur 100% of the time by the model and was achieved in ~87% of the *in vitro* experiments (Figure 4D). Overall, these results demonstrate that *in silico* modeling accurately classified and predicted desired pattern formation achieved by hiPSC self-organization *in vitro*.

To determine how robust the predicted parameters were within the *in vitro* system, the proposed mixing ratios of the populations were incrementally varied by 10 percent ($n = 16$ per condition) (Figures 4E and S6). Robustness scores for each of the mixing ratios were calculated (Figures 4Ei and 4Eii) to compare how well each condition produced patterns similar to the respective target (bullseye or multi-island). In bullseye patterns, a 50 percent change in mixing ratio from the predicted parameters (80% CDH1KD:20% ROCK1 KD) resulted in significant decrease in pattern robustness scores ($p < 0.05$) (Figure 4Ei). Despite an increase in the robustness scores for the multi-island patterns in the *in silico* experiments, there were no significant differences in the robustness scores calculated for the parallel *in vitro* experiments with varying population ratios (Figure 4Eii). Robustness scores produced by the TSSL algorithm *in vitro* were uniformly lower and had higher variance than the comparable *in silico* conditions (Figures 4Ei and 4Eii), reflecting the greater difficulty in classifying natural images over the synthetic images generated by the CPM. Because of the domain change from *in silico* to *in vitro* images, the TSSL algorithm was less able to confidently recognize patterns and explain variability both within and across experiments, resulting in reduced discrimination between mixing ratios. Additionally, differences could be due to the fact that the CPM used is a 2D model that does not account for possible vertical movement within a hiPSC colony. However, as the primary goal of the TSSL was to enable *in silico* pattern optimization, the decreased classification power for *in vitro* images did not adversely impact the ability of pattern optimization to predict conditions that resulted in the desired target patterns.

Colony Organization Impacts Initial Patterns of iPSC Differentiation

During human development, cell-autonomous pattern formation is intimately coupled with cell fate decisions that lead to complex tissue structures. Therefore, we interrogated how the experimentally generated multicellular patterns affected subsequent hiPSC differentiation. We examined the initial cell fate commitment after treatment with BMP4 for 48 h (Figures S7A and S7B) with a panel of markers descriptive of different differentiation stages (Fig-

ure S7C). In brief, hiPSCs were marked by high OCT4 and SOX2 expression in the pluripotent state, and then as differentiation proceeded, the first lineage fate decision was marked by up-regulation of markers associated with the gastrulating primitive streak (Brachyury (BRA(T)) and SNAIL). Cells then transitioned through a mesendodermal fate (EOMES) before displaying mesoderm (GATA4) or endoderm (SOX17) specific markers. The ectoderm lineage remained SOX2(+). Additionally, CDX2 was used to mark both extra embryonic lineages and presumptive neural plate cells within the neuroectoderm (Niwa et al., 2005; Tewary et al., 2017; Wang et al., 2012; Warmflash et al., 2014). WT colonies displayed a radial differentiation pattern with central SOX2(+) OCT4(+) SNAIL(+) cells and a ring of EOMES(+) cells around the periphery indicating the beginning of mesendodermal specification (Figures S7C and S7D). The lack of robust BRA(T) expression was likely due to the transient nature of BRA(T) expression during mesendoderm induction so the time point examined (48 h) in this experiment may have captured the tail end of expression. WT colonies displayed a slight increase in SOX17 at the center of the colonies, while GATA4 and CDX2 remained low throughout the colonies (Figures S7C and S7D). A similar radial pattern of cell differentiation was maintained in island-patterned colonies, although SOX17 expression was reduced and GATA4 and CDX2 expression increased (Figures S7C and S7D). The bullseye patterned colonies displayed a slight increase in BRA(T) expression at the center of the colonies overlapping with the central island that defines a bullseye pattern. Additionally, for the bullseye patterns, GATA4 expression was increased across the entire colony; the radial ring of EOMES was expanded to the entire colony; and high levels of SOX2, OCT4, and SNAIL were displayed in the center of the colony. These results suggest that the central region of bullseye colonies underwent lineage transitions through the mesendodermal fate to the mesoderm lineage and displayed an expansion of the CDX2-positive cells. The bullseye EOMES expression pattern was distinctly different from the control and island-patterned colonies that formed a ring of EOMES expression, indicating a positional change in fate acquisition dictated by the establishment of the bullseye pattern. Thus, the genetic manipulations used to control multicellular organization of human PSCs also influenced the initial differentiated phenotypes of the patterned colonies.

DISCUSSION

Cell-intrinsic patterning of multicellular stem and progenitor populations is a critical feature of morphogenic events that occur throughout early development (Deglincerti et al., 2016; Ducibella

Figure 4. Computational Synthesis of De Novo Spatial Patterns and In Vitro Validation

(A–D) Comparisons of three simulations of patterns predicted *in silico* and the resulting patterns seen *in vitro* under the same experimental conditions (scale bars = 200 μm). Pluripotent colonies stained for DAPI (blue) and CDH1 (red (A) or orange (C)) to distinguish populations by the presence or absence of CDH1. TSSL robustness scores show how well a simulation matches our specification compared to the parallel control *in silico* or *in vitro* experiment. Scores are only comparable if they are calculated in the same environment (simulation versus simulation but not simulation versus experimental image). Image classification in different environments is a well-known limitation in machine learning. (B) and (D) Successful pattern creation rates, comparing *in silico* to *in vitro* results (bullseye $n = 286$ colonies and multi-island $n = 168$ colonies).

(E) Proposed KD populations for the bullseye pattern and the multi-island pattern were varied by 10% *in silico* and *in vitro* ($n = 10$). Example target patterns (left) from the image set used to train the image classifier to identify and score multi-island and bullseye patterns. (i and ii) Robustness scores for the respective *in vitro* colonies as KD populations were varied by 10% where an increase in robustness score indicates more similarity to the target pattern. The predicted parameters from the *in silico* optimization are highlighted in gray and the *in vitro* are in black. Significance is indicated by \times with p values < 0.05 where $n = 16$ colonies per condition and error bars indicate standard deviations.

and Anderson, 1975; Montero and Heisenberg, 2004; Sasai, 2013). Thus, systems in which multicellular organization can be robustly controlled and perturbed will help to elucidate key mechanisms in development and symmetry breaking events. Currently, the study of symmetry breaking events often involves the manipulation of cell extrinsic factors, for example, varying morphogen gradients (Demers et al., 2016; Chung et al., 2005), changes in substrate patterning (Hsiao et al., 2009; Théry et al., 2006), and/or the creation of restrictive boundary conditions (Tewary et al., 2017; Théry, 2010; Warmflash et al., 2014). In contrast, attempts to influence patterning events using synthetic biology approaches often rely on implementation of an artificial circuit that uses reaction-diffusion gradients to establish multicellular patterns (Greber and Fussenegger, 2010; Sekine et al., 2018; Sohka et al., 2009; Toda et al., 2018).

In this study, we demonstrate the induction of active multicellular organization through controlled perturbation of intrinsic cell mechanisms without imposing exogenous boundary conditions. We developed a computational model capable of predicting empirically testable experimental perturbations (combinations of time-dependent gene KDs and mixing ratios) to generate desired multicellular spatial patterns in hiPSC colonies. Using agent-based model predictions of spatiotemporal pattern formation, we were able to predict and achieve new patterns *in silico* and *in vitro* without using extrinsic patterning methods (i.e., hydrogels and micropatterning). Optimized design parameters achieved desired organization of cells within a colony, providing a new platform to interrogate questions of cell patterning and lineage fate acquisition. Ultimately, these results demonstrate that machine learning and mathematical optimization enable predictive and controlled spatial self-organization of heterogeneous populations of pluripotent cells, which is a critical first step for hiPSC self-assembly prior to lineage commitment and subsequent organoid and tissue formation.

Previous attempts to pair computational models with experimental morphogenic systems have been largely observational and rarely demonstrate the ability to design phenotypes *in silico* that can be recapitulated *in vitro*. In this study, both the *in silico* and *in vitro* aspects can be adapted to additional parameters, truly taking advantage of machine learning and optimization to generate desired multicellular patterns. With respect to extending *in vitro* perturbations, CRISPR technology can be adapted to repress or activate any accessible genes related to cell patterning and organogenesis. As additional biological parameters are considered, we can quantitatively characterize the effect on cell patterning, and the *in silico* model can be refined to take those factors into account (Briers et al., 2016; White et al., 2015), enabling interrogation beyond cell mechanics and into other realms of cell-cell communication such as paracrine signaling gradients and gap junction connectivity to allow for more robust pattern formation than that described by only manipulating cellular mechanics (Glen et al., 2018; White et al., 2013). Ultimately, the combination of agent-based modeling, machine learning, and directed symmetry breaking provides a unique route to engineer complex multicellular tissue structures that go far beyond simple observation of pattern formation and facilitate targeted mechanistic studies that address fundamental principles of development and morphogenesis, leading to robust practices for complex *in vitro* tissue formation.

Key Changes Prompted by Reviewer Comments

For context, the complete Transparent Peer Review Record is included within the Supplemental Information as Document S2.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCE TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
 - Materials Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell Lines
- METHOD DETAILS
 - Generation of CRISPRi Knockdown iPSC Lines
 - Mixed Colony Generation
 - Immunofluorescence Staining and Imaging
 - Protein Quantification
 - mRNA Quantification
 - Time Lapse Imaging
 - Comparison of *In Vitro* and *In Silico* Spatial Patterns
 - BMP4 Differentiations
 - Cellular Potts Model Environment
 - Cellular Potts Model Dynamics
 - CPM Configuration Energy
 - Physical Units and Other Cellular Phenomena of CPM
 - Model Fitting to Empirical data
 - TSSL Scoring and Pattern Optimization
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY
 - Software
 - Data Availability

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cels.2019.10.008>.

ACKNOWLEDGMENTS

We would like to thank the developers of the multi-scale simulator Morpheus for providing technical support and feedback on modeling. We also thank the Gladstone Light Microscopy and Histology Core, the Gladstone Stem Cell Core, and the Gladstone Art Department for their support and experimental expertise. This work was supported by the National Science Foundation; Center Emergent Behaviors of Integrated Cellular Systems (CBET 0939511); the National Science Foundation, Center Cyber Physical Systems Frontier grant: Collaborative Research: bioCPS for Engineering Living Cells (CNS-1446607); the California Institute of Regenerative Medicine (LA1_C14-08015); and the National Heart Lung and Blood Institute (1F31HL140907-01).

AUTHOR CONTRIBUTIONS

D.B., A.R.G.L., C.B., and T.C.M. were responsible for conceptualization of the project and designing the experiments. A.R.G.L. designed and conducted the *in vitro* experiments and data collection with help from D.A.J. B.R.C. provided supervision and resources for the *in vitro* experiments. D.B. designed and implemented the *in silico* computational model with input from A.R.G.L., D.A.J., and I.H. I.H. implemented the machine learning optimization of the *in silico* experiments with input from D.B. D.B. and D.A.J. wrote the image analysis script with contributions from A.R.G.L. D.B. and A.R.G.L. wrote the original manuscript draft with input from all co-authors. D.B., A.R.G.L., I.H., and D.A.J. wrote

the Supplemental Information. D.B. and A.R.G.L. prepared the figures and tables with input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 2, 2019

Revised: July 17, 2019

Accepted: October 23, 2019

Published: November 20, 2019

REFERENCES

- Bartocci, E., Gol, E.A., Haghghi, I., and Belta, C. (2016). A formal methods approach to pattern recognition and synthesis in reaction diffusion networks. *IEEE Trans. Control Network Syst.* *5*, 308–320.
- Bentley, K., Franco, C.A., Philippides, A., Blanco, R., Dierkes, M., Gebala, V., Stanchi, F., Jones, M., Aspalter, I.M., Cagna, G., Gerhardt, H., et al. (2014). The role of differential VE-cadherin dynamics in cell rearrangement during angiogenesis. *Nat. Cell Biol.* *16*, 309–321.
- Bredenoord, A.L., Clevers, H., and Knoblich, J.A. (2017). Human tissues in a dish: the research and ethical implications of organoid technology. *Science* *355*, eaaf9414.
- Briers, D., Haghghi, I., White, D., Kemp, M.L., and Belta, C. (2016). Pattern synthesis in a 3D agent-based model of stem cell differentiation. In proceedings of IEEE 55th conference on Decision and Control (CDC), 4202–4207.
- Brodland, G.W. (2002). The differential interfacial tension hypothesis (DITH): a comprehensive theory for the self-rearrangement of embryonic cells and tissues. *J. Biomech. Eng.* *124*, 188–197.
- Chandra, R.A., Douglas, E.S., Mathies, R.A., Bertozzi, C.R., and Francis, M.B. (2005). Programmable cell adhesion encoded by DNA hybridization. *Angew. Chem.* *118*, 910–915.
- Chung, B.G., Flanagan, L.A., Rhee, S.W., Schwartz, P.H., Lee, A.P., Monuki, E.S., and Jeon, N.L. (2005). Human neural stem cell growth and differentiation in a gradient-generating microfluidic device. *Lab Chip* *5*, 401–406.
- Cohen, W.W. (1995). Fast effective rule induction. In *Machine Learning Proceedings 1995* (Elsevier), pp. 115–123.
- Czirók, A., Varga, K., Méhes, E., and Szabó, A. (2013). Collective cell streams in epithelial monolayers depend on cell adhesion. *New J. Phys.* *15*, 075006.
- Deglincerti, A., Croft, G.F., Pietila, L.N., Zernicka-Goetz, M., Siggia, E.D., and Brivanlou, A.H. (2016). Self-organization of the *in vitro* attached human embryo. *Nature* *533*, 251–254.
- Demers, C.J., Soundararajan, P., Chennampally, P., Cox, G.A., Briscoe, J., Collins, S.D., and Smith, R.L. (2016). Development-on-chip: *in vitro* neural tube patterning with a microfluidic device. *Development* *143*, 1884–1892.
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* *34*, 184–191.
- Ducibella, T., and Anderson, E. (1975). Cell shape and membrane changes in the eight-cell mouse embryo: prerequisites for morphogenesis of the blastocyst. *Dev. Biol.* *47*, 45–58.
- Eberhart, R., and Kennedy, J. (1995). A new optimizer using particle swarm theory. Proceedings of the sixth international symposium on micro machine and human science (Minnesota Historical Society), pp. 39–43.
- Finkel, R.A., and Bentley, J.L. (1974). Quad trees a data structure for retrieval on composite keys. *Acta Inform.* *4*, 1–9.
- Foty, R.A., and Steinberg, M.S. (2005). The differential adhesion hypothesis: A direct evaluation. *Dev. Biol.* *278*, 255–263.
- Glen, C.M., McDevitt, T.C., and Kemp, M.L. (2018). Dynamic intercellular transport modulates the spatial patterning of differentiation during early neural commitment. *Nat. Commun.* *9*, 4111.
- Graner, F., and Glazier, J.A. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.* *69*, 2013–2016.
- Greber, D., and Fussenegger, M. (2010). An engineered mammalian band-pass network. *Nucleic Acids Res.* *38*, e174.
- Haghghi, I., Jones, A., Kong, Z., Bartocci, E., Gros, R., and Belta, C. (2015). SpaTeL: a novel spatial-temporal logic and its applications to networked systems. Proceedings of the 18th international conference on Hybrid Systems: computation and control, pp. 189–198.
- Heisenberg, C.P. (2017). D'Arcy Thompson's 'on growth and form': from soap bubbles to tissue self-organization. *Mech. Dev.* *145*, 32–37.
- Heisenberg, C.P., and Bellaïche, Y. (2013). Forces in tissue morphogenesis and patterning. *Cell* *153*, 948–962.
- Hookway, T.A., Butts, J.C., Lee, E., Tang, H., and McDevitt, T.C. (2016). Aggregate formation and suspension culture of human pluripotent stem cells and differentiated progeny. *Methods* *101*, 11–20.
- Hsiao, S.C., Shum, B.J., Onoe, H., Douglas, E.S., Gartner, Z.J., Mathies, R.A., Bertozzi, C.R., and Francis, M.B. (2009). Direct cell surface modification with DNA for the capture of primary cells and the investigation of myotube formation on defined patterns. *Langmuir* *25*, 6985–6991.
- Jackins, C.L., and Tanimoto, S.L. (1983). Quad-trees, oct-trees, and K-trees: A generalized approach to recursive decomposition of euclidean space. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-5*, 533–539.
- Jia, D., Dajusta, D., and Foty, R.A. (2007). Tissue surface tensions guide *in vitro* self-assembly of rodent pancreatic islet cells. *Dev. Dyn.* *236*, 2039–2049.
- Jones, E., Oliphant, T., Peterson, P., and others (2001). SciPy: Open source scientific tools for Python.
- Kay, R.R., and Thompson, C.R.L. (2009). Forming patterns in development without morphogen gradients: scattered differentiation and sorting out. *Cold Spring Harbor Perspect. Biol.* *1*, a001503.
- Kennedy, J. (2011). Particle Swarm Optimization. In *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, eds. (Springer). https://doi.org/10.1007/978-0-387-30164-8_630.
- Krieg, M., Arboleda-Estudillo, Y., Puech, P.H., Käfer, J., Graner, F., Müller, D.J., and Heisenberg, C.P. (2008). Tensile forces govern germ-layer organization in zebrafish. *Nat. Cell Biol.* *10*, 429–436.
- Libby, A.R., Joy, D.A., So, P.L., Mandegar, M.A., Muncie, J.M., Mendoza-Camacho, F.N., Weaver, V.M., Conklin, B.R., and McDevitt, T.C. (2018). Spatiotemporal mosaic self-patterning of pluripotent stem cells using CRISPR interference. *eLife* *7*, e36045.
- Ludwig, T.E., Levenstein, M.E., Jones, J.M., Berggren, W.T., Mitchen, E.R., Frane, J.L., Crandall, L.J., Daigh, C.A., Conard, K.R., Piekarczyk, M.S., et al. (2006). Derivation of human embryonic stem cells in defined conditions. *Nature Biotechnology* *24*, 185–187.
- MacKay, J.L., Sood, A., and Kumar, S. (2014). Three-dimensional patterning of multiple cell populations through orthogonal genetic control of cell motility. *Soft Matter* *10*, 2372–2380.
- Magno, R., Grieneisen, V.A., and Marée, A.F. (2015). The biophysical nature of cells: potential cell behaviours revealed by analytical and computational studies of cell surface mechanics. *BMC Biophys.* *8*, 8.
- Maître, J.L., Berthoumieux, H., Krens, S.F.G., Salbreux, G., Jülicher, F., Paluch, E., and Heisenberg, C.P. (2012). Adhesion functions in cell sorting by mechanically coupling the cortices of adhering cells. *Science* *338*, 253–256.
- Maître, J.L., and Heisenberg, C.P. (2013). Three functions of cadherins in cell adhesion. *Curr. Biol.* *23*, R626–R633.
- Mandegar, M.A., Huebsch, N., Frolow, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., Conklin, B.R., et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell* *18*, 541–553.
- Marcon, L., Diego, X., Sharpe, J., and Müller, P. (2016). High-throughput mathematical analysis identifies Turing networks for patterning with equally diffusing signals. *eLife* *5*, e14022.
- Marée, A.F.M., Grieneisen, V.A., and Hogeweg, P. (2007). The cellular Potts model and biophysical properties of cells, tissues and morphogenesis. In *Single-Cell-Based Models in Biology and Medicine*, M.A.J. Chaplain and K.A. Rejniak, eds., pp. 107–136.

- Merks, R.M.H., Perryn, E.D., Shirinifard, A., and Glazier, J.A. (2008). Contact-inhibited chemotaxis in de novo and sprouting blood-vessel growth. *PLoS Comput. Biol.* *4*, e1000163.
- Molitoris, J.M., Paliwal, S., Sekar, R.B., Blake, R., Park, J., Trayanova, N.A., Tung, L., and Levchenko, A. (2016). Precisely parameterized experimental and computational models of tissue organization. *Integr. Biol.* *8*, 230–242.
- Montero, J.A., and Heisenberg, C.P. (2004). Gastrulation dynamics: cells move into focus. *Trends Cell Biol.* *14*, 620–627.
- Nielsen, A.A.K., Der, B.S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E.A., Ross, D., Densmore, D., and Voigt, C.A. (2016). Genetic circuit design automation. *Science* *352*, aac7341.
- Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* *123*, 917–929.
- Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* *9*, 62–66.
- Ouchi, N.B., Glazier, J.A., Rieu, J.-P., Upadhyaya, A., and Sawada, Y. (2003). Improving the realism of the cellular Potts model in simulations of biological cells. *Phys. A* *329*, 451–458.
- Palsson, E. (2008). A 3-D model used to explore how cell adhesion and stiffness affect cell sorting and movement in multicellular systems. *J. Theor. Biol.* *254*, 1–13.
- Park, S., Kim, D., Jung, Y.-G., and Roh, S. (2015). Thiazovivin, a Rho kinase inhibitor, improves stemness maintenance of embryo-derived stem-like cells under chemically defined culture conditions in cattle. *Animal Reproduction Science* *161*, 47–57.
- Pir, P., and Le Novère, N. (2016). Mathematical models of pluripotent stem cells: at the dawn of predictive regenerative medicine. *Methods Mol. Biol.* *1386*, 331–350.
- Sasai, Y. (2013). Cytosystems dynamics in self-organization of tissue architecture. *Nature* *493*, 318–326.
- Sekine, R., Shibata, T., and Ebisuya, M. (2018). Synthetic mammalian pattern formation driven by differential diffusivity of Nodal and Lefty. *Nat. Commun.* *9*, 5456.
- Sharpe, J. (2017). Computer modeling in developmental biology: growing today, essential tomorrow. *Development* *144*, 4214–4225.
- Sohka, T., Heins, R.A., Phelan, R.M., Greisler, J.M., Townsend, C.A., and Ostermeier, M. (2009). An externally tunable bacterial band-pass filter. *Proc. Natl. Acad. Sci. USA* *106*, 10135–10140.
- Steinberg, M.S. (1975). Adhesion-guided multicellular assembly: A commentary upon the postulates, real and imagined, of the differential adhesion hypothesis, with special attention to computer simulations of cell sorting. *J. Theor. Biol.* *55*, 431–443.
- Szabó, A., Ünnepp, R., Méhes, E., Twal, W.O., Argraves, W.S., Cao, Y., and Czirók, A. (2010). Collective cell motion in endothelial monolayers. *Phys. Biol.* *7*, 046007.
- Taylor, H.B., Khuong, A., Wu, Z., Xu, Q., Morley, R., Gregory, L., Poliakov, A., Taylor, W.R., and Wilkinson, D.G. (2017). Cell segregation and border sharpening by Eph receptor–ephrin-mediated heterotypic repulsion. *J. R. Soc. Interface* *14*, 20170338.
- 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.
- Théry, M. (2010). Micropatterning as a tool to decipher cell morphogenesis and functions. *J. Cell Sci.* *123*, 4201–4213.
- Théry, M., Racine, V., Piel, M., Pépin, A., Dimitrov, A., Chen, Y., Sibarita, J.B., and Bornens, M. (2006). Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. *Proc. Natl. Acad. Sci. USA* *103*, 19771–19776.
- Toda, S., Blauch, L.R., Tang, S.K.Y., Morsut, L., and Lim, W.A. (2018). Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science* *361*, 156–162.
- Van Liedekerke, P., Palm, M.M., Jagiella, N., and Drasdo, D. (2015). Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. *Comp. Part. Mech.* *2*, 401–444.
- van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T. (2014). Scikit-image: image processing in Python. *PeerJ* *2*, e453.
- Voss-Böhme, A. (2012). Multi-scale modeling in morphogenesis: a critical analysis of the cellular Potts model. *PLoS One* *7*, e42852.
- 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.
- Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K., et al. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnology* *25*, 681–686.
- White, D.E., Kinney, M.A., McDevitt, T.C., and Kemp, M.L. (2013). Spatial pattern dynamics of 3D stem cell loss of pluripotency via rules-based computational modeling. *PLoS Comput. Biol.* *9*, e1002952.
- White, D.E., Sylvester, J.B., Levario, T.J., Lu, H., Strelman, J.T., McDevitt, T.C., and Kemp, M.L. (2015). Quantitative multivariate analysis of dynamic multicellular morphogenic trajectories. *Integr. Biol.* *7*, 825–833.
- Witten, I.H., Frank, E., Hall, M.A., and Pal, C.J. (2016). Data mining: practical machine learning tools and techniques (Morgan Kaufmann).
- Ziomek, C.A., and Johnson, M.H. (1980). Cell surface interaction induces polarization of mouse 8-cell blastomeres at compaction. *Cell* *21*, 935–942.

STAR★METHODS

KEY RESOURCE TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|---|---|
| Antibodies | | |
| E-cadherin Antibody | Abcam Cat# ab1416 | RRID:AB_300946 |
| Oct-3/4 (C-20) antibody | Santa Cruz Biotechnology Cat# sc-8629 | RRID:AB_2167705 |
| SOX2 antibody [9-9-3] | Abcam Cat# ab79351 | RRID:AB_10710406 |
| Hoechst 33258 antibody | Thermo Fisher Scientific Cat# H3569 | RRID:AB_2651133 |
| Experimental Models: Cell Lines | | |
| CRISPRi-Gen1C | Conklin Lab (Mandegar et al., 2016) | NA |
| CRISPRi-Gen2 | Conklin Lab (Mandegar et al., 2016) | NA |
| Oligonucleotides | | |
| ROCK1 guide sequence : CGGGGCGCGG ACGCTCGGAA | Integrated DNA Technologies | NA |
| CDH1 guide sequence: GCAGTCCGACG CCACTGAG | Integrated DNA Technologies | NA |
| CDH1 (90%) guide sequence: TCACCGCG TCTATGCGAGGC | Integrated DNA Technologies | NA |
| CDH1 (75%) guide sequence: CCCG TACCGCTGATTGGCTG | Integrated DNA Technologies | NA |
| CDH1 (70%) guide sequence: TCAGCCA ATCAGCGGTACGG | Integrated DNA Technologies | NA |
| Software and Algorithms | | |
| Python Programming Language Modules: scipy, numpy, matplotlib, pandas, seaborn, scikit-image, numpy, scipy, scikit-image | http://www.python.org | RRID:SCR_008394 |
| Morpheus, v1.9.1 | https://imc.zih.tu-dresden.de/wiki/morpheus | RRID:SCR_014975 |
| MATLAB | http://www.mathworks.com/products/matlab/ | RRID:SCR_001622 |
| WEKA | http://www.cs.waikato.ac.nz/ml/weka/ ; Witten et al., 2016 | RRID:SCR_001214 |
| TSSL and all custom data analysis code | This work | https://github.com/dmarcbriers/Multicellular-Pattern-Synthesis |

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources, reagents, or source code should be directed to the Lead Contact, Todd McDevitt (todd.mcdevitt@gladstone.ucsf.edu).

Materials Availability

This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

All work with human iPSC lines was approved by the University of California, San Francisco Human Gamete, Embryo and Stem Cell Research (GESCR) Committee. Cell lines were derived from the parent line WTC (Coriell Cat. # GM25256) where the species of origin was confirmed by a LINE assay. All cell lines tested negative for mycoplasma using a MycoAlert Mycoplasma Detection Kit (Lonza). All human induced pluripotent stem cells (hiPSCs) were cultured at 37°C, seeded at a density of 12,000 cells per cm² in feeder-free media conditions on growth factor-reduced matrigel (BD Biosciences), and daily fed MTeSRTM medium (STEMCELL Technologies)

(Ludwig et al., 2006). When hiPSC confluency reached 75%, cells were dissociated and singularized using Accutase (StemCell Technologies). Single cells were counted using a Invitrogen Countess Automated Cell Counter (ThermoFisher Scientific), re-plated at previously described density, and in the first 24hrs after passaging, fed with MTeSR™ medium supplemented with the small molecule Rho-associated coiled-coil kinase (ROCK) inhibitor Y-276932 (10 μ M; Selleckchem) to promote survival (Park et al., 2015; Watanabe et al., 2007).

METHOD DETAILS

Generation of CRISPRi Knockdown iPSC Lines

CRISPRi knockdown lines were previously generated as described in (Mandegar et al., 2016), where 20 base pair guides were designed using the Broad Institute sgRNA design website (Doench et al., 2016). 20-base pair sequences were cloned into the gRNA-CNKB vector using restriction enzyme BsmBI digestions, followed by ligation with T4 DNA ligase as described in (Mandegar et al., 2016). 200,000 cells of the CRISPRi-Gen1C or CRISPRi-Gen2 hiPSC lines from the Conklin Lab were nucleofected with individual gRNA vectors using the Human Stem Cell Nucleofector Kit 1 solution with the Amaxa Nucleofector 2b device (Lonza). Cells were then plated at increasing dilutions into 3 wells of a 6-well plate coated with growth factor-reduced Matrigel (BD Biosciences) in MTeSR™ supplemented with Y-276932 (10 μ M) for 2 days. Then the nucleofected hiPSCs were treated with blasticidin (10 μ g/ml) for a selection period of 7 days. Surviving colonies for each gRNA were pooled and passaged in MTeSR™ with blasticidin (10 μ g/ml) and Y-27632 (10 μ M) for a single day then transitioned to MTeSR™ media only. After stable polyclonal populations of hiPSCs were established for each gRNA, cells were karyotyped by Cell Line Genetics (Libby et al., 2018) (Figure S8). Finally, knock-down efficiency was tested by the addition of doxycycline (2 μ M) to the culture media for 6 days and subsequent qPCR of mRNA levels of respective genes compared to time matched controls of the same line without CRISPRi induction.

Mixed Colony Generation

Mixed population hiPSC colonies were generated using forced aggregation via PDMS microwells in a 24-well tissue culture plate (~975 400 \times 400- μ m wells per well) (Hookway et al., 2016; Libby et al., 2018). hiPSCs were dissociated and singularized using Accutase (StemCell Technologies) and subsequently counted using a Invitrogen Countess Automated Cell Counter (ThermoFisher Scientific). The proper ratios of cells to create 100 cell aggregates were then seeded into PDMS wells in MTeSR™ with Y-27632 (10 μ M), centrifuged at 200g for 5 minutes, and allowed to compact overnight (~18hrs). Aggregates were then washed-out of the PDMS wells with fresh MTeSR™ and re-plated into a growth factor reduced Matrigel (BD Biosciences) coated PerkinElmer CellCarrier™-96 plates at ~10/aggregates/cm² and fed daily with MTeSR™.

Immunofluorescence Staining and Imaging

Human iPSCs were fixed for 25 minutes with 4% paraformaldehyde (VWR) and subsequently washed 3 times with PBS. Fixed colonies were simultaneously blocked and permeabilized with a 1 \times PBS solution with 0.3% Triton X-100 (Sigma Aldrich) and 5% Normal Donkey Serum (Jackson ImmunoResearch) for 1 hour at room temperature. Samples were then incubated with primary antibodies overnight at 4°C in a 1 \times PBS solution with 1% bovine serum albumin (Sigma Aldrich) and 0.3% Triton-X. Samples were washed 3 times and then incubated for 1 hour at room temperature with secondary antibodies. Primary antibodies used were: anti-OCT4 (SantaCruz 1:400), anti-SOX2 (AbCAM 1:400), and anti-Ecadherin (AbCAM 1:200). All secondary antibodies were used at 1:1000 and purchased from Life Technologies. Images were taken in one focal plane on the apical surface of hiPSC colonies.

Mixed colonies were imaged using a Zeiss Observer.Z1 (Zeiss) and an InCell Analyzer2000 (GE Healthcare) with a 10 \times objective, and confocal images were obtained using a Zeiss LSM880 Confocal w/ Airyscan (Zeiss) microscope with a 10 \times objective. Images were analyzed in ImageJ and in python using the skimage package (van der Walt et al., 2014).

Protein Quantification

Protein quantification for CDH1 KD was first quantified by immunofluorescence imaging of mixed colonies of WT-GFP hiPSCs and CDH1 KD colonies (Libby et al., 2018). Total fluorescence of CDH1 was measured by a python script that compared fluorescence of the CDH1 channel normalized to the amount of WT cells vs KD cells (determined by GFP fluorescence) (Figure S2). This data was supplemented by Western blot data from the previously published KD of CDH1 and ROCK1 in (Libby et al., 2018).

mRNA Quantification

The relative gene expression following CRISPRi knockdown was previously reported in (Libby et al., 2018) and used as a reference to establish knockdown timing curves used in our *in silico* simulations. As previously reported (Libby et al., 2018), total mRNA isolation from dissociated hiPSCs was performed using an RNeasy Mini Kit (Qiagen) according to manufacturer's instructions and quantified with a Nanodrop 2000c Spectrometer (ThermoFisher). Obtained mRNA was then used to synthesize cDNA using an iScript cDNA Synthesis kit (BioRad). A StepOnePlus Real-Time PCR system (Applied Biosciences) was used to quantify and detect gene expression by Fast SYBR Green Master Mix (Thermo Fisher Scientific). Relative gene expression was determined by normalizing comparative threshold(Ct) values to the house keeping gene 18S rRNA. Gene expression was then displayed as a fold change comparison to the day 0 control before the start of gene knockdown. The NCBI Primer-BLAST website was used to design the primers. Statistical analysis was conducted using a two-tailed unpaired t-test between any two groups (p<0.05, n=3).

Time Lapse Imaging

Mixed hiPSC colonies were imaged at the basal surface on optically clear PerkinElmer CellCarrier™-96 plates on an inverted AxioObserver Z1 (Zeiss) with an ORCA-Flash4.0 digital CMOS camera (Hamamatsu) with a 10× objective, where that single plane was used for parameter estimations. Using ZenPro software, colony locations were mapped and a single colony was imaged every 30 minutes over the course of 12 hours. Time lapse imaging occurred from hours 24–36 and from hours 96–108 after mixed colony plate down. The 12 hours series of images were then used to compare *in silico* to *in vitro* pattern formation and organization of cells. Additionally, mixed colonies of wildtype and CRISPRi-Gen1C without knockdown guides were imaged for 6 hours every 5 minutes with a 20× objective from hours 60–66 after plate down. These 6 hours image series were used to generate velocity values as previously described in Section 2.2 (Velocity Characterization).

Comparison of *In Vitro* and *In Silico* Spatial Patterns

We used *in vitro* and *in silico* images to calculate the total number of cells in an image, the number of clusters, and the circularity of each cluster (Figure S1). Our work-flow for comparing patterns (Figure S9) involved splitting the images into single color RGB channels using the python module scikit-image (van der Walt et al., 2014).

For *in silico* images each channel represented a different cell type. After splitting the image into color channels we detected the number of islands in a colony. For *in silico* images, cells were separated by a black border so we sequentially masked out the border, dilated the image, removed small objects, then removed small holes in the mask with scikit-image. Contiguous regions (8-connected) were considered clusters. We then overlaid a mask of individual cells onto each cluster using a logical AND comparison of the image masks to determine if the cell cluster met our criteria to be considered an island. Using only the cell clusters we considered islands, we then calculated the number of cells per island and the circularity of the islands using the formula $Circularity = 4\pi * Area/Perimeter^2$.

In contrast to the work-flow for simulation images, for *in vitro* images one channel represented all cell nuclei and the other channel represented cells stained for the protein CDH1, which delineated the CDH1 knockdown cells from the WT or ROCK1 knockdown cells. For the *in vitro* images, the CDH1 channel was thresholded and then dilated to create a CDH1+ cell mask followed by removal of small objects and holes to create a smooth segmentation. To generate the island masks, isolated CDH1 negative clusters were identified using the “label” function on the inverse of the CDH1+ mask. Individual cells were localized by detecting local maximum intensity in the DAPI channel images then the number of DAPI peaks per island were calculated using the logical AND of the island and CDH1 negative masks. Finally, we used the function “regionprops” to calculate the cluster area and perimeter for each island, which were then employed to calculate the island circularity with the above formula.

BMP4 Differentiations

Successfully patterned hiPSC colonies were differentiated for 48hrs in MTeSR™ cell culture medium (StemCell Technologies) supplemented with BMP4 (R&D Systems, Minneapolis, United States of America) at a 50 μM/ml concentration. The colonies were then fixed with 4% PFA for 25 min and subsequently analyzed.

Cellular Potts Model Environment

We modeled the mechanical properties of interacting human induced pluripotent stem cells (hiPSCs) with an extended cellular Potts model (CPM). In the model of mechanically driven self-organization in hiPSCs, cell–cell interaction mechanics were explained by four physical properties of cells. 1) cell-cell adhesion, 2) cortical tension, 3) conservation of volume, 4) and directionally persistent cell migration. Below, we describe how the extended CPM was used to recapitulate spatiotemporal patterns and predictively design de novo spatiotemporal behaviors.

We defined the environment of a CPM simulation S on a 2D square lattice domain $S \in \mathbb{Z}_+^2$. Each lattice site, $x = (m, n) \in S$, represented a coordinate location where $m \in \mathbb{Z}_+$ and $n \in \mathbb{Z}_+$ were the horizontal and vertical coordinates of each lattice site respectively. The spatial resolution of each lattice site was $1 \mu\text{m}^2$ so that each square region of the grid is equal to 1 square micrometer.

To represent the location of hiPSCs, each lattice site x was assigned a value σ_x , conventionally called the *spin* or cell index (cell ID) of a site, from the set of cell indices $k \in K$ given $K = \{1, \dots, N(t)\}$ where $N(t)$ was the number of cells in the simulation at time t . Lattice sites that represent empty space where there is no hiPSC covering the lattice site were assigned a cell index of 0. In the CPM, a cell C_k was composed of multiple lattice sites where each lattice site represents a partial region of a cell or the surrounding media. A cell C_k was defined as the set of lattice sites with the same cell ID $C_k = \{x \in S : \sigma_x = k\}$. Since a single cell was composed of multiple lattice sites, the CPM was able to capture fluctuations in a cell’s shape with a granularity that is not possible with Type A cellular automaton or center-based models (Van Liedekerke et al., 2015). Each cell was also assigned a cell type τ to notate the type of genetic perturbation (i.e. KD) of that cell which determined its intracellular and extracellular behaviors.

Next, we summarize two common metrics to describe cell morphology in a CPM simulation; cell area and cell membrane length. These metrics are important since their values in the model were directly measured from microscopy images. For a discussion of these metrics see Magno et al. (2015) and Voss-Böhme (2012).

Given that each lattice site had an area of $1 \mu\text{m}$, the area of a cell at time t in the simulation was defined as the number of lattice sites encompassed by a cell:

$$a_{k,t} = |\{x \in S : \sigma_x = k\}|, \quad (\text{Eq } 1)$$

where $|\cdot|$ denoted the cardinality of a set.

The time varying membrane length of a cell $p_{k,t}$, synonymously called the perimeter or surface length in other studies, was defined as the number of lattice interfaces bordering other cells or empty space:

$$l_{k,t} = 0.5 \times \sum_{\text{interfaces}\{x,x'\}} \delta(k, \sigma_x), \quad (\text{Eq 2})$$

where x' represented any of the lattice sites adjacent to x , $(m \pm 1, n) \vee (m, n \pm 1)$ in 2D.

The Kronecker symbol δ was defined by $\delta(u, v) = 1$ if $u = v$ and $\delta(u, v) = 0$ if $u \neq v$. An interface (x, x') was a shared border between lattice sites. To avoid counting adjacent lattice sites inside a cell, the CPM only summed interfaces between lattice sites with different cell ID's; when $\delta(\sigma_x, \sigma_{x'}) = 0$. Put simply, we were measuring the perimeter of each stem cell.

Cellular Potts Model Dynamics

The CPM uses a function called the **Hamiltonian** H to describe the energy (favorable behaviors) for any *configuration* of cells. Cell motility evolved by choosing a random lattice site x , a region of a cell-cell interface or a cell-media interface and attempted to copy it to a random neighboring lattice site x' . The Hamiltonian was defined as the sum of four constraints that represent four physical properties of simulated stem cells: 1) conservation of cell area, 2) locally polarized cell migration, 3) cell-cell adhesion, 4) and cell membrane length which commonly represents cortical tension. In the CPM, the goal was to minimize the Hamiltonian or minimize violations of the desired cellular behaviors. Therefore, each constraint calculated a decrease (reward) or increase (penalty) in the configuration energy due the collective properties of cells in the simulation.

When a change in a lattice site was proposed, this affects H . If the proposed change was accepted, the change in H was defined as ΔH . A proposed change for a cell's lattice site was accepted with the following probability:

$$\text{if } \Delta H < -Y, P(\sigma \rightarrow \sigma') = 1 \quad (\text{Eq 3})$$

$$\text{otherwise, } P(\sigma \rightarrow \sigma') = e^{-(\Delta H + Y)/T}, \quad (\text{Eq 4})$$

where the yield $Y = 0.1$ and the temperature $T = 10$.

Simply, if the proposed change in local cell position resulted in less energy, then the change was accepted. If the proposed update would have resulted in greater energy (ΔH), then the change was only accepted with a very low probability. In this way, complex behaviors such as preferential cell-cell adhesions, cortical tension, and cell migration, are represented by a score and a weight, where the score represents a reward or penalty depending on the divergence of a cell from its target behavior, while the weight represents the relative importance of the respective cell behavior.

CPM Configuration Energy

The free energy for a configuration of cells was defined as the sum of four constraints: local cell-cell/cell-ECM adhesion, cell area conservation, cell membrane length, and locally polarized cell migration:

$$H = H_{\text{adhesion}} + H_{\text{area}} + H_{\text{membraneLength}} + H_{\text{migration}} \quad (\text{Eq 5})$$

For a configuration of cells, the free energy due to cell adhesion was

$$H_{\text{adhesion}} = \sum_{k \in K} J_{\tau(\sigma_x), \tau(\sigma_{x'})} (1 - \delta(\sigma_x, \sigma_{x'})), \quad (\text{Eq 6})$$

where $J_{\tau(\sigma_x), \tau(\sigma_{x'})}$ represented the cell adhesion strength between lattice sites σ_x and $\sigma_{x'}$ that was defined by their cell type $\tau(\sigma_{x'})$.

$(1 - \delta_{\sigma_x, \sigma_{x'}})$ restricted these calculations to interfaces between cells instead of all lattice sites, and improved the efficiency of the simulation. Although not explicit in our notation, the cell adhesion strength was a time-dependent function controlled by protein expression to mimic changes in cell behavior with inducible gene knockdown. The energy due to cells resisting changes from their resting area was defined as

$$H_{\text{area}} = \sum_{k \in K} \lambda_a (a_{k,t} - A_{k,t})^2, \quad (\text{Eq 7})$$

where $A_{k,t}$ represented the target area of a cell, $a_{k,t}$ represented the current area of a cell, and λ_a was the relative strength of area conservation term.

The cortical tension constraint was defined as:

$$H_{\text{membraneLength}} = \sum_{k \in K} \lambda_l (l_{k,t} - L_{k,t})^2, \quad (\text{Eq 8})$$

where $l_{k,t}$ represented the current membrane length of a cell at time t , $L_{k,t}$ was the target membrane length, and λ_l was the strength of the cortical tension constraint.

As a proxy for increasing or decreasing the cell membrane length, the Equivalent Circular Perimeter (ECP) was used to set the membrane length for a cell given its current area. The ECP of a non-circular 2D object was defined as the perimeter of a circle with equivalent surface area as the non-circular object:

$$ECP(k) = 2\sqrt{a_{k,t}\pi}.$$

The target membrane length was calculated using a membrane length proportionality constant:

$$L_{k,t} = r_k ECP(k), \quad (\text{Eq 9})$$

where r_k was the membrane proportionality constant.

To find r_k the membrane length and area of cells were measured and divided by the ECP of the cell. This ratio of membrane length to ECP was equal to the membrane proportionality constant r_k . The ECP allowed us to calculate the membrane length of a cell of any area that would have a comparable shape to empirical measurements.

To capture directionally persistent cell migration, we modeled "polarized cell migration" as the tendency of cells to bias their movement in the same direction as their previous direction of movement as described in Czirók et al. (2013) and Szabó et al. (2010). Cells had a target direction $\vec{t} \rightarrow$ based on previous movement where CPM updates in this direction were preferred (they decreased the energy in H). For each copy attempt $x \rightarrow x'$, the cell center was displaced in direction \vec{s}' . The change in energy due to migration in this direction was defined as:

$$H_{migration} = -\mu a_{k,t} (\vec{t} \cdot \vec{s}'), \quad (\text{Eq 10})$$

where μ was the strength of cell migration, $a_{k,t}$ was the cell area at time t , \vec{t} was a unit vector giving the target direction, and \vec{s}' was a unit vector giving the current direction of a stem cell if the CPM update ($x \rightarrow x'$) was to be accepted.

The function was multiplied by -1 since updates in the direction of \vec{t} have a dot product that approached +1 as the angle between \vec{t} and \vec{s}' approached zero. Multiplying by -1 resulted in decreased configuration energy for cells moving in the same direction as the target direction vector.

For every MCS, the target direction at any time (\vec{t}_t) was updated continuously given the displacement of a cell's centroid $w\Delta O = \mathbf{O}_t - \mathbf{O}_{t-1}$ then transformed into a unit vector $\vec{s} = \Delta O / |\Delta O|$. This target direction included how 'decay-time' D of the previous direction and the current cell displacements contributed to the current polarity of the cell (Szabó et al., 2010):

$$\vec{t}_t = (1 - D)\vec{t}_{t-1} + D\vec{s}. \quad (\text{Eq 11})$$

Physical Units and Other Cellular Phenomena of CPM

Cell division was symmetrical (the parent cell divided into 2 equally sized daughter cells), and the timing of cell division was asynchronous. This was achieved by assigning a uniformly distributed "division counter" d_c for each cell at $t = 0$ between 0 and the division time d_t . This counter was incremented at each time step of the simulation, and a cell would divide when $d_c = d_t$. d_c was then reset to 0 for both daughter cells. Cell division was assumed to be asynchronous amongst the population, and cell division times specific to each type of knockdown were incorporated into the model to provide an accurate depiction of population growth kinetics. Cell division times were calculated from *in vitro* doubling rates and modeled to be 18 hours for CDH1(-) cells, and 20 hours for all other cells.

Model Fitting to Empirical data

In the main text we provide a brief explanation of the characterization experiments to fit our computational model. Here we describe the mathematical transformations, mapping functions, and model parameters associated with these characterization experiments. The parameters fit during this process are summarized in Table S2.

Morphology Characterization

Three types of colonies were characterized; purely wildtype, wildtype and CDH1 knockdown in a 1:1 ratio, and wildtype:ROCK1 knockdown in a 1:1 ratio. We measured the cell area, perimeter, and ECP at the center and periphery of colonies (Figure S1). The median cell area was used to set the target cell area ($A_{k,t}$) in our simulations (Figure S1; Table S2). The median membrane proportionality constant ($r_k = \text{perimeter} / \text{ECP}$) was used to set the target membrane length $L_{k,t}$ in our simulations (Figure S1; Table S2).

Velocity Characterization (Space versus Time)

We characterized space vs. time by measuring the velocity (change in distance over time) of wildtype cells in dense colonies. Mixed aggregates of 90% WT and 10% CRISPRi cells without a targeting guide were generated. With the addition of doxycycline (DOX) to the cell culture media, the CRISPRi no guide population expressed a cytoplasmic mCherry marker which allowed individual cells to be distinguished from the untagged WT background (Figure 2B). 24 colonies were imaged for 6 hours at 5 minutes/image at 20x magnification creating a time series of 73 frames. Each frame was individually normalized and thresholded using non-local means (Otsu, 1979). Cell migration tracks were generated by following matching contours between frames where matching contours share at least ten pixels overlap. We used watershed segmentation to separate adjacent cells. Instantaneous frame to frame velocity was calculated as

$$V_{inst} = [(x_{cm,2} - x_{cm,1}) / \Delta t, (y_{cm,2} - y_{cm,1}) / \Delta t], \quad (\text{Eq 12})$$

where $x_{cm,2}$ was the center of mass of each segmented cell body at the currently observed frame and $x_{cm,1}$ was the center of mass of each segmented cell body at the previous frame, and Δt was 5 minutes.

Taking the average magnitude of the per-cell instantaneous velocity over 24 colonies gave a median velocity of 0.29 $\mu\text{m}/\text{minute}$. We then ran parameter sweeps to fit model parameters that affect cell migration (Table S1):

- MCS - copy attempts per simulation hour
- $J_{WT,WT}$ - adhesion energy or reward per micrometer of cell border between wildtype cells
- μ - strength of self-propulsion
- λ_j - strength of cortical tension

We chose the parameter combination where the simulation velocity distribution matched the empirical velocity distribution and remained a dense colony (0.34 $\mu\text{m}/\text{minute}$). It is important to note that we chose optimal model parameters using the distribution of cell velocities and not the median cell velocity. We ran 24 simulations to mimic the the experimental setup of the *in vitro* characterization. Using the Mann-Whitney U test, there was no significant difference in the distribution of cell velocities; p-value threshold of 0.05 and p-value for 24 *in silico* colonies was 0.051. After fitting the model to empirical data of cell morphology and velocity, we could recapitulate the cell morphology and collective cell migration of wildtype stem cell colonies without genetic modulation (Videos S1 and S2).

Temporal Knockdown Characterization (Protein Expression versus Time) - CDH1

We characterized the time-dependent knockdown of CDH1 expression which was responsible for changes in cell-cell adhesion (J). CDH1 was knocked down using CRISPRi, and the relative expression was measured for 6 consecutive days. The relative mRNA expression of CDH1 was quantified by quantitative PCR ($n=3$) and protein expression of CDH1 was measured by immuno-fluorescence microscopy ($n=10$), which displayed a 24 hours delay from the mRNA knockdown. The data was min-max normalized to a domain of [0,1] using the median expression for each day:

$$y(t_k)' = (y(t_k) - y_{min}) / (y_{max} - y_{min}), \quad (\text{Eq 13})$$

where t_k was the time since the knockdown, $y(t_k)$ was the expression at k hours after knockdown, y_{max} was the max expression from all days, and y_{min} was the minimum expression over all days. It is important to note that t_k is the time since knockdown and t is time since the initiation of co-culture experiments. Using least squares optimization (Python `scipy.optimize.curve_fit` function), the K_m (repression coefficient) and n (Hill coefficient) of the Hill Function for repression were fit to the normalized median expression to create a response function using least squares optimization:

$$F(t) = 1 / (1 + (K_m/t_k)^n), \quad (\text{Eq 14})$$

where K_m was the time half expression occurs, and n was the hill coefficient. Functions were fit to the normalized median expression (per day) to create a time-dependent response function for CDH1 knocked down to 90%, 75%, 70%, and 2% of the original mRNA expression (Figures 2F and S3). This range of knockdown efficiencies allowed us to computationally model how differing levels of CDH1 expression could impact spatial patterning. Given value for the parameters k_m and n (Table S2) we now had a continuous response function for the expression of CDH1 given a knockdown time that we could modulate. It is important to note that normalizing the relative expression to a domain of [0, 1] allows us to stretch the response function to different parameter ranges in the spatial pattern characterization experiments.

Temporal Knockdown Characterization (Protein Expression versus Time) - ROCK1

Using the same approach as the CDH1 knockdown, we created a Hill response function for ROCK1 knocked down to 20% mRNA expression. ROCK expression is represented by the strength of cortical tension (λ_j) parameter in our computational model. We assumed mRNA expression changed 24 hours ahead of protein expression, so we shifted the time axis forward by one day to account for the delay. The median expression for each day was min-max normalized to a domain of [0,1] (Equation S13). The K_m (repression coefficient) and n (Hill coefficient) of the hill function for repression were fit to the normalized median expression to create a response function using least squares optimization (Equation S14; Figure 2; Table S2). Due to the delay in protein knockdown compared to mRNA levels, the Hill functions were shifted by 24 hours to account for the delay in protein loss (Figure 2G), allowing us to model the average change in ROCK1 protein expression for individual cells over time.

Spatial Pattern Characterization (Protein Expression versus Space)

Given the previous characterization experiments, we were able to model cell proliferation, cell morphology, wildtype cell migration, and temporal changes in the expression of CDH1 and ROCK1. However, the time-dependent modulation of protein express had to be mapped to changes in multicellular self-organization. To model the the time-dependent modulation of cell-cell adhesion via CDH1 and cortical tension via ROCK1, fluorescent microscopy images were collected 96 hours after mixing either ROCK1 KD or CDH1 KD cells with wildtype hiPSCs. Then *in silico* parameter sweeps were run overlaying a range of parameters controlling the strength of adhesion or membrane stiffness. These two parameters rescaled their respective Hill Functions and produced a range of spatial patterns due to progressively weaker cell-cell adhesion or progressively stiffer cell membrane parameter values. We then conducted

double-blind experiments to fix adhesion strength and membrane stiffness parameters which most closely matched *in vitro* spatial patterning for CDH1 and ROCK1 knockdowns respectively (Figure 2H and 2I).

$$F^i(t) = y_{knockdown} + (y_{wildtype} - y_{knockdown}) * (1 + (K_m/t)^n), \quad (\text{Eq 15})$$

where $y_{knockdown}$ was the adhesion strength (J) or target membrane length (L_k, t) of knockdown cell lines in the model, and $y_{wildtype}$ was the adhesion strength (J) or target membrane length of wildtype cells in the model.

In Equation 15, we scaled the normalized response function from $[0, 1]$ to the range of model parameters $[y_{knockdown}, y_{wildtype}]$. Given the characterization experiments of cell morphology, cell migration velocity, time-dependent modulation of cell mechanics, and the resulting spatial organization, the computational model was able to recapitulate the spatial patterning due to the CDH1 and ROCK1 knockdowns (Videos S3, S4, S5, and S6).

TSSL Scoring and Pattern Optimization

In order to automatically compare patterns produced by the model from different parameterizations and determine optimal parameter values, we needed a measure capable of quantifying how close any given pattern was to the desired one. A very effective algorithm was proposed in Bartocci et al. (2016) for this purpose.

Quad-Tree Representation of an Image

Consider an RGB representation of an $m \times n$ image as the matrix A where the element $a_{ij} = \langle a_{ij}^{(r)}, a_{ij}^{(g)}, a_{ij}^{(b)} \rangle$ is the normalized RGB values for the pixel located on the i th row and j th column of the image. Thus,

$$0 \leq a_{ij}^{(c)} \leq 1 \text{ for } c \in \{r, g, b\}.$$

Given a matrix A , $A[i_s, i_e; j_s, j_e]$ was used to denote the submatrix created by selecting rows with indices from i_s to i_e and columns from j_s to j_e . Following, we represented the matrix A as a quad-tree. A quad-tree $Q = (\mathcal{V}, R)$ is a quaternary tree representation of matrix A where each vertex $v \in \mathcal{V}$ represents a submatrix of A and the relation $R \subset \mathcal{V} \times \mathcal{V}$ defines four children for each vertex that is not a leaf.

Figures 3Ai and 2Aii demonstrates how a quad-tree is built from a matrix. In this figure, we label each edge in the quad-tree with the direction of the sub-matrix represented by the child: north west (NW), north east (NE), south west (SW), and south east (SE). In Figure 3Ai:

- v_0 represents the complete matrix A at quadrant level 1
- v_1 represents the first quadrant of level 2 or $A[1, \lfloor m/2 \rfloor; 1, \lfloor n/2 \rfloor]$, where m is the total number of rows and n is the total number of columns in A
- v_2 represents $A[\lfloor m/2 \rfloor + 1, m; 1, \lfloor n/2 \rfloor]$
- v_3 represents $A[\lfloor m/2 \rfloor + 1, m; \lfloor n/2 \rfloor + 1, n]$
- v_4 represents $A[1, \lfloor m/2 \rfloor; \lfloor n/2 \rfloor + 1, n]$
- v_5 represents represents the first quadrant of Level 3 or $A[1, \lfloor m/4 \rfloor; 1, \lfloor n/4 \rfloor]$,
- etc

We used the procedure described in Briers et al. (2016) to construct quad-trees, which is slightly different from Bartocci et al. (2016). In Bartocci et al. (2016), the assumption was made that A has a size of $2^k \times 2^k$ so that each submatrix could be divided into four equal-sized partitions. Here, we relaxed this requirement by allowing non-equal submatrices to be children of a node. Furthermore, Bartocci et al. (2016) defined a leaf as a vertex of the quad-tree for which all the elements of a submatrix had the same values. While this approach works perfectly for the 32×32 network that is studied in that paper, it can be problematic for larger images since the number of vertices in a quad-tree grows exponentially as more levels are added to it. In this paper, we constructed quad-trees with a fixed depth of 5, regardless of the size and other characteristics of A .

The representation function $\mu^{(c)}(v) : \mathcal{V} \rightarrow [0, b] \times [0, b]$ was defined for sub-matrix $A[i_s, i_e; j_s, j_e]$ represented by vertex $v \in \mathcal{V}$ of the quad-tree $Q = (\mathcal{V}, R)$ as follows :

$$\mu^{(c)}(v) = \left(\mu_1^{(c)}, \mu_2^{(c)} \right)$$

$$\mu_1^{(c)} = \frac{1}{(i_e - i_s + 1)(j_e - j_s + 1)} \sum_{i_j \in \{i_s, \dots, i_e\} \times \{j_s, \dots, j_e\}} a_{ij}^{(c)}, \quad (\text{Eq 16})$$

$$\mu_2^{(c)} = \frac{1}{(i_e - i_s + 1)(j_e - j_s + 1)} \sum_{i_j \in \{i_s, \dots, i_e\} \times \{j_s, \dots, j_e\}} \left(a_{ij}^{(c)} - \mu_1^{(c)} \right)^2,$$

where $c \in \{r, g, b\}$ was an RGB color. The function $\mu^{(c)}$ provided the mean value and variance for the concentration of RGB colors in a particular region of the space represented by the vertex v .

Quad-trees can be interpreted as multi resolution representation of images, as the nodes that appear in deeper levels provide statistical information for higher resolutions and nodes that appear on higher levels correspond to more global characteristics of an image.

Tree Spatial Superposition Logic

In, a formal logic, called tree spatial superposition logic (TSSL), was introduced. TSSL is capable of formally specifying global patterns in a network of locally interacting agents. The authors showed that this logic is sophisticated enough to describe complicated patterns such as Turing patterns in biochemical reaction-diffusion systems. In this paper, we used this logic to express various patterns that are studied here (Figure 4). First, we present a brief introduction to TSSL. The reader can refer to Bartocci et al. (2016) for a thorough explanation of this logic, definitions of syntax and semantics, and its properties.

A TSSL formula is recursively constructed using the following:

- Linear predicates over valuations for the representation function (Equation S16). For example: $\mu_1^{(r)} > 0.8$ or $\mu_1^{(b)} < 0.5$.
- Boolean operators, such as $\neg\varphi$, $\varphi_1 \wedge \varphi_2$, and $\varphi_1 \vee \varphi_2$.
- Spatial operators: $\exists_B \circ \varphi$, $\forall_B \circ \varphi$, where B is a nonempty subset of the set of directions $\{NW, NE, SW, SE\}$.

The spatial operators $\exists_B \circ$ and $\forall_B \circ$ are read as *there exists in directions B next* and *for all directions B next*, respectively. $\exists_B \circ \varphi$ is interpreted as follows: for at least one of the nodes located in the next level of the quad-tree labeled with one of the directions in B , φ must be satisfied. $\forall_B \circ \varphi$ specifies that for all such nodes φ must be satisfied. We demonstrate how TSSL can be used to express spatial patterns through an example.

Consider a 4×4 pattern as illustrated in Figure 3A. This pattern can be expressed as the following TSSL formula ϕ . A portion of the quad-tree satisfying this formula is shown in Figure 3B.

$$\begin{aligned} \phi = & \\ & \forall_{\{NE, SW, SE\}} \circ [\forall_{\{NW, SE\}} \circ \phi_{white} \wedge \forall_{\{NE, SW\}} \circ \phi_{colored}] \wedge \\ & \forall_{NW} \circ [\forall_{NW} \circ \phi_{white} \wedge \forall_{\{NE, SW\}} \circ \phi_{blue} \wedge \\ & \forall_{SE} \circ (\forall_{\{NW, SE\}} \circ \phi_{white} \wedge \forall_{\{NE, SW\}} \circ \phi_{blue})], \end{aligned} \quad (\text{Eq 17})$$

where

$$\begin{aligned} \phi_{white} &= \mu_1^{(r)} = 1 \wedge \mu_1^{(g)} = 1 \wedge \mu_1^{(b)} = 1, \\ \phi_{colored} &= \mu_1^{(r)} < 1 \vee \mu_1^{(g)} < 1 \vee \mu_1^{(b)} < 1, \\ \phi_{blue} &= \mu_1^{(r)} = 0 \wedge \mu_1^{(g)} = 0 \wedge \mu_1^{(b)} \geq 0.5. \end{aligned}$$

TSSL formulas can be viewed as formal pattern descriptors or pattern classifiers. For instance, the formula of Equation S178 accepts a quad-tree derived from a checkerboard pattern and rejects any other quad-tree. Although TSSL is capable of describing complicated spatial behaviors in an image, it is difficult in general to write a formula that describes a complex pattern. In Bartocci et al. (2016), the authors proposed to use machine learning techniques in order to find such a formula from a given set of positive and negative examples.

Assume a set of positive images (\mathcal{Y}_+), illustrating a desirable pattern, and a set of negative images (\mathcal{Y}_-), in which the desirable pattern was not present, were available. We created a set \mathcal{L} from these images as:

$$\mathcal{L} = \{ (Q_y, +) \mid y \in \mathcal{Y}_+ \} \cup \{ (Q_y, -) \mid y \in \mathcal{Y}_- \},$$

where Q_y was the quad-tree generated from image y .

The set \mathcal{L} was separated into a learning set \mathcal{L}_L (used to train a classifier) and a testing set \mathcal{L}_T (used to test the classifier obtained from \mathcal{L}_L) such that $\mathcal{L} = \mathcal{L}_L \cup \mathcal{L}_T$. A rules-based learner called RIPPER (Cohen, 1995) was used to learn a set of classification rules from \mathcal{L}_L . Each of these rules was in the form:

$$R_i : C_i \Rightarrow \text{Label}_i,$$

where C_i was a Boolean formula over linear predicates over the representation values of the nodes of a quad-tree and $\text{Label}_i \in \{\pm\}$.

We used the Weka workbench for implementing RIPPER. Each C_i was then translated into an equivalent TSSL formula Φ_i . Since the classification rules were interpreted as nested if-else statements, the TSSL formula equivalent to the entire set of classification rules corresponding to the positive class was written as:

$$\Phi_+ = \bigvee_{j \in R_+} (\Phi_j \wedge \bigwedge_{i=1, \dots, j-1} \neg \Phi_i), \quad (\text{Eq 18})$$

where R_+ was the set of indices of rules labeled positive.

Quantitative Robustness

A TSSL formula can be created for any desired spatial pattern by following the procedure described in the previous section. If this formula is evaluated as true for a given image, it means that the image contains the required pattern. On the other hand, a false evaluation of the formula means that the pattern does not exist. However, this qualitative evaluation of TSSL descriptors does not provide any information about how strongly an image demonstrates the required pattern.

In order to provide information about how strongly an image satisfies or violates the given property, TSSL was also equipped with a recursive quantitative semantics definition which assigned a real value to a TSSL formula φ with respect to vertex $v \in \mathcal{V}$ of quad-tree $Q = (\mathcal{V}, R)$; denoted by $\rho(\varphi, v)$. The TSSL quantitative valuation was derived recursively as follows:

- $\rho(\mu_i^{(c)} \geq d, v) = \mu_i^{(c)}(v) - d$
- $\rho(\mu_i^{(c)} \leq d, v) = d - \mu_i^{(c)}(v)$
- $\rho(\neg\varphi, v) = -\rho(\varphi, v)$
- $\rho(\varphi_1 \wedge \varphi_2, v) = \min(\rho(\varphi_1, v), \rho(\varphi_2, v))$
- $\rho(\varphi_1 \vee \varphi_2, v) = \max(\rho(\varphi_1, v), \rho(\varphi_2, v))$
- $\rho(\exists_B \circ \varphi, v) = 0.25 \max_{b \in B}(\rho(\varphi, v_b))$ where v_b was the child vertex of v with label b
- $\rho(\forall_B \circ \varphi, v) = 0.25 \min_{b \in B}(\rho(\varphi, v_b))$ where v_b was the child vertex of v with label b

It was proven in that TSSL quantitative semantics are sound. In other words, a quad-tree Q satisfied a formula φ ($Q \models \varphi$) if $\rho(\varphi, v_0) > 0$ where v_0 was the root of Q , and Q violated φ ($Q \not\models \varphi$) if $\rho(\varphi, v_0) < 0$. Therefore, the problem of checking whether an image contains a pattern expressed as a TSSL formula was reduced to computing its quantitative valuation $\rho(\varphi, v_0)$. Moreover, the absolute value of $\rho(\varphi, v_0)$ was viewed as a measure of how strongly φ was satisfied (or violated) by Q . Hence, the quantitative valuation of a formula with respect to a quad-tree was called its *robustness*. This property is demonstrated in Figure 3.

Particle Swarm Optimization

Consider an agent-based model with a set of parameters $p \in \Omega \subset \mathbb{R}^{N_p}$, where Ω was the possible set of parameter ranges and N_p was the total number of parameters. For instance, in the model described in Section S1, we had $N_p = 5$ parameters with ranges specified in Table S1.

The output of the model was a sequence of T images where $A[t]$ was the image corresponding to time step $t \in \{0, 1, \dots, T\}$ and T was the total duration of simulation. Our goal was to determine parameter values that result in emergence of a required pattern in the sequence of images derived from the model. Recall that we could specify the pattern using a TSSL formula $\Phi_{Pattern}$. Moreover, each image $A[t]$ could be translated into a corresponding quad-tree $Q[t]$ with root $v_0[t]$. Therefore, for a fixed parameterization p , we could quantify the resulting sequence of images with $S(p)$ using the following equation:

$$S(p) = \max_{0 \leq t \leq T} \rho(\Phi_{Pattern}, v_0[t]), \quad (\text{Eq 19})$$

where ρ was the TSSL robustness as described in the previous section.

Note that since the model was stochastic in nature, $S(p)$ was a random variable and would have a different value every time a sample simulation was produced using the model with the parameters p . If $S(p) > 0$, there exists at least one image in that particular sequence for which the TSSL robustness was positive and the pattern was present. On the other hand, the pattern had not emerged in the sample simulation if $S(p) < 0$. We called $S(p)$ the *robustness* degree for parametrization p . Now, the problem became finding the parameterization p^* that maximized the score $S(p)$. Since $S(p)$ was a random variable, we choose to maximize its expected value:

$$p^* = \operatorname{argmax}_{p \in \Omega} \mathbb{E}(S(p)), \quad (\text{Eq 20})$$

which means that we were looking for the parameterization p^* that *on average* produced patterns with highest possible robustness score.

If we simulated the model n times from parameters p , the expected value could be approximated with the sample mean:

$$\mathbb{E}(S(p)) \approx \tilde{S}(p) = \frac{1}{n} \sum_{i=1}^n S_i(p), \quad (\text{Eq 21})$$

where $S_i(p)$ was the robustness score for parameters p in the i th simulation.

In general, a large sample is needed to achieve an accurate approximation. however, it was shown in that in practice, a relatively small n suffices for the purpose of optimization in Equation S20. In this paper, we computed the average robustness for three sample simulations in every case ($n = 3$).

Many optimization methods can be used to solve this optimization problem. Inspired by Bartocci et al. (2016), we employed particle swarm optimization (PSO) (Kennedy, 2011) to solve this problem. PSO is a heuristic solution to unconstrained optimization problems that is capable of solving problems with irregular search spaces, is easily distributable, and does not require the objective function to be differentiable.

The PSO algorithm worked as follows: the procedure began by randomly initializing a set of M particles with positions $z_i \in \Omega$ and velocities \dot{z}_i . The position of a particle was a candidate solution to Equation S20, and the velocity was a search direction from the current solution. Next, n simulations were produced and n sequences of quad-trees $Q[t](z_i)$ were created for each particle and the average robustness degree $\tilde{S}(z_i)$ was evaluated for each set of simulations represented by particle z_i . The position of the i th

particle that had performed best so far was stored in the variable z_i^* , and the optimal value of z_i^* was denoted by z^* . After all particles had been evaluated, the positions and velocities were updated according to the following relations:

$$\begin{aligned} z_i' &\leftarrow Wz_i + \eta(r_p)(z_i^* - z_i) + \eta(r_g)(z^* - z_i) \\ z_i &\leftarrow z_i + z_i', \end{aligned} \quad (\text{Eq 22})$$

where $\eta(r_i)$ was a random number uniformly distributed over $[0, r_i]$ and the parameters $W \in \mathbb{R}$, r_p , r_g are tuned by the user. This iterative process continued until a termination criterion was met.

If $\hat{S}(p^*)$ was positive or negative but sufficiently close to zero, we had found the optimal parameterization of the model for the required pattern. This occurred for the bullseye and Multi-Islands patterns. The optimal parameterization is shown in Figure 4. On the other hand, $S(p^*) \ll 0$ indicated that even for the best possible parameterization of the model, the required pattern did not emerge, meaning that the model was not capable of producing that pattern at all. This occurred for the Janus (left-right) pattern (Figure 4C iii).

Figure S4A demonstrates two sample simulations, one for the bullseye pattern and one for the Multi-Islands pattern. Figure S4B shows how the corresponding TSSL scores evolve over time for each simulation. It is seen in this figure that the scores gradually improve until at some points the desired patterns are formed.

QUANTIFICATION AND STATISTICAL ANALYSIS

Mann-Whitney U-tests were used to compare *in vitro* and *in silico* experimental populations in Figure 2. Unpaired T-tests with Welch's correction were used to compare *in vitro* and *in silico* experimental populations in Figure 4. Error bars depicted in graphical representations signify 1 standard deviation unless otherwise specified in the figure legend. For each statistical test the number of replicates is described in the figure legend. Throughout the manuscript the symbol * signifies statistical significance at the 0.05 level unless otherwise specified.

DATA AND CODE AVAILABILITY

Software

Model fitting of single-cell morphology, cell velocities, temporal knockdown characterizations, and spatial pattern characterizations were performed with custom Python code generated for these studies (modules: scipy, numpy, matplotlib, pandas, seaborn, scikit-image) (Jones et al., 2001). Image preprocessing, segmentation, and quantification of cell and colony morphology was performed with custom Python code (modules: numpy, scipy, scikit-image). Quantification of pattern similarity and pattern optimization were performed with custom code for TSSL. All custom code can be accessed From: <https://github.com/dmarcbriers/Multicellular-Pattern-Synthesis>.

Data Availability

All data is available in the main text or supplementary materials. A GitHub repository of analysis code can be found at: <https://github.com/dmarcbriers/Multicellular-Pattern-Synthesis>.