

Title: Incomplete cell sorting creates engineerable structures with long term stability

Authors: Jesse Tordoff^{1*}, Matej Krajnc², Nicholas Walczak³, Matthew Lima⁴, Jacob Beal³, Stanislav Shvartsman^{2,5,6}, Ron Weiss^{4†}

Author Affiliations:

- (1) Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
- (2) The Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
- (3) Raytheon BBN Technologies, Waltham, MA, 02451, USA
- (4) Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
- (5) Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA
- (6) Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

* Corresponding author and lead contact address: jessetordoff@gmail.com

† Corresponding author address: rweiss@mit.edu

Summary

Adhesion-mediated cell sorting has long been considered an organizing principle in developmental biology. While most computational models have emphasized the dynamics of segregation to fully sorted structures, cell sorting can also generate a plethora of transient, incompletely sorted states. The time scale of such states in experimental systems is unclear: if they are long lived, they can be harnessed by development or engineered in synthetic tissues. Here we use experiments and computational modeling to demonstrate how such structures can be systematically designed by quantitative control of cell composition. By varying the number of highly adhesive and less adhesive cells in multicellular aggregates, we find the cell type ratio and total cell count control pattern formation, with resulting structures maintained for several days. Our work takes a step towards mapping the design space of self-assembling structures in development, and provides guidance to the emerging field of shape engineering with synthetic biology.

Introduction

In 1907, H. V. Wilson reported a remarkable phenomenon: when sponges were pushed through a fine cloth and completely dissociated, they could re-assemble themselves back into complete living organisms¹. Steinberg proposed that a simple physical mechanism - differences in cell-cell adhesion - could account for this behavior². Since then, there has been a large body of work that addresses the molecular mechanisms of self-assembling biological structures^{3,4}, culminating in a recent drive to direct self-organizing shape formation synthetically with the goal of engineering artificial living structures for materials engineering and organ regeneration^{5,6,7,8,9}.

Although self-organization has been studied extensively, the general question of what can be constructed through multicellular self-organization remains unsolved. Here we present a systematic study of a range of structures that can be constructed using a ubiquitous tool for self-organization in animal development: cell sorting. Cell sorting, or the process by which mixtures of cell types can physically rearrange themselves into distinct populations, is driven by differences in tissue surface tension between populations of cells. Tissue surface tension is determined primarily by an interplay between cell-cell adhesion and cell cortex tension^{10,11}. Complete cell sorting can result in a range of structure types, including an engulfed cluster of one cell type surrounded by another cell type, or total separation of two cell types to distinct sections. Partial or incomplete cell sorting was discussed in the seminal work by Glazier and Graner¹² in which they developed a Cellular Potts Model for simulating differential adhesion. They demonstrated computationally that when partially sorted clusters have limited diffusion, they can be stable over long periods without significant merging. Most other theoretical work on cell-sorting-based structure formation has been primarily focused on studying segregation dynamics and the resulting fully segregated steady-state patterns^{13,14,15,16,17}, with some recent exceptions that examined the periodic Turing-like patterns that can emerge as a result of cell adhesion combined with phenotypic switching¹⁸.

Although many computational models and experimental systems exist for cell sorting-based pattern formation^{3,11,12,19,20}, they lack a precise set of design rules for controlling the features of the resulting patterns and structures, especially of incompletely sorted forms. Here we address this issue and establish several design rules by systematic exploration of two features of the design space: total number of cells and relative composition of highly adhesive and less adhesive cells. We find a set of self-organizing structures that are distinct from fully sorted steady-state structures, and are maintained over the course of multiple days. In aggregates with highly adhesive and less adhesive cells mixtures, we find that the ratio of cell types precisely controls a transition between a completely sorted structure and the formation of many semi-regular

clusters that are robustly tunable. Due to the finite time scale of cell rearrangement, we argue that these states may be more relevant for tissue design than the thermodynamically stable fully-sorted states. In our system, we predicted that steady state sorted behavior would emerge in some aggregates on a time scale that is typically longer than stages of development in a developing organism. Although incompletely sorted, patterns with multiple clusters persist over the course of many days and have predictable, controllable features. These structures are designable and stable over biologically relevant time scales, paving the way for synthetic multicellular shape engineering.

Results

Ratio of highly adhesive to less adhesive cells precisely controls pattern features

Given enough time to sort out, a mixture of motile highly self-adhesive and less adhesive cells is theoretically expected to assemble into the lowest energy arrangement: a single central sphere of the more adhesive cells engulfed by a shell of less adhesive cells^{2,11}. To establish whether other biologically relevant patterns can be created using cell sorting between highly adhesive cells and less adhesive cells, we created aggregates composed of two cell types with different adhesive properties: HEK293FT (HEK) cells and CHO K1 (CHO) cells. CHO cells express low levels of cadherin²¹, are minimally self-adhesive, and form loose aggregates in 3D cell culture. HEK293 cells express E-cadherin and N-cadherin and are more self-adhesive, forming compacted spheroids in 3D cell culture²². To make CHO and HEK 3D aggregates, cells were first maintained in adherent 2D cell culture, trypsinized to make single-cell suspensions, and added in precise numbers to low-adherent U-bottom plates using fluorescence activated cell sorting (FACS) (**Figure 1a**). After seeding, plates were mixed with pipetting and centrifuged at a low speed to encourage cells to fall to the center of the U-bottom. To distinguish cell types with microscopy, CHO and HEK cells were engineered with genetically integrated expression of fluorescent proteins tagBFP and eYFP, respectively, both driven by a constitutive human EF1a promoter.

Using this method to create co-cultured aggregates, we systematically varied the ratio of HEK and CHO cells and imaged pattern formation after an initial sorting period of approximately 24 hours. We analyzed ten technical replicates of nine different HEK to CHO ratios, seeded at 7000 cells in total. Pattern formation was found to be consistent across replicates and highly dependent on HEK to CHO ratio, showing different modes of behavior at ratios around a critical 50% threshold (**Figure 1b**). At high HEK to CHO ratios, aggregates self-organized into the predicted form, with a compact HEK spheroid engulfed by non-adherent diffuse CHO cells (**Figure 1b, rows 6-9**). With 40% or fewer HEK cells, aggregates assembled into a different pattern: instead of coalescing into a

central sphere, the HEK cells grouped into many distinct clusters, separated by CHO cells (**Figure 1b, rows 1-4**). Aggregates with intermediate numbers of HEK and CHO cells (**Figure 1b, row 5**) had more variable patterning, with some replicates forming a complete engulfed sphere and others with several merging central clusters surrounded by multiple smaller clusters. **Supplemental Videos 1-6** show 3D Z-stacks of representative aggregates. Although clusters have a Z height greater than one cell, the aggregates are flattened, and are wider across than in height. Further, clusters appear to contact the plate and do not lie on top of each other. The central Z-slice captures almost all clusters, and we therefore use only this slice and not the whole Z-stack in further analysis.

To quantify the aggregate's behavior, we analyzed the central Z-slice microscopy images with a custom image analysis pipeline that measures cluster number and area. **Figure 1c** and **Figure S1** shows the sizes of each cluster in the images from **Figure 1b** combined across replicates and grouped by the HEK:CHO ratio of the aggregate. Cluster size abruptly transitions at 40-50% HEK from exclusively small (less than 0.03 mm²) to a mixture of large (0.03 mm² to 0.12 mm²) and small clusters. The images show that the few small clusters in the >50% HEK wells primarily exist either on the border of the aggregates or apart from the aggregate itself. The plates are coated with a cell-resistant coating, and therefore all cells should not be able to stick to the plate and should fall to the bottom of the U-shaped well due to gravity. The individual clusters that are not in contact with the rest of the aggregate are therefore likely due to small imperfections in the cell-resistance coating on the U-bottom plate, and can fall down to the outside of the aggregate over the course of hours. To exclude these flaws, we filtered out any cluster that was smaller than 5% of the largest cluster in the aggregate. It would be ideal to set an absolute size filter to remove small clusters created from flaws in the plate, but the sizes of these external clusters are comparable to the small clusters formed through sorting in the 10% and 20% HEK aggregates. **Figure S1** shows the same data without this percent filtering.

Figure 1d shows the number of clusters in each aggregate, grouped by percent HEK composition. The number of clusters in an aggregate of a specific HEK:CHO ratio was tightly distributed, especially between replicates with more than 50% HEK cells when a single central cluster forms (**Figure 1d**).

Although incompletely sorted, these aggregates assemble into structures with reliable and tunable features. The size and number of clusters are consistent and the type of patterning - whether a single central cluster or many distinct clusters - is dependent on a precise ratio of highly adhesive to less adhesive cell populations.

A particle-based model predicts scaling behavior

To explore how aggregate patterns can be further systematically designed, we employed a two-dimensional particle-based model of cell sorting, based on a generalized Vicsek model of interacting self-propelled particles^{12,23}. Although our experimental system uses 3D cell culture in which cells are not constricted to a monolayer, we chose to create a 2D model because we found that experimentally the aggregates are flattened and clusters exist in the same plane even though cells themselves can pile on top of each other (**Supplementary videos 1-6**), and because it is significantly more efficient to simulate than a 3D model.

The model describes collective behavior of cells within the aggregate assuming simple rules for local interactions. In particular, cells are assumed to move with constant magnitude of velocity, v_0 , in a direction given by a polar angle θ_i (**Figure 2a**), which is aligned at all times with the net force acting on the cell, $F_i = \sum_{j \neq i} f_{ij} + f_i^{(ext)}$ (Refs. 24,25). Here the pairwise intercellular force f_{ij} describes the excluded-volume repulsion at cell-cell distances shorter than the cell diameter σ , mid-range attraction due to cell-cell adhesion, and random forces at the cell-cortex level (**Figure 2a**, full equation described in methods). The adhesion term is associated with three parameters: U_{11} , U_{22} , and U_{12} , which represent cell-type-dependent adhesion strengths between cells of the same type (11 and 22) and cells of different types (12). The external force ($f_i^{(ext)}$) describes the effect of gravity (g) and the curved bottom of the underlying surface, preventing cells from escaping the aggregate (**Figure 2a**). In our model, cells move by 0.05σ during a unit time, set by cell size and velocity as σ/v_0 .

To systematically search for incompletely sorted configurations, we first considered a two-component cell aggregate of 1000 cells with an initially well-mixed mixture of 20% type-1 cells and 80% type-2 cells, in which the adhesion between type-1 cells was assumed the strongest ($U_{11} = 30mg$ and $U_{12}, U_{22} < U_{11}$). We explored configurations at time $t = 10^7\sigma/v_0$ after the beginning of sorting, which roughly corresponds to the time scale at which the configurations at low values of U_{12} and U_{22} fully segregate. By measuring the final number of type-1-cell clusters, we were able to identify regions in the parameter space (U_{12}, U_{22}) with incomplete sorting, displaying patterns with multiple clusters (**Figure 2b-f**).

We next check whether the model captures the transition from the single- to many-cluster patterns at a critical composition ratio, as found experimentally (**Figure 1c**). To do so we performed a parameter scan, varying the ratio at fixed values of adhesion strengths to find at least one simulation parameter combination that produces a phase transition at a similar ratio to the observed data. We measured the final number of clusters and their mean size and found that for $U_{11}=30$, $U_{12}=1$, $U_{22}=5$ the

simulation indeed approximated the behavior observed experimentally, with the critical ratio being at about 40% (**Figure 2g**).

Next, we used the model to predict how the patterns change upon varying both the composition ratio and the total cell number at seeding. We found that a given composition ratio had a characteristic mean cluster size that was approximately constant in any aggregate above a certain number of cells (**Figure 3a**). The characteristic cluster size varied according to the ratio, with high-fraction type-1 cell aggregates sorting into larger clusters than low-fraction type-1 cell aggregates. The inflection point from linear to constant rising as the density of strongly adherent cells rises. This behavior is expected, based on the theoretical results of prior studies^{12,13}. For 70% density and higher, no inflection point is observed, and at some point above this level we may expect a phase transition to a percolating state in which the inflection point rises to infinity and thus cluster size will always rise linearly. The number of clusters followed a similarly consistent rule, remaining at one below a density-dependent cell-count threshold, and approximately linearly increased according to number of cells above that threshold (**Figure 3B**). For 30% density and lower, obtaining a single cluster becomes unreliable and at some point below this level, we may expect that the threshold drops to a level where it cannot be expected to be obtained with more than one strongly adherent cell.

To investigate this prediction experimentally, we repeated the ratio sweep while seeding at varying total cell numbers (**Figure 3C,D**). As predicted by the model, the average cluster size at a given ratio is approximately constant after an inflection point that rises with the density of HEK cells, with no inflection point in range for 70% and higher. Likewise, for each ratio, there is a linear relationship between the number of clusters and total number of cells beyond a certain threshold. In effect, we see that each HEK ratio has its own characteristic maximum cluster size. If there are too few cells to achieve this cluster size, then one central cluster forms (though not reliably for 40% HEK and below). If there are more than enough cells to achieve this size, then multiple clusters can form, and the number of clusters scales with the total aggregate size. Correspondingly, at cell ratios that transition from one cluster to multiple clusters (40, 50, and 60% HEK aggregates), the aggregate size at which multiple clusters form corresponds to the aggregate size at which average cluster size levels out to a consistent value (**Figure 3D**). These relations can be to a large extent explained using dimensional analysis (**Note S1**), which further confirms the robustness of pattern manipulation via easily controllable parameters. It is important to note that the trends for the mean cluster size between the model and experiment do not match exactly. In the model, we see a decrease in the mean cluster size with an increasing cell number, whereas in experiments cluster sizes appear to stabilize. It is hard to say which of the simplifications of physical mechanisms considered by our model contribute to this

discrepancy. The decrease here is much weaker than the inverse of the number of cell, $1/N$, so in this respect the relation is closer to being independent on the number of cells N than it is to having a significant dependency on N . This is in close agreement with the predicted scaling relations obtained by simple geometric scaling arguments (**Note S1**; note that geometric scaling provides only an extremely abstract model, which thus serves as a check on results but cannot explain other phenomena). The geometry itself is therefore likely the main determinant of how cluster size and number of clusters scale with total cell count. This suggests that we should see similar overall trends for the two plots regardless the underlying source of fluctuations that give rise to fluid-like behavior of tissues (e.g., active cell motility, proliferation and death, or tension fluctuations in the cell cortex).

Incomplete sorting between highly adhesive cells and less adhesive cells creates long-term stable clusters

We have established that incompletely sorted structures can have precise engineerable features controllable by cell ratio and total number of cells. To determine whether these structures are relevant for shape engineering and biological development, we next sought to determine their persistence after the initial sorting period. At steady state, it is expected that clusters of highly adhesive, moving cells will merge with each other to form one single cluster. This prediction comes from the fluid nature of tissues: by random cell movement, clusters should eventually meet and merge.²⁶

To predict the timescale of cluster merging across different cell ratios, we simulated cluster formation with the particle-based model and tracked the number and size of clusters over time. This data shows there is an initial sorting period after which the number of clusters remains relatively stable, with slow decreases in cluster numbers thereafter (**Figure 4a**). We found that while the aggregates with large percentages of strongly adhering cells (>60%) reach completely sorted states on experimental time scales, aggregates with small percentages (<60%) undergo consistent but slow sorting. Extrapolating the observed kinetics predicts time scale of complete sorting on the order of years (**Figure S4**).

To evaluate this time scale of merging in our experimental system, we tracked structure formation over the course of many days. Surprisingly, most of the multi-cluster aggregates were stable after the initial sorting period and retained a consistent number of clusters for several days, at which point CHO cells began to lose fluorescence (**Figure S2a**). Structure stability was found to vary based on cell type ratio. For higher HEK to CHO ratios, the number of clusters at the end of the experiment, 66 hours, was similar to the number of clusters after an initial sorting period of 10 hours (**Figure S2b**). The number of clusters in aggregates with lower HEK percentages (10-30%) declined at

a steeper and steadier rate, especially as CHO cells began to lose fluorescence at 45 hours. We observed that after 35 hours, aggregates of the lower HEK ratios (10-30%) were large enough that clusters expanded out of the region of view. **Figure 4b** shows clusters up to this time point, and the full 65 hour time course is shown in **Figure S2b**). As expected, the experimental and simulation results show very similar dynamics, though the exact transition point and number of clusters varies for intermediate ratios (40%-60%) because we are not adjusting the simulation parameters for a post-facto fit.

One potential explanation of stability could be that clusters, although motile, simply diffuse too slowly to merge in a reasonable time frame. Glazier and Graner¹² simulated some partially sorted aggregates that were stable over long term, and found that strong surface energy between two cell types can block the diffusion of small clusters, which prevented cluster merging over a long timescale after an initial sorting period. To evaluate the speed of cluster diffusion in our system, we tracked the position and velocity of each cluster in incompletely sorted aggregates over time and measured their mean squared displacement (MSD). **Figure S3** shows the average of these displacement values over time for each cluster across five replicate aggregates. This analysis tracked only clusters that did not merge with another cluster over the examined time period, with the reasoning that this would simplify velocity calculations and the minority of clusters that did merge in a specific time period would not significantly change the average velocity overall. The timescale for cluster merging was estimated by first measuring the average smallest distance to the nearest cluster (**Figure 4c**), and the effective diffusion coefficient, D_{eff} , (**Figure 4d**, from the slopes of **Figure S3**). The ratio of D_{min}^2 to D_{eff} (**Figure 4e**) represents the expected time scale on which a cluster could cover the distance to its nearest neighbor and come into contact and merge, based on their diffusion speed. Aggregates with 10%, 20%, and 30% HEK cells had an expected merging time of 5-10 hours, and aggregates with 40% and 50% were higher, between 15 hours and 40. This is consistent with the cluster tracking in **Figure 4B** and **Figure S2b**, which shows slow decreases in the number of clusters in 10%, 20%, and 30% HEK aggregates and approximately constant numbers of clusters of 40% and 50% HEK aggregates over the time course. The stability of the incompletely sorted aggregates could therefore be explained by the slow speed of cluster diffusion relative to the distance between clusters.

Together, these results demonstrate that on the time scale of days, the sorting dynamics that follow the initial sorting period are slow enough to not be observable in many cases. This property gives rise to structures that are stable over the course of biologically relevant time scales, and can therefore be relevant for establishing structure during development and in synthetic shape engineering.

Discussion

We have shown that incomplete cell sorting can result in an array of different, engineerable structures with dynamic properties that are relevant on experimental and developmental time scales. We demonstrated both experimentally and computationally that the features of these structures can be precisely manipulated by tuning the ratio and total number of cells in the aggregate, and remain mostly stable for several days. This represents a step towards mapping the design space of stable self-assembling structures in developmental biology and in synthetic living structures. Although the stability of the structures on longer time scales is not certain, their persistence over the course of days suggests that other mechanisms could be used during such times to solidify transient patterned states so that they would become trapped and could no longer change their architecture. In particular, one could at the right time and amplitude increase cell-cell adhesion strength for both cell populations, decrease cell motility, or increase cortical contractility after the initial pattern formation. Predictable design might then be supported by extraction of an inverse model from results such as we have presented here.

Further work is needed to access the full space of structures and features that can be made with cell sorting-based self-organization. Local spatial control of cell sorting, for example using optogenetics, would allow for more controllable heterogeneity in structures with specifically targeted regions of cell-cell adhesion. Additionally, our results suggest that controlling other features in addition to cell-cell adhesion, including tensile forces and surface tension, which influence sorting *in vivo*²⁷, would likely also allow for control over sorted structures. Finer temporal control of cell sorting might allow for multi-step assembly where one feature builds on another, for example by establishing multiple adhesive layers of cells, potentially using synthetic adhesion toolboxes like those developed in the recent work of Glass and Riedel-Kruse⁹. Ultimately, with multiple cell types and spatiotemporally controlled self-assembly, it may be possible to recreate and control the robustness and complexity of cell-sorting-influenced biological patterns, like the patterning of zebrafish stripes²⁸ or generation of regular hair follicles²⁹. Controlled pattern formation is then likely to be useful in a range of tissue engineering applications, such as construction of organoids or culturing of patient-derived grafts, in which the ratio of cell types being combined can be controlled.

As we demonstrated, the protocols for the design of synthetic tissues can successfully be informed by computational simulations, which often provide simple and robust rules for the control of relevant parameters. Moreover, the coherence between simulation and experiments and the conclusions of prior theoretical work^{12,14,16,17} both suggest that these results are likely to be transferrable to any other mixture of strains with highly differentiated adhesion levels, given characterization of the relevant adhesion parameters. It is therefore expected that these protocols will continue to be complemented by computational models, which will need to be improved by including

more realistic and detailed descriptions of cell-cell interactions. This could be done by employing models that describe tissues at the level of individual-cell shapes, e.g., the vertex model^{30,31}. In particular, using such an approach would address the role of cell-scale activity that could give rise to a variety of cell behaviors, e.g., cell motility, cell proliferation and death, and active cell-shape changes. These phenomena are known to influence overall tissue fluidity^{32,33,34,35} and could therefore be used as additional mechanisms contributing to the guided design of shapes.

Experimental Procedures

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jesse Tordoff (jesselordoff@gmail.com).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The software generated during this study is available at our github repository, <https://github.com/TASBE/TASBEImageAnalytics>.

Cell Lines and Culturing

CHO (sex: female) and HEK (sex: female) cells referred to in this work are CHO K1 cells genetically integrated with expression of eYFP and a Bxb1 integration site at the AAVS1 locus and HEK293FT genetically integrated with expression of eYFP and a Bxb1 integration site at the Rosa B1 locus, obtained from Duportet et al³⁶. Cells were cultured at 37°C and 5% CO₂ in Dulbecco's Modified Eagle Media supplemented with 10% fetal bovine serum and 1% non-essential amino acids. To make them fluoresce blue, Bxb1 CHO cells were co-transfected with 300ng of a pHeF1a-tagBFP plasmid with a Bxb1 recognition site and puromycin resistance gene and 300ng of Bxb1 expression vector. Puromycin was added 3 days after transfection and puro selection continued for approximately two weeks, during which media was changed and cells were passaged as needed. The HEK cells obtained from Duportet et al³⁶ expressed eYFP, and were used without further modification.

Self-propelled particle model

Our model is a two dimensional model based on the Vicsek model of self-propelled disk-like particles²³. Within this description, the collective cellular behavior is described by simple rules of local interaction between neighboring cells. In particular, cells are assumed to move with velocity $v_i = v_0(\cos \theta_i, \sin \theta_i)$, where v_0 is the constant magnitude of velocity, and θ_i is the polar angle defining the direction of motion. Following previously proposed extensions of the original Vicsek model, we assume that the direction of motion of cell i , θ_i , instantaneously aligns with the net force F_i acting on the cell^{12,13,24} such that $\theta_i(t) = \arg [F_i(t)]$. Positions of cells r_i are then updated as

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t,$$

where with no loss of generality, we set $\Delta t = 1$. Per time step, cells are allowed to move only by a small fraction of their size, which constrains the magnitude of velocity v_0 ; here $v_0 = 0.05$.

Cells, modeled as disks with unit diameter and experience the overall force

$$F_i = \sum_{j \neq i} f_{ij} + f_i^{(ext)},$$

where f_{ij} is the pairwise intercellular force between cells i and j , whereas $f_i^{(ext)}$ is the external force that captures combined effects of the gravity and curvature of the underlying surface. The force f_{ij} depends on the intercellular pairwise distance $|r_{ij}|$, where $r_{ij} = r_j - r_i$, whereas $f_i^{(ext)}$ depends only on the position of the cell in question (i). All forces are measured in units of mg , where m and g are the cell mass and the gravitational acceleration, respectively.

At cell-cell distances shorter than the cell diameter ($|r_{ij}| < 1$), f_{ij} describes short-range hard-core repulsion due to excluded volume. In particular, $f_{ij} = -U_{hc}r_{ij}/|r_{ij}|$, where $U_{hc} = 10^4$. Next, mid-range attraction due to cell-cell adhesion is described by cell-type-dependent adhesion strength $U_{ij} \in \{U_{11}, U_{22}, U_{12}\}$, such that within the adhesion range ($1 < |r_{ij}| < r_0$),

$$f_{ij} = U_{ij}(|r_{ij}| - r_e) \frac{r_{ij}}{|r_{ij}|} + \alpha \hat{\xi}_{ij},$$

where $r_0 = 1.6$ and $r_e = 1.01$ are the adhesion range and the equilibrium intercellular distance, respectively. The random pairwise force $\alpha \hat{\xi}_{ij}$ acts due to noise at the cell cortex level; here $\hat{\xi}_{ij}$ is a random unit vector and α is the magnitude of the noise.

Pairwise forces obey Newton's III law of motion and therefore $\hat{\xi}_{ij} = -\hat{\xi}_{ji}$.

A static two-dimensional force field $f_i^{(ext)}$ drives cells towards the center of the simulation domain, mimicking the experimentally observed persistent cellular movements toward the bottom of the plate. In particular, $f_i^{(ext)} = -\hat{r}_i h' / [1 + (h')^2]$, where $h(r)$ is the height profile of the plate, $\hat{r}_i = r_i/|r_i|$ and $h' = dh/dr$. In our model, the profile is approximated by a hemispherical well with the radius $R = 325$ such that

$f_i^{(ext)} = -(r_i/R)\sqrt{1 - (|r_i|/R)^2}$. The model was implemented in C++. Typical simulations ran for about half a day in real time.

| Parameter | Value | Units |
|--|--------|-------------------|
| instantaneous cell velocity v_0 | 0.05 | $\sigma/\Delta t$ |
| strength of hard-core repulsion U_{hc} | 10^4 | mg |
| adhesion range r_0 | 1.6 | σ |
| equilibrium cell-cell distance | 1.01 | σ |
| radius of the plate | 325 | σ |
| magnitude of the noise | 10 | mg |

Table 1: Values of model (fixed) parameters used in simulations. These values were determined by varying them manually so as to minimize the mismatch with the experimental results.

Structure formation assays

To make 3D aggregates, adherent cells growing in tissue-culture-treated 10cm dishes were trypsinized with 2.5 mL trypsin for 5 minutes. Cells were sorted into ultra-low attachment 96-well plates with a BD FACSaria II flow cytometer at the Swanson Biotechnology Center Flow Cytometry Core Facility at MIT. Cells were gated on for positive eYFP or tagBFP expression. Immediately after FACS sorting, cells were mixed with gentle pipetting and spun with a centrifuge at 300rcf for 4 minutes. All cells were grown and assayed in DMEM supplemented with 10% FBS and 1% non-essential amino acids, with 200uL media per well.

Microscopy and Image Analysis

Aggregates were imaged with a Leica TCS SP5 II Confocal Laser Scanning Microscope in an incubation chamber at 37°C at 5% CO₂. Each image represents one Z slice through the approximate center of the aggregate, as measured by where the edges of

clusters were most crisply in focus. For the time lapse experiment, images were taken every 20 minutes over the course of 66 hours. At approximately hour 13, microscopy was stopped for 68 minutes to refocus on each well and restarted immediately. The image processing program FIJI was used to adjust image brightness and merge fluorescent channels. In order to quantify the cell clusters, we developed an image processing pipeline, described in³⁷, which leverages FIJI³⁸, a distribution of ImageJ³⁹. The pipeline is built around the Particle Analyzer⁴⁰ plugin for FIJI which is able to perform connected components on a binary image, grouping adjacent pixels into a single cluster. These clusters are filtered based on size, excluding anything smaller than 350 um². This was chosen as a threshold because it is much smaller than the area of a single cell and excluded small pixel-level noise. In order to create the binary mask that is input into the Particle Analyzer Plugin, the fluorescent images are thresholded. The thresholds are computed on the pixel intensity values for image channel corresponding to the fluorescent channel (i.e. for blue fluorescent images, only the blue intensity channel is used, for yellow the intensity channel is created by averaging the red and green intensity channels). In order to make this analysis easily repeatable, a processing script was created in Jython. This script allows the user to specify various configuration parameters, and is able to read Leica metadata to help setup the experiment. The cluster detection method was also developed into a custom TrackMate detector plugin to track the clusters over time to allow for diffusion analysis

Diffusion-based estimation of merging time scale

The effective diffusion coefficient of the clusters was calculated by finding the slope of the best fit line of the mean squared displacement across a sweep of cell ratios. To estimate the average distance a cluster would need to cover before hitting another cluster, we iterated through every pair of clusters in an aggregate and calculated their edge distance, as measured by the distance between their XY centers minus the lengths of their radii. We identified the smallest edge distance for each cluster and averaged it with all clusters of a given cell type ratio across 5 technical replicates.

Acknowledgements

This work has been supported by the Defense Advanced Research Projects Agency under Contract No. W911NF-17-2-0098. The views, opinions, and/or findings expressed are of the author(s) and should not be interpreted as representing official views or policies of the Department of Defense or the U.S. Government. This document does not contain technology or technical data controlled under either U.S. International Traffic in Arms Regulation or U.S. Export Administration Regulations.

Author Contributions

JT designed experiments, set up 3D aggregates, and performed data analysis. JT and ML maintained and prepared cell lines and performed microscopy. NW and JB performed image analysis and data analysis. MK and SS created and analyzed the particle-based model. All authors wrote the manuscript and contributed to the interpretation and analysis of the results.

Declaration of Interests

The authors declare no competing interests.

References

- ¹ Wilson, H.V., 1907. On some phenomena of coalescence and regeneration in sponges. *Journal of Experimental Zoology*, 5(2), pp.245-258.
- ² Steinberg, M.S., 1963. Reconstruction of tissues by dissociated cells. *Science*, 141(3579), pp.401-408.
- ³ Duguay, D., Foty, R.A. and Steinberg, M.S., 2003. Cadherin-mediated cell adhesion and tissue segregation: qualitative and quantitative determinants. *Developmental biology*, 253(2), pp.309-323.
- ⁴ Werner, S., Vu, H.T.K. and Rink, J.C., 2017. Self-organization in development, regeneration and organoids. *Current opinion in cell biology*, 44, pp.102-109.
- ⁵ Cachat, E., Liu, W., Martin, K.C., Yuan, X., Yin, H., Hohenstein, P. and Davies, J.A., 2016. 2-and 3-dimensional synthetic large-scale de novo patterning by mammalian cells through phase separation. *Scientific reports*, 6, p.20664.
- ⁶ Davies, J.A., 2008. Synthetic morphology: prospects for engineered, self-constructing anatomies. *Journal of anatomy*, 212(6), pp.707-719.
- ⁷ Toda, S., Blauch, L.R., Tang, S.K., Morsut, L. and Lim, W.A., 2018. Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science*, 361(6398), pp.156-162.
- ⁸ Cerchiari, A.E., Garbe, J.C., Jee, N.Y., Todhunter, M.E., Broaders, K.E., Peehl, D.M., Desai, T.A., LaBarge, M.A., Thomson, M. and Gartner, Z.J., 2015. A strategy for tissue self-organization that is robust to cellular heterogeneity and plasticity. *Proceedings of the National Academy of Sciences*, 112(7), pp.2287-2292.
- ⁹ Glass, David S., and Ingmar H. Riedel-Kruse. "A synthetic bacterial cell-cell adhesion toolbox for programming multicellular morphologies and patterns." *Cell* 174, no. 3 (2018): 649-658.
- ¹⁰ Manning, M.L., Foty, R.A., Steinberg, M.S. and Schoetz, E.M., 2010. Coaction of intercellular adhesion and cortical tension specifies tissue surface tension. *Proceedings of the National Academy of Sciences*, 107(28), pp.12517-12522.
- ¹¹ Heisenberg, C.P., 2017. D'Arcy Thompson's 'on Growth and form': From soap bubbles to tissue self-organization. *Mechanisms of development*, 145, pp.32-37.
- ¹² Glazier, J.A. and Graner, F., 1993. Simulation of the differential adhesion driven rearrangement of biological cells. *Physical Review E*, 47(3), p.2128.

- ¹³ Belmonte, J.M., Thomas, G.L., Brunnet, L.G., De Almeida, R.M. and Chaté, H., 2008. Self-propelled particle model for cell-sorting phenomena. *Physical Review Letters*, 100(24), p.248702.
- ¹⁴ Beatrici, C.P. and Brunnet, L.G., 2011. Cell sorting based on motility differences. *Physical Review E*, 84(3), p.031927.
- ¹⁵ Nakajima, A. and Ishihara, S., 2011. Kinetics of the cellular Potts model revisited. *New Journal of Physics*, 13(3), p.033035.
- ¹⁶ Strandkvist, C., Juul, J., Baum, B., Kabla, A. J., & Duke, T. (2014). A kinetic mechanism for cell sorting based on local variations in cell motility. *Interface Focus*, 4, 20140013.
- ¹⁷ Beatrici, C. P., de Almeida, R. M. C., & Brunnet, L. G. (2017). Mean-cluster approach indicates cell sorting time scales are determined by collective dynamics. *Physical Review E*, 95, 032402.
- ¹⁸ Bonforti, A., Duran-Nebreda, S., Montañez, R., & Solé, R. (2016). Spatial self-organization in hybrid models of multicellular adhesion. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(10), 103113.
- ¹⁹ Nielsen, A.V., Gade, A.L., Juul, J. and Strandkvist, C., 2015. Schelling model of cell segregation based only on local information. *Physical Review E*, 92(5), p.052705.
- ²⁰ Cochet-Escartin, O., Locke, T.T., Shi, W.H., Steele, R.E. and Collins, E.M.S., 2017. Physical mechanisms driving cell sorting in Hydra. *Biophysical journal*, 113(12), pp.2827-2841.
- ²¹ Niessen, C.M. and Gumbiner, B.M., 2002. Cadherin-mediated cell sorting not determined by binding or adhesion specificity. *The Journal of cell biology*, 156(2), pp.389-400.
- ²² Blumlein, A., Williams, N. and McManus, J.J., 2017. The mechanical properties of individual cell spheroids. *Scientific reports*, 7(1), p.7346.
- ²³ Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. and Shochet, O., 1995. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6), p.1226.
- ²⁴ Grégoire, G., Chaté, H. and Tu, Y., 2003. Moving and staying together without a leader. *Physica D: Nonlinear Phenomena*, 181(3-4), pp.157-170.
- ²⁵ Grégoire, Guillaume, and Hugues Chaté. "Onset of collective and cohesive motion." *Physical review letters* 92, no. 2 (2004): 025702..
- ²⁶ Castro, M. G., Leggett, S. E., & Wong, I. Y. (2016). Clustering and jamming in epithelial–mesenchymal co-cultures. *Soft Matter*, 12(40), 8327-8337.
- ²⁷ Krieg, Michael, Y. Arboleda-Estudillo, P-H. Puech, J. Käfer, F. Graner, D. J. Müller, and C-P. Heisenberg. "Tensile forces govern germ-layer organization in zebrafish." *Nature cell biology* 10, no. 4 (2008): 429.
- ²⁸ Volkening, A. and Sandstede, B., 2018. Iridophores as a source of robustness in zebrafish stripes and variability in Danio patterns. *Nature communications*, 9(1), p.3231.
- ²⁹ Shyer, A.E., Rodrigues, A.R., Schroeder, G.G., Kassianidou, E., Kumar, S. and Harland, R.M., 2017. Emergent cellular self-organization and mechanosensation initiate follicle pattern in the avian skin. *Science*, 357(6353), pp.811-815.

- ³⁰ Sahu, P., Sussman, D.M., Marchetti, M.C., Manning, M.L. and Schwarz, J.M., 2019. Large-scale mixing and small-scale demixing in a confluent model for biological tissues. *arXiv preprint arXiv:1905.00657*.
- ³¹ Staple, Douglas B., Reza Farhadifar, J-C. Röper, Benoit Aigouy, Suzanne Eaton, and Frank Jülicher. "Mechanics and remodelling of cell packings in epithelia." *The European Physical Journal E* 33, no. 2 (2010): 117-127.
- ³² Yang, Xingbo, Dapeng Bi, Michael Czajkowski, Matthias Merkel, M. Lisa Manning, and M. Cristina Marchetti. "Correlating cell shape and cellular stress in motile confluent tissues." *Proceedings of the National Academy of Sciences* 114, no. 48 (2017): 12663-12668.
- ³³ Bi, Dapeng, Xingbo Yang, M. Cristina Marchetti, and M. Lisa Manning. "Motility-driven glass and jamming transitions in biological tissues." *Physical Review X* 6, no. 2 (2016): 021011.
- ³⁴ Ranft, Jonas, Markus Basan, Jens Elgeti, Jean-François Joanny, Jacques Prost, and Frank Jülicher. "Fluidization of tissues by cell division and apoptosis." *Proceedings of the National Academy of Sciences* 107, no. 49 (2010): 20863-20868.
- ³⁵ Krajnc, Matej, Sabyasachi Dasgupta, Primož Ziherl, and Jacques Prost. "Fluidization of epithelial sheets by active cell rearrangements." *Physical Review E* 98, no. 2 (2018): 022409.
- ³⁶ Duportet, Xavier, Liliana Wroblewska, Patrick Guye, Yingqiang Li, Justin Eyquem, Julianne Rieders, Tharathorn Rimchala, Gregory Batt, and Ron Weiss. "A platform for rapid prototyping of synthetic gene networks in mammalian cells." *Nucleic acids research* 42, no. 21 (2014): 13440-13451.
- ³⁷ Walczak, N, Beal, J, Tordoff, J, Weiss, R. "TASBE Image Analytics: a processing pipeline for quantifying cell organization from fluorescent microscopy." Under review.
- ³⁸ Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch et al. "Fiji: an open-source platform for biological-image analysis." *Nature methods* 9, no. 7 (2012): 676.
- ³⁹ Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T. and Eliceiri, K.W., 2017. ImageJ2: ImageJ for the next generation of scientific image data. *BMC bioinformatics*, 18(1), p.529.
- ⁴⁰ Tinevez, Jean-Yves, Nick Perry, Johannes Schindelin, Genevieve M. Hoopes, Gregory D. Reynolds, Emmanuel Laplantine, Sebastian Y. Bednarek, Spencer L. Shorte, and Kevin W. Eliceiri. "TrackMate: An open and extensible platform for single-particle tracking." *Methods* 115 (2017): 80-90.
- ⁴¹ CloudCompare [GPL software]. (2019). Retrieved from <http://www.cloudcompare.org/>

Figure Titles and Legends

Figure 1: *Ratio of highly adhesive to less adhesive cells precisely controls pattern features.* (A) Setup of 3D structure formation workflow. HEK and CHO cells were grown in 2D culture, trypsinized to a single cell suspension, and sorted with fluorescent

activated cell sorting (FACS) into round bottom ultra-low adherent plates. HEK cells are shown in yellow and CHO cells are shown in blue. (B) A ratio sweep of HEK cells (yellow) and CHO cells (blue). Aggregates were seeded at 7000 cells total and imaged after approximately 24 hours. Images show a confocal 2D central Z-slice of the aggregates. Columns represent technical replicates. (C) Quantification of all clusters across replicates in (B), with each point representing the area of a single cluster. The percent HEK label indicates the percent of HEK cells at $t=0$ when aggregates were seeded. Clusters less than 5% the area of the largest cluster in each aggregate are filtered out. (D) Quantification of the number of clusters in each aggregate in (B) according to the percent HEK cells.

Figure 2: Phenomenological model of cell sorting captures structural behaviors.
 (A) Schematic of the model. Cells, represented by disks of two types (yellow and blue circles), move with velocity v_i along a direction defined by a polar angle θ_i . Cell-cell interactions are described by pairwise forces due to steric repulsion, adhesion, and random noise at the cell-cortex level. Additionally, cells experience an external force due to an effective 2D gravitational field (red arrow). (B) Heatmap showing number of clusters in 20:80 mixtures, $t = 10^7$ after the beginning of sorting in adhesion strength parameter space (U_{12}, U_{22}) at fixed $U_{11}=30$. Each box in the phase diagram represents a single simulation result. (C-F) Representative simulation snapshots of basic types of patterns found within the range of parameters explored in (B). (G) Mean cluster size and number of clusters of type-1 cells vs. fraction of type-2 cells at $U_{11}=30$, $U_{12}=1$, $U_{22}=5$. Transition from single- to many-cluster regime occurs at fraction ≈ 0.4 . Panel G shows results from a single representative set of runs.

Figure 3: Adhesive subpopulation ratio and total cell number precisely control cluster features

(A,B) Simulated results of the mean cluster size (A) and mean number of clusters (B) of aggregates at $U_{11}=30$, $U_{12}=1$, $U_{22}=5$ and various cell-type ratios across different number of cells. Lines represent the average values of 5 simulations at $t = 10^7$ and error bars represent the standard deviation between these values. (C,D) Experimental quantification of the cluster size (C) and mean number of clusters (D) of aggregates with various HEK to CHO ratios across different total cell numbers, as counted and seeded at $t=0$ and imaged at 24 hours across three technical replicates. Clusters less than 5% the area of the largest cluster in an aggregate are excluded. (C) Violin plots of each cluster across all replicates. Black bars represent the mean. (D) Total number of clusters in an aggregate at a given ratio for three technical replicates.

Figure 4: *Incomplete sorting of highly adhesive cells and less adhesive cells creates long-term stable clusters.* (A) A particle-based model simulation of cluster formation over time with parameters $U_{11}=30$, $U_{12}=1$, $U_{22}=5$ in 1000 cells. Lines represent the mean number of clusters at various highly adhesive to less adhesive cell population ratios, averaged over 5 simulations. The shaded region indicates the standard deviation between these simulations. The x axis represents the number of hours in the simulation, scaled to time steps by a factor of $28,571^{-1}$ and starting at time step 5×10^5 . This scaling was determined by calibrating by eye the time axis in panel A with that in panel B. (B) Image analysis of experimental data of cluster formation over 35 hours. Lines represent the average over 5 replicates of the specified HEK:CHO ratio with 8000 cells in total, with the shaded region representing the standard deviations between replicates. Clusters less than 5% of the maximum cluster in an aggregate are excluded. (C) The average distance, D_{\min} , between a cluster and its nearest neighbor at different ratios of strongly to weakly adhesive cells. (D) The effective diffusion coefficient D_{eff} of clusters across three time ranges. (E) The time predicted for a cluster to cover the average distance between itself and its nearest cluster neighbor, calculated as the ratio of D_{\min}^2 to D_{eff} .

Supplementary Videos

- Video 1: VID_0_perc_HEK
- Video 2: VID_10_perc_HEK
- Video 3: VID_30_perc_HEK
- Video 4: VID_60_perc_HEK
- Video 5: VID_90_perc_HEK
- Video 6: VID_100_perc_HEK

Supplemental Video 1: *3D Z-Stacks of representative aggregate with 100% CHO cells*, related to **Figure 1**. The aggregate was imaged approximately 24 hours after seeding using a Leica SP5 confocal microscope. In total, 7000 CHO and HEK cells were seeded. Z-slices were taken $5\mu\text{m}$ apart. Z-stacks were converted to 3D renderings by first converting them to point clouds, which were created by turning each pixel in the image into a 3D point. Using the microscope parameters to get the physical size and depth of a pixel, the X,Y,Z coordinates of each pixel in the image were transformed into microns. To remove background fluorescence, a manual color intensity threshold of 120 was used (with colors being in the range of 0-255). Points were saved in a PLY file and visualized using CloudCompare⁴¹. Images are 1.55mm in width and 1.55 in length.

Supplemental Video 2: *3D Z-Stacks of representative aggregate with approximately 10% HEK cells*, related to **Figure 1**. The aggregate was imaged approximately 24 hours after seeding using a Leica SP5 confocal microscope. In total, 7000 CHO and HEK cells were seeded. Z-slices were taken $5\mu\text{m}$ apart. Z-stacks were converted to 3D renderings by first converting them to point clouds, which were created by turning each pixel in the image into a 3D point. Using the microscope parameters to get the physical size and depth of a pixel, the X,Y,Z coordinates of each pixel in the image were transformed into microns. To remove background fluorescence, a manual color intensity threshold of 120 was used (with colors being in the range of 0-255). Points were saved in a PLY file and visualized using CloudCompare⁴¹. Images are 1.55mm in width and 1.55 in length.

Supplemental Video 3: *3D Z-Stacks of representative aggregate with approximately 30% HEK cells*, related to **Figure 1**. The aggregate was imaged approximately 24 hours after seeding using a Leica SP5 confocal microscope. In total, 7000 CHO and HEK cells were seeded. Z-slices were taken $5\mu\text{m}$ apart. Z-stacks were converted to 3D renderings by first converting them to point clouds, which were created by turning each pixel in the image into a 3D point. Using the microscope parameters to get the physical size and depth of a pixel, the X,Y,Z coordinates of each pixel in the image were transformed into microns. To remove background fluorescence, a manual color intensity threshold of 120 was used (with colors being in the range of 0-255). Points were saved in a PLY file and visualized using CloudCompare⁴¹. Images are 1.55mm in width and 1.55 in length.

Supplemental Video 4: *3D Z-Stacks of representative aggregate with approximately 60% HEK cells*, related to **Figure 1**. The aggregate was imaged approximately 24 hours after seeding using a Leica SP5 confocal microscope. In total, 7000 CHO and HEK cells were seeded. Z-slices were taken $5\mu\text{m}$ apart. Z-stacks were converted to 3D renderings by first converting them to point clouds, which were created by turning each pixel in the image into a 3D point. Using the microscope parameters to get the physical size and depth of a pixel, the X,Y,Z coordinates of each pixel in the image were transformed into microns. To remove background fluorescence, a manual color intensity threshold of 120 was used (with colors being in the range of 0-255). Points were saved in a PLY file and visualized using CloudCompare⁴¹. Images are 1.55mm in width and 1.55 in length.

Supplemental Video 5: *3D Z-Stacks of representative aggregate with approximately 90% HEK cells*, related to **Figure 1**. The aggregate was imaged approximately 24 hours after seeding using a Leica SP5 confocal microscope. In total, 7000 CHO and HEK cells were seeded. Z-slices were taken $5\mu\text{m}$ apart. Z-stacks were converted to 3D renderings by first converting them to point clouds, which were created by turning each pixel in the image into a 3D point. Using the microscope parameters to get the physical size and

depth of a pixel, the X,Y,Z coordinates of each pixel in the image were transformed into microns. To remove background fluorescence, a manual color intensity threshold of 120 was used (with colors being in the range of 0-255). Points were saved in a PLY file and visualized using CloudCompare⁴¹. Images are 1.55mm in width and 1.55 in length.

Supplemental Video 6: *3D Z-Stacks of representative aggregate with 100% HEK cells*, related to **Figure 1**. The aggregate was imaged approximately 24 hours after seeding using a Leica SP5 confocal microscope. In total, 7000 CHO and HEK cells were seeded. Z-slices were taken 5 μ m apart. Z-stacks were converted to 3D renderings by first converting them to point clouds, which were created by turning each pixel in the image into a 3D point. Using the microscope parameters to get the physical size and depth of a pixel, the X,Y,Z coordinates of each pixel in the image were transformed into microns. To remove background fluorescence, a manual color intensity threshold of 120 was used (with colors being in the range of 0-255). Points were saved in a PLY file and visualized using CloudCompare⁴¹. Images are 1.55mm in width and 1.55 in length.







