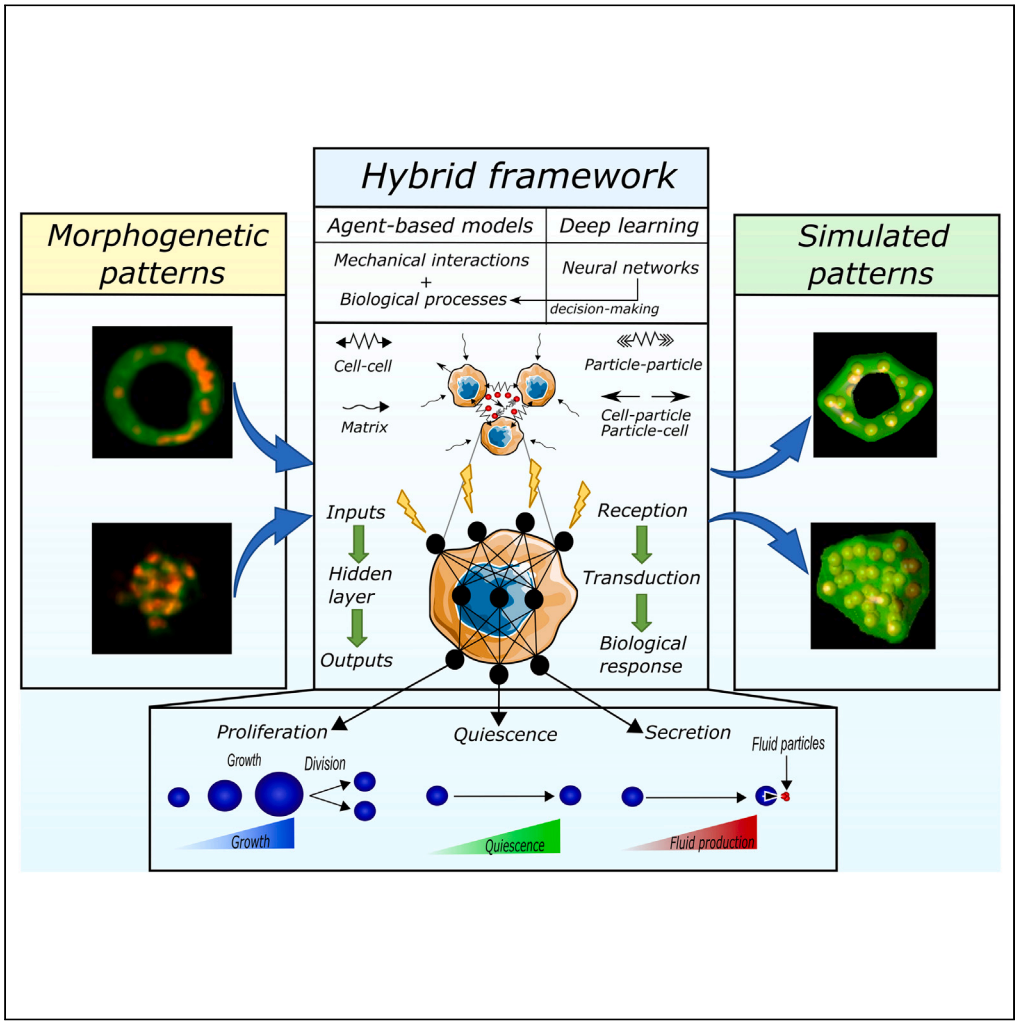
iScience

CellPress
OPEN ACCESS

Article

A hybrid physics-based and data-driven framework for cellular biological systems: Application to the morphogenesis of organoids



Daniel Camacho-Gomez, Ioritz Sorzabal-Bellido, Carlos Ortiz-de-Solorzano, Jose Manuel Garcia-Aznar, Maria Jose Gomez-Benito

gomezmj@unizar.es

Highlights

Framework combining agent-based modeling and deep learning

Deep learning manages cell decisions of biological functions in the agent-based model

Model reproduces morphogenetic patterns from image data metrics

Simulates organoid development, unraveling cell coordination

Camacho-Gomez et al., iScience 26, 107164
July 21, 2023
https://doi.org/10.1016/j.isci.2023.107164

iScience

CellPress
OPEN ACCESS

Article

A hybrid physics-based and data-driven framework for cellular biological systems: Application to the morphogenesis of organoids

Daniel Camacho-Gomez,[1] Ioritz Sorzabal-Bellido,[2] Carlos Ortiz-de-Solorzano,[2] Jose Manuel Garcia-Aznar,[1] and Maria Jose Gomez-Benito[1,3,*]

SUMMARY

How cells orchestrate their cellular functions remains a crucial question to unravel how they organize in different patterns. We present a framework based on artificial intelligence to advance the understanding of how cell functions are coordinated spatially and temporally in biological systems. It consists of a hybrid physics-based model that integrates both mechanical interactions and cell functions with a data-driven model that regulates the cellular decision-making process through a deep learning algorithm trained on image data metrics. To illustrate our approach, we used data from 3D cultures of murine pancreatic ductal adenocarcinoma cells (PDAC) grown in Matrigel as tumor organoids. Our approach allowed us to find the underlying principles through which cells activate different cell processes to self-organize in different patterns according to the specific microenvironmental conditions. The framework proposed here expands the tools for simulating biological systems at the cellular level, providing a novel perspective to unravel morphogenetic patterns.

INTRODUCTION

Organ morphogenesis can be studied *in situ* through the analysis of stained tissue sections taken at different phases of embryonic development. This provides relevant, albeit static, information about the evolution of the morphology and the interactions between the elements of the developing organ. Morphogenesis can also be studied using 2D *in vitro* models built from stem cells properly stimulated with growing factors and/or cocultured cells. These models replicate some of the cellular interactions that occur during organ development, providing important information about the signaling pathways involved in normal organ formation. 2D culture models, however, lack a key element of organ morphogenesis, namely, the complex ensemble of three-dimensional spatial interactions that occur during the developmental process. These interactions, which determine the successful series of events that end with the formation of a mature organ, include not only interactions between cells but also, very importantly, interactions between the cells and their local microenvironment. This has caused the birth of a growing trend in the study of morphogenesis, which is the use of 3D *in vitro* cellular models or organoids that recreate *in vitro* the process of organ formation.[1–3] Some organotypic models have also been developed that replicate the growth and development of tumors (tumor organoids) from cancer stem cells and the appropriate stimulation of growth factors and biomechanical cues.[4,5] These complex models replicate the interactions that exist between cancer cells and with other cells of the tumor microenvironment, most relevantly the cells of the immune system. Tumor organoids can be of different complexity, from simple spheroids made of one or a few cell types suspended in medium enriched with the appropriate growth factors[6] to more complex models comprising several cell types embedded in a biomimetic hydrogel. Finally, more advanced models are being developed based on microfabrication-bioprinting and microfluidic technologies that allow the simulation of interstitial flows.[7] Tumor organoids are becoming invaluable tools to study the response of tumors to therapies, including those that stimulate the host immune system.

There are also different approaches to simulate morphogenesis *in silico*. On the one hand, continuum models permit the simulation of large cell populations at the tissue level. This is the case for reaction-diffusion systems based on partial differential equations (PDEs)[8] or positional information (PI).[9] However, these methods neglect the individuality of cells in favor of larger scales, disregarding the important role played by cell-to-cell

[1]Department of Mechanical Engineering, Multiscale in Mechanical and Biological Engineering (M2BE), Aragon Institute of Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain

[2]Solid Tumors and Biomarkers Program, IDISNA, and CIBERONC, Center for Applied Medical Research, University of Navarra, Zaragoza, Spain

[3]Lead contact

*Correspondence: gomezmj@unizar.es

https://doi.org/10.1016/j.isci.2023.107164

interactions and the phenotypic diversity of the biological system. Consequently, these methods struggle to reproduce the way cell communities develop into complex structures, especially in three dimensions. On the other hand, discrete approaches with agent-based modeling consider cells as autonomous entities that interact among themselves and with the microenvironment.[10] These methods are usually classified into lattice models, in which cell positions are fixed within a spatial grid,[11] and lattice-free models, which allow continuous cell positioning in space.[12] Based on the way the cell shape is represented in lattice-free models, there are center-based models, in which cells are represented as spheres and are described by their centers and radii,[13] and deformable cell models and vertex models, which account for the cell shape and deformability.[14,15] These models have been widely employed, for instance, to simulate tumor growth *in vitro,*[16] to study the role of extracellular matrix density in cell migration within tumor spheroids,[17] collective cell migration,[18] or tissue regeneration.[19] These approaches seem appropriate to simulate morphogenesis, it being a process that evolves from a single cell, and the emerging pattern emerges from cell interactions, cell heterogeneity, and cell processes.[20,21] Furthermore, agent-based modeling focuses on the cell level; therefore, it very flexibly simulates *in vitro* experiments. However, these models have some challenges related to the designation of rules and fitting parameters that govern decision-making heuristics and the behavior of cells.[20,21] In addition, they usually have a large number of parameters compared to continuous models, especially when considering molecular events per cell,[22] which makes their calibration difficult.

Data-driven models are widely used to predict the behavior of complex physical systems.[23,24] They can unravel unknown phenomena and reduce the degree of abstraction of computational models. In computational biology, there are examples of data-driven models used to find the optimal parameter values in gene regulatory networks to reproduce the growing limb bud[25] or Bayesian inference of agent-based cellular automaton to study kidney branching morphogenesis.[26] These methods combining data-driven and computational models aim to obtain the required parameters of the computational model, which is an optimization method that lacks generalization, as in the case of the data changing, the parameters must change as well. In the context of data-driven methods, there is an increasing use of artificial intelligence and, in particular, machine learning and deep learning algorithms.[27] However, data-driven models are considered black boxes and are difficult to relate to physical interpretations, because they are mainly focused on finding input-output relationships.

In this work, we propose a hybrid approach that combines deep learning and agent-based models as a novel computational framework to simulate cellular biological systems. In other approaches, the coordination between cells is fixed beforehand and then adjusted to reproduce a certain behavior. On the one hand, in continuum models, cell processes are included in constitutive equations and these cell functions are performed at the same time.[28,29] Then, they are adjusted through a sensitivity analysis to obtain the fitting solution. However, there is no competition between cell functions, because they occur simultaneously. On the other hand, in stochastic models, cells change between cell states through cell state rules, Boolean circuits, gene regulatory circuits (GRNs), and biochemical reactions from the more abstract to the more detailed.[30,31] These relationships between cell states are complex to establish and describe.[32] Both approaches are usually made to simulate one particular case or median cases from experiments. However, it is difficult to reproduce different cases since their parameters are adjusted to reproduce one observed behavior and the variability is obtained by the introduction of noise. Moreover, they usually fail to reproduce dynamical processes because of the intricacy of recreating distinct behaviors temporarily with fixed parameters and even more taking into account the variability between cells. In this new work, we do not determine how cells behave before defining the relationships between cell states to perform a sensitivity analysis to find the parameters that allow us to reproduce an observed behavior. Instead, we eliminate all the parameters in the physics-based model that preestablish the behavior of cells and incorporate the mechanical cell interactions and biological cell functions in an agent-based model that uses a neural network to drive the cellular decision-making of the agents during the simulation, replicating how cells make decisions and behave *in vivo*. This deep learning algorithm coordinates cell decision-making by evaluating the simulation in real time and comparing it with metrics that define the final architecture of the morphogenetic pattern extracted from cell-level images of real organoids or *in vivo* data.

To validate and illustrate this framework, we show three different applications. First, we explain how to particularize this approach to simulate the formation and growth of tumor organoids from a mouse pancreatic adenocarcinoma cell line grown in a 3D biometric Matrigel matrix. These organoids form cystic structures, namely, an organized architecture of cells around a fluid-filled space called the lumen. Then, we show the simulations of these experiments. In addition, we show the application of the approach to simulate the formation and growth

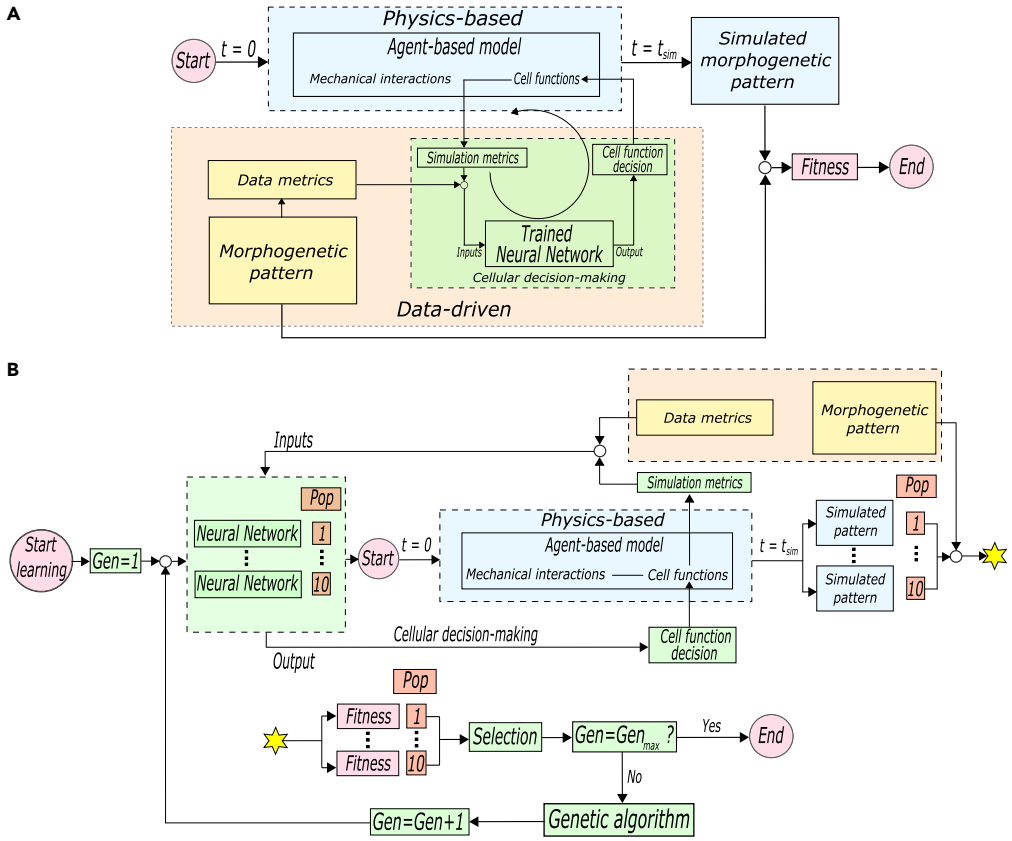


Figure 1. Framework of the hybrid physics-based and data-driven model

(A) Schematic overview of the hybrid data-driven and physics-based algorithm. It consists of an agent-based model that incorporates mechanical interactions and cell biological functions. To orchestrate the cell functions within the simulation, the agent-based model employs a data-driven algorithm. Thus, to determine which function the cell has to perform, the agent-based model goes into the data-driven algorithm. The data-driven algorithm extracts from the desired morphogenetic pattern data metrics that characterize the pattern, which are introduced, together with simulation metrics, into the neural network as inputs. The neural network evaluates the inputs and determines the cell function that the cell must perform. This process is repeated until the simulation time of the agent-based model reaches (t_{sim}). Finally, a fitness value is given to the simulation by comparing the simulated and desired morphogenetic patterns.

(B) Schematic overview of the training algorithm of the neural network. The training starts with the creation of a random generation (*Gen*) composed of a population (*Pop*) of ten neural networks. Then, the physics-based algorithm performs a simulation with each neural network, following a similar methodology as (A). Finally, the genetic algorithm selects the two best populations based on their fitness and creates a new generation through crossover and mutation algorithms. The training is concluded when the number of generations reaches the specified number of generations (Gen_{max}).

of solid tumor organoids from other mouse pancreatic adenocarcinoma cell lines grown in a 3D biometric Matrigel matrix, which consists of dense aggregates of cells. Finally, we apply this framework to simulate a theoretical case of the evolution of solid tumor organoids with time-dependent quantitative data.

RESULTS

Hybrid physics-based and data-driven framework

We present a novel hybrid physics-based and data-driven framework that combines agent-based modeling and deep learning to simulate morphogenetic patterns (Figure 1). The purpose of the physics-based model is to replicate a specific morphogenetic pattern. The data-driven algorithm in turn coordinates, through a neural network, the cell functions required in the physics-based model to achieve that pattern (Figure 1A). Thus, the physics-based part consists of an agent-based model that couples mechanical interactions and the cell biological functions that are decisive in the morphogenetic process, namely, proliferation, quiescence, secretion, migration, cell death, etc. The selection of these cell functions depends on the morphogenetic process being simulated.

The physics-based model coordinates the biological behavior of cells to form a specific pattern, which also requires the appropriate mechanical interactions. For that coordination, instead of using parameters that fix the relationships between them, we introduce a data-driven algorithm that performs the cellular decision-making of the cell functions. Therefore, during the simulation of the agent-based model, the neural network determines which cell function each cell has to perform (Figure 1A). To this end, given a morphogenetic pattern, we first define the data metrics that characterize the pattern and establish target values. These data metrics set the objective of the simulation. Then, when a cell needs to decide which function to perform, the neural network receives as inputs the predefined data metrics and the simulation metrics from the agent-based model that evaluates the actual state of the simulation. In this way, the neural network compares the simulation with the target values and makes a decision (see the data-driven box in Figure 1A). Each node of the output layer is related to one cell function; therefore, the activated node determines the biological function that the cell will perform. This decision is fed to the agent-based model, which initiates the corresponding cell function for that cell. This process is repeated during the predefined time of simulation every time a cell finishes the biological function that it was performing. When the simulation finishes, a fitness value is reported, which indicates how close the simulated morphogenetic pattern is compared to the desired morphogenetic pattern. Therefore, the result depends on how well the neural network evaluates the simulation and coordinates cell functions through adequate decision-making.

Learning by the neural network framework

To simulate the morphogenetic pattern, the neural network needs to learn how to evaluate the simulation and, depending on the target values, decide which biological function the cell has to perform. To make the neural network learn, we propose a learning methodology based on a genetic algorithm (Figure 1B). Initially, the genetic algorithm creates a random population of ten neural networks with different weight values assigned, i.e., a generation. Then, we perform a simulation of the agent-based model with each neural network of the generation. When the simulation finishes, we obtain a simulated pattern for each population. Subsequently, each neural network of the generation is given a fitness value, which represents how well they performed in the simulation. Finally, it selects from the current generation the two fittest neural networks and obtains, through a genetic algorithm (further details are included in learning algorithm in STAR Methods), the next generation of neural networks. This process is repeated for a predefined number of generations.

Application of the physics-based framework to the simulation of cystic organoids

To show the potential of our approach, we simulated cystic tumor organoids, i.e., organoids formed by an outer shell made of cancer cells surrounding a lumen (Figure 2). Lumen morphogenesis implies the development of an organized architecture of cells around a fluid-filled space. This inner space, called the lumen, is found in many parts of metazoan organisms and allows them to perform specific functions. The development of a lumen requires sophisticated coordination of several cell processes. Specifically, three basic mechanisms are crucial[33]: Initially, cells proliferate and polarize to determine the landmark where the lumen will be initiated. Next, cells secrete fluid into that central site and generate luminal hydrostatic pressure that makes the lumen enlarge.[34–37] To preserve the overall cystic structure, a balance between strong cell-cell junctions and spatially controlled cell mitosis is needed. Finally, the stiffness of the extracellular matrix (ECM) plays an important role[38] by balancing the luminal hydrostatic pressure and providing physical cues that regulate cell behavior.[39]

To simulate the morphogenesis of the lumen, we built a three-dimensional agent-based model. This physics-based model is based on a previous three-dimensional lattice-free, center-based model.[40] Here, we use the structure of that model, but we eliminate the parameters that coordinate the relationships between cell functions, the effect of the cell net forces on the cell cycle, and multiluminal polarization mechanisms. To reproduce the morphogenesis of organoids composed of cells enclosing a fluid-filled lumen, we consider two types of agents: cells, which are the biological entities, and particles, which are secreted by cells and simulate the lumen fluid. The mechanical equilibrium between cells anchored to the ECM and the luminal pressure ensures the maintenance of the luminal architecture (further details are included in agent-based modeling in STAR Methods). To model this mechanical equilibrium, we use agents that interact mechanically to generate and maintain the lumen. Thus, cells interact among themselves based on pairwise potential functions in an adhesive-repulsive manner. The repulsion between agents mimics cell resistance to deformation when their membranes touch, and the attractive forces are the result of the cell junctions exerted through specialized protein complexes. Moreover, particles also interact via pairwise potential functions and interact with cells in a repulsive manner. This interaction mimics the luminal hydrostatic pressure generated by the cells' fluid secretion,

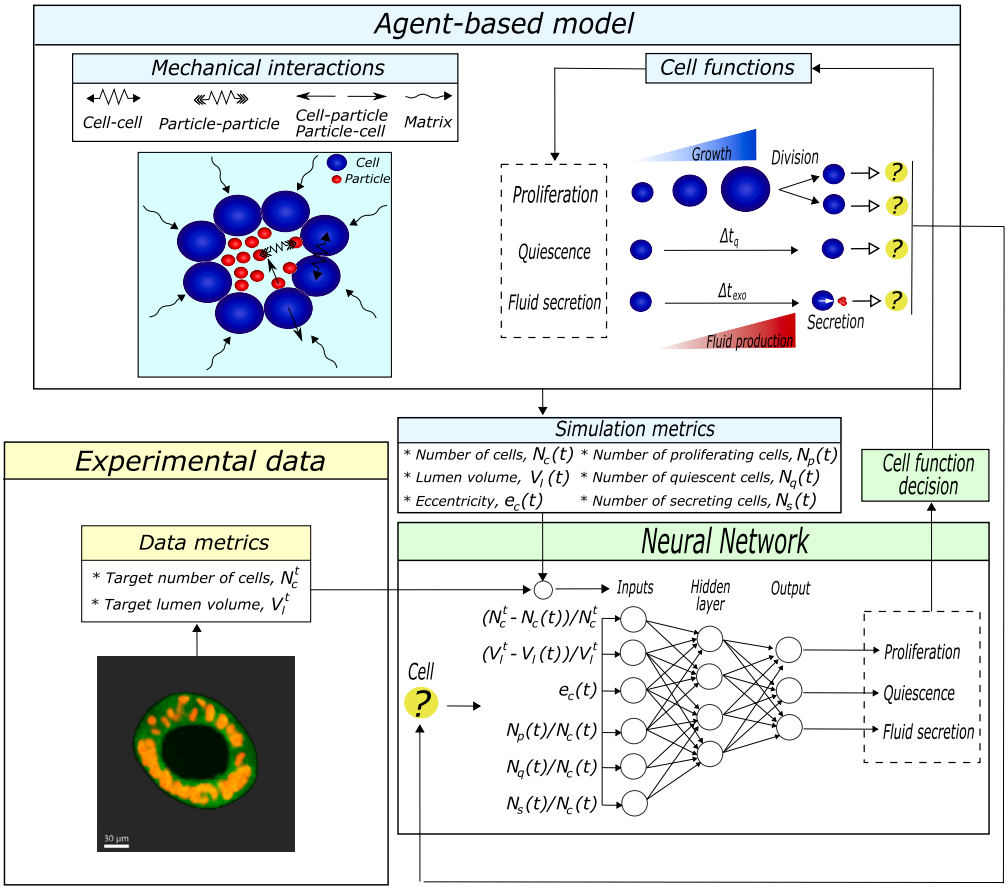


Figure 2. Particularization of the framework for organoids with lumen

The physics-based model consists of an agent-based model that integrates mechanical interactions and cell functions. Cells are biological entities, and the particles simulate the lumen fluid. Three cell functions are considered: proliferation, quiescence, and fluid secretion. Δt_q is the minimum period of time that a cell remains quiescent and Δt_{exo} is the fluid production time.

and it is responsible for the movement of cells and generation of the luminal space. Finally, agents interact with the extracellular matrix by means of a friction coefficient that represents the dynamic viscosity of the matrix (Figure 2 top left).

Three decisive cell functions are considered to achieve the cystic morphogenetic pattern: proliferation, quiescence, and fluid secretion (Figure 2 top right). First, we define a mathematical model for simulating the cell cycle that regulates cell proliferation. Overall, it consists of a growth phase, in which the cell progressively increases its volume as a result of DNA replication, and a mitosis phase, in which the cell divides into two daughter cells (see cell division in STAR Methods). In addition, we include a quiescent state in which the cell remains inactive for a period of time (Δt_q). Finally, cells can produce fluid during a period of time (Δt_{exo}) and then secrete it inside the lumen, increasing the lumen volume. To simulate this process, we assume that cells generate particles that simulate the lumen fluid (further details are included in characterization of the lumen fluid in STAR Methods). Consequently, the lumen is in a state of hydrostatic pressure because of this cell secretion.

Application of the data-driven framework to the simulation of cystic organoids

In the data-driven framework, we first define the metrics that characterize the pattern using target values. In the case of organoids with lumen, we defined the target number of cells (N_c^t) and target lumen volume (V_l^t) to describe the morphology of the organoid (Figure 2 middle). We also define several simulation metrics that are updated when a cell has to decide its next cell function. Specifically, these simulation metrics

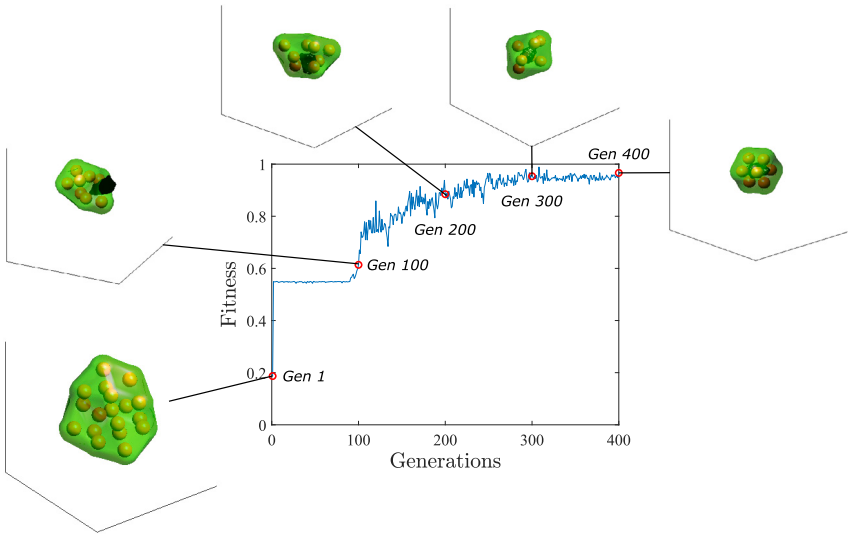


Figure 3. Learning of the neural network

The fitness value represented corresponds with the highest fitness of the population within the generation. Snapshots of the achieved organoid inserted in a cube with 100 μm sides featuring one population at generations 1, 100, 200, 300, and 400 are represented. The spheres represent the nuclei of cells with radius $R_c/2$ and the green hull is an estimation of the cell membrane through alpha shapes of the cells with $\alpha = 2R_c$.[41] The lumen fluid is represented in black through the alpha shapes of particles with the smallest alpha that produces an alpha shape enclosing all of the particles.

are the total number of cells, the lumen volume, the eccentricity between the cells center of mass and the lumen fluid center of mass (see eccentricity in STAR Methods), and the number of proliferating, quiescent and secreting cells. These values are given as inputs to a neural network (Figure 2 bottom right). We assume that each cell can only perform one cell function at a time. Hence, the neural network determines the appropriate cell function that cells must perform. The neural network consists of an input layer with six nodes, one hidden layer with four nodes and a hyperbolic tangent activation function, and a layer with three outputs and a sigmoid activation function. Each node of the output layer identifies the function that the cell must perform: proliferation, quiescence, or fluid secretion.

Learning of the neural network to simulate cystic organoids

The success of the morphogenetic pattern depends on how well the neural network coordinates cell functions during the simulation. To make the neural network learn to form organoids with a lumen, we trained the neural networks for 400 generations. For this purpose, we did not use any of the experimental data; instead, we used a synthetic pattern consisting of an organoid with 8 cells and a lumen of 1000 μm^3. To illustrate this process, we show the increase in performance for each generation of neural networks (Figure 3). As shown, the first generation of neural networks did not coordinate the cell functions properly, because the fitness (see fitness function in STAR Methods) of this generation was 0.1869. In fact, the neural network only allowed cells to proliferate, thus generating an organoid without a lumen. Then, the performance increased to 0.6136 in generation 100, because cells were also secreted to generate the lumen. However, the fluid of the lumen leaked because of the poor coordination between proliferation and secretion. In generation 200, the fitness increased to 0.8846, and better coordination than in generation 100 was found, in that lumen fluid did not leak, but the organoid was too asymmetrical. Generation 300 achieved a symmetrical organoid with lumen with a better approximation of the number of cells and lumen volume. Finally, in generation 400, an organoid composed of 7 cells and a lumen volume of 1245 μm^3 was achieved.

Simulation of pancreatic cystic tumor organoids

We simulated 3D cultures of murine pancreatic ductal adenocarcinoma cells grown in a biomimetic matrix made of Matrigel (4 *mg/ml*). Real 3D cultures of PDAC93-GFP cells were generated and incubated for 5 days. In the experimental setup, we generated a randomly distributed population of organoid seeds ranging from single cells to small clusters of 2–3 cells (Figure S2). Then, multiphoton microscopy images of the grown organoids were segmented and quantified to characterize their morphology. In particular,

Table 1. Experiment and simulation of pancreatic tumor organoids with lumen

	Experiment		Simulation	
Organoid	N_c^t	$V_l^t(\mu m^3)$	N_c	$V_l(\mu m^3)$
1	29	11263	30	11330
2	37	13233	41	14097
3	13	1985	14	2174
4	59	42762	59	43239

we obtained the number of cells in the organoid and the lumen volume, which were used as target values for the simulation. We chose organoids of different sizes to test the generalization potential of the methodology (Table 1). Specifically, we used two intermediate-size samples (organoids #1 and #2), one small-size sample (organoid #3), and one large-size sample (organoid #4). We initiated the simulations with one cell and ran the simulations for 7 days. Although the PDAC93-GFP cells were incubated for 5 days, we extended the simulation time to 7 days to prove that the model was able to reach steady-state patterns. The parameter values used in the simulations are described in Table S1.

We show images of the experimental organoids used as targets of the simulation, the achieved solution of one of the simulations performed, and the coordination of cell functions (Figure 4) corresponding to the organoids in Table 1. As shown, sophisticated coordination of cell functions is required to form an organoid with a lumen (Figure 4B). In our simulations, cells first proliferated, generating a closed volume. Then, some cells started secreting fluid, whereas other cells continued proliferating to create more luminal space to secrete and prevent fluid leakage. In organoid #1, 2 out of 16 cells (13%) started secreting after 84 h, and then at 105 h, the rest of the cells also contributed to lumen enlargement by producing and secreting fluid (Video S1). In the case of organoid #2, 23 out of 32 cells (72%) were secreting at 105 h, whereas 9 cells continued proliferating. After these 9 cells finished proliferating, the target lumen volume was already reached, so no more fluid secretion was needed (Video S2). In organoid #3, cells started secreting earlier (63 h) than in organoids #1, #2 and #4. In this situation, only two cells were secreting over approximately 8 h since the lumen volume of this organoid was small (Video S3). In the case of organoid #4, 5 cells out of 32 cells (16%) started secreting at 105 h, and 21 h later, the remaining cells that initially were proliferating were also secreting. Finally, cells entered a quiescent state in all four organoids when they achieved the final pattern.

To demonstrate the robustness of the neural network to properly coordinate cell functions, we performed 10 simulations with the same neural network for each organoid. Both the number of cells (Figure 4C) and the lumen volume (Figure 4D) obtained in the simulations consistently approximated the target data. The fitness of the simulations was evaluated similarly to the fitness function used for the training of the neural network but neglected the penalization of the eccentricity, since we were not aiming to approximate the eccentricity of the *in vitro* experiments, because it was not used as input data (Figure 4E). The minimum median value of the fitness was 0.978 for organoid #2. The median values for organoids #1, #3, and #4 were all above 0.983.

Moreover, a sensitivity analysis was performed to investigate the influence of the main parameters of the model on the coordination of cells. For that purpose, organoid #1 was simulated in situations where the cell cycle time was decreased and increased by 20% from 30 h to 24 h and 36 h respectively (Figure S7). Originally, the cells finished proliferating after 105 h. A cell cycle time reduction makes cells proliferate faster; therefore, cells finished proliferating after 84 h. Conversely, cells proliferate slower when the cell cycle time increases, thus requiring approximately 125 h to finish proliferation. In addition, a similar sensitivity analysis for the fluid production time (Δt_{exo}) was performed (Figure S8). When the fluid production time was 15 min, cells secreted during 23 h. A reduction of the fluid production time to 6 min caused cells to secrete during 21 h, only 2 h less than in the previous case. However, in this case, only two cells were secreting, since they were able to produce more fluid over the same time period. Finally, when the fluid production time increased to 60 h, cells were secreting over 34 h. Although the variation in these parameters produced changes in the temporal evolution of the process, neither the variation in the cell cycle time nor the differences in the fluid production time altered the coordination of cell functions. Therefore, there might exist a specific organization of cell functions independent of cell type.

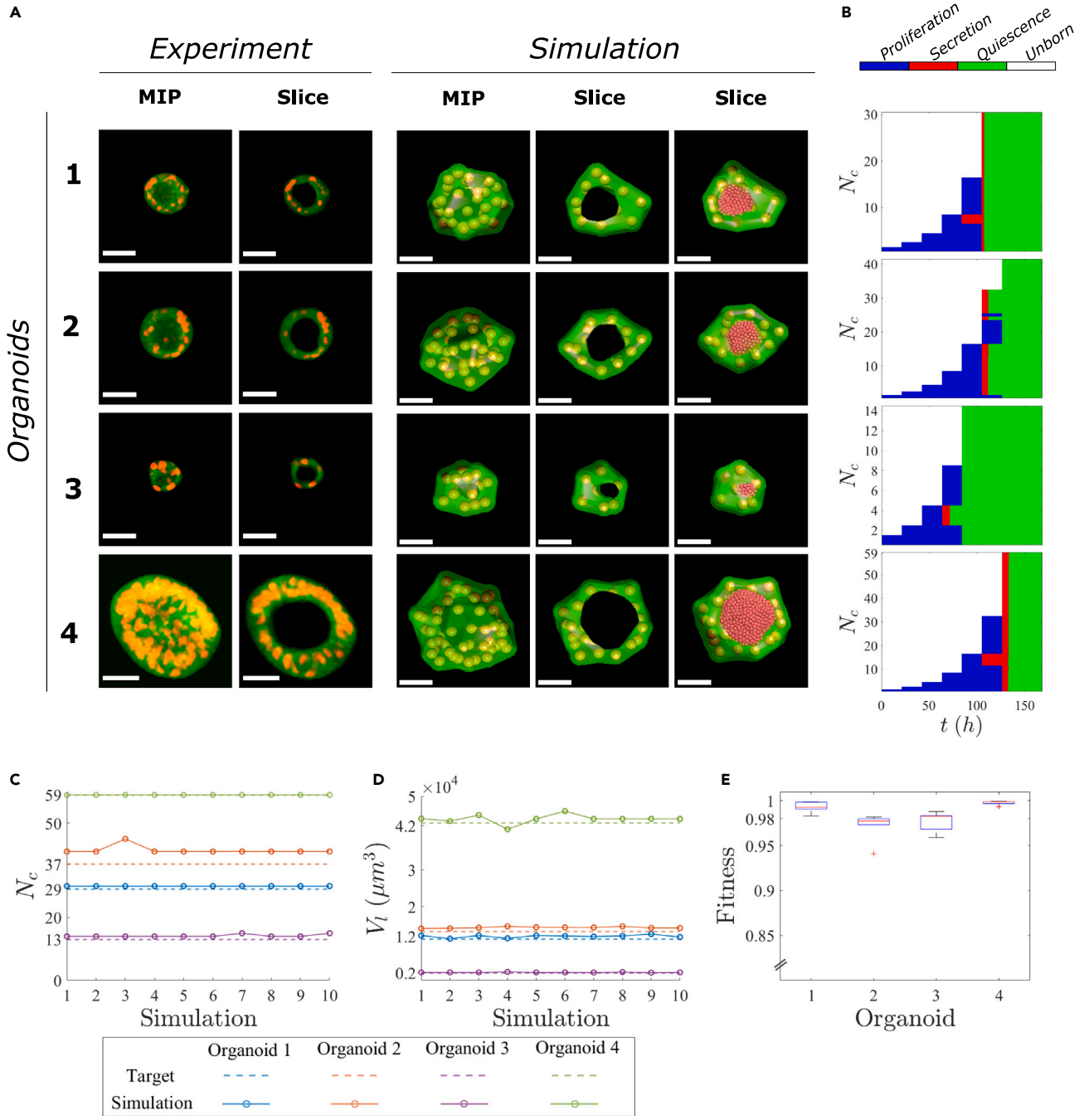


Figure 4. Simulation of experimental pancreatic tumor organoids with lumen

(A) Image of *in vitro* organoids and a slice view. Maximum Intensity Projection (MIP). Snapshots of the full view, a slice of the simulated organoids, and a slice of the simulated organoids with the lumen particles representation. The spheres represent the nuclei of cells with radius $R_c/2$ and the green hull is an estimation of the cell membrane through alpha shapes of the cells with $\alpha = 2R_c$. The lumen fluid is represented in black (fourth column) through the alpha shapes of particles with the smallest alpha that produces an alpha shape enclosing all of the particles. All scale bars are 30 μm.

(B) Coordination of cell functions, in which blue represents proliferation, red represents secretion, green represents quiescence, and white represents an unborn cell.

(C) Target number of cells (N_c^t) and the number of cells (N_c) in the 10 simulations for each organoid.

(D) Target lumen volume (V_l^t) and lumen volume (V_l) in the 10 simulations for each organoid.

(E) Boxplot of the fitness value of the 10 simulations for each organoid.

Table 2. Experiment and simulation of solid tumor organoids

	Experiment	Simulation
Organoid	N_c^t	N_c
1	23	23
2	34	34
3	53	53
4	60	60

Finally, an analysis of the effect of the noise on the parameters was performed. Accordingly, the cell cycle time and fluid production time parameters varied ±20%. Thus, every time a cell is set to perform either proliferation or fluid secretion, a random variation of the parameter for that individual cell is made between [-20, +20] % of the value from Table S1 and assigned for its cell cycle time or fluid production time. To determine how this variation could affect the results, we simulated an intermediate cystic organoid (organoid #2). Thus, we performed 10 simulations of this organoid to determine the effect of the fluctuation of the parameter. We found that the coordination of cell functions when the parameters are variable is similar to the case without any variability (Figure S9). However, there is a desynchronization between cells in terms of cell functions caused by the variability of the duration of their functions. To better compare the coordination of the cell functions, we represented the evolution of the normalized number of proliferating cells, secreting cells, and quiescent cells over time (Figure S9C). In the three cases, the evolution of the functions performed by the cells in the case with variable parameters was aligned with the case without variation. Despite the introduction of this noise, the median value of the fitness of the ten simulations was 0.9956, and the minimum was 0.9537 (Figure S9D).

Application of the framework to solid tumor organoids

To illustrate the possibility of applying this framework to other morphogenetic patterns, we applied our methodology to simulate the formation of solid tumor organoids (Figure S10). These solid tumor organoids consist of dense spherical aggregates of cells. To simulate the formation of this pattern, we considered two cell functions: proliferation and quiescence. Furthermore, we consider cell-cell mechanical interactions based on pairwise potential functions and the interaction of cells with the extracellular matrix by means of a friction coefficient. In the data-driven part of the model, we employed the target number of cells (N_c^t) as the metric that defines the size of the organoid and fixes the objective of the simulation. We also extracted some simulation metrics, which are the total number of cells, the number of proliferating cells, and the number of quiescent cells. The neural network consists of an input layer of three nodes, a hidden layer of two nodes, and an output layer of two nodes, which represents either proliferation or quiescent states of the cell. Therefore, the neural network coordinates both processes to achieve the target size of the organoid.

Simulation of solid tumor organoids

We simulated the corresponding experiments of 3D cultures of PM12500-GFP tumor cells grown also in a biomimetic matrix (Matrigel 4 *mg/ml*). This cell line is characterized by the formation of organoids composed of aggregates of cells that do not form a lumen. Here, real 3D organoids of PM12500-GFP tumor cells were also generated and incubated for 5 days. Then, their morphology was segmented and quantified from multiphoton microscopy images of the grown organoids. In this case, we obtained the number of cells to quantify the size of the organoids. Different-size organoids were chosen again to prove the adaptive response of the computational-based methodology to the data. In particular, we used one small sample (organoid #1), one intermediate sample (organoid #2), and two large samples (organoids #3 and #4) (Table 2). The simulations were initiated with one cell and run for 7 days to prove that the model is able to reach steady-state patterns. The parameters of the simulations can be found in Table S2.

We show images of the experimental organoids, the simulation of the organoids, and the coordination of cell functions (Figure 5), which represent the organoids in Table 2. The simulated organoids matched the target number of cells of the experimental data by day 5. Here, precise coordination between proliferative and quiescent cells is needed, because the organoids approximate the target number of cells (Figure 5B). In organoid #1, after 67 h, 7 cells continued proliferating, and 9 cells entered quiescence, reaching the target number of cells at 84 h. In organoid #2, most of the cells (93.75%) became quiescent at 84 h, whereas

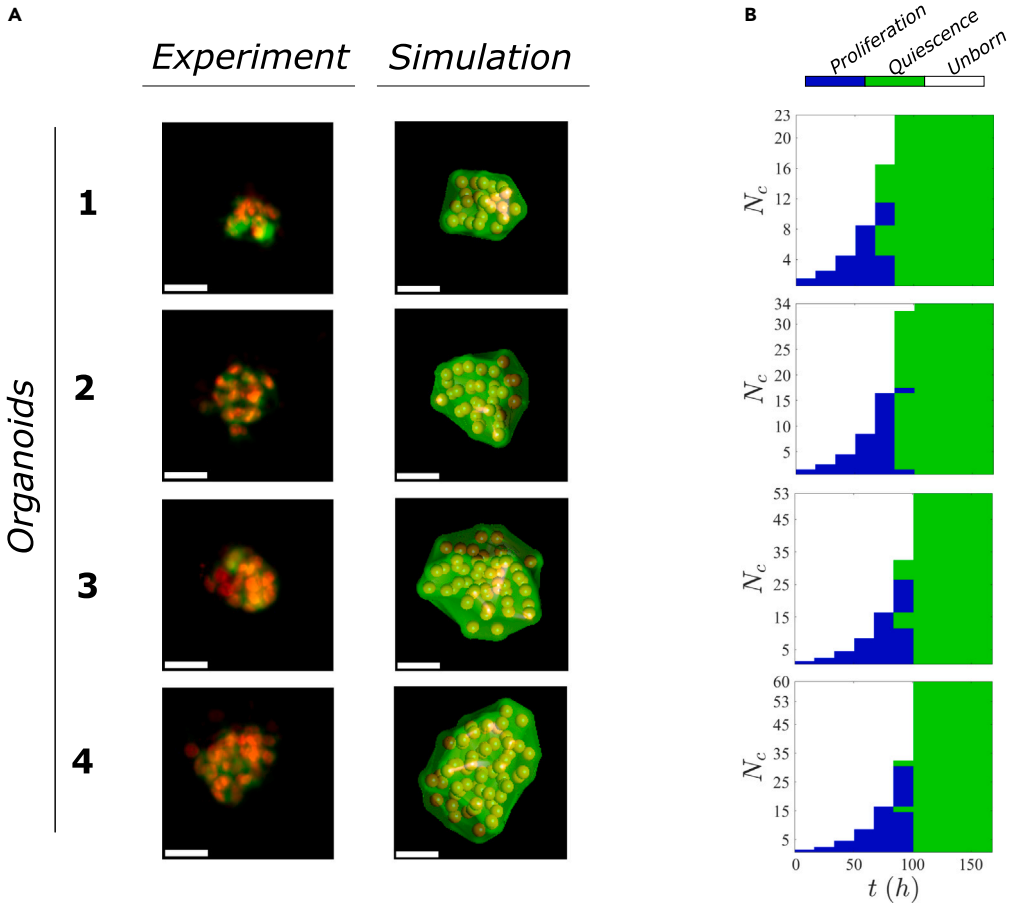


Figure 5. Simulation of experimental solid tumor organoids

(A) Image of *in vitro* organoids. Snapshots of the simulated organoids. The spheres represent the nuclei of cells with radius $R_c/2$ and the green hull is an estimation of the cell membrane through alpha shapes of the cells with $\alpha = 2R_c$. All scale bars are 30 μm.

(B) Coordination of cell functions, in which blue represents proliferation and green represents quiescence.

2 cells continued proliferating to finally form the target pattern at 101 h. In organoid #3. Twenty-two cells (68.75%) were proliferating after 84 h to increase the number of cells up to 53 at 101 h. Organoid #4 was the case in which more cells continued proliferating after 84 h, 28 out of 32 (87.5%).

Application of the framework to the evolution of solid organoids

To demonstrate not only the application of the methodology to other patterns but also the possibility to mimic the temporal evolution of patterns, we show an extrapolation of our methodology to simulate the evolution of solid organoids without lumen, which consists of dense spherical aggregates of cells (Figure S11). With this methodology, we aim to reproduce the growth of solid organoids so that the evolution of this organoid matches the target evolution. To do that, target data is specified at different time points ($\boldsymbol{t}^t$). In this case, the target data consisted of the number of cells ($\boldsymbol{N}_c^t$). In the agent-based framework, we consider cell mechanical interactions and the cell interactions with the extracellular matrix. Moreover, cells can either proliferate or stay in a quiescent state during a period of time (Δt_q). When a cell finishes the cell function that it is performing, the neural network decides the next cell function that the cell has to perform. Thus, we extract some simulation metrics (number of cells, number of proliferating cells, number of quiescent cells, and the simulation time) that characterizes the state of the simulation. Then, we obtain from the target time ($\boldsymbol{t}^t$) the target time immediately superior to the simulation time (t) and the number of cells ($\boldsymbol{N}_c^t$) corresponding to that target time. These data are transferred to the input layer of the neural network in the following manner: the first node receives the comparison between the target number of cells and the number of cells of the simulation $\left(\left(\boldsymbol{N}_c^t - N_c(t)\right)/\boldsymbol{N}_c^t\right)$; the second node receives the normalized number of

Table 3. Target data for the evolution of solid organoids

	Target number of cells (N_c^t)		
Organoid	Day 3	Day 5	Day 7
A	7	15	40
B	5	5	17
C	6	20	60

quiescent cells, and the third node receives the normalized number of proliferating cells. Finally, the neural network determines the next function to perform.

Simulation of the evolution of solid organoids

To show the possibilities of the methodology, we consider three synthetic cases of solid organoids with different temporal evolutions (Table 3). Organoid A represents a proliferative organoid. Case B represents an organoid that proliferates, then remains quiescence between day 3 and day 5, and proliferates again between day 5 and 7. Finally, case C corresponds to a proliferative organoid with a higher proliferation rate than case A. The parameters used for the simulation are collected in Table S3.

The simulations matched the evolution of the number of cells in the three cases (Figure 6), adapting to the changes in input data over time. In organoid A, as cells proliferate, they become quiescent whereas some cells still proliferate to match the target values (Figure 6A and Video S4). In the case of the organoid B, cells proliferate until day 3, and between day three and five they enter into a quiescent state to finally proliferate after day 5 (Figure 6B and Video S5). In organoid C, cells have a higher proliferative capacity than organoids A and B, so cells tend to proliferate against quiescence (Figure 6C and Video S6). Also, we simulated the evolution of a larger solid tumor organoid to demonstrate that the model works for any size. In this case, the target number of cells was established to 100 cells on day 7 and 500 cells on day 10 (Table S5). The simulation matched the evolution of the number of cells (Figure S12). Thus, this methodology is also able to mimic the temporal evolution of morphogenetic patterns of any size as long as the final result is feasible to achieve.

DISCUSSION

We present a novel hybrid physics-based and data-driven approach that combines agent-based modeling and deep learning for the simulation of developmental biology at the cellular level. Thus, we hypothesized that the different patterns that cells form are consequences of the distinct spatial and temporal coordination of their cellular functions. The following question is then how cells determine which cellular function to perform, since if we are able to replicate the decision-making of cells, we can reproduce the coordination of cell functions to achieve any pattern. Thus, we built a novel computational framework in which a neural network manages the decision-making of biological cell functions within a physics-based model to achieve an objective morphogenetic pattern. To accomplish the simulation of morphogenetic patterns, we do not fix the relationships between cell functions beforehand; instead, we consider cell functions as independent processes that cannot occur simultaneously, and the deep learning algorithm decides for each cell which cellular function they will be performing during the developmental process, properly coordinating all cellular functions for pattern formation. Thus, this artificial intelligence-based method learns the intrinsic mechanisms of the morphogenetic process and, therefore, is capable of reproducing similar morphogenetic patterns with different dynamical processes without optimizing the parameters *ad hoc* just by changing the target data about the pattern transferred to the input of the neural network. Therefore, with this approach, we can understand how cells must behave both spatially and temporarily to form different patterns.

In this work, we showed how to integrate a physics-based model with a deep learning algorithm to simulate morphogenetic patterns in 3D organoids. This modular framework offers a wide flexibility to change and adapt the physics-based and the data-driven models involved. On the one hand, in the physics-based part, agent-based modeling offers a broad range of types of models: from center-based models to deformable or vertex models.[10] This allows us to take advantage of a chosen model to better adapt to a specific application. On the other hand, regarding the data-driven part of the framework, different types of neural networks and activation functions can be considered. Indeed,

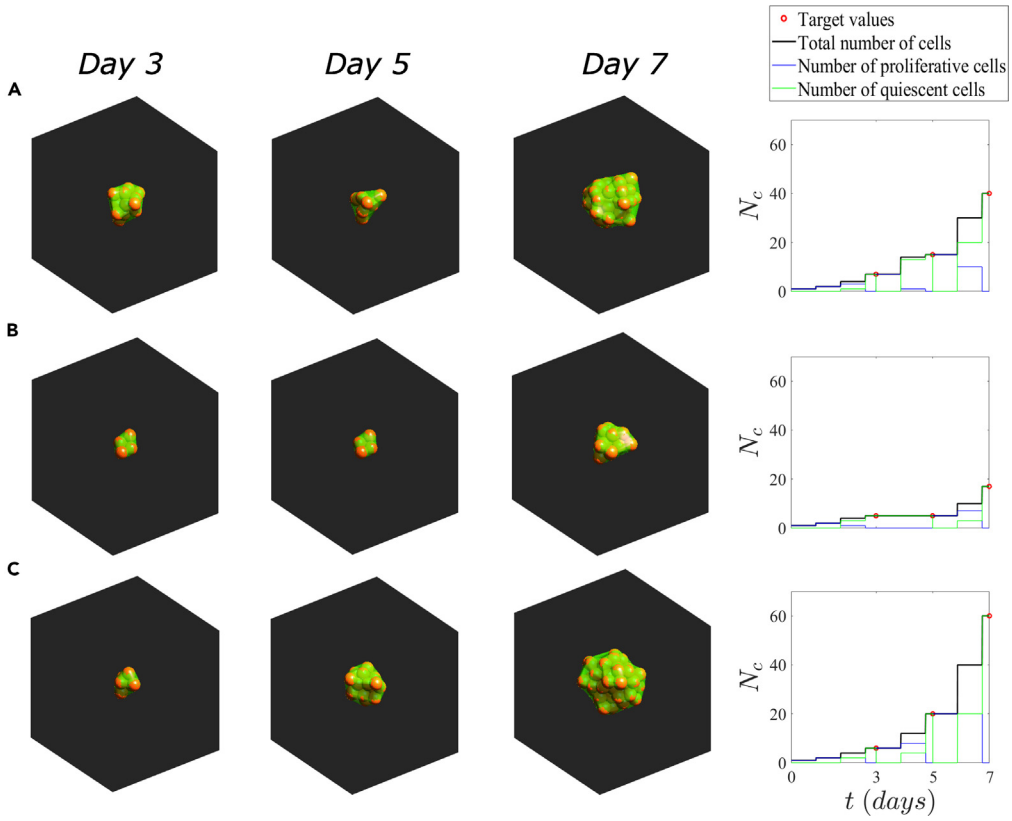


Figure 6. Simulation of the evolution of solid organoids

Snapshots at days 3, 5, and 7 of the temporal evolution of the solid organoids inserted in a cube of 200 μm side and the evolution of their number of cells. The spheres represent the full cell volume, and the green hull is an estimation of the cell membrane through alpha shapes of the cells with $\alpha = 2R_c$.

(A) Organoid A.

(B) Organoid B.

(C) Organoid C.

the architecture of the neural network could be simplified to reduce the computational cost of the learning process or incremented to unravel complex features, trying to find the optimal conditions for each case depending on the type of morphogenetic pattern, the quantity of input data, and the different cell states considered.[42] In addition, there exist different learning algorithms that can be integrated into this approach (genetic algorithm, gradient descent, conjugate gradient, Levenberg-Marquardt algorithm, among others[43]) for different computational speeds and memory requirements given a set number of parameters of the chosen neural network. Therefore, our framework is scalable with different types of models and permits modification to improve the computational cost and to better capture the specific biological process.

To evaluate the potential of the approach, we showed two applications mimicking *in vitro* data of the final pattern of organoids and one application reproducing synthetic time-lapse data. Thus, one advantage of this approach is that it requires little information about the morphogenetic processes, being able to undertake pattern formation using just final morphological information about the pattern extracted visually. In pattern formation, generally, the information is obtained from the observed final pattern (in general image data), and the efforts move to unravel how it was formed. In particular, time-lapse experiments could alter the viability of cell cultures by inducing cellular damage (e.g., imaging-induced phototoxicity, leading to cell-cycle arrest or cell death[44]). However, it is possible not only to obtain the coordination from the data of the final pattern but also to reproduce the temporal evolution of patterns. This permits us to mimic the exact evolution of the pattern formation when having time-lapse data, since it matches the known temporal constraints.

Regarding the simulation of *in vitro* data of lumen morphogenesis in PDAC organoids, physiologically, the generation of a cystic organoid requires the orchestration of proliferation, fluid secretion, and quiescence. On the one hand, if cells secrete fluid too early, the preapical patch, which is the closed volume necessary to secrete fluid, will not be formed, resulting in fluid leakage. On the other hand, if cells secrete too much fluid, the hydrostatic pressure excessively increases, and the lumen opens. In addition, if cells proliferate too much, they will not fit in the monolayer around the lumen, and cystic structures will not be formed. In contrast, if cells do not proliferate, they will not create enough inner space for the fluid, resulting in leakage. Only if cell functions are coordinated in a sophisticated manner will cystic structures form. In particular, first cells must proliferate to create an inner volume and then secrete fluid while continuing to proliferate to create more inner space as the fluid increases, before finally entering into a quiescent state. Therefore, the success of lumen generation depends on the proper coordination between cell proliferation, cell secretion, and quiescence. In fact, what the neural network learns is that if it generates too much pressure through fluid secretion with an insufficient number of cells, it leaks. In addition, it learns that cells must first proliferate to generate the organoid and, as the organoid reaches its final size, enter into a quiescent state. Despite the great biological variability in terms of the final size of *in vitro* experiments (Figure S13), we were able to obtain the developmental process to finally predict their size in each case. This quantitative validation with *in vitro* data proves that this approach does not produce *ad hoc* solutions and confirms that the different coordination of processes determines the final pattern. This is very suitable to simulate biological systems, which display variability and fluctuations around a solution. In the case of solid tumor organoids, the neural network coordinates proliferation and quiescence to achieve the final size of the organoid. This regulation of the dynamics of the population of cells is a helpful tool that could be used in many types of simulations. This permits the integration of data into simulations to determine cell proliferation, which is sometimes a difficult task, especially when the dynamics of the population of cells change over time. Therefore, our model is designed to predict the order in which cells behave to achieve a specific pattern.

Moreover, from a biological perspective, we found in our simulations that some cells are predisposed to perform a specific function over others. This is the case for the initial cell or first cells, which tend to proliferate more than daughter cells (Figure 4B). These initial precursor cells might have a higher proliferative potential, and as they divide, the daughter cells are more specialized with a higher differentiation grade. Inside an organoid, there exists heterogeneity of cellular functions, and we hypothesize that the cellular functions that cells perform in a morphogenetic process are not governed by the maturation of cells but by the stimuli to which they are exposed when they determine which function to perform. This implies that some cells are more specialized in division and others in secretion within the organoid. We also found that the final size of the organoid depends on the morphological characteristics of the organoid when the lumen is initiated. In our simulations, the initiation of the lumen, determined by the beginning of the fluid secretion, was different for each case, being earlier the smaller the final size of the organoid was (Figure 4B). We were able to establish that, initially, cells have a less differentiated phenotype and have a higher proliferative capacity. Then, as they proliferate, the daughter cells adopt a higher differentiation grade and specialize into secretory cells because of the stimuli that they are exposed when they are born. Thus, when they start secreting, the proliferative capacity of the organoid is reduced, so the final size of the organoid is limited by its size at the lumen initiation time. Therefore, we hypothesize that the proliferation capacity of cells may be hindered by the initiation of lumen formation and that the final size can be predicted at the initiation of lumen formation. These hypotheses could be validated by performing cell culture immunostaining with a mitotic cell marker to determine the number of proliferative cells in the biological system.[45]

Another potential of this approach is that once the neural network has learned how to orchestrate cells to form the morphogenetic pattern, the parameters of the physics-based model can be modified to study the morphogenetic process under different scenarios. This allows the simulation of morphogenesis with different extracellular matrix properties, cell cycle times, or even organ-specific morphogenesis. In the case of morphogenesis, similar patterns are found in many parts of the organism, but the dynamic process differs in each organ since cell cycle time phases vary considerably between different types of cells and secretory cells may require different amounts of time to produce fluid. This implies that similar morphogenetic patterns can result in different sizes and require different lengths of time to achieve the final shape. However, although morphogenesis depends on cell type and the dynamics of the process vary, the requirements of the morphogenetic process are conserved among similar patterns. To this aim, we conducted a sensitivity analysis (Figures S7 and S8) varying the cell cycle time and fluid production time for the

pancreatic cystic tumor organoids. This variation in these two parameters represents either the fluctuation of the behavior between the same type of cells or other types of cells that can also form cystic organoids. We also performed an analysis of the influence of the introduction of noise in the cell cycle and fluid secretion parameters (Figure S11), showing that the model is robust in forming the patterns and that the neural network can orchestrate cells despite the fluctuations of these parameters. With these analyses, we showed that in addition to the alteration of the dynamical processes of cells, the coordination to form the same pattern with the same morphological features (in this case, the number of cells and lumen volume) is conserved.

The potential of the proposed framework combining physics-based and data-driven modeling opens the door to a novel way of performing biological simulations at a cellular level, understanding how cell functions must coordinate to generate morphogenetic patterns by replicating how cells make decisions and adapting dynamically their responses to the variability found in biology.

Limitations of the study

In the applications shown in this work, some simplifications have been assumed. First, we employed a center-based lattice-free model to simulate the formation and growth of organoids, which entails some simplifications. The most important one is that cells are assumed to be nondeformable spheres, and their shape was not represented accurately. This simplification allowed us to approximate the size of the organoids through their number of nuclei with a reduced computational cost. Since the deformation of cells is not simulated, there are differences between the volume of the simulated organoids and the experiments (Figure 5A). However, this simplification does not influence the coordination that cells must follow to form morphogenetic patterns. Here, we did not aim to reproduce the exact shape and volume of the organoids; instead, we focused on showing the potential of the framework to unravel the orchestration of cell functions from a morphogenetic pattern, which is independent of the cell shape and volume. Second, we represented the lumen fluid through a particle-based model comparable to other works.[40,46] Although this representation is a simplification, we showed that our approach accurately reproduces the expected velocity and shear stress profiles for a Newtonian fluid, similar to water (see the characterization of the lumen fluid in STAR Methods). This confirms the capability of our method to model real fluids, serving as a minimal model that aims to recapitulate certain features such as mass conservation and momentum exchange. However, it does not rigorously reproduce bulk-scale physical properties. Furthermore, our approach enables us to simulate the interaction between cells and the fluid during a morphogenetic process, in which the lumen initiates *de novo* and grows through a fluid deposition process by cells, thus evolving the interface between cells and fluid over time. In addition, we investigated the influence of particle radius on the rheological properties of the fluid (further details are included in particle radius influence on the rheological properties of the fluid in STAR Methods). We have justified the choice of the particle radius and demonstrated that the rheological properties of the fluid play a crucial role in cystic organoid morphogenesis (see particle radius influence on fitness in STAR Methods).

Another limitation of the approach relates to the selection of cell functions and the question of whether the hypothesized functions are sufficient to generate the specific pattern. In our model, the decision-making possibilities are limited to the cell functions considered, which in this work were proliferation, secretion, and quiescence. Indeed, there are many cellular functions and processes that could be simulated. However, including more cell functions does not guarantee that all of them might be activated to reach a specific morphogenetic pattern. This allows us to underestimate them and reduce the complexity of the neural network and the number of outputs. Furthermore, in this first work, we created this data-driven approach with a neural network that receives the quantitative data as inputs. However, the neural network aims to reproduce the cell signaling through which cells sense, receive, and transduce external factors from their microenvironment (such as pressure, chemical signals, nutrients, etc.), or even internal factors (such as cell damage and cell deformation). Therefore, this approach permits the introduction of those factors as inputs of the neural network to enrich cell signaling and response. The final concern is related to finding the optimal neural network architecture for each specific application. Although some rules help to select the number of nodes and hidden layers, trial and error are commonly used. The neural network architecture determines the capacity to solve the problem and must be appropriate for the application. Moreover, the choice of the fitness function is fundamental to make the neural network learn. The fitness function evaluates the neural network performance depending on how close the simulated pattern is compared to the objective pattern. Therefore, this interplay between the neural network and the fitness function regulates

the learning of the morphogenetic pattern. Despite the simplifications made in the applications of this framework, it allowed us to reproduce the morphogenetic patterns of organoids, understating how cells orchestrate their cellular functions to achieve those patterns.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Murine PDAC cells culture and maintenance
  - 3D cultures of murine PDAC cells in matrigel
- METHOD DETAILS
  - Learning algorithm
  - Selection
  - Crossover
  - Mutation
  - Agent-based modeling
  - Cell functions
  - Cell division
  - Fitness function
  - Eccentricity
  - Characterization of the lumen fluid
  - Particle radius influence on the rheological properties of the fluid
  - Particle radius influence on fitness
  - Multiphoton microscopy of PDAC organoids
  - Quantification of 3D PDAC organoid morphology
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107164.

ACKNOWLEDGMENTS

This research was funded by MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe (Grant No. RTI2018-094494-B-C21, I.S.B., C.O.S.A., Grant No. PID2021-122409OB-C21 J.M.G.A., Grant No. PID2021-124271OB-I00 M.J.G.B.), the European Research Council (ICoMICS Adv grant agreement: 101018587, J.M.G.A., M.J.G.B.), and MCIN/AEI/10.13039/501100011033 and Next Generation EU (ProCanAid Grant No. PLEC2021-007709, D.C.G., J.M.G.A., M.J.G.B.).

AUTHOR CONTRIBUTIONS

D.C.G. conceived and designed the study, implemented the framework and performed all coding, performed the simulations and analysis, prepared the figures, and wrote the original manuscript. I.S.B. carried out the experimental work, wrote the experimental methods, contributed with experimental figures, and reviewed the manuscript. C.O.S.A. supervised the experimental work, reviewed the manuscript, and obtained the resources. J.M.G.A. conceived and designed the study, supervised the computational work, contributed to the figures and manuscript preparation, and obtained the resources. M.J.G.B. conceived and designed the study, supervised the computational work, contributed to the figures and manuscript preparation, and obtained the resources. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 14, 2022
Revised: March 30, 2023
Accepted: June 13, 2023
Published: June 19, 2023

REFERENCES

1. Lancaster, M.A., and Knoblich, J.A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. Science *345*, 1247125. https://doi.org/10.1126/science.1247125.

2. Blutt, S.E., and Estes, M.K. (2022). Organoid Models for Infectious Disease. Annu. Rev. Med. *73*, 167–182. https://doi.org/10.1146/annurev-med-042320-023055.

3. Sahu, S., and Sharan, S.K. (2020). Translating embryogenesis to generate organoids: Novel approaches to personalized medicine. iScience *23*, 101485. https://doi.org/10.1016/j.isci .2020.101485.

4. Dart, A. (2018). Organoid diversity. Nat. Rev. Cancer *18*, 404–405. https://doi.org/10.1038/s41568-018-0018-3.

5. Tuveson, D., and Clevers, H. (2019). Cancer modeling meets human organoid technology. Science *364*, 952–955. https://doi.org/10.1126/science.aaw6985.

6. Ishiguro, T., Ohata, H., Sato, A., Yamawaki, K., Enomoto, T., and Okamoto, K. (2017). Tumor-derived spheroids: relevance to cancer stem cells and clinical applications. Cancer Sci. *108*, 283–289. https://doi.org/10.1111/cas.13155.

7. Hofer, M., and Lutolf, M.P. (2021). Engineering organoids. Nat. Rev. Mater. *6*, 402–420. https://doi.org/10.1038/s41578-021-00279-y.

8. Gierer, A., and Meinhardt, H. (1972). A theory of biological pattern formation. Kybernetik *12*, 30–39. https://doi.org/10.1007/BF00289234.

9. Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. *25*, 1–47. https://doi.org/10.1016/S0022-5193(69)80016-0.

10. 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. Comput. Part. Mech. *2*, 401–444. https://doi.org/10.1007/s40571-015-0082-3.

11. Engelberg, J.A., Datta, A., Mostov, K.E., and Hunt, C.A. (2011). MDCK cystogenesis driven by cell stabilization within computational analogues. PLoS Comput. Biol. *7*, e1002030. https://doi.org/10.1371/journal.pcbi.1002030.

12. Palsson, E. (2008). A 3-D model used to explore how cell adhesion and stiffness Palsson, E. J. Theor. Biol. *254*, 1–13. https://doi.org/10.1016/j.jtbi.2008.05.004.

13. Sepúlveda, N., Petitjean, L., Cochet, O., Grasland-Mongrain, E., Silberzan, P., and Hakim, V. (2013). Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model. PLoS Comput. Biol. *9*, e1002944. https://doi.org/10.1371/journal.pcbi.1002944.

14. Van Liedekerke, P., Ghysels, P., Tijskens, E., Samaey, G., Roose, D., and Ramon, H. (2011). Mechanisms of soft cellular tissue bruising. A particle based simulation approach. Soft Matter *7*, 3580–3591. https://doi.org/10.1039/C0SM01261K.

15. Okuda, S., Inoue, Y., Eiraku, M., Adachi, T., and Sasai, Y. (2015). Vertex dynamics simulations of viscosity-dependent deformation during tissue morphogenesis. Biomech. Model. Mechanobiol. *14*, 413–425. https://doi.org/10.1007/s10237-014-0613-5.

16. Drasdo, D., and Höhme, S. (2005). A single-cell-based model of tumor growth in vitro: monolayers and spheroids. Phys. Biol. *2*, 133–147. https://doi.org/10.1088/1478-3975/2/3/001.

17. Gonçalves, I.G., Garcia-Aznar, J.M., and Garcia-Aznar, J.M. (2021). Extracellular matrix density regulates the formation of tumour spheroids through cell migration. PLoS Comput. Biol. *17*, e1008764. https://doi.org/10.1371/journal.pcbi.1008764.

18. González-Valverde, I., and García-Aznar, J.M. (2018). Mechanical modeling of collective cell migration: an agent-based and continuum material approach. Comput. Methods Appl. Mech. Eng. *337*, 246–262. https://doi.org/10.1016/j.cma.2018.03.036.

19. Van Liedekerke, P., Neitsch, J., Johann, T., Warmt, E., Gonzàlez-Valverde, I., Hoehme, S., Grosser, S., Kaes, J., and Drasdo, D. (2020). A quantitative high resolution computational mechanics cell model for growing and regenerating tissues. Biomech. Model. Mechanobiol. *19*, 189–220. https://doi.org/10.1007/s10237-019-01204-7.

20. Thorne, B.C., Bailey, A.M., DeSimone, D.W., and Peirce, S.M. (2007). Agent based modeling of multicell morphogenic processes during development. Birth Defects Res. C Embryo Today. *81*, 344–353. https://doi.org/10.1002/bdrc.20106.

21. Glen, C.M., Kemp, M.L., and Voit, E.O. (2019). Agent-based modeling of morphogenetic systems: Advantages and challenges. PLoS Comput. Biol. *15*, e1006577. https://doi.org/10.1371/journal.pcbi.1006577.

22. Bergman, D., Sweis, R.F., Pearson, A.T., Nazari, F., and Jackson, T.L. (2022). A global method for fast simulations of molecular dynamics in multiscale agent based models of biological tissues. iScience, 104387. https://doi.org/10.1016/j.isci.2022.104387.

23. Montáns, F.J., Chinesta, F., Gómez-Bombarelli, R., and Kutz, J.N. (2019). Data-driven modeling and learning in science and engineering. Compt. Rendus Mec. *347*, 845–855. https://doi.org/10.1016/j.crme.2019.11.009.

24. An, D., Kim, N.H., and Choi, J.-H. (2015). Practical options for selecting data driven or physics-based prognostics algorithms with reviews. Reliab. Eng. Syst. Saf. *133*, 223–236. https://doi.org/10.1016/j.ress.2014.09.014.

25. Uzkudun, M., Marcon, L., and Sharpe, J. (2015). Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud. Mol. Syst. Biol. *11*, 815. https://doi.org/10.15252/msb.20145882.

26. Lambert, B., MacLean, A.L., Fletcher, A.G., Combes, A.N., Little, M.H., and Byrne, H.M. (2018). Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. J. Math. Biol. *76*, 1673–1697. https://doi.org/10.1007/s00285-018-1208-z.

27. Lusch, B., Kutz, J.N., and Brunton, S.L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. Nat. Commun. *9*, 4950–5010. https://doi.org/10.1038/s41467-018-07210-0.

28. Takeda, H., Kameo, Y., and Adachi, T. (2021). Continuum modeling for neuronal lamination during cerebral morphogenesis considering cell migration and tissue growth. Comput. Methods Biomech. Biomed. Eng. *24*, 799–805. https://doi.org/10.1080/10255842.2020.1852554.

29. Stein, A.M., Demuth, T., Mobley, D., Berens, M., and Sander, L.M. (2007). A mathematical model of glioblastoma tumor spheroid invasion in a threedimensional in vitro experiment. Biophys. J. *92*, 356–365. https://doi.org/10.1529/biophysj.106.093468.

30. Sharpe, J. (2017). Computer modeling in developmental biology: growing today, essential tomorrow. Development *144*, 4214–4225. https://doi.org/10.1242/dev.151274.

31. Tripathi, S., Kessler, D.A., and Levine, H. (2020). Biological networks regulating cell fate choice are minimally frustrated. Phys. Rev. Lett. *125*, 088101. https://doi.org/10.1103/PhysRevLett.125.088101.

32. Tripathi, S., Levine, H., and Jolly, M.K. (2020). The physics of cellular decision making during epithelial–mesenchymal transition. Annu. Rev. Biophys. *49*, 1–18. https://doi.org/10.1146/annurev-biophys-121219-081557.

33. Datta, A., Bryant, D.M., and Mostov, K.E. (2011). Molecular regulation of lumen morphogenesis. Curr. Biol. *21*, R126–R136. https://doi.org/10.1016/j.cub.2010.12.003.

34. Latorre, E., Kale, S., Casares, L., Gómez-González, M., Uroz, M., Valon, L., Nair, R.V., Garreta, E., Montserrat, N., Del Campo, A., et al. (2018). Active superelasticity in three-dimensional epithelia of controlled shape. Nature *563*, 203–208. https://doi.org/10.1038/s41586-018-0671-4.

35. Navis, A., and Bagnat, M. (2015). Developing pressures: fluid forces driving morphogenesis. Curr. Opin. Genet. Dev. *32*, 24–30. https://doi.org/10.1016/j.gde.2015.01.010.

36. Dasgupta, S., Gupta, K., Zhang, Y., Viasnoff, V., and Prost, J. (2018). Physics of lumen growth. Proc. Natl. Acad. Sci. USA *115*, E4751–E4757. https://doi.org/10.1073/pnas.1722154115.

37. Duclut, C., Sarkar, N., Prost, J., and Jülicher, F. (2019). Fluid pumping and active flexoelectricity can promote lumen nucleation in cell assemblies. Proc. Natl. Acad. Sci. USA *116*, 19264–19273. https://doi.org/10.1073/pnas.1908481116.

38. Paszek, M.J., Zahir, N., Johnson, K.R., Lakins, J.N., Rozenberg, G.I., Gefen, A., Reinhart-King, C.A., Margulies, S.S., Dembo, M., Boettiger, D., et al. (2005). Tensional homeostasis and the malignant phenotype. Cancer Cell *8*, 241–254. https://doi.org/10.1016/j.ccr.2005.08.010.

39. Checa, S., Rausch, M.K., Petersen, A., Kuhl, E., and Duda, G.N. (2015). The emergence of extracellular matrix mechanics and cell traction forces as important regulators of cellular self-organization. Biomech. Model. Mechanobiol. *14*, 1–13. https://doi.org/10.1007/s10237-014-0581-9.

40. Camacho-Gómez, D., Garćıa-Aznar, J.M., and Gómez-Benito, M.J. (2022). A 3D multi-agent-based model for lumen morphogenesis: the role of the biophysical properties of the extracellular matrix. Eng. Comput. 1–15. https://doi.org/10.1007/s00366-022-01654-1.

41. Edelsbrunner, H., and Mücke, E.P. (1994). Three-dimensional alpha shapes. ACM Trans. Graph. *13*, 43–72. https://doi.org/10.1145/174462.156635.

42. Wilamowski, B. (2009). Neural network architectures and learning algorithms. EEE. Ind. Electron. Mag. *3*, 56–63. https://doi.org/10.1109/MIE.2009.934790.

43. Ray, S. (2019). A quick review of machine learning algorithms. In International conference on machine learning, big data, cloud and parallel computing (COMITCon) (IEEE.), pp. 35–39. https://doi.org/10.1109/COMITCon.2019.8862451.

44. Fei, K., Zhang, J., Yuan, J., and Xiao, P. (2022). Present Application and Perspectives of Organoid Imaging Technology. Bioengineering *9*, 121. https://doi.org/10.3390/bioengineering9030121.

45. Yu, C.C., Woods, A.L., and Levison, D.A. (1992). The assessment of cellular proliferation by immunohistochemistry: a review of currently available methods and their applications. Histochem. J. *24*, 121–131. https://doi.org/10.1007/BF01047461.

46. Liedekerke, P. van, Gannoun, L., Loriot, A., Lemaigre, F., and Drasdo, D. (2021). Influence of cell mechanics in embryonic bile duct lumen formation: insight from quantitative modeling. https://hal.inria.fr/hal-03135722.

47. Boj, S.F., Hwang, C.-I., Baker, L.A., Chio, I.I.C., Engle, D.D., Corbo, V., Jager, M., Ponz-Sarvise, M., Tiriac, H., Spector, M.S., et al. (2015). Organoid models of human and mouse ductal pancreatic cancer. Cell *160*, 324–338. https://doi.org/10.1016/j.cell.2014.12.021.

48. Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (2020). Star-convex polyhedra for 3d object detection and segmentation in microscopy. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3666–3673. https://doi.org/10.1109/WACV45572.2020.9093435.

49. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

50. Evans, E.A. (1985). Detailed mechanics of membrane-membrane adhesion and separation. II. Discrete kinetically trapped molecular cross-bridges. Biophys. J. *48*, 185–192. https://doi.org/10.1016/S0006-3495(85)83771-1.

51. Zheng, Z., Zhu, H., Wan, Q., Liu, J., Xiao, Z., Siderovski, D.P., and Du, Q. (2010). LGN regulates mitotic spindle orientation during epithelial morphogenesis. J. Cell Biol. *189*, 275–288. https://doi.org/10.1083/jcb.200910021.

52. Martin-Belmonte, F., Gassama, A., Datta, A., Yu, W., Rescher, U., Gerke, V., and Mostov, K. (2007). PTEN-mediated apical segregation of phosphoinositides controls epithelial morphogenesis through Cdc42. Cell *128*, 383–397. https://doi.org/10.1016/j.cell.2006.11.051.

53. Jaffe, A.B., Kaji, N., Durgan, J., and Hall, A. (2008). Cdc42 controls spindle orientation to position the apical surface during epithelial morphogenesis. J. Cell Biol. *183*, 625–633. https://doi.org/10.1083/jcb.200807121.

54. Qin, Y., Meisen, W.H., Hao, Y., and Macara, I.G. (2010). Tuba, a Cdc42 GEF, is required for polarized spindle orientation during epithelial cyst formation. J. Cell Biol. *189*, 661–669. https://doi.org/10.1083/jcb.201002097.

55. Rodriguez-Fraticelli, A.E., Vergarajauregui, S., Eastburn, D.J., Datta, A., Alonso, M.A., Mostov, K., and Martín-Belmonte, F. (2010). The Cdc42 GEF Intersectin 2 controls mitotic spindle orientation to form the lumen during epithelial morphogenesis. J. Cell Biol. *189*, 725–738. https://doi.org/10.1083/jcb .201002047.

56. Hao, Y., Du, Q., Chen, X., Zheng, Z., Balsbaugh, J.L., Maitra, S., Shabanowitz, J., Hunt, D.F., and Macara, I.G. (2010). Par3 controls epithelial spindle orientation by aPKC-mediated phosphorylation of apical Pins. Curr. Biol. *20*, 1809–1818. https://doi.org/10.1016/j.cub.2010.09.032.

57. Ghaffarizadeh, A., Heiland, R., Friedman, S.H., Mumenthaler, S.M., and Macklin, P. (2018). PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLoS Comput. Biol. *14*, e1005991. https://doi.org/10.1371/journal.pcbi.1005991.

58. Suveges, S., Chamseddine, I., Rejniak, K.A., Eftimie, R., and Trucu, D. (2021). Collective cell migration in a fibrous environment: a hybrid multi-scale modelling approach. Front. Appl. Math. Stat. *7*, 680029. https://doi.org/10.3389/fams.2021.680029.

59. Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K.W., Schindelin, J., Cardona, A., and Sebastian Seung, H. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics *33*, 2424–2426. https://doi.org/10.1093/bioinformatics/btx180.

60. Bloice, M.D., Stocker, C., and Holzinger, A. (2017). Augmentor: an image augmentation library for machine learning. Preprint at: arXiv. https://doi.org/10.48550/arXiv.1708.04680.

61. Haase, R., Royer, L.A., Steinbach, P., Schmidt, D., Dibrov, A., Schmidt, U., Weigert, M., Maghelli, N., Tomancak, P., Jug, F., and Myers, E.W. (2020). CLIJ: GPU accelerated image processing for everyone. Nat. Methods *17*, 5–6. https://doi.org/10.1038/s41592-019-0650-1.

62. Legland, D., Silva, J.V., Cauty, C., Irina, K., and Floury, J. (2016). Quantitative image analysis of binary microstructures: Application to the characterisation of dairy systems. In European Microscopy Congress 2016: Proceedings (Wiley Online Library), pp. 591–592. https://doi.org/10.1002/9783527808465.EMC2016.6806.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Advanced DMEM F12	Gibco	12634-010
HEPES	Lonza	17-737E
GlutaMax	Gibco	35050-038
Penicillin/Streptomycin	Gibco	15140-122
A83-01	TOCRIS	2939
EGF	Life Technologies	PMG8041
FGF	Peprotech	100-26
Gastrin I	TOCRIS	3006
Noggin	Peprotech	250-38
Y-27632	Sigma	Y0503
N-acetylcysteine	Sigma	A9165
Nicotinamide	Sigma	N0636
B-27	Life Technologies	17504-044
Experimental models: Cell lines		
Mouse ductal pancreatic cancer organoids	Boj, et al.[47]	https://doi.org/10.1016/j.cell.2014.12.021
Deposited data		
Original data and code	Github, Zenodo	https://github.com/daniel-camacho-gomez/AIOrganoids https://doi.org/10.5281/zenodo.8019619
Software and algorithms		
StarDist 3D	Weigert, et al.[48]	https://github.com/stardist/stardistl
ImageJ	Schneider, et al.[49]	https://imagej.nih.gov/ij/
MATLAB	MathWorks	https://es.mathworks.com/products/matlab.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Maria Jose Gomez-Benito (gomezmj@unizar.es).

Materials availability

This study did not generate unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- Original codes are available on request to the corresponding authors and are also publicly available at https://github.com/daniel-camacho-gomez/AIOrganoids. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Murine PDAC cells culture and maintenance

Cystic organoids formed from PDAC93-GFP and solid organoids formed from PM12500-GFP pancreatic ductal adenocarcinoma (PDAC) cells derived from Pdx1-Cre; Kras+/LSL-G12D; Trp53+/LSL-R172H male

mice (KPC) were kindly donated by Dr. Mariano Ponz and Dr. Silvestre Vicent. Authentication of cell lines was not required. The cell line used to generate the organoids was a primary line donated by a collaborator, generated from a transgenic mouse model of pancreatic cancer. It is not a commercial line, thus authentication is not possible. The organoid cell line was tested for mycoplasma contamination using MycoAlert® Mycoplasma Detection Kit (Lonza). Cells were thawed from frozen stock in 50 ml Falcon tubes containing 10 ml of DMEM and centrifuged to remove any traces of the cryoprotectant DMSO. Then, the cells were incubated in T75 flasks (TC treated; Nunc EasyFlask, Thermo Scientific) containing 10 ml of DMEM supplemented with 10% FetalClone III (SH30109.02, Cytiva) at 37 °C and 5% CO2. Once cells reached 90% confluence, cell cultures were passed to a new T75 flask using 0.05% Trypsin-EDTA (25300-096, Gibco) followed by centrifugation and resuspension in 10 ml of fresh DMEM supplemented with serum and subsequent incubation at 37 °C and 5% CO2.

3D cultures of murine PDAC cells in matrigel

PDAC93-GFP and PM12500-GFP cells were grown in T75 flasks (TC treated; Nunc EasyFlask, Thermo Scientific) containing 10 ml of DMEM supplemented with 10% FetalClone III (SH30109.02, Cytiva). Once the cells reached 90% confluence, they were detached using 0.05% trypsin-EDTA (25300-096, Gibco), centrifuged, and resuspended in 1 *ml* of fresh DMEM supplemented with serum. Then, the cell suspensions were embedded in 4 mg/ml growth factor reduced (GFR) Matrigel at a final concentration of 300,000 cells/ml. A total of 20 μl of this cell-hydrogel mixture was added to individual 5 mm diameter wells of a homemade PDMS device followed by a 15 min incubation at 37 °C to complete the Matrigel gelation process. Finally, 800 μl of organoid feeding media (a detailed description of the cell culture media used is available in Table S4) was added to the organoid culture device, and the 3D cell culture was incubated at 37 °C and 5% CO_2 as needed.

METHOD DETAILS

Learning algorithm

In order to make the neural network learn, we employed a genetic algorithm based on selection, crossover, and mutation algorithms. For the sake of simplicity, we illustrate how the weights of one node of the neural network are updated.

Selection

selection is how the progenitors are chosen from the generation. Thus, the two fittest progenitors of the actual generation are chosen as parents of the next generation. Let A be the best of its generation and B the second best. These two are also selected as offspring (parent A = off.1 and parent B = off.2). Then, the rest of the offspring (off.3, off.4, …, off.10) are obtained through crossover and mutation algorithms.

Crossover

parent A and parent B are mixed to obtain the offspring. Four-different crossover algorithms are implemented:

single point crossover: beginning with one parent ending with other:

$$\overbrace{\begin{bmatrix}\textcircled{1}\\ \textcircled{2}\\ 3\\ 4\end{bmatrix}}^{\text{Parent A}} + \overbrace{\begin{bmatrix}5\\ 6\\ \textcircled{7}\\ \textcircled{8}\end{bmatrix}}^{\text{Parent B}} = \overbrace{\begin{bmatrix}1\\ 2\\ 7\\ 8\end{bmatrix}}^{\text{Off. 3}} \qquad \overbrace{\begin{bmatrix}1\\ 2\\ \textcircled{3}\\ \textcircled{4}\end{bmatrix}}^{\text{Parent A}} + \overbrace{\begin{bmatrix}\textcircled{5}\\ \textcircled{6}\\ 7\\ 8\end{bmatrix}}^{\text{Parent B}} = \overbrace{\begin{bmatrix}5\\ 6\\ 3\\ 4\end{bmatrix}}^{\text{Off. 4}}$$

two point crossover: first and final values are chosen from one parent and middle values from the other parent:

$$\overbrace{\begin{bmatrix}\textcircled{1}\\ 2\\ 3\\ \textcircled{4}\end{bmatrix}}^{\text{Parent A}} + \overbrace{\begin{bmatrix}5\\ \textcircled{6}\\ \textcircled{7}\\ 8\end{bmatrix}}^{\text{Parent B}} = \overbrace{\begin{bmatrix}1\\ 6\\ 7\\ 4\end{bmatrix}}^{\text{Off. 5}} \qquad \overbrace{\begin{bmatrix}1\\ \textcircled{2}\\ \textcircled{3}\\ 4\end{bmatrix}}^{\text{Parent A}} + \overbrace{\begin{bmatrix}\textcircled{5}\\ 6\\ 7\\ \textcircled{8}\end{bmatrix}}^{\text{Parent B}} = \overbrace{\begin{bmatrix}5\\ 2\\ 3\\ 8\end{bmatrix}}^{\text{Off. 6}}$$

arithmetic crossover: sum of parents:

$$\overbrace{\begin{bmatrix}1\\2\\3\\4\end{bmatrix}}^{Parent\ A} + \overbrace{\begin{bmatrix}5\\6\\7\\8\end{bmatrix}}^{Parent\ B} = \overbrace{\begin{bmatrix}6\\8\\10\\12\end{bmatrix}}^{Off.\ 7}$$

uniform crossover: offspring values are randomly copied from A or B parent:

$$\overbrace{\begin{bmatrix}1\\2\\③\\4\end{bmatrix}}^{Parent\ A} + \overbrace{\begin{bmatrix}⑤\\⑥\\7\\⑧\end{bmatrix}}^{Parent\ B} = \overbrace{\begin{bmatrix}5\\6\\3\\8\end{bmatrix}}^{Off.\ 8}$$

Mutation

The purpose of mutation is to introduce diversity into the sampled population, avoid local minima and get faster to the solution ($rand(-1,1)$ is a function that chooses the value -1 or 1 and mut_{rate} is a parameter):

$$off.\ 9 = \overbrace{\begin{bmatrix}1\\2\\3\\4\end{bmatrix}}^{Parent\ A} + mut_{rate} \cdot [\,rand(-1,1) \quad rand(-1,1) \quad rand(-1,1) \quad rand(-1,1)\,] \cdot \overbrace{\begin{bmatrix}1\\2\\3\\4\end{bmatrix}}^{Parent\ A} \quad \text{(Equation 1)}$$

$$off.\ 10 = \overbrace{\begin{bmatrix}5\\6\\7\\8\end{bmatrix}}^{Parent\ B} + mut_{rate} \cdot [\,rand(-1,1) \quad rand(-1,1) \quad rand(-1,1) \quad rand(-1,1)\,] \cdot \overbrace{\begin{bmatrix}5\\6\\7\\8\end{bmatrix}}^{Parent\ B} \quad \text{(Equation 2)}$$

Agent-based modeling

Mechanical interactions between agents make them move and change their positions. To describe these interactions we follow our previous work.[40] We illustrate how the cell position $\boldsymbol{x}_c$ and particle position $\boldsymbol{x}_p$ are calculated. Let $\mathbf{N}_c$ be the set of cells $\mathbf{N}_c = \{1,\ldots,N_c\}$, and let $\mathbf{N}_p$ be the set of particles $\mathbf{N}_p = \{1,\ldots,N_p\}$. First, the velocity of each i-cell $\boldsymbol{v}_{c_i}$ and each k-particle $\boldsymbol{v}_{p_k}$ are calculated from the balance of forces:

$$m_{c_i}\frac{d\boldsymbol{v}_{c_i}}{dt} = \sum_{j\in \mathbf{N}_c}\left(\boldsymbol{F}_{c_ic_j}\right) + \sum_{j\in \mathbf{N}_p}\left(\boldsymbol{F}_{c_ip_j}\right) + \boldsymbol{F}_{c_{i_{drag}}} \approx 0, \quad \text{(Equation 3)}$$

$$m_{p_k}\frac{d\boldsymbol{v}_{p_k}}{dt} = \sum_{j\in \mathbf{N}_p}\left(\boldsymbol{F}_{p_kp_j}\right) + \sum_{j\in \mathbf{N}_c}\left(\boldsymbol{F}_{p_kc_j}\right) + \boldsymbol{F}_{p_{k_{drag}}} \approx 0. \quad \text{(Equation 4)}$$

Here, m_{c_i} and m_{p_k} are the cell and particle masses, respectively, $\boldsymbol{F}_{c_ic_j}$ represents cell-cell interaction force, $\boldsymbol{F}_{c_ip_j}$ indicates the cell-particle interaction force, $\boldsymbol{F}_{p_kp_j}$ is the particle-particle interaction force, $\boldsymbol{F}_{p_kc_j}$ denotes the particle-cell interaction force and $\boldsymbol{F}_{c_{i_{drag}}}$ and $\boldsymbol{F}_{p_{k_{drag}}}$ are the friction of the cell and particle with the extracellular matrix, respectively. The inertial terms $m_{c_i}\frac{d\boldsymbol{v}_{c_i}}{dt}$ and $m_{p_k}\frac{d\boldsymbol{v}_{p_k}}{dt}$ are neglected because $Re \ll 1$. The drag forces $\boldsymbol{F}_{c_{i_{drag}}}$ and $\boldsymbol{F}_{p_{k_{drag}}}$ are obtained from Stoke's law:

$$\boldsymbol{F}_{c_{i_{drag}}} = -6\pi\eta R_{c_i}\boldsymbol{v}_{c_i}, \quad \text{(Equation 5)}$$

$$\boldsymbol{F}_{p_{k_{drag}}} = -6\pi\eta R_p\boldsymbol{v}_{p_k}, \quad \text{(Equation 6)}$$

where η is the dynamic viscosity of the extracellular matrix, R_{c_i} is the radius of the i-cell, R_p is the radius of the particle and $\boldsymbol{v}_{c_i}$ and $\boldsymbol{v}_{p_k}$ are the velocities of the i-cell and k-particle.

Cell-cell interaction forces are usually modeled as repulsive-attractive forces. The repulsion between cells arises from cell resistance to deformation when their membranes touch, and the attractive forces are the result of the junctions that cells form between themselves through specialized protein complexes ([50]). In the case of particles, the repulsive-attractive forces represent the intermolecular forces in fluids. Accordingly, we modeled the interaction forces $\boldsymbol{F}_{\delta_i\gamma_j}$ (both subindexes δ and γ denote *c* or *p*, depending on whether the *i* and *j* agents are cells (*c*) or particles (*p*)) following,[12] as follows:

$$\boldsymbol{F}_{\delta_i\gamma_j} = F_{\delta\gamma}\frac{\boldsymbol{r}_{ij}}{\|\boldsymbol{r}_{ij}\|} \quad \text{(Equation 7)}$$

where:

$$\boldsymbol{r}_{ij} = \boldsymbol{x}_{\gamma_j} - \boldsymbol{x}_{\delta_i}, \quad \text{(Equation 8)}$$

and:

$$F_{\delta\gamma} = \begin{cases} F_{rep_{\delta\gamma}}\chi(-s)^{3/2}, & s<0 \text{ (repulsion)}, \\ -F_{adh_{\delta\gamma}}\chi\left\{(s+s_0)e^{-\lambda(s+s_0)^2} - v_0e^{-\lambda s^2}\right\}, & s\geq 0 \text{ (adhesion)}. \end{cases} \quad \text{(Equation 9)}$$

Consequently, χ, *s*, x_0 and v_0 are defined as:

$$\chi = \frac{R_{\delta_i}}{2}\left(\frac{1}{R_{\delta_i}}+\frac{1}{R_{\gamma_j}}\right), s = \frac{d - min_{dist}}{R_{\delta_i}}, \quad \text{Equation (10)}$$

$$x_0 = \sqrt{\frac{1}{2\lambda}}, v_0 = x_0e^{-\lambda x_0^2}. \quad \text{(Equation 11)}$$

$F_{rep_{\delta\gamma}}$ and $F_{adh_{\delta\gamma}}$ are the strengths of the adhesive and repulsive forces, respectively. $\boldsymbol{r}_{ij}$ is the distance between the centers of the agents, and R_{δ_i} and R_{γ_j} are the radii of the corresponding agents. x_0, v_0 and λ are matching constants, and χ is a geometric correction factor. The value of $min_{dist} = -0.1R_{\delta_i}$ is chosen such that the equilibrium state where the adhesive and repulsive forces are balanced is slightly less than zero, following,[12] and $d = \|\boldsymbol{r}_{ij}\| - R_{\delta_i} - R_{\gamma_j}$ is the distance between the agents' surfaces.

Since cells do not present any attraction towards the lumen fluid, the interaction force between cells and particles is only repulsive ($F_{adh_{cp}} = F_{adh_{pc}} = 0$).

Finally, the velocity of the *i*-cell and the *k*-particle at time *t* can be calculated explicitly:

$$\frac{d\boldsymbol{x}_{c_i}(t)}{dt} = \boldsymbol{v}_{c_i}(t) = \frac{1}{6\pi\eta R_{c_i}}\left(\sum_{j\in N_c}\left(\boldsymbol{F}_{c_ic_j}\right) + \sum_{j\in N_p}\left(\boldsymbol{F}_{c_ip_j}\right)\right), \quad \text{(Equation 12)}$$

$$\frac{d\boldsymbol{x}_{p_k}(t)}{dt} = \boldsymbol{v}_{p_k}(t) = \frac{1}{6\pi\eta R_p}\left(\sum_{j\in N_p}\left(\boldsymbol{F}_{p_kp_j}\right) + \sum_{j\in N_c}\left(\boldsymbol{F}_{p_kc_j}\right)\right). \quad \text{(Equation 13)}$$

Cell functions

To reproduce the biological characteristics of tumor pancreatic organoids with lumen, three cell functions are considered within the agent-based model: proliferation, quiescence, and fluid secretion. Proliferation is modeled by a growth phase, in which the cell progressively increases its volume as a result of DNA replication, and by a mitosis phase, in which the cell divides into two daughter cells. Thus, we track the volume of each cell $V_i(t)$ and calculate the volume growth from[40]:

$$\frac{dV_i(t)}{dt} = \alpha V_i(t), \quad \text{(Equation 14)}$$

where $\alpha = 1/T_c$ is the growth rate, which is related to the cell cycle time T_c. When the volume of the cell reaches twice its initial volume, DNA replication is concluded, and the cell divides. In this regard, cell division is performed using a random cleavage plane that contains the line that passes through the cell center and the lumen center of mass.

Quiescent cells remain inactive for 6 minutes (Δt_{bio}) and do not perform any cell function.

Finally, through the fluid secretion function, cells generate particles inside the lumen after fluid production time Δt_{exo}. First, when the lumen does not yet have any fluid, the cell secretes into the cell's center of mass. Then, when another cell secretes, some random particles are duplicated to recreate the increment in the fluid volume.

Cell division

When the volume of the cell reaches twice the value of its initial volume, DNA replication is concluded, and the cell divides. A spatially controlled division is a fundamental condition to maintain the lumen architecture and to enhance its growth by enlarging the lumen volume. In this regard, a complex molecularly controlled process regulates the spindle orientation, so mitosis occurs in the plane of the monolayer.[51–56] Here, we distinguish the division between nonpolarized cells and polarized cells. Nonpolarized cells are those that have not yet formed a lumen, and polarized cells are those that belong to a lumen and face it. In the case of nonpolarized cells, the division direction is chosen randomly. Polarized cell division is performed using a random cleavage plane that contains the line that passes through the cell center and the lumen center of mass. The position of the two daughter cells ($\boldsymbol{x}_{daughters}$) are calculated similarly to other models[57,58] from the center of the parent cell $\boldsymbol{x}_{parent}$ at:

$$\boldsymbol{x}_{daughters} = \boldsymbol{x}_{parent} \pm \left(R_c - \frac{1}{\sqrt[3]{2}} R_c \right) \boldsymbol{n}, \qquad \text{(Equation 15)}$$

where R_c is the radius of the parent cell and $\boldsymbol{n}$ is the unit orientation vector. When a nonpolarized cell divides, the unit orientation vector $\boldsymbol{n}$ is chosen randomly. However, when a polarized cell divides, the unit orientation vector $\boldsymbol{n}$ is normal to the random cleavage plane that contains the line that passes through the cell center and the lumen center of mass.

Fitness function

The fitness function is an objective function that assesses neural network performance. When the simulation finishes, a fitness value is given to the neural network depending on how close the obtained result is with respect to the target solution. Let N_c^t and V_l^t be the number of cells and lumen volume target values. The fitness function for cystic organoids is formed by three components: the first relates to the number of cells, the second to the lumen volume and the third is a penalty term for the eccentricity. Therefore, the fitness is evaluated through this equation:

$$Fitness(N_c, V_l, d_{cm}) = 0.5 \cdot 100^{-\frac{\left(N_c - N_c^t\right)^2}{2\left(N_c^t\right)^2}} + 0.5 \cdot 100^{-\frac{\left(V_l - V_l^t\right)^2}{2\left(V_l^t\right)^2}} - 250\left(\frac{e_c}{\left(V_l^t/10\right)}\right)^2, \qquad \text{(Equation 16)}$$

where N_c^t is the target number of cells, V_l^t is the target lumen volume and e_c is the eccentricity (the distance between the cells' center of mass and the lumen's center of mass). Thus, the better the neural network approximates the target number of cells and target lumen volume, the better the solution. However, the higher the eccentricity e_c is, the worse the solution. In the case of the fitness function for solid organoids, the fitness function is only dependent on the number of cells.

Eccentricity

In the process of the generation of a cystic organoid, fluid secretion, proliferation, and quiescence must be orchestrated. If cells secrete fluid too early, the preapical patch, which is the closed volume necessary to secrete fluid would not form, so the fluid would leak. Also, if cells secrete too much the lumen would open. Also, if they do not proliferate, they would not create more inner space for the fluid and it would leak. To maintain the structure, there is competition between these processes. Therefore, to make the neural network understand that these cell functions must coordinate to preserve the structure, we introduce a metric called eccentricity that permits us to take into account the structure. This metric refers to the distance between the cells' center of mass and the lumen's center of mass. The lower this value is, the more spherical the organoid is (Figure S1A). Conversely, the higher this value is, the less spherical the organoid is, reaching a certain moment that it opens (Figure S1B). Therefore, the neural network uses this metric to evaluate the state of the cystic structure.

Characterization of the lumen fluid

We conducted a shear test to validate that our approach represents an actual fluid. In this regard, we used a bulk domain consisting of fluid particles with a radius $R_p = 2\,\mu m$ (Figure S3) The top layer of particles (depicted in black in Figure S3) was subjected to a velocity u in the X direction, while the bottom layer (depicted in blue in Figure S3) had fixed displacements in the X and Y directions.

Then, to obtain the steady flow curve to assess the behavior of particles, we set the shear stress experiment with a time duration of $t_{exp} = 30$ min and varied the velocity of the top layer within the values $u = [0.01,0.05, 0.1,0.2,0.5]/t_{exp}\ \mu m/min$. Finally, we were able to determine the viscosity of the fluid (η) by measuring the shear stress (τ) as the sum of the reaction forces (F_x) of the particles forming the bottom XY-plane of the cube of side length l in the X direction $\left(\tau = \frac{\sum F_x}{l^2}\right)$ for each shear rate ($\dot{\gamma} = \partial v/\partial z$).

$$\eta = \frac{\tau}{\frac{\partial v}{\partial z}}. \qquad \text{(Equation 17)}$$

We show the linear regression of the shear stress on the deformation rate (Figure S4). We obtained that the shear stress is linearly proportional to the shear strain, resulting in a constant viscosity of $1.072 \cdot 10^{-3}$ Pa.s, with an R-Squared value of 0.9996. Therefore, this analysis confirms that our approach accurately reproduces the expected velocity and shear stress profiles for a Newtonian fluid similar to water ($\sim 10^{-3}$ Pa.s) and demonstrates the capability of our method to model real fluids, serving as a minimal model that aims to recapitulate certain features such as mass conservation and momentum exchange. However, it does not rigorously reproduce bulk-scale physical properties.

Particle radius influence on the rheological properties of the fluid

The particle radius in a particle-based model has a great influence on the mechanical and rheological properties of the fluid. Viscosity is a physical property of a fluid that describes its resistance to deformation or flow. In particle-based approaches, the viscosity coefficient is typically calculated from the smoothing length and the particle spacing and depends on the strength of the interaction forces between neighboring particles.

Reducing the particle radius while keeping all other properties constant, results in a change in the mechanical and rheological properties of the fluid. Therefore, to maintain the rheological properties of the fluid while changing the particle radius it is necessary to modify other parameters such as the interaction forces between the particles to maintain a consistent viscosity coefficient so that the macroscopic behavior of the fluid remains the same. On the one hand, if the particle radius is reduced for a given volume, the number of particles increases. Consequently, the fluid becomes more viscous than when the particle radius is greater since more particles oppose the shear stress. On the other hand, increasing the particle radius reduces the viscosity of the fluid. Thus, it is necessary to adjust the interacting forces so that the behavior remains the same when integrating the response of the fluid to deformation.

We have modified the interacting forces between particles for particle radii of 1, 4 and 6 μ m, through the optimization of the parameters $F_{rep_{pp}}$ and $F_{adh_{pp}}$ in Equation 9 so we obtain the same response of the rheological properties of the fluid as for the particle radius of $R_p = 2$ μ m (Figure S5A). In this case, the mean value of the viscosity obtained for all particle sizes is $1.078 \cdot 10^{-3}$ Pa.s with a standard error of $1.210 \cdot 10^{-5}$. Finally, we show a comparison between the interacting forces for $R_p = 2$ μ m and the modified interacting forces for particle radii of 1, 4 and 6 μ m to obtain the same response (Figure S5B). Since the number of particles increases when the particle radius decreases, the interacting forces reduce so when integrating the contribution of the particles to the viscosity it is equivalent across different particle sizes.

Particle radius influence on fitness

We analyze the effect of the variation of the particle radius. For this purpose, we simulated an intermediate-size cystic organoid (organoid 2 from Table 1 of the manuscript) with different particle radii to determine how it affects the results. We show the influence of particle radii on the fitness value and computation time while keeping the interacting forces the same as for the case of $R_p = 2$ μ m (Figure S6A). We found that increasing the particle radius from 1 μm to 2 μm causes an absolute decrease of the fitness value of 0.15% and a relative decrease in the computation of 96.02% (from 1029 min to 41 min). An increase of

the particle radius from 2 μm to 4 μm entails an absolute worsening of the fitness of the approximation of 4.16% and a relative improvement of the computing time of 93.88% (from 41 to 3 min). Therefore, this analysis shows that the rheological properties of the fluid are important in the cystic morphogenetic process and require proper interaction to develop the organoid correctly. In particular, making the fluid less viscous (increasing the particle radius) entails a worsening of the simulated organoid. Then, we studied the influence of the particle radii on the fitness value and computation time with the interacting forces obtained in Figure S5B to have the same viscosity as water. In this case (Figure S6B) the fitness is above 0.91 for all particle sizes simulated. However, increasing the particle radius results in a slight decrease in fitness due to a higher increase in lumen volume per fluid secretion compared to smaller particle sizes, leading to a reduced sensitivity in the secretion process. Therefore, a particle radius of 2 μ m is optimal for maintaining low computational costs while having sufficient sensitivity in the fluid secretion process.

Multiphoton microscopy of PDAC organoids

3D cultures of PDAC93-GFP and PM12500-GFP cells were generated and incubated for 5 days as described in the previous section. Then, samples were fixed with 4% paraformaldehyde in DPBS solution at 37 °C for 30 minutes and washed thoroughly with DPBS followed by cell nuclei staining with a 5 $\mu g/ml$ solution of Hoechst 33342 for 16 h at room temperature. Image acquisition was performed using a Zeiss LSM 880 AxioObserver inverted confocal microscope equipped with a 25x LD LCI Plan-Aprochomat 0.8 NA W objective. A total sample volume of approximately 1065x1065x200 μm^3 was acquired using a Mai Tai® DeepSee™ Ti:Sapphire laser sequentially set to 740 nm and 920 nm for multiphoton microscopy of stained nuclei and endogenous cytoplasmic GFP, respectively.

Quantification of 3D PDAC organoid morphology

Nuclei and cytoplasmic GFP segmentation masks were generated using StarDist 3D[48] and the Trainable Weka Segmentation 3D[59] plugin for Fiji, respectively. The StarDist 3D model was trained from scratch for 50 epochs on 45 paired image patches (patch size: (72,72,32), batch size 1, number of rays: 32, augmentation: true),[60] accelerated using an NVIDIA Quadro P1000 GPU. Afterward, the generated masks were analyzed using a homemade script for Fiji. Briefly, cytoplasmic GFP masks were preprocessed using a GPU-accelerated 3D median filter and binary closing available in the CLIJ2 library[61]). Then, each individual organoid was labeled, and the intersection between the cytoplasmic GFP mask and the nuclear mask was calculated using a logical AND operator. The number of nuclei and morphological descriptors, such as the organoid volume or sphericity, were quantified using MorpholibJ.[62]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses and figure generation were performed using MATLAB (MathWorks). The exact number of simulations conducted to test the robustness in Figure 4 was indicated in the legend. The R-squared values of the linear regressions in Figures S4 and S5 were also specified in the legend.