Digitize your Biology! Modeling multicellular systems through interpretable cell behavior

Jeanette A.I. Johnson[1,2+], Genevieve L. Stein-O'Brien[1,2,3+], Max Booth[1], Randy Heiland[4], Furkan Kurtoglu[4], Daniel R. Bergman[1,2], Elmar Bucher[4], Atul Deshpande[1,2], André Forjaz[5], Michael Getz[4], Ines Godet[6], Melissa Lyman[1,2], John Metzcar[4,7], Jacob Mitchell[2,8], Andrew Raddatz[9], Heber Rocha[4], Jacobo Solorzano[10], Aneequa Sundus[4], Yafei Wang[4], Danielle Gilkes[1], Luciane T. Kagohara[1,2], Ashley L. Kiemen[1,2,11], Elizabeth D. Thompson[11], Denis Wirtz[1,2,5,11,12], Pei-Hsun Wu[2,5], Neeha Zaidi[1,2], Lei Zheng[1,2], Jacquelyn W. Zimmerman[1,2], Elizabeth M. Jaffee[1,2], Young Hwan Chang[13], Lisa M. Coussens[14], Joe W. Gray[13], Laura M. Heiser[13], Elana J. Fertig[1,2,15,16*], Paul Macklin[4,*]

[1]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University. Baltimore, MD USA.
[2]Convergence Institute, Johns Hopkins University. Baltimore, MD USA.
[3]Department of Neuroscience, Johns Hopkins University. Baltimore, MD USA.
[4]Department of Intelligent Systems Engineering, Indiana University. Bloomington, IN USA.
[5]Department of Chemical and Biomolecular Engineering, Johns Hopkins University. Baltimore, MD USA
[6]Memorial Sloan Kettering Cancer Center. New York, NY USA.
[7]Department of Informatics, Indiana University. Bloomington, IN USA
[8]Department of Human Genetics, Johns Hopkins University. Baltimore, MD USA.
[9]Department of Biomedical Engineering, Georgia Tech University. Atlanta, GA USA.
[10]Centre de Recherches en Cancerologie de Toulouse. Toulouse, France.
[11]Department of Pathology, Johns Hopkins University. Baltimore, MD USA
[12]Department of Materials Science and Engineering, Johns Hopkins University. Baltimore, MD USA
[13]Department of Biomedical Engineering, Oregon Health & Science University. Portland, OR USA.
[14]Department of Cell, Developmental and Cancer Biology, Oregon Health & Science University. Portland, OR USA.
[15]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA.
[16]Department of Biomedical Engineering, Johns Hopkins University. Baltimore, MD USA.

[+]Contributed equally
[*]Corresponding authors. ejfertig@jhmi.edu , macklinp@iu.edu

ABSTRACT

Cells are fundamental units of life. Recent technical advances have revolutionized our ability to quantify the state and identity of individual cells, and intercellular regulatory programs. However, these static measurements alone are limited in their ability to predict the complex collective behaviors that emerge from populations of many interacting cells over time. Mathematical models have a proven record of successfully predicting the behaviors of dynamic biological systems, e.g., therapeutic responses in cancer. Simulations from these models enable *in silico* visualization, examination, and refinement of biological models and can be used to generate new hypotheses about cells and their collective behavior. Agent-based modeling is particularly well-suited to studying communities of interacting cells, as it is intuitive to map a single cell to a single agent. Thus, we have developed a conceptual framing (with a reference implementation in the widely-used PhysiCell agent-based modeling framework) that can be initialized directly from single cell and spatial transcriptomic data, and that can be easily populated with interactive rules. Because the expert mathematical and computational knowledge required to build agent-based models has limited their widespread adoption in the biomedical research community, we engineered this framework to specify complex cellular responses to signals (or stimuli) using a single line of human-readable text. This plain language text encodes cellular phenotypes and regulatory mechanisms from high throughput data and published literature, using a novel concept of hypothesis grammar. We motivate and fully describe this grammar and its philosophy, and then present a series of five example reference models of tumor growth and response to immunotherapy. Biologically, these examples demonstrate how mathematical models can predict from single cell and spatial transcriptomic data the cellular phenotypes responsible for tumor cell invasion and the simulation of immunotherapy treatment to overcome tumor cell growth.

Computationally, these examples are designed to demonstrate how this conceptual framing and software implementation empower interdisciplinary teams to build an agent-based model of their experimental system, leveraging prior biological knowledge alone or in combination with information from spatial multiomics technologies. Altogether, this approach provides an interface to bridge biological, clinical, and systems biology researchers for mathematical modeling of biological systems at scale, allowing the community to extrapolate from single-cell characterization to emergent multicellular behavior.

INTRODUCTION AND BACKGROUND

Every cell has an identity and internal state defined by its DNA, epigenetic profile, and molecular signaling pathways. Quantifying those internal state variables and how they change due to signals received from their environment has enabled the classification of cell types and behaviors[1-4]. However, the necessity of destroying the cell to obtain those measurements results in static snapshots of a single moment in time. Thus, the ability to connect those snapshots into cellular movies remains an area of strong community interest and an open computational challenge[5-7]. The large number of cellular and molecular interactions within biological systems make temporal predictions impossible to perform solely through experimentation. While techniques such as trajectory inference and RNA velocity can estimate transitions of cellular phenotypes[8], they cannot account for more complex temporal changes throughout the diverse cellular ecosystems of biological systems. Beyond filling the gaps between measurement times, computational techniques for temporal analysis of cellular states provide the potential to predict the future state of multicellular systems. While machine learning from single-cell datasets can make these predictions for individual cell types[9,10], they cannot account for changes resulting from cell-cell interactions. Moreover, these predictive methods are data-driven and unable to predict cellular states from prior biological knowledge or mechanism alone. More advanced computational tools are needed to model unseen emergent behaviors in complex biological systems comprised of multiple cell types.

Agent-based modeling is a powerful mathematical modeling technique that enables prediction of emergent complex behaviors by populations of individual "agents" in a system as they follow predefined rules based on their identity, state, and nearby conditions[11]. Agent-based models (ABMs) are well-suited to studying the dynamics of multicellular biology, as they can map a cell type to their set of internal states and rules of behavior[12-15]. At each time increment, each agent evaluates its surroundings and internal state variables and considers this information to calculate its next action. When applied to individual cells, each agent's rules can represent hypotheses of single-cell behaviors, including their actions upon or in response to nearby cells (i.e., cell-cell interactions). By encoding the rules and relationships governing the stochastic behavior of cellular agents in complex systems, ABMs empower *in silico* experimentation and modeling of multiscale dynamic knowledge representations even in the absence of temporal measurements[13,16].

ABMs can thus be used to test hypotheses in human development and disease where comprehensive, subject based experimentation is not possible. Applications to understanding human disease and in particular cancer have already been proven particularly powerful[17-32]. By predicting the future state of cells and the impact of perturbations, ABMs provide a powerful toolset to generate digital twins and virtual clinical trials[7,14,33-41]. Furthermore, the ability to run simulations at scale across diverse biological conditions with ABMs can refine biological understanding and predict future cellular behaviors in these complex systems. As a result, they provide a computational means to prioritize bench experiments or clinical trials to address the costs and practical constraints of real-world experimentation.

Widespread application of ABMs for modeling biological systems currently has two major limitations; the highly technical nature of most software implementations, and the limited ability to integrate molecular data to ground simulations in the real-world. The former issue gatekeeps ABM away from those without

significant mathematical and computer-programming experience. This technical requirement limits widespread application, including by many potential users who have extensive working knowledge of the biological system they are investigating. Even for users with advanced computational expertise, the custom coding required can limit reproducibility. Disease etiology and operating biological hypotheses are often hidden deep in source code, obscuring the assumptions about the system and, thus, the full set of hypotheses that are being simulated and tested. These technical limitations to ABM modeling also limit the ability to embed molecular datasets, which are often too high-dimensional to manually encode into equations of agents and rules. A conceptual framing that can generally abstract cellular phenotypes and their interactions—combined with a simplifying coding infrastructure—is essential for the integration of molecular measurements to personalize model predictions.

Here, we sought a way for an expert's knowledge base to be condensed and standardized into a set of clear biological statements that can be readily mapped onto biophysically motivated ABM parameters that describe their biological system, with or without molecular and cellular measurements. Facilitating this requires an intuitive language to concisely express expert knowledge as plain text descriptions of the rules of cell interactions that "encode" a system of study, and software to translate these plain text descriptions into mathematical expressions and executable models for immediate exploration of a digitized copy of the biological system[42]. While we use PhysiCell[42] as a reference implementation, the grammar can independently represent expert knowledge for further analysis or import into other simulation frameworks[43-48]. Thus, the grammar allows for broad application of ABMs and addresses the need to be reproducible, modular, and extensible. We demonstrate how this grammar can encode expert biological knowledge through sample models in cancer biology spanning tumor cell growth, invasion, and response to immunotherapy. We note that because this language focuses on fundamental cell behavioral responses to biophysical signals, it can be applied far beyond cancer biology such as to immunology, developmental biology, tissue engineering, and related areas.

Building on the successful efforts by the community to compile unified definitions and ontologies of cell behavior[2,5,49-53], this work represents the first major integration of those ontologies with high-throughput single cell and spatial omics data. Our grammar defines the components of ABMs based on annotated cell types and behaviors, and empowers integration of cellular phenotypes estimated from genomics data. Rules can be both knowledge-driven (e.g., based on expert statements that observe changes in cell behavior in response to chemical stimuli) or data-driven (e.g., observing changes in expression of immune cells in the tumor microenvironment). We demonstrate how the data-driven rules enable for the first time direct transfer of bulk and single-cell RNA-seq and spatial transcriptomics into ABMs for personalized predictions. Assimilating multi-omics and spatial molecular data into agent-based models can ground simulations in the real-world in such a way that ABMs can serve as an abstract digital copy of the biological system as it changes over time.

RESULTS

Agent-based simulation frameworks[12] like PhysiCell[42] model individual cells as software agents with independent states (e.g., position, volume, cycle status) and processes (e.g., volume regulation, motility, secretion); see **Fig. 1A.** Each cell agent respond to stimuli (signals) in their microenvironment that effect changes in their behaviors (**Fig. 1B**). Cell hypotheses relating cell behavioral responses to signals are written in a grammar that can be translated into mathematics and executable code. (**Fig. 1C**). Hypotheses can be drawn from a variety of sources, including domain expertise, natural language processing / mining of prior literature, and analysis of transcriptomic and other data. Due to our uniform knowledge representation, all these rules can be compatibly integrated into a growing library of cell hypotheses. (**Fig. 1D**).

We implement the grammar using the PhysiCell agent-based modeling framework[42], however it can be translated to other agent-based modeling systems[43,45-48,54-56].

[Fig. 1]

After describing our new hypothesis grammar, we demonstrate five progressive case studies that express complex multicellular systems behaviors using the behavior hypothesis grammar. The models represent incrementally more complex systems of interacting Eukaryotic cells, and each seeks to highlight different capabilities and features of PhysiCell and the cell hypothesis rules grammar. (**Figures 2-6, Tables 4-8**) All five models simulate a tumor microenvironment with increasing cell diversity, including immune cells, culminating in the fifth reference model which performs a virtual clinical trial to predict the effect of combination immune-targeted therapies on tumor clearance using genomics-derived rules (**Fig. 6**). This approach is generalizable to diverse biological systems.

A grammar encoding cell behavioral responses to extracellular signals

Behavioral statements

For any cell type $\boldsymbol{T}$, we construct simple statements that relate changes in a single behavior $\boldsymbol{B}$ to an exogenous signal $\boldsymbol{S}$: "In cell type $\boldsymbol{T}$, $\boldsymbol{S}$ increases/decreases $\boldsymbol{B}$ [with optional arguments]." Here $\boldsymbol{B}$ is a well-defined biophysical parameter in our dictionary of behaviors (see **Methods** and **Supplementary Information**), $\boldsymbol{S}$ is a well-defined biophysical variable in our dictionary of signals, and optional arguments further specific the mathematical behavior of the responses. For example:

In malignant epithelial cells, oxygen increases cycle entry.

In MCF-7 breast cancer cells, doxorubicin increases apoptosis.

A full description of the grammar, optional arguments, and examples can be found in the **Supplementary Information**.

Mathematical representation: individual rules

With clearly defined behaviors and signals and the grammar to connect them, we can uniquely map human-interpretable cell hypothesis statements onto mathematical expressions that make the grammar both human interpretable and computable. Each individual rule modulates a single behavioral parameter b as a function of a signal s. Given a response function R, we then mathematically represent the individual rule as a function $b(s)$:

$$b(s) = b_0 + (b_M - b_0)R(s), \tag{1}$$

where b_0 is the base value of the parameter in the absence of signal, and b_M is the maximally changed value of the parameter with large signals. By default, we use sigmoidal (Hill) response functions R, due to their familiar use in signaling network models and pharmacodynamics, as well as their smooth variation between 0 (at no response) and 1 (at maximum response). However, capped linear response functions (varying between 0 and 1) are also possible. See **Fig. 1C** for a typical rule. Full mathematical details and additional detailed examples are available in the **Supplementary Information**.

Generalized mathematical representation: multiple rules

Our full mathematical formulation allows new hypotheses to be directly added to models without modifying prior hypotheses, making our framing extensible and scalable as new knowledge is acquired. Suppose that a behavior $\boldsymbol{B}$ (with corresponding behavioral parameter b) is controlled by multiple rules subject to promoting (up-regulating) signals $\boldsymbol{u}$ and inhibiting (down-regulating) signals $\boldsymbol{d}$:

- u_1 increases B (with half-max u_1^* and Hill power p_1)
- u_2 increases B (with half-max u_2^* and Hill power p_2)

- ...
- u_m increases B (with half-max u_m^* and Hill power p_m)
- d_1 decreases B (with half-max d_1^* and Hill power q_1)
- d_2 decreases B (with half-max d_2^* and Hill power q_2)
- …
- d_n decreases B (with half-max d_m^* and Hill power q_n)

Here, let $b_M$be the maximum value of the behavior parameter b (under the combined influence of the up-regulating signals $\boldsymbol{u}$), let b_0 be its base value in the absence of signals, and let b_m be its minimum value (under the combined influence of the down-regulating signals $\boldsymbol{d}$).

We define the total up response as:

$$U = H_M(\boldsymbol{u}; \boldsymbol{u}_{\text{half}}, \boldsymbol{p}) = \frac{\left(\frac{u_1}{u_1^*}\right)^{p_1} + \left(\frac{u_2}{u_2^*}\right)^{p_2} + \cdots + \left(\frac{u_m}{u_m^*}\right)^{p_m}}{1 + \left(\frac{u_1}{u_1^*}\right)^{p_1} + \left(\frac{u_2}{u_2^*}\right)^{p_2} + \cdots + \left(\frac{u_m}{u_m^*}\right)^{p_m}} \tag{2}$$

and the total down response as:

$$D = H_M(\boldsymbol{d}; \boldsymbol{d}_{\text{half}}, \boldsymbol{q}) = \frac{\left(\frac{d_1}{d_1^*}\right)^{q_1} + \left(\frac{d_2}{d_2^*}\right)^{q_2} + \cdots + \left(\frac{d_n}{d_n^*}\right)^{q_n}}{1 + \left(\frac{d_1}{d_1^*}\right)^{q_1} + \left(\frac{d_2}{d_2^*}\right)^{q_2} + \cdots + \left(\frac{d_n}{d_n^*}\right)^{q_n}}. \tag{3}$$

We combine the overall response of the behavioral parameter as bilinear interpolation in the nonlinear up- and down-responses U and D:

$$b(\boldsymbol{u}, \boldsymbol{d}) = (1 - D) \cdot [(1 - U) \cdot b_0 + U \cdot b_M\ ] + D \cdot b_m \tag{4}$$

Notice that:

- In the presence of up-regulating signals only, U reduces to a multivariate Hill response function $H_M(\boldsymbol{u}; \boldsymbol{u}^*, \boldsymbol{p})$. In the presence of a single up-regulating signal u only, $b(\boldsymbol{u},\boldsymbol{d})$ reduces to a Hill response curve *b(u)* familiar to systems biology and pharmacodynamics.
- In the presence of down-regulating signals only, D reduces to a multivariate Hill response function $H_M(\boldsymbol{d}; \boldsymbol{d}^*, \boldsymbol{q})$. In the presence of a single down-regulating signal d only, $b(\boldsymbol{u},\boldsymbol{d})$ reduces to a Hill response curve *b(d)* familiar to systems biology and pharmacodynamics.
- Generally, the combined up-regulating signals sets a "target" value of the parameter, which can then be inhibited by the combined down-regulating signals.

Note also that adding and removing individual rules to the form does not require alteration to the remaining rules. In this release, we use multivariate Hill response functions for clarity, but mixed linear and Hill responses could be used in the future. The PhysiCell implementation of this generalized response, additional mathematical details, and expanded examples can be found in the **Supplementary Information**.

Automated model annotation

To help drive reproducibility, we generate and save a full description of all rules in HTML and text formats after initial parsing. Future revisions will generate these annotations in alternative formats, such as Microsoft Word tables in DOCX format and LaTeX tables.

Example 1: Model the progression of hypoxia in a metastatic tumor

In cancer, cell growth becomes unchecked, which exhausts oxygen and nutrients in non-vascularized tumors. Modeling this resource consumption problem has provided a foundation for mathematical modeling of tumors, and this serves as the base example of tumor cell behavior in the absence of an immune response, from which our other modeling examples are built.

As a first example, we show that the grammar can express common mathematical models for oxygen-dependent tumor growth[42,57-63], with inclusion of additional mechanobiologic feedback on cycle entry as in recent tumor growth models[57,58,64]. Following prior work[57], we model hypoxia-induced migration, where low oxygen conditions can "reprogram" tumor cells to a transient, post-hypoxic phenotype of increased chemotactic migration, and subsequent prolonged exposure to high oxygen conditions can "revert" those cells back to a less motile phenotype. Consistent with prior modeling predictions and experimental validation[57], these motile cells fail to exit the tumor and invade the surrounding tissue when their hypoxic adaptations do not persist in higher oxygen conditions (**Fig. 2b**, later simulation times). Note that rather than stating "low oxygen increases necrosis" and "hypoxia increases transformation into motile tumor cells", the language uses "oxygen decreases necrosis" and "oxygen decreases transformation to motile tumor cells", as there are currently no symbols for "no", "lack of", or "low."

We used these rules to simulate 5 days of growth of a 2-D tumor in an environment of 38 mmHg oxygenation (physioxic conditions[65]), starting from 2000 viable cells seeded randomly in a virtual disk with a 400 μm radius. Results are shown in **Fig. 2**.

[Fig. 2]

This leads to an oxygen-poor necrotic tumor core, while hypoxic cells are disseminated throughout the tumor. Here we show a model of a transient post-hypoxic phenotype of increased chemotactic migration, where cells eventually return to their baseline phenotype upon reoxygenation. The full model is stored as **example1** in the user_projects directory of the PhysiCell git repository, and an automated model annotation is in the **Supplementary Materials**.

Cell Hypothesis Rules (detailed)

In tumor cells:
oxygen increases cycle entry from 1.7e-05 towards 0.0007 with a Hill response, with half-max 21.5 and Hill power 4.
pressure decreases cycle entry from 1.7e-05 towards 0 with a Hill response, with half-max 0.25 and Hill power 3.
oxygen decreases necrosis from 0.0028 towards 0 with a Hill response, with half-max 3.75 and Hill power 8.
oxygen decreases transform to motile tumor from 0.001 towards 0 with a Hill response, with half-max 6.75 and Hill power 8.

In motile tumor cells:
oxygen increases cycle entry from 1.7e-05 towards 0.0007 with a Hill response, with half-max 21.5 and Hill power 4.
pressure decreases cycle entry from 1.7e-05 towards 0 with a Hill response, with half-max 0.25 and Hill power 3.
oxygen decreases necrosis from 0.0028 towards 0 with a Hill response, with half-max 3.75 and Hill power 8.
oxygen increases transform to tumor from 0 towards 0.005 with a Hill response, with half-max 6.75 and Hill power 8.

Example 2: Forecast tumor progression in pancreatic cancer from an initial state defined by spatial transcriptomics

We sought to extend our rules framework from our toy model of tumor cell growth in Example 1 to a personalized, data-driven model of tumor progression. For these models, we leveraged the rules-based framework to encode rules of cell-cell interactions and initial conditions of cells from spatial transcriptomics and single-cell RNA-seq datasets. Due to the dense stroma that is characteristic of PDAC, we sought to use our computational model to investigate the influence of cancer-associated fibroblast (CAF) content on tumor progression (**Fig. 3a**). Our previous studies of cell-cell interactions in single-cell RNA-seq analysis and organoid co-cultures demonstrated two distinct transcriptional phenotypes in neoplastic cells[66].

Fibroblast density and co-culture were associated with co-occurrence of inflammatory signaling and EMT in neoplastic cells, and attributed to cell-cell communication from integrins in fibroblasts (ITGB1) to epithelial cells. This neoplastic cell phenotype was mutually exclusive with proliferative signaling in epithelial cells. These observations led us to hypothesize that fibroblasts promote the epithelial-to-mesenchymal transition (EMT), facilitating and enabling cell invasion in PDAC. We leverage the rules-based framework to test this hypothesis in computational simulations. Our model boils down key features of interest from this transcriptional analysis into a simplified model of a pancreatic tumor microenvironment, where the major actors are epithelial and mesenchymal cells, fibroblasts, the collagen-rich ECM, and other pancreatic cells, here primarily modeled in their role as structure and scaffolding within which the other cell types interact. (Future work will extend the focus to additional processes and roles played by these cells.) The model has a simple pro-inflammatory factor which is pro-tumorigenic and which inhibits the normal mesenchymal-to-epithelial transition, while contact with fibroblasts increases the rate of EMT. In this way the fibroblasts in this model can shift the microenvironment in favor of tumorigenesis consistent with the hypothesis generated from our transcriptional signatures.

[Fig. 3]

For this model, we initialized cell positions in a virtual tissue based on Visium spatial transcriptomic data from two pancreatic ductal adenocarcinoma (PDAC) lesions with varying fibroblast density (**Fig. 3b**)[67]. Our bioinformatics methods for three-way integration between H&E imaging data, spatial transcriptomics, and transcriptional signatures of cellular phenotypes[67] along with pathology curation (see **Methods**) were used to categorize the epithelial cell phenotypes and fibroblasts in our model. ECM density was initialized using a heuristic from image-derived collagen and cancer-associated fibroblast annotations from the H&E imaging, using a machine learning method for tissue annotation called CODA[68], and the bounding cells were assumed to contain a similarly dense collagen matrix, forming a niche around the known sample and abstractly reflecting the character of the solid pancreatic tissue. Other pancreatic cells in the spatial transcriptomics data were approximated as steady state (no net proliferation, death, motility, or secretion) and are assumed to be essentially inert with regards to carcinogenesis. Agents were initialized from these cellular annotations of the Visium data matrix as described in Methods. The tumor's development was forecasted for 90 days, initialized with 3983 cells (PDAC01) or 3482 cells (PDAC02). Included in this count is a ring of bounding pancreatic cells. Simulation results are shown in (**Fig. 3c)**. We observe a transitory state in which the tumor cells transition from mixed epithelial and mesenchymal states to become nearly uniformly mesenchymal due to interactions with the fibroblasts. Subsequently, masses of epithelial tumor cells arise in both models, consistent with extensive tumor growth and the observed transition between PDAC transcriptional subtypes during tumorigenesis and invasion[69]. An interface of mesenchymal tumor cells is maintained between the epithelial tumor cell and fibroblast cell masses. In PDAC01, rapidly dividing epithelial tumor cell clusters arise from lesions not surrounded by fibroblasts and invade the bounding pancreatic cells. In contrast, the dense, uniform fibroblast surrounding all of the lesions in PDAC02 slows invasion. The reduced rate of invasion results in smaller invasive lesions at 90 days in the PDAC02 sample compared to PDAC01. Whereas all the lesions in PDAC01 are invading bounding pancreatic cells, PDAC02 develops an epithelial tumor cell mass that is constrained from further motility by the surrounding CAFs and the dense ECM they have constructed. These computational predictions show that the hypothesized epithelial-fibroblast interactions inducing the epithelial transition between classical (epithelial-like) and basal (mesenchymal-like) pancreatic transcriptional subtypes observed in primary human pancreatic tumor progression, that return to the more epithelial-like classical subtype in metastatic sites[69], highlighting the power of using the grammar, genomics-driven rules, and integrating spatial molecular data into dynamical modeling to contextualize and explain complex cellular interactions observed in single-cell data and organoid models from spatial.

Cell Hypothesis Rules (detailed)

In epithelial_normal cells:
contact with fibroblast increases transform to mesenchymal_normal from 0 towards 0.01 with a Hill response, with half-max 0.01 and Hill power 4.

In mesenchymal_normal cells:
ecm decreases migration speed from 0 towards 0 with a Hill response, with half-max 0.5 and Hill power 4.
inflammatory_signal decreases transform to epithelial_normal from 0.01 towards 0 with a Hill response, with half-max 0.2 and Hill power 4.

In epithelial_tumor cells:
pressure decreases cycle entry from 0.001 towards 0 with a Hill response, with half-max 1 and Hill power 4.
contact with fibroblast increases transform to mesenchymal_tumor from 0 towards 0.01 with a Hill response, with half-max 0.01 and Hill power 4.

In mesenchymal_tumor cells:
ecm decreases migration speed from 0 towards 0 with a Hill response, with half-max 0.5 and Hill power 4.
inflammatory_signal decreases transform to epithelial_tumor from 0.01 towards 0 with a Hill response, with half-max 0.2 and Hill power 4.

A full copy of the automated model annotation can be found in the **Supplementary Materials**, while the code and configuration are included in the GitHub repository as **example2_pdac01** and **example2_pdac02**.

Example 3: Tumor attackers and tumor defenders

The models shown thus far allow for tumor cell growth to continue unchecked by the immune system. In these models, the only constraints on the tumor are mechanical barriers and its own size preventing the core from maintaining sufficient oxygenation. However, in reality, tumors have a microenvironment composed of immune cells that modulate tumor growth and are important components to model and key contributors to therapeutic response.

To introduce virtual immune cells, we extend our model by including CD8+ T cell agents capable of contact-mediated killing, and pro- and anti- inflammatory factors that modulate the probability that killing will occur after a given cell contact. CD8+ T cells perform the cytotoxic killing, while macrophages switch between promoting and suppressing tumor killing (secreting pro- or anti-inflammatory factor) depending on the oxygenation in their immediate surroundings, and consistent with the literature[70-72]. Macrophages are also responsible for phagocytosing dead cells and can increase secretion of pro-inflammatory factors, which in turn attracts CD8+ T cells who home to the tumor by following this chemokine. CD8+ T cells can attack and damage malignant epithelial cells, and accumulated damage can cause tumor cell death.

We used these rules to simulate 5 days of growth of a 2-D tumor in a virtual environment of 38 mmHg oxygenation (physioxic conditions[65]), starting from 1000 viable tumor cells seeded randomly, surrounded by a ring of virtual immune cells where 50 of each non-tumor cell type are seeded. See **Fig. 4**. Through these simulations, we observe that CD8+ T cells cluster together and migrate throughout the tumor along with macrophages to accomplish tumor clearance. As such, this model shows the cooperation of the innate and adaptive immune system in the task of tumor sensing and clearance and demonstrates a simple, continuous, plastic M1-M2 axis of macrophage behavior. However, it does not account for unidirectional programs of macrophage differentiation or for the dynamics of T cell activation and clonal expansion. Since macrophages in this simulation all begin external to the tumor, they initially have the M1 phenotype which predisposes them toward an inflammatory response. Later examples further promoting immune response with immunotherapy treatment will show behavior of a tumor thoroughly infiltrated by anti-inflammatory macrophages.

[Fig. 4]

A full copy the automated model annotation can be found in the **Supplementary Materials**, while the code is included in the GitHub repository as **example3**.

Cell Hypothesis Rules (detailed)

In tumor cells:
oxygen increases cycle entry from 0 towards 0.00072 with a Hill response, with half-max 21.5 and Hill power 4.
pressure decreases cycle entry from 0 towards 0 with a Hill response, with half-max 1 and Hill power 4.
oxygen decreases necrosis from 0.0028 towards 0 with a Hill response, with half-max 3.75 and Hill power 8.
damage increases apoptosis from 7.2e-05 towards 0.072 with a Hill response, with half-max 180 and Hill power 2.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In macrophage cells:
oxygen increases pro-inflammatory factor secretion from 0 towards 10 with a Hill response, with half-max 5 and Hill power 4.
oxygen decreases anti-inflammatory factor secretion from 10 towards 0 with a Hill response, with half-max 5 and Hill power 4.

In CD8 T cell cells:
anti-inflammatory factor decreases attack tumor from 0.1 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
pro-inflammatory factor increases attack tumor from 0.1 towards 1 with a Hill response, with half-max 0.5 and Hill power 8.
anti-inflammatory factor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
contact with tumor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 2.

Example 4: T cell activation, expansion, and exhaustion in a diverse tumor microenvironment

New therapies are developing to harness the immune system to attack tumor cells and are leading to unprecedented benefit in subsets of patients with immune-rich tumors[73]. In spite of this promise, many other tumor subtypes have immunosuppressive microenvironments that limit immunotherapy response. Together these observations have motivated the development of therapeutic strategies designed to promote effective tumor response by favorably modulating the tumor immune microenvironment[74].

Literature derived from preclinical and correlatives in clinical studies indicate the impact of therapeutics on specific cell types[75-78]. However, given the complexity of the tumor immune microenvironment, it is difficult to extend this estimated effect on individual cell types to the emergent, holistic response in complex tumor tissues. Modeling provides an important tool to estimate the influence of therapy on the overall tumor ecosystem and facilitates identification of mechanisms of therapeutic response and resistance that emerge through the interactions of different cell types[73]. To explore these dynamics, we performed a PubMed search to curate rules of cell-cell interactions implicated in modulating response to diverse therapies and input them to PhysiCell using our defined grammar[79-82]. This simulation shows T cell activation in response to tumor contact.

The previous example included a model of virtual macrophages with a continuous M1-M2 axis of behavior[83]. We now show an alternative model with discrete M0, M1, and M2 states, as well as more developed maturation of naïve $CD8^+$ T cells into activated $CD8^+$ T cells, in the context of an early tumor-immune microenvironment[84]. The rules also include observations about the functional interactions between these cell types via IL-10 and IFNG. This model of immune invasion shows the establishment of permissive, resident macrophages simultaneously with the initiation of an adaptive response.

We used these rules to simulate 5 days of growth of a 2-D tumor in a virtual environment of 38 mmHg oxygenation (physioxic conditions[65]), starting with 1000 viable tumor cells, surrounded by a ring comprised of 200 Naive $CD8^+$ T cells and 200 M0 Macrophages (**Fig. 5**). These results show a highly successful anti-tumor response in terms of T cell activation and expansion, but complete clearance of the tumor core is not achieved by the simulation endpoint. Moreover, as the simulation progresses, a population of M2-like, tumor-protective macrophages emerge, and the number of naïve cells wanes as they activate and expand into effector T cells. The M2 macrophages secrete an anti-inflammatory factor, deterring

T cells and protecting the remaining tumor from complete destruction through the duration of the simulation.

[Fig. 5]

A full copy the automated model annotation can be found in the **Supplementary Materials**, while the code is included in the GitHub repository as **example4**.

Cell Hypothesis Rules (detailed)

In tumor cells:
oxygen increases cycle entry from 0 towards 0.00072 with a Hill response, with half-max 21.5 and Hill power 4.
pressure decreases cycle entry from 0 towards 0 with a Hill response, with half-max 1 and Hill power 4.
oxygen decreases necrosis from 0.0028 towards 0 with a Hill response, with half-max 3.75 and Hill power 8.
damage increases apoptosis from 7.2e-05 towards 0.072 with a Hill response, with half-max 180 and Hill power 2.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.
IFN-gamma decreases migration speed from 0.5 towards 0 with a Hill response, with half-max 0.25 and Hill power 2.

In M0 macrophage cells:
contact with dead cell increases transform to M1 macrophage from 0 towards 0.05 with a Hill response, with half-max 0.1 and Hill power 10.
contact with dead cell decreases migration speed from 1 towards 0.1 with a Hill response, with half-max 0.1 and Hill power 4.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In M1 macrophage cells:
contact with dead cell decreases migration speed from 1 towards 0.1 with a Hill response, with half-max 0.1 and Hill power 4.
oxygen decreases transform to M2 macrophage from 0.01 towards 0 with a Hill response, with half-max 5 and Hill power 4.
IFN-gamma increases cycle entry from 7.2e-05 towards 0.00036 with a Hill response, with half-max 0.25 and Hill power 2.
IFN-gamma increases phagocytose dead cell from 0.01 towards 0.05 with a Hill response, with half-max 0.25 and Hill power 2.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In M2 macrophage cells:
contact with dead cell decreases migration speed from 1 towards 0.1 with a Hill response, with half-max 0.1 and Hill power 4.
IFN-gamma decreases cycle entry from 7.2e-05 towards 0 with a Hill response, with half-max 0.25 and Hill power 2.
IFN-gamma increases phagocytose dead cell from 0.01 towards 0.05 with a Hill response, with half-max 0.25 and Hill power 2.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In naive T cell cells:
IL-10 decreases transform to CD8 T cell from 0.001 towards 0 with a Hill response, with half-max 0.25 and Hill power 2.
IFN-gamma increases transform to CD8 T cell from 0.001 towards 0.01 with a Hill response, with half-max 0.25 and Hill power 2.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In CD8 T cell cells:
IFN-gamma increases cycle entry from 7.2e-05 towards 0.00041 with a Hill response, with half-max 0.25 and Hill power 2.
IL-10 decreases attack tumor from 0.01 towards 0 with a Hill response, with half-max 0.25 and Hill power 2.
IL-10 decreases migration speed from 1 towards 0.1 with a Hill response, with half-max 0.25 and Hill power 2.
contact with tumor decreases migration speed from 1 towards 0.1 with a Hill response, with half-max 0.1 and Hill power 2.
IL-10 increases transform to exhausted T cell from 0 towards 0.005 with a Hill response, with half-max 0.25 and Hill power 4.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In exhausted T cell cells:
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

Example 5: Using experimental insight to model combination immune-targeted therapies in the pancreatic cancer microenvironment

While the literature contains a wealth of prior knowledge of cellular roles in cancer, distilling cell behaviors into quantifiable networks for modeling previously required extensive manual curation. New single-cell datasets are uncovering novel cellular mechanisms of tumors in the microenvironment and their response to therapy. As we showed in example 2, data-driven cell-cell interaction rules derived from these datasets provide the opportunity to refine knowledge-driven or literature-based rules in the context of

individual tumors. Integrating genomics datasets into ABMs can further personalize virtual prediction of therapeutic outcomes.

An ongoing platform neoadjuvant clinical trial[85] is evaluating combination therapies in human pancreatic cancer to rewire the immune system through systematic, rational combination therapeutic strategies. Building on the essential tumor-immune paradigm demonstrated in preceding examples, we sought to model and understand behaviors observed from human pancreatic tumor biospecimens.

Multi-omics data from our previous neoadjuvant clinical trial of GVAX and Nivolimab demonstrated that immunotherapy activates chemokine signaling in CD4 T cells signaling to CD8 T cells to promote changes in lymphocyte chemotaxis[86]. In this example, we (computationally) simulate therapy for a cohort of patients using our behavior rules and initial cell numbers derived from the cellular proportions in our reference atlas of single-cell RNA-seq data from untreated pancreatic tumors[87]. These simulations demonstrate inter-patient heterogeneity of treatment effects, which we hypothesize resulted from differences in initial immune cell ratios in the simulated tumor microenvironment **(Fig. 6b)**. Indeed, when comparing predicted tumor cell killing between baseline tissues, tumor growth correlates negatively with baseline CD8+ T cell content, and is attenuated when simulated therapy is applied **(Fig. 6c)**. These results demonstrate a correlation between baseline CD8+ T cell numbers and tumor volume, consistent with our previous analysis of overall survival in the PDAC immunotherapy clinical trial[86].

[Fig. 6]

The limited responsiveness of PDAC to immunotherapy has led to new combination therapeutic strategies to further enhance T cell function. To test the mechanisms whereby different therapies can alter T cells, we next simulate monotherapy and combination therapy on a single patient's hypothetical pancreatic microenvironment.

Our group has recently observed tumor response associated with prolonged disease-free survival through the addition of CD137 agonist therapy to a combination of an irradiated, granulocyte-macrophage colony-stimulating factor (GM-CSF)–secreting, allogeneic PDAC vaccine (GVAX)[88] and immune checkpoint inhibitor therapy[85], consistent with the promise shown by this approach in preclinical studies. Previously, our ligand-receptor analysis of the PDAC atlas single-cell data further predicted increased interferon expression in CD137 high CD8 T cells[86]. We hypothesize that this difference in the cellular regulatory network will increase CD8+ T cells ability to infiltrate the tumor and maintain an anti-tumor state, impacting immunotherapy outcome when the proportion of CD137 high CD8 T cells is increased through the addition of an agonist therapy. We can represent these genomics-driven rules in PhysiCell to inform a model to predict and explain the impact of therapeutic combination between GVAX, Nivolumab (anti PD-1), and Urelumab (CD137 agonist) therapies.

We used our genomics observations of cell-cell interactions to motivate the rules of cellular agents in PhysiCell and built four parallel models of tumor-immune interactions. To reflect the dense immunosuppressive nature of pancreatic tumors, the models are initialized as having recruited resident suppressors with dense internal macrophages and external T cells, excluded from the tumor. As inferred in the bulk RNA-seq data, PD-1lo CD4+ T cells secrete chemokines that can attract CD8+ T cells to cause a bias in migration. Based on the single-cell RNA-seq-derived rules, CD137hi CD8+ T cells further secrete pro-inflammatory factor, allowing for further motility bias (i.e., more directed migration) and tumor sensing. We model treatment by including CD4+ T cells following the vaccine treatment, and further altering the proportion of PD-1lo cells to simulate Nivolumab therapy and CD137hi cells to simulate the impact of Urelumab. In contrast to example 2, fibroblasts were excluded from our model due to their underrepresentation in the reference bulk and single-cell datasets from which these rules of cell-cell interactions were derived.

In the first simulation, we model tumor-immune interactions occurring after immune priming via GVAX cancer vaccine alone; this is the first row in **Fig. 6e**. In the second, we simulate a combination of GVAX and anti-PD-1 (Nivolumab), showing increased killing but poor and uneven infiltration, with some tumor foci completely escaping immune detection; see the second row in **Fig. 6e**. In the third simulation (third row in **Fig. 6e**), we model the combination of GVAX and CD137 agonist, and shows superior tumor infiltration by T cells but inefficient killing. And in the fourth simulation (fourth row in **Fig. 6e**), we modeled the combination of GVAX, Checkpoint, and CD137 agonist, which in our model resulted in effective tumor clearance through effective infiltration and killing **(Fig. 6c)**.

When we apply this approach, we observe that immune cells aggregate or "clump" together after vaccine therapy, which limits the effective infiltration of T cells throughout the tumor area and subsequent success of tumor killing. This observation is consistent with clinical observations of the formation of proximal sites of lymphoid activation (tertiary lymphoid structures or TLS) observed after neoadjuvant vaccination in PDAC[89]. Though there is clear T cell priming and expansion occurring at the TLS, presumably comprising anti-tumor neoantigen or vaccine neoantigen-specific T cells, immune infiltration into the tumor mass is limited in both our mathematical model and clinical studies. The addition of anti-PD-1 blocking antibody and CD137 agonist antibody allows T cells to effectively traffic into the tumor, leading to enhanced tumor cell killing in the triple combination in our mathematical model. Our early "toy" models of tumor-T cell interactions—which did not incorporate further immune cell communication—also predicted T cell aggregation, inconsistent infiltration, and poor response[90,91]; this further highlights the key importance of immune cell coordination across time and space for effective immunotherapeutic responses, and shows the utility of the grammar and models in elucidating these key communications.

A full copy the automated model annotation can be found in the **Supplementary Materials**, while the code is included in the GitHub repository as **example5_gvax** (case 1), **example5_gvax_ipi** (case 2), **example5_gvax_uru** (case 3), and **example5_gvax_ipi_uru** (case 4).

Cell Hypothesis Rules (detailed)

In PD-L1lo_tumor cells:
oxygen increases cycle entry from 0 towards 0.00072 with a Hill response, with half-max 21.5 and Hill power 4.
pressure decreases cycle entry from 0 towards 0 with a Hill response, with half-max 1 and Hill power 4.
oxygen decreases necrosis from 0.0028 towards 0 with a Hill response, with half-max 3.75 and Hill power 8.
damage increases apoptosis from 7.2e-05 towards 0.072 with a Hill response, with half-max 180 and Hill power 2.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In PD-L1hi_tumor cells:
oxygen increases cycle entry from 0 towards 0.00072 with a Hill response, with half-max 21.5 and Hill power 4.
pressure decreases cycle entry from 0 towards 0 with a Hill response, with half-max 1 and Hill power 4.
oxygen decreases necrosis from 0.0028 towards 0 with a Hill response, with half-max 3.75 and Hill power 8.
damage increases apoptosis from 7.2e-05 towards 0.072 with a Hill response, with half-max 180 and Hill power 2.
dead increases debris secretion from 0 towards 0.017 with a Hill response, with half-max 0.1 and Hill power 10. Rule applies to dead cells.

In macrophage cells:
oxygen increases pro-inflammatory factor secretion from 0 towards 1 with a Hill response, with half-max 5 and Hill power 4.
oxygen decreases anti-inflammatory factor secretion from 10 towards 0 with a Hill response, with half-max 5 and Hill power 4.

In PD-1hi_CD137lo_CD8_Tcell cells:
contact with PD-L1hi_tumor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 2.

In PD-1lo_CD137lo_CD8_Tcell cells:
anti-inflammatory factor decreases attack PD-L1hi_tumor from 1e-06 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
pro-inflammatory factor increases attack PD-L1hi_tumor from 1e-06 towards 1 with a Hill response, with half-max 0.5 and Hill power 8.
anti-inflammatory factor decreases attack PD-L1lo_tumor from 1e-05 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
pro-inflammatory factor increases attack PD-L1lo_tumor from 1e-05 towards 1 with a Hill response, with half-max 0.5 and Hill power 8.
anti-inflammatory factor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
contact with PD-L1hi_tumor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 2.

In PD-1hi_CD137hi_CD8_Tcell cells:
contact with PD-L1hi_tumor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 2.

In PD-1lo_CD137hi_CD8_Tcell cells:
anti-inflammatory factor decreases attack PD-L1hi_tumor from 1e-06 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
pro-inflammatory factor increases attack PD-L1hi_tumor from 1e-06 towards 1 with a Hill response, with half-max 0.5 and Hill power 8.
anti-inflammatory factor decreases attack PD-L1lo_tumor from 1e-05 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
pro-inflammatory factor increases attack PD-L1lo_tumor from 1e-05 towards 1 with a Hill response, with half-max 0.5 and Hill power 8.
anti-inflammatory factor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.
contact with PD-L1hi_tumor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 2.

In PD-1hi_CD4_Tcell cells:
anti-inflammatory factor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.

In PD-1lo_CD4_Tcell cells:
anti-inflammatory factor decreases migration speed from 1 towards 0 with a Hill response, with half-max 0.5 and Hill power 8.

DISCUSSION

The real-world limitations inherent to human-focused research, especially for clinical pathology samples and population- or trial-level data, do not exist *in silico*. Thus, scientists are turning to computational models to guide and supplement lab experiments. For example, the NCI digital twins initiative aims to develop models of patient tumors to predict which therapies will most benefit each individual[33,35]. Mathematical models allow investigators to simulate many replicates of their system's behavior under different sets of conditions. The ability to perform large numbers of replicates and numerous iterations cheaply and easily maximizes the chance of capturing extremely rare critical events, something key to those who study cancer etiology. An agent-based model is an abstraction that can be run thousands or millions of times, and whose parameters and in-built hypotheses are all readily modifiable by the user. Here, we introduced a new conceptual framing (a grammar) for specifying cell behavior hypotheses, which can systemize and facilitate our thinking of how cells interact to drive tissue ecosystems. The grammar made it possible to introduce new capabilities in the PhysiCell Agent Based Modeling (ABM) framework that simplify the workflow for investigators to build a simulation of their experimental system.

The hypothesis grammar gives the ability to encode complex behaviors and responses to signals into model agents via a single line of human readable text. Previously, custom hand-written code and a high level of technical knowledge were required to implement even basic models. Now, a scientist can easily create, specify, and modify cell behaviors without writing code or hand-edited markup languages, particularly when used in combination of graphical and cloud-based modeling frameworks. In this implementation, it is simple to modulate and apply behaviors to different agents in the system and to build hypotheses into the model. Any multi-cell model system can be translated into an ABM. Further, PhysiCell is open-source, built and maintained by a large community of mathematicians, statisticians, and scientists. From their efforts and from the user community, a vast amount of knowledge is encoded in the system at baseline; however, everything is completely customizable, extensible, and modifiable. Even when hypotheses are not fully executed in simulations, the cell behavior grammar affords an opportunity to systematically collect, annotate, curate, and grow our knowledge of behaviors of many cell types.

Additionally, we have added the ability to directly translate parameters from spatial transcriptomics data—namely spatial location and annotated identity—to an ABM, and therefore use spatial data to initialize an ABM directly. Thus, models can now exactly match the tissue structure and transcriptional profile of samples directly. Spatial relationships between cells matter immensely and spatial relationships can have a great impact on simulated (and real) outcomes. This strong dependence of many cancer systems and ABM trajectories on initial conditions can complicate model inquiry and impact critical system behaviors

and model parameters obtained through inference; by leveraging robust single-cell spatial transcriptomics tissue profiles as initial conditions in the digital modeling stage, the hypothesis-driven rules modeling paradigm is grounded in precise referential data but also offers a path to both stronger model inquiry and more confident mathematical inference. These models nonetheless still require annotation of a finite number of agents identified in spatial molecular data, often annotating cells into broad phenotypes and abstracting cellular subtypes. Future work must evaluate the sensitivity of models to the granularity of cellular phenotypes in these high-throughput datasets and accurate inference of parameters in the resulting higher-dimensional models. Moreover, future work must facilitate simple and robust identification of cell types and model parameterization, starting by identifying which parameters are most critical, and curating best parameter estimates for community reuse.

We demonstrate a variety of models built using the hypothesis grammar, and two also informed by genomics-derived rules, which we hope will serve as helpful references for users. These examples cover multicellular behavior, demonstrated through the case studies of carcinogenesis and immune response to tumor growth. In some cases, all agents follow the same rules and their fate is decided by the actions of those around them. In other cases, cell agents are acting at cross purposes and actively seeking to outcompete, evade, or hunt and kill each other. We model immune processes such as macrophage plasticity, T cell activation and expansion, antigen recognition, and inflammation. All of these examples are available and can be re-run on any user machine within minutes. While these provide biologically and clinically relevant models to test hypotheses, we note that future work must determine the sensitivity of models to cellular resolution, initial conditions, and model parameters. Metrics to benchmark mathematical models both qualitatively and quantitatively are essential to fully leverage these models to predict personalized tumor conditions and to empower virtual clinical trials.

In future work, we hope to further refine the hypothesis grammar to expand its usability and flexibility. We are considering incorporating the ability to specify "wild card" rules (*) and other special cases using regular expression-type syntax, as well as negation symbols (e.g., "low oxygen" or "hypoxia") that can simplify the examples presented in this paper. However, we first wish to consult with the wider user community to receive input and feedback on such extensions. Moreover, emerging large language models (LLMs) such as Chat-GPT may facilitate "translation" of familiar language (e.g., "fibrosis") into the smaller set of symbols in the current grammar. The language currently treats all statements as independent (inclusive OR), but we may need additional language operators to signify relationships between rules such as AND or REQUIRES. Other generalizations and improvements to the forms of response curves, consensus process models, and default parameter values are likely to emerge from widespread community use, feedback, and discussion. We envision formal community governance from key stakeholders including experimentalists, clinicians, bioinformaticians, mathematicians, and software developers. We also envision that the hypothesis grammar could become a common language to unite current non-compatible agent-based and multicellular simulation platforms, allowing greater reproducibility.

Additionally, we aspire to establish a repository to collect and curate biological hypothesis statements grouped as digital cell lines[50], enabling users to contribute and share cell behavior statements for future reuse in other models of the same system. This repository can serve as a valuable resource for scientists to draw upon when digitizing their system. Through collaboration and collective contributions, we can foster the growth, expansion, and maintenance of this repository. Community discussion, curated validation data, better leverage of high-throughput computing resources, and uncertainty quantification analyses techniques should lead to better understanding of the reliability and sensitivity of model parameters. We also hope to achieve fuller harmonization with other omics data analysis software and high-throughput data types as we move towards integrating data directly into agent-based models. By achieving this integration, we can enhance the model's capability to incorporate relevant biological information and improve its alignment with the broader computational biology field.

This language framework will be useful to those seeking to build models of multicellular systems, and we are excited to continue to move toward fuller biological completeness and more complete integration with omics data, to increasingly define agent behavior in an automated and data-driven fashion. These advancements expand the functionality, usability, and compatibility of our approach, empowering interdisciplinary researchers in their computational or systems biology endeavors. These advancements expand the functionality, usability, and compatibility of our approach, empowering researchers across disciplines to unlock the full potential of their single-cell data. Armed with this conceptual framing and tools, they can extrapolate beyond single-cell characterizations to predict tissue dynamics, and ultimately perform virtual experiments to steer tissues from disease to health.

ACKNOWLEDGEMENTS

Thanks to Laura Wood and Jennifer Elisseeff for feedback on this study.

FUNDING SOURCES

Funding was provided by P01CA247886 (to EMJ, NZ, LTK, JZ, EJF), K08CA248624 (to NZ), the Lustgarten Foundation 'A Translational Convergence Program of Personalized Immunotherapy for Pancreatic Cancer Patients at Johns Hopkins' (to EMJ, NZ, LTK, JZ, EJF), a Lustgarten Foundation-AACR Career Development Award for Pancreatic Cancer Research, in honor of Ruth Bader Ginsburg (ATK), GI SPORE P50CA062924 (EMJ, EJF), U01CA253403 (EJF, AD, EMJ), U54CA274371 (EJF, AK, DW), U01CA212007 (EJF, LTK), U54CA268083 (DW, AK, PH, EMJ, EJF), R00NS122085 (GSO), NSF 1720625 (RH, YW, PM), the Jayne Koskinas Ted Giovanis Foundation for Health and Policy (RH, MG, IG, J. Metzcar, DG, HR, PM), U01CA232137 (RH, FK, PM), NSF 1818187 (RH, AS, PM), Leidos Biomedical Research contract 75N91019D00024 (RH, HR, PM), and a Luddy Faculty Fellowship (HR, PM).

METHODS

PhysiCell agent-based modeling framework

PhysiCell[42] is an open source, agent-based modeling framework written in C++ that can run on a broad variety of desktop platforms, in the cloud[92], and on high performance computing resources[90,91,93]. PhysiCell simulates each cell as an agent with lattice-free position and volume, individual birth and death rates, and motion driven by the balance of mechanical forces and biased random migration. In more recent versions of PhysiCell, agents can also interact with built-in models of phagocytosis, effector attack, fusion, and elastic cell-cell adhesion. PhysiCell is coupled to a reaction-diffusion solver (BioFVM[94]) that models secretion and uptake (consumption) of diffusible factors by individual cell agents at their individual positions, as well as diffusion and decay of these substrates through extracellular spaces. PhysiCell bundles its key cell behavioral parameters as a *phenotype object* for simpler representation. Modelers simulate biological hypotheses by writing custom C++ functions that dynamically vary the cell agent's phenotype parameters based on conditions at the cell's position, such as contact with other cells, mechanical pressure, and concentrations and gradients of signaling factors.

Cell behaviors

To build this grammar, we require clear abstractions of key cell behaviors that frequently occur in multicellular observations and corresponding reference models. In this context, a ***cell behavior*** is a cell-scale process, such as cycling, death, or phagocytosis. Generally, each behavior can be represented by a small number of continuous phenotypic parameters, describing the rate, magnitude, or frequency of the behavior. In earlier work, Sluka et al. developed the Cell Behavior Ontology (CBO)[49] as a controlled vocabulary

of individual cell behaviors. More recently, we worked with a multidisciplinary coalition to extend and structure behaviors from the CBO and other sources into MultiCellDS[50] (multicellular data standard). In particular, this work defined a *behavioral cell phenotype* that collects of biophysical characterizations of a cell's behavior, organized hierarchically by function: cycling, death, volume, mechanics, secretion (including uptake), and motility. Since releasing MultiCellDS as a preprint, we have tested this approach to cell behavior through a variety of agent-based simulation and modeling projects[42,57,58,90,91,95-100]. Based upon recent immunologic modeling work[96-99], we extended phenotype to include cell-cell interactions (phagocytosis, effector attack, and fusion), as well as transformations between cell types (e.g., differentiation, transdifferentiation, and other state changes that persist even when exogenous signals are removed).

The grammar's supported cell behaviors and key biophysical parameters are summarized in **Table 1**. See the **Supplementary Information** for a full description of these cell behaviors, including reference model implementation details in the PhysiCell framework.

Behavior	Main controllable behavior parameters
Cycling	Exit rates from each cycle phase
Death	Apoptotic (controlled / non-inflammatory) death rate Necrotic (uncontrolled / inflammatory) death rate
Secretion and Uptake	Secretion rates and targets (for each diffusible extracellular substrate) Uptake (consumption) rates (for each diffusible extracellular substrate) Generalized net export rates (for each diffusible extracellular substrate)
Migration and Chemotaxis	Migration speed, bias, and persistence time Chemotactic sensitivities (for each diffusible extracellular substrate)
Cell-cell adhesion	Interaction distances, strength Attachment and detachment rates Maximum number of cellular adhesions Adhesion affinities (to each cell type)
Resistance to deformation	strength of cell repulsion
Transformation (changing cell type)	Rate of changing to each cell type
Fusion	Rate of fusing with each cell type
Phagocytosis (predation or ingestion)	Rate of phagocytosing dead cells Rate of phagocytosing each live cell type
(Effector) attack	Attack rates (for each cell type) Rate of damaging attacked cells Immunogenicity (to each potential attacking cell type)
Other	User-defined custom cell behavior parameters

TABLE 1. Key cell behaviors included in the grammar, with a summary of controllable behavior parameters.

Signals

Signals are (typically exogeneous but sometimes internal) stimuli or information that can be interpreted by a cell to drive behavioral or state changes. In the context of mathematical modeling, signals are inputs to constitutive laws or agent rules. We broadly surveyed mathematical and biological models from cancer biology[40,61,101-110], tissue morphogenesis[101,111-115], immunology[40,96-98,116,117], and microbial ecosystems[118,119], to generalize classes of inputs to cell behavioral rules, generally including chemical factors, mechanical cues, cell volume (e.g., for volume-based cycle checkpoints), physical contact with cells, live/dead status, current simulation time (for use in triggering events), and accumulated damage (e.g., from effector attack[120-122]). The signals are summarized in **Table 2**. See the **Supplementary Information** for a full description.

Signal types symbol	accessible variables
Diffusible chemical substrates	extracellular concentration and gradient (for each diffusible substrate) total internalized substrate (for each diffusible substrate)
Cell mechanics/physics	total cell volume, mechanical pressure
Contact	Total numbers of live and dead cells in contact Total number of each live cell type in contact, contact with basement membrane
Effector attack	Total cell damage and accumulated attack time by effector cells
Death	Status as dead. Finer-grained status as apoptotic or necrotic
Other	Current elapsed simulation time, access to user-defined parameters

Table 2. A summary of the signals that can be used to drive behavioral or state changes in cells.

Coordinate transforms and spatial transcriptomics integration

To simulate cellular profiles based on initial conditions defined from VISIUM single-cell spatial transcriptomics data in PhysiCell, we interpret a given tissue profile $\mathcal{A}$—a collection of the form $(x_i, y_i) \in \mathbb{R}^2$ arranged in a triangular lattice with uniform spacing $\mathrm{d} = 71.000\ \mu m$—as an $(n \times 2)$ $(n \times 2)$ matrix.

$$X = \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix} \in \mathbb{R}^{(n\times 2)}.$$

Choosing a prospective center location $X_0 = (x_0, y_0) \in \mathbb{R}^2$ and using a hypothesized average cell diameter d_f in the tissue, we apply a linear transformation $T_{(X_0, d_f)}$ on X to produce a scaled cell coordinate matrix $X_f = T_{(X_0, d_f)}(X)$ describing a digital tissue with center $(0,0) = T_{(X_0, d_f)}(X_0)$ and ideal lattice spacing on the order of d_f to encode cell signals and behaviors in the PhysiCell stage at scale.

Example:

To generate the initial cell coordinate matrix $X_f$used in Example 2, we start with a collection $\mathcal{A}$ and the corresponding matrix $X \in \mathbb{R}^{(n\times 2)}$(as above) of coordinates of the form $(x_i, y_i) \in \mathbb{R}^2$ from the spatial transcriptomics data, which contain a subset $\mathcal{H} \subset \mathcal{A}$ of n_h coordinates of agents of type other_cells in the main tumor mass. We compute a shift

$$(x_k, y_k) = \sum_{\vec{x} \in \mathcal{H}} \left(\frac{\vec{x}}{n_h} \right)$$

representing the coordinate-wise center of the main tissue region and the corresponding matrix translation

$$X_0 = (x_k, y_k)^n = \begin{bmatrix} x_k & y_k \\ \vdots & \vdots \\ x_k & y_k \end{bmatrix} \in \mathbb{R}^{(n\times 2)}.$$

The transformed coordinate matrix X_f then takes the form

$$X_f = K(X - X_0),$$

where K is the $(n \times n)$ scaling matrix of the form

$$K = s \cdot \operatorname{diag}\left(\frac{d_f}{d}, \ldots, \frac{d_f}{d} \right)$$

determined by a global spacing bias $s \approx 0.97$, hypothesized average cell diameter $d_f = 16.825\ \mu m$, and VISIUM spacing $d = 71.000\ \mu m$ from above.

Spatial transcriptomics data for example 2

Two resected pancreatic lesions were subjected to the commercial Visium spatial transcriptomics sequencing FFPE protocol. Slides were stained with H&E and imaged prior to RNA extraction, and image analysis was performed in parallel with transcriptomic analysis. An artificial intelligence method for annotation of pancreatic tumor tissue regions called CODA[68] was used to annotate acinar cells, islet cells, smooth muscle cells, and the distribution of collagen. This method was also used to distinguish normal ductal, neoplastic, and tumor cells from the H&E imaging, which were further visually confirmed by a pathologist (E.D.T.). Spots with greater than 70% purity of ductal cells were further annotated to assign agent types for the associated tumor and normal cells in each spot. For this annotation, we used our transfer learning method ProjectR[123] version 1.8.0 to distinguish proliferative signaling (modeled as an epithelial phenotype) from co-occurrence of EMT and inflammatory signaling (modeled as the mesenchymal phenotype) as defined in CoGAPS non-negative matrix factorization analysis of our reference scRNA-seq atlas of PDAC tumors using methods described previously[67,87,124]. To locate fibroblasts, Seurat version 4.1.0 was used to compute module scores from a pan-CAF gene signature as described previously[67]. Two python scripts were used to translate from annotated ST data into a PhysiCell-readable csv file containing coordinates and categorical cell types available in the github repository as extract_visium_coordinates.py and pull_physicell_init.py. Submission of spatial transcriptomics data to dbGAP is in progress.

[125,126]Single-cell RNA-seq PDAC atlas data as a reference dataset for example 5

This example uses our PDAC single-cell RNA-seq atlas data to further define immune cell subtypes in reference tumors [127,128]. Cell-cell communication analysis was performed using the Domino package[129-131]. To determine the by-tissue cell counts, the single-cell RNA sequencing data was preprocessed, clustered, and annotated using the Seurat R package[132]. Cell identity clusters of interest ("Activated_CD4", "B cell", "CD4", "CD8", "Effector_CD8", "Epithelial_cancer", "Macrophage", "Mast", "Neutrophil", "NK/CTL", "T cell", "Treg cell") were then thresholded based on median normalized expression of genes of interest (here CD274/PD-L1, PDCD1/PD-1, TNFRSF9/CD137) and the number of cells falling into lo/hi categories were reported as described previously[86]. These cell numbers were then used to initialize the PhysiCell model in example 5, while Domino analysis informed the model rules. Whereas the cell phenotypes from the single-cell PDAC atlas data for example 2 are based on data from GSE155698[128] and CRA001160, example 5 uses only the former dataset due the immune enrichment of this sample cohort[127].

Code availability

PhysiCell Version 1.12.0[133] and later includes a full reference implementation of the grammar and grammar-based simulation modeling. The specific modeling examples presented in the results section presented below are available on GitHub at https://github.com/physicell-models/grammar_samples. To get a list of all the example models:

```
make list-user-projects
```

To load and compile an example named **myproject**, use:

```
make load PROJ=myproject && make
```

Similar to our prior work to create cloud-based training materials[95] and cloud-based model dissemination[92], and inspired by other recent advances on "zero-install" models[134], we have created a cloud-based version[135] of PhysiCell based on the nanoHUB platform[136]. This cloud implementation allows scientists to create, execute, visualize, and explore grammar-based models interactively in a web browser, without need for programming expertise or software setup. (See documentation and training materials in the

supplementary information). The cloud-hosted model is available at https://nanohub.org/tools/pcstudio. Alternatively, scientists can download the latest release of the PhysiCell Studio desktop application at https://github.com/PhysiCell-Tools/PhysiCell-Studio/releases. Assuming an executable model has been compiled, the Studio allows interactive creation and editing of rules (**Fig. 7**), running a simulation, and visualizing results. Refer to the Studio user guide at https://github.com/rheiland/PhysiCell-Studio/blob/main/user-guide/ for more information.

References

1 Regev, A. *et al.* The Human Cell Atlas. *Elife* **6** (2017). https://doi.org:10.7554/eLife.27041

2 HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187-192 (2019). https://doi.org:10.1038/s41586-019-1629-x

3 Rozenblatt-Rosen, O. *et al.* The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* **181**, 236-249 (2020). https://doi.org:10.1016/j.cell.2020.03.053

4 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). https://doi.org:10.1126/science.1260419

5 Stein-O'Brien, G. L., Le, D. T., Jaffee, E. M., Fertig, E. J. & Zaidi, N. Converging on a Cure: The Roads to Predictive Immunotherapy. *Cancer Discov* **13**, 1053-1057 (2023). https://doi.org:10.1158/2159-8290.CD-23-0277

6 Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**, 547-554 (2019). https://doi.org:10.1038/s41587-019-0071-9

7 Fertig, E. J., Jaffee, E. M., Macklin, P., Stearns, V. & Wang, C. Forecasting cancer: from precision to predictive medicine. *Med (N Y)* **2**, 1004-1010 (2021). https://doi.org:10.1016/j.medj.2021.08.007

8 Ding, J., Sharon, N. & Bar-Joseph, Z. Temporal modelling using single-cell transcriptomics. *Nat Rev Genet* **23**, 355-368 (2022). https://doi.org:10.1038/s41576-021-00444-7

9 Qiu, X. *et al.* Mapping transcriptomic vector fields of single cells. *Cell* **185**, 690-711 e645 (2022). https://doi.org:10.1016/j.cell.2021.12.045

10 Erbe, R., Stein-O'Brien, G. & Fertig, E. J. Transcriptomic forecasting with neural ordinary differential equations. *Patterns (N Y)* **4**, 100793 (2023). https://doi.org:10.1016/j.patter.2023.100793

11 Materi, W. & Wishart, D. S. Computational systems biology in cancer: modeling methods and applications. *Gene Regul Syst Bio* **1**, 91-110 (2007).

12 Metzcar, J., Wang, Y. F., Heiland, R. & Macklin, P. A Review of Cell-Based Computational Modeling in Cancer Biology. *Jco Clin Cancer Info* **3** (2019). https://doi.org:10.1200/Cci.18.00069

13 Vodovotz, Y. & An, G. *Translational Systems Biology*. (Academic Press, 2014).

14 Vodovotz, Y. & An, G. Agent-based models of inflammation in translational systems biology: A decade later. *Wiley Interdiscip Rev Syst Biol Med* **11**, e1460 (2019). https://doi.org:10.1002/wsbm.1460

15 An, G., Mi, Q., Dutta-Moscato, J. & Vodovotz, Y. Agent-based models in translational systems biology. *Wiley Interdiscip Rev Syst Biol Med* **1**, 159-171 (2009). https://doi.org:10.1002/wsbm.45

16 An, G. Introduction of an agent-based multi-scale modular architecture for dynamic knowledge representation of acute inflammation. *Theoretical Biology and Medical Modelling* **5** (2008). https://doi.org:10.1186/1742-4682-5-11

17 Scott, J. G., Fletcher, A. G., Anderson, A. R. & Maini, P. K. Spatial Metrics of Tumour Vascular Organisation Predict Radiation Efficacy in a Computational Model. *PLoS Comput Biol* **12**, e1004712 (2016). https://doi.org:10.1371/journal.pcbi.1004712

18 Bozic, I. *et al.* Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife* **2**, e00747 (2013). https://doi.org:10.7554/eLife.00747

19 Cavenee, W. K. *et al.* Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305**, 779-784 (1983). https://doi.org:10.1038/305779a0

20 Diaz, L. A., Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540 (2012). https://doi.org:10.1038/nature11219

21 Frank, S. A., Iwasa, Y. & Nowak, M. A. Patterns of cell division and the risk of cancer. *Genetics* **163**, 1527-1532 (2003). https://doi.org:10.1093/genetics/163.4.1527

22 Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci U S A* **105**, 4283-4288 (2008). https://doi.org:10.1073/pnas.0712345105

23 Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820-823 (1971). https://doi.org:10.1073/pnas.68.4.820
24 Komarova, N. L. & Wodarz, D. Drug resistance in cancer: principles of emergence and prevention. *Proc Natl Acad Sci U S A* **102**, 9714-9719 (2005). https://doi.org:10.1073/pnas.0501870102
25 Leder, K. *et al.* Mathematical modeling of PDGF-driven glioblastoma reveals optimized radiation dosing schedules. *Cell* **156**, 603-616 (2014). https://doi.org:10.1016/j.cell.2013.12.029
26 Lenaerts, T., Pacheco, J. M., Traulsen, A. & Dingli, D. Tyrosine kinase inhibitor therapy can cure chronic myeloid leukemia without hitting leukemic stem cells. *Haematologica* **95**, 900-907 (2010). https://doi.org:10.3324/haematol.2009.015271
27 Sanga, S. *et al.* Mathematical modeling of cancer progression and response to chemotherapy. *Expert Rev Anticancer Ther* **6**, 1361-1376 (2006). https://doi.org:10.1586/14737140.6.10.1361
28 Sherratt, J. A. & Nowak, M. A. Oncogenes, anti-oncogenes and the immune response to cancer: a mathematical model. *Proc Biol Sci* **248**, 261-271 (1992). https://doi.org:10.1098/rspb.1992.0071
29 Swanson, K. R., Alvord, E. C., Jr. & Murray, J. D. Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. *Br J Cancer* **86**, 14-18 (2002). https://doi.org:10.1038/sj.bjc.6600021
30 Lu, Y. *et al.* Resolution of tissue signatures of therapy response in patients with recurrent GBM treated with neoadjuvant anti-PD1. *Nat Commun* **12**, 4031 (2021). https://doi.org:10.1038/s41467-021-24293-4
31 Pereira, E. J. *et al.* Sporadic activation of an oxidative stress-dependent NRF2-p53 signaling network in breast epithelial spheroids and premalignancies. *Sci Signal* **13** (2020). https://doi.org:10.1126/scisignal.aba4200
32 Kim, M. *et al.* A protein interaction landscape of breast cancer. *Science* **374**, eabf3066 (2021). https://doi.org:10.1126/science.abf3066
33 Stahlberg, E. A. *et al.* Exploring approaches for predictive cancer patient digital twins: Opportunities for collaboration and innovation. *Front Digit Health* **4**, 1007784 (2022). https://doi.org:10.3389/fdgth.2022.1007784
34 Laubenbacher, R. *et al.* Building digital twins of the human immune system: toward a roadmap. *NPJ Digit Med* **5**, 64 (2022). https://doi.org:10.1038/s41746-022-00610-z
35 Hernandez-Boussard, T. *et al.* Digital twins for predictive oncology will be a paradigm shift for precision cancer care. *Nat Med* **27**, 2065-2066 (2021). https://doi.org:10.1038/s41591-021-01558-5
36 Madhavan, S. *et al.* Envisioning the future of precision oncology trials. *Nat Cancer* **2**, 9-11 (2021). https://doi.org:10.1038/s43018-020-00163-8
37 An, G. In silico experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. *Crit Care Med* **32**, 2050-2060 (2004). https://doi.org:10.1097/01.ccm.0000139707.13729.7d
38 Saint-Pierre, P. & Savy, N. Agent-based modeling in medical research, virtual baseline generator and change in patients' profile issue. *Int J Biostat* (2023). https://doi.org:10.1515/ijb-2022-0112
39 Cosgrove, J. *et al.* Agent-Based Modeling in Systems Pharmacology. *CPT Pharmacometrics Syst Pharmacol* **4**, 615-629 (2015). https://doi.org:10.1002/psp4.12018
40 Norton, K. A., Gong, C., Jamalian, S. & Popel, A. S. Multiscale Agent-Based and Hybrid Modeling of the Tumor Immune Microenvironment. *Processes (Basel)* **7** (2019). https://doi.org:10.3390/pr7010037
41 Rockne, R. C. *et al.* The 2019 mathematical oncology roadmap. *Phys Biol* **16**, 041005 (2019). https://doi.org:10.1088/1478-3975/ab1a09
42 Ghaffarizadeh, A., Heiland, R., Friedman, S. H., Mumenthaler, S. M. & Macklin, P. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. *PLOS Computational Biology* **14**, e1005991 (2018). https://doi.org:10.1371/journal.pcbi.1005991
43 Mirams, G. R. *et al.* Chaste: an open source C++ library for computational physiology and biology. *PLoS Comput Biol* **9**, e1002970 (2013). https://doi.org:10.1371/journal.pcbi.1002970
44 Kang, S., Kahan, S., McDermott, J., Flann, N. & Shmulevich, I. Biocellion: accelerating computer simulation of multicellular biological system models. *Bioinformatics* **30**, 3101-3108 (2014). https://doi.org:10.1093/bioinformatics/btu498

45 Swat, M. H. *et al.* Multi-scale modeling of tissues using CompuCell3D. *Methods Cell Biol* **110**, 325-366 (2012). https://doi.org:10.1016/B978-0-12-388403-9.00013-8
46 Breitwieser, L. *et al.* BioDynaMo: a modular platform for high-performance agent-based simulation. *Bioinformatics* **38**, 453-460 (2022). https://doi.org:10.1093/bioinformatics/btab649
47 Starruss, J., de Back, W., Brusch, L. & Deutsch, A. Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology. *Bioinformatics* **30**, 1331-1332 (2014). https://doi.org:10.1093/bioinformatics/btt772
48 Daub, J. T. & Merks, R. M. Cell-based computational modeling of vascular morphogenesis using Tissue Simulation Toolkit. *Methods Mol Biol* **1214**, 67-127 (2015). https://doi.org:10.1007/978-1-4939-1462-3_6
49 Sluka, J. P. *et al.* The cell behavior ontology: describing the intrinsic biological behaviors of real and model cells seen as active agents. *Bioinformatics* **30**, 2367-2374 (2014). https://doi.org:10.1093/bioinformatics/btu210
50 Friedman, S. H. *et al.* MultiCellDS: a community-developed standard for curating microenvironment-dependent multicellular data. *bioRxiv* ***[preprint]*** **090456** (2016). https://doi.org:10.1101/090456
51 Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25-29 (2000). https://doi.org:10.1038/75556
52 Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224** (2023). https://doi.org:10.1093/genetics/iyad031
53 Börner, K. *et al.* Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nature Cell Biology* **23**, 1117-1128 (2021). https://doi.org:10.1038/s41556-021-00788-6
54 Germann, P., Marin-Riera, M. & Sharpe, J. ya||a: GPU-Powered Spheroid Models for Mesenchyme and Epithelium. *Cell Systems* **8**, 261-266.e263 (2019). https://doi.org:10.1016/j.cels.2019.02.007
55 Bravo, R. R. *et al.* Hybrid Automata Library: A flexible platform for hybrid modeling with real-time visualization. *PLoS Comput Biol* **16**, e1007635 (2020). https://doi.org:10.1371/journal.pcbi.1007635
56 Hoehme, S. & Drasdo, D. A cell-based simulation software for multi-cellular systems. *Bioinformatics* **26**, 2641-2642 (2010). https://doi.org:10.1093/bioinformatics/btq437
57 Rocha, H. L. *et al.* A persistent invasive phenotype in post-hypoxic tumor cells is revealed by fate mapping and computational modeling. *iScience* **24**, 102935 (2021). https://doi.org:10.1016/j.isci.2021.102935
58 Wang, Y. *et al.* Impact of tumor-parenchyma biomechanics on liver metastatic progression: a multi-model approach. *Sci Rep* **11**, 1710 (2021). https://doi.org:10.1038/s41598-020-78780-7
59 Jagiella, N., Muller, B., Muller, M., Vignon-Clementel, I. E. & Drasdo, D. Inferring Growth Control Mechanisms in Growing Multi-cellular Spheroids of NSCLC Cells from Spatial-Temporal Image Data. *PLoS Comput Biol* **12**, e1004412 (2016). https://doi.org:10.1371/journal.pcbi.1004412
60 Hamis, S. *et al.* Combining hypoxia-activated prodrugs and radiotherapy in silico: Impact of treatment scheduling and the intra-tumoural oxygen landscape. *PLoS Comput Biol* **16**, e1008041 (2020). https://doi.org:10.1371/journal.pcbi.1008041
61 Anderson, A. R. A. in *Single-Cell-Based Models in Biology and Medicine Mathematics and Biosciences in Interaction* Ch. Chapter 1, 3-28 (2007).
62 Kingsley, J. L., Costello, J. R., Raghunand, N. & Rejniak, K. A. Bridging cell-scale simulations and radiologic images to explain short-time intratumoral oxygen fluctuations. *PLOS Computational Biology* **17** (2021). https://doi.org:10.1371/journal.pcbi.1009206
63 Szabó, A. & Merks, R. M. H. Cellular Potts Modeling of Tumor Growth, Tumor Invasion, and Tumor Evolution. *Frontiers in Oncology* **3** (2013). https://doi.org:10.3389/fonc.2013.00087
64 Van Liedekerke, P. *et al.* Quantitative cell-based model predicts mechanical stress response of growing tumor spheroids over various growth conditions and cell lines. *PLOS Computational Biology* **15** (2019). https://doi.org:10.1371/journal.pcbi.1006273
65 McKeown, S. R. Defining normoxia, physoxia and hypoxia in tumours-implications for treatment response. *Br J Radiol* **87**, 20130676 (2014). https://doi.org:10.1259/bjr.20130676

66 Kinny-Köster, B. *et al.* Inflammatory Signaling in Pancreatic Cancer Transfers Between a Single-cell RNA Sequencing Atlas and Co-Culture. *bioRxiv [preprint]*, 2022.2007.2014.500096 (2022). https://doi.org:10.1101/2022.07.14.500096
67 Bell, A. T. F. *et al.* PanIN and CAF Transitions in Pancreatic Carcinogenesis Revealed with Spatial Data Integration. *bioRxiv [preprint]* **2022.07.16.500312** (2022). https://doi.org:10.1101/2022.07.16.500312
68 Kiemen, A. L. *et al.* CODA: quantitative 3D reconstruction of large tissues at cellular resolution. *Nat Methods* **19**, 1490-1499 (2022). https://doi.org:10.1038/s41592-022-01650-9
69 Chan-Seng-Yue, M. *et al.* Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat Genet* **52**, 231-240 (2020). https://doi.org:10.1038/s41588-019-0566-9
70 El-Kenawi, A. *et al.* Acidity promotes tumour progression by altering macrophage phenotype in prostate cancer. *Br J Cancer* **121**, 556-566 (2019). https://doi.org:10.1038/s41416-019-0542-2
71 Henze, A. T. & Mazzone, M. The impact of hypoxia on tumor-associated macrophages. *J Clin Invest* **126**, 3672-3679 (2016). https://doi.org:10.1172/JCI84427
72 Leblond, M. M. *et al.* Hypoxia induces macrophage polarization and re-education toward an M2 phenotype in U87 and U251 glioblastoma models. *Oncoimmunology* **5**, e1056442 (2016). https://doi.org:10.1080/2162402X.2015.1056442
73 Waldman, A. D., Fritz, J. M. & Lenardo, M. J. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* **20**, 651-668 (2020). https://doi.org:10.1038/s41577-020-0306-5
74 Schoenfeld, A. J. & Hellmann, M. D. Acquired Resistance to Immune Checkpoint Inhibitors. *Cancer Cell* **37**, 443-455 (2020). https://doi.org:10.1016/j.ccell.2020.03.017
75 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012). https://doi.org:10.1038/nature11003
76 Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570-575 (2012). https://doi.org:10.1038/nature11005
77 Johnson, B. E. *et al.* An omic and multidimensional spatial atlas from serial biopsies of an evolving metastatic breast cancer. *Cell Rep Med* **3**, 100525 (2022). https://doi.org:10.1016/j.xcrm.2022.100525
78 Tatarova, Z. *et al.* A multiplex implantable microdevice assay identifies synergistic combinations of cancer immunotherapies and conventional drugs. *Nat Biotechnol* **40**, 1823-1833 (2022). https://doi.org:10.1038/s41587-022-01379-y
79 Muller, E. *et al.* Toll-Like Receptor Ligands and Interferon-gamma Synergize for Induction of Antitumor M1 Macrophages. *Front Immunol* **8**, 1383 (2017). https://doi.org:10.3389/fimmu.2017.01383
80 Coller, S. P. & Paulnock, D. M. Signaling pathways initiated in macrophages after engagement of type A scavenger receptors. *J Leukoc Biol* **70**, 142-148 (2001).
81 Acosta-Iborra, B. *et al.* Macrophage oxygen sensing modulates antigen presentation and phagocytic functions involving IFN-gamma production through the HIF-1 alpha transcription factor. *J Immunol* **182**, 3155-3164 (2009). https://doi.org:10.4049/jimmunol.0801710
82 Ruffell, B. *et al.* Macrophage IL-10 blocks CD8+ T cell-dependent responses to chemotherapy by suppressing IL-12 expression in intratumoral dendritic cells. *Cancer Cell* **26**, 623-637 (2014). https://doi.org:10.1016/j.ccell.2014.09.006
83 Mantovani, A., Allavena, P., Marchesi, F. & Garlanda, C. Macrophages as tools and targets in cancer therapy. *Nat Rev Drug Discov* **21**, 799-820 (2022). https://doi.org:10.1038/s41573-022-00520-5
84 van der Leun, A. M., Thommen, D. S. & Schumacher, T. N. CD8(+) T cell states in human cancer: insights from single-cell analysis. *Nat Rev Cancer* **20**, 218-232 (2020). https://doi.org:10.1038/s41568-019-0235-4
85 Heumann, T. *et al.* A platform trial of neoadjuvant and adjuvant antitumor vaccination alone or in combination with PD-1 antagonist and CD137 agonist antibodies in patients with resectable pancreatic adenocarcinoma. *Nat Commun* **14**, 3650 (2023). https://doi.org:10.1038/s41467-023-39196-9

86 Li, K. *et al.* Multi-omic analyses of changes in the tumor microenvironment of pancreatic adenocarcinoma following neoadjuvant treatment with anti-PD-1 therapy. *Cancer Cell* (2022). https://doi.org:10.1016/j.ccell.2022.10.001

87 Kinny-Köster, B. *et al.* Inflammatory Signaling in Pancreatic Cancer Transfers Between a Single-cell RNA Sequencing Atlas and Co-Culture. *bioRxiv* ***[preprint]*** **2022.07.14.500096** (2022). https://doi.org:10.1101/2022.07.14.500096

88 Jaffee, E. M. *et al.* Novel allogeneic granulocyte-macrophage colony-stimulating factor-secreting tumor vaccine for pancreatic cancer: a phase I trial of safety and immune activation. *J Clin Oncol* **19**, 145-156 (2001). https://doi.org:10.1200/JCO.2001.19.1.145

89 Lutz, E. R. *et al.* Immunotherapy converts nonimmunogenic pancreatic tumors into immunogenic foci of immune regulation. *Cancer Immunol Res* **2**, 616-631 (2014). https://doi.org:10.1158/2326-6066.CIR-14-0027

90 Ozik, J., Collier, N., Heiland, R., An, G. & Macklin, P. Learning-accelerated discovery of immune-tumour interactions. *Mol Syst Des Eng* **4**, 747-760 (2019). https://doi.org:10.1039/c9me00036d

91 Ozik, J. *et al.* High-throughput cancer hypothesis testing with an integrated PhysiCell-EMEWS workflow. *BMC Bioinformatics* **19**, 483 (2018). https://doi.org:10.1186/s12859-018-2510-x

92 Heiland, R., Mishler, D., Zhang, T., Bower, E. & Macklin, P. xml2jupyter: Mapping parameters between XML and Jupyter widgets. *J Open Source Softw* **4** (2019). https://doi.org:10.21105/joss.01408

93 Saxena, G., Ponce-de-Leon, M., Montagud, A., Vicente Dorca, D. & Valencia, A. in *Computational Methods in Systems Biology Lecture Notes in Computer Science* Ch. Chapter 18, 266-279 (2021).

94 Ghaffarizadeh, A., Friedman, S. H. & Macklin, P. BioFVM: an efficient, parallelized diffusive transport solver for 3-D biological simulations. *Bioinformatics* **32**, 1256-1258 (2016). https://doi.org:10.1093/bioinformatics/btv730

95 Sundus, A. *et al.* PhysiCell training apps: Cloud hosted open-source apps to learn cell-based simulation software. *bioRxiv* ***[preprint]*** **10.1101/2022.06.24.497566** (2022). https://doi.org:10.1101/2022.06.24.497566

96 Jenner, A. L. *et al.* Agent-based computational modeling of glioblastoma predicts that stromal density is central to oncolytic virus efficacy. *iScience* **25**, 104395 (2022). https://doi.org:10.1016/j.isci.2022.104395

97 Islam, M. A., Getz, M., Macklin, P. & Versypt, A. N. F. An agent-based modeling approach for lung fibrosis in response to COVID-19. *bioRxiv* ***[preprint]*** (2022). https://doi.org:10.1101/2022.10.03.510677

98 Getz, M. *et al.* Iterative community-driven development of a SARS-CoV-2 tissue simulator. *bioRxiv* ***[preprint]*** (2021). https://doi.org:10.1101/2020.04.02.019075

99 Risner, K. H. *et al.* Maraviroc inhibits SARS-CoV-2 multiplication and s-protein mediated cell fusion in cell culture. *bioRxiv* ***[preprint]*** (2020). https://doi.org:10.1101/2020.08.12.246389

100 Letort, G. *et al.* PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. *Bioinformatics* **35**, 1188-1196 (2019). https://doi.org:10.1093/bioinformatics/bty766

101 Hoehme, S., Friebel, A., Hammad, S., Drasdo, D. & Hengstler, J. G. Creation of Three-Dimensional Liver Tissue Models from Experimental Images for Systems Medicine. *Methods Mol Biol* **1506**, 319-362 (2017). https://doi.org:10.1007/978-1-4939-6506-9_22

102 Finley, S. D. & Popel, A. S. Effect of tumor microenvironment on tumor VEGF during anti-VEGF treatment: systems biology predictions. *J Natl Cancer Inst* **105**, 802-811 (2013). https://doi.org:10.1093/jnci/djt093

103 Swan, A., Hillen, T., Bowman, J. C. & Murtha, A. D. A Patient-Specific Anisotropic Diffusion Model for Brain Tumour Spread. *Bull Math Biol* **80**, 1259-1291 (2018). https://doi.org:10.1007/s11538-017-0271-8

104 Chaplain, M. A., Graziano, L. & Preziosi, L. Mathematical modelling of the loss of tissue compression responsiveness and its role in solid tumour development. *Math Med Biol* **23**, 197-229 (2006). https://doi.org:10.1093/imammb/dql009

105 Alarcon, T., Byrne, H. M. & Maini, P. K. A cellular automaton model for tumour growth in inhomogeneous environment. *J Theor Biol* **225**, 257-274 (2003). https://doi.org:10.1016/s0022-5193(03)00244-3
106 Scott, J. G., Basanta, D., Anderson, A. R. & Gerlee, P. A mathematical model of tumour self-seeding reveals secondary metastatic deposits as drivers of primary tumour growth. *J R Soc Interface* **10**, 20130011 (2013). https://doi.org:10.1098/rsif.2013.0011
107 Kaznatcheev, A., Vander Velde, R., Scott, J. G. & Basanta, D. Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *Br J Cancer* **116**, 785-792 (2017). https://doi.org:10.1038/bjc.2017.5
108 Poleszczuk, J., Hahnfeldt, P. & Enderling, H. Biphasic modulation of cancer stem cell-driven solid tumour dynamics in response to reactivated replicative senescence. *Cell Prolif* **47**, 267-276 (2014). https://doi.org:10.1111/cpr.12101
109 Powathil, G. G., Adamson, D. J. & Chaplain, M. A. Towards predicting the response of a solid tumour to chemotherapy and radiotherapy treatments: clinical insights from a computational model. *PLoS Comput Biol* **9**, e1003120 (2013). https://doi.org:10.1371/journal.pcbi.1003120
110 Hamis, S., Nithiarasu, P. & Powathil, G. G. What does not kill a tumour may make it stronger: In silico insights into chemotherapeutic drug resistance. *J Theor Biol* **454**, 253-267 (2018). https://doi.org:10.1016/j.jtbi.2018.06.014
111 Fortuna, I. *et al.* CompuCell3D Simulations Reproduce Mesenchymal Cell Migration on Flat Substrates. *Biophys J* **118**, 2801-2815 (2020). https://doi.org:10.1016/j.bpj.2020.04.024
112 Dunn, S. J. *et al.* A two-dimensional model of the colonic crypt accounting for the role of the basement membrane and pericryptal fibroblast sheath. *PLoS Comput Biol* **8**, e1002515 (2012). https://doi.org:10.1371/journal.pcbi.1002515
113 Glen, C. M., Kemp, M. L. & Voit, E. O. Agent-based modeling of morphogenetic systems: Advantages and challenges. *PLoS Comput Biol* **15**, e1006577 (2019). https://doi.org:10.1371/journal.pcbi.1006577
114 Schubert, M. *et al.* Computational modelling unveils how epiblast remodelling and positioning rely on trophectoderm morphogenesis during mouse implantation. *Plos One* **16** (2021). https://doi.org:10.1371/journal.pone.0254763
115 Camacho-Gómez, D., García-Aznar, J. M. & Gómez-Benito, M. J. A 3D multi-agent-based model for lumen morphogenesis: the role of the biophysical properties of the extracellular matrix. *Engineering with Computers* **38**, 4135-4149 (2022). https://doi.org:10.1007/s00366-022-01654-1
116 Cess, C. G. & Finley, S. D. Multi-scale modeling of macrophage—T cell interactions within the tumor microenvironment. *PLOS Computational Biology* **16** (2020). https://doi.org:10.1371/journal.pcbi.1008519
117 Ruiz-Martinez, A. *et al.* Simulations of tumor growth and response to immunotherapy by coupling a spatial agent-based model with a whole-patient quantitative systems pharmacology model. *PLoS Comput Biol* **18**, e1010254 (2022). https://doi.org:10.1371/journal.pcbi.1010254
118 Ni, C. & Lu, T. Individual-Based Modeling of Spatial Dynamics of Chemotactic Microbial Populations. *ACS Synth Biol* **11**, 3714-3723 (2022). https://doi.org:10.1021/acssynbio.2c00322
119 Hellweger, F. L. & Bucci, V. A bunch of tiny individuals—Individual-based modeling for microbes. *Ecological Modelling* **220**, 8-22 (2009). https://doi.org:10.1016/j.ecolmodel.2008.09.004
120 Osinska, I., Popko, K. & Demkow, U. Perforin: an important player in immune response. *Cent Eur J Immunol* **39**, 109-115 (2014). https://doi.org:10.5114/ceji.2014.42135
121 Farhood, B., Najafi, M. & Mortezaee, K. CD8(+) cytotoxic T lymphocytes in cancer immunotherapy: A review. *J Cell Physiol* **234**, 8509-8521 (2019). https://doi.org:10.1002/jcp.27782
122 Raskov, H., Orhan, A., Christensen, J. P. & Gogenur, I. Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy. *Br J Cancer* **124**, 359-367 (2021). https://doi.org:10.1038/s41416-020-01048-4
123 Sharma, G., Colantuoni, C., Goff, L. A., Fertig, E. J. & Stein-O'Brien, G. projectR: an R/Bioconductor package for transfer learning via PCA, NMF, correlation and clustering. *Bioinformatics* **36**, 3592-3593 (2020). https://doi.org:10.1093/bioinformatics/btaa183

124 Johnson, J. *et al.* Inferring cellular and molecular processes in single-cell data with non-negative matrix factorization using Python, R, and GenePattern Notebook implementations of CoGAPS. *bioRxiv [preprint]* **2022.07.09.499398** (2022). https://doi.org:10.1101/2022.07.09.499398
125 Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**, D991-995 (2013). https://doi.org:10.1093/nar/gks1193
126 Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210 (2002). https://doi.org:10.1093/nar/30.1.207
127 Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res* **29**, 725-738 (2019). https://doi.org:10.1038/s41422-019-0195-y
128 Steele, N. G. *et al.* Multimodal Mapping of the Tumor and Peripheral Blood Immune Landscape in Human Pancreatic Cancer. *Nat Cancer* **1**, 1097-1112 (2020). https://doi.org:10.1038/s43018-020-00121-4
129 Cherry, C. *et al.* Computational reconstruction of the signalling networks surrounding implanted biomaterials from single-cell transcriptomics. *Nat Biomed Eng* **5**, 1228-1238 (2021). https://doi.org:10.1038/s41551-021-00770-5
130 Sherman, T. D., Gao, T. & Fertig, E. J. CoGAPS 3: Bayesian non-negative matrix factorization for single-cell analysis with asynchronous updates and sparse data structures. *BMC Bioinformatics* **21**, 453 (2020). https://doi.org:10.1186/s12859-020-03796-9
131 Fertig, E. J., Ding, J., Favorov, A. V., Parmigiani, G. & Ochs, M. F. CoGAPS: an R/C++ package to identify patterns and biological process activity in transcriptomic data. *Bioinformatics* **26**, 2792-2793 (2010). https://doi.org:10.1093/bioinformatics/btq503
132 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048
133 PhysiCell. *PhysiCell Version 1.12.0*, <https://github.com/MathCancer/PhysiCell/releases/tag/1.12.0> (2023).
134 Wortel, I. M. & Textor, J. Artistoo, a library to build, share, and explore simulations of cells and tissues in the web browser. *Elife* **10** (2021). https://doi.org:10.7554/eLife.61288
135 Heiland, R. & Macklin, P. *PhysiCell Studio Cloud (Version 1.0)*, <https://nanohub.org/tools/pcstudio> (2023).
136 Madhavan, K. *et al.* nanoHUB.org: cloud-based services for nanoscale modeling, simulation, and education. *Nanotechnology Reviews* **2**, 107-117 (2013). https://doi.org:10.1515/ntrev-2012-0043

Digitize your Biology! Modeling multicellular systems through interpretable cell behavior: Figures

Jeanette A.I. Johnson[1,2+], Genevieve L. Stein-O'Brien[1,2,3+], Max Booth[1], Randy Heiland[4], Furkan Kurtoglu[4], Daniel Bergman[1,2], Elmar Bucher[4], Atul Deshpande[1,2], André Forjaz[5], Michael Getz[4], Ines Godet[6], Melissa Lyman[1,2], John Metzcar[4,7], Jacob Mitchell[2,8], Andrew Raddatz[9], Heber Rocha[4], Jacobo Solorzano[10], Aneequa Sundus[4], Yafei Wang[4], Danielle Gilkes[1], Luciane T. Kagohara[1,2], Ashley L. Kiemen[1,2,11], Elizabeth D. Thompson[11], Denis Wirtz[1,2,5,11,12], Pei-Hsun Wu[2,5], Neeha Zaidi[1,2], Lei Zheng[1,2], Jacquelyn W. Zimmerman[1,2], Elizabeth M. Jaffee[1,2], Young Hwan Chang[13], Lisa Coussens[14], Joe Gray[13], Laura M. Heiser[13], Elana J. Fertig[1,2,15,16*], Paul Macklin[4,*]

[1]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University. Baltimore, MD USA.
[2]Convergence Institute, Johns Hopkins University. Baltimore, MD USA.
[3]Department of Neuroscience, Johns Hopkins University. Baltimore, MD USA.
[4]Department of Intelligent Systems Engineering, Indiana University. Bloomington, IN USA.
[5]Department of Chemical and Biomolecular Engineering, Johns Hopkins University. Baltimore, MD USA
[6]Memorial Sloan Kettering Cancer Center. New York, NY USA.
[7]Department of Informatics, Indiana University. Bloomington, IN USA
[8]Department of Human Genetics, Johns Hopkins University. Baltimore, MD USA.
[9]Department of Biomedical Engineering, Georgia Tech University. Atlanta, GA USA.
[10]Centre de Recherches en Cancerologie de Toulouse. Toulouse, France.
[11]Department of Pathology, Johns Hopkins University. Baltimore, MD USA
[12]Department of Materials Science and Engineering, Johns Hopkins University. Baltimore, MD USA
[13]Department of Biomedical Engineering, Oregon Health and Science University. Portland, OR USA.
[14]Department of Cell, Development and Cancer Biology, Oregon Health and Science University. Portland, OR USA.
[15]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA.
[16]Department of Biomedical Engineering, Johns Hopkins University. Baltimore, MD USA.

[+]Contributed equally
[*]Corresponding authors. ejfertig@jhmi.edu , macklinp@iu.edu

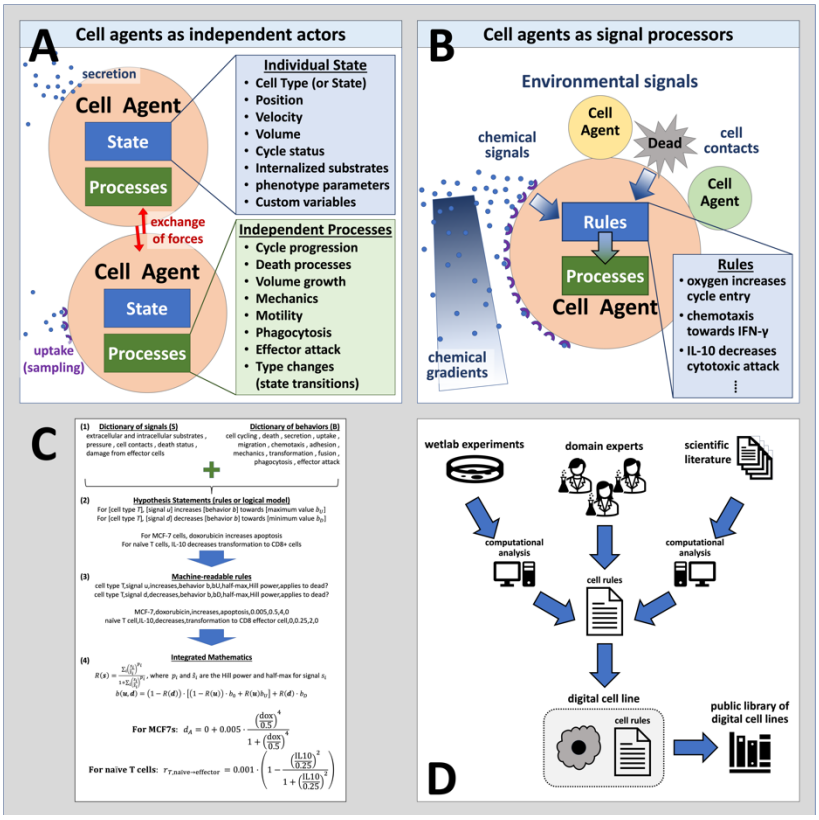

Figure 1: Using agent-based models to digitize cell knowledge. (A) Agent-based models simulate cells as individual objects with separate states and processes. (B) Cells agents use rules that process biophysical signals in their microenvironment—including other cells—to drive changes in their behaviors. These rules are based on our biological hypotheses. (C) The cell behavior grammar combines signals and behaviors from well-defined dictionaries (1) to create interpretable hypothesis statements (2) that can be automatically transformed into computer-readable code (3) and mathematical models (4). (D) Rules can integrate knowledge gained from novel experiments, domain expertise, and literature mining to create digital cell lines. Over time, libraries of digital cell lines accumulate, curate, and systemize our knowledge.

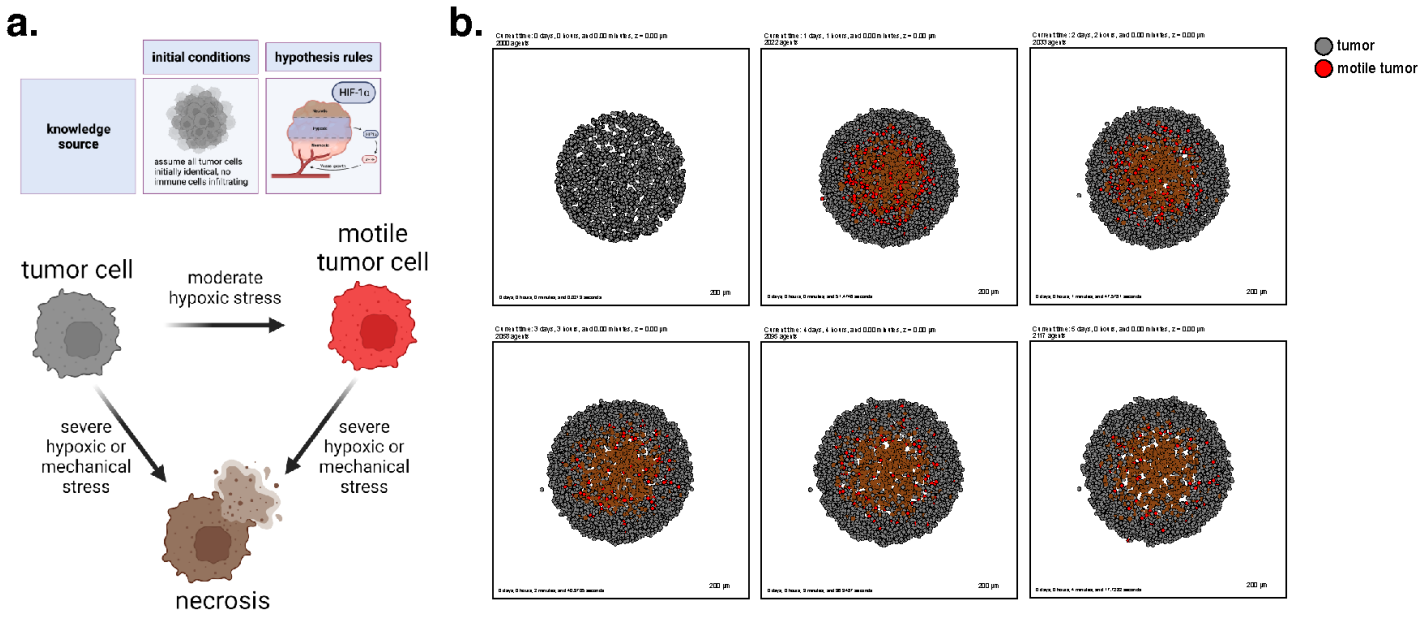

Figure 2. A transient, hypoxia-induced migratory phenotype induced within a homogenous tumor. (a) A cartoon showing the biology in this model and the possible cell type transitions. (b) Simulation snapshots at intervals throughout 5 days. Observe the development of a necrotic core, and the failure of motile tumor cells to reach the tumor boundary before reverting to their prior phenotype.

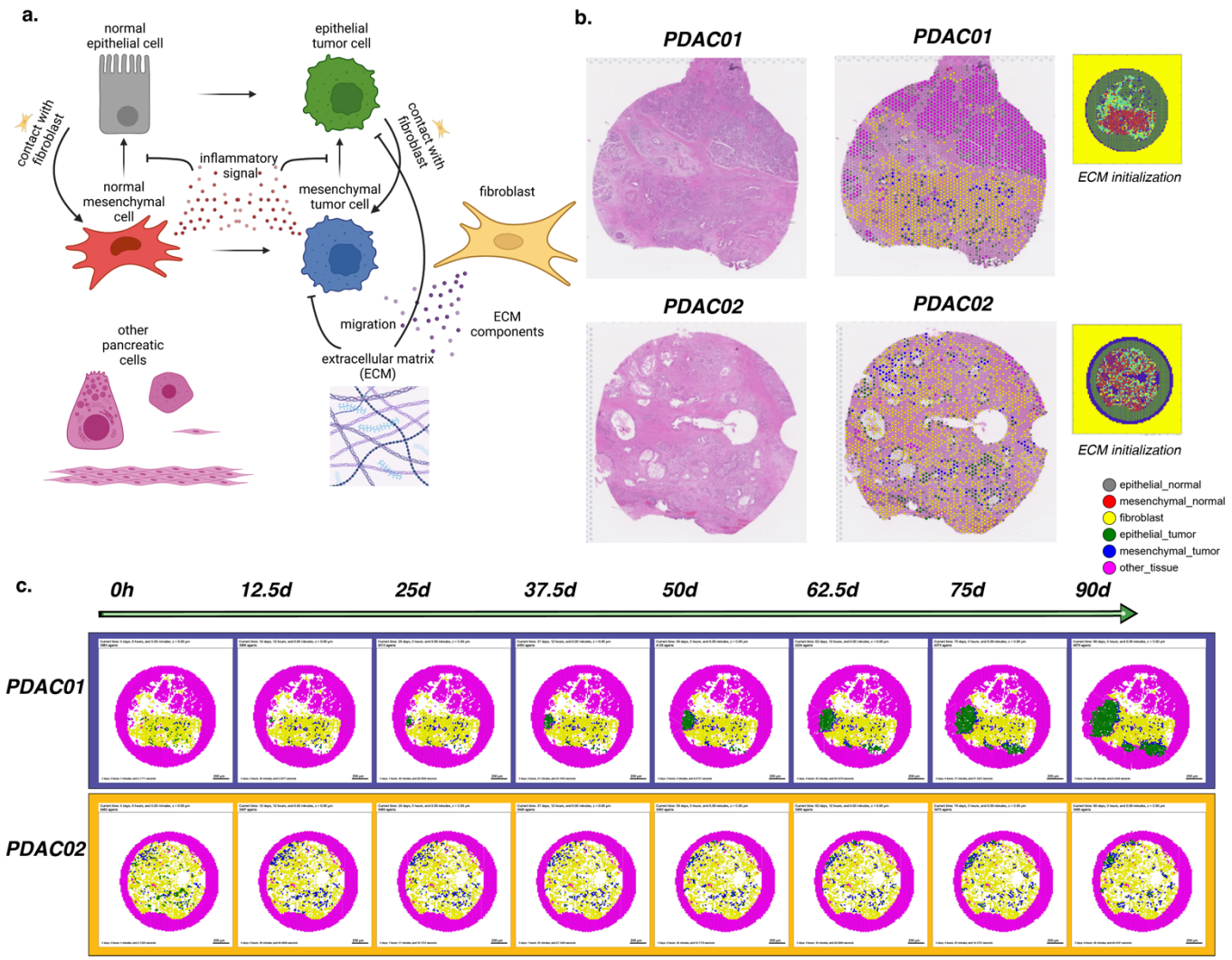

Figure 3. Forecasting tissue outcomes at spatial resolution in the pancreatic tumor microenvironment. (a) Diagram representing the agent types, substrates, and interactions in the model. (b) PDAC tissues selected for modeling and the assigned categorical spot annotations. (c) Snapshots from 90 days of simulated tumor progression in PDAC01 and PDAC02.

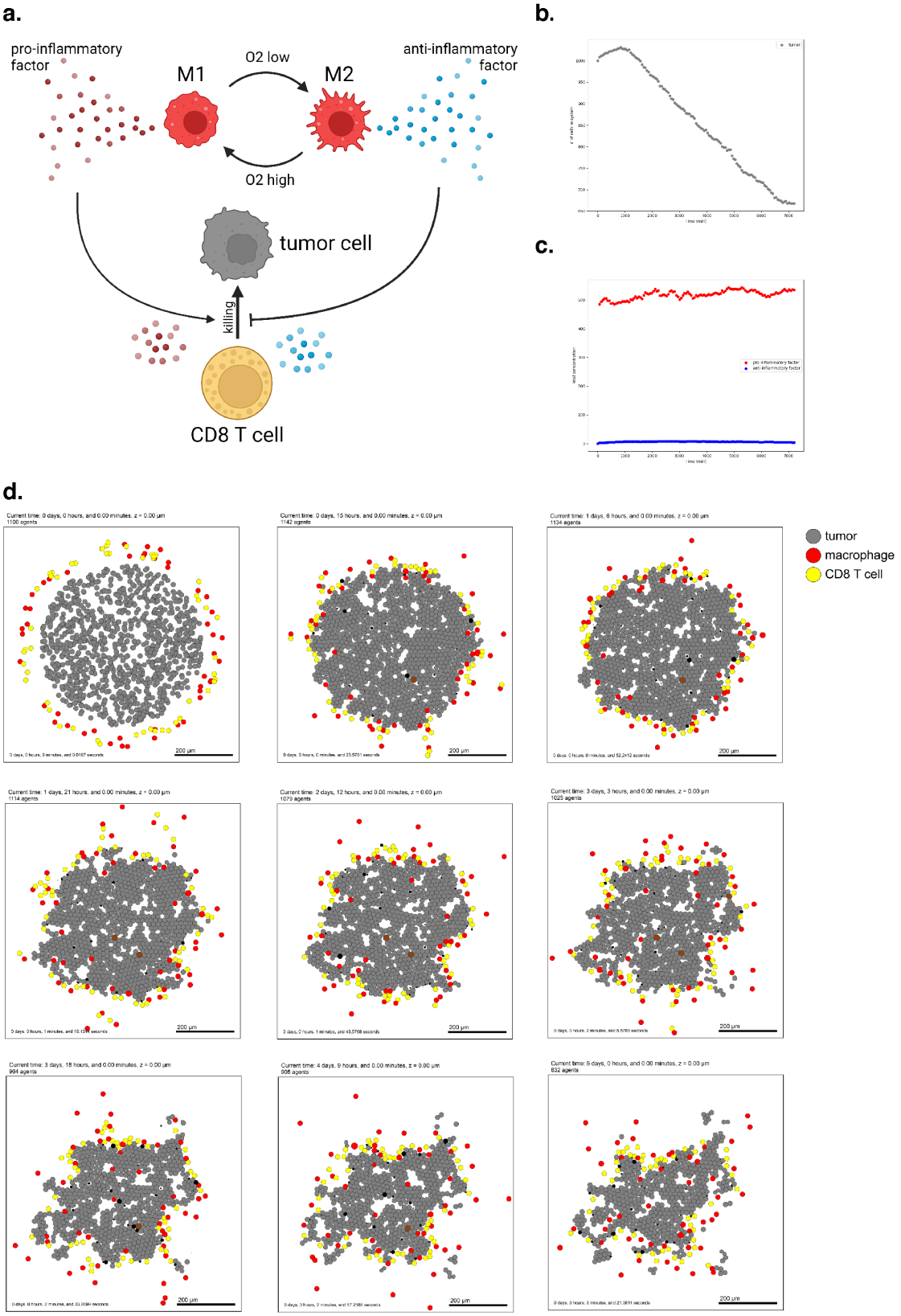

Figure 4. **Simulating a simple antitumor immune response.** (a) A schematic of cell types and substrates within the simulated tumor microenvironment (b) Tumor cell count dwindles over time as the immune response progresses. (c) Concentration of pro- and anti-inflammatory factor throughout the simulation. (d) Snapshots from 5 days of simulated immune response.

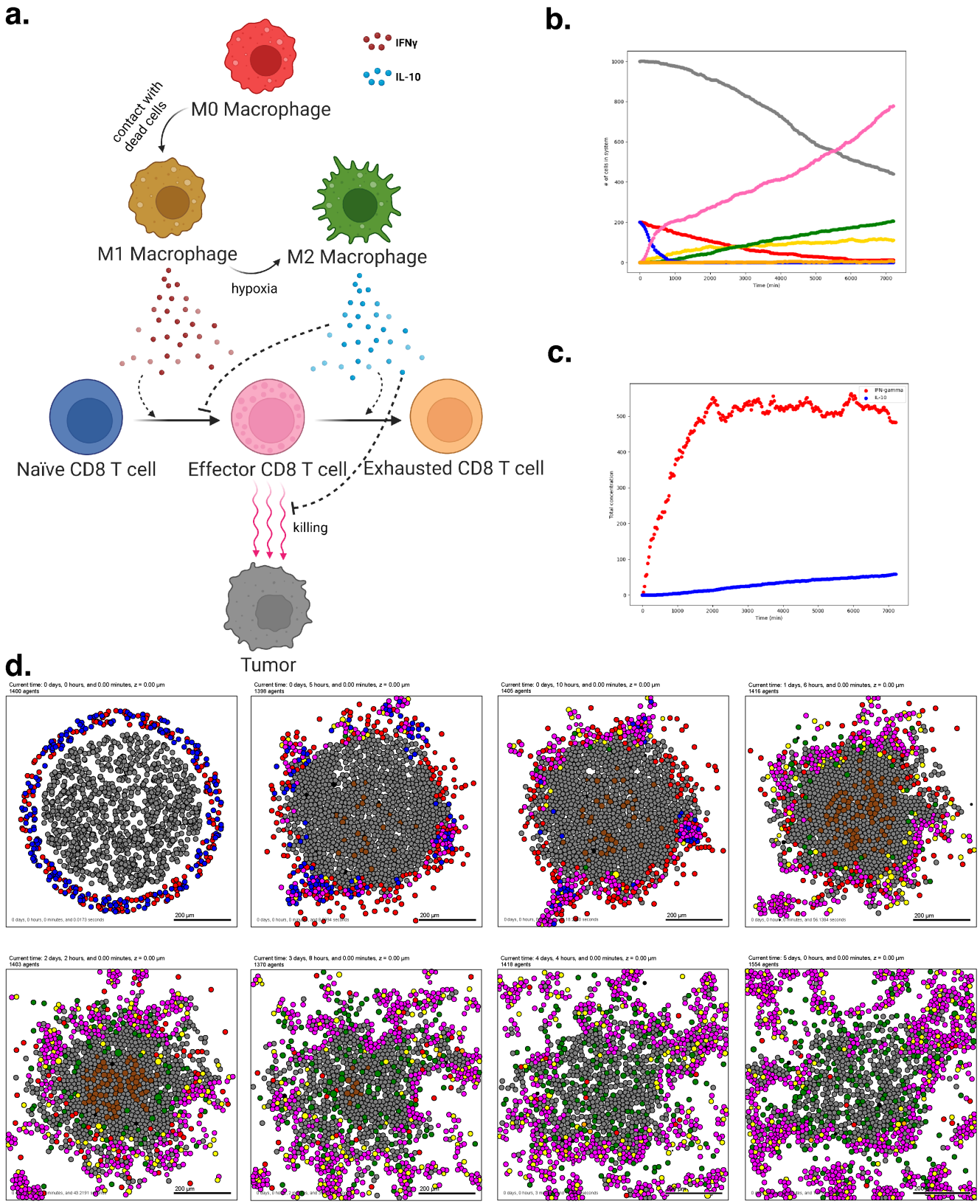

Figure 5. T cell activation and expansion (a) A schematic of the cell types and states in this model. Note that macrophages can now occupy three distinct states, and transitions are unidirectional. (b) Cell counts for each cell type over time. Note the marked expansion of the CD8 T cell population and the corresponding decline in tumor burden. (c) Pro- and anti-inflammatory factor concentration throughout the simulation (d) Interval snapshots of the simulated immune response.

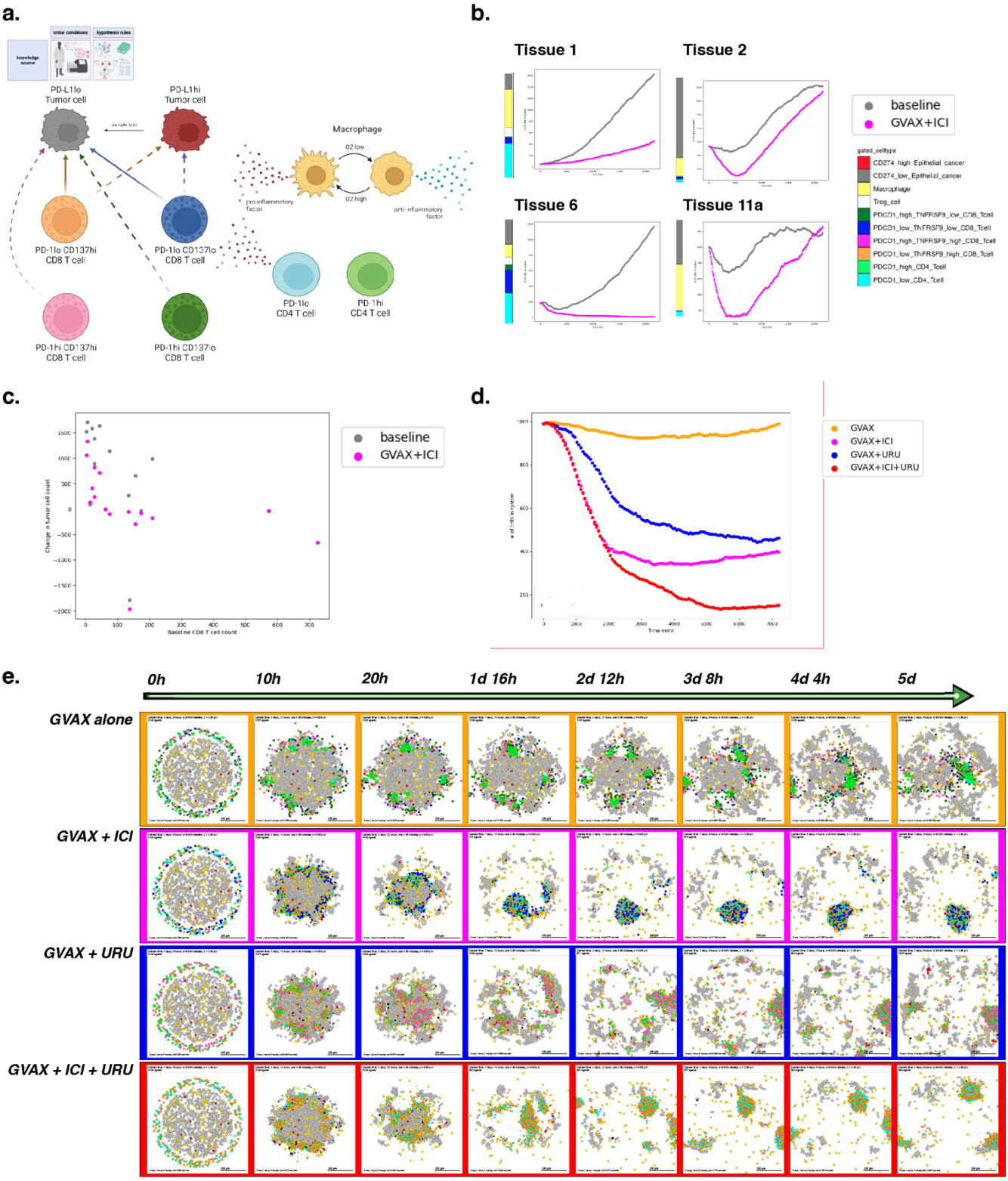

Figure 6. **A simulation of combination immune-targeted anti-cancer therapies.** (a) Schematic of cell types and states present in the model. Note differential killing ability and factor secretion between CD8 T cell subtypes. Therapy is modeled by shifting cell phenotypes according to the agonist or blocking antibodies they have received. (b) Tumor cell count over time simulated at baseline and with therapy. Four tissues were chosen to represent observed patterns of immune response across 16 total tissues. (c) Relationship between simulated tumor volume change over 15 days and baseline CD8 counts for each pancreas tissue identified from the PDAC atlas. (d) Tumor cell counts across simulated time for each therapeutic condition. Note tumor clearance following simulated triple combination therapy. (e) Snapshots over five days of simulation for four therapy conditions.

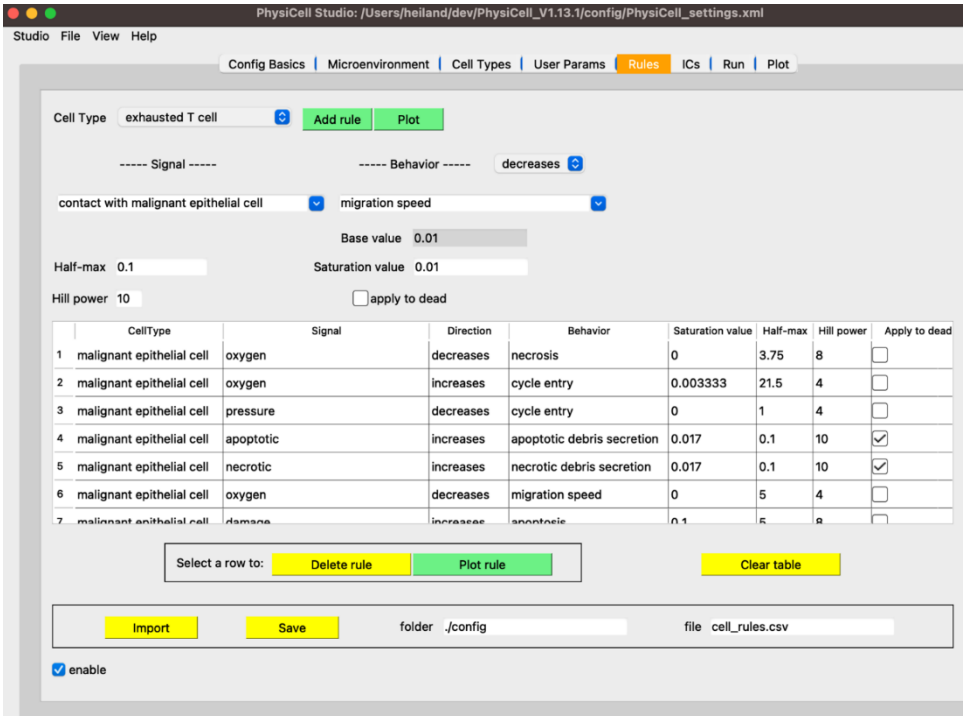

Figure 7. The Rules tab in PhysiCell Studio. Rules can be created and edited interactively. The choices available in Signals and Behaviors are dynamically updated based on the microenvironment and cell types.