TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2024

476

Daniel Camacho Gómez

Unveiling the potential of combining machine learning and mechanistic models: from cell behavior to tumor growth

Director/es

Gómez Benito, María José

García Aznar, José Manuel

ISSN 2254-7606

Prensas de la Universidad Universidad Zaragoza

ISSN 2254-7606

Universidad Zaragoza

Tesis Doctoral

UNVEILING THE POTENTIAL OF COMBINING MACHINE LEARNING AND MECHANISTIC MODELS: FROM CELL BEHAVIOR TO TUMOR GROWTH

Autor

Daniel Camacho Gómez

Director/es

Gómez Benito, María José
García Aznar, José Manuel

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Ingeniería Mecánica

2024

Universidad Zaragoza

Unveiling the potential of combining machine learning and mechanistic models: from cell behavior to tumor growth

Dissertation

submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Mechanical Engineering

by

Daniel Camacho Gómez

Supervisors:

Prof. Dr. María José Gómez Benito

Prof. Dr. José Manuel García Aznar

Department of Mechanical Engineering, University of Zaragoza, Spain

May 22, 2024

Only the disciplined ones are truly free because they are capable of doing what they must at every moment. Those who lack discipline are enslaved by their passions and desires.

Acknowledgements/Agradecimientos

El camino de mi tesis quizá no haya sido el que cabe esperar cuando uno inicia esta aventura. Justo cuando comencé, sucedió lo inesperado: el confinamiento debido la COVID-19. Eso hizo que, a pesar de la ilusión por esta nueva etapa y por compartirla con nuevos compañeros, eso no fuese posible en un principio. Sin embargo, no dejé que la situación me frenara en absoluto; al contrario, lo tomé como una gran oportunidad en la que la motivación no decayó.

Una parte fundamental de que eso fuese posible y que no me desviase del camino que quería seguir fueron mis directores de tesis, Manu y MªJosé. Quiero empezar agradeciéndole a ellos por cómo han sabido guiarme en todo momento, a pesar de que, a la propia incertidumbre que conlleva una tesis, se unió inicialmente la distancia. Siempre han estado más que dispuestos a ayudarme y a aconsejarme desde la honestidad. De ellos he aprendido en estos años, entre muchas otras cosas, a investigar, realizar ciencia y, de forma más importante, el valor de comunicar, lo cual trataré de poner en práctica en mi carrera profesional. Pero también, a nivel personal, me llevo muchos buenos momentos compartidos y una gran relación en la sé que puedo confiar.

Seguidamente, me gustaría agradecer al grupo M2BE, a MªAngeles, Carlos, Angela, Oscar, Inês, Pedro, Fran, David, Jaime y al resto de integrantes por el buen ambiente de trabajo y compañerismo que siempre ha habido.

No quiero olvidarme de Raúl, Diego y Silvia por su ayuda, por escucharme (y aguantarme), por todos esos ratos al margen de la tesis y por todos los cafés juntos, que por desgracia (o tal vez por suerte), no han podido ser desde el principio. Ellos, junto con la complicidad del resto de la sala, Nico, Álvaro y Elena, han hecho de estos últimos meses toda una "experiencia" y han logrado que ir a trabajar sea un placer (y una tortura). También quiero agradecer a todos con los que he compartido mi día a día a lo largo de la tesis: José, Ángel, Mikel, Lucas, Ricardo, Carlos, Alicia, Benedetta, Quercus, Itziar, Iulen, Bea y Patri.

At this point, I would like to express my gratitude to Prof. Mark Girolami for hosting me in his research group at Cambridge. Not only was the experience technically inspiring, but I also met excellent people who embraced me and made my stay personally unforgettable: Justin, Arnaud, Connor, Yannis, Alex, Raris, and Ben. Additionally, I met great people from all over the world who enriched me as a person, such as Wenjie, Josephine, Thilanka, and many more. Special thanks go to Alberto and Ele for making everyday life truly enjoyable and memorable.

También quiero agradecer a los amigos de toda la vida, Borja, Yolanda, Marta, Nuria, Bascu, Juani, Fleky, Flamen y Navas quienes hacen de mis vueltas a Málaga una vuelta a mi hogar.

Ni mucho menos quiero terminar sin agradecer a mi hermano Franci, María, y a Guille, Gonzalo y Germán, por su alegría y por quererme con esa energía que,

por un momento, hace que el resto de las cosas queden aparte.

No podrían faltar mis gracias a Carmen por ser mi confidente todos estos años. Hemos recorrido nuestros caminos creciendo y avanzando juntos. Ha sido una parte esencial sin la cual, sin duda, no podría estar escribiendo estas palabras. Además, agradezco a Carmen y Jesús, y a Marta, Cris, Laura, Jesús, María y Jesús por su constante apoyo y cariño.

Finalmente, me gustaría dedicar este libro a mis padres. Son ellos la razón por la que estoy aquí hoy. Me han enseñado, mediante su ejemplo, a trabajar, esforzarme e intentar dar lo mejor de mí. Estaré eternamente agradecido por todos los valores que me inculcaron desde pequeño y que me han definido como persona. Por todo esto, siempre serán corresponsables de mis logros en la vida.

A mis padres.

Abstract

Cancer is one of the most prevalent and challenging diseases worldwide, affecting millions of people, and with a projected increase in future prevalence. Therefore, it is crucial to enhance our understanding of the mechanisms underlying its initiation and development, improve current monitoring methods, and pursue more effective treatments.

This thesis presents a computational approximation to cancer, deconstructing the disease into its components to gain a comprehensive understanding. It examines various aspects of cancer across different scales, from the cellular level to the organ level to elucidate the factors influencing cell behavior, understand the mechanisms driving cells from healthy to tumoral states, and predict tumor growth. In this regard, computational models are powerful tools for their capacity to provide cost-effective means to abstract phenomena, explore hypothetical scenarios, analyze factors meticulously, determine causality, and predict potential outcomes. Thus, this thesis employs diverse computational modeling approaches, ranging from discrete models with agent-based models to continuum models with partial differential equations. Additionally, it investigates the integration of artificial intelligence methods into these physics-based models, demonstrating their combined capability to reproduce complex phenomena and enhance predictions with meaningful interpretations.

Following this methodology, the thesis initially examines movement patterns of experiments of the immune system cells and engineered immune cells used in cancer immunotherapy. The goal is to assess their migratory capacity and identify factors that may diminish the effectiveness of immunotherapy in cancer. Subsequently, the focus shifts to the morphogenesis of cystic structures to understand their formation and the factors contributing to their transition from healthy to tumoral structures, which are found in the most common type of cancer. Here, the impact of the extracellular matrix and fluid hydrostatic pressure on the process is investigated, followed by quantitative reproduction of experimental assays involving pancreatic cells. Finally, the thesis addresses the prediction of prostate tumor growth in real patients incorporating clinical information. Thus, a methodology is presented to predict prostate cancer prognosis, offering a clinical tool for comprehensive evaluation and advancing toward more accurate, faster, and inexpensive monitoring methods, potentially reducing the need for clinical tests such as magnetic resonance scans.

Therefore, this thesis aims to provide insights into some aspects of cancer, improve therapy approaches, and provide new perspectives on the integration of artificial intelligence into the field of computational biology.

Resumen

El cáncer es una de las enfermedades más prevalentes y desafiantes, afectando a millones de personas en todo el mundo y con una incidencia en aumento. Por tanto, es crucial avanzar en el entendimiento de los mecanismos subyacentes de su inicio y desarrollo, así como mejorar los métodos de monitorización y buscar tratamientos más efectivos.

Esta tesis presenta una aproximación computacional al cáncer, desgranando la enfermedad en sus algunos de sus componentes para obtener una comprensión integral de la misma. Así, se examinan varios aspectos del cáncer en diferentes escalas, desde el nivel celular hasta el nivel de órgano, para dilucidar los factores que influyen en el comportamiento celular, entender los mecanismos que alteran su comportamiento sano a uno tumoral y, finalmente, predecir el crecimiento de tumores. En este sentido, los modelos computacionales son una herramienta poderosa por su capacidad para abstraer fenómenos, explorar escenarios hipotéticos, analizar los factores involucrados, determinar causalidad y realizar predicciones. Por ello, esta tesis emplea diversos enfoques de modelado computacional, que van desde modelos discretos con modelos basados en agentes hasta modelos continuos con ecuaciones diferenciales parciales. Además, investiga la integración de métodos de inteligencia artificial en estos modelos físicos, demostrando su capacidad combinada para reproducir fenómenos complejos y mejorar las predicciones.

Siguiendo esta metodología, esta tesis examina inicialmente los patrones de movimiento de células del sistema inmunitario y células inmunitarias modificadas utilizadas en tratamientos de inmunoterapia en el cáncer. El objetivo es evaluar su capacidad migratoria e identificar los factores que puedan disminuir la efectividad de dicho tratamiento. Posteriormente, el foco se traslada hacia la morfogénesis de estructuras císticas para comprender su formación y los factores que contribuyen a su transición de estructuras sanas a tumorales típicas del cáncer más común. Así, se investiga el impacto de la matriz extracelular y la presión hidrostática del fluido en el proceso, seguido de la reproducción de ensayos experimentales con células pancreáticas. Finalmente, la tesis aborda el cáncer de próstata en pacientes reales incorporando información clínica. De este modo, se presenta una metodología para predecir su pronóstico, ofreciendo una herramienta clínica para una evaluación integral y avanzanda hacia métodos de monitoreo más precisos, rápidos y económicos, potencialmente reduciendo la necesidad de pruebas clínicas.

Por lo tanto, esta tesis tiene como objetivo proporcionar información sobre algunos aspectos del cáncer, mejorar los enfoques terapéuticos y ofrecer nuevas perspectivas sobre la integración de inteligencia artificial en el campo de la biología computacional.

Contents

List of Figures

List of Tables

Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
ABM	Agent-Based Model
ADC	Apparent Diffusion Coefficient
AI	Artificial Intelligence
ANN	Artificial Neural Network
CA	Cellular Automata
CAR	Chimeric Antigen Receptor
CBM	Center-Based Model
CDF	Cumulative Distribution Function
CFD	Computational Fluid Dynamics
CNN	Convolutional Neural Network
CPM	Cellular Potts Model
DCE	Dynamic Contrast Enhanced
DCE-MRI	Dynamic Contrast-Enhanced Magnetic Resonance Imaging
DDFT	Dynamic Density Functional Theory
DL	Deep Learning
DNA	Deoxyribonucleic Acid
DNN	Deep Neural Network
DT	Digital Twin
DW	Diffusion Weighted
ECM	Extracellular Matrix
FDM	Finite Difference Method
FE	Finite Element
FEM	Finite Element Method
GA	Genetic Algorithm
GNN	Graph Neural Network
GP	Gaussian Process
GPE	Gaussian Process Emulator
GPUs	Graphics Processing Units

IFP	Interstitial Fluid Pressure
IoT	Internet of Things
LGCA	Lattice Gas Cellular Automata
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MDCK	Madin-Darby canine kidney
ML	Machine Learning
MRI	Magnetic Resonance Imaging
NN	Neural Network
ODE	Ordinary Differential Equation
PAP	Preapical Patch
PCa	Prostate Cancer
PDAC	Pancreatic Ductal Adenocarcinoma
PDE	Partial Differential Equation
PDEC	Pancreatic Ductal Epithelial Cells
PI-RADS	Prostate Imaging Reporting and Data System
PINN	Physics-Informed Neural Network
PSA	Prostate-Specific Antigen
RL	Reinforcement Learning
SEM	Subcellular Element Model
SPH	Smoothed-Particle Hydrodynamics
T2WI	T2-Weighted Image
TME	Tumor Microenvironment
US	United States

Chapter 1

Introduction

The years wrinkle our skin, but lack of enthusiasm wrinkles our soul.

– ***Socrates***

Contents

1.1 Motivation

Cancer is one of the most prevalent and challenging diseases of our time, impacting millions of people worldwide. It is the second leading cause of death globally [1], as well as in Europe [2] and the United States (US) [3], accounting for an estimated 10 million deaths in 2020 alone [4]. This rate imposes a significant economic burden on healthcare systems [5, 6], individuals [7], and societies [8] at large, with a consistent upward trajectory observed in recent years [9]. Despite the continuous decrease in the risk of death from cancer, resulting in an overall mortality rate drop of 32% in US since 1991 [10], the future prevalence of cancer is projected to increase [11–13]. This rise in cancer cases could be attributed to various factors, including the aging of society, sedentary lifestyles, or unhealthy dietary habits [14], which may contribute to an increased susceptibility to cancer, exacerbating the economic and social challenges associated with cancer. Moreover, as life expectancy rises, there is a concerning possibility that new manifestations and forms of cancer may emerge in the future, underscoring the enduring and evolving nature of cancer as an unending disease. Therefore, it is crucial to enhance our understanding of cancer. This encompasses not only the pursuit of more effective treatments for advanced stages but also a deep exploration of the mechanisms underlying its initiation. By comprehending how cells transition from a healthy to a tumoral behavior and identifying the factors that drive this transformation, we can gain profound insights into the fundamental processes governing cell behavior. This approach might pave the way for advancements in early detection, targeted interventions, and ultimately the potential to halt cancer's devastating impact on individuals and societies.

1.2 Biological background

In the last two decades, a new perspective has emerged in cancer research, highlighting the significant impact of mechanical forces and cues on tumor initiation, growth, invasion, metastasis, and response to therapy. This prominent branch, known as mechanobiology [15], specifically focuses on unraveling the role played by mechanical forces and properties in biological processes [16]. These processes encompass not only diseases such as cancer but also physiological phenomena such as tissue development, immune response, or wound healing. Understanding this interplay between mechanical cues and cellular behavior offers valuable insights into the complex mechanisms driving cancer progression and vital physiological processes.

In response to mechanical cues from both their environment and neighboring cells, cells exhibit mechanisms to capture these signals in a reception process (Figure 1.1**A**) and subsequently convert them through a process called mechanotransduction into a biological response (Figure 1.1**B**). Cells are immersed in

a complex and dynamic three-dimensional network called the Extracellular Matrix (ECM). The ECM, composed of proteins, carbohydrates, and other molecules, provides structural and biochemical support to cells [17]. The ECM is a basic component in various physiological processes, including cell differentiation, migration, adhesion, and tissue development. Thus, cells possess around their membranes a protein called integrins, which attach them to the ECM (Figure 1.1**A**, detail 1) [18, 19]. This connection, often referred to as focal adhesions, not only provides structural support but also acts as a signaling hub. These integrins are connected to the cell's cytoskeleton, a dynamic protein network within the cytoplasm that provides structural support, determines cell shape, and regulates cellular behavior. Comprising elements such as actin filaments, myosin motors, microtubules, and intermediate filaments, the cytoskeleton plays a crucial role in generating, transmitting, and responding to mechanical signals over both short and long timescales [20–23] (Figure 1.1**A**, detail 2). Thus, mechanical forces are transmitted through the integrins between the ECM and the cytoskeleton, triggering biochemical events and activating signaling molecules within the cell. In this way, the mechanical properties of the ECM, the integrins and the cytoskeleton mediate the mechanical forces transmitted, regulating the biological response of the cell [24].

Besides the ECM, the surroundings of the cell also include the presence of interstitial fluid, which generates a fluid-cell interaction (Interstitial Fluid Pressure (IFP)) that derives to a cellular biomolecular response [25] (Figure 1.1**A**, detail 3). Cells employ various mechanisms to capture and respond to fluid-related signals, detecting and interpreting mechanical forces associated with fluid pressure. We can highlight the presence of ion channels located in the cell membrane that are sensitive to mechanical forces, including fluid pressure. These channels open or close in response to mechanical cues, allowing the influx or efflux of ions, which triggers signaling events [26, 27] (Figure 1.1**A**, detail 4). Also, the cell membrane itself can act as a sensor for fluid pressure, leading to alterations in membrane and cytoskeleton tensions to then initiate signaling pathways. Among membrane-associated proteins and complexes involved, some cells also have a hair-like structure named primary cilia that extends from their surface. It serves as antennae for detecting fluid flow and pressure changes [28] (Figure 1.1**A**, detail 5).

Furthermore, cells actively participate in mechanical interactions with neighboring cells. Much like integrins, cells are equipped with cell surface receptors known as cadherins, which establish physical connections between adjacent cells [29] and their cytoskeletons through intermediate proteins [30, 31] (Figure 1.1**A**, detail 6). These adhesion molecules contribute to the structural integrity and stability of cellular arrangements within tissues. This cell aggregation and sorting involve the orchestrated movement and alignment of cells, eventually leading to the generation of mechanical forces or tension transmitted through the cadherins or direct membrane contact.

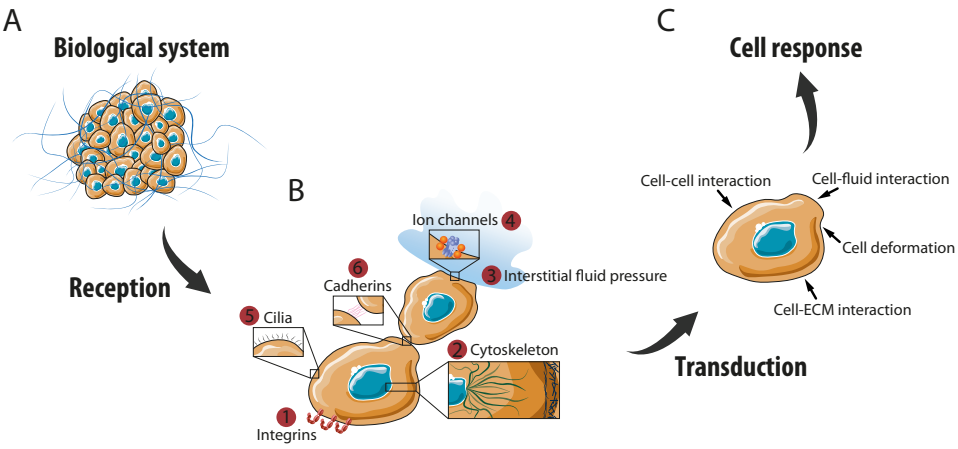

Figure 1.1: ***Overview of cell signaling.*** ***A*** *Biological system composed of a cluster of cells and the ECM.* ***B*** *Receptors of the cell. The integrins attach cells to the ECM and mediate the mechanical forces transmitted between the ECM and the cell (1). The cytoskeleton provides support to the cell, playing an essential role in processing the mechanical forces that the cell receives (2). The cell surroundings may include interstitial fluid, which generates a fluid-cell interaction (3). Additionally, the cell possesses ion channels that are sensitive to cues, including fluid pressure (4). Some cells also have hair-like structures named primary cilia extending from their surface to detect fluid flow and pressure changes (5). Finally, the cadherins establish physical connections between adjacent cells and their cytoskeleton, transmitting mechanical forces through cells (6).* ***C*** *Activation of the biological cell response. These reception mechanisms allow cells to receive cues from other cells and the environment, such as cell-cell, cell-fluid, and cell-ECM interactions. Finally, cells initiate complex intracellular signaling cascades that transduce these signals into a cellular response or biological behavior.*

These are basic mechanisms through which cells receive the mechanical information of the ECM and neighboring cells, although there exist many more mechanisms that allow them to capture distinct mechanical signals, and certainly chemical signals. After these reception mechanisms of signals, cells initiate complex intracellular signaling cascades that transduce these into biological behaviors (Figure 1.1**B**). From a direct standpoint, the tensions resulting from mechanical interactions induce deformation in the cell's cytoskeleton, influencing specific cellular behaviors. Moreover, the mechanisms of ECM and cell recognition modulate the expression of specific genes associated with diverse cellular responses in the nucleus, including changes in cell shape, migration, proliferation, and differentiation. Therefore, the orchestrated integration of these mechanotransduction mechanisms allows cells to dynamically adapt to their mechanical microenvironment, generating context-specific biological outcomes.

This mechanical perspective underscores the need for an engineering approach to comprehensively analyze and bridge the apparent gap between engi-

neering and biology. Leveraging mathematical, computational biology, and engineering techniques not only approximates the physical nature of biological problems but also offers insights into unraveling these intricate processes. Next, we will see a mechanobiological interpretation of important biological processes that will be discussed later in the following chapters of this thesis.

1.2.1 Cell migration

Cell migration is the process in which cells move in our body and it plays an important role in many physiological aspects [32]. Cell migration is essential in embryonic development, where cells must move to form tissues and organs [33]. Regarding the immune system, these cells possess a remarkable ability to migrate within our body, traveling long distances to reach infection or injury sites and orchestrate a defensive response [34]. In wound healing, cells migrate coordinately to close gaps and restore tissue integrity [35, 36]. Nonetheless, it is not only important in physiological processes, but it is also a key feature of cancer. Aberrant cell migration is implicated in cancer metastasis, where cells acquire the ability to migrate within the tissue and even invade distant tissues. Therefore, understanding the mechanisms of cell migration holds a critical implication for addressing both physiological and disease processes.

From a simplified standpoint, cell migration involves a series of regulated events tightly related to the generation of protrusive and contractile forces by the actin filament cytoskeleton [37]. Cells control the polymerization and depolymerization of actin filaments to rearrange and generate a protrusion of the plasma membrane at the leading edge of the migration direction [38]. These protrusions, such as the lamellipodium and filopodia, attach to the fibers of the ECM to facilitate cell movement and also act as a sensing system to explore the microenvironment [39]. Subsequently, a contractile force is generated in the cell protrusions that drive forward cell movement, followed by rear retraction. In this process, myosin, a molecular protein, contributes to cell contractility [40]. Another element of cell-substrate adhesion, integrins, which attach to the ECM, also contribute to the pulling forces necessary for migration. During cell movement, the cytoskeleton and nucleus reassemble, coordinating and adapting to the changing cellular environment [41].

Depending on the degree of involvement of the explained mechanisms, different modes of migration are characterized [42]. Ameboid cell migration is characterized by a more rounded cell shape, lacking protrusion generation. In this case, high-speed myosin contractility is the driving force of the process, rather than cell-matrix adhesions. Mesenchymal cell migration, however, requires higher cell deformation with the creation of protrusions led by dynamic changes in the cytoskeleton, low myosin contractility but high cell-matrix adhesions. Other migration modes, such as lobopodial and swimming, span between these two modes.

Besides the forces generation of the cell, the physical microenvironment also is key in this process. Hence, cells can initiate the movement from a stationary state motivated by signals such as the ECM molecules, growth factors, chemokines, or mechanical cues. Once the locomotion is initiated, the ECM provides both physical support and a substrate for cell adhesion and its topological aspects such as its porosity or the fiber alignment can highly impact cell migration, affecting their speed, persistence, and directionality. But also, the mechanical properties of the ECM such as the stiffness and viscoelasticity are crucial in the way the cells migrate. Lo et al. [43] found that cells migrated preferentially toward stiffer directions in collagen substrates with stiffness gradients, concluding that cells can receive the mechanical input generated by substrate deformation to regulate the formation and retraction of protrusions. This phenomenon, which they called "durotaxis", has been widely studied afterward [44–46], showcasing that not only chemical cues (chemotaxis) guide cell migration, but tissue rigidity and strain also play an important role controlling cell locomotion.

Additionally, mechanical cues such as shear stress and fluid flow also influence cell migration. Prentice-Mott et al. [47] demonstrated that cells tend to move through the lower hydraulic resistance paths in confined environments, a phenomenon known as barotaxis. These asymmetrical hydraulic resistance paths create pressure gradients that guide cell directional migration towards the higher pressure gradient [48, 49]. Furthermore, alterations in pressure and external forces impact the force equilibrium at the cell cortex, leading to the activation of mechanosensitive ion channels and potentially establishing a feedback system that interacts with the actin network [50]

In summary, the intricate interplay between cellular mechanisms, including actin cytoskeleton dynamics, myosin contractility, and integrin-mediated adhesion, orchestrates diverse modes of cell migration. The physical microenvironment, governed by factors such as ECM composition, topography, and mechanical properties, significantly influences cell motility. Moreover, emerging insights into phenomena such as durotaxis, influenced by substrate stiffness, and barotaxis, responding to hydraulic resistance paths, underscore the multifaceted nature of cellular responses to mechanical cues. Understanding these complexities holds the key to unraveling the fundamental principles governing both physiological and pathological cell migration dynamics

1.2.2 Morphogenesis

Morphogenesis refers to the developmental process through which cells organize to form complex structures. From the earliest stages of embryonic development to tissue remodeling in adulthood, morphogenesis guides the creation of tissues and organs, ensuring the precise spatial arrangement of cells for functional structures. To achieve this, cells undergo proliferation, coordinated movements,

and differentiation, orchestrating temporary and spatial changes from the initial single cell to generate diverse tissues and organs.

Behind this process, distinct gradients of chemical signaling molecules, known as morphogens, guide cell fate, and control tissue patterning. In 1952, Alan Turing established the mechanisms of pattern formation in morphogenesis [51]. The structure formed, referred to as pattern formation, requires satisfying two conditions: local activation and long-range inhibition [52]. Thus, it is necessary for one of the substances, referred to as the activator, to be autocatalytic. This means that once a slight deviation from the homogeneous state occurs, there is an increase in its production, leading to a consequent rise in its concentration. On the other hand, long-range inhibition is required from what is termed the inhibitor. The inhibitor counteracts the autocatalysis of the activator, preventing its increase from tending to infinity and allowing the system to reach a steady-state pattern. Importantly, this inhibitor must possess the characteristic of diffusing at a faster rate than the activator.

However, the control over this process is not solely relegated to morphogens, mechanical cues also have considerable influence over the outcome [53–56], regulating the process and providing feedback about size and shape. The ultimate success in constructing and stabilizing cell structures directly relies on precise mechanical interactions throughout the developmental process. Firstly, mechanical forces generated by cells' cytoskeleton produce forces that can drive morphogenetic processes [57]. Thus, apical constriction, mediated by the contraction of actin-myosin cytoskeletal networks, plays a critical role in the formation of complex tissue structures, such as tissue folding, bending, and invagination [58]. In this context, cell-cell adhesions are crucial not only in supporting the structure and guaranteeing its integrity [59] but also can lead to distinct cell sorting for tissue patterning and organization [60]. Differential adhesive strength through different expressions of cadherins determines the pattern of cell sorting [61, 62]. Additionally, interactions between cells and the ECM serve as mechanical cues, regulating cell behavior [63], or serving as a scaffold for tissue organization, guiding and stabilizing tissue expansion through force anisotropies resulting from its heterogeneous distribution [64]. Furthermore, fluid flow is a major factor in morphogenesis, affecting morphogenesis in distinct ways. It can create gradients of chemical signaling molecules, generate physical forces including pressures and shear stresses that contribute to shaping tissues or it can activate biological pathways in mechanosensitive cells [65]. Thus, the interplay between the dynamical rearrangements of the cytoskeleton, the ECM, fluid flow, and cell adhesion strength [66] underscores the intricate relationship between mechanical forces and cellular behavior, shaping tissue development and morphogenesis through mechanotransduction pathways.

Therefore, understanding morphogenesis tackles the question of how structures are formed *de novo*, and understanding this process allows us to understand

not only how healthy structures are formed, but also how they can be disrupted and the circumstances under those transform their correct formation to a pathological one.

1.2.3 Cancer

Cancer is a complex disease characterized by the uncontrolled growth and division of cells in the body. It encompasses a diverse group of diseases with distinct characteristics, but they share several common features. The prevalent feature of the disease is the uncontrolled growth and division of cancer cells. Unlike healthy cells, which undergo programmed death (apoptosis) when damaged or old, cancer cells evade the control mechanisms, conferring them a limitless proliferative potential. Another hallmark of cancer is uncontrolled cell migration, which gives the capacity of malignant cells to escape and invade other tissues, producing the expansion of the disease. Therefore, understanding the factors that contribute to this uncontrolled growth and migration of cells is fundamental to taking a step closer to addressing cancer.

The initiation of cancer can be interpreted as a morphogenetic process in which abnormal cells begin organizing to form a pathological structure. Thus, mutations in healthy cells could induce altered transduction of normal cues from other cells and environmental factors, modifying their normal biological functioning to a tumorous state. Alternatively, an altered cascade of signals, rather than a mutation within the cell, can modify cell behavior, promoting their transformation into a tumoral phenotype. For instance, cells dynamically adapt to force by modifying their behavior and remodeling their microenvironment [67]. Paszek et al. [68] found that non-cancerous mammary cells, when cultured in a high-stiffness collagen matrix, formed aberrant tumoral structures. Provenzano et al. [69] demonstrated that increased stromal collagen in mouse mammary tissue enhances tumor formation and invasion by increasing ECM stiffness, resulting in higher mechanical loads and resistance to cellular contractility, ultimately promoting proliferation and a more transformed phenotype. Additionally, Messal et al. [70] employed three-dimensional whole-organ imaging techniques to analyze pancreatic tissue architecture in mice and showed that mechanical tension imbalances are determinants of epithelial tumorigenesis.

As a result, cells form a tumor structure. In this situation, there is a consolidation of a tumor microenvironment that promotes the growth of the tumor. Thus, the complex nature of this process, not only lies in the compound interactions that are occurring but also in the changing properties of the agents involved in this process. For example, within cancer, ECM remodeling, involving alterations in its composition, structure, and organization, driven by cancer-activated fibroblasts. This remodeled and stiffened ECM serves as a physical cue transduced by tumor cells, affecting their proliferation, invasiveness, morphology, and respon-

siveness to treatments [71, 72]. Moreover, residual solid stress is accumulated within tumors during progression [73], influencing physical cues for tumor cells, representing distinct abnormalities in tumor behavior alongside matrix stiffness, impacting tumor progression and metastasis [74]. The growth of the tumor produces physical forces compressing blood and lymphatic vessels [75], reducing nutrient delivery and creating hypoxia, affecting the delivery of chemotherapy [76] and hypoxia reduces the effectiveness of radiotherapy [77, 78]. Also, there is an increased IFP, a significant hallmark of solid tumor growth. This elevation stems from various factors, including the chaotic and aberrant vasculature within tumors, high permeability of these vessels causing leakages, matrix decomposition, and secretion of factors by tumor cells. Elevated IFP not only serves as a predictive and prognostic biomarker of tumors but also poses a therapeutic barrier hindering effective drug delivery [79, 80]. In this environment, tumor cells induce the generation of new vascularization networks to transport nutrients such as oxygen and glucose to obtain energetic support to maintain their functions and keep growing [81, 82].

These physical properties (altered mechanical properties of the ECM, compression, solid stress, increased interstitial fluid pressure, and hypoxia), together with other types of cells such as cells of the immune system, chemical factors such as growth factors or chemokines, constitute the Tumor Microenvironment (TME), whose composition is crucial in tumor initiation, growth, and invasion, and metastasis [83, 84]. Ultimately, at some point, tumor expansion and progression can derive into an uncontrolled cell migration, spreading from the primary tumor to surrounding tissues and even distant organs (metastasis) [85], which is the main cause of cancer-related death. Therefore, studying aspects such as cell morphogenesis, how cells transition from healthy behaviors to pathological ones and the reasons behind it, tumor-contributing factors, and cell migration, can allow us to gain insights into the initiation, establishment, and expansion of cancer.

1.3 Experimental biology

One potent approach for advancing our understanding of biology is through an experimental perspective. Experimental models serve as invaluable tools in studying biology owing to their tangible and hands-on nature. The tangible and concrete nature of experimental models not only enhances our comprehension of fundamental biological principles but also facilitates the development of innovative strategies and interventions. In the spectrum of experimental models, a distinction arises based on the level of abstraction or proximity to a whole organism.

At a lower level, *in vitro* models aim to mimic characteristics of living organisms to study cell behavior, cell migration, tumor growth, or drug response in

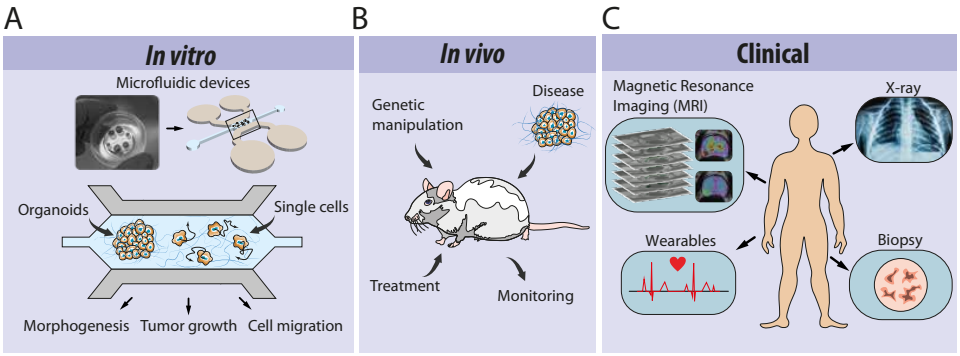

Figure 1.2: ***Experimental model approaches.*** ***A*** *In vitro experiments. These experiments aim to reproduce the characteristics of living organisms in a controlled context. We highlight the use of microfluidic devices for cell culturing, which allows the study of processes such as morphogenesis, tumor growth, or cell migration in a closely physiological environment.* ***B*** *In vivo models. These commonly utilize mice, providing a more realistic exploration of biological aspects. Typically, it involves the induction of a disease or genetic manipulation, followed by monitoring the signs of the disease or response to treatment.* ***C*** *Clinical models. Here, we consider methods to study or evaluate biological processes and diseases in human subjects. These data acquisition methodologies encompass MRI, X-ray, biopsies, and wearables to detect and predict the prognosis.*

a controlled, physiologically relevant context. In this regard, cell cultures are a technique in which cells are seeded in a suitable medium outside their natural environment, allowing for their manipulation. These cultures can be maintained as monolayers, a simple and low-cost method of cell culture maintenance where cells adhere to a flat surface. However, these 2D cell monolayers have limited ability to mimic the real characteristics of tissues or tumors, and the interactions between cells and the environment [86]. To overcome this limitation, 3D culture models have emerged to better mimic the complexity of tissues and bridge the gap between cell cultures and real tissues [87]. Among 3D culture models, cell spheroids are a cell culture model where cells aggregate and grow in a spherical shape, forming a multicellular structure with cell-cell interactions similar to those found within organisms. This model allows the study of various aspects of cellular behavior, such as proliferation, differentiation, migration, and response to drugs or environmental cues, in a more physiologically relevant context [88]. Thus, it has enabled the generation of *in vitro* models of tumors [89] and cancer [90]. Additionally, the advancement of cell spheroids has led to the development of cell organoids, which maintain the structure and key functions of specific organs or tissues. Organoids typically consist of multiple cell types organized in a manner resembling the architecture of the organ they mimic. These *organ-on-chip* models show promise in studying human physiology and cancer in an organ-specific context [91].

Among *in vitro* techniques for culturing cell spheroids or organoids [92, 93], microfluidic devices have emerged as powerful tools due to their capacity to provide a high level of control over microenvironmental conditions [94, 95]. This capability enables the study and manipulation of specific chemical and mechanical cues to explore more detailed cell responses. Moreover, the well-defined environment in microfluidic devices closely mimics physiological conditions, allowing for the exploration of how cells interact with their surroundings. These attributes, coupled with lower budget requirements compared to other experimental methods, confer value to microfluidic devices in the study of cell behavior.

Next, at a higher level of abstraction, *in vivo* models, commonly utilizing mice, provide a more in-depth exploration of biological aspects. In this context, various techniques, not only those employed in clinical approaches but also exclusive methods inapplicable to human subjects, come into play. Typically, it involves the induction of a disease or genetic manipulation, followed by monitoring the signs of the disease or response to treatment. These models are invaluable instruments for studying intricate biological processes and testing pharmaceutical interventions. However, ethical considerations related to mice models prompt a gradual transition toward alternative methodologies, aiming ultimately to diminish reliance on these models over time.

Finally, in addition to traditional laboratory-based experiments conducted in controlled settings, clinical approaches focus on investigating biological processes and disease mechanisms in human subjects. In these clinical approaches, acquiring data from patients to assess the status and progression of their diseases is fundamental. Common data acquisition techniques include established diagnostic tools such as Magnetic Resonance Imaging (MRI), X-ray imaging, and blood tests. While these methods predominantly offer macroscopic insights, advancements in T2-Weighted MRI techniques allow the extraction of microscopic data, enabling the derivation of cell concentrations and assessments of vascularization levels. Biopsies also constitute another valuable clinical tool that permits the microscopic observation of tissue samples under a microscope to evaluate diseases at the cellular level. Thus, clinical approaches serve as indispensable tools for disease diagnosis and prognosis. However, they play a crucial role in predicting the evolution of diseases rather than providing in-depth insights into their underlying nature.

1.4 Computational biology

Computational models have risen as a powerful tool in unraveling the intricacies of cell behavior and tumor growth. Their capacity to simulate complex systems, generate hypotheses, and incorporate diverse data sources makes them indispensable in advancing our understanding of cellular processes and facilitating the development of innovative approaches for disease diagnosis, treatment,

and prevention.

While experimental models remain integral in studying biological systems, computational models excel in their ability to simulate intricate biological processes, offering distinct advantages that are challenging to achieve through experimental models alone. These simulations provide a comprehensive view of dynamic interactions within cellular systems, offering insights into mechanisms that might evade observation in a laboratory setting. Additionally, computational models afford researchers the opportunity for rapid and cost-effective exploration of diverse scenarios. Unlike experimental models, which might be constrained by practical limitations such as time, cost, and ethical considerations, computational models can simulate a wide range of conditions and perturbations, enabling the investigation of hypothetical scenarios and accelerating the pace of discovery. One of the inherent strengths of computational models lies in their capacity to isolate variables and meticulously analyze their individual effects, a feat often challenging within experimental models. In experimental setups, manipulating a single variable without impacting others can be a daunting task, leading to complex interactions that confound observations. Conversely, computational models offer a controlled environment, where variables can be systematically adjusted and their impacts studied in isolation. This ability to disentangle and scrutinize individual variables provides a clearer understanding of their specific influences on cellular behavior and tumor dynamics. Consequently, computational models offer a distinct advantage in discerning the true effects of variables, a task often muddled in experimental settings, where disentangling the effects of multiple interacting variables can obscure the real underlying relationships. Furthermore, these models serve as invaluable predictive tools. By integrating data from various sources and incorporating theoretical constructs, they facilitate the generation of hypotheses that can guide experimental design. This predictive capability aids in streamlining experimental efforts by focusing attention on the most promising avenues, potentially reducing the time and resources required for empirical validation. Importantly, computational models can bridge gaps between experimental findings, providing a framework for interpreting complex data sets and elucidating underlying mechanisms. They complement experimental observations by offering a platform to test hypotheses, refine theoretical models, and generate new insights that can guide and inform experimental investigations. Moreover, these models serve as predictive platforms, enabling the generation of hypotheses that can be experimentally validated. Through their iterative refinement, they offer a means to test theoretical constructs and explore scenarios that might be challenging or impossible to investigate solely through experimental approaches.

In essence, computational models stand as catalysts in pushing the boundaries of our understanding of cellular dynamics and tumor growth. Their role in integrating diverse datasets, simulating intricate biological processes, and guid-

ing experimental inquiry underscores their significance in driving advancements in biomedical research and therapeutic innovation. There exist broad types of computational models, whose selection would depend on the particular application. In general, depending on the spatial scale of resolution, we can distinguish between discrete and continuum models from the microscale (cell level) to the macroscale (tissue level) respectively.

1.4.1 Discrete models

Discrete models are a specific type of mathematical model widely used to study cell behavior and biological systems [96, 97]. It involves representing cells as distinct, separate entities rather than as continuous entities. Within this discrete approach, Agent-Based Models (ABMs) stand out as a versatile framework for simulating complex biological systems. In ABMs, individual cells are represented as autonomous agents that interact with their environment and other agents. This methodology allows for the study of collective behavior, emergent properties, heterogeneity, variability in behaviors between cells, and spatial dynamics, offering a microscopic perspective on individual cell behavior. Through the incorporation of experimental data and established biological principles, these models can provide insights into the impact of cell-cell and cell-environment interactions, mechanical forces, and biochemical signaling on cellular behaviors such as migration, proliferation, and differentiation.

Multiple types of ABMs exist, varying based on how cells are represented and the level of detail considered (Figure 1.3**A**). One common classification depends on how the agent's position is described. On the one hand, lattice models fix cell positions within a mesh or lattice, resulting in a discrete cell position. This implies that cells can only exist in specific fixed positions in space and their movement is discrete. On the other hand, off-lattice models allow cell positions to be free in space, enabling continuous movement.

Among lattice models, we can further classify them based on how many agents can occupy the same lattice. In the case of Lattice Gas Cellular Automata (LGCA), several agents can coexist in a single lattice, where each agent within the lattice represents a subpopulation of cells. While this reduces computational costs, it compromises the accuracy of cell representation and its individuality. Cellular Automata (CA) enhances the representation of individual agents by maintaining a lattice size equal to the agent size, therefore, each occupied lattice is itself an agent. Lastly, in Cellular Potts Model (CPM), the lattice size is smaller than the agent size, allowing a cell to be represented by multiple agents spanning different lattice positions. This enables an accurate representation of cell shapes at the expense of increased complexity.

Numerous computational models have been developed to study biological processes. Mallet and De Pillis [98] built a CA to describe tumor growth and

its interaction with the immune system, while Rubenstein and Kaufman [99] presented a CPM to analyze the growth and invasion of glioma, taking into account the fibrous component of the ECM. A notable aspect of CPMs is their effectiveness in reproducing spatial patterns based on experimental data. Engelberg et al. [100] successfully used a CPM to simulate cystogenesis, reproducing the temporal evolution of the pattern of Madin-Darby canine kidney (MDCK) cyst growth in collagen cultures, and Oers et al. [101] integrated a CPM model with a finite element model to study, from a biomechanical perspective, the formation of blood vessel-like structures in cell cultures. Additionally, lattice models are not limited to 2D representations, some works have also developed 3D models. Shirinifard et al. [102] developed a 3D CPM model of tumor growth and tumor-induced vascularization in hypoxic conditions, Scianna et al. [103] studied 3D cell migration with a CPM, providing a relationship between ECM topology and cell locomotion directionality, and Zhang et al. [104] studied cancer heterogeneity in brain tumors with a 3D CA.

Regarding off-lattice models, agents are described by a radius and the position of their center, allowing them to adopt any position in space. Here, we can make a similar subclassification based on the relationship between the size of the agent and the cell. In cluster models, the size of the agent is greater than that of the cell, so each agent represents a subpopulation of cells. Thus, the interactions between agents represent interactions between populations, and the agents condense the information about population behavior, reducing the computational cost. Then, the Center-Based Model (CBM) maintains an equal size for both the agent and the cell, providing a more detailed description of the cell population since each agent precisely represents an individual cell. The SEM involves representing a cell with multiple smaller agents. In this scenario, the cell's actual shape representation arises from the interactions among agents within the cell, increasing the complexity of interactions both within cells and between agents of different cells. Additional off-lattice models aiming for a dynamic representation of populations of interacting cells include those based on Delaunay-Voronoi tessellations to estimate cell shapes from agent centers or deformable cell models, in which cell surfaces are discretized and triangulated with viscoelastic elements to reproduce the mechanical response of the cytoskeleton.

Off-lattice models have been also highly employed in diverse biological contexts. Schlüter et al. [105] used a CBM to investigate cell migration, focusing on the interactions between individual cells and matrix fibers, Macklin et al. [106] developed a CBM to study ductal carcinoma, utilizing a patient-specific calibration method based on histopathologic data achievable from a single biopsy, or Kempf et al. [107] simulated radiotherapy on a 3D tumor spheroid with a CBM, taking into account cell cycle-specific radiosensitivity. Furthermore, there are other off-lattice models dedicated to capturing cell shape. Gardiner et al. [108] utilized a SEM to simulate, with multiple particles, the cell spreading experiments of single

cells and the deformation of epithelial layers composed of multiple cells. Schaller and Meyer-Hermann [109] developed a 3D Voronoi-Delaunay cell model to evaluate the induction of necrosis based on the ratio of oxygen and glucose uptake rates in tumor spheroids. Van Liedekerke et al. [110] presented a 3D deformable cell model to emphasize the importance of detailed representations of cell shape and mechanics in computational models, applying their model to the regeneration of hepatocytes after drug-induced liver damage.

Nonetheless, utilizing ABMs often presents certain challenges [111]. A direct challenge emerges in abstracting biological phenomena into the ABM. ABMs require decomposing observed behaviors into cell-level behaviors, and translating this information from the biological system is inherently intricate. Localized information, such as cell-cell, cell-matrix interactions and biochemical or metabolic pathways can be challenging. Consequently, ABMs often resort to heuristic-type laws (if-conditions), introducing predefined rules to govern cell behavior, which may reduce the model's realism. Related to this, ABMs frequently incorporate numerous parameters, and their translation into physically meaningful terms can be non-trivial. This introduces an element of uncertainty, questioning whether the established rules are robust enough to yield the expected emergent behavior and if the driving characteristics behind that behavior are accurately identified and sufficient. Another limitation in ABMs is the computational cost, acting as a significant factor. This is especially true for models aiming for a nuanced representation of cell shape or deformation, or when increasing the number of simulated agents. Thus, the computational burden escalates, as the number of interactions to calculate scales quadratically with the number of agents. However, with the increasing adoption of Graphics Processing Units (GPUs) for its efficiency in handling a multitude of small calculations required for each agent in ABMs and advancements in parallel computation [108, 112], the computational burden has decreased, allowing for an expansion in the number of simulated agents.

In summary, despite the limitations, ABMs offer a powerful framework to simulate cells, capturing the heterogeneity among individual cells and also for simulating experimental setups in which the data is obtained at the cell level.

1.4.2 Continuum models

The continuum approach presents a distinct representation of cell populations wherein individual cells are not depicted separately but as a collective entity. This collective representation often involves employing a continuum variable that describes the concentration or density of cells within a given region. Thus, the continuum approach offers a more macroscopic view of cellular dynamics, focusing on population-level behaviors rather than the discrete interactions of individual cells or cell heterogeneity. This abstraction allows for the analysis of large-scale phenomena and system-level responses, particularly in scenarios

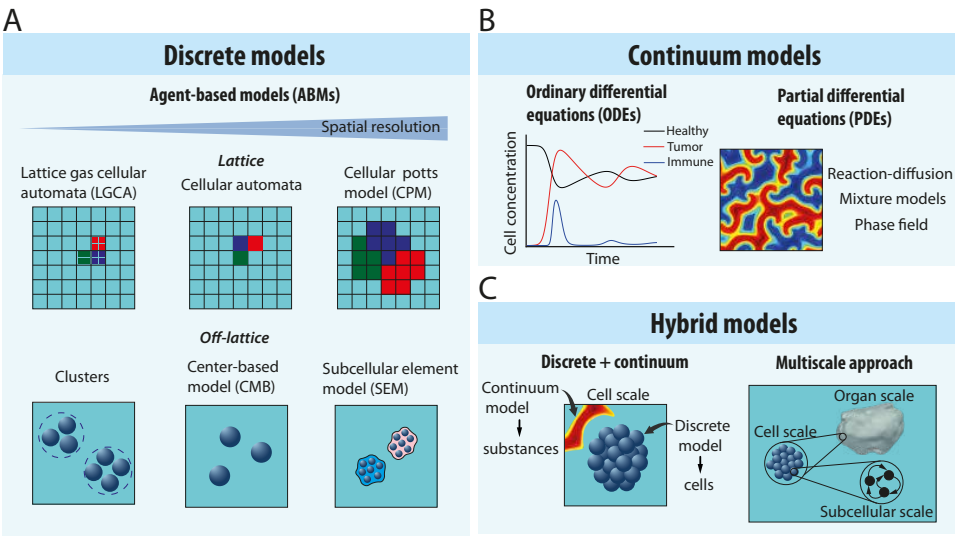

Figure 1.3: ***Review of computational models in biology.*** ***A*** *Classification of discrete models through ABMs. This categorization distinguishes between lattice and off-lattice models, displaying some of the most characteristic models within each subcategory.* ***B*** *Classification of continuum models based on mathematical representation. This division includes ODE models, describing only temporal evolution, and PDE models, allowing both temporal and spatial representations.* ***C*** *Hybrid models. Here, the integration of discrete and continuum approaches to simulate biological cells and chemical species respectively at the cell level is outlined, alongside fully multiscale approaches, where distinct spatial scales are considered.*

where the detailed behaviors of individual cells might not be the primary focus. Additionally, the continuum approach facilitates the application of mathematical and computational models, providing valuable insights into population dynamics, spatial patterns, and the overall trends of cellular behaviors within complex biological systems.

Continuum models in biology are classified based on several criteria. One criterion to classify them could depend on the mathematical representation of the population (Figure 1.3**B**). First, on a very low level, we can distinguish models based on Ordinary Differential Equations (ODEs), in which the evolution of the variables is only described in time, so the evolution in space is neglected. In this case, tumor growth is usually modeled through mathematical functions such as logistic, Malthusian, Alle, Bertalanffy, or exponential law [113]. Interactions of cells with chemicals or the immune system can be represented by adding additional terms to the ODEs, hindering or favoring tumor growth. Hence, ODEs have been historically used to simulate solid tumor growth [114], immune response [115], and radiation and chemotherapy treatments [116, 117]. Sarkar and Banerjee [118] simulated the effect of the immune system response in the tumor with a prey-predator model, DePillis et al. [119] simulated cancer growth considering

the immune system response coupled with treatments such as chemotherapy and monoclonal antibodies or Browning et al. [120] developed a predictive model of patient-specific tumor response to radiotherapy.

Models based on Partial Differential Equations (PDEs) gained popularity since they can reproduce spatial variability, obtaining a better representation of biological phenomena. In this category, PDEs based on reaction-diffusion equations have been used in many areas of biology. Cell populations and chemicals are quantified as concentrations, with their variations modeled using reactive terms. Additionally, cell movement and the homogenization of substances are represented by a diffusion process, usually assumed to follow Fick's second law. This approach has proven effective in simulating pattern formation and morphogenetic processes [121]. Ferreira Jr et al. [122] introduced a reaction-diffusion model for avascular cancer growth including competition for nutrients among normal and cancer cells, Chaplain and Lolas [123] presented a reaction-diffusion model to study cell invasion in which cell migration is affected by a chemotaxis and haptotaxis process. Mechanochemical models that also couple to the reaction-diffusion equation with PDEs, constitutive laws to reproduce mechanical behavior, Weis et al. [124] developed a mechanically coupled model to consider tissue stiffness restricts cell diffusion to predict breast tumor response to chemotherapy therapy or Chen et al. [125] also considered biomechanical properties of the kidney and surrounding tissues to study patient's specific kidney tumor.

Then, we distinguish mixture models in which components are considered distinct mixtures rather than concentration-based representations. Ciarletta et al. [126] studied solid stress inhibition in avascular tumor growth by employing a mathematical model that incorporates a single-phase continuum with dissipative evolution and a multi-scale approach considering surface growth. Other models also consider poroelastic material properties as Roose et al. [127], who modeled a tumor spheroid in poroelastic media with mechanoregulation of cell proliferation in response to the solid stress or Fraldi and Carotenuto [128] explored tumor growth dynamics by incorporating nonlinear poroelastic behavior and investigated how stress and fluid content dynamically change within the tumor, influencing the nourishment and starvation processes.

Finally, in phase field models, a distinctive approach is taken where, rather than explicitly representing cells and chemicals with concentrations, they are conceptualized as interfaces and boundaries within the system. Lorenzo et al. [129] used a phase field method to consider the transformation between healthy and tumor cells with a reaction-diffusion model for nutrient consumption for prostate cancer, suggesting that the particular geometry of the tumor as well as the specific anatomy of the prostate influences tumor morphology. Xu et al. [130] employed the same idea to study vascular tumor growth, coupling the phase field method to represent the tumor interface with reaction-diffusion equations to simulate the oxygen and the tumor angiogenic factor released by malignant cells. Vi-

lanova et al. [131] used the phase field method to investigate the creation of new vascularization networks (angiogenesis) induced by tumors and their regression.

Continuum models have certain limitations. While simulating large populations can be advantageous, it also entails drawbacks. Simulating these large scales involves neglecting the individuality of cells in favor of capturing broader collective behaviors. Consequently, they encounter challenges in capturing emergent phenomena resulting from individual cell behaviors or interactions. Stochastic effects, cell polarization, mutations, and other cellular processes are often difficult to model accurately within continuum frameworks. Furthermore, continuum models struggle to depict spatial heterogeneity, a common feature of many biological systems where behaviors or properties vary across different regions or cell populations. Hence, representing cell populations as homogeneous entities oversimplifies the system's complexity, failing to capture population heterogeneity accurately. Lastly, simulating *in vitro* cell experiments with continuum models may present difficulties, as these experiments typically gather data at the cell scale, making it challenging to relate to macroscopic properties

Overall, continuum models are essential in computational biology, serving as mathematically rigorous tools to gain insights into macroscopic phenomena and conduct patient-specific simulations

1.4.3 Discrete vs continuum approaches

The simulation of biological systems often involves a choice between employing discrete or continuum approaches. Determining the most suitable model type relies heavily on the specific nature of the problem. Both discrete and continuum approaches offer distinct advantages and drawbacks. Typically, the decision on the model will fall within the objectives of the study and the type of data available.

Discrete models excel in capturing detailed interactions between individual components within the system. These models are particularly advantageous when studying phenomena reliant on individual behaviors, cellular interactions, or spatial heterogeneity. Also, they are straightforward to integrate data from cell experiments. However, they might demand more computational resources and data to capture the complexity of the system accurately.

Conversely, continuum models provide a more macroscopic view, representing populations or fields, and are advantageous in describing overall system trends and behaviors. These models are beneficial when examining large-scale processes, such as tissue-level phenomena or global system responses. Moreover, they are more simple than discrete models to integrate with real patient data, which usually are on a macroscopic level. However, they might oversimplify individual interactions or fail to capture fine-scale heterogeneities inherent in biological systems.

The choice between these approaches hinges on the study's specific aims,

the scale of the biological process under investigation, and the available experimental or observational data. For instance, when dealing with aggregated data, continuum models might be more suitable for providing a broader understanding of the system's behavior. Conversely, if detailed cellular-level insights are crucial and sufficient data are available to support individual interactions, discrete models could offer a more comprehensive view. Ultimately, the decision on model selection requires a careful balance between the intricacy needed to represent the biological system accurately and the computational feasibility within the constraints of available resources.

1.4.4 Hybrid models

Hybrid models arise as an effort to obtain the benefits from the discrete and continuum perspectives and reduce the disadvantages they have separately [132, 133]. However, this terminology is usually employed indiscriminately and includes different possibilities of hybridization (Figure 1.3**C**)

On the one hand, hybrid models sometimes refer to the combination of discrete models with ABMs for simulating cells and ODEs or PDEs to simulate, with a continuum approach, the presence of chemicals and substances in the environment, such as oxygen or glucose. Thus, the scale of simulation is at the cell level, but two different approaches are employed for cell simulation and chemical representation. Ramis-Conde et al. [134] proposed a hybrid cancer cell invasion model with off-lattice discrete cells that interact with the surrounding environment composed of the ECM and degrading enzymes, modeled with PDEs. Jeon et al. [135] also developed a hybrid model to study tumor growth and invasion with an off-lattice center-based approach for the cells. The ECM, the matrix-degrading enzymes secreted by cells, and the oxygen concentration are simulated with PDEs. Vilanova et al. [136] created a model to simulate the formation of new blood vessels (angiogenesis) in tumors to study the role of fluid flow. The vasculature network is simulated using a continuum approach, with the creation of new branches being driven by discrete agents that represent tip endothelial cells. On the other hand, hybrid models are also used to refer to approaches that integrate scales from the nanoscale (molecules), microscale (cells), and macroscale (tissues), communicating information between scales [137, 138]. These models usually aim to simulate large populations of cells at the tissue level while preserving the events that occur at the cell level. In this multiscale approach, continuum models are employed at the macroscale, integrated with discrete models to simulate the cells. Chauviere et al. [139] presented a theoretical framework showcasing its capabilities for tumor growth, incorporating a discrete model to represent the individual dynamics of cells, coupled with a continuum Dynamic Density Functional Theory (DDFT) model operating at the tissue scale. Borau et al. [140] created a multiscale model to simulate tumor growth by integrat-

ing cell behavior across different scales. They simulated cells at the tissue level with a reaction-diffusion model that communicates with an agent-based model to determine cell behavior at the cell level. Rocha et al. [141] proposed a hybrid three-scale model composed of an agent-based model for cell dynamics, which also incorporates a molecular model for signaling pathways, and a PDE model for nutrients and growth factors at the tissue scale to study tumor dynamics

1.4.5 Digital twins

The culmination of computational models in biology might be the creation of Digital Twins (DTs). A DT serves as a virtual replica of the physical entity, offering a detailed and dynamic representation of its real-world counterpart. In the field of biology, a digital twin involves generating a model for patients' tissues, incorporating patient-specific data spanning various scales, including genetic aspects, tissue geometries and information, and other relevant data. This digital twin, designed as a multiscale model, provides insights not only into macroscopic behaviors but also sheds light on behaviors across different scales. This represents a paradigm shift in healthcare, enabling predictions about patients' diseases and the identification of optimal treatments [142]. Currently, DT models have been developed to predict specific disease prognoses and optimize treatments [143]. For instance, Borau et al. [140] incorporated patient-specific Magnetic Resonance Imaging (MRI) data into their multiscale model to predict the evolution of neuroblastoma cancer, and Chaudhuri et al. [144] developed a DT methodology for optimizing radiation therapy for glioma patients.

DTs will pave the way for a future where each individual possesses their digital counterpart, enabling continuous health monitoring, evaluation, and prediction. In this regard, real-time acquisition wearables, such as those monitoring heart rate, glucose levels, or oxygen levels, will enhance the DT by providing real-time health status information through Internet of Things (IoT) technology. Thus, a clear challenge to keep advancing in the development of DT arises from managing large volumes of multiscale, individual-specific data that dynamically changes over time and between different people [145, 146]. Incorporating such vast amounts of data implies that DT models can adapt precisely between individuals and also in real-time to data variability. Therefore, the near future of DT involves developing calibration methods incorporating ML to dynamically and precisely adapt DT models.

1.5 Model calibration and validation in computational biology

One imperative aspect of computational models in biology is their reliability and accuracy. Ensuring that simulated biological behaviors closely align with

observed realities is essential for obtaining meaningful predictions. Model calibration and validation address the question: when faced with a known biological phenomenon, how can we guarantee that our computational model accurately reproduces these observations to make reliable predictions? This typically involves fine-tuning the model's parameters to align with empirical data, followed by an assessment of the computational results against observed analogs. This process not only enhances the credibility of simulations but also provides us with solid conclusions and a more precise understanding of biological systems and has the potential to enhance clinical tasks such as patient stratification and survival prognosis [147, 148].

However, in computational biology, the diverse range of model types poses a challenge when designing optimization procedures. Computational biology models span from discrete [96, 97] to continuum models, with hybrid models combining elements of both approaches, each with numerous subcategories. Additionally, the biological data used to evaluate the outcome of computational models originates from distinct sources, such as cell-level or tissue-level data, and exhibits deterministic or stochastic characteristics, further complicating the evaluation and robustness assessment of model performance. Moreover, the multitude of alternative optimization methods, ranging from gradient-based methods to sampling and other non-gradient-based methods, adds another layer of complexity. As a result, researchers may encounter difficulty in gaining a comprehensive perspective on employing appropriate calibration techniques across different model types for specific applications.

In addition to these classical approaches, ML methods are gaining considerable attention and have emerged as game-changing tools. Their ability to learn from nonlinear problems and adapt to data variability makes them particularly appealing to biology. However, despite their growing popularity in other fields, their adoption in computational biology is not widespread. Therefore, there is an opportunity to leverage ML methods and apply them to computational biology to enhance the adaptability of our models to biological data variability.

In this section, we will present some classical approaches for parameter estimation in computational biology that can be used in both discrete and continuum models. Also, we will discuss the integration of machine-learning methods into computational biology and explore new opportunities for combining modern machine-learning techniques with computational models in biology.

1.5.1 Classical approaches

Classical approaches in optimization or calibration involve well-established methods used historically in various fields to estimate the model's parameters. These approaches vary in their computational requirements, convergence speed, and applicability to different optimization or calibration problems. Here, we will

distinguish between sampling methods, gradient-based methods, and other non-gradient-based methods and discuss some of the main techniques of each one and the ones employed in this doctoral thesis.

Sampling methods

Sampling methods explore the parameter space by generating points and then evaluating the performance using an objective function. At its simplest, the *grid search* method offers a straightforward approach where parameters are evaluated across a predefined set of values, exploring all possible combinations within the range. However, this method can be highly computationally expensive, especially when the parameter's range is wide, as its cost increases significantly with the number of parameters. In contrast, the *random search* introduces randomness to the grid search by randomly sampling parameter combinations from the grid, thereby avoiding the exhaustive requirements of the grid search [149]. However, the random search may not effectively explore the parameter space or find the optimal solution since it does not utilize prior knowledge of generated samples, rendering the process inefficient.

To address these limitations, prior knowledge can be incorporated into the exploration of the parameter space. This is exemplified by Monte Carlo (MC) method, where parameter values are randomly sampled from probability distributions. In this way, the efficiency and effectiveness of the parameter combination can be enhanced. Using a MC method Ruiz-Arrebola et al. [150] investigate the growth of multicellular tumor spheroids using an on-lattice agent-based model. Within this model, MC sampling is utilized to simulate stochastic cell processes such as cell proliferation and to sample parameter values from Gaussian distributions for parameter estimation. Also, it is worth highlighting that MC methods are widely employed for estimating radiation doses (dosimetry) and have become a standard in radiation treatment planning [151–153]. Similarly, the *inverse transform method* samples values between 0 and 1, treating them as probabilities, and calculates the corresponding parameter values using cumulative distribution functions.

Alternatively, *Bayesian optimization* offers a more sophisticated approach that does not assume any specific functional form for the objective function. Instead, it approximates the objective function, typically using a Gaussian Process (GP), which serves as a surrogate. The GP is iteratively updated, guiding the exploration of regions with high uncertainty and exploitation of regions with expected high performance. This makes Bayesian optimization an efficient optimization technique that systematically explores the parameter space while minimizing the number of evaluations of the objective function. Hervas-Raluy et al. [154] use Bayesian optimization for parameter inference in a continuum multiphase porous media model to reproduce the growth of neuroblastoma tumor spheroids in 3D

microfluidic devices. In their work, the authors conveniently demonstrated how Bayesian calibration effectively captures the uncertainty arising from experimental data. In general, special care must be taken when dealing with stochasticity in the model, since the prediction of the model does not depend solely on the parameters combination but also on randomness, therefore, it has to be taken into account when evaluating model accuracy. In this regard, Lima et al. [155] develop a coarse-grained hybrid off-lattice agent-based model to reproduce tumor growth with cell process (cell movement, growth, death, and phenotypic transitions) represented as stochastic processes. Here, the Bayesian calibration employed a generalized likelihood function to quantify the probability of the observed data given a set of model parameters for assessing the associated uncertainties while handling with stochasticity of the model.

While these are some of the prominent sampling methods, it is important to note that there are numerous other techniques available, each with its unique characteristics and applications. Additionally, within these methods, there exist further subdivisions and variations in their functioning and applicability. Despite their diversity, sampling methods remain indispensable tools in various fields, offering versatile approaches to exploring parameter spaces, optimizing models, and gaining insights into complex systems.

Gradient-based methods

Gradient-based methods focus on moving along the minimizing direction of the objective function rather than generating samples. Therefore, the objective function must be differentiable. One basic first-order optimization algorithm within this category is *gradient descent*. The main idea is to iteratively update the model's parameters to move in the negative direction of the objective function with respect to those parameters. Thus, it efficiently explores high-dimensional spaces to find the minimum of objective functions, although it can encounter convergence issues and become stuck in local minima. A variant of gradient descent is *stochastic gradient descent*. This method randomly samples subsets of the dataset to compute the gradient, introducing stochasticity into the optimization process and improving computational cost and convergence. However, both methods are susceptible to oscillations, especially in non-convex problems, as well as issues related to the learning rate, which determines the magnitude of parameter updates based on the gradients.

Newton-Raphson method, a second-order method that utilizes the Hessian matrix representing the curvature of the function, converges toward the minimum of the function more efficiently than first-order methods such as gradient descent. Moreover, the Newton-Raphson method adjusts the step size dynamically based on the curvature of the objective function, providing more accurate parameter updates. However, second-order methods are computationally more expensive and

may not be suitable for large-scale optimization problems. An alternative to the Newton-Raphson method is *Quasi-Newton methods*, such as the *Broyden-Fletcher-Goldfarb-Shanno algorithm*, which, instead of computing the Hessian matrix, uses an approximation from gradient evaluations, reducing the computational complexity.

In summary, gradient-based methods offer efficient approaches for optimizing objective functions in high-dimensional spaces.

Other gradient-free methods

There exist other techniques besides sampling methods that do not rely on gradient calculation to minimize an objective function. Evolutionary algorithms, such as *genetic algorithms*, find optimal parameters by iteratively combining them, inspired by the principles of natural selection. They create an initial candidate of parameter combinations (population), and after evaluating their performance in solving the optimization problem, propose a new population using naturally-inspired operations like selection, crossover, and mutation. Genetic algorithms are capable of finding optimal or near-optimal solutions in complex optimization problems, including nonlinear and multimodal objective functions. Additionally, they are inherently parallelizable and can effectively handle problems with large search spaces. Qods et al. [156] utilized the genetic algorithm to minimize the tumor size predicted at the end of therapy by finding the optimal drug regimens for patients with colorectal cancer using a multi-drug tumor growth ODE model. Akasiadis et al. [157] employed a genetic algorithm to optimize the parameters of an agent-based model that integrates a Boolean network model for simulating intracellular signaling and cell fate in the presence of drugs. The study showcased various types of objective functions (commonly referred to as fitness functions in the context of genetic algorithms) and their integration in a parallel execution context to reduce time complexity. However, genetic algorithms may require a large number of iterations to converge, leading to longer computation times. *Pattern search*, on the other hand, proposes an initial set of parameter combinations but then explores the parameter space differently. It determines the new parameter set along predefined directions from the previous best combination, adjusting the step size as it approximates the optimal values. Thus, pattern search algorithms are relatively easy to implement and are easily parallelizable, allowing for non-convex, multimodal optimization without the need for gradient computations. Knopoff et al. [158] presented a pattern search method to estimate the parameters of the PDE model to replicate the radius of spherical tumor growth under drug treatment. They illustrated how incorporating time data into the objective function, rather than solely calibrating the data at a final time, leads to improved parameter estimation. Pattern search main weaknesses are that they have limited scalability when applied to high-dimensional problems, slow con-

vergence compared to gradient-based methods, and may be prone to converging to a local optimum.

1.5.2 Machine learning approaches

Artificial Intelligence (AI) is a promising field that is becoming more and more relevant in different disciplines. AI is a general concept that refers to performing tasks typically associated with humans. A subset of AI is Machine Learning (ML), which are the techniques that permit learning from existing data to generate insights or predictions when encountering new data [159]. Inside ML, Deep Learning (DL), also called deep neural learning or deep NNs, represents a specific subdomain that utilizes Artificial Neural Networks (ANNs). ANNs are a computational model inspired by the brain's biological Neural Network (NN). It consists of nodes organized through layers, which receive input signals, process the information through activation functions, and propagate the output to nodes in the next layer through connections. By iteratively adjusting the connection strengths (weights) between nodes during training, ANNs can learn to approximate complex functions and make predictions on new data.

ML techniques have wide applications in biology and have shown to be extremely useful in this context [160]. For instance, for protein structure prediction [161], disease prediction [162], or biomedical image segmentation [163] among many more applications. The increasing quantity of data collected over time, in addition to new generative models that generate more data, contributes to the fact that mechanistic models are maybe being relegated in favor of ML models. Moreover, the way that ML handles complex non-linear relationships between the data, the automation, and efficiency in predictions that lead to real-time analyses are making these models attractive compared to their mechanistic counterparts. However, most current works that integrate biological and medical data focus on identifying correlations among data to then infer the behavior of the system, ignoring the fundamental laws of physics and lacking physical interpretation [164].

The integration of ML with mechanistic models has the potential to benefit from both approaches. While ML excels in uncovering complex patterns and relationships within the data, mechanistic models aim to identify causality between data and outcomes [165]. Together, they can enhance prediction accuracy, efficiently adapting to dynamical data and resulting in solutions with meaningful interpretations. However, there is a lack of progress toward integrating ML and mechanistic modeling in biology [166]. Therefore, there is a promising opportunity to bridge the gap between mechanistic modeling and ML in biology that needs to be pushed forward [167]. Here, we will discuss the integration of ML models in computational biology models and provide some perspectives on the future of the field.

Machine learning for computational biology

Combined with computational biology models, ML frameworks offer a promising avenue to replicate complex dynamical behaviors and intricate biological phenomena. We can distinguish various approaches in this combination.

The first approach involves generating a surrogate of the mechanistic model by using an ML model to learn how the inputs relate to their outputs [168–170]. This approach is particularly relevant when obtaining a direct solution with the mechanistic model becomes computationally prohibitive due to high computational costs. Thus, this surrogate model can be used for the prediction of responses instead of the mechanistic model, enabling fast computation. Additionally, it facilitates more robust sensitivity analyses while also reducing computational time [171]. For instance, Liang et al. [172] exemplify the effectiveness of generating a surrogate of a Finite Element (FE) model for estimating the stress distributions of the aorta through DNNs. Their study demonstrates minimal errors in stress distributions with the surrogate model and a significant improvement in computing time (from 30 min to 1 s). Karabelas et al. [173] develop a patient-specific four-chamber heart Computational Fluid Dynamics (CFD) model based on the Navier-Stokes-Brinkman equations. They then generate a surrogate of the physics-based model using a Gaussian Process Emulator (GPE). Subsequently, they conduct the first variance-based global sensitivity analysis (GSA) using this surrogate model to identify whole-heart CFD features, a task not achievable with traditional models in realistic timeframes. Thus, this approach holds promise in reproducing mechanistic models to make real-time decision-making without compromising predictive power in biology [174], which is especially relevant for clinical decision-making.

The next step would be to employ NNs for reduced-order modeling, aiming to extract critical information and alleviate the complexity of biological simulations. These methods facilitate the identification of important factors and reduce the number of variables involved in the process, thereby enhancing comprehension of their significance. Consequently, eliminating non-relevant variables can indirectly impact the effective number of parameters of the model.

Moreover, prior physics knowledge can be fully integrated into NNs to respect the physical laws described by non-linear PDEs through Physics-Informed Neural Networks (PINNs) [175]. This integration of physics laws into the NN architecture allows PINNs to capture complex nonlinear relationships in the data while ensuring that the learned solutions remain physically consistent. This is achieved by incorporating the residual of the approximated physics-based model, i.e., the PDE model, into the loss function (physics-informed loss), along with the loss between the predicted and the known solution (data-driven loss) [176]. Subsequently, the NN weights are optimized to minimize this combined loss. Additionally, PINNs are applicable for parameter optimization with minimal additional

effort alongside learning the solution. Sainz-DeMena et al. [177] explore PINNs to fit diffusion-corrected pharmacokinetic models to synthetic Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) data. This imaging technique assesses the vascular properties of tissues, proving valuable for diagnosing and monitoring tumors. The study shows that PINNs outperform traditional algorithms based on FEMs in accurately retrieving vascularization parameters from diffusion-corrected tracer-kinetic models, exhibiting greater robustness against noisy or incomplete data. Zhu [178] utilize PINNs to solve a time-dependent reaction-diffusion PDE model of glioblastoma tumor growth and infer biophysical parameters. The study showcases the significant speed-up that PINNs offer compared to traditional simulation methods, particularly in the context of solving the inverse problem to replicate patient-specific evolutions, which typically demands numerous evaluations to calibrate the mechanistic model. Zhang et al. [179] also simulate glioblastoma tumor growth using a reaction-diffusion PDE model. In this work, they propose a sequential methodology: as a first step, an initial characteristic solution by solving the PDE with Finite Difference Method (FDM) is obtained, employing non-representative tumor growth and diffusion rates parameters. Subsequently, they train a PINN to map this initial solution to a better approximation to finally fine-tune another PINN to estimate the patient-specific tumor growth and diffusion rate from the improved solution.

Finally, at the lowest level of abstraction, NNs are employed solely to approximate the model's parameters [180]. In this approach, the NN approximates a set of parameters used by mechanistic models based on observed data. Subsequently, this model is executed with the estimated parameters to generate predicted data. The prediction is then compared with the observed data to calculate the loss, which is backpropagated to optimize the NN weights and improve parameter approximation. Gaskin et al. [181] developed this methodology for parameter estimation in ABMs with stochastic processes and uncertainty quantification, considering various data types such as time-series data, steady-state data with only a single time frame, noiseless and noisy data. They demonstrate its potential in modeling epidemic infection diffusion and economic activity, outperforming other classical Markov Chain Monte Carlos (MCMCs) in terms of accuracy and speed.

Machine learning for reproducing cell behavior

Previously, we discussed current opportunities in combining biological mechanistic models (either continuum or discrete models) with ML. However, integrating ML with discrete models offers another perspective, which has the potential to represent a paradigm shift in the simulation methodologies for biological systems. Thus, ML can grant cells in discrete models the capacity for autonomous decision-making [182, 183] based on the interactions with their environment and

other cells, dynamically adapting the behavior to the evolving conditions.

Considering cellular behavior, cells possess specialized receptors that capture diverse information from their microenvironment. These receptors enable cells to perceive chemical signals and mechanical cues from the environment and neighboring cells. Subsequently, the transduction of this multifaceted information entails intricate interactions between received signals and the cell's internal machinery, encompassing Deoxyribonucleic Acid (DNA) and signaling pathways. Within the cell, this integration initiates a cascade of biochemical reactions and signaling pathways, eventually culminating in the activation of specific cellular responses.

In this context, we can clearly state an analogy of this process with NNs (Figure 1.4). The information that the cell receives from the receptors that capture the information from their microenvironment and other cells is represented in the input layer. Then, the processing of this information takes place through the hidden layers of the NN, which analogously simulates the intricate protein pathways within the cell. Lastly, the cell activates its biological response in the same way that the NN activates the outputs of the NN.

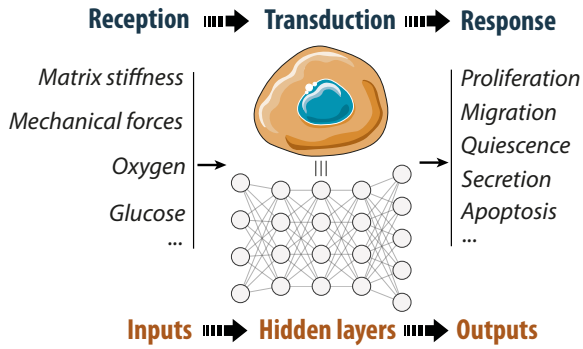

Figure 1.4: ***Representation of cell signaling using neural networks.*** *The process of cell signaling resembles the functioning of NNs. Cells receive signals from the environment and other cells, similar to how NNs receive data in the input layer. Subsequently, cells activate cascades of pathways to transduce these signals, comparable to the data processing that occurs in the hidden layers of NNs. Finally, the biological response of the cell is activated, analogous to the activation of the output layer of the NN.*

Consequently, through the training of this NN, we can develop a model that learns and predicts cell behavior by leveraging the data they receive. In this approach, the NN weights symbolize the intricate internal machinery, including the cell's DNA, proteins, and pathways. These components function akin to a translation mechanism, where given specific information, they orchestrate biological responses within the cell. The diversity in NN architectures and their corresponding weights offers the capability to simulate various cell types or even model the behavior of tumor cells. By manipulating these architectures and weights, we can

emulate distinct cellular behaviors, reflecting the incredible potential of this approach to replicate and predict cellular responses across different scenarios and cell types.

Here, Reinforcement Learning (RL) or Deep RL, when used with NNs, may play a crucial role in learning complex behaviors [184]. In RL, agents learn by interacting with the environment to maximize a reward signal. This reward in biology is not external to the environment but an internal motivation for the organism. For cells, it may encompass homeostasis, correct physiological functioning, survival, or, for tumor cells, maximize proliferation or invasion. By defining this reward, agents make actions to achieve the highest reward, leading to collective behaviors and emergent phenomena. Additionally, this reward can alternatively measure the similarity between the simulation and observed data, enabling the inverse solution of actions through trial and error to replicate observed data. The potential of this methodology lies in its ability to eliminate unnecessary parameters and heuristic or rule-based conditions of discrete models that predefine cell behavior. Instead, agents learn from experience and adjust their individual behavior accordingly. In this way, we advance in creating parsimonious ABM models that consider only physical properties and eliminate rule-based conditions, thereby enhancing interpretability, which is one of the main criticisms of ABM [111]. Thus, this fusion holds the potential to propel us toward the development of an intelligent cell model that processes information from its microenvironment to activate its biological response.

Bray [185] first proposes the use of NNs to simulate cell signaling pathways and discuss the biological adequacy of the model. Then, Vohradsky [186]) focuses on testing the validity of using NNs to analyze genetic networks and utilize it to simulate experiments involving bacterial infection by viruses with an ODE system. Within this ODE system, a NN is employed to map the evolution of genes between time steps with predefined parameters to consider positive, negative, or non-interactions between. Thus, the author shows the model not only accurately replicates observed behavior but also predicts a scenario that cannot be achieved experimentally, providing a coherent explanation consistent with observed principles. Finally, Gerlee and Anderson [187] introduces the use of NN to simulate cell behavior in a ABM (CA model) for studying early tumor development. In this approach, the NN processes microenvironmental conditions defined by the number of neighbors, oxygen, glucose, and hydrogen ion concentration to coordinate proliferation, quiescence, apoptosis, metabolism, and movement processes. Each cell utilizes a separate NN, and when a cell divides, it copies the NN to the daughter cells with a mutation of its weights. However, the NN weights are also chosen to capture a user-predefined behavior of cells. Thus, these works employ a NN solely as a function approximator with predefined weight values and do not utilize any ML techniques, therefore, failing to unravel cell behavior.

Conclusions

Although ML is employed in biology for image-based diagnosis and prognosis prediction [188], the combination of mechanistic modeling with ML is still in its early stages. Moreover, the combination of ABM with ML is not adopted in biology despite the growing interest in other fields [148, 189]. Therefore, ML in computational biology supposes a promising path to propel the field forward with the potential to mark the next leap in advancement.

1.6 Objectives and thesis outline

The main objective of this thesis is to develop computational methods to understand and reproduce cell behavior from the microscale to the macroscale. Through these computational methods, the aim is to gain insights into the factors that govern cell behavior and comprehend the mechanisms that drive cells from healthy to tumoral states. By unraveling these factors, we aim to predict tumor initiation and prognosis, to then analyze strategies to intervene effectively in future research. For this purpose, we explore cell behavior across different scales, ranging from the cellular level to the tissue level. Throughout these scales, we consider various cell biological processes and mechanical interactions to investigate phenomena such as cell migration, morphogenesis, and tumor growth. In this way, we aim to construct a multiperspective view of the factors governing different phenomena across scales.

We study this biological complexity developing both discrete and continuum computational models. These models are designed to accurately replicate data obtained from experimental biology, spanning from *in vitro* cell experiments to real patient data. Thus, the developed computational models are integrated into frameworks that enable the automation of simulating experimental data, thereby adapting effectively to data variability. To achieve this, we employ various approaches, including traditional calibration methods and ML techniques. Additionally, these frameworks are designed to be modular, allowing for easy adaptation to explore other types of data acquired from experimental methods or to integrate them with other mechanistic models. This versatility facilitates the adaptation of the approaches to the needs of other researchers.

To comprehensively scrutinize each of the presented objectives, this book is structured in the following chapters (Figure 1.5):

- ***Chapter 2 - Individual cell migration***. Here, we investigate individual 3D cell migration in microfluidic devices using an ABM combined with a sampling method. To accomplish this, we propose a framework to predict 3D migration patterns based solely on 2D imaging data from *in vitro* cell migration assays, which can be obtained even through basic microscopes. We demonstrate that 2D cell migration data may not accurately represent

real 3D trajectories and predict significant differences in 3D movement patterns among various immune cell types. This framework offers researchers a tool to reduce the need for sophisticated and expensive microscopes, as well as the computational burden associated with generating and analyzing 3D experimental data.

- ***Chapter 3 - Lumen morphogenesis***. In this study, we investigate the role of extracellular matrix ECM density in the process of lumen morphogenesis using a 3D ABM. In our simulation, we consider the luminal hydrostatic pressure resulting from cell fluid secretion as the driving force behind this morphogenetic process. We propose a discrete model to represent the fluid and simulate cell-fluid interactions. Our model predicts normal lumen formation in low-density matrices, while it indicates disruption of healthy morphogenetic formation when the matrix density increases, consistent with experimental evidence. Thus, it gives a mechanistic explanation for alterations in cell behavior and underscores the role of the ECM as a mechanoregulator in transitioning from healthy to tumorous formations

- ***Chapter 4 - Organoid morphogenesis***. This work presents a framework to reproduce *in vitro* experiments of cystic and solid organoids with an ABM. It introduces a novel methodology that combines ML techniques with ABM to unravel cell behavior in experimental assays. For this purpose, it employs a NN to guide cell decision-making of cell biological functions during development to reproduce the final organoid structure. This approach enables the identification of underlying principles governing cell activation and self-organization into various patterns under specific microenvironmental conditions. Therefore, the framework provides an automated tool for researchers to simulate their *in vitro* organoid experiments, which allows for the customization of both the physics-based and machine-learning models to suit their experimental needs.

- ***Chapter 5 - Prostate cancer***. The objective of this work is to predict prostate tumor growth in real patients using a continuum PDE model based on a biomarker obtained from blood tests. This framework generates a digital twin of the patient's prostate from the MRI and incorporates it into a physics-informed ML model that predicts tumor growth. The ML model regulates tumor dynamics in the physics-based model spatially and temporally, considering the 3D geometrical interactions between physiological data. This framework offers a novel approach to faster and more cost-effective monitoring of prostate cancer prognosis, reducing reliance on MRI for a comprehensive evaluation of tumor prognosis and the uncertainty associated with tumor growth predictions from blood tests.

Finally, ***Chapter 6*** provides a conclusive summary of the results achieved

throughout this dissertation and presents future research lines that arise from this thesis.

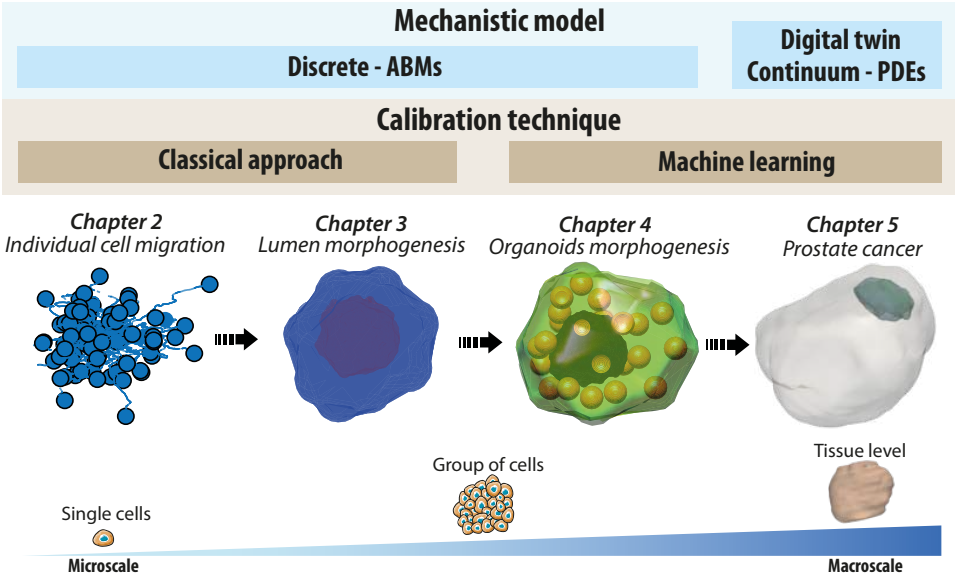

Figure 1.5: ***Visual summary of the dissertation's research lines.*** *The common thread of this dissertation is based on the methods developed and the resolution of the spatial scale. Firstly, it focuses on studying individual cell migration and lumen morphogenesis in the microscale through ABMs. Then, it integrates an ML algorithm into the ABM to reproduce the morphogenesis of organoids. Finally, it employs a ML method combined with a continuum PDE model of a DT to create a macroscopic, patient-specific model for prostate cancer.*

Chapter 2

Individual cell migration

The excitement of learning separates youth from old age. As long as you're learning you're not old.

– ***Rosalyn Sussman Yalow***

Contents

This chapter is based on:

Camacho-Gomez, D., Movilla, N., Borau, C., Martin, A., Oñate-Salafranca, C., Pardo, J., Gomez-Benito, M. J., & Garcia-Aznar, J. M. (2024). *An agent-based method to estimate 3D cell migration trajectories from 2D measurements: Quantifying and comparing T vs CAR-T 3D cell migration.* Under review.

Abstract

Immune cell migration is one of the key features that enable immune cells to find invading pathogens, control tissue damage, and eliminate primary developing tumors. Chimeric Antigen Receptor (CAR) T-cell therapy is a novel strategy in the battle against various cancers. It has been successful in treating hematological tumors, yet it still faces many challenges in the case of solid tumors. In this work, we evaluate the 3D migration capacity of T and CAR-T cells within dense collagen-based hydrogels. Quantifying 3D cell migration requires microscopy techniques that may not be readily accessible. In this context, we introduce a straightforward mathematical model designed to infer 3D trajectories of cells from experimental microscopy-based measurements of two-dimensional (2D) cell trajectories. We apply the aforementioned methodology to assess the migration capabilities of T and CAR-T cells within collagen-based matrices of varying concentrations. Our findings indicate that CAR-T cells exhibit distinct migration behaviors compared to native T cells in the absence of exogenous chemical stimuli. However, upon the introduction of a CXCL12 chemical gradient, CAR-T cells demonstrate migration patterns that closely resemble those of T cells. Thus, this approach offers an automation method that could be widely adopted by researchers.

2.1 Introduction

Cell migration is a key biological process that plays a crucial role in many physiological processes [32] such as wound healing [35], tissue repair [36], morphogenesis [190] and the functionality of the immune system [34]. In the context of immune response, immune cells, such as T cells and macrophages, navigate through tissues to detect and eliminate pathogens or abnormal cancer cells. However, despite their innate high migration capacity, in some cases, the natural T cells of the immune system struggle to recognize cancer cells because these can evade detection or suppress the immune response [191]. Cancer cells often employ various mechanisms to evade the immune system's surveillance, inhibiting T-cell function within the tumor microenvironment. To overcome this challenge, CAR-T cell therapy emerges as an innovative treatment strategy, leveraging engineered immune cells tailored to target specific antigens found on cancer cells [192]. This method involves the modification of a patient's T cells to express CARs, thereby reprogramming these cells for enhanced recognition and targeted destruction of cancer cells with exceptional specificity. Notably, CAR-T cell therapy has demonstrated promising results in blood cancers, including lymphomas, certain forms of leukemia, and, more recently, multiple myeloma [193–195].

Nevertheless, the transformation of T cells into engineered CAR-T cells may influence their migratory capabilities. These alterations may potentially compro-

mise their natural ability to navigate the body and reach tumor sites, thus challenging their capacity to infiltrate solid tumors effectively. This issue represents a significant obstacle to the success of CAR-T cell therapy in treating solid tumors, as opposed to its demonstrated efficacy in blood cancers. Therefore, it is crucial to comprehensively understand how these modifications impact the migratory behavior and infiltration potential of CAR-T cells to optimize their therapeutic effectiveness in solid tumor contexts.

The study of cell migration has traditionally been performed using *in vitro* [196] (e.g., the Boyden chamber migration assay [197] or scratch assay [198]) and *in vivo* (e.g., intravital microscopy [199, 200] and *in vivo* imaging [201, 202]) methods [203, 204]. However, the use of microfluidic devices emerged as a powerful tool to study cell migration, providing a high degree of control over the microenvironment, and allowing for the study of the response of cells to specific chemical and mechanical cues [49, 205, 206]. Microfluidic devices facilitate the analysis of the impact of different ECM properties, such as matrix density and stiffness, fiber alignment, and pore size [207, 208]. Another advantage of these devices is the ability to recreate 3D environments, which can provide more physiologically relevant information than 2D migration assays. Furthermore, by integrating imaging systems, microfluidic devices allow for real-time monitoring of cell migration, providing information on the kinematics of cell movement and allowing for a better characterization of their motile behavior. Many different cell types have been analyzed and studied in these microfluidic-based chips: tumor cells [209, 210], cancer-associated fibroblasts [211], dermal human fibroblasts [212], osteoblasts [213], leukocyte [214] among others. However, the use of microfluidic devices to study 3D immune cell migration poses a significant challenge due to their high migration capacity. While these devices can recreate 3D environments, quantification of 3D cell trajectories requires time-lapse studies integrating 2D imaging and Z-stack acquisition from the assays. Given the highly dynamic nature of 3D immune cell migration [215], it is essential to quantify a substantial number of events within a brief timeframe, which poses a risk of phototoxicity when using standard confocal microscopes. Light sheet microscopy offers a good alternative by providing low phototoxicity, high temporal resolution, and live sample imaging [215]. The primary challenge resides in managing the extensive data produced, which is complex, computationally heavy, and requires sophisticated tools for automated analysis. Apart from the technical hurdles, these microscopes represent a significant investment and are not commonly found in standard laboratory settings.

In this study, we propose a novel methodology based on ABMs to estimate 3D cell trajectories within microfluidic devices, relying only on 2D imaging data that can be acquired even from the most basic bright-field microscope. ABMs offer discrete computational representations for simulating cellular behavior at a cell level, and they have been extensively employed in different areas of cell mi-

gration. For instance, Saucedo-Mora et al. [216] simulated individual and collective cell migration in glioblastoma considering the influence of the oxygen field, Gonçalves and Garcia-Aznar [217]studied the role of collagen density in regulating cell migration and spheroid formation, Bretti and De Gaetano [218] investigated the influence of tumor cell-produced chemical signals on immune cell dynamics within microfluidic chips, and Peng et al. [219] explored the dynamics of cell migration within flexible channels and how mechanical interactions with the microenvironment and neighboring cells influence this process. While these approaches have focused on simulating cell migration within microfluidic devices, these works often struggle to extrapolate 3D behavior and lack a comprehensive understanding of cell migration in 3D environments.

Our lattice-free center-based ABM aims to predict 3D cellular behavior from 2D imaging data. To achieve this, we consider that cell locomotion direction is not entirely arbitrary but exhibits a directional bias influenced by the cell's previous orientation. To determine this directional memory, we propose an inverse sampling method to track its evolution from the previous direction. By deriving cumulative distribution functions through this sampling method, we gain insights into how past directions influence future ones. Consequently, this approach facilitates the simulation of a wide spectrum of migration behaviors, spanning from entirely persistent movement in a single direction when the cumulative distribution function approximates a Heaviside function, to non-persistent random movement devoid of directional persistence when the function takes on a linear shape. Finally, we apply the proposed methodology to simulate migration assays involving CAR-T cells and natural T cells (serving as the control group) within distinct collagen concentration matrices within microfluidic-based devices. The resulting 3D migration patterns are then compared and validated with experimental 3D measurements obtained through a lattice light-sheet microscope. Our results underscore the variances between 2D projections and genuine 3D motion, revealing that CAR-T cells display diminished motility in 3D relative to natural T cells.

2.2 Material and methods

2.2.1 Computational framework for predicting 3D migration patterns

We present a computational framework for predicting 3D migration of cells in microfluidic devices using 2D *in vitro* data (Figure 2.1). In this experimental setup, cells are seeded individually into microfluidic devices, and their movements are live-tracked within a single plane by capturing images at regular time intervals of Δt. These tracked cells are adequately spaced apart to prevent interactions between them, enabling the study of their individual migration dynamics. Then, we

acquire individual cell trajectories and instantaneous velocities for each cell. Subsequently, we calculate the mean velocity of each cell and determine the overall mean of these individual mean velocities across the entire population $\bar{v}_c$.

The computational model consists of a 3D center-based lattice-free ABM. Thus, the balance of forces acting on a single-moving cell is:

$$m_c \frac{d\boldsymbol{v}_c(t)}{dt} = \boldsymbol{F}_{loc}(t) + \boldsymbol{F}_{drag}(t) \approx 0. \tag{2.1}$$

Here, m_c is the mass of the cell, $\boldsymbol{v}_c(t)$ is its velocity, $\boldsymbol{F}_{loc}(t)$ represents the cell's locomotive force for migration, and $\boldsymbol{F}_{drag}(t)$ represents the friction force of the cell with the matrix. In this expression, the inertia term is typically neglected due to the small mass and velocity of the cell. The friction force is calculated using Stokes' law, which is appropriate for modeling the movement of spherical particles in a viscous fluid under laminar flow conditions:

$$\boldsymbol{F}_{drag}(t) = 6\pi R_c \eta \boldsymbol{v}_c(t), \tag{2.2}$$

with $R_c = 10\ \mu\text{m}$ the radius of the cell and η the dynamic viscosity of the collagen hydrogel obtained from Valero et al. Considering that the magnitude of the locomotive force is constant but variable in direction:

$$\boldsymbol{F}_{loc}(t) = F_{loc}\boldsymbol{e}(t) = 6\pi R_c \eta \bar{v}_c \boldsymbol{e}(t), \tag{2.3}$$

where $\bar{v}_c$ is the overall mean velocity of the experiment and $\boldsymbol{e}(t) = (e_\theta(t), e_\varphi(t))$ is the unit orientation vector in spherical coordinates $(r, \theta(t), \varphi(t))$. We simulate the temporal changes in the migration direction $(\theta(t), \varphi(t))$ through a rotational transformation:

$$\begin{aligned} \theta(t + \Delta t) &= \theta(t) + \theta_{rot}, \\ \varphi(t + \Delta t) &= \varphi(t) + \varphi_{rot}. \end{aligned} \tag{2.4}$$

Here, θ_{rot} and φ_{rot} are the rotations of the azimuthal and polar angles respectively. To determine these rotations, we consider that the migration direction exhibits a directional bias influenced by its prior orientation and propose an inverse transform sampling method. This method involves generating pseudo-random numbers from a uniform distribution and applying the inverse CDF associated with the desired probability distribution. Specifically, we utilize the CDF of the Cauchy function:

$$C(a; a_0, \gamma) = \frac{1}{\pi} \arctan\left(\frac{a - a_0}{\gamma}\right) + \frac{1}{2}, \tag{2.5}$$

with $a_0 = 0$ to be centered at $x = 0$, and γ a parameter that controls the function's shape. To account for potential variations in rotations for $\theta(t)$ and $\varphi(t)$, we

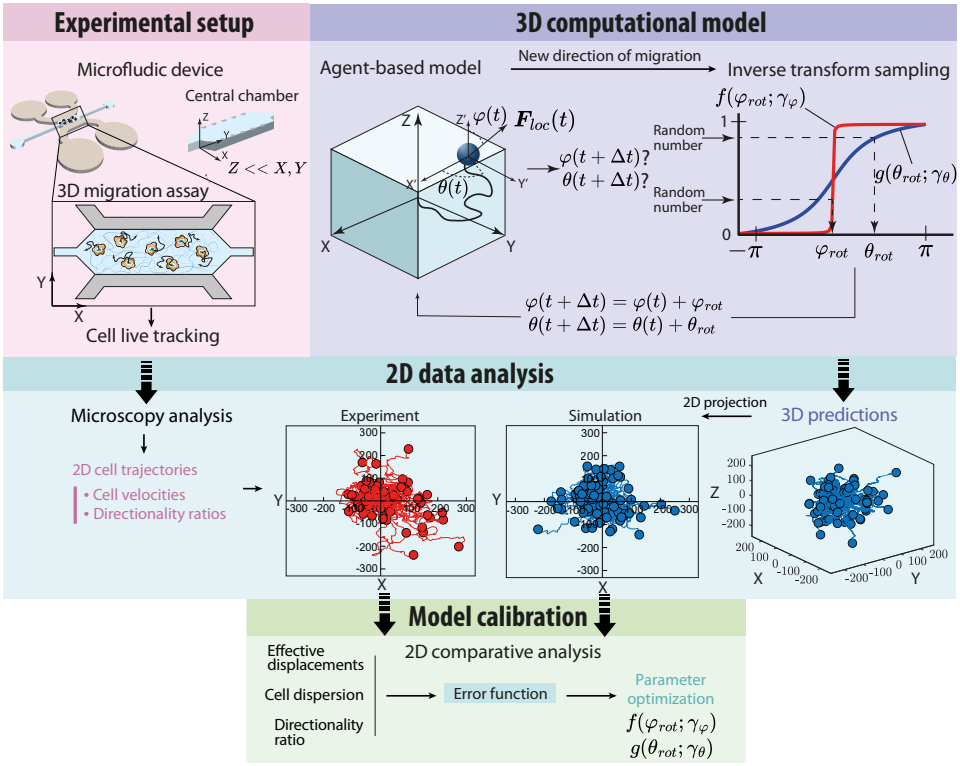

Figure 2.1: ***Overview of the 3D migration framework****. In the experimental setup, N individual cells are initially seeded onto a microfluidic device and tracked over t_{sim} duration, capturing 2D images at intervals of Δt. Concurrently, N 3D simulations of a single moving cell are performed using the ABM. The direction of the locomotive force is determined by angles $\theta(t)$ and $\varphi(t)$. To simulate the temporal direction of the locomotive force $\boldsymbol{F}_{loc}(t)$, we implement an inverse transform sampling method. This technique calculates rotational angles θ_{rot} and φ_{rot} using cumulative distribution functions $f(\varphi_{rot};\gamma_\varphi)$ and $g(\theta_{rot};\gamma_\theta)$ respectively. After completing the simulations, we evaluate the accuracy of the simulations by projecting the 3D predictions into 2D and comparing them with the experiments. Thus, we calculate the error between the simulation and experiment of the mean effective displacement, cell dispersion, and the directionality ratio. Subsequently, we update the functions $f(\varphi_{rot};\gamma_\varphi)$ and $g(\theta_{rot};\gamma_\theta)$ until the error between the simulation and experiment is minimized.*

utilize two analogous functions, denoted as $f(\varphi_{rot};\gamma_\varphi)$ and $g(\theta_{rot};\gamma_\theta)$, derived from the same form as the CDF in equation (2.5) within the inverse transform sampling method. Thus, we use this technique to generate values for θ_{rot} and φ_{rot} within the interval $(-\pi,\pi)$, allowing for the calculation of the new directional angles $\theta(t+\Delta t)$ and $\varphi(t+\Delta t)$ through equation (2.4).

After simulating N single migrating cells in 3D, we project the trajectories in 2D for a direct comparison with the experimental data. For model calibration, we employ prediction metrics involving the cell trajectories ($\boldsymbol{x}_c(t)$) and directionality ratio ($d_r(t)$) (see Supplementary 1 for directionality ratio details). In particular, we determine the mean effective displacement of cells along the X ($s_x = (\bar{x}_c(t = t_{sim}))$) and Y ($s_y = (\bar{y}_c(t = t_{sim}))$) axes and integrate it into the metric $s_{xy} = \sqrt{s_x^2 + s_y^2}$. Similarly, we calculate the standard deviation of cells' final positions along the X (σ_x) and Y (σ_y) directions to estimate cell invasion and incorporate this information into a comprehensive metric, $\sigma_{xy} = \sqrt{\sigma_x^2 + \sigma_y^2}$, to capture cell dispersion over the XY plane. We also calculate the area under the curve of the mean XY directionality ratio ($\bar{d}_r(t)$) of cells $\left(D_r = \int_{t=0}^{t=t_{sim}} \bar{d}_r(t)\, dt\right)$ to evaluate the cell trajectories' straightness. Then, we compute the error between the simulation and experimental metrics through Gaussian-like functions of the form:

$$h(a,b) = \exp\left(-\frac{(a-b)^2}{2\,(b^2)}\right). \tag{2.6}$$

Thus, the error is calculated as:

$$\mathcal{E} = 1 - \left(\frac{1}{3}h(s_{xy}^{sim}, s_{xy}^{exp}) + \frac{1}{3}h(\sigma_{xy}^{sim}, \sigma_{xy}^{exp}) + \frac{1}{3}h(D_r^{sim}, D_r^{exp})\right). \tag{2.7}$$

Finally, we minimize the error ($\mathcal{E}$) using a direct search optimization method, iteratively updating γ_φ and γ_θ while minimizing the error ($\mathcal{E}$) to obtain the functional shapes of $f(\varphi_{rot};\gamma_\varphi)$ and $g(\theta_{rot};\gamma_\theta)$ that accurately reproduce cell migration within the microfluidic devices.

2.2.2 Fabrication of microfluidic devices

Microfluidic chips were fabricated in PDMS (polydimethylsiloxane-Dow Corning GMbH Sylgard 184) at a 10:1 ratio of base to curing agent, following the methodology described by Shin et al. [205]. PDMS microdevices were plasma-bonded to 35 mm glass-bottom Petri dishes (IBIDI) and treated with poly-D-lysine (PDL) solution (Sigma- Aldrich) at 1 mg/mL to enhance the surface-matrix attachment. The geometry of the microfluidic devices consisted of a central chamber in which we introduced the hydrogel with the cells and two adjacent media

channels that allowed the introduction of cell culture medium and the generation of chemical gradients through the addition of a growth factor CXCL12 in one of the channels [221]. The study previously published by Moreno-Arotzena et al. [206] presents further details about the geometry of these microfluidic devices.

2.2.3 Cell culture and transduction of primary human T cells

PBMCs were obtained by Ficoll-Paque gradient centrifugation from blood obtained from Healthy Donors (CEICA, C.I. PI 11/006). They were activated with anti-CD3 (OKT3 clone) and anti-CD28 antibodies (BD Pharmingen) in complete medium; 44% RPMI1640 (PAN Biotech) + 44% Click´s médium (Sigma) + 10% FBS (Sigma) + 1% Glutamax (Gibco) + 1% Pen/Strep, (Sigma). The next day, the medium was supplemented with IL-7 and IL-15 (Miltenyi Biotec) at a final concentration of $10\,\mathrm{ng/mL}$. The transduction was performed on the third day with lentiviral vector supernatants as described by Davies et al. [222] and Vectofusin-1 (Miltenyi Biotec) according to the manufacturer´s recommendations. The activated PBMCs followed the same steps but without lentiviral vector supernatants. After centrifugation, the medium was replaced with fresh medium with IL-4 at $30\,\mathrm{ng/mL}$ (Miltenyi Biotec). The medium was replaced every two or three days with fresh medium with IL-4.

2.2.4 Hydrogel preparation and 3D cell culture

The hydrogels were prepared using collagen gel solution type I (BD Bioscience) to a final concentration of 4 and 6 $\mathrm{mg/mL}$, following the methodology proposed by Shin et al. [205]. The dilution was brought to pH 7.4 with 0.5M NaOH. T cells and CAR-T4 were suspended in a culture medium and mixed with the collagen hydrogel to a final dilution of 1.5 x 10^5 cells/ml to follow individual migration in 3D. For light-sheet microscopy (LLSM) 6 $\mathrm{mg/mL}$ collagen gels were used. The final dilution of cells was 1 x 10^6 cells/mL which were stained with 1 mM Vibrant DiO (Molecular Probes) at $37\,^{\circ}\mathrm{C}$ for 30 min. Then, the cells were spun down, rinsed, and resuspended in a fresh medium. The dilution of cells was then pipetted into the central gel chamber and the hydrogel was confined by surface tension. Once in place, the collagen gel solution was polymerized in a humidity chamber at $37\,^{\circ}\mathrm{C}$ and 5% CO_2 for 20 minutes. After that, the matrix was hydrated with RPMI 1640 (Lonza) supplemented with 10% FBS, 1% penicillin/streptomycin (Sigma-Aldrich), 1% GlutaMAX (Thermo Fischer), and $30\,\mathrm{ng/mL}$ IL4 (Milteny Biotec).

2.2.5 Image acquisition and quantification of T cells and CAR-T4 migration

Cell migration was recorded using time-lapse live microscopy (Nikon D-Eclipse Ti Microscope and Zeiss Axio Observer 7) with a 10X objective for migration, acquiring phase contrast images every $30\,\mathrm{s}$ for $1\,\mathrm{h}$. The focal plane was chosen to be in the middle along the z-axis of the device ensuring that the tracked cells were embedded within the 3D network. Cell trajectories were obtained using IMAGEJ and further analyzed using MATLAB scripts.

3D live T cell and CAR-T4 migration with light-sheet microscopy (LLSM) on a ZEISS Lattice Light Sheet 7 microscope (ZEISS, Oberkochen, Germany) was followed during $1.5\,\mathrm{h}$ exciting the cells at $488\,\mathrm{nm}$. For all experiments, the datasets acquired were deskewed using the ZEISS Zen (blue edition 3.7) software by a linear interpolation and a Cover-glass transformation. Deconvolution was performed using a constrained iterative algorithm. The resulting images had a voxel size of 0.14 x 0.14 x 0.14 $\mu\mathrm{m}^3$ and image stacks had a mean temporal resolution of $169.8 \pm 26.34\,\mathrm{s}$ per frame. For further processing, sample image stacks were resized with the resampling method of ZEISS's software to 1/8 of their original size, reaching a voxel size of 1.16 x 1.16 x 1.16 $\mu\mathrm{m}^3$, reducing significantly data size for easier management. During the experiment, the incubation conditions were controlled and held at $37\,^{\circ}\mathrm{C}$, 5% CO_2, and 95% of humidity. The experiments with $4\,\mathrm{mg/mL}$ collagen matrices were performed with two independent experiments, with four technical replicas for T cells and for CAR-T4, and the experiments with $6\,\mathrm{mg/mL}$ with two independent experiments and four technical replicas with T cells and with four independent experiments with six technical replicas with CAR-T4. For the assays with CXCL12 in $6\,\mathrm{mg/mL}$ collagen, two experiments were performed with three technical replicas. 3D experiment with light-sheet microscopy was performed once with two technical replicas with T cells and three with CAR-T4.

2.3 Results

2.3.1 CAR-T cell migration is more sensitive to collagen concentration than T cells

To demonstrate the potential of the computational framework in accurately replicating cell movement patterns, we first conduct cell migration assays of T and CAR-T4 cells (hereinafter referred to as CAR-T) within microfluidic devices embedded in collagen-based hydrogels. These cells are individually seeded in two distinct concentrations of type I collagen (4 and $6\,\mathrm{mg/mL}$), replicating the extracellular environment. The microfluidic device consists of a central chamber with an embedded collagen matrix containing cells and two lateral channels serving

as reservoirs for growth media. In some cases, the chemoattractant CXCL12 is introduced into one of the channels to establish a chemotactic gradient within the microfluidic devices with $6\,\mathrm{mg/mL}$ of collagen. Live-cell imaging techniques in 2D are employed, capturing images at 30-second intervals for 1 hour to track cell movement. Subsequently, we characterize the migration patterns using the proposed calibration metrics (Table 2.1, experiments column).

Then, we apply the proposed methodology to gain insights into the 3D behavior of cells, replicating the 2D microfluidic device migration assays. The calibration metrics derived from the experimental data are matched with the simulations (Table 2.1, simulation column), and the shape of the functions $f(\varphi_{rot};\gamma_\varphi)$ and $g(\theta_{rot};\gamma_\theta)$, which determine the directionality of cell migration, is obtained (Figure 2.2**B**). Note how the optimized functions exhibit very subtle differences, yet these nuances have a significant impact on the outcomes.

Table 2.1: ***Experiments and simulation data***. *N is the number of cells, $\bar{v}_c$ is the overall mean velocity of the cells, s_{xy} is the module of the mean effective displacement of cells along the X and Y axes, the σ_{xy} indicates the module of the standard deviation of cells' final positions along the X and Y directions, D_r is the area under the curve of the mean XY directionality ratio of cells, and $\mathcal{E}$ is the error between the experiments and the simulation calculated through Gaussians-like functions from (2.7).*

			Experiments				Simulations			
Cell type	Medium	N	$\bar{v}_c$ (μm/min)	s_{xy} (μm)	σ_{xy} (μm)	D_r	s_{xy} (μm)	σ_{xy} (μm)	D_r	$\mathcal{E}$ (%)
T	4 mg/mL	75	6.80	81.12	110.64	22.32	84.00	108.41	23.46	0.33
	6 mg/mL	44	5.26	52.40	81.64	19.20	55.89	73.09	19.10	1.17
CAR-T	4 mg/mL	31	6.53	80.17	122.04	20.41	86.24	109.40	21.98	1.70
	6 mg/mL	49	3.94	41.57	70.15	18.11	45.76	59.66	20.35	3.60
	6 mg/mL + CXCL12	27	4.30	51.22	78.42	26.67	51.70	71.36	26.51	0.63

The 2D projections of our 3D simulations replicated the cell trajectories of T and CAR-T cells in various collagen matrices (Figure 2.2**C**). Our observations reveal that at a collagen concentration of $4\,\mathrm{mg/mL}$, T cells and CAR-T cells exhibit a similar mean displacement in the XY plane. However, T cells show more directional movement (8.54% greater) and less dispersion (9.3% lower) than CAR-T cells. At a collagen concentration of $6\,\mathrm{mg/mL}$, T cells reduce their mean displacement by 35.5%, while CAR-T cells exhibit a more substantial reduction of 48.2%. Similarly, dispersion decreases more in the case of CAR-T (42.6%) compared to T cells (26.2%), and directionality is also lower in CAR-T compared to T cells. Furthermore, although the velocities are similar at $4\,\mathrm{mg/mL}$, at $6\,\mathrm{mg/mL}$ the overall mean velocity of T cells decreases by 22.7%, while for CAR-T cells, the decrease is 39.6%. This highlights a markedly higher sensitivity of CAR-T cells to collagen density alterations, resulting in a 20.6% smaller mean displacement, 14.1% smaller dispersion, and 25.10% smaller velocity compared to T cells under similar $6\,\mathrm{mg/mL}$ conditions. Nonetheless, the introduction of the chemo-attractant CXCL12 to the $6\,\mathrm{mg/mL}$ matrix leads to a 23.25% increase in XY plane displacement, an 11.7% increase in dispersion, and 9.14% increase in velocity, narrowing

the gap between CAR-T and T cells' mean displacement and dispersion to 2.2% and 3.9%, respectively, and making migration more persistent (47.3% greater than the condition without CXCL12).

To further validate the accuracy of our replications of 2D cell movement, we analyze the effective displacements along the X and Y axes, revealing no statistically significant differences between simulation predictions and experimental displacements derived from *in vitro* data (Figure 2.2**D**). Additionally, we assess the directionality ratio for T and CAR-T cells in different matrices, confirming its ability to capture the temporal evolution of the experimental data (Figure 2.2**E**).

2.3.2 CAR-T cells tend to move in 2D, in contrast to the patterns observed in T cells

In this section, we present model predictions of 3D trajectories for T and CAR-T cells in different matrices (Figure 2.3**A**), providing an isometric view of the 2D projected trajectories in Figure 2.2**C**. Despite the apparent similarity in 2D invasion patterns, as shown before, significant disparities are observed in the 3D movement patterns between T and CAR-T cells. While T cells exhibit multi-directional migration, CAR-T cells show a limited movement along the Z-axis, suggesting predominantly 2D movement at both 4 and $6\,\mathrm{mg/mL}$. To quantify these results, we compare the effective displacement along the X and Y-axis with the effective displacements along the Z-axis (Figure 2.3**B**). At both 4 and $6\,\mathrm{mg/mL}$ matrices, T cells demonstrate a more balanced movement across the X, Y, and Z axes, with a ratio between the median displacement in the Z direction and the XY directions of 1.01 and 0.95, respectively. In contrast, CAR-T cells at the same concentration matrices exhibit predominantly favored movement in the XY plane, with a ratio between the median displacement in the Z direction and the X and Y directions of about 0.60 for both 4 and $6\,\mathrm{mg/mL}$. However, the introduction of CXCL12 to the $6\,\mathrm{mg/mL}$ matrix significantly boosts the Z-axis movement for CAR-T cells, resulting in the most remarkable 3D movement observed in this scenario, with a median displacement in the Z direction 2.16 times higher than in the X and Y. We find no statistically significant differences between the effective displacements on the XY plane and Z in T cells but identify significant differences for CAR-T cells, suggesting the tendency of CAR-T cells to move two-dimensionally in the 4 and $6\,\mathrm{mg/mL}$ matrices not enriched with CXCL12.

To confirm these predictions of the 3D movement patterns, we conduct new 3D *in vivo* migration experiments, tracking T and CAR-T cells in real-time for 1.5 hours within a $6\,\mathrm{mg/mL}$ hydrogel. This 3D experiment, utilizing light-sheet microscopy, provides a more detailed understanding of cell migration within a 3D environment, allowing for the study of their spatial behavior. Subsequently, we calculate the ratio between the movement in the Z direction and the planar

XY movement for each displacement (Figure 2.3**C**). Our observations reveal that T cells exhibited 3D movement, with a ratio between Z and XY movements of 0.93. In contrast, CAR-T cells predominantly move in the XY plane with a ratio between Z and XY movements of 0.57. We find no statistically significant differences between simulations and experiments but we do find a significant difference in movement patterns between T cells and CAR-T cells, confirming the model predictions of a more predominant 2D movement pattern of CAR-T cells compared to T cells.

2.4 Discussion

In this work, we developed a computational framework to estimate 3D migration patterns of cells within microfluidic environments using an ABM, relying exclusively on 2D *in vitro* measurements. To assess the predictive capabilities of our model, we simulated *in vitro* migration experiments of T and CAR-T cells in hydrogels with different concentrations of type I collagen in microfluidic devices. We showed that the 2D projections of our 3D predictions successfully reproduced the 2D migration patterns of both T and CAR-T cells in the different density matrices, with no statistically significant differences between the experiments and simulations (Figure 2.2). Moreover, we observed that although CAR-T and T cells behaved similarly at $4\,\mathrm{mg/mL}$, CAR-T cells showed higher susceptibility to the increment in collagen concentration compared to T cells, resulting in a more pronounced reduction in their invasiveness.

Furthermore, our computational model revealed notable differences in the 3D movement patterns between T and CAR-T cells. T cells displayed migratory behavior in three dimensions within the two collagen concentration matrices, aligning with the characteristic high motility observed in immune cells that enables their access to intricate regions within the body. In contrast, our simulations unveiled a distinctive pattern in CAR-T cells. They appeared to predominantly move within the XY plane and exhibited limited movement in the Z direction, indicative of primarily 2D migration. To quantify the differences between movements in each direction, we calculated the displacement ratio between the Z direction and the X and Y directions. T cells presented a ratio around 1 in both density matrices, indicating a balanced three-dimensional movement, while CAR-T cells presented a ratio around 0.6 in both matrices, emphasizing their predisposition to move more over the XY plane compared to the Z direction. We found no statistically significant difference between the Z and XY displacements in the case of T cells, but we did find notable differences in the Z movement compared to the XY movement in the case of CAR-T cells.

To validate these results of our new approach, we conducted 3D cell migration experiments at $6\,\mathrm{mg/mL}$ using light-sheet microscopy, allowing us to quantify cell migration patterns in 3D. Similarly, we calculated, for each displacement, the

ratio between the Z movement and the mean X and Y movements, showing that in 3D experiments, T cells exhibited three-dimensional movement (Z movement versus mean XY movement of 0.93), whereas CAR-T cells exhibited limited movement along the Z axis (Z movement versus mean XY movement of 0.57). We compared these experimental observations with our simulations and found no statistically significant differences between them. However, we did observe differences between T and CAR-T cells, confirming the model's predictions of 3D migration patterns derived from 2D data. Additionally, our investigation revealed that the presence of the chemical factor CXCL12 not only stimulated CAR-T cells to migrate within the XY plane but also induced a shift in behavior, promoting 3D movement similar to T cells. Therefore, enhancing the migrative capacity of CAR-T cells might be a promising strategy to improve the efficacy of therapies in solid tumors [223].

The planar migration pattern observed in CAR-T cells may be attributed to two primary factors. Firstly, the exposure of T-cells to chimeric antigen receptors during their conversion into CAR-T cells could influence their migration capabilities, potentially impeding their ability to navigate through intricate and confined spaces. Secondly, the geometrical characteristics of microfluidic devices must be considered. The vertical dimension (Z) significantly differs in scale from the other two dimensions, and the potential influence of fiber alignment occurring in the XY [224] plane during media loading may impact CAR-T cell migration. This influence could lead to a reduction in migration, potentially transitioning from 3D movement to a mode resembling 2D migration. The combination of these factors may account for the observed behavior, further limiting the cells' ability to migrate in the Z direction within this complex environment.

Nevertheless, we have to keep in mind that our novel predictive approach is based on several simplifications. Initially, we assumed cells exert constant magnitude locomotive force, with temporal changes attributed solely to changes in migration direction. Cells may generate locomotive forces varying in both magnitude and direction. However, to calculate this force magnitude, we used the temporal mean velocity of the entire cell population. This mean velocity represents the temporal velocity variations across the cell population. Therefore, considering a mean locomotive force effectively encompasses the temporal variations in force magnitude across the population of cells, offering a comprehensive measure of these temporal fluctuations and modeling it with a mean representative value. However, temporal variations in the locomotive force magnitude can be easily incorporated, for instance, making it variable within a certain range. Temporal variations in the magnitude of the locomotive force can be easily incorporated by allowing it to vary within a certain range, based on probability distributions derived from velocity measurements. Additionally, we utilized the displacements, dispersion, and directionality ratio as predictive metrics. While projecting trajectories in 2D can yield the same trajectories across multiple po-

sitions, the combination of these predictive metrics, in conjunction with cell velocity, constrains the feasible cell positions that result in XY projections. This synergy between metrics effectively constrains the problem space and enhances predictive accuracy. Nonetheless, further refinements could be incorporated during the calibration phase, as the model can be easily tailored to the user's needs. For example, additional criteria could be added to better capture the directionality ratio over time, taking into account not only the area under the curve but the value of the ratio at different time steps. Similarly, our approach possesses the flexibility to incorporate additional metrics based on the specific *in-vitro* data collected, demonstrating its adaptability for various types of migration assays.

2.5 Conclusions

Therefore, the methodology here proposed can be used by researchers to estimate 3D migration patterns from 2D experimental data. which can be easily obtained with automatic quantification algorithms [225, 226]. This approach helps reduce the need for sophisticated and expensive microscopies required in laboratories, as well as the computational burden involved in producing and analyzing 3D experimental data.

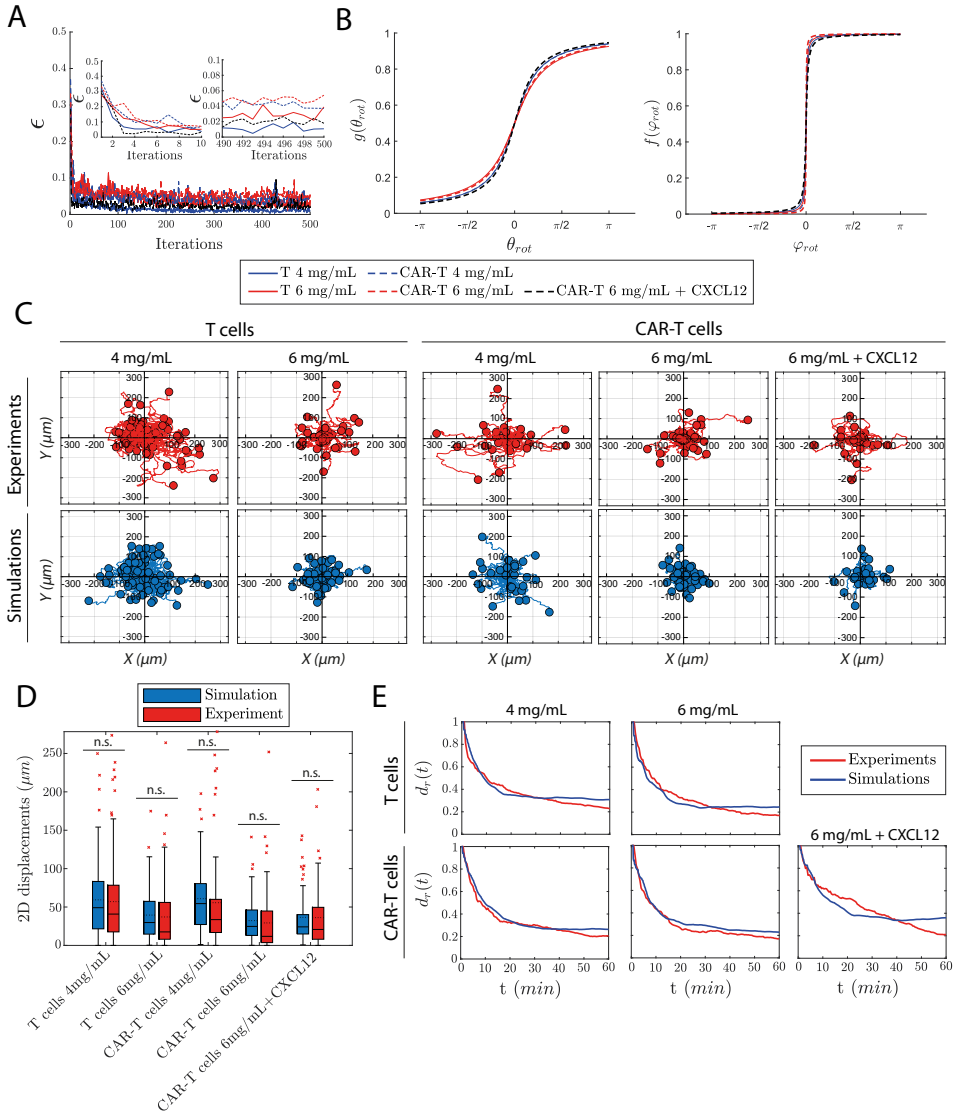

Figure 2.2: ***Comparative analysis of 2D cell migration patterns between experiments and simulations.*** ***A.*** *Calibration errors for each iteration of minimization, with two details at the first and last 10 iterations.* ***B.*** *Obtained directional migration functions $g(\theta_{rot};\gamma_\theta)$ and $f(\varphi_{rot};\gamma_\varphi)$ from the inverse transform sampling method that minimize calibration error.* ***C.*** *Experimental and simulated 2D cell trajectories of T and CAR-T cells in different collagen density matrices.* ***D.*** *Analysis of effective displacements along X and Y axes between experiments and simulations. ANOVA followed by post hoc Tukey–Kramer tests are performed to determine statistical significance. $***P < 0.001$; $**P < 0.01$; $*P < 0.05$. We find no statistically significant differences between experiment and simulation results in any of the conditions.* ***E.*** *Comparison of simulated and experimental directionality curves.*

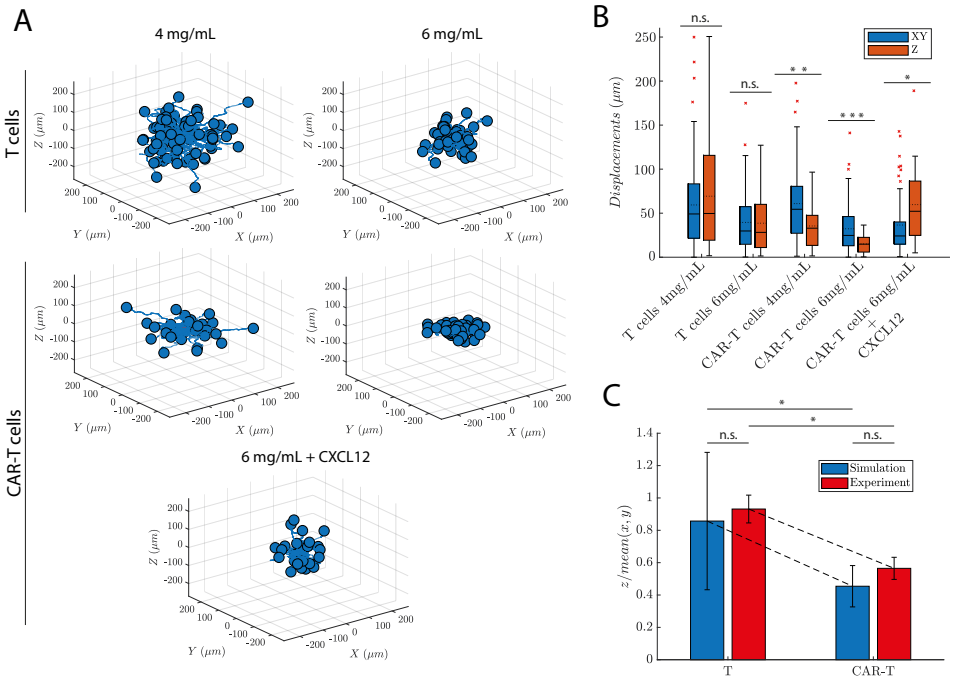

Figure 2.3: ***Analysis of 3D cell migration patterns.*** ***A.*** *Isometric view of the 3D predictions of cell trajectories for T and CAR-T cells corresponding to the 2D simulated trajectories in Figure 2.2* ***C.*** ***B.*** *Analysis of effective displacements along the X and Y axes versus the Z axis in the simulations for T and CAR-T cells in different matrices. ANOVA followed by post hoc Tukey–Kramer tests are performed to determine statistical significance. $***P < 0.001$; $**P < 0.01$; $*P < 0.05$. We find statistically significant differences between the CAR-T movement pattern in the X and Y directions and the Z direction in all conditions, in contrast with the T cells which present similar patterns.* ***C.*** *Comparison of the ratio of effective displacement in the Z direction and the mean effective displacement in the X and Y directions between simulations and the 3D in vivo experiment in the 6 $\mathrm{mg/mL}$ matrix. We find no statistically significant differences between simulations and experiments, but we do find statistically significant differences between T and CAR-T cells.*

Chapter 3

Lumen morphogenesis

You cannot teach a man anything; you can only help him discover it in himself.

– ***Galileo Galilei***

Contents

This chapter is based on:

Camacho-Gomez, D., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2022). *A 3D multi-agent-based model for lumen morphogenesis: the role of the biophysical properties of the extracellular matrix.* Engineering with Computers, 38(5), 4135-4149. https://doi.org/10.1007/s00366-022-01654-1.

Abstract

The correct function of many organs depends on proper lumen morphogenesis, which requires the orchestration of both biological and mechanical aspects. However, how these factors coordinate is not yet fully understood. Here, we focus on the development of a mechanistic model for computationally simulating lumen morphogenesis. In particular, we consider the hydrostatic pressure generated by the cells' fluid secretion as the driving force and the density of the ECM as regulators of the process. For this purpose, we develop a 3D ABM for lumen morphogenesis that includes cells' fluid secretion and the density of the ECM. Moreover, this computer-based model considers the variation in the biological behavior of cells in response to the mechanical forces that they sense. Then, we study the formation of the lumen under different mechanical scenarios and conclude that an increase in the matrix density reduces the lumen volume and hinders lumen morphogenesis. Finally, we show that the model successfully predicts normal lumen morphogenesis when the matrix density is physiological and aberrant multilumen formation when the matrix density is excessive.

3.1 Introduction

Lumen morphogenesis is a key process in the development of tissues and organs. Luminal structures are found in many parts of metazoan organisms, and they allow these organisms to perform specific functions, such as blood flow, digestion, and renal system activity. The formation involves the coordination of cells in a sophisticated manner. Three basic principles are required to create a lumen: spatially controlled cellular mitosis, cell-cell and cell-matrix interaction, and cell fluid pumping [227]. However, how these mechanisms coordinate is not fully understood.

Fluid-solid interaction is critical in vascular biology [228], in hemodynamics [229], in angiogenesis [136] and, in general, in physiology and pathophysiology [230]. Thus, luminal structures are in a state of hydrostatic pressure [231], and this pressure, which is developed from fluid secretion, might be the driving force in luminal structures during morphogenesis [232], as shown in ductal network formation in the pancreas [233]. Furthermore, the mechanical properties of the ECM play an important role in lumen formation. The mechanical interaction of cells with the ECM provides physical cues that are relevant to cellular self-organization [234]. In fact, an increment in tissue stiffness is related to the tumorous behavior of cells, as tumors are generally stiffer than normal tissues [235]. In the case of luminal structures, matrix stiffening compromises tissue organization, inhibits lumen formation, and enhances growth [68]. This perturbation in the tissue architecture as a result of matrix stiffening produces a dysfunctional lumen associated with disease and may be related to the aberrant structures found in

carcinomas [236, 237]. Therefore, understanding how these two factors influence lumen morphogenesis might provide insight into not only normal formation but also pathological formation.

Previous computer-based models have been developed to study different aspects of luminal structures. Specifically, on-lattice models are the most prevalent. In this type of model, cells occupy sites in a regular lattice, and mechanical interactions are usually represented by minimizing the energy of the system [96]. Thus, Cerruti et al. [238] analyzed dynamic regimes of epithelial growth, Engelberg et al. [100] developed a predictive model that matches their *in vitro* data, Boas and Merks [239] investigated two alternative mechanisms of lumen formation, Liedekerke et al. [240] studied cell mechanics in the embryonic bile duct, Rejniak and Anderson [241] simulated the development of an acinar structure with five different subpopulations of cells, and then the maintenance and stability of the epithelial monolayer and the hollow lumen [242]. Continuum approaches have been also adopted to simulate developmental processes. In the case of lumen morphogenesis, Dasgupta et al. [243] studied the balance between paracellular leaks and the build-up of osmotic pressure in the lumen, and Duclut et al. [244] built a continuum model to explore the role of the coupling of mechanical, electrical, and hydraulic phenomena in tissue lumen formation. Nonetheless, in this type of model, individual cells are neglected in favor of a larger scale of simulation, so it is difficult to take into account how heterogeneities in cell behaviors affect lumen initiation and formation from individual cells [245].

Although important achievements have been made, to the best of our knowledge, some crucial aspects have not yet been considered in lumen morphogenesis. Cell-cell interaction is essential and neither on-lattice models nor continuum models can directly represent the mechanical interaction between individual cells. Moreover, the forces produced by the luminal hydrostatic fluid pressure are essential in lumen morphogenesis, as they permit the lumen to initiate and expand. Finally, the mechanical properties of the ECM are usually neglected, even though it regulates the process, as it opposes the expansion produced by hydrostatic pressure. Therefore, we introduce a 3D multi-ABM for lumen morphogenesis that integrates the cell-cell and cell-matrix interactions, the hydrostatic pressure generation, and the role of the ECM. ABMs have been widely used to study cell and tumor growth [246–248], or tissue regeneration mimicking cell deformation [110]. Here, we aim to create a model that mimics lumen morphogenesis, taking into account the internal luminal hydrostatic pressure generated by cells' secretion and the effect of matrix density. To evaluate the predictive capacity of the model, we qualitatively replicate the experimental results achieved by [68], who found that an increase in matrix stiffness inhibits lumen formation and enhances cell colony size, resulting in an aberrant multiluminal architecture.

3.2 Materials and methods

We formulate a 3D multi-ABM for lumen morphogenesis, in which agents are spheres that do not change shape (Figure 3.1). The objective of this computer-based model is to mimic the morphogenesis of an organoid composed of cells enclosing a fluid-filled lumen by means of numerical simulations. To accomplish this task, we consider two types of agents: cells, which are the biological entities, and particles, which are secreted by cells and simulate the lumen fluid.

Therefore, we define a computational model for simulating the cell cycle that regulates cell proliferation and fluid secretion to form the lumen. To simulate how cells secrete fluid, we assume that cells generate particles inside the lumen, thereby increasing the lumen volume. Consequently, the lumen is in a state of hydrostatic pressure [231] due to this cell secretion [232, 249].

Mechanical equilibrium between cells anchored to the ECM and the luminal pressure ensures the maintenance of the luminal architecture. To model this mechanical equilibrium, we use agents that interact mechanically to generate and maintain the lumen. Thus, cells interact among themselves based on pairwise potential functions in an adhesive-repulsive manner. Moreover, particles also interact among themselves via pairwise potential functions and interact with cells in a repulsive manner. This interaction mimics the luminal hydrostatic pressure generated by cells' fluid secretion, and it is responsible for the movement of cells and generation of the luminal space. Finally, agents interact with the ECM by means of a friction coefficient that represents the dynamic viscosity of the matrix.

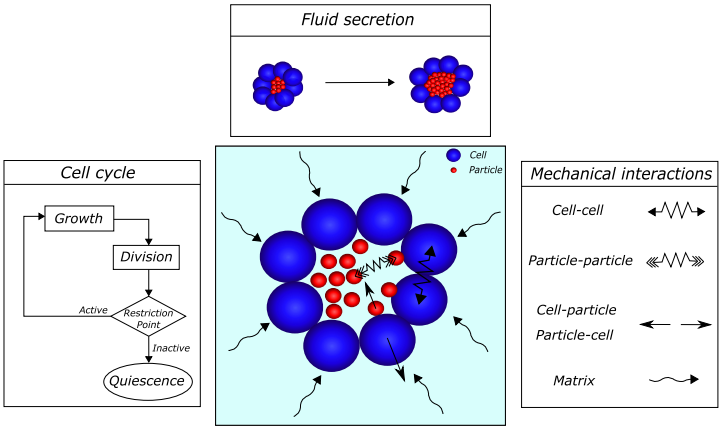

Figure 3.1: ***Sketch of the lumen model****. It consists of two main agents and their interactions. First, for the agent that simulates the cells, we consider a mathematical model for the cell cycle, which handles proliferation. Second, the fluid secretion agent makes cells secrete fluid into the lumen through particles, increasing the hydrostatic pressure. Finally, both agents interact mechanically, and the lumen expands.*

3.2.1 Modeling the cell cycle

The cell cycle is a complex process that occurs that involves the growth and proliferation of cells, organismal development, regulation of DNA damage repair, tissue hyperplasia in response to injury, and diseases such as cancer [250]. Overall, it can be described by a growth phase, in which the cell progressively increases its volume as a result of DNA replication, and by a mitosis phase, in which the cell divides into two daughter cells. To simulate the growth phase, we consider that each *i*-cell has a total volume $V_i(t)$, whose temporal evolution is given by:

$$\frac{dV_i(t)}{dt} = \alpha_i(P_{in}(t), F_{c_{net_i}}(t))V_i(t), \tag{3.1}$$

where $\alpha_i(P_{in}(t), F_{c_{net_i}}(t))$ is the growth rate of the *i*-cell, which we assume that depends on the luminal pressure ($P_{in}(t)$) and the net force exerted on the cell ($F_{c_{net_i}}(t)$). In this expression, cell growth is exponential and dependent on cell size according to experimental evidence that growth rates increase with cell size throughout the cell cycle [251–253].

Here, we assume the value of the growth rate is influenced by the luminal pressure $P_{in}(t)$ and by the net force exerted on the cell ($F_{c_{net_i}}(t)$). The variability of the growth rate during lumen morphogenesis has been experimentally observed in cell cultures in which MDCK [100] and primary Pancreatic Ductal Epithelial Cells (PDEC) [254] slowed their proliferation when they initiated the lumen. The onset of the lumen is determined by the initiation of the luminal pressure. Therefore, when there is no luminal pressure, the cell cycle time decreases so it can generate a closed volume to secrete fluid and form a lumen *de novo*. By contrast, when the luminal pressure increases, the cells do not need to divide as quickly as before, so the growth rate decreases. The growth rate $\alpha_i(P_{in}(t), F_{c_{net_i}}(t))$ is obtained from:

$$\alpha_i(P_{in}(t), F_{c_{net_i}}(t)) = \begin{cases} 1/T_{ini}(1 + a_i(F_{c_{net_i}}(t))), & P_{in}(t) = 0, \\ 1/T_{pol}(1 + a_i(F_{c_{net_i}}(t))), & P_{in}(t) > 0, \end{cases} \tag{3.2}$$

where $P_{in}(t)$ is the luminal pressure and T_{pol} and T_{ini} are constants related to the cell cycle time ($T_{pol} > T_{ini}$). Second, we consider that the value of $\alpha_i(P_{in}(t), F_{c_{net_i}}(t))$ is influenced by $a_i(F_{c_{net_i}}(t))$ depending on the net cell force supported by the cell ($F_{c_{net_i}}(t)$). In this regard, the ECM stiffness regulates the magnitude of the net cell force, and an increase in cytoskeletal tension, mediated by sustained matrix stiffness, promotes growth [68]. Moreover, the compliance of the matrix acts as a cell-cycle inhibitor and matrix stiffening increases cell proliferation [255–258] and cell cycle progression [259]. To model this mechanoregulation, we include a variation in the growth rate as a function of the net cell force as follows:

$$a_i(F_{c_{net_i}}(t)) = \begin{cases} 0, & F_{c_{net_i}}(t) < F_{bot}, \\ \left(F_{c_{net_i}}(t) - F_{bot}\right) \dfrac{a_{max}}{F_{top} - F_{bot}}, & F_{bot} \leq F_{c_{net_i}}(t) \leq F_{top}, \\ a_{max}, & F_{c_{net_i}}(t) > F_{top}, \end{cases} \tag{3.3}$$

where $F_{c_{net_i}}(t)$ is the net cell force, a_{max} is the maximum variation in the growth rate, and F_{bot} and F_{top} are approximately the mean values of net cell forces when matrix density is low and high respectively.

Last, to account for biological variability, we let the daughter cell growth rate $\alpha_i(P_{in}(t), F_{c_{net_i}}(t))$ vary randomly between $[-20, 20]\%$ based on a normal probability distribution around the progenitor cell's value.

When the volume of the cell reaches twice the value of its initial volume, DNA replication is concluded, and the cell divides. Spatially controlled division is a fundamental condition to maintain the lumen architecture and to enhance its growth by enlarging the lumen volume. In this regard, a complex molecularly controlled process regulates the spindle orientation, so mitosis occurs in the plane of the monolayer [260–265]. Here, we distinguish division between nonpolarized cells and polarized cells. Nonpolarized cells are those that have not yet formed a lumen, and polarized cells are those that belong to a lumen and face it. In the case of nonpolarized cells, the division direction is chosen randomly. Polarized cell division is performed using a random cleavage plane that contains the line that passes through the cell center and the lumen center of mass. The position of the two daughter cells ($\boldsymbol{x}_{daughters}$) are calculated similarly to other models [266, 267] from the center of the parent cell $\boldsymbol{x}_{parent}$ at:

$$\boldsymbol{x}_{daughters} = \boldsymbol{x}_{parent} \pm \left(R_c - \frac{1}{\sqrt[3]{2}} R_c\right) \boldsymbol{n}, \tag{3.4}$$

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

After cell division, the cell decides whether to remain active and continue in the cell cycle to divide again or to become inactive within the cell cycle, namely restriction point [268]. In the model, at this point, either or both of the daughter cells can enter a quiescent state. Thus, we introduce a variable called $state_i$ to register for each cell whether they are active or inactive. The probability that a i-cell enters a quiescent state is

$$P(Q)_i = b \cdot c_{steps_i}, \tag{3.5}$$

where b is a probability parameter and c_{steps_i} is the number of times the cell has divided. Each time a cell divides, the value of c_{steps_i} of its daughter cells increases by one. Thus, a random number in the interval [0,1] is generated for each daughter cell, and if it is lower than their probability $P(Q)_i$, the corresponding daughter cell enters into a quiescence state. When a cell becomes inactive, it implies that it does not grow ($\alpha_i = 0$) and, therefore, its growth rate does not follow Equation 3.2. Otherwise, the cell continues in the cycle and starts growing to double its volume again and then divides.

3.2.2 Fluid secretion

One of the key aspects of the model is how cells create the lumen. To generate a luminal domain *de novo*, neighboring cells must coordinate to secrete fluid into a common site, and that common site could be the midbody created during mitosis [227]. The midbody is a transient structure formed in the last phases of cell division to complete the separation between cells [269]. This landmark determines the apical-basal polarization of the cells and, therefore, the site where the cells will secrete to create the lumen. To contemplate the polarization of cells, we introduce a variable called $polarized_i$ that registers for each cell whether they are polarized or nonpolarized. Consider an initial active nonpolarized cell (Figure 3.2a). When this cell divides, it generates a midbody, and the two daughter cells polarize with respect to that point. Then, the cells need to form a closed volume, the Preapical Patch (PAP) [270], to be able to generate hydrostatic pressure. We consider that the PAP is formed when the number of cells is equal to a specific value (n_{t_c}). With subsequent cell division, the number of cells increases, and the cells form the PAP. Once the initiation site for the lumen is created, the cells secrete fluid into that point to open the lumen. Each polarized cell, after a period of time of fluid production Δt_{exo}, secretes fluid into the lumen. To model the luminal fluid, cells generate a certain number of particles. First, when the lumen does not yet have any fluid, the cell secretes into the initiation site where the lumen will be generated. Then, when another cell secretes, some random particles are duplicated inside the lumen to recreate the increment in the fluid volume. The position of the new particles is obtained from the position of the particle that duplicates in a similar way to cell division. As a result of the cells' secretion, the luminal hydrostatic pressure increases, and the lumen grows. The luminal hydrostatic pressure (P_{in}) is estimated by the mean of the particle's net force (F_{in}). We hypothesize that there is a maximum luminal hydrostatic pressure above which cells cannot pump any more fluid into the lumen. Thus, cells can only secrete when the force generated by hydrostatic pressure, estimated through the mean particle net force, is below a threshold F_{lim}. This makes lumen formation a dynamic process of phases in which polarized cells can secrete, thereby increasing the hydrostatic, and phases in which polarized cells are not able to secrete due to

the high pressure.

Cells remain polarized as long as they face either the midbody or the lumen. However, due to subsequent cell division or mechanical interactions, a cell can leave the lumen and no longer face it (green cell in Figure 3.2b). This cell is now nonpolarized, and depending on whether it is active or not within the cell cycle, it may create a secondary lumen. If it is active, the process is equivalent to the previous case: a new midbody is established when the cell divides and the cells polarize with respect to that point, create a preapical patch, and secrete into the area to generate the lumen. In this case, when the initial nonpolarized cell polarizes to create a new lumen, the number of times that the cell has divided c_{steps_i} is reset to prevent its daughter cells from entering quiescence and being unable to form the new lumen. On the other hand, if it is not active, the cell will not form a new lumen.

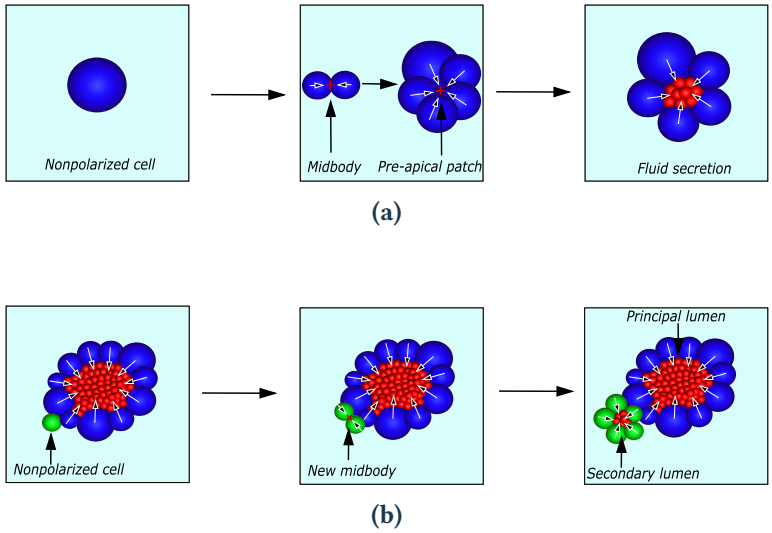

Figure 3.2: ***Sketch of the lumen morphogenesis model.*** *Case (a) shows how the lumen is formed beginning from an individual cell, and case (b) shows the formation of a secondary lumen. The arrows indicate the apical-basal polarization of cells.*

3.2.3 Mechanical interactions

Mechanical interactions between agents make them move and change their positions. We illustrate how the cell position $\boldsymbol{x}_c$ and particle position $\boldsymbol{x}_p$ are calculated. Let $\mathbf{N}_c$ be the set of cells $\mathbf{N}_c = \{1, ..., N_c\}$, and let $\mathbf{N}_p$ be the set of particles $\mathbf{N}_p = \{1, ..., N_p\}$. First, the velocity of each *i*-cell $\boldsymbol{v}_{c_i}$ and each *k*-particle

$\boldsymbol{v}_{p_k}$ are calculated from the balance of forces:

$$m_{c_i}\frac{d\boldsymbol{v}_{c_i}}{dt} = \sum_{j\in\mathbf{N}_c}\left(\boldsymbol{F}_{c_ic_j}\right) + \sum_{j\in\mathbf{N}_p}\left(\boldsymbol{F}_{c_ip_j}\right) + \boldsymbol{F}_{c_{i_{drag}}} \approx 0, \tag{3.6}$$

$$m_{p_k}\frac{d\boldsymbol{v}_{p_k}}{dt} = \sum_{j\in\mathbf{N}_p}\left(\boldsymbol{F}_{p_kp_j}\right) + \sum_{j\in\mathbf{N}_c}\left(\boldsymbol{F}_{p_kc_j}\right) + \boldsymbol{F}_{p_{k_{drag}}} \approx 0. \tag{3.7}$$

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

$$\boldsymbol{F}_{c_{i_{drag}}} = -6\pi\eta R_{c_i}\boldsymbol{v}_{c_i}, \tag{3.8}$$

$$\boldsymbol{F}_{p_{k_{drag}}} = -6\pi\eta R_p\boldsymbol{v}_{p_k}, \tag{3.9}$$

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

Cell-cell interaction forces are usually modeled as repulsive-attractive forces. The repulsion between cells arises from cell resistance to deformation when their membranes touch, and the attractive forces are the result of the junctions that cells form between themselves through specialized protein complexes [271]. In the case of particles, the repulsive-attractive forces represent the intermolecular forces in fluids. Accordingly, we modeled the interaction forces $\boldsymbol{F}_{\delta_i\gamma_j}$ (both subindexes δ and γ denote *c* or *p*, depending on whether the *i* and *j* agents are cells (*c*) or particles (*p*)) following [272], as follows:

$$\boldsymbol{F}_{\delta_i\gamma_j} = F_{\delta\gamma}\frac{\boldsymbol{r}_{ij}}{||\boldsymbol{r}_{ij}||}, \tag{3.10}$$

where:

$$\boldsymbol{r}_{ij} = \boldsymbol{x}_{\gamma_j} - \boldsymbol{x}_{\delta_i}, \tag{3.11}$$

and:

$$F_{\delta\gamma} = \begin{cases} F_{rep_{\delta\gamma}}\chi(-s)^{3/2}, & s < 0 \text{ (repulsion)}, \\ -F_{adh_{\delta\gamma}}\chi\{(s+s_0)e^{-\lambda(s+s_0)^2} - v_0e^{-\lambda x^2}\}, & s \geq 0 \text{ (adhesion)}. \end{cases} \tag{3.12}$$

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

$$\chi = \frac{R_{\delta_i}}{2}\left(\frac{1}{R_{\delta_i}} + \frac{1}{R_{\gamma_j}}\right), \qquad s = \frac{d - min_{dist}}{R_{\delta_i}}, \tag{3.13}$$

$$x_0 = \sqrt{\frac{1}{2\lambda}}, \qquad v_0 = x_0 e^{-\lambda x_0^2}. \tag{3.14}$$

$F_{rep_{\delta\gamma}}$ and $F_{adh_{\delta\gamma}}$ are the strengths of the adhesive and repulsive forces, respectively. $\boldsymbol{r}_{ij}$ is the distance between the centers of the agents, and R_{δ_i} and R_{γ_j} are the radii of the corresponding agents. x_0, v_0 and λ are matching constants, and χ is a geometric correction factor. The value of $min_{dist} = -0.1R_{\delta_i}$ is chosen such that the equilibrium state where the adhesive and repulsive forces are balanced is slightly less than zero, following [272], and $d = ||\boldsymbol{r}_{ij}|| - R_{\delta_i} - R_{\gamma_j}$ is the distance between the agents' surfaces.

Since cells do not present any attraction towards the lumen fluid, the interaction force between cells and particles is only repulsive ($F_{adh_{cp}} = F_{adh_{pc}} = 0$). Thus, the cell net forces $F_{c_{net_i}}(t)$ are computed from:

$$F_{c_{net_i}}(t) = |\sum_{j\in\mathbf{N}_c}\left(\boldsymbol{F}_{c_ic_j}\right) + \sum_{j\in\mathbf{N}_p}\left(\boldsymbol{F}_{c_ip_j}\right)|, \tag{3.15}$$

and the mean interactive cells' net force during the simulation as follows:

$$\bar{F}_c = mean\left(\frac{\sum_{j\in\mathbf{N}_c}\left(F_{c_{net_i}}(t)\right)}{n_c(t)}\right), \tag{3.16}$$

with $n_c(t)$ the number of cells.

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

$$\frac{d\boldsymbol{x}_{c_i}(t)}{dt} = \boldsymbol{v}_{c_i}(t) = \frac{1}{6\pi\eta R_{c_i}}\left(\sum_{j\in\mathbf{N}_c}\left(\boldsymbol{F}_{c_ic_j}\right) + \sum_{j\in\mathbf{N}_p}\left(\boldsymbol{F}_{c_ip_j}\right)\right), \tag{3.17}$$

$$\frac{d\boldsymbol{x}_{p_k}(t)}{dt} = \boldsymbol{v}_{p_k}(t) = \frac{1}{6\pi\eta R_p}\left(\sum_{j\in\mathbf{N}_p}\left(\boldsymbol{F}_{p_kp_j}\right) + \sum_{j\in\mathbf{N}_c}\left(\boldsymbol{F}_{p_kc_j}\right)\right). \tag{3.18}$$

3.2.4 Implementation

Mechanical interactions occur faster than biological processes ($\Delta t_{mech} < \Delta t_{bio}$), which allows them to be uncoupled and implemented with different time steps [266, 273].

The fluid secretion and agent dynamics are solved every $\Delta t_{mech} = 0.01$ min, and when the current simulated time (t) increases $\Delta t_{bio} = 6$ min ($t = t_{bio}$), the cell cycle is solved for each cell along with the fluid secretion and the agent dynamics. Figure 3.3 presents a simplified flowchart of the implemented algorithm. Initially, we begin the simulations with an active cell. The variable $state_i$ stores for each cell whether they are active or inactive within the cell cycle. After division, in the restriction point, cells can reenter in the cell cycle or become inactive, so the variable $state_i$ is updated for each daughter cell. If the cell becomes inactive, it enters into a quiescent phase and does not grow anymore ($\alpha_i = 0$). In the fluid secretion part, the variable $polarized_i$ accounts whether cells are polarized or nonpolarized, and n_l refers to the number of lumens in the organoid, therefore, $n_c(n_l)$ is the number of cells in the n_l lumen. Moreover, we track the fluid production time t_{exo_i} for each polarized cell and the luminal hydrostatic pressure $F_{in}(n_l)$ in the n_l lumen. If the cell secretes fluid, we reset its fluid production time t_{exo_i}. Finally, after fluid secretion, we solve the agent dynamics.

To evaluate the predictive capacity of the model, we replicate the experiment developed by Paszek et al. [68], who studied the influence of matrix rigidity in the lumen formed by MCF10A cells and concluded that matrix stiffening compromises tissue organization, inhibits lumen formation and enhances growth. However, there was no clear conclusion about the intrinsic mechanisms that regulate this process. Here, we focus on reproducing the results by comparing the lumen formation in a low-density matrix and a high-density matrix. Different techniques have been employed to model the ECM (e.g., the Smoothed-Particle Hydrodynamics (SPH) method [274]). Here. we model the ECM through a uniform dynamic viscosity and relate the density of the ECM with the dynamic viscosity based on Valero et al. [275], which presents a characterization of crosslinked collagen-based hydrogels. The low-density matrix represents the 4 $\mathrm{mg/ml}$ collagen concentration, with a dynamic viscosity of approximately 20 $\mathrm{Pa\,s}$ (η_{low}) [275]. Regarding the high-density matrix, we consider a dynamic viscosity of 100 $\mathrm{Pa\,s}$ (η_{high}), which corresponds to a five-fold increase in the dynamic viscosity with respect to the low-density matrix. Moreover, to analyze the trend of the lumen evolution under different-density matrices, we include an intermediate-density matrix with a dynamic viscosity of 50 $\mathrm{Pa\,s}$ for comparative purposes.

Finally, we initiate the simulations with one cell of radius 10 μm, then create a random seed and run the simulation for 7 days. Thus, we first performed 20 simulations with the low-density matrix and then, with the random seeds generated for each, we executed the equivalent simulations for the intermediate- and

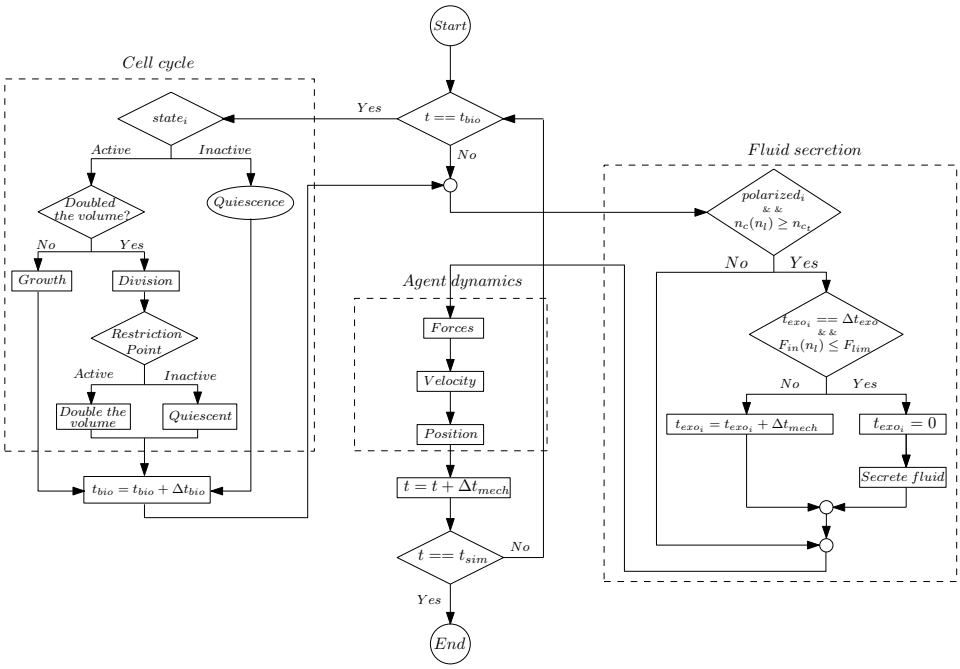

Figure 3.3: ***A simplified flowchart of the implemented algorithm****. First, we check if the current simulated time t is equal to the time at which the cell cycle has to be solved t_{bio}. If that is the case, the cell cycle is solved for every cell. In the cell cycle, for each cell, we check whether it is active or inactive. If it is active, we evaluate if the cell has already doubled in volume and is ready to divide or if it has to continue growing. After division, the restriction point decides whether each daughter cell remains active and prepares to double in volume or enters into a quiescent state. If the cell is inactive, it remains quiescent. After the cell cycle has finished, t_{bio} is increased. Then, we check if cells are polarized and if the number of cells in each lumen is higher than n_{c_t}. For each lumen that meets that condition, we check for each cell polarized around that lumen if it is time to secrete fluid and if the lumen pressure is lower than F_{lim}. If the cell secretes, we reset its time of fluid production t_{exo_i} and secrete fluid. If not, the cell produces fluid by increasing t_{exo_i}. Afterward, we solve the agent dynamics. We obtain the forces produced by mechanical interactions and the velocity of the agents and update the agents' positions. Finally, we increase the simulated time; if the stipulated simulation time t_{sim} is reached, the simulation is complete; otherwise, we repeat the algorithm.*

Table 3.1: ***Parameters of the lumen morphogenesis model.***

Parameter	Description	Value	Source
Δt_{exo}	Fluid production time	5 h	Estimated
η_{low}	Dynamic viscosity of the low-density matrix	20 Pa s	[275]
η_{high}	Dynamic viscosity of the high-density matrix	100 Pa s	[68, 275], estimated
F_{bot}	Mean value of the cells' net force in the low-density matrix	0.42 pN	Estimated
F_{top}	Mean value of the cells' net force in the high-density matrix	1.17 pN	Estimated
T_{ini}	Cell cycle time constant	30 h	Estimated
T_{pol}	Cell cycle time constant	100 h	Estimated
n_{t_c}	Number of cells to form the PAP	5	Estimated
a_{max}	Maximum variation in the growth rate	60%	[275], estimated
b	Quiescence probability	0.1	Estimated
R_p	Particle radius	2 μm	Estimated
$F_{rep_{cc}}$	Cell-cell repulsive force	-4.80 pN	[272], estimated
$F_{adh_{cc}}$	Cell-cell adhesive force	24 pN	[272], estimated
$F_{rep_{pp}}$	Particle-particle repulsive force	-4.80 pN	[272], estimated
$F_{adh_{pp}}$	Particle-particle adhesive force	24 pN	[272], estimated
$F_{rep_{cp}}$	Cell-particle repulsive force	0.60 pN	[272], estimated
$F_{adh_{cp}}$	Cell-particle adhesive force	0	[272], estimated
λ	Matching constant of the potential function	7	[272])
F_{lim}	Force threshold for cell secretion	0.07 pN	Estimated

high-density matrices. The agent surfaces were discretized and then processed with alpha shapes [276]. The parameters used for the simulations are shown in Table 3.1. The code was fully implemented in MATLAB R2019a.

3.3 Results

3.3.1 Lumen morphogenesis requires a low-density matrix to be accomplished

First, we studied lumen formation in the low-density matrix with a dynamic viscosity of 20 Pa s (η_{low}). We simulated the evolution of one lumen during 7 days (Figure 3.4). The simulation began with only one cell, and on the first day, it grew and then divided into two cells (day 1). At this moment, a midbody was

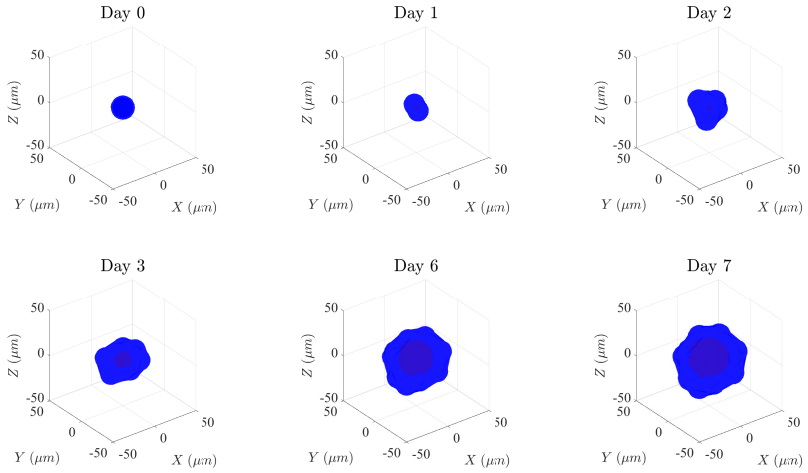

Figure 3.4: ***Evolution of a simulated cyst in the low-density matrix*** *($\eta = 20$ Pas). Cells are represented in blue and the lumen fluid in red.*

created (red dot), which determined the apical-basal polarization of the cells and the common site where cells would secrete. At day 2, the cells divided around the midbody to create a closed volume. Then, cells began to secrete and create a small lumen. At day 3, the lumen was formed by 8 cells and was already in a state of hydrostatic pressure. Between day 3 and day 6 the lumen continued growing, increasing its volume by a factor of 9.7 times on day 6 compared to day 3, and the number of cells increased to 16. Finally, at day 7, a cyst with a single lumen formed by a monolayer of 16 cells was achieved.

3.3.2 Increasing the matrix density hinders lumen morphogenesis and produces an aberrant multiluminal architecture

Then, we investigated, with the same seed of random numbers, the evolution of the lumen during 7 days of simulation in a dense matrix with a dynamic viscosity of 100 Pas (η_{high}) (Figure 3.5). Again, the simulation began with one cell at day 0, which grew and divided. At day 1, the midbody was already generated. At day 2, due to the fact that in the high-density matrix the cells' net forces are higher than those in the low-density matrix, the growth rate of cells had increased, so the number of cells increased to 8. At this point, a closed volume for lumen initiation was created, and cells secreted fluid into the lumen, generating a small lumen. In the next 24 hours, cells continued secreting, and the lumen volume increased slightly. Between day 3 and day 6, significant changes were observed. The number of cells increased from 8 to 24. However, the lumen volume increased by only a factor of 3 (the increment in the low-density matrix was

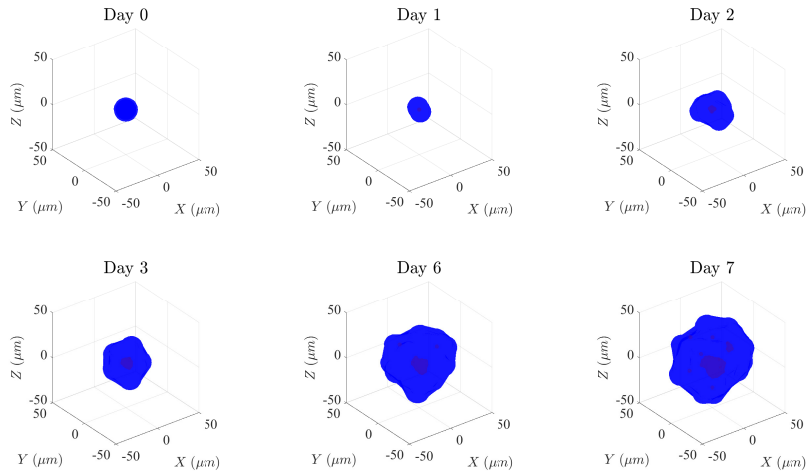

Figure 3.5: ***Evolution of a simulated cyst in the high-density matrix*** *(η = 100 Pas). Cells are represented in blue and the lumen fluid in red.*

approximately 9.7). The high-density matrix opposes cell movement, so the luminal hydrostatic pressure cannot displace the cells. Consequently, the pressure increases with successive secretions without enlarging the lumen, at times reaching the hydrostatic pressure limit at which cells cannot pump more fluid into the lumen. Moreover, as the lumen is small, cell divisions caused some cells to move into a second layer where they ceased to face the lumen. This change made them polarize to create two midbodies and later two new lumens. Finally, at day 7, a dysfunctional structure composed of 7 lumens and 53 cells was observed.

3.3.3 Matrix density reduces the lumen volume

To compare the lumen formed in the low and high-density matrices, we show the structure achieved (Figure 3.6a and Figure 3.6b (left)) and a view of the lumen with only the cells that are situated approximately below the center of the organoid in both matrices (Figure 3.6a and Figure 3.6a (middle)) after 7 days of lumen formation. In the case of the low-density matrix, a cyst with a single regular lumen and 16 cells was developed. However, in the high-density matrix, 7 lumens were developed. The volume of the largest lumen for the high-density matrix was 80.64% smaller than that in the case of the low-density matrix, and the secondary lumen volume formed in the high-density matrix was 94.2% less than that of the principal lumen of the cyst. Moreover, a non-linear increase of the lumen volume was obtained in the case of the low-density matrix (Figure 3.6a (right)) since there were more cells to secrete fluid as the simulation time increased due to cell proliferation. In contrast, in the high-density matrix, the principal lumen

volume grew discontinuously (Figure 3.6b (right)), with higher increases in lumen volume as the simulation proceeded. The evolution of the pressure and the number of cells is represented in the low- and high-density matrix (Figure 3.6c and Figure 3.6d). The pressure was obtained by integrating the interacting forces between the cells and the lumen fluid ($\boldsymbol{F}_{cp}$) divided by the lumen surface. The pressure increased after day two with successive cell secretions. Thus, the pressure in the low-density matrix (Figure 3.6c) increased when the cell secreted fluid and then decreased as a consequence of lumen expansion. After day 5, a significant decrease in pressure was developed since cells proliferated and generated more inner volume. In contrast, in the high-density matrix (Figure 3.6d), with each cell secretion the pressure increased more sharply. In this situation, little lumen expansion was produced and, therefore, the pressure increased with time. In this case, the total population of cells increased 3.3 times compared to the low-density matrix, from 16 cells to 53 cells.

To analyze the robustness of the model, we show (Figure 3.7) the lumen volume after 7 days for the 20 simulations performed in the low and high-density matrices[1] (see Figure A.1 in *Appendix* for a representation of the lumen volume after 7 days for each simulation performed in an intermediate-density matrix ($\eta = 50$ Pas)). In the case of the low-density matrix (Figure 3.7a), 9 of 20 simulations formed a single large lumen, and 11 simulations formed a principal large lumen with some additional small ones. The median value of the number of lumens was 2.5. When the sum of the lumens coincides with the higher lumen volume value, the volumes of the secondary lumens are insignificant compared to that of the principal lumen. In the case of the high-density matrix (Figure 3.7b), only 2 cases produced a single lumen, and the median number of lumens increased to 6. In all simulations in the high-density matrix, the higher lumen volume is significantly smaller than that in the previous cases. Moreover, two main cases are identified, those with a small principal lumen with various lumens of negligible volume and those with a small principal lumen, some secondary lumens smaller than the principal but still significant and some minor lumens (for instance, simulations 1, 3, 6, 13, 15 and 19).

We compare the higher lumen volume of the organoid obtained in each simulation after 7 days in the low-, intermediate- and high-density matrices for the 20 simulations performed (Figure 3.8). The lumen volume obtained in the simulations in the low-density matrix presents a higher variability than that in the high-density matrix. Thus, the standard deviation is 2.8 times higher than that in the high-density matrix. However, the lumen volume achieved in most cases in the low-density matrix is significantly greater than that in the high-density matrix. Additionally, the median lumen volume difference between the low-density and high-density matrix is approximately 74.2%, between the low-density

[1]The cases of lumen morphogenesis in the low and high-density matrices presented previously correspond to the second simulation ($n_{sim} = 2$) in Figure 3.7a and Figure 3.7b respectively.

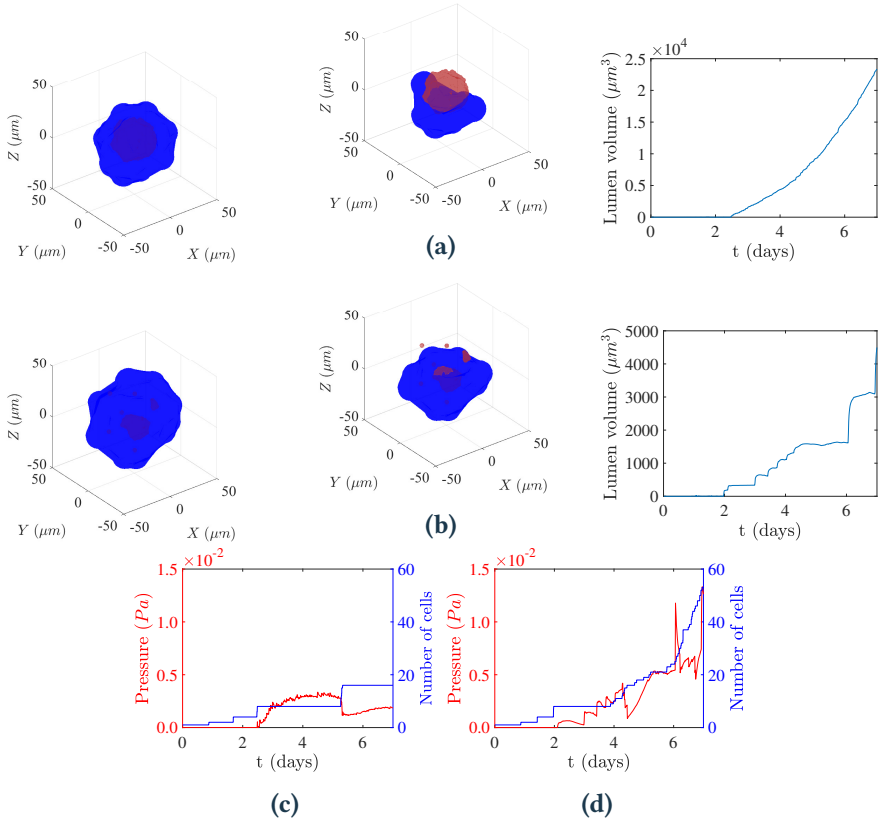

Figure 3.6: ***Comparison of lumen formation in the low and high-density matrices.*** *Cyst formed after 7 days of simulation in the low-density matrix in (a), and in the high-density matrix in (b). Full view of the cyst (left), a perspective of the cyst with only the cells positioned approximately below the center of the organoid (middle), and the evolution of the lumen volume (right). Cells are represented in blue and the lumen fluid in red. Evolution of the pressure and number of cells in the low-density matrix in (c) and in the high-density matrix in (d).*

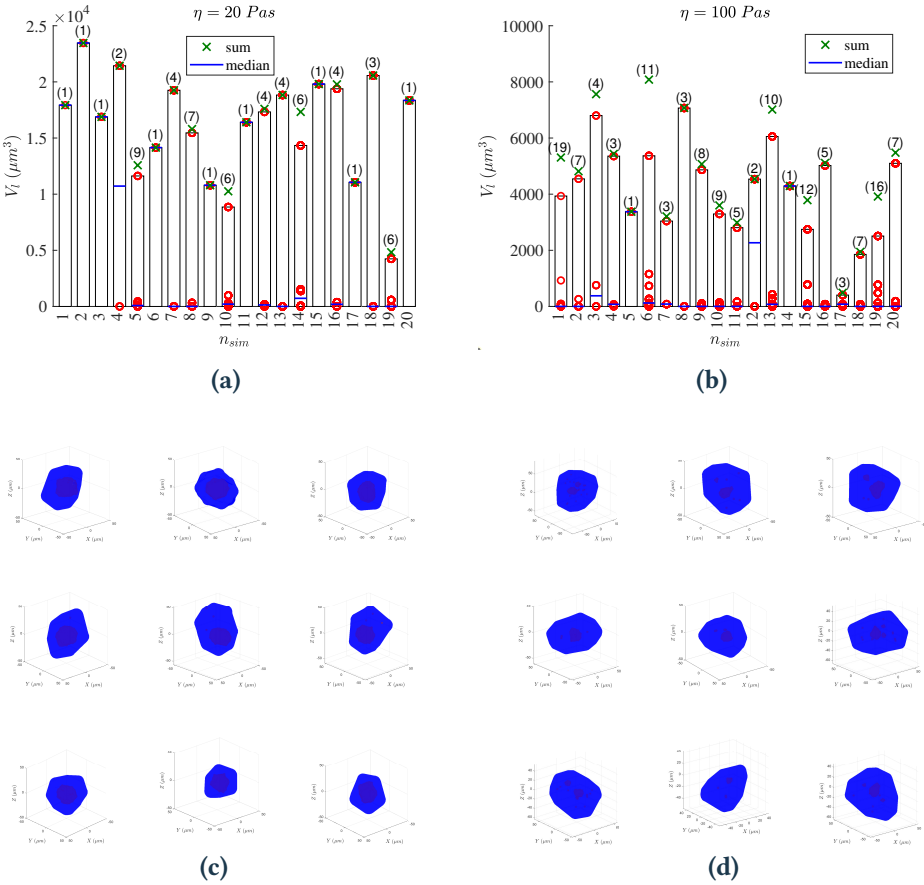

Figure 3.7: ***Comparison of the lumen volume in the low and high-density matrices.*** *Volume of the lumen after 7 days for each of the twenty simulations in the low-density matrix in (a) and in the high-density matrix in (b). The numbers in parentheses indicate the number of lumens produced, the red circles represent the discrete volume of each lumen, the green crosses show the sum of the volumes of all lumens in each simulation and the blue line is the median. The bars represent the volume of the largest lumen, which are plotted for the sake of the visualization (note the difference between the scales of the graphs). A representation of the cyst formed in simulations 1, 2, 3, 4, 7, 13, 15, 17, and 20 is shown in the case of the low-density matrix in (c) and in the case of the high-density matrix in (d).*

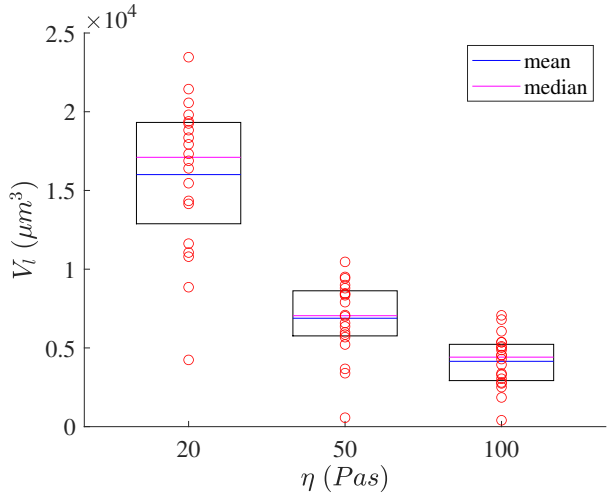

Figure 3.8: ***Higher lumen volume of the organoid after 7 days for each simulation in the low-, intermediate- and high-density matrices.*** *Red circles are the discrete values of each simulation. In cases in which multiple lumens were produced, the red circle corresponds to the volume of the largest lumen. The top line of the box represents the 75th percentile and the bottom line the 25th percentile. The magenta line shows the mean and the blue line the median.*

and intermediate-density matrix is approximately 57.0%, and 39.7% between the intermediate-density and high-density matrix.

Finally, the mean interactive cells' net force during the 7 days simulated for each simulation in the case of the low-, intermediate- and high-density matrices is presented in Figure 3.9. The forces acting on cells are higher in the high-density matrix than in the low-density matrix. In particular, the median in the high-density matrix is approximately 4.5 times greater than that in the low-density matrix. Moreover, the standard deviation in the high-density matrix is 2.98 times higher than that in the low-density matrix. In the case of the intermediate-density matrix, the median is 2.3 times greater than in the low-density matrix but 0.51 times lower than in the high-density matrix.

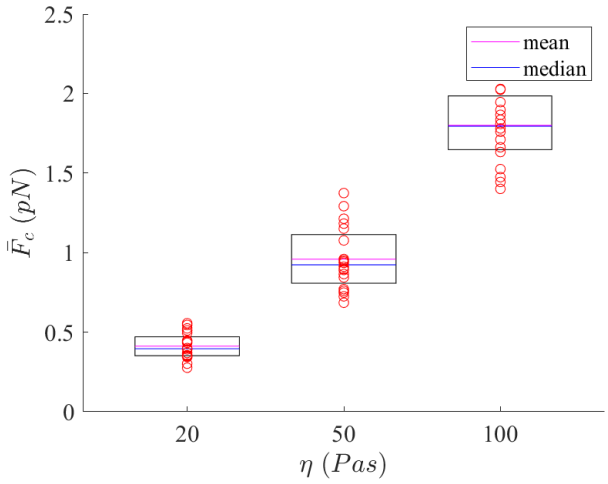

Figure 3.9: ***Mean value of the cells' net forces for each simulation in the low-, intermediate- and high-density matrices.*** *Red circles are the discrete values of each simulation. The top line of the box represents the 75th percentile and the bottom line the 25th percentile. The magenta line shows the mean value and the blue line the median.*

3.4 Discussion

We present a 3D multi-ABM for lumen morphogenesis that introduces the regulatory role of the luminal hydrostatic pressure generated by the cells' fluid secretion and the interaction with the ECM. These biophysical effects allow the simulation of lumen morphogenesis under three matrices with different densities and the determination of how the density influences its formation. We show that increasing the density of the matrix hinders lumen morphogenesis, increases the number of cells, and produces a dysfunctional architecture. Moreover, the model predicts normal lumen morphogenesis and aberrant multilumen formation related to tumor formation as a result of the increased density of the matrix.

In a high-density matrix, we find alteration of the lumen formation process that produces an aberrant structure. The main contribution to the disruption of normal lumen morphogenesis is the balance between the matrix and the forces derived from the luminal hydrostatic pressure. When the matrix density is high, the luminal hydrostatic pressure generated by the cell secretion cannot displace the cells as easily as when the matrix density is low since the matrix opposes the cells' movement. Consequently, the lumen remains diminished, and the hydrostatic pressure continues increasing with the successive cell secretion to a value at which point cells cannot pump more fluid into the lumen. In this situation, cell division does not contribute to creating luminal space because the lumen size is

small, and its surface is fully covered by overlapped cells. Conversely, as there is no space around the lumen surface, cell division forces some cells to leave the monolayer. At this point, the cells that do not face the lumen attempt to polarize to create a new lumen. Finally, a dysfunctional structure with numerous small lumens is formed. This aberrant architecture is associated with tumor pathogenesis and is found in many carcinomas [236]. Therefore, we conclude that the relationship between the matrix density and the luminal hydrostatic pressure is crucial and might help to determine the development of tumor pathogenesis [68].

Moreover, the matrix density not only acts as a damping and determines the timescale of the problem, but also provides important mechanobiological feedback to cells. From a mechanobiological perspective, the second aspect that contributes to lumen malformation is the cell net force. We showed that cells' net forces increase with matrix density (Figure 3.9), and this increment in the net forces transforms the cells towards a malignant phenotype, which enhances the cell growth rate. The interplay between mechanical forces, ECM, and growth factors controls the cell cycle progression [259]. Thus, the compliance of the matrix acts as a cell-cycle inhibitor and matrix stiffening increases cell proliferation [255–258]. Our model includes this effect and when the matrix density is higher, the cell net forces are higher and it makes cells increase their growth rate, therefore, the total number of cells increases with matrix density. This contributes to some cells escaping from the lumen they are facing and attempting to form a new lumen. Moreover, we showed that there is low variability in mean cell net force; therefore, it is strongly influenced by matrix density. Therefore, the density of the matrix acts not only as a relaxation time of the system, but also as a regulator of cells' net forces, luminal pressure, morphology, and the formation of multiple lumens.

Our work replicates the experimental observations of [68]. They found that matrix rigidity compromises tissue organization, inhibits lumen formation, increases colony size, and increases cell forces. These four tendencies are observed in our results; in the case of the high-density matrix, an aberrant architecture with multiple lumens was obtained, the lumen volume was significantly smaller than that in the low-density matrix, and the number of cells and the cells' forces increased.

In the proposed model, several simplifications have been considered. Cells are assumed to be nondeformable spheres; therefore, cell shape was not represented accurately. In a deformable model, we would have cell deformation around the lumen. However, apart from the shape representation, little differences in cells' position are expected. This simplification reduces the computational cost of the simulations, but it does not affect the process of lumen morphogenesis. Here, we focused on simulating lumen morphogenesis from the beginning with an individual cell; therefore, we are focused on determining the position of each cell to understand their coordination rather than obtaining the exact representation

of the cell shape. Moreover, we represented the lumen fluid through a particle-based simulation. With this approach, the mechanical properties of the fluid are not accurately described, especially concerning the rheological behavior of the lumen. Although this representation is a simplification, it allowed us to simulate the interaction between cells and the fluid in a developmental process in which the interface between cells and the fluid changes with time. Also, our model permits the simulation of the formation of lumens *de novo* and the increase of the luminal fluid.

In our model, we characterized the ECM by means of the dynamic viscosity for the whole domain. With this approximation, we are considering the viscous properties of the matrix but not the elastic ones. Despite this simplification, our model allows us to study how the biophysical properties of the ECM, represented through its viscosity, affect the lumen size and morphology in a three-dimensional simulation. Both cells and particles experiment with a drag force that opposes their relative motion with the matrix. It may be reasonable to expect variation in the dynamic viscosity between the lumen and the outside of the cyst. A lower dynamic viscosity inside the lumen implies lower resistance to the motion of particles, and we might expect variation in the growth and size of the lumen. However, this aspect is already considered through the parameter representing the strength of the repulsive and adhesive forces in the particle-particle and particle-cell interactions. Thus, a reduction in dynamic viscosity in the lumen facilitates particle movement, which is equivalent to an increase in the magnitude of these interacting forces.

In this study, we suggest that there is a maximum luminal pressure above which the cells cannot secrete. On the one hand, an increase in the force threshold for cell secretion allows higher hydrostatic pressure, and the lumen might open. On the other hand, if the threshold is excessively restricted, cells cannot secrete, and the lumen remains diminished. Therefore, this parameter is set such that it allows the lumen to grow without opening.

This computational model is conceptualized as a tool for simulating lumen morphogenesis by cells in different organs. Thus, the parameters used in the model could be experimentally quantified and, therefore, enable adjustment of the model to simulate lumen morphogenesis in different scenarios. Specifically, the cell cycle constants of the model can be experimentally estimated to characterize diverse cell types that form a lumen. The fluid production time can be estimated by monitoring the volume of the lumen in cell cultures such that the evolution of the lumen volume obtained can be fitted with the experimental results.

3.5 Conclusions

This work finds a strong correlation between the density of the ECM and lumen morphogenesis. Thus, an optimal density that provides adequate biomechanical conditions to form the lumen and reach a proper structure may exist. We showed that cells require a low-density matrix to form a normal lumen. In this case, cell division and cell fluid secretion act in a coordinated manner to form a normal lumen. However, an increase in matrix density disrupts this coordination and promotes an aberrant multiluminal architecture. Therefore, matrix density provides crucial properties to regulate the deviation from normal lumen morphogenesis to tumorigenesis. Finally, we would like to highlight that this computer-based model has allowed us to investigate different mechanical scenarios representative of a relevant biological process as lumen morphogenesis, clearly demonstrating the possibilities of computational simulations in biological engineering research.

Chapter 4

Organoid morphogenesis

Imagination is more important than knowledge.

– ***Albert Einstein***

Contents

This chapter is based on:

Camacho-Gomez, D., Sorzabal-Bellido, I., Ortiz-de-Solorzano, C., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2023). *A hybrid physics-based and data-driven framework for cellular biological systems: Application to the morphogenesis of organoids.* iScience, 26(7). https://doi.org/10.1016/j.isci.2023.107164.

Abstract

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

4.1 Introduction

Organ morphogenesis can be studied *in situ* through the analysis of stained tissue sections taken at different phases of embryonic development. This provides relevant, albeit static, information about the evolution of the morphology and the interactions between the elements of the developing organ. Morphogenesis can also be studied using 2D *in vitro* models built from stem cells properly stimulated with growing factors and/or cocultured cells. These models replicate some of the cellular interactions that occur during organ development, providing important information about the signaling pathways involved in normal organ formation. 2D culture models, however, lack a key element of organ morphogenesis, namely, the complex ensemble of three-dimensional spatial interactions that occur during the developmental process. These interactions, which determine the successful series of events that end with the formation of a mature organ, include not only interactions between cells but also, very importantly, interactions between the cells and their local microenvironment. This has caused the birth of a growing trend in the study of morphogenesis, which is the use of 3D *in vitro* cellular models or organoids that recreate *in vitro* the process of organ formation [277–279]. Some organotypic models have also been developed that replicate the growth and development of tumors (tumor organoids) from cancer stem cells and the appropriate stimulation of growth factors and biomechanical cues [280, 281]. These complex models replicate the interactions that exist between cancer cells and with other cells of the tumor microenvironment, most relevantly the cells of the immune system. Tumor organoids can be of different complexity, from simple spheroids made of one or a few cell types suspended in medium enriched with the

appropriate growth factors [282] to more complex models comprised of several cell types embedded in a biomimetic hydrogel. Finally, more advanced models are being developed based on microfabrication-bioprinting and microfluidic technologies that allow the simulation of interstitial flows [283]. Tumor organoids are becoming invaluable tools to study the response of tumors to therapies, including those that stimulate the host immune system.

There are also different approaches to simulate morphogenesis *in silico*. On the one hand, continuum models permit the simulation of large cell populations at the tissue level. This is the case for reaction-diffusion systems based on PDEs [284] or positional information (PI) [285]. However, these methods neglect the individuality of cells in favor of larger scales, disregarding the important role played by cell-to-cell interactions and the phenotypic diversity of the biological system. Consequently, these methods struggle to reproduce the way cell communities develop into complex structures, especially in three dimensions. On the other hand, discrete approaches with agent-based modeling consider cells as autonomous entities that interact among themselves and with the microenvironment [286]. These methods are usually classified into lattice models, in which cell positions are fixed within a spatial grid [100], and lattice-free models, which allow continuous cell positioning in space [272]. Based on the way the cell shape is represented in lattice-free models, there are center-based models, in which cells are represented as spheres and are described by their centers and radii [287], and deformable cell models and vertex models, which account for the cell shape and deformability [288, 289]. These models have been widely employed, for instance, to simulate tumor growth *in vitro* [246], to study the role of ECM density in cell migration within tumor spheroids [217], collective cell migration [290], or tissue regeneration [291]. These approaches seem appropriate to simulate morphogenesis, it being a process that evolves from a single cell, and the emerging pattern emerges from cell interactions, cell heterogeneity, and cell processes [111, 245]. Furthermore, agent-based modeling focuses on the cell level; therefore, it very flexibly simulates *in vitro* experiments. However, these models have some challenges related to the designation of rules and fitting parameters that govern decision-making heuristics and the behavior of cells [111, 245]. Additionally, they usually have a large number of parameters compared to continuous models, especially when considering molecular events per cell [292], which makes their calibration difficult.

Data-driven models are widely used to predict the behavior of complex physical systems [293, 294]. They can unravel unknown phenomena and reduce the degree of abstraction of computational models. In computational biology, there are examples of data-driven models used to find the optimal parameter values in gene regulatory networks to reproduce the growing limb bud [295] or Bayesian inference of agent-based cellular automaton to study kidney branching morphogenesis [296]. These methods combining data-driven and computational models

aim to obtain the required parameters of the computational model, which is an optimization method that lacks generalization, as in the case of the data changing, the parameters must change as well. In the context of data-driven methods, there is an increasing use of artificial intelligence and, in particular, ML and DL algorithms [297]. However, data-driven models are considered black boxes and are difficult to relate to physical interpretations, as they are mainly focused on finding input-output relationships.

In this work, we propose a hybrid approach that combines DL and ABMs as a novel computational framework to simulate cellular biological systems. In other approaches, the coordination between cells is fixed beforehand and then adjusted to reproduce a certain behavior. On the one hand, in continuum models, cell processes are included in constitutive equations and these cell functions are performed at the same time [298, 299]. Then, they are adjusted through a sensitivity analysis to obtain the fitting solution. However, there is no competition between cell functions, as they occur simultaneously. On the other hand, in stochastic models, cells change between cell states through cell state rules, boolean circuits, gene regulatory circuits, and biochemical reactions from the more abstract to the more detailed [300, 301]. These relationships between cell states are complex to establish and describe [302]. Both approaches are usually made to simulate one particular case or median cases from experiments. However, it is difficult to reproduce different cases since their parameters are adjusted to reproduce one observed behavior and the variability is obtained by the introduction of noise. Moreover, they usually fail to reproduce dynamical processes due to the intricacy of recreating distinct behaviors temporarily with fixed parameters and even more taking into account the variability between cells. In this new work, we do not determine how cells behave before defining the relationships between cell states in order to perform a sensitivity analysis to find the parameters that allow us to reproduce an observed behavior. Instead, we eliminate all the parameters in the physics-based model that preestablish the behavior of cells and incorporate the mechanical cell interactions and biological cell functions in an ABM that uses a NN to drive the cellular decision-making of the agents during the simulation, replicating how cells make decisions and behave *in vivo*. This DL algorithm coordinates cell decision-making by evaluating the simulation in real time and comparing it with metrics that define the final architecture of the morphogenetic pattern extracted from cell-level images of real organoids or *in vivo* data.

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

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

4.2 Methods

4.2.1 Hybrid physics-based and data-driven framework

We present a novel hybrid physics-based and data-driven framework that combines agent-based modeling and DL to simulate morphogenetic patterns (Figure 4.1). The purpose of the physics-based model is to replicate a specific morphogenetic pattern. The data-driven algorithm in turn coordinates, through a NN, the cell functions required in the physics-based model to achieve that pattern (Figure 4.1**a**). Thus, the physics-based part consists of an agent-based model that couples mechanical interactions and the cell biological functions that are decisive in the morphogenetic process, namely, proliferation, quiescence, secretion, migration, cell death, etc. The selection of these cell functions depends on the morphogenetic process being simulated.

The physics-based model coordinates the biological behavior of cells to form a specific pattern, which also requires the appropriate mechanical interactions. For that coordination, instead of using parameters that fix the relationships between them, we introduce a data-driven algorithm that performs the cellular decision-making of the cell functions. Therefore, during the simulation of the agent-based model, the NN determines which cell function each cell has to perform (Figure 4.1**a**). To this end, given a morphogenetic pattern, we first define the data metrics that characterize the pattern and establish target values. These data metrics set the objective of the simulation. Then, when a cell needs to decide which function to perform, the NN receives as inputs the predefined data metrics and the simulation metrics from the agent-based model that evaluates the actual state of the simulation. In this way, the NN compares the simulation with the target values and makes a decision (see the data-driven box in Figure 4.1**a**). Each node of the output layer is related to one cell function; therefore, the activated node determines the biological function that the cell will perform. This decision is fed to the agent-based model, which initiates the corresponding cell function for that cell. This process is repeated during the predefined time of simulation every time a cell finishes the biological function that it was performing. When the simulation finishes, a fitness value is reported, which indicates how close the simulated morphogenetic pattern is compared to the desired morphogenetic pattern. Therefore, the result depends on how well the NN evaluates the simulation and coordinates cell functions through adequate decision-making.

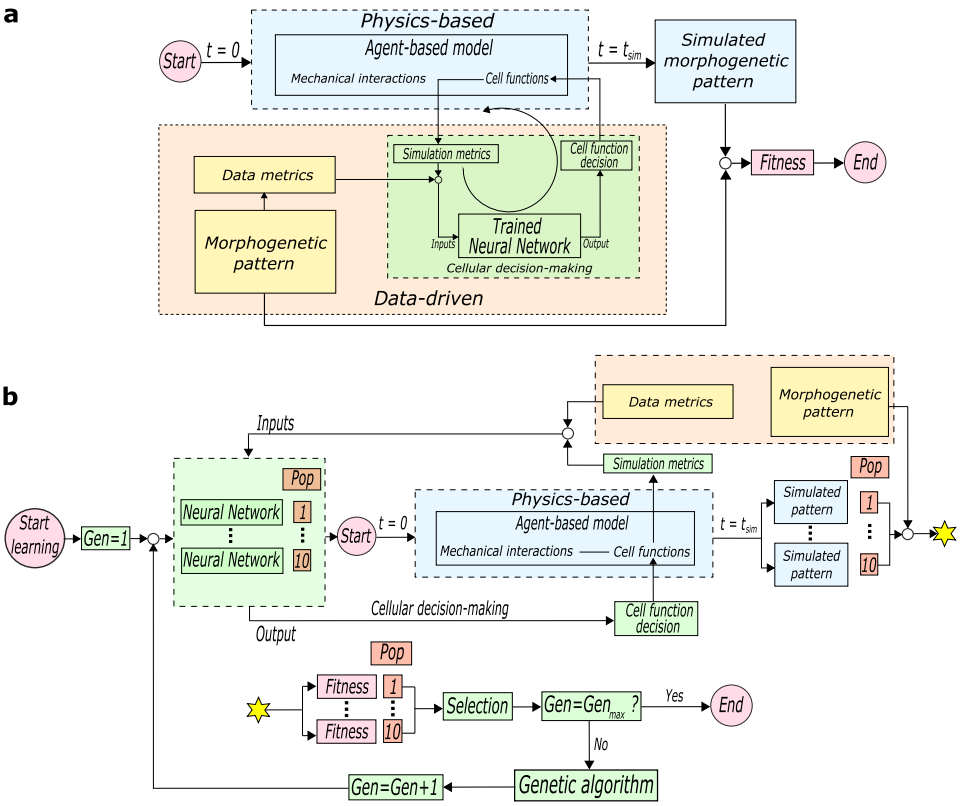

Figure 4.1: ***Framework of the hybrid physics-based and data-driven model.*** ***a*** *Schematic overview of the hybrid data-driven and physics-based algorithm. It consists of an agent-based model that incorporates mechanical interactions and cell biological functions. To orchestrate the cell functions within the simulation, the agent-based model employs a data-driven algorithm. Thus, to determine which function the cell has to perform, the agent-based model goes into the data-driven algorithm. The data-driven algorithm extracts from the desired morphogenetic pattern data metrics that characterize the pattern, which are introduced, together with simulation metrics, into the NN as inputs. The NN evaluates the inputs and determines the cell function that the cell must perform. This process is repeated until the simulation time of the agent-based model reaches (t_{sim}). Finally, a fitness value is given to the simulation by comparing the simulated and desired morphogenetic patterns.* ***b*** *Schematic overview of the training algorithm of the NN. The training starts with the creation of a random generation (Gen) composed of a population (Pop) of ten NNs. Then, the physics-based algorithm performs a simulation with each NN, following a similar methodology as* ***a****. Finally, the genetic algorithm selects the two best populations based on their fitness and creates a new generation through crossover and mutation algorithms. The training is concluded when the number of generations reaches the specified number of generations (Gen_{max}).*

4.2.2 Learning by the neural network framework

To simulate the morphogenetic pattern, the NN needs to learn how to evaluate the simulation and, depending on the target values, decide which biological function the cell has to perform. To make the NN learn, we propose a learning methodology based on a genetic algorithm (Figure 4.1**b**). Initially, the genetic algorithm creates a random population of ten NNs with different weight values assigned, i.e., a generation. Then, we perform a simulation of the agent-based model with each NN of the generation. When the simulation finishes, we obtain a simulated pattern for each population. Subsequently, each NN of the generation is given a fitness value, which represents how well they performed in the simulation. Finally, it selects from the current generation the two fittest NNs and obtains, through a genetic algorithm (further details are included in *Learning algorithm* in *Appendix* A.2.6), the next generation of NNs. This process is repeated for a predefined number of generations.

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

To show the potential of our approach, we simulated cystic tumor organoids, i.e., organoids formed by an outer shell made of cancer cells surrounding a lumen (Figure 4.2). Lumen morphogenesis implies the development of an organized architecture of cells around a fluid-filled space. This inner space, called the lumen, is found in many parts of metazoan organisms and allows them to perform specific functions. The development of a lumen requires sophisticated coordination of several cell processes. Specifically, three basic mechanisms are crucial [227]: Initially, cells proliferate and polarize to determine the landmark where the lumen will be initiated. Next, cells secrete fluid into that central site and generate luminal hydrostatic pressure that makes the lumen enlarge [231, 232, 243, 244]. To preserve the overall cystic structure, a balance between strong cell-cell junctions and spatially controlled cell mitosis is needed. Finally, the stiffness of the ECM plays an important role [68] by balancing the luminal hydrostatic pressure and providing physical cues that regulate cell behavior [234].

To simulate the morphogenesis of the lumen, we built a 3D ABM. This physics-based model is based on a previous 3D lattice-free center-based model [303]. Here, we use the structure of that model, but we eliminate the parameters that coordinate the relationships between cell functions, the effect of the cell net forces on the cell cycle, and multiluminal polarization mechanisms. To reproduce the morphogenesis of organoids composed of cells enclosing a fluid-filled lumen, we consider two types of agents: cells, which are the biological entities, and particles, which are secreted by cells and simulate the lumen fluid. The mechanical equilibrium between cells anchored to the ECM and the luminal pressure ensures the maintenance of the luminal architecture (further details are included in *Agent-*

based modeling in *Appendix* A.2.7). To model this mechanical equilibrium, we use agents that interact mechanically to generate and maintain the lumen. Thus, cells interact among themselves based on pairwise potential functions in an adhesive-repulsive manner. The repulsion between agents mimics cell resistance to deformation when their membranes touch, and the attractive forces are the result of the cell junctions exerted through specialized protein complexes. Moreover, particles also interact via pairwise potential functions and interact with cells in a repulsive manner. This interaction mimics the luminal hydrostatic pressure generated by the cells' fluid secretion, and it is responsible for the movement of cells and generation of the luminal space. Finally, agents interact with the ECM by means of a friction coefficient that represents the dynamic viscosity of the matrix (Figure 4.2 top left).

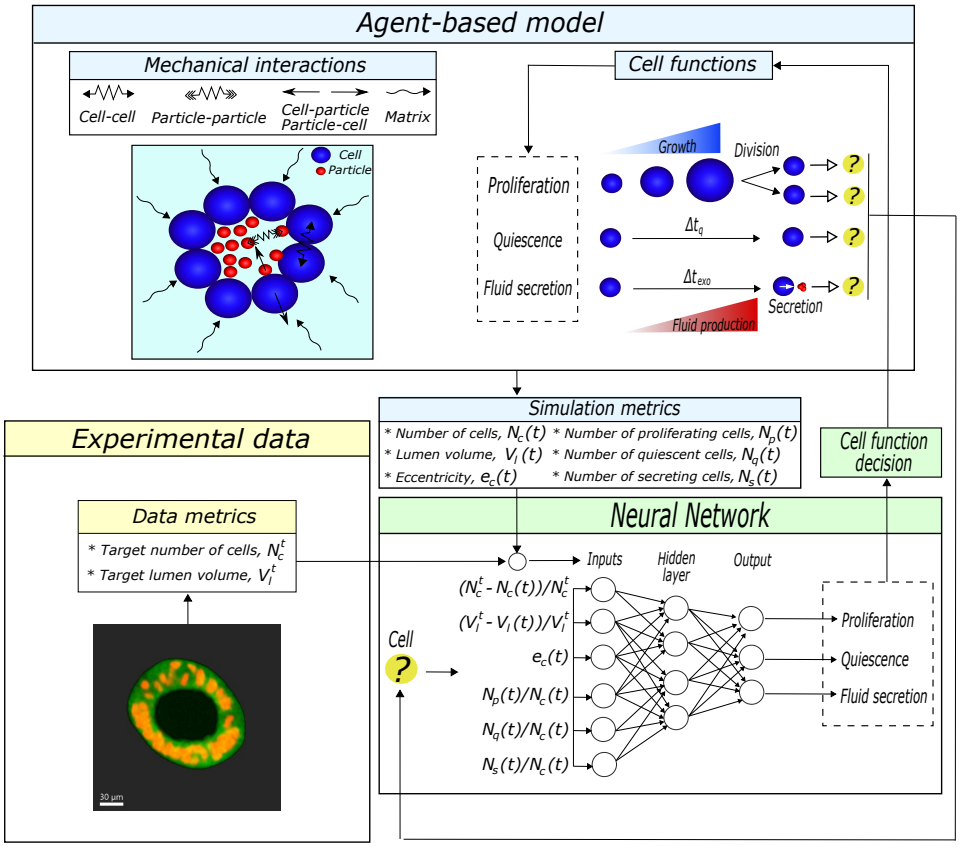

Figure 4.2: ***Particularization of the framework for organoids with lumen.*** *The physics-based model consists of an ABM that integrates mechanical interactions and cell functions. Cells are biological entities, and the particles simulate the lumen fluid. Three cell functions are considered: proliferation, quiescence, and fluid secretion.* Δt_q *is the minimum period of time that a cell remains quiescent and* Δt_{exo} *is the fluid production time.*

Three decisive cell functions are considered to achieve the cystic morphogenetic pattern: proliferation, quiescence, and fluid secretion (Figure 4.2 top right). First, we define a mathematical model for simulating the cell cycle that regulates cell proliferation. Overall, it consists of a growth phase, in which the cell progressively increases its volume as a result of DNA replication, and a mitosis phase, in which the cell divides into two daughter cells (see *Cell division* in *Appendix* A.2.8). In addition, we include a quiescent state in which the cell remains inactive for a period of time (Δt_q). Finally, cells can produce fluid during a period of time (Δt_{exo}) and then secrete it inside the lumen, increasing the lumen volume. To simulate this process, we assume that cells generate particles that simulate the lumen fluid (further details are included in the *Characterization of the lumen fluid* section in *Appendix* A.2.13). Consequently, the lumen is in a state of hydrostatic pressure due to this cell secretion.

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

In the data-driven framework, we first define the metrics that characterize the pattern using target values. In the case of organoids with lumen, we defined the target number of cells (N_c^t) and target lumen volume (V_l^t) to describe the morphology of the organoid (Figure 4.2 middle). We also define several simulation metrics that are updated when a cell has to decide its next cell function. Specifically, these simulation metrics are the total number of cells, the lumen volume, the eccentricity between the cells center of mass and the lumen fluid center of mass (see *Eccentricity* in *Appendix* A.2.10), and the number of proliferating, quiescent and secreting cells. These values are given as inputs to a NN (Figure 4.2 bottom right). We assume that each cell can only perform one cell function at a time. Hence, the NN determines the appropriate cell function that cells must perform. The NN consists of an input layer with six nodes, one hidden layer with four nodes and a hyperbolic tangent activation function, and a layer with three outputs and a sigmoid activation function. Each node of the output layer identifies the function that the cell must perform: proliferation, quiescence, or fluid secretion.

4.3 Results

4.3.1 Learning of the neural network to simulate cystic organoids

The success of the morphogenetic pattern depends on how well the NN coordinates cell functions during the simulation. To make the NN learn to form organoids with a lumen, we trained the NNs for 400 generations. For this purpose, we did not use any of the experimental data; instead, we used a synthetic

pattern consisting of an organoid with 8 cells and a lumen of $1000\,\mu m^3$. To illustrate this process, we show the increase in performance for each generation of NNs (Figure 4.3). As shown, the first generation of NNs did not coordinate the cell functions properly, as the fitness (see *Fitness function* in *Appendix* A.2.9) of this generation was 0.1869. In fact, the NN only allowed cells to proliferate, thus generating an organoid without a lumen. Then, the performance increased to 0.6136 in generation 100, as cells were also secreted to generate the lumen. However, the fluid of the lumen leaked due to the poor coordination between proliferation and secretion. In generation 200, the fitness increased to 0.8846, and better coordination than in generation 100 was found, in that lumen fluid did not leak, but the organoid was too asymmetrical. Generation 300 achieved a symmetrical organoid with lumen with a better approximation of the number of cells and lumen volume. Finally, in generation 400, an organoid composed of 7 cells and a lumen volume of $1245\,\mu m^3$ was achieved.

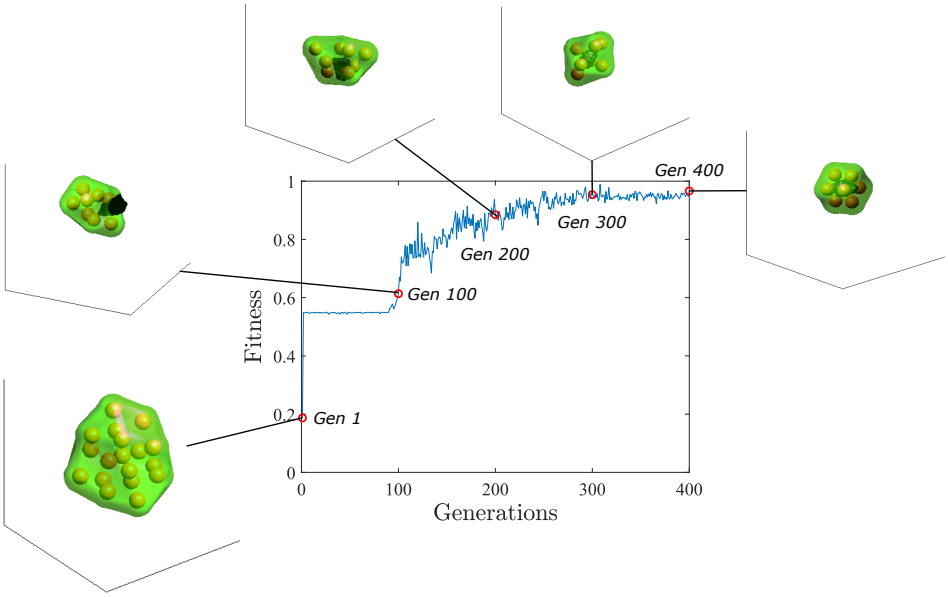

Figure 4.3: ***Learning of the neural network of cystic organoid formation.*** *The fitness value represented corresponds with the highest fitness of the population within the generation. Snapshots of the achieved organoid inserted in a cube with* $100\,\mu m$ *sides featuring one population at generations 1, 100, 200, 300, and 400 are represented. The spheres represent the nuclei of cells with radius* $R_c/2$ *and the green hull is an estimation of the cell membrane through alpha shapes of the cells with* $\alpha = 2R_c$ *[276]. The lumen fluid is represented in black through the alpha shapes of particles with the smallest alpha that produces an alpha shape enclosing all of the particles.*

4.3.2 Simulation of pancreatic cystic tumor organoids

We simulated 3D cultures of murine pancreatic ductal adenocarcinoma cells grown in a biomimetic matrix made of Matrigel ($4\,\mathrm{mg/ml}$). Real 3D cultures of PDAC93-GFP cells were generated and incubated for 5 days. In the experimental setup, we generated a randomly distributed population of organoid seeds ranging from single cells to small clusters of 2-3 cells (see Figure A.3 in *Organoid seeds* in *Appendix* A.2.11). Then, multiphoton microscopy images of the grown organoids were segmented and quantified to characterize their morphology. In particular, we obtained the number of cells in the organoid and the lumen volume, which were used as target values for the simulation. We chose organoids of different sizes to test the generalization potential of the methodology (Table 4.1). Specifically, we used two intermediate-size samples (organoids #1 and #2), one small-size sample (organoid #3), and one large-size sample (organoid #4). We initiated the simulations with one cell and ran the simulations for 7 days. Although the PDAC93-GFP cells were incubated for 5 days, we extended the simulation time to 7 days to prove that the model was able to reach steady-state patterns. The parameter values used in the simulations are described in Table A.2 in *Simulation parameters for pancreatic tumor cystic organoids* in *Appendix* A.2.12.

Table 4.1: ***Experiment and simulation results of pancreatic cystic tumor organoids.*** *N_c^t is the number of target cells and V_l^t the target lumen volume.*

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

We show images of the experimental organoids used as targets of the simulation, the achieved solution of one of the simulations performed, and the coordination of cell functions (Figure 4.4) corresponding to the organoids in Table 4.1. As shown, sophisticated coordination of cell functions is required to form an organoid with a lumen (Figure 4.4**b**). In our simulations, cells first proliferated, generating a closed volume. Then, some cells started secreting fluid, while other cells continued proliferating to create more luminal space to secrete and prevent fluid leakage. In organoid #1, 2 out of 16 cells (13%) started secreting after $84\,\mathrm{h}$, and then at $105\,\mathrm{h}$, the rest of the cells also contributed to lumen enlargement by producing and secreting fluid. In the case of organoid #2, 23 out of 32 cells (72%) were secreting at $105\,\mathrm{h}$, while 9 cells continued proliferating. After these 9 cells finished proliferating, the target lumen volume was already reached, so no more fluid secretion was needed. In organoid #3, cells started secreting earlier ($63\,\mathrm{h}$)

than in organoids #1, #2 and #4. In this situation, only two cells were secreting over approximately $8\,\mathrm{h}$ since the lumen volume of this organoid was small. In the case of organoid #4, 5 cells out of 32 cells (16%) started secreting at $105\,\mathrm{h}$, and $21\,\mathrm{h}$ later, the remaining cells that initially were proliferating were also secreting. Finally, cells entered a quiescent state in all four organoids when they achieved the final pattern.

To demonstrate the robustness of the NN to properly coordinate cell functions, we performed 10 simulations with the same NN for each organoid. Both the number of cells (Figure 4.4**c**) and the lumen volume (Figure 4.4**d**) obtained in the simulations consistently approximated the target data. The fitness of the simulations was evaluated similarly to the fitness function used for the training of the NN but neglected the penalization of the eccentricity, since we were not aiming to approximate the eccentricity of the *in vitro* experiments, as it was not used as input data (Figure 4.4**e**). The minimum median value of the fitness was 0.978 for organoid #2. The median values for organoids #1, #3, and #4 were all above 0.983.

Moreover, a sensitivity analysis was performed to investigate the influence of the main parameters of the model on the coordination of cells. For that purpose, organoid #1 was simulated in situations where the cell cycle time was decreased and increased by 20% from $30\,\mathrm{h}$ to $24\,\mathrm{h}$ and $36\,\mathrm{h}$ respectively (see Figure A.8 in *Sensitivity analysis* in *Appendix* A.2.16). Originally, the cells finished proliferating after $105\,\mathrm{h}$. A cell cycle time reduction makes cells proliferate faster; therefore, cells finished proliferating after $84\,\mathrm{h}$. Conversely, cells proliferate slower when the cell cycle time increases, thus requiring approximately $125\,\mathrm{h}$ to finish proliferation. Additionally, a similar sensitivity analysis for the fluid production time (Δt_{exo}) was performed (see Figure A.9 in *Sensitivity analysis* in *Appendix* A.2.16). When the fluid production time was $15\,\mathrm{min}$, cells secreted during $23\,\mathrm{h}$. A reduction of the fluid production time to 6 min caused cells to secrete during 21 h, only two hours less than in the previous case. However, in this case, only two cells were secreting, since they were able to produce more fluid over the same time period. Finally, when the fluid production time increased to $60\,\mathrm{h}$, cells were secreting over $34\,\mathrm{h}$. Although the variation in these parameters produced changes in the temporal evolution of the process, neither the variation in the cell cycle time nor the differences in the fluid production time altered the coordination of cell functions. Therefore, there might exist a specific organization of cell functions independent of cell type.

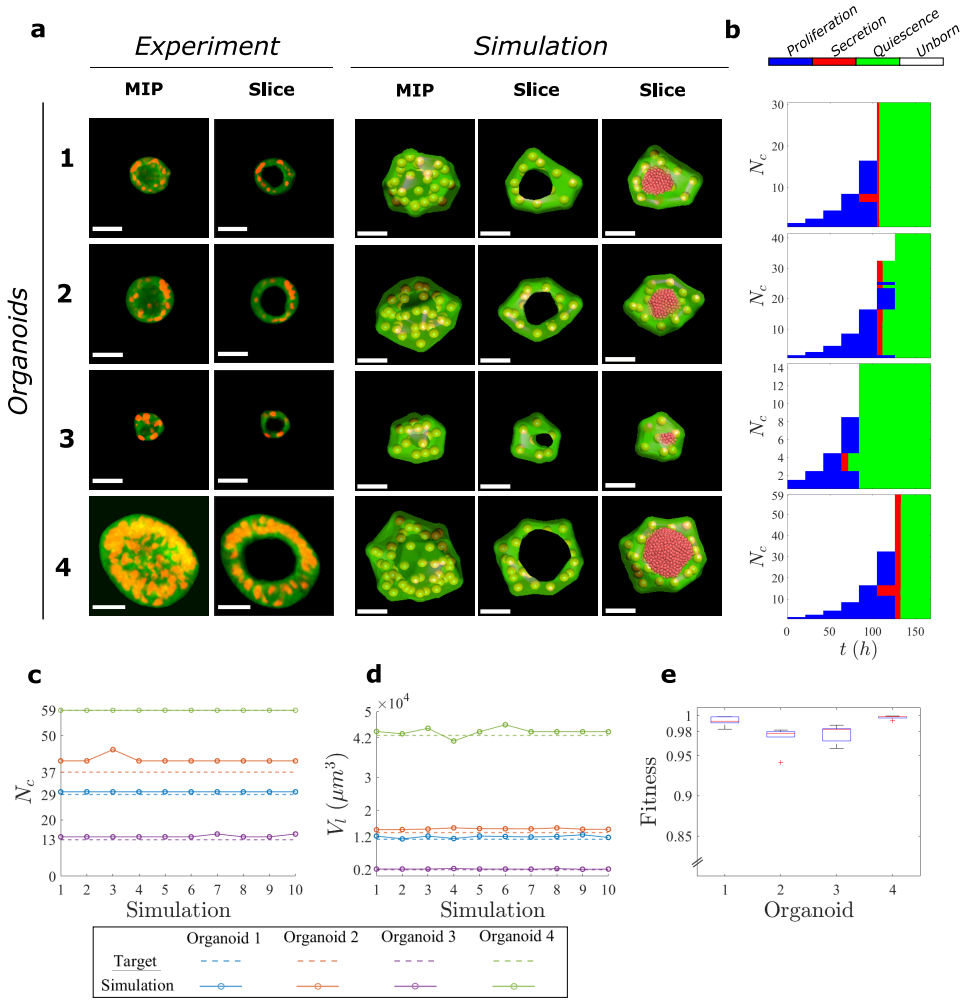

Figure 4.4: ***Simulation of experimental pancreatic tumor organoids with lumen.*** ***a*** *Image of in vitro organoids and a slice view. Maximum Intensity Projection (MIP). Snapshots of the full view, a slice of the simulated organoids, and a slice of the simulated organoids with the lumen particles representation. The spheres represent the nuclei of cells with radius $R_c/2$ and the green hull is an estimation of the cell membrane through alpha shapes of the cells with $\alpha = 2R_c$. The lumen fluid is represented in black (fourth column) through the alpha shapes of particles with the smallest alpha that produces an alpha shape enclosing all of the particles. All scale bars are 30 μm.* ***b*** *Coordination of cell functions, in which blue represents proliferation, red represents secretion, green represents quiescence, and white represents an unborn cell.* ***c*** *Target number of cells (N_c^t) and the number of cells (N_c) in the 10 simulations for each organoid.* ***d*** *Target lumen volume (V_l^t) and lumen volume (V_l) in the 10 simulations for each organoid.* ***e*** *Boxplot of the fitness value of the 10 simulations for each organoid.*

Finally, an analysis of the effect of the noise on the parameters was performed. Accordingly, the cell cycle time and fluid production time parameters varied ±20%. Thus, every time a cell is set to perform either proliferation or fluid secretion, a random variation of the parameter for that individual cell is made between [-20, +20]% of the value from Table A.2 and assigned for its cell cycle time or fluid production time. To determine how this variation could affect the results, we simulated an intermediate cystic organoid (organoid #2). Thus, we performed 10 simulations of this organoid to determine the effect of the fluctuation of the parameter. We found that the coordination of cell functions when the parameters are variable is similar to the case without any variability (see Figure A.10**a** and Figure A.10**b** in *Random effect analysis* in *Appendix* A.2.17). However, there is a desynchronization between cells in terms of cell functions caused by the variability of the duration of their functions. To better compare the coordination of the cell functions, we represented the evolution of the normalized number of proliferating cells, secreting cells, and quiescent cells over time (Figure A.10**c**). In the three cases, the evolution of the functions performed by the cells in the case with variable parameters was aligned with the case without variation. Despite the introduction of this noise, the median value of the fitness of the ten simulations was 0.9956, and the minimum was 0.9537 (Figure A.10**d**).

4.3.3 Application of the framework to solid tumor organoids

To illustrate the possibility of applying this framework to other morphogenetic patterns, we applied our methodology to simulate the formation of solid tumor organoids (further details of the methodology can be found in Figure A.11 in *Application of the framework to solid tumor organoids* in *Appendix* A.2.18). These solid tumor organoids consist of dense spherical aggregates of cells. To simulate the formation of this pattern, we considered two cell functions: proliferation and quiescence. Furthermore, we consider cell-cell mechanical interactions based on pairwise potential functions and the interaction of cells with the ECM by means of a friction coefficient. In the data-driven part of the model, we employed the target number of cells (N_c^t) as the metric that defines the size of the organoid and fixes the objective of the simulation. We also extracted some simulation metrics, which are the total number of cells, the number of proliferating cells, and the number of quiescent cells. The NN consists of an input layer of three nodes, a hidden layer of two nodes, and an output layer of two nodes, which represents either proliferation or quiescent states of the cell. Therefore, the NN coordinates both processes to achieve the target size of the organoid.

4.3.4 Simulation of pancreatic solid tumor organoids

We simulated the corresponding experiments of 3D cultures of PM12500-GFP tumor cells grown also in a biomimetic matrix (Matrigel $4\,\mathrm{mg/ml}$). This cell line is

characterized by the formation of organoids composed of aggregates of cells that do not form a lumen. Here, real 3D organoids of PM12500-GFP tumor cells were also generated and incubated for 5 days. Then, their morphology was segmented and quantified from multiphoton microscopy images of the grown organoids. In this case, we obtained the number of cells to quantify the size of the organoids. Different-size organoids were chosen again to prove the adaptive response of the computational-based methodology to the data. In particular, we used one small sample (organoid #1), one intermediate sample (organoid #2), and two large samples (organoids #3 and #4) (Table 4.2). The simulations were initiated with one cell and run for 7 days to prove that the model is able to reach steady-state patterns. The parameters of the simulations can be found in Table A.3 in *Simulation parameters for pancreatic solid tumor organoids* in *Appendix* A.2.19

We show images of the experimental organoids, the simulation of the organoids, and the coordination of cell functions (Figure 4.5), which represent the organoids in Table 4.2. The simulated organoids matched the target number of cells of the experimental data by Day 5. Here, precise coordination between proliferative and quiescent cells is needed, as the organoids approximate the target number of cells (Figure 4.5**b**). In organoid #1, after 67 h, 7 cells continued proliferating, and 9 cells entered quiescence, reaching the target number of cells at 84 h. In organoid #2, most of the cells (93.75%) became quiescent at 84 h, while 2 cells continued proliferating to finally form the target pattern at 101 h. In organoid #3. Twenty-two cells (68.75%) were proliferating after 84 h to increase the number of cells up to 53 at 101 h. Organoid #4 was the case in which more cells continued proliferating after 84 h, 28 out of 32 (87.5%).

Table 4.2: ***Experiment and simulation results of pancreatic solid tumor organoids.*** *N_c^t is the number of target cells.*

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

4.3.5 Application of the framework to the evolution of solid organoids

To demonstrate not only the application of the methodology to other patterns but also the possibility to mimic the temporal evolution of patterns, we show an extrapolation of our methodology to simulate the evolution of solid organoids without lumen, which consists of dense spherical aggregates of cells (further de-

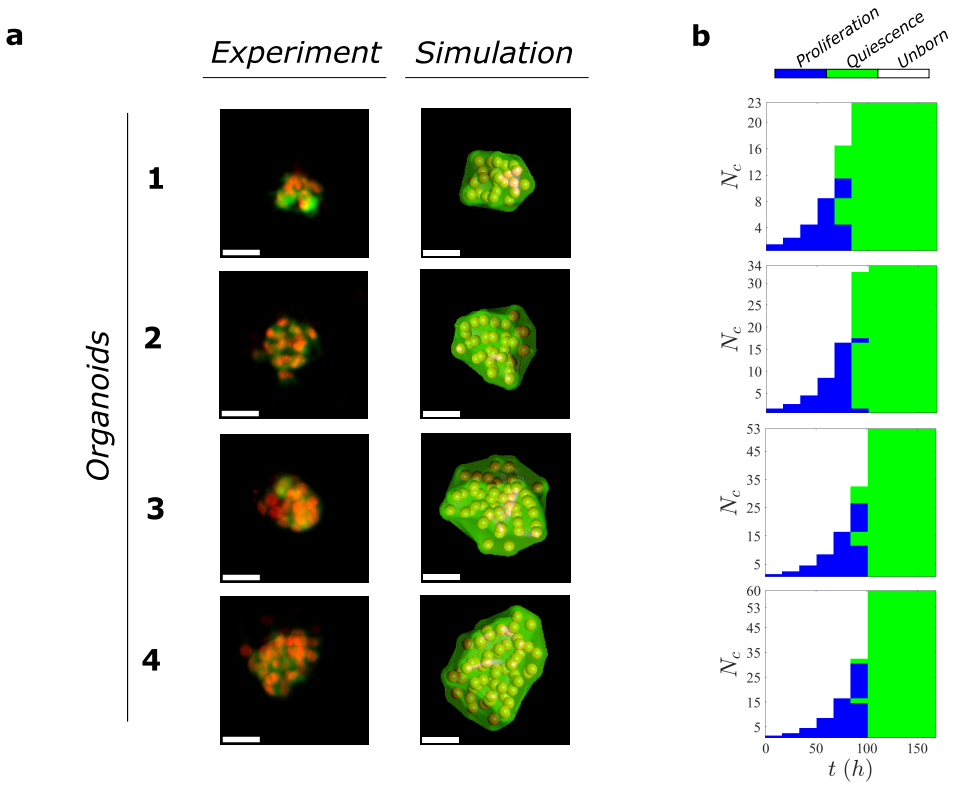

Figure 4.5: ***Simulation of experimental solid tumor organoids.*** ***a****. Image of in vitro organoids. Snapshots of the simulated organoids. The spheres represent the nuclei of cells with radius $R_c/2$ and the green hull is an estimation of the cell membrane through alpha shapes of the cells with $\alpha = 2R_c$. All scale bars are 30 μm.* ***b*** *Coordination of cell functions, in which blue represents proliferation and green represents quiescence.*

tails of the methodology can be found in Figure A.12 in *Application of the framework to the evolution of solid organoids* in *Appendix* A.2.20). With this methodology, we aim to reproduce the growth of solid organoids so that the evolution of this organoid matches the target evolution. To do that, target data is specified at different time points ($\boldsymbol{t}^t$). In this case, the target data consisted of the number of cells ($\boldsymbol{N}_c^t$). In the agent-based framework, we consider cell mechanical interactions and the cell interactions with the ECM. Moreover, cells can either proliferate or stay in a quiescent state during a period of time (Δt_q). When a cell finishes the cell function that it is performing, the NN decides the next cell function that the cell has to perform. Thus, we extract some simulation metrics (number of cells, number of proliferating cells, number of quiescent cells, and the simulation time) that characterizes the state of the simulation. Then, we obtain from the target time ($\boldsymbol{t}^t$) the target time immediately superior to the simulation time (t) and the

number of cells ($\boldsymbol{N}_c^t$). These data are transferred to the input layer of the NN in the following manner: the first node receives the comparison between the target number of cells and the number of cells of the simulation ($(\boldsymbol{N}_c^t - \boldsymbol{N}_c)/\boldsymbol{N}_c^t$); the second node receives the normalized number of quiescent cells, and the third node receives the normalized number of proliferating cells. Finally, the NN determines the next function to perform.

4.3.6 Simulation of the evolution of solid organoids

In order to show the possibilities of the methodology, we consider three synthetic cases of solid organoids with different temporal evolutions (Table 4.3). Organoid A represents a proliferative organoid. Case B represents an organoid that proliferates, then remains quiescence between day 3 and day 5, and proliferates again between day 5 and 7. Finally, case C corresponds to a proliferative organoid with a higher proliferation rate than case A. The parameters used for the simulation are collected in Table A.4 in *Simulation parameters for the evolution of solid organoids* in *Appendix* A.2.21

The simulations matched the evolution of the number of cells in the three cases (Figure 4.6), adapting to the changes in input data over time. In organoid A, as cells proliferate, they become quiescent while some cells still proliferate to match the target values (Figure 4.6**a**). In the case of the organoid B, cells proliferate until day 3, and between day three and five they enter into a quiescent state to finally proliferate after day 5 (Figure 4.6**b**). In organoid C, cells have a higher proliferative capacity than organoids A and B, so cells tend to proliferate against quiescence (Figure 4.6**c**). Also, we simulated the evolution of a larger solid tumor organoid to demonstrate that the model works for any size. In this case, the target number of cells was established to 100 cells on day 7 and 500 cells on day 10 (Table A.5 in *Simulation of the evolution of larger solid organoids* in *Appendix* A.2.22). The simulation matched the evolution of the number of cells (Figure 4.6). Thus, this methodology is also able to mimic the temporal evolution of morphogenetic patterns of any size as long as the final result is feasible to achieve.

Table 4.3: ***Target data for the evolution of solid organoids.***

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

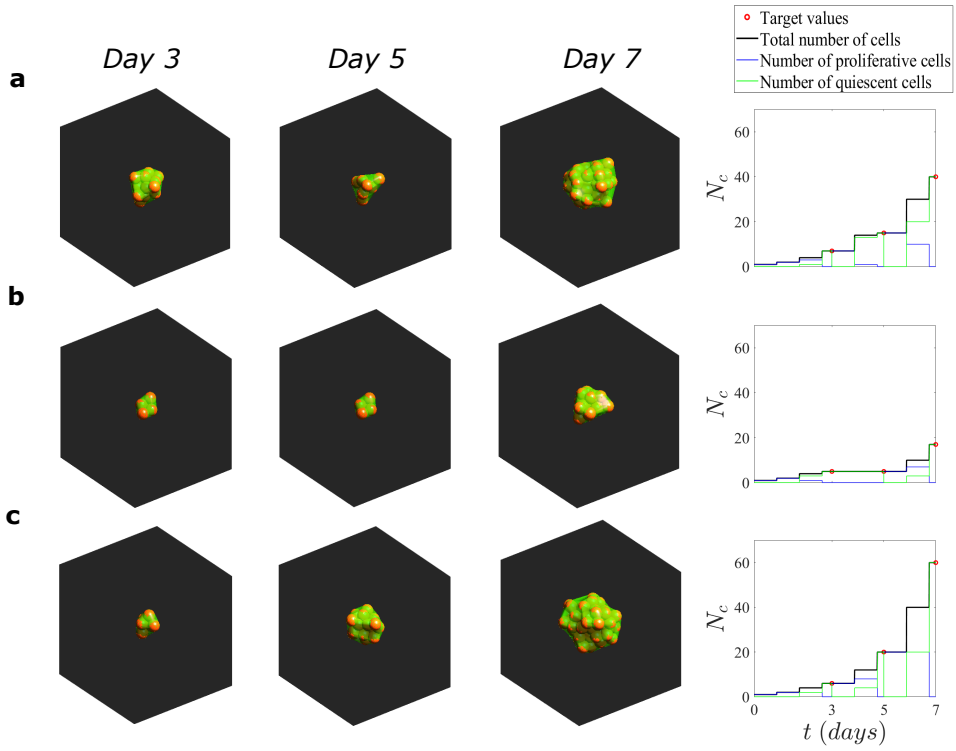

Figure 4.6: ***Simulation of the evolution of solid organoids.*** *Snapshots at days 3, 5, and 7 of the temporal evolution of the solid organoids inserted in a cube of 200 μm side and the evolution of their number of cells.* ***a*** *Organoid A.* ***b.*** *Organoid B.* ***c*** *Organoid C. The spheres represent the full cell volume, and the green hull is an estimation of the cell membrane through alpha shapes of the cells with $\alpha = 2R_c$.*

4.4 Discussion

We present a novel hybrid physics-based and data-driven approach that combines agent-based modeling and DL for the simulation of developmental biology at the cellular level. Thus, we hypothesized that the different patterns that cells form are consequences of the distinct spatial and temporal coordination of their cellular functions. The following question is then how cells determine which cellular function to perform, since if we are able to replicate the decision-making of cells, we can reproduce the coordination of cell functions to achieve any pattern. Thus, we built a novel computational framework in which a NN manages the decision-making of biological cell functions within a physics-based model to achieve an objective morphogenetic pattern. To accomplish the simulation of morphogenetic patterns, we do not fix the relationships between cell functions beforehand; instead, we consider cell functions as independent processes that

cannot occur simultaneously, and the DL algorithm decides for each cell which cellular function they will be performing during the developmental process, properly coordinating all cellular functions for pattern formation. Thus, this artificial intelligence-based method learns the intrinsic mechanisms of the morphogenetic process and, therefore, is capable of reproducing similar morphogenetic patterns with different dynamical processes without optimizing the parameters *ad hoc* just by changing the target data about the pattern transferred to the input of the NN. Therefore, with this approach, we can understand how cells must behave both spatially and temporarily to form different patterns.

In this work, we showed how to integrate a physics-based model with a DL algorithm to simulate morphogenetic patterns in 3D organoids. This modular framework offers a wide flexibility to change and adapt the physics-based and data-driven models involved. On the one hand, in the physics-based part, agent-based modeling offers a broad range of types of models: from center-based models to deformable or vertex models [286]. This allows us to take advantage of a chosen model to better adapt to a specific application. On the other hand, regarding the data-driven part of the framework, different types of NNs and activation functions can be considered. Indeed, the architecture of the NN could be simplified to reduce the computational cost of the learning process or incremented to unravel complex features, trying to find the optimal conditions for each case depending on the type of morphogenetic pattern, the quantity of input data, and the different cell states considered [304]. Additionally, there exist different learning algorithms that can be integrated into this approach (genetic algorithm, gradient descent, conjugate gradient, Levenberg-Marquardt algorithm, among others [305]) for different computational speeds and memory requirements given a set number of parameters of the chosen NN. Therefore, our framework is scalable with different types of models and permits modification to improve the computational cost and to better capture the specific biological process.

To evaluate the potential of the approach, we showed two applications mimicking *in vitro* data of the final pattern of organoids and one application reproducing synthetic time-lapse data. Thus, one advantage of this approach is that it requires little information about the morphogenetic processes, being able to undertake pattern formation using just final morphological information about the pattern extracted visually. In pattern formation, generally, the information is obtained from the observed final pattern (in general image data), and the efforts move to unravel how it was formed. In particular, time-lapse experiments could alter the viability of cell cultures by inducing cellular damage (e.g., imaging-induced phototoxicity, leading to cell cycle arrest or cell death [306]). However, it is possible not only to obtain the coordination from the data of the final pattern but also to reproduce the temporal evolution of patterns. This permits us to mimic the exact evolution of the pattern formation when having time-lapse data, since it matches the known temporal constraints.

Regarding the simulation of *in vitro* data of lumen morphogenesis in PDAC organoids, physiologically, the generation of a cystic organoid requires the orchestration of proliferation, fluid secretion, and quiescence. On the one hand, if cells secrete fluid too early, the preapical patch, which is the closed volume necessary to secrete fluid, will not be formed, resulting in fluid leakage. On the other hand, if cells secrete too much fluid, the hydrostatic pressure excessively increases, and the lumen opens. Additionally, if cells proliferate too much, they will not fit in the monolayer around the lumen, and cystic structures will not be formed. In contrast, if cells do not proliferate, they will not create enough inner space for the fluid, resulting in leakage. Only if cell functions are coordinated in a sophisticated manner will cystic structures form. In particular, first cells must proliferate to create an inner volume and then secrete fluid and while continuing to proliferate to create more inner space as the fluid increases, before finally entering into a quiescent state. Therefore, the success of lumen generation depends on the proper coordination between cell proliferation, cell secretion, and quiescence. In fact, what the NN learns is that if it generates too much pressure through fluid secretion with an insufficient number of cells, it leaks. Additionally, it learns that cells must first proliferate to generate the organoid and, as the organoid reaches its final size, enter into a quiescent state. Despite the great biological variability in terms of the final size of *in vitro* experiments (see Figure A.14 in *Cell count analysis of the in vitro cystic organoids* in *Appendix* A.2.23), we were able to obtain the developmental process to finally predict their size in each case. This quantitative validation with *in vitro* data proves that this approach does not produce *ad hoc* solutions and confirms that the different coordination of processes determines the final pattern. This is very suitable to simulate biological systems, which display variability and fluctuations around a solution. In the case of solid tumor organoids, the NN coordinates proliferation and quiescence to achieve the final size of the organoid. This regulation of the dynamics of the population of cells is a helpful tool that could be used in many types of simulations. This permits the integration of data into simulations to determine cell proliferation, which is sometimes a difficult task, especially when the dynamics of the population of cells change over time. Therefore, our model is designed to predict the order in which cells behave to achieve a specific pattern.

Moreover, from a biological perspective, we found in our simulations that some cells are predisposed to perform a specific function over others. This is the case for the initial cell or first cells, which tend to proliferate more than daughter cells (Figure 4.4**b**). These initial precursor cells might have a higher proliferative potential, and as they divide, the daughter cells are more specialized with a higher differentiation grade. Inside an organoid, there exists heterogeneity of cellular functions, and we hypothesize that the cellular functions that cells perform in a morphogenetic process are not governed by the maturation of cells but by the stimuli to which they are exposed when they determine which function to

perform. This implies that some cells are more specialized in division and others in secretion within the organoid. We also found that the final size of the organoid depends on the morphological characteristics of the organoid when the lumen is initiated. In our simulations, the initiation of the lumen, determined by the beginning of the fluid secretion, was different for each case, being earlier the smaller the final size of the organoid was (Figure 4.4**b**). We were able to establish that, initially, cells have a less differentiated phenotype and have a higher proliferative capacity. Then, as they proliferate, the daughter cells adopt a higher differentiation grade and specialize into secretory cells due to the stimuli that they are exposed when they are born. Thus, when they start secreting, the proliferative capacity of the organoid is reduced, so the final size of the organoid is limited by its size at the lumen initiation time. Therefore, we hypothesize that the proliferation capacity of cells may be hindered by the initiation of lumen formation and that the final size can be predicted at the initiation of lumen formation. These hypotheses could be validated by performing cell culture immunostaining with a mitotic cell marker to determine the number of proliferative cells in the biological system [307].

Another potential of this approach is that once the NN has learned how to orchestrate cells to form the morphogenetic pattern, the parameters of the physics-based model can be modified to study the morphogenetic process under different scenarios. This allows the simulation of morphogenesis with different ECM properties, cell cycle times, or even organ-specific morphogenesis. In the case of morphogenesis, similar patterns are found in many parts of the organism, but the dynamic process differs in each organ since cell cycle time phases vary considerably between different types of cells and secretory cells may require different amounts of time to produce fluid. This implies that similar morphogenetic patterns can result in different sizes and require different lengths of time to achieve the final shape. However, although morphogenesis depends on cell type and the dynamics of the process vary, the requirements of the morphogenetic process are conserved among similar patterns. To this aim, we conducted a sensitivity analysis (Figure A.8 and Figure A.9) varying the cell cycle time and fluid production time for the pancreatic cystic tumor organoids. This variation in these two parameters represents either the fluctuation of the behavior between the same type of cells or other types of cells that can also form cystic organoids. We also performed an analysis of the influence of the introduction of noise in the cell cycle and fluid secretion parameters (Figure A.10), showing that the model is robust in forming the patterns and that the NN can orchestrate cells despite the fluctuations of these parameters. With these analyses, we showed that in addition to the alteration of the dynamical processes of cells, the coordination to form the same pattern with the same morphological features (in this case, the number of cells and lumen volume) is conserved.

In the applications shown in this work, some simplifications have been as-

sumed. First, we employed a center-based lattice-free model to simulate the formation and growth of organoids, which entails some simplifications. The most important one is that cells are assumed to be nondeformable spheres, and their shape was not represented accurately. This simplification allowed us to approximate the size of the organoids through their number of nuclei with a reduced computational cost. Since the deformation of cells is not simulated, there are differences between the volume of the simulated organoids and the experiments (Figure 4.5**a**). However, this simplification does not influence the coordination that cells must follow to form morphogenetic patterns. Here, we did not aim to reproduce the exact shape and volume of the organoids; instead, we focused on showing the potential of the framework to unravel the orchestration of cell functions from a morphogenetic pattern, which is independent of the cell shape and volume. Second, we represented the lumen fluid through a particle-based model comparable to other works [240, 303]. Although this representation is a simplification, we showed that our approach accurately reproduces the expected velocity and shear stress profiles for a Newtonian fluid, similar to water (see *Characterization of the lumen fluid* in *Appendix* A.2.13). This confirms the capability of our method to model real fluids, serving as a minimal model that aims to recapitulate certain features such as mass conservation and momentum exchange. However, it does not rigorously reproduce bulk-scale physical properties. Furthermore, our approach enables us to simulate the interaction between cells and the fluid during a morphogenetic process, in which the lumen initiates *de novo* and grows through a fluid deposition process by cells, thus evolving the interface between cells and fluid over time. Additionally, we investigated the influence of particle radius on the rheological properties of the fluid (further details are included in *Particle radius influence on the rheological properties of the fluid* in *Appendix* A.2.14). We have justified the choice of the particle radius and demonstrated that the rheological properties of the fluid play a crucial role in cystic organoid morphogenesis (see *Particle radius influence on fitness* in *Appendix* A.2.15).

Another limitation of the approach relates to the selection of cell functions and the question of whether the hypothesized functions are sufficient to generate the specific pattern. In our model, the decision-making possibilities are limited to the cell functions considered, which in this work were proliferation, secretion, and quiescence. Indeed, there are many cellular functions and processes that could be simulated. However, including more cell functions does not guarantee that all of them might be activated to reach a specific morphogenetic pattern. This allows us to underestimate them and reduce the complexity of the NN and the number of outputs. Furthermore, in this first work, we created this data-driven approach with a NN that receives the quantitative data as inputs. However, the NN aims to reproduce the cell signaling through which cells sense, receive, and transduce external factors from their microenvironment (such as pressure, chemical signals, nutrients, etc.), or even internal factors (such as cell damage and cell

deformation). Therefore, this approach permits the introduction of those factors as inputs of the NN to enrich cell signaling and response. The final concern is related to finding the optimal NN architecture for each specific application. Although some rules help to select the number of nodes and hidden layers, trial and error are commonly used. The NN architecture determines the capacity to solve the problem and must be appropriate for the application. Moreover, the choice of the fitness function is fundamental to make the NN learn. The fitness function evaluates the NN performance depending on how close the simulated pattern is compared to the objective pattern. Therefore, this interplay between the NN and the fitness function regulates the learning of the morphogenetic pattern. Despite the simplifications made in the applications of this framework, it allowed us to reproduce the morphogenetic patterns of organoids, understating how cells orchestrate their cellular functions to achieve those patterns.

4.5 Conclusions

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

Chapter 5

Prostate cancer

Those who can imagine anything, can create the impossible.

– ***Alan Turing***

Contents

This chapter is based on:

Camacho-Gomez, D., Borau, C., Garcia-Aznar, J. M., Gomez-Benito, M. J., Girolami, M., & Perez, M. A. (2024). *A physics-informed machine learning digital twin framework for prostate cancer: predicting tumor growth via serum PSA blood tests.* Submitted.

Abstract

Prostate cancer is a significant global health concern, with millions of new cases diagnosed annually. Existing monitoring methods, reliant on Prostate-Specific Antigen (PSA) measurements in blood tests, exhibit limited precision, often failing to detect tumor progression especially when PSA levels remain stable. Thus, this work presents a computational framework integrating physics-based modeling and ML in digital twins to predict tumor prognosis using PSA blood tests and image-based biomarkers. Patient-specific tumor evolution is simulated in the digital twin using a physics-based model that considers PSA secretion and flux from tissue to blood, depending on the tumor vascular distribution. Furthermore, this physics-based model is enhanced by a deep learning model that spatially and temporally regulates tumor growth dynamics through the patient's PSA blood tests and three-dimensional spatial interactions of physiological variables. We test and validate our framework by accurately predicting tumor progression in real patients over the years. Our results reveal scenarios of hidden tumor growth, where serum PSA levels may not rise despite tumor progression. Therefore, our framework provides a promising tool in clinical prostate cancer monitoring, predicting tumor growth and aggressiveness, reducing the uncertainty of PSA blood tests, and minimizing the need for performing magnetic resonance imaging scans.

5.1 Introduction

Prostate Cancer (PCa) is one of the most prevalent forms of cancer that affect men [308]. It is estimated that prostate cancer accounted for 1.4 million new cases globally and resulted in more than 370,000 deaths in 2020 alone [309]. Prostate cancer is characterized by the uncontrolled growth and division of luminal cells within the prostate gland. Over time, these cancerous cells can invade nearby tissues and potentially spread to other parts of the body, mainly to the bones, lymph nodes, liver, and lungs [310–313] in a process called metastasis [314]. Therefore, predicting the evolution of prostate cancer is essential for timely detection of growth and halting the expansion of the disease.

The diagnosis of prostate cancer typically relies on the Prostate Imaging Reporting and Data System (PI-RADS), which assigns scores on a five-point scale to lesions observed in MRI sequences [315], and the Gleason score, which assesses the differentiation of cells in biopsy samples [316], serving as an indicator of tumor prognosis [317]. Subsequently, prostate cancer monitoring is commonly based on the Prostate-Specific Antigen (PSA) biomarker [318]. The PSA is a protein produced by both normal and cancerous cells within the prostate gland. Its main function is to liquefy semen, aiding in the mobility and transportation of sperm during ejaculation [319]. PSA levels in the blood have been extensively

employed as a biomarker for both the detection and ongoing monitoring of prostate cancer. Elevated levels of PSA may indicate various prostate conditions, including prostate cancer [320]. While the exact reason for increased PSA levels in prostate cancer is not fully understood, it is believed that cancerous cells can disrupt the normal architecture of the prostate gland, leading to increased production and leakage of PSA into the bloodstream. Consequently, measuring PSA levels in the blood can help in the early detection of prostate cancer and monitoring its progression over time. Yet, tumor progression often occurs without a significant rise in PSA levels, thereby obscuring the prognosis of the tumor. The limited specificity and sensitivity of the PSA as biomarker [321–323] are linked with poor diagnosis, as well as treatment and screening-related adverse effects [324]. Consequently, it is essential to deepen our understanding of the relationship between PSA levels and tumor development.

Significant improvements have been made in the comprehension of prostate cancer through computational models. Jain et al. [325] studied through a system of differential equations the progression of prostate cancer under continuous and intermittent antiandrogen treatment regimes, emphasizing the heterogeneous nature of the disease and adjusting personalized parameters to the "average" patients. Colli et al. [326] analyzed prostate cancer growth with chemotherapy and antiangiogenic therapy effects, suggesting that, while cytotoxic drugs may suffice to treat mild tumors, the combination of cytotoxic action with the reduction of intratumoral nutrient availability is essential to kill the aggressive tumor. Lorenzo et al. [327] formulated a model for the PSA dynamics after radiotherapy treatment, showing good agreement with the patients' PSA. However, these works do not incorporate the anatomical and physiological characteristics of the prostate gland itself, nor do they fully capture personalized aspects of prostate cancer. By considering the specific features of the prostate through image-based biomarkers, such as its shape, vascularization, or the spatial distribution of cancer cells, computational models could provide more detailed and accurate insights into tumor growth dynamics and treatment response. This enhanced precision would allow for more personalized and effective treatment strategies that would ultimately benefit the patients.

Advancing along these lines, digital twins are emerging as a promising approach. Digital twins [328] are virtual representations of physical objects, processes, or systems, that mimic the behavior, characteristics, and functionalities of their physical counterparts. In the context of biology, digital twins offer a means to integrate the anatomical and physiological characteristics of patients' tissues, together with medical records, and other relevant information into computational models [329–331]. Thus, digital twins are a powerful tool for providing a more accurate and personalized understanding of tumor growth and treatment response in prostate cancer, allowing for a more comprehensive representation of the disease. Various authors have made efforts in this direction. For exam-

ple, Boubaker et al. [332] performed a finite element simulation of the interaction of the prostate and the surrounding organs to predict their deformation to improve radiotherapy delivery. Lorenzo et al. [129] simulated a patient's specific prostate generated from computed tomography (CT) images. Lorenzo et al. [333] extended the model to include the mechanical interaction of prostate cancer and benign prostatic hyperplasia. Despite significant advances, the aforementioned works still present several limitations. First, using only CT imaging to determine prostate geometry falls short in precisely identifying prostate cancer and pinpointing its location. Second, these models lack the capability to incorporate detailed information at the cellular level within the prostate tissue, such as cell density or cellularity, crucial for accurately simulating tumor growth. Moreover, the lack of perfusion data eliminates the possibility of accounting for the spatial arrangement of the vascular network. Addressing these limitations holds the promise of enhancing personalized diagnosis, treatment planning, and disease management for prostate cancer patients, which could significantly contribute to improved patient outcomes.

In this work, we present a physics-informed machine-learning framework to predict the prognosis of prostate cancer using digital twins. Initially, we generate patient-specific digital twins of the prostate from T2-Weighted Image (T2WI), incorporating cellular-level data such as cellularity, vascularization, and tumor location. These digital twins are then introduced into the computational framework to obtain the tumor's progresion over time. In this physics-based model of prostate cancer, we consider the production of tissue PSA by cancer cells and propose a mathematical model to calculate serum PSA levels based on the prostate's vascularization. Consequently, the rise of serum PSA levels is caused by tumor cell proliferation. To determine the tumor's evolution from patient data, we incorporate a patient-agnostic deep-learning method that regulates the dynamics of tumor growth in the physics-based model, considering the spatial interaction of physiological variables and other relevant factors. We calibrate the physics-based model to reflect patient-specific tumor growth dynamics by integrating one additional follow-up MRI data. This process enables the determination of personalized physical parameters, enhancing the accuracy of tumor prognoses. Subsequent tumor growth predictions are made using only serum PSA and non-MRI data from follow-up assessments. We validate our approach with real patient data sourced from the La Fe Hospital (HULAFE) database. Our findings demonstrate the efficacy of our methodology in accurately predicting the evolution of prostate cancer, based on patient follow-ups using serum PSA data, thereby improving the interpretability of PSA monitoring and advancing toward a potential method to serve as a surrogate for MRI.

5.2 Materials and methods

5.2.1 Physics-informed machine learning digital twin framework for prostate cancer

We present a computational framework to predict the evolution of prostate cancer in digital twins, combining a physics-based model for tumor progression with a machine-learning model that regulates its growth dynamics (Figure 5.1). This patient-specific digital twin of the prostate is generated from the T2WI sequences on MRI, which include Diffusion Weighted (DW) and Dynamic Contrast Enhanced (DCE) image sequences, as well as the tumor segmented by expert radiologists from HULAFE, all common procedures in the evaluation and diagnosis of prostate cancer (Figure 5.1**a**). Using these MRI sequences, a 3D voxelized geometry is generated, containing the patient's data such as the cellularity obtained from the Apparent Diffusion Coefficient (ADC) [334] and tumor location. Also, the spatial distribution of the vascularization is considered through the k_{trans}, derived from the DCE-MRI [335, 336]. Also, the spatial distribution of vascularization is considered through k_{trans}, which represents the volume transfer constant of contrast agent between the blood plasma and extravascular extracellular space in DCE-MRI [335, 336], providing crucial information about tissue perfusion and permeability.

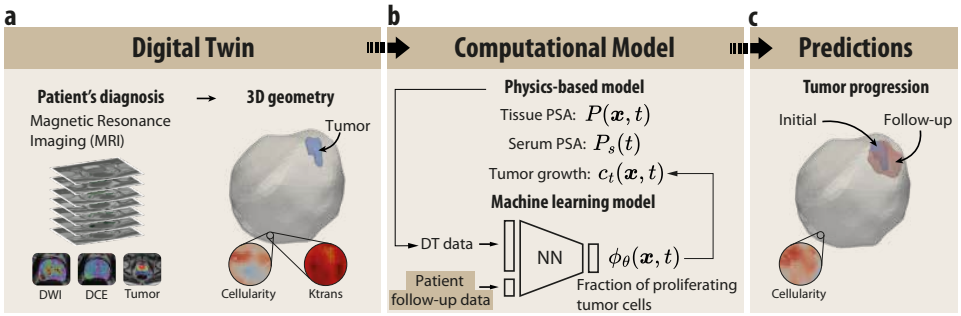

Figure 5.1: ***Prostate cancer digital twin framework.*** ***a*** *The digital twin geometry is reconstructed from the T2-weighted image sequences on magnetic resonance imaging, including the spatial distributions of cellularity, k_{trans}, and the tumor mask.* ***b*** *The computational model consists of two main parts. First, a physics-based model simulates the evolution of tissue PSA $P(\boldsymbol{x},t)$, serum PSA $P_s(t)$, and tumor growth $c_t(\boldsymbol{x},t)$. Second, a ML model based on NN determines the fraction of proliferating tumor cells $\phi_\theta(\boldsymbol{x},t)$ in the tumor growth equation based on the data from the digital twin and the patient follow-up serum PSA blood test.* ***c*** *The outcome of the model is the patient's tumor evolution from diagnosis to the follow-up date.*

We propose a physics-informed ML model to simulate the evolution of prostate cancer (Figure 5.1**b**). The purpose of the physics-based model is to repre-

sent the biological processes involved in prostate cancer within the digital twin framework to predict the progression of the diagnosed tumor. In this model, we simulate the tissue PSA distribution $P(\boldsymbol{x}, t)$ as a consequence of the PSA leakage from cancer cells. Additionally, we incorporate the exchange of PSA between the tissue and the bloodstream. To achieve this, we consider the capillaries located in the tissue, represented through spatial maps of $k_{\text{trans}}(\boldsymbol{x})$. This exchange depends on the concentration difference between the tissue PSA distribution ($P(\boldsymbol{x}, t)$) and the serum PSA ($P_s(t)$), as well as the permeability of the capillaries. We also consider the natural decay of the tissue PSA. The serum PSA ($P_s(t)$) is obtained by integrating the flux between the bloodstream and the tissue, also considering its natural decay. Finally, we model the evolution of the concentration of tumor cells $c_t(\boldsymbol{x}, t)$, which represents tumor growth. This tumor cell concentration is responsible for the production of tissue PSA, and therefore, for the variations in the simulated serum PSA concentration. Hence, the dynamics of tumor growth cause the variations in serum PSA (see *Physics-based model for prostate cancer* in *Appendix* A.3.1 for further details of the physics-based model).

In this physics-based model, we integrate a ML model to determine the dynamics of tumor growth ($c_t(\boldsymbol{x}, t)$) based on the patient's serum PSA tests from their follow-ups. This ML model approximates, through a fully connected NN, the fraction of proliferating tumor cells $\phi_\theta(\boldsymbol{x}, t)$, which is incorporated into the equation describing tumor growth (see next section for further details). To accomplish this, the deep-learning model receives data from the physics-based model and the patient follow-up data based on serum PSA blood tests. Consequently, it regulates the tumor growth dynamics in the physics-based model so the simulated PSA reproduces the patient's serum PSA values accurately. Finally, we predict the tumor's evolution from the time of diagnosis to the follow-up date which is responsible for the observed serum PSA variations.

5.2.2 Machine learning model for tumor growth dynamics

We propose a deep-learning algorithm to determine tumor growth by regulating the fraction of proliferating tumor cells ($\phi_\theta(\boldsymbol{x}, t)$) in equation (A.18) based on the serum PSA of the patient's follow-up (Figure 5.2). To approximate this function that controls tumor growth, we construct an input matrix with data from the digital twin and the patient. From each voxel, we extract the cellularity $c(\boldsymbol{x}, t)$, the normalized tissue PSA $P(\boldsymbol{x}, t)/\max(P(\boldsymbol{x}, t))$, the normalized k_{trans} ($k_{\text{trans}}(\boldsymbol{x})/\max(k_{\text{trans}}(\boldsymbol{x}))$, tumor mask ($T_{\text{mask}}(\boldsymbol{x}, t)$), and the fraction of proliferating tumor cells ($\phi_\theta(\boldsymbol{x}, t)$). Subsequently, we collect this information from the adjacent voxels, adhering to a standardized sequence. Given the voxel-based geometry, each voxel is surrounded by up to 26 neighbors, except for edge voxels in which we insert -1 values to substitute for the missing voxel data. This method allows us to estimate the proportion of proliferating cells not solely through in-

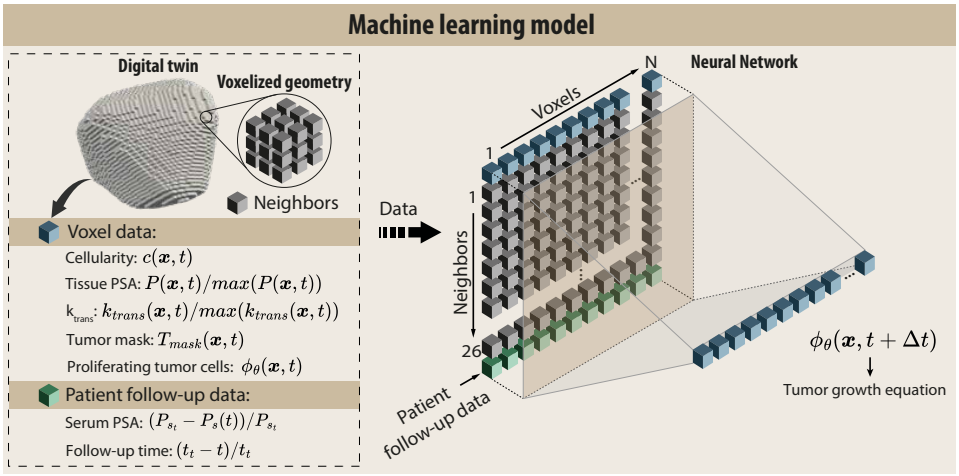

Figure 5.2: ***Machine learning model for tumor growth dynamics.*** *This model approximates the evolution of the fraction of proliferating tumor cells in each location ($\phi_\theta(\boldsymbol{x},t)$). To do that, it employs data from the digital twin and the patient's follow-up serum PSA test to construct an input matrix. The matrix is then processed through a NN to determine the fraction of proliferating tumor cells. Spatial interactions are considered by incorporating data from neighboring voxels. This process is repeated at every time step in the physics-based simulation.*

dividual voxel data but also by considering the context of their surrounding environment, thereby incorporating spatial dynamics into our analysis. Additionally, the voxelization of geometry standardizes voxel connections and the number of inputs, rendering the NN adaptable to the prostate geometry across different patients. This makes the NN agnostic of the digital twin's geometry, simplifying the application of the model without requiring retraining due to geometrical changes.

In addition, for each voxel, we concatenate the patient's data, consisting of the serum PSA obtained in the follow-up blood test (P_{s_t}) compared with the current simulated serum PSA at the time step when the NN is called. This data is then normalized as $(P_{s_t} - P_s(t))/P_{s_t}$. Similarly, we include the time at which the follow-up blood test was conducted (t_t), compared with the simulated time, and normalize it as $(t_t - t)/t_t$. Finally, this matrix is fed forward to the NN to obtain ($\phi_\theta(\boldsymbol{x}, t + \Delta t)$). This process is repeated at every time step in the physics-based simulation.

5.2.3 Calibration for patient-specific tumor growth dynamics

The calibration of the framework involves two phases. The initial phase entails training the patient-independent NN of the ML model to learn serum PSA dynamics. The second phase involves obtaining patient-specific parameters for the physics-based model for personalized tumor growth (Figure 5.3).

In the initial phase, the weights of the NN (θ) are obtained to learn serum PSA dynamics through adequate regulation of tumor growth (Figure 5.3**a**). It is important to note that this training step uses a single case of initial conditions (either synthetic or from an actual patient) which is subjected to a multi-objective simulation: reach a certain blood level of PSA (target serum PSA dataset ($\boldsymbol{P}_{s_t}$)) within a defined period (target times dataset ($\boldsymbol{t}_t$)). Then, we calculate the loss between the simulated serum PSA and the target serum PSA at the target times. This loss is then backpropagated to optimize the NN weights (see *Model calibration* in *Appendix* A.3.3). In this way, the NN learns the underlying relationships between PSA production and the spatial distribution of cellularity, k_{trans}, and tumor growth that result in the different evolutions of the serum PSA. This deep-learning model addresses the inverse problem of predicting tumor growth from serum PSA at a certain date. At this point, the physics-based parameters are not required to be patient-specific, so this step is performed only once during the model setup.

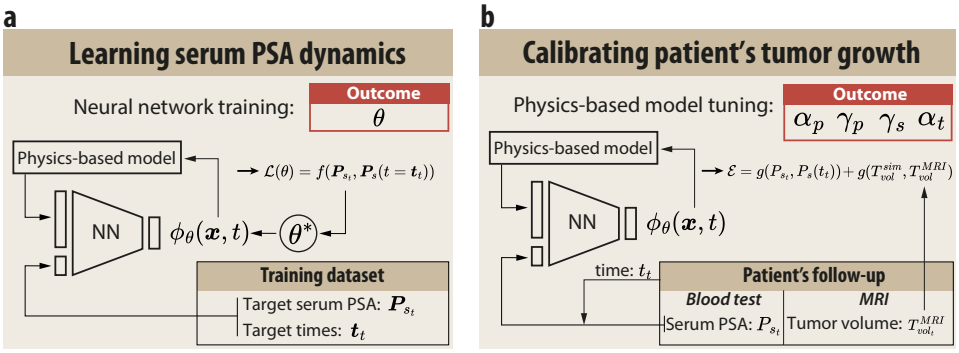

Figure 5.3: ***Calibration of the framework to reproduce patient-specific tumor growth.*** ***a*** *Training of the NN to learn to replicate serum PSA dynamics by controlling tumor growth. This involves conducting simulations with each serum PSA dataset, followed by minimizing the loss between the simulated and dataset serum PSA.* ***b*** *Calibration of the patient-specific parameters of the physics-based model for personalized predictions. The serum PSA is integrated into the computational model to predict tumor growth at the follow-up. Subsequently, the error between the predicted tumor volume and serum PSA and the clinical outcomes is minimized.*

Once the ML model is set up with the trained NN, the next phase involves obtaining the patient's specific parameters of the physics-based model to make personalized predictions from their blood test (Figure 5.3**b**). This ensures that the predicted tumor aligns with the patient's actual tumor when replicating serum PSA levels. For this purpose, we conduct a simulation incorporating the patient's follow-up data as targets for the NN, resulting in the personalized evolution of the tumor and serum PSA. Subsequently, we calculate the error between the predicted volume and serum PSA and the actual values obtained from the follow-up and minimize it by optimizing the parameters of the physics-based model (refer

to *Model calibration* in *Appendix* A.3.3). After determining the patient-specific parameters of the physics-based model with a single MRI follow-up, we can make further predictions of tumor growth using only subsequent serum PSA follow-ups.

5.3 Results

5.3.1 Unveiling patient-specific tumor growth

To demonstrate the potential of the framework for predicting patient-specific prostate cancer progression, we utilized data from two different anonymized patients from HULAFE, whom we will henceforth name Patient A and Patient B. Patient A was diagnosed at the age of 68, with a PI-RADS category 5 and a Gleason score of 3+3. Patient B, diagnosed at the age of 60, had a PI-RADS category 4. Patient B's biopsy revealed a Gleason score of 3+3 and a 2% concentration of tumor cells. Since the concentration of tumor cells was not measured in the biopsy of Patient A, we assumed it to be the same as that of Patient B.

We applied the calibration method outlined in Figure 5.3**b** to determine the patient-specific parameters of the physics-based model which replicate the serum PSA and tumor volume evolution from the diagnosis to the first follow-up (Figure 5.4 **a**). Notably, the obtained parameters outline substantial differences between the two patients. The tissue PSA production rate (α_p) is 49.05% lower in the case of Patient A compared to Patient B. Conversely, the tumor growth rate (α_t) is 124.07% higher in Patient A than in Patient B. This indicates that while Patient A's tumor secretes less PSA, it exhibits a higher growth rate. Regarding decay rates, Patient B exhibited higher decay rates both in tissue and serum, although our parametric analysis suggested that the serum PSA decay rate (γ_s) was the least influential parameter in the computational model.

We illustrate the tumor evolution of Patient A from diagnosis to follow-up (Figure 5.4**b**). The serum PSA level at diagnosis is $6.60\,\mathrm{ng/mL}$ with a corresponding tumor volume of $579\,\mathrm{mm}^3$. Subsequently, after 393 days, the clinical tumor volume increases to $862\,\mathrm{mm}^3$, accompanied by a rise in serum PSA to $7.40\,\mathrm{ng/mL}$. We reproduce this clinical evolution with our computational model, with a simulated serum PSA of $7.70\,\mathrm{ng/mL}$ and a tumor volume of $862\,\mathrm{mm}^3$, resulting in a relative error of 2.82% compared to the clinical tumor volume at follow-up 1. Furthermore, our model not only accurately replicates the serum PSA levels and tumor volume, but also captures the shape of the tumor as shown in the comparison of the tumors in Figure 5.4**b**. Additionally, we reliably reproduce the cellularity histogram of the patient's follow-up (Figure 5.4**b**).

We also depict the tumor evolution of Patient B from diagnosis to the follow-up 1 after 710 days (Figure 5.4**c**). In this case, the diagnosed tumor volume is $249\,\mathrm{mm}^3$ with a serum PSA level of $7.64\,\mathrm{ng/mL}$. Subsequently, at the follow-

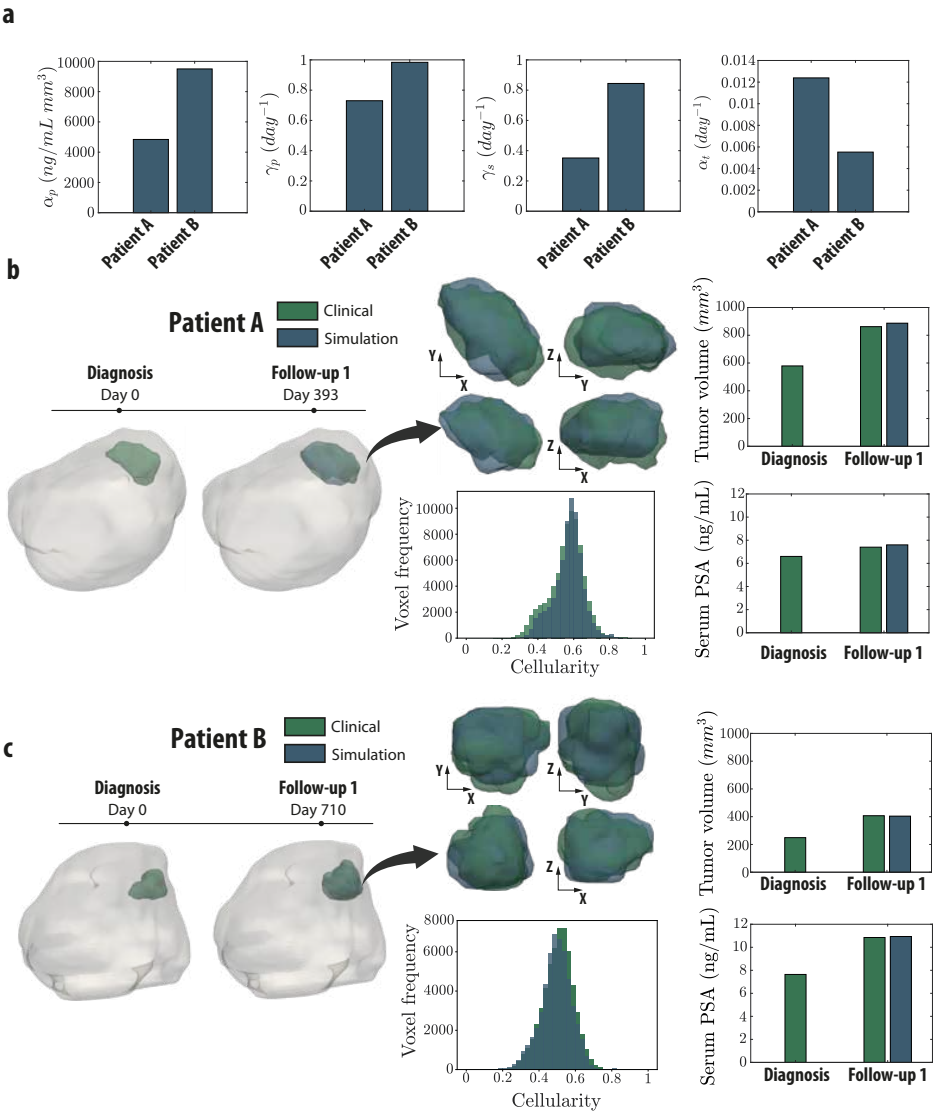

Figure 5.4: ***Patient-specific prostate tumor growth.*** ***a*** *Calibration of the physics-based parameters for Patients A and B.* ***b*** *Tumor growth in Patient A.* ***c*** *Tumor growth in Patient B. For both patients, representations of the actual prostate geometry with the tumor at diagnosis and comparisons between the clinical and simulated tumors at follow-up are provided. Additionally, comparisons between tumor volume, serum PSA, and prostate cellularity data from clinical and simulation records at follow-up are presented.*

up 1, the clinical serum PSA increases to $10.85\,\mathrm{ng/mL}$ and the tumor volume to $407\,\mathrm{mm}^3$. Here, we also replicate both the serum PSA ($10.94\,\mathrm{ng/mL}$) and tumor volume ($404\,\mathrm{mm}^3$), with relative errors of 0.7% and 0.8%, respectively. Moreover, the simulated tumor mimics the real shape of the tumor at the follow-up, also matching the histogram of prostate cellularity (Figure 5.4**c**).

Interestingly, we can observe that the patients have different tumor and PSA evolutions. Patient A's tumor volume increased from diagnosis to follow-up 1 at a rate of $0.72\,\mathrm{mm}^3/\mathrm{day}$ with only a $2.0\times10^{-3}\,\mathrm{ng/mL/day}$ increase in serum PSA, while Patient B's tumor volume and serum PSA increased at rates of $0.22\,\mathrm{mm}^3/\mathrm{day}$ and $4.5\times10^{-3}\,\mathrm{ng/mL/day}$, respectively. Thus, the growth rate of Patient A's tumor was 3.27 times higher, while the serum PSA increase rate was 2.25 times lower. These prognoses of the patients align with the calibration results of the physics-based parameters, which indicate that Patient A's tumor grows faster but secretes less than Patient B's tumor. In Patient B, given the higher rate of PSA secretion by tumor cells, variations in tumor size lead to significant fluctuations in PSA levels, establishing a direct relationship between tumor growth and serum PSA increase. However, in the case of Patient A, variations in tumor size result in smaller fluctuations in serum PSA levels, whereas an equivalent increase in serum PSA suggests a greater extent of tumor growth. Hence, discerning the prognosis of Patient A's tumor based on serum PSA variations alone might be misleading. This finding is consistent with the higher aggressiveness of Patient A's tumor as indicated by PI-RADS compared to Patient B. Consequently, our model demonstrates the capability to precisely simulate the progression of prostate cancer. It enables the extraction of patient-specific tumor growth parameters, thereby providing valuable insights into the tumor's characteristics and facilitating predictions about its aggressiveness or potential risk.

5.3.2 Predicting long-term patient-specific tumor progression from serum PSA tests

Our framework's greatest contribution is the capability to predict long-term tumor progression solely from serum PSA measurements in subsequent follow-up blood tests, potentially reducing the need for frequent MRI scans and thereby optimizing patient care. To do this, we fixed the parameters of the physics-based model obtained from each patient and inferred tumor growth from serum PSA levels at their subsequent follow-ups (Figure 5.5). Using this approach, we predicted the evolution of Patient A (Figure 5.5**a**). The clinical follow-ups for Patient A's serum PSA show a 5.94% increase from Follow-up 1 ($7.40\,\mathrm{ng/mL}$) to Follow-up 2 ($7.84\,\mathrm{ng/mL}$) at day 582. Subsequently, in Follow-up 3, the serum PSA increases by 7.22% ($8.45\,\mathrm{ng/mL}$) from Follow-up 2. We incorporate these measurements into our computational model and accurately replicate the serum PSA evolution, obtaining the evolution of the tumor that produces these varia-

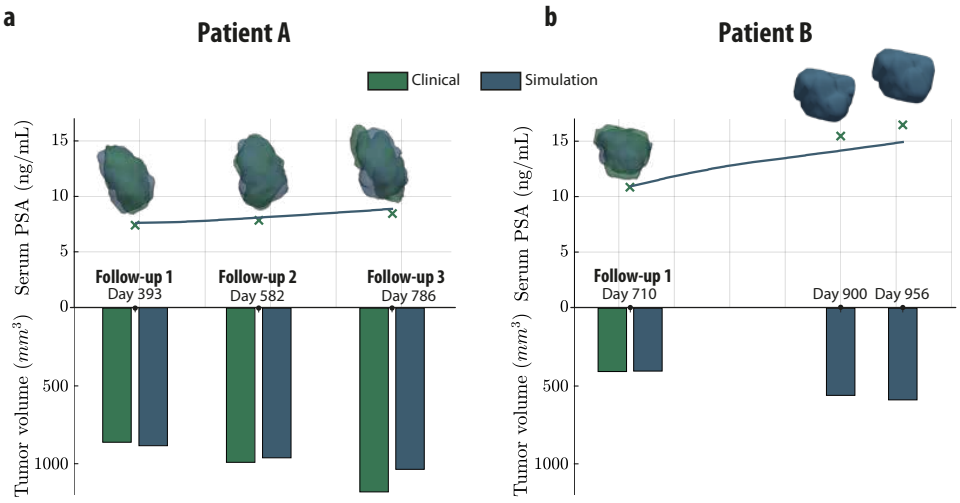

Figure 5.5: ***Prediction of patient-specific prostate tumor progression. a*** *Predictions for Patient A.* ***b*** *Predictions for Patient B. For both patients, the clinical serum PSA measurements from follow-up blood tests and the simulated evolution of serum PSA are shown. Additionally, a comparison between the clinical and predicted tumor volumes is provided (no subsequent image-based follow-up was conducted for Patient B after day 710).*

tions. For Patient A, follow-ups included both blood tests and MRI scans. Thus, we could compare not only the PSA levels but the tumors to further validate our predictions. At follow-up 1, the clinical tumor volume was $862\,\mathrm{mm}^3$. By follow-up 2, it increased to $991\,\mathrm{mm}^3$ (a 13% increase from follow-up 1), and at follow-up 3, it further grew to $1181\,\mathrm{mm}^3$ (a 16% increase from follow-up 2). As shown in Figure 5.5**a**, our predictions were in agreement with the actual evolution of Patient A, confirming that the nature of Patient A's tumor produces minimal PSA elevation despite continuous tumor growth and the potential of our model to infer patient-specific parameters and future tumor progression solely from serum PSA data.

We also predicted the evolution of Patient B based on blood tests conducted on days 900 and 956 (Figure 5.5**b**). Patient B's follow-ups rely solely on serum PSA tests (no MRI data), so predicted tumor volume at follow-ups 2 and 3 cannot be compared with clinical data. In this case, the serum PSA levels increased by 29.82% on day 900 compared to Follow-up 1, followed by a further increase of 6.53% on day 956. Our approach predicted a tumor volume increase of 38.61% on day 900 compared to Follow-up 1, followed by an additional increase of 5.18% over the next 56 days, consistent with the percentage increases in serum PSA. We conclude that Patient B's tumor growth is directly correlated with the increase in serum PSA levels. Therefore, the correlation between tumor growth and serum PSA levels allows for better detection in Patient B compared to Patient A.

5.4 Discussion

We have developed an innovative framework that integrates a digital twin for the prostate with a physics-informed machine-learning model to predict the prognosis of prostate cancer. Through this model, we can predict a patient's specific prostate tumor evolution based on an initial MRI and subsequent PSA blood tests. This approach involves employing a deep learning model to regulate tumor dynamics within the physics-based model, determining the fraction of proliferating tumor cells in each location of the prostate. The deep NN method we have developed takes into account the three-dimensional spatial distribution of physiological variables in the prostate, including cellularity levels, vascularization, tumor presence, and other relevant data, to capture the spatial-physical interactions driving disease progression. Importantly, our approach relies on a deep NN independent of the prostate's geometry (geometry agnostic), thereby enhancing its ability to generalize and apply effectively across patients. We validated our methodology by predicting tumor progression in two real patients and demonstrated its potential to accurately reproduce not only clinical tumor volume and shape over time, but also to offer insights into the aggressiveness or potential risk for the patient given their particular growth dynamics. The determination of physics-based patient parameters enabled us to characterize two distinct patient profiles. In one scenario (Patient B), tumor growth was directly correlated with the increase in serum PSA levels, which could be readily detected through blood tests. Conversely, in Patient A, the production rate of PSA was lower while the tumor growth rate was higher. Consequently, Patient A's tumor grew without a significant elevation in PSA levels, highlighting the limitation of relying solely on serum PSA for screening. This mechanism elucidates the well-known lack of sensitivity of PSA as a standalone biomarker for screening [321–323]. With further characterization of these parameters across more patients, we could potentially establish correlations between the model's parameters and tumor aggressiveness. Therefore, our model allows for the identification of patients with hidden tumor growth, where serum PSA levels do not rise despite tumor development, thus addressing the challenge of monitoring the disease solely through serum PSA and enabling more timely and tailored therapeutic interventions.

Furthermore, our computational framework allows for the prediction of tumor prognosis from periodic blood tests, offering the potential to predict long-term outcomes over periods spanning months or even years with remarkable computational speed (a simulation spanning 956 real-world days takes approximately 23 seconds to complete (see *Model implementation* in *Appendix* A.3.4 for further details). This efficiency facilitates swift, accurate prognostic evaluations, significantly reducing the reliance on invasive or more costly procedures and marking a leap forward in personalized cancer management. In our proposed physics-based model, we integrated biologically relevant aspects of prostate can-

cer, including the increase in tissue PSA caused by the PSA leakage from cancer cells. Under normal conditions, PSA produced by healthy cells is delivered to the urethra, however, in cancer pathological conditions, tumor cells proliferate without generating the necessary ducts for PSA delivery to the urethra, resulting in leakage. Depending on the prostate's vascularization, as described by the k_{trans} distribution, this abnormal concentration of PSA passes into the bloodstream. Thus, the k_{trans} distribution determines the spatial distribution of vascularization, tissue perfusion, and permeability in DCE-MRI [335]. The flux of fluid and substances between tissue and blood through the k_{trans} distribution has already been modeled [140, 337, 338]. However, to the best of our knowledge, this is also the first computational work in prostate cancer considering the intravasation of PSA from the tissue to the blood while taking into account the spatial distribution of vascularization through the k_{trans} distribution.

We proposed a mathematical multiscale model to simulate these biological processes and perform long-term simulations with reduced computational burden (see *Multiscale physics-based model* in *Appendix* A.3.2) for further details of the temporal multiscale model). In this model, certain assumptions were made. We inferred the concentration of tumor cells from cellularity and the percentage of tumor cells from biopsy results. If the biopsy is randomly sampled from the prostate, there may be samples without tumor, leading to an underestimation of the percentage. Ideally, our estimation would be more precise if the biopsy were guided to the tumor location. Additionally, the updating algorithm of the tumor mask considers an upper threshold for expansion, representing the carrying capacity that the voxel cannot exceed in that area. Although this upper threshold aligns with evidence that the cell population reaches a maximum tissue carrying capacity [339, 340], other algorithms could be introduced, such as considering this threshold as a stochastic parameter or increasing a certain percentage of the initial concentration. Furthermore, clinical segmented tumors may have inaccuracies, resulting in small deviations in clinical tumor volumes, which can be minimized with the utilization of available automated segmentation tools for prostate cancer [341, 342] and for geometry reconstruction [343]. Regarding the ML model, larger datasets of serum PSA at different time points could be employed, enhancing the capacity to reproduce various dynamic behaviors or training on different geometries. However, despite these simplifications, we have demonstrated the robust predictive capabilities of our model in determining the patient's prognosis.

5.5 Conclusions

Therefore, this framework, which integrates physics-informed ML models into the prostate digital twin, offers a novel approach to creating a potential computational surrogate of MRI imaging to monitor the prognosis of prostate cancer and reduce the uncertainty of tumor growth from serum PSA blood tests.

Chapter 6

Conclusions

It does not matter how slowly you go as long as you do not stop.

– ***Confucius***

Contents

6.1 Original contributions and general conclusions

This thesis has presented computational models designed to replicate cell behavior and simulate biological systems to advance toward a more comprehensive understanding of cancer. Various phenomena have been investigated across the scales, such as cell migration, morphogenesis, and tumor growth. To achieve this goal, different modeling approaches have been employed, including discrete models utilizing ABM and continuum approaches involving PDEs. Furthermore, this thesis emphasizes the validation of these models against experimental data and develops tools to ensure the computational models accurately capture biological phenomena. To this end, diverse calibration techniques have been utilized, from classical methodologies to ML techniques. Additionally, these models are designed within frameworks to facilitate their use by other researchers in similar applications, offering adaptability and modularity to integrate with other models and experiments.

Different works have been presented throughout this thesis. Firstly, we study individual cell migration in microfluidic devices (*Chapter 2*). Here, we aim to understand whether the conversion of T cells into CAR-T cells affects their migration capacity, potentially hindering their infiltration into solid tumors, which could impede the success of the therapy. To achieve this, we develop a framework to predict 3D migration from 2D experimental microscopy-based data using a 3D ABM. In this model, we simulate the temporal changes in the migration direction with a sampling method based on inverse transform sampling. This method approximates the rotation of the direction of migration over time by generating pseudo-random numbers from a uniform distribution and applying the inverse CDF associated with the desired probability distribution. Then, we propose an optimization procedure based on the pattern search method to obtain the CDFs that correctly orient cell migration over time and minimize the error between the simulation and experimental metrics characterizing the migration patterns.

With this model, we simulate cell migration assays of T and CAR-T cells in microfluidic devices conducted under hydrogels with different concentrations of type I collagen. We demonstrate that the 2D projections of the simulated 3D trajectories accurately reproduce the experimental 2D trajectories in both cell types across different density matrices. Hence, our simulations indicate that CAR-T cell migration is more sensitive to collagen concentration increases than T cells, resulting in a more pronounced reduction in their invasiveness. Notably, our computational model reveals significant differences in 3D movement patterns between T and CAR-T cells. T cells exhibit migratory behavior in 3D within the collagen concentration matrices, consistent with the high motility typically observed in immune cells that enable them to access intricate body regions. In contrast, our simulations show that CAR-T cells predominantly move within the XY plane, with limited movement in the Z direction, suggesting primarily 2D

migration. However, the presence of the chemical factor CXCL12 not only stimulates CAR-T cells to migrate within the XY plane but also induces a behavioral shift, promoting 3D movement similar to T cells. Therefore, enhancing the migrative capacity of CAR-T cells with chemical factors may improve the efficacy of therapies in solid tumors. We validate these findings with cell migration experiments using light-sheet microscopy to quantify 3D cell migration patterns, aligning with our predictions. To explain this behavior, we hypothesize that exposure to chimeric antigen receptors during the conversion of T cells into CAR-T cells, along with the geometrical characteristics of microfluidic devices where the vertical dimension (Z) significantly differs in scale from the other two dimensions, and the potential influence of fiber alignment occurring in the XY plane, could influence their migration capabilities, potentially impeding their ability to navigate through intricate and confined spaces.

Therefore, this work demonstrates that 2D projections of 3D trajectories may not accurately represent real migration patterns, highlighting the importance of analyzing migration in 3D. Moreover, it offers a tool to estimate 3D migration patterns from 2D experimental data, which can be easily obtained with automatic quantification algorithms. This approach helps reduce the need for sophisticated and expensive microscopy equipment required in laboratories, as well as the computational burden involved in producing and analyzing 3D experimental data.

Next, we continue at the microscale, focusing on the morphogenetic process of lumen formation starting from an individual cell to understand, from a mechanobiological perspective, how it is formed and the reasons behind its deviation into the tumoral structures found in carcinomas (*Chapter 3*). Particularly, we investigate the role of luminal hydrostatic pressure generated during the process due to fluid secretion by cells, which acts as the driving force, and the regulatory function of ECM density. Previous studies had neglected the interaction between fluid and cells and the ECM, mainly concentrating on 2D-lumen formation. To address these gaps, we develop a 3D multi-ABM that incorporates various types of agent particles. The first type simulates biological cells and integrates a mathematical model for cell proliferation with mechanoregulation feedback from matrix density. We also incorporate a cell fluid secretion process into this biological agent to generate fluid in the lumen domain. The fluid, generated by this process, is simulated using the second type of particles, adopting a discrete approach to simulate fluid dynamics. This discrete approach facilitates the simulation of fluid-cell interaction, occurring within a deformable interface that expands over time due to cell movement driven by mechanical interactions and cell proliferation. Moreover, the particle-based fluid description enables an effective simulation of fluid deposition resulting from cell secretion, facilitating the formation of lumens *de novo* and the gradual increase in fluid volume over time. This aspect is particularly challenging to capture accurately using both grid-based methods such as CFD and particle-based methods such as SPH.

We studied lumen formation under different matrix densities and successfully replicated the experimental observations achieved by Paszek et al., offering a mechanical explanation for the regulatory role of matrix density and fluid secretion. A physiological-like matrix density provides adequate biomechanical conditions for lumen formation and a proper structure. However, an increase in matrix density disrupts normal coordination and promotes aberrant formation, characterized by an increase in luminal hydrostatic pressure and the formation of multiple small lumens. When matrix density is high, the luminal hydrostatic pressure generated by cell secretion cannot displace cells as easily as when matrix density is low, as the matrix opposes the cells' movement. Consequently, the lumen remains diminished, and hydrostatic pressure increases with successive cell secretions. In this situation, cell division does not contribute to creating luminal space because the lumen size is small, and its surface is fully covered by overlapped cells. Conversely, with no space around the lumen surface, cell division forces some cells to leave the monolayer, where they polarize to create a new lumen. Finally, a dysfunctional structure with numerous small lumens is formed.

Therefore, this model underscores the capacity of mechanistic models to replicate behaviors, identify causality, and subsequently address biological questions, such as the role of matrix density in lumen formation and its deviation towards aberrant architectures associated with tumor pathogenesis observed in carcinomas.

Subsequently, we extend the lumen morphogenesis model to replicate real *in vitro* experiments of pancreatic cystic tumor organoids (*Chapter 4*). However, the experimental results display great variability in terms of the final size of the pattern, which is characteristic of the intrinsic stochasticity of biology. This complicates their simulation because it implies that the organoids have different dynamical processes and cannot be effectively reproduced with parameters that coordinate cell behavior, as this behavior must change in each case. Therefore, we aim to build a model that efficiently adapts to this observed variability in the organoid patterns to decipher cell behavior. To achieve this, we develop a framework combining ABM and ML to simulate morphogenetic patterns of 3D organoids. Here, we eliminate any rule-based parameters that model cell behavior in the ABM and instead introduce a DL model that learns to coordinate temporary cell behavior to achieve the final morphogenetic pattern. Thus, the ABM only includes physically relatable parameters, advancing to make ABM more parsimonious, enhancing its interpretability, and addressing the criticism of reduced realism induced by the heuristic-type laws typically incorporated. In this framework, after a cell completes a biological function in the ABM, the NN receives metrics evaluating the state of the physics-based simulation and data describing the final morphogenetic pattern to achieve. Subsequently, the NN activates the corresponding next biological cell function for that cell.

To make the NN learn a morphogenetic pattern, we propose an approach sim-

ilar to RL. In this method, we simulate an example of the desired pattern, evaluate the accuracy of the results using a fitness function, and optimize the NN parameters to maximize the fitness function. This optimization of the NN is achieved through a GA, although other algorithms can be implemented. One advantage of this approach is that it requires minimal information about the morphogenetic processes and only needs the final morphological information about the pattern extracted visually. Once the NN is trained, we demonstrate the framework's potential to accurately replicate the morphogenesis of organoids, both cystic tumor organoids and solid organoids from the pancreas. In both cases, we showcase the accuracy of the simulated patterns compared to *in vitro* experiments, and the robustness of the NN in coordinating cell functions despite variations in cell cycle time or fluid secretion time, as well as the introduction of noise in these processes. Also, it is demonstrated that this methodology can benefit not only from reproducing a final morphogenetic pattern but also from reproducing its evolution through the incorporation of time-series data. This artificial intelligence-based method learns the intrinsic mechanisms of the morphogenetic process, enabling it to generalize and reproduce similar morphogenetic patterns with different dynamical processes without the need for *ad hoc* parameter optimization.

Moreover, this work provides further demonstration of the capability of the developed discrete model to simulate fluid-like behavior. We conduct a shear test of the particle-based fluid model and illustrate how to maintain the rheological properties of the fluid as a Newtonian fluid, despite changes in the discretization of particle size.

From a biological perspective, our simulations reveal that certain cells are predisposed to perform specific functions. We observe a heterogeneity of cellular functions within the organoid, with some cells proliferating more while others are more inclined to secrete fluid. It appears that initial cells possess a higher proliferative potential, and as they divide, the daughter cells become more specialized with a higher level of differentiation. Furthermore, in cystic organoids, the final size depends on morphology at the initiation of the lumen, with proliferative capacity diminishing as the lumen expands. We hypothesize that initially, cells exhibit a less differentiated phenotype and have a greater proliferative capacity. Subsequently, as they proliferate, the daughter cells adopt a higher degree of differentiation and specialize into secretory cells. Consequently, once secretion begins, the proliferative capacity of the organoid diminishes, and the final size becomes limited by its size at the time of lumen initiation. Thus, the proliferation capacity of cells may be constrained by the initiation of lumen formation and the final size may be predicted at the beginning of lumen formation.

Therefore, this work first showcases the potential of combining discrete ABM and ML in computational biology to reproduce cell behavior. It demonstrates the synergistic power of ML to learn to replicate data effectively without the need for recalibrations and solve the inverse problem of deciphering cell behavior from a

final pattern, along with the mechanistic model to obtain biological interpretations.

Finally, we transition from the microscale to the macroscale to investigate patient-specific prostate cancer at the organ level (*Chapter 5*). In clinical practice, prostate cancer is typically diagnosed through MRI, and tumor progression is monitored through routine blood tests to measure the biomarker PSA. However, there is controversy surrounding this biomarker as an indicator of tumor evolution due to its lack of specificity and sensitivity. In many cases, tumors progress without variations in PSA levels, obscuring their monitoring. Hence, this work presents a novel framework that integrates a physics-based model with a ML method for predicting patient-specific prostate tumor growth in DTs from PSA blood tests.

Here, we illustrate the methodology of constructing a DT from MRI scans, incorporating physiological information. Then, the patient's DT is introduced into a physics-based model employing a continuum representation with PDEs to predict tumor prognosis. This model incorporates biologically relevant aspects of prostate cancer, including the increase in tissue PSA resulting from PSA leakage from cancer cells. Additionally, we consider PSA intravasation from tissue to blood, taking into account the spatial distribution of vascularization, which had been neglected in existing works. Furthermore, the model encompasses a mathematical representation of tumor growth. To regulate growth dynamics within the physics-based model, we introduce a novel ML approach not previously proposed. This DL model considers the 3D prostate geometry and spatial interactions among physiological data to control tumor growth based on patient-specific PSA blood test results. Importantly, this developed methodology is independent of the individual patient's geometry to ensure generalizability across patients without the need for refittings, offering the potential to forecast long-term outcomes over periods spanning months or even years with remarkable computational speed.

We validate our methodology by predicting tumor progression in two real patients and demonstrate its potential to accurately reproduce not only clinical tumor volume and shape over time, but also to offer insights into the aggressiveness or potential risk for the patient given their particular growth dynamics. We propose a calibration method for the physics-based model to reproduce a patient's specific tumor growth. Through mechanistic modeling, we demonstrate that the parameters of the physics-based model can provide patient-specific insights, particularly regarding the aggressiveness and potential risk of the patient's tumor. Thus, our model identifies a clinical scenario of patients with hidden tumor growth, where serum PSA levels do not rise despite tumor development, addressing the challenge of monitoring the disease solely through serum PSA and enabling more timely and tailored therapeutic interventions.

Therefore, this framework, which integrates physics-informed machine learning models into the prostate DT, paves a new way of predicting tumor progres-

sion with DT and exploring the benefits of mechanistic modeling and ML. Moreover, it offers a novel approach to creating a potential computational surrogate of MRI imaging to monitor the prognosis of prostate cancer and reduce the uncertainty of tumor growth from serum PSA blood tests. Thus, this work represents a step towards employing computational biology for clinical applications, offering patient-specific and accurate real-time predictions of tumor prognosis.

6.2 Future lines

Different future research directions have emerged throughout this thesis to address some of the limitations presented in the works and to explore new opportunities and paths opened by this research.

One aspect that could benefit from improvement in the developed ABM models is the representation of the ECM. In our simulations, we considered the viscous component of the matrix and its effect on cells using Stoke's drag force while neglecting the elastic component. However, the ECM is a compound material with a viscoelastic rheological response. Considering this, a more accurate representation of the ECM mechanical properties [344] could enhance our ability to simulate the lumen formation process and better understand its effects. Additionally, incorporating the elastic component would allow us to calculate stresses and gain a comprehensive understanding of the mechanical environmental variables involved, not only during healthy formation but also during its deviation into tumoral formation. This could lead to a more detailed description of the mechanical forces arising from cell-ECM interactions and result in greater heterogeneity in mechanical forces between cells.

Furthermore, future work arising from the framework that combines ABM and ML for organoid morphogenesis, presented in *Chapter 4*, involves informing the NN with physical data. In this model, the NN was provided with data metrics that evaluate the state of the simulation, facilitating comparison and correlation with experiments. This enabled effective simulation of experiments as the NN controlled cell behavior based on differences between experiments and simulations. However, another approach involves considering physical and chemical aspects such as cell-cell forces, cell-matrix forces, or glucose levels, among others, to predict cellular functions with the NN. In this case, the loss function would integrate the comparison of data metrics between experiments and simulation results. Therefore, with this approach, when having sufficient data from experiments to train the NN to make it learn multiple behaviors, we can advance into a ML that integrates all possible cell behaviors behind pattern morphogenesis. Alternatively, another perspective on the combination of ABM and ML involves employing Convolutional Neural Networks (CNNs) to predict cell behavior directly from experimental images in simulations. Here, instead of providing physics-based information to the NN, a CNN would receive an image of the experimen-

tal outcome to guide cell behavior in the ABM. In this case, the loss function would integrate the differences between the input images and the images of the simulation results from the ABM so that we could approximate the experiments spatially.

Lastly, the integration of ABM with RL schemes offers a wide range of possibilities for simulating biological systems. This approach, in which agents learn by interacting with the environment and other agents to maximize a reward signal, aligns nicely and integrates naturally with ABMs. This reward can represent cell fate and encompass homeostasis, correct physiological functioning, survival, or, for tumor cells, maximize proliferation or invasion. Thus, this approach can help us understand cell migration and how mechanical and chemical cues from the environment guide it, simulate the competition of cells and tumor cells for nutrients such as glucose or oxygen, reproduce tumor growth, or analyze mechanical interactions regulating developmental processes. Therefore, RL in biology is a promising field with clear potential for exploration.

Moving to the organ scale, another approach to predict cancer evolution in DTs involves a fully ML model with Graph Neural Networks (GNNs) [345]. Here, a graph from the DT is generated, with each node condensing all the information derived from the MRI. The GNN then creates an embedding of the graph, capturing the geometrical interactions between the physiological data from the MRI. Subsequently, an ML model with a NN can make predictions about tumor evolution considering the holistic information of the graph representing the DT. This model can be trained using a dataset of diagnostic MRIs to generate the graph and their corresponding follow-up MRIs to validate the predictions. Furthermore, certain GNN models can also consider changes in the graph over time, predicting not only tumor growth but also changes in the geometry. Therefore, this approach holds promise for improved prediction of cancer growth, leveraging all the information condensed in the MRI.

6.3 Thesis contributions

This section presents the contributions of this thesis.

6.3.1 Publications in peer-reviewed journals

The publications in international peer-reviewed journals achieved during the thesis period are listed below:

1. Juste-Lanas, Y., Guerrero, P. E., **Camacho-Gomez, D.**, Hervas-Raluy, S., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2022). *Confined Cell Migration and Asymmetric Hydraulic Environments to Evaluate the Metastatic Potential of Cancer Cells.* ASME. Journal of Biomechanical Engineering, 144(7): 074502. https://doi.org/10.1115/1.4053143.

2. **Camacho-Gomez, D.**, Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2022). *A 3D multi-agent-based model for lumen morphogenesis: the role of the biophysical properties of the extracellular matrix*. Engineering with Computers, 38(5), 4135-4149. https://doi.org/10.1007/s00366-022-01654-1.

3. **Camacho-Gomez, D.**, Sorzabal-Bellido, I., Ortiz-de-Solorzano, C., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2023). *A hybrid physics-based and data-driven framework for cellular biological systems: Application to the morphogenesis of organoids*. iScience, 26(7). https://doi.org/10.1016/j.isci.2023.107164.

4. **Camacho-Gomez, D.**, Movilla, N., Borau, C., Martin, A., Oñate-Salafranca, C., Pardo, J., Gomez-Benito, M. J., & Garcia-Aznar, J. M. (2024). *An agent-based method to estimate 3D cell migration trajectories from 2D measurements: Quantifying and comparing T vs CAR-T 3D cell migration*. Under review.

5. **Camacho-Gomez, D.**, Borau, C., Garcia-Aznar, J. M., Gomez-Benito, M. J., Girolami, M., & Perez, M. A. (2024). *A physics-informed machine learning digital twin framework for prostate cancer: predicting tumor growth via serum PSA blood tests*. Submitted.

6.3.2 Conferences contributions

The following communications have been presented at national and international conferences and meetings (underlines indicate the work was presented by the author):

1. Camacho-Gomez, D., Borau, C., Garcia-Aznar, J. M., Gomez-Benito, M. J., Girolami, M., & Perez, M. A. *A data-driven physics-based model for predicting prostate cancer progression from the PSA blood test*. 9th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS). 3 - 7 June 2024. Lisbon, Portugal. In-person oral participation.

2. Gomez-Benito, M. J., Camacho-Gomez, D., Aparicio, R., Garcia-Aznar, J. M. Advancing our understanding of pathological and physiological processes through Computational Mechanobiology. 7th Barcelona VPH Summer School. 7 - 9 June 2023. Barcelona, Spain. Co-author oral presentation.

3. Camacho-Gomez, D., Garcia-Aznar, J. M., Bellido-Sorzabal, I., Ortiz de Solorzano Aurusa, C., & Gomez-Benito, M. J. *A hybrid model of organoids morphogenesis*. International Symposium on Computer Mechanics in Biomechanics and Biomedical Engineering (CMBBE2023). 3 - 5 May 2023, Paris, France. In-person oral participation.

4. Camacho-Gomez, D., Guerrero, P. E., Garcia-Aznar, J. M., Gomez-Benito, M. J. *An agent-based model to simulate DIPG migration in microfluidic devices*. Virtual Physiological Human (VPH22). 6 - 9 September 2022, Porto, Portugal. In-person oral participation.

5. Camacho-Gomez, D., Perez, A., Borau, C., Gomez-Benito, M. J., Garcia-Aznar, J. M., Perez, M. A. *Simulation of Tumors using Computational Models*. Disfrutar Divulgando Desinteresadamente Conference. 2022. Zaragoza.

6. Camacho-Gomez, D., Garcia-Aznar, J. M., Gomez-Benito, M. J. *A computational model to study the interplay between cell fluid secretion and matrix stiffness in lumen morphogenesis*. 11th European Solid Mechanics Conference (ESMC22). 4 - 8 July 2022, Galway, Ireland. In-person oral participation.

7. Camacho-Gomez, D., Guerrero, P. E., Garcia-Aznar, J. M., Gomez-Benito, M. J. *Hybrid discrete-continuum model to evaluate DIPG cells invasion and matrix degradation*. USACM Thematic Conference on the Role of Mathematical and Computational Modeling in Cancer Research. Virtual oral participation.

8. Camacho-Gomez, D., Garcia-Aznar, J. M., Gomez-Benito, M. J. *Computational model for cell-matrix interaction as a regulator of lumen formation*. X Reunión del Capítulo Español de la Sociedad Europea de Biomecánica (ESB). 25-26 de octubre de 2021, Granada. In-person oral participation.

9. Camacho-Gomez, D., Garcia-Aznar, J. M., Gomez-Benito, M. J. *A 3D agent-based model to explore DIPG cell invasion in different stiffness matrices*. 6th U.S. National Congress on Computational Mechanics (USNCCM16). 2021. Virtual oral participation

10. Guerrero, P. E., Camacho-Gomez, D., Mendoza, N., Schuhmacher, A., Gomez-Benito, M. J., Garcia-Aznar, J. M. Mechanobiology of extracellular matrix on diffuse intrinsic pontine glioma (DIPG) tumor growth and migration. American Society of Cell Biology (ASCB). 2021. Co-author - Poster Presentation.

11. Guerrero, P. E., Camacho-Gomez, D., Mendoza, N., Schuhmacher, A., Gomez-Benito, M. J., Garcia-Aznar, J. M. Diffuse Intrinsic Pontine Glioma (DIPG) tumor growth and migration using 3D biomimetic hydrogels. Organoids in Cancer Research an Agora - EPFL Workshop. 2021. Co-author - Poster Presentation.

12. Goncalves, I., Camacho-Gomez, D., P. E. Guerrero, Gomez-Benito, M. J., Garcia-Aznar, J. M. Mechanobiology of extracellular matrix on the tumor

growth. 26th Congress of the European Society of Biomechanics. 2021. Co-author - Virtual Oral Presentation.

6.3.3 Open-source software

The author has contributed with three open repositories of the frameworks developed:

- https://github.com/daniel-camacho-gomez/AIOrganoids
- https://github.com/daniel-camacho-gomez/2D_Data-to-3D_Migration
- https://github.com/daniel-camacho-gomez/ProstateNet

6.4 Teaching, supervising and mentoring activities

The author has co-supervised one Master's thesis entitled *A hybrid physics-based and deep learning model for the simulation of cell migration in confined channels*, one Bachelor's thesis entitled *Study of cell dynamics through agent-based modeling*, and has fully supervised one Bachelor's thesis entitled *Computational study of the influence of extracellular matrix anisotropy on cell migration.*

The author has also taught the following courses:

1. 2022. Pre-course "Agent-Based Models in Biomedical Applications" at the XI Meeting of the Spanish Chapter of the European Society of Biomechanics (ESB), Zaragoza.
2. 2022-2024. Interdisciplinary seminar course titled 'Agent-based Models for Cell Simulation' within the Master's Degree Program in Biomedical Engineering at the University of Zaragoza.

6.5 Collaborations

During the development of the thesis, the author has participated and established the following collaborations:

Individual and Collective Migration of Immune Cellular Systems (ICoMICS) project. This EU-funded project aims to develop a modeling approach capable of predicting how therapeutic immune cells migrate, interact with the tumor microenvironment, and contribute to the improvement of immunotherapy outcomes. The results of this work are presented in *Chapter 2*.

Microfluidics mechanobiology and image analysis for the simulation of tumoral growth in 3D (TUMOR-ON-CHIP) project. This project establishes a collaboration between Fundación para la Investigación Médica Aplicada (FIMA), University of Navarra, experts in image analysis and quantitative microscopy applied

to the study of biomedical problems, and the University of Zaragoza. The objective is to study the growth of tumor organoids cultured in microfluidic devices. This collaboration involves complementary work, where the computational models developed by the University of Zaragoza would be experimentally validated using the quantitative microscopy tools developed by FIMA. The outcomes of this joint project are presented in *Chapter 4*.

ProCanAid project. This initiative aims to utilize a digital twin of the prostate for aiding in detection, diagnosis, and simulating the effects and effectiveness of various oncological treatments for prostate cancer. The project entails collaboration between La Fe Hospital, which provides patient data, the company Quibim, which employs image analysis to automate segmentation and extraction of imaging biomarkers, and the University of Zaragoza, which utilizes this data to develop computational models for predicting prostate cancer growth. *Chapter 5* presents the work carried out within this project.

Moreover, the author conducted a research stay lasting three months at the Computational Statistics and Machine Learning Group at the Department of Engineering, University of Cambridge, led by Prof. Mark Girolami. During this stay, the author developed the machine learning model for the ProCanAid project presented in *Chapter 5*.

Additionally, the author contributed by performing computational fluid simulations of asymmetric hydraulic environments in microfluidic devices in the study to evaluate the metastatic potential of cancer cells within confined channels, aiming to understand the influence of pressure gradients on cell migration [49].

Finally, the author has also participated in the European PRIMAGE project. This project, financed by the European Commission, has 16 European partners that participate in the consortium and has an implementation duration of 4 years. This project proposes an open cloud-based platform to support decision-making in the clinical management of two pediatric cancers, Neuroblastoma (NB), the most frequent solid cancer of early childhood, and Diffuse Intrinsic Pontine Glioma (DIPG) the leading cause of brain tumor-related death in children. PRIMAGE platform implements the latest advancement of in silico imaging biomarkers and modeling of tumor growth towards a personalized diagnosis, prognosis, and therapy follow-up.

6.6 Funding

This doctoral thesis has been supported by PRedictive In-silico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers (PRIMAGE) (G.A. no. 826494) and Procanaid Next Generation EU (Grant No. PLEC2021-007709). The research stay at the University of Cambridge was additionally supported by the Erasmus+ program managed by Campus Iberus and Fundación Ibercaja-Cai (No IT 5/23).

Chapter 7

Conclusiones

Cuando llegue la inspiración, que me encuentre trabajando.

– **Pablo Ruiz Picasso**

Contents

7.1 Contribuciones originales y conclusiones generales

En esta tesis se han presentado modelos computacionales para replicar el comportamiento celular y simular sistemas biológicos con el fin de avanzar nuestro conocimiento sobre el cáncer. Se han investigado diversos fenómenos en distintas escalas, tales como la migración celular, la morfogénesis y el crecimiento de tumores. Para ello, se han empleado diferentes enfoques de modelado, incluyendo modelos discretos que utilizan *Modelos Basados en Agentes* (ABM, de sus siglas en inglés) y enfoques continuos que involucran *Equaciones en Derivadas Parciales* (PDE, de sus siglas en inglés). Además, esta tesis hace hincapié en la validación de estos modelos con datos experimentales y desarrolla herramientas para garantizar que los modelos computacionales capturen con precisión los fenómenos biológicos. Con este fin, se han utilizado diversas técnicas de calibración, desde metodologías clásicas hasta técnicas de *Aprendizaje Automático* (ML, de sus siglas en inglés). Además, estos modelos se han diseñado dentro de entornos automatizados para facilitar su uso por parte de otros investigadores en aplicaciones similares, ofreciendo adaptabilidad y modularidad para integrarse con otros modelos y experimentos.

Diversos trabajos han sido presentados a lo largo de esta tesis. En primer lugar, se ha estudiado la migración celular individual en dispositivos de microfluídica (*Capítulo 2*). El objetivo es comprender si la conversión de células T en células CAR-T afecta a su capacidad de migración, potencialmente obstaculizando su infiltración en tumores sólidos, lo que podría dificultar el éxito de la terapia. Con tal fin, se ha desarrollado un marco de trabajo para predecir la migración tridimensional a partir de datos experimentales basados en microscopía bidimensional utilizando un modelo tridimensional basado en agentes. En este modelo, se han calculado los cambios temporales en la dirección de migración con un método de muestreo basado en muestreo de transformada inversa. Este método aproxima la rotación de la dirección de migración en el tiempo generando números pseudoaleatorios a partir de una distribución uniforme y aplicando la función de distribución acumulativa inversa asociada con la distribución de probabilidad deseada. Posteriormente, se ha propuesto un procedimiento de optimización basado en el método de búsqueda de patrones para obtener las funciones de distribución acumulativa que orientan correctamente la migración celular en el tiempo y que minimizan el error entre la simulación y las métricas experimentales que caracterizan los patrones de migración.

Con este modelo, se han simulado ensayos de migración celular de células T y CAR-T en dispositivos de microfluídica llevados a cabo con hidrogeles de diferentes concentraciones de colágeno tipo I. Así, se ha demostrado que las proyecciones 2D de las trayectorias 3D simuladas reproducen con precisión las trayectorias 2D experimentales en ambos tipos de células en las distintas matrices. Por tanto, las

simulaciones indican que la migración de las células CAR-T es más sensible a los aumentos de concentración de colágeno que las células T, lo que resulta en una reducción más pronunciada de su invasividad. Cabe destacar que el modelo computacional revela diferencias significativas en los patrones de movimiento 3D entre las células T y CAR-T. Las células T muestran un comportamiento migratorio en 3D dentro de las matrices de colágeno, consistente con la alta movilidad típicamente observada en las células inmunes que les permite acceder a regiones corporales intrincadas. Por el contrario, las simulaciones muestran que las células CAR-T se mueven predominantemente dentro del plano XY, con movimiento limitado en la dirección Z, sugiriendo principalmente una migración 2D. Sin embargo, la presencia del factor químico CXCL12 no solo estimula a las células CAR-T a migrar dentro del plano XY, sino que también induce un cambio de comportamiento, promoviendo un movimiento 3D similar al de las células T. Por lo tanto, mejorar la capacidad migratoria de las células CAR-T con factores químicos puede mejorar la eficacia de las terapias en tumores sólidos. Así, se han validado estos hallazgos con experimentos de migración celular utilizando microscopía de fluorescencia de lámina de luz, los cuales confirman los patrones de migración celular predichos por las simulaciones. Para explicar este comportamiento, se ha hipotetizado que la exposición a receptores antigénicos quiméricos durante la conversión de células T en células CAR-T, junto con las características geométricas de los dispositivos microfluídicos donde la dimensión vertical (Z) difiere significativamente en escala de las otras dos dimensiones, y la posible influencia de la alineación de fibras que ocurre en el plano XY, podrían influir en sus capacidades de migración, potencialmente obstaculizando su capacidad para navegar a través de espacios intrincados y confinados.

De este modo, este trabajo demuestra que las proyecciones bidimensionales de las trayectorias tridimensionales pueden no representar con precisión los patrones reales de migración, resaltando la importancia de analizar la migración en 3D. Además, el modelo desarrollado constituye una herramienta computacional para estimar los patrones de migración tridimensional a partir de datos experimentales bidimensionales, los cuales pueden obtenerse fácilmente con algoritmos de cuantificación automática. Este enfoque, por tanto, ayuda a reducir la necesidad de equipos de microscopía sofisticados y costosos requeridos en laboratorios, así como la carga computacional involucrada en la producción y análisis de datos experimentales tridimensionales.

A continuación, siguiendo en la escala microscópica, el objetivo se ha centrado en el proceso morfogénesis de formación del lumen a partir de una célula individual para comprender, desde una perspectiva mecanobiológica, cómo se forma y las razones detrás de su desviación hacia las estructuras tumorales encontradas en carcinomas (*Capítulo 3*). En particular, se ha investigado el papel de la presión hidrostática luminal generada durante el proceso debido a la secreción de fluido por las células, que actúa como la fuerza impulsora, y la función reguladora de

la densidad de la matriz extracelular. Los estudios previos habían despreciado la interacción entre el fluido y las células y la matriz extracelular, concentrándose principalmente en la formación de lúmenes en 2D. Para abordar estas limitaciones, se ha desarrollado un modelo multiagente 3D que incorpora varios tipos de partículas agentes. El primer tipo simula células e integra un modelo matemático para la proliferación celular con una retroalimentación mecanorregulada por la densidad de la matriz. También se ha incorporado al agente biológico un proceso de secreción de fluido para generar fluido en el dominio del lumen. El fluido, generado por este proceso, se ha simulado utilizando el segundo tipo de partículas, adoptando un enfoque discreto para esta simulación de la dinámica de fluidos. Este enfoque discreto facilita la simulación de la interacción fluido-célula, la cual ocurre en una interfaz deformable que se expande en el tiempo debido al movimiento celular consecuencia de las interacciones mecánicas y proliferación celular. Además, la descripción del fluido basada en partículas permite una simulación efectiva de la deposición de fluido resultante de la secreción celular, facilitando la formación de lúmenes *de novo* y el aumento gradual del volumen de fluido con el tiempo. Este aspecto es particularmente delicado de capturar con precisión utilizando tanto métodos basados en malla como *Dinámica de Fluidos Computacional* (CFD, de sus siglas en inglés) como en métodos basados en partículas como *Hidrodinámica de Partículas Suavizadas* (SPH, de sus siglas en inglés).

Así, se ha estudiado la formación del lumen bajo diferentes densidades de matriz y se ha logrado replicar con éxito las observaciones experimentales realizadas por Paszek et al., ofreciendo una explicación mecánica del papel regulador de la densidad de la matriz y la secreción de fluido en el lumen. Una densidad de matriz fisiológicamente similar proporciona condiciones biomecánicas adecuadas para la formación de una estructura apropiada del lumen. Sin embargo, un aumento en la densidad de la matriz impide la coordinación normal y promueve una formación aberrante, caracterizada por un aumento en la presión hidrostática luminal y la formación de múltiples lúmenes pequeños. Cuando la densidad de la matriz es alta, la presión hidrostática luminal generada por la secreción de fluido intersticial no puede desplazar a las células tan fácilmente como cuando la densidad de la matriz es baja, ya que la matriz se opone al movimiento de las células. En consecuencia, el lumen permanece disminuido y la presión hidrostática aumenta con las sucesivas secreciones celulares. En esta situación, la división celular no contribuye a crear espacio luminal porque el tamaño del lumen es pequeño y su superficie está completamente cubierta por células superpuestas. Por el contrario, al no haber espacio alrededor de la superficie del lumen, la división celular obliga a algunas células a abandonar la monocapa, donde se polarizan para crear un nuevo lumen. Finalmente, se forma una estructura disfuncional con numerosos lúmenes pequeños.

Por lo tanto, este modelo subraya la capacidad de los modelos mecanicistas para replicar comportamientos, identificar causalidad y posteriormente abordar

preguntas biológicas, como el papel de la densidad de la matriz en la formación del lumen y su desviación hacia arquitecturas aberrantes asociadas con la patogénesis tumoral observada en carcinomas.

Posteriormente, se ha extendido el modelo de morfogénesis del lumen para replicar experimentos *in vitro* reales de organoides tumorales quísticos pancreáticos (*Capítulo 4*). En este sentido, los resultados experimentales muestran una gran variabilidad en cuanto al tamaño final del patrón, lo cual es característico de la estocasticidad intrínseca de la biología. Esto complica su simulación porque implica que los organoides tienen diferentes procesos dinámicos por lo que no pueden reproducirse eficazmente con parámetros fijos que coordinen el comportamiento celular en todos los casos, ya que este comportamiento debe cambiar en cada uno de ellos. Por lo tanto, el objetivo es construir un modelo que se adapte eficientemente a esta variabilidad observada en los patrones de organoides para descifrar el comportamiento celular. Para tal fin, se ha desarrollado un marco de trabajo que combina ABM y ML para simular patrones morfogenéticos de organoides 3D. Así, este modelo ABM elimina cualquier parámetro basado en reglas que modele el comportamiento celular y, en su lugar, se ha introducido un modelo de *Aprendizaje Profundo* (DL, de sus siglas en inglés) que aprende a coordinar el comportamiento temporal de las células para lograr el patrón morfogenético final. Por lo tanto, el ABM solo incluye parámetros con significado físico, avanzando en hacer modelos ABMs más parsimoniosos, mejorando su interpretabilidad y abordando la crítica de reducción de realismo inducida por las leyes tipo heurística típicamente incorporadas. En este marco de trabajo, después de que una célula completa una función biológica en el ABM, la *Red Neuronal* (NN, de sus siglas en inglés) recibe métricas que evalúan el estado de la simulación basada en física y datos que describen el patrón morfogenético final a alcanzar. Posteriormente, la NN activa la siguiente función biológica correspondiente para esa célula.

Para hacer que la NN aprenda un patrón morfogenético, se ha propuesto un enfoque similar al *Aprendizaje por Refuerzo* (RL, de sus siglas en inglés). En este método desarrollado, se simula un ejemplo del patrón deseado, se evalúa la precisión de los resultados utilizando una función de recompensa y se optimiza los parámetros de la NN para maximizar esta función de recompensa. La optimización de la NN se ha realizado a través de un *Algoritmo Genético* (GA, de sus siglas en inglés), aunque se pueden implementar otros algoritmos. Una ventaja de este enfoque es que requiere información mínima sobre los procesos morfogenéticos y solo necesita la información morfológica sobre el patrón final extraído visualmente. Una vez que la NN está entrenada, se ha demostrado el potencial del marco de trabajo para replicar con precisión la morfogénesis de los organoides, tanto los organoides tumorales quísticos como los organoides sólidos del páncreas. En ambos casos, se muestra la precisión de los patrones simulados en comparación con los experimentos *in vitro*, y la robustez de la NN en la coordinación de las funciones celulares a pesar de las variaciones en el tiempo del ciclo celular o el tiempo

de secreción de fluidos, así como la introducción de ruido en estos procesos. Además, se demuestra que esta metodología puede beneficiarse no solo de reproducir un patrón morfogenético final, sino también de reproducir su evolución mediante la incorporación de datos de series temporales. Este método basado en inteligencia artificial aprende los mecanismos intrínsecos del proceso morfogenético, lo que le permite generalizar y reproducir patrones morfogenéticos similares con diferentes procesos dinámicos sin necesidad de optimización de parámetros *ad hoc*.

Además, este trabajo proporciona una demostración de la capacidad del modelo discreto desarrollado para simular el comportamiento similar al de un fluido. Para ello, se ha llevado a cabo un ensayo de cizalladura del modelo de fluido basado en partículas para obtener la curva de flujo estacionaria y se ha ilustrado cómo mantener las propiedades reológicas del fluido como un fluido newtoniano, a pesar de los cambios en la discretización del tamaño de las partículas.

Por lo tanto, este novedoso trabajo muestra el potencial de combinar modelos discretos mediante ABM y ML en biología computacional para reproducir el comportamiento celular. Demuestra el poder sinérgico de ML para aprender a replicar datos de manera efectiva sin necesidad de recalibraciones y resolver el problema inverso de descifrar el comportamiento celular a partir de un patrón final, junto con el modelo mecanicista para obtener interpretaciones biológicas.

Finalmente, el foco se ha trasladado de la microescala a la macroescala para investigar el cáncer de próstata específico del paciente a nivel de órgano (*Capítulo 5*). En la práctica clínica, el cáncer de próstata se diagnostica típicamente mediante *Imagen por Resonancia Magnética* (MRI, de sus siglas en inglés), y la progresión del tumor se monitoriza a través de análisis de sangre rutinarios para medir el biomarcador *Antígeno Prostático Específico* (PSA, de sus siglas en inglés). Sin embargo, existe controversia en torno a este biomarcador como indicador de la evolución del tumor debido a su falta de especificidad y sensibilidad. En muchos casos, los tumores progresan sin variaciones en los niveles de PSA, lo que dificulta su monitoreo. Por lo tanto, este trabajo presenta un nuevo marco computacional que integra un modelo basado en la física con un método de ML para predecir el crecimiento específico del tumor de próstata del paciente en *Gemelos Digitales* (DTs, de sus siglas en inglés) a partir de pruebas de sangre de PSA.

En este trabajo, se ilustra una metodología de construcción de un DT a partir de imágenes de MRI, incorporando información fisiológica. Luego, el DT del paciente se introduce en un modelo basado en física que emplea una representación continua con PDEs para predecir el pronóstico del tumor. Este modelo incorpora aspectos biológicamente relevantes del cáncer de próstata, incluido el aumento en el PSA tisular resultante de la filtración de PSA de las células cancerosas. Además, consideramos la intravasación de PSA desde el tejido hacia la sangre, teniendo en cuenta la distribución espacial de la red vascular, la cual no ha sido considerada en los trabajos existentes en la literatura. Además, el modelo

incorpora una representación matemática del crecimiento tumoral. Para regular la dinámica de crecimiento dentro del modelo basado en física, se ha adoptado un enfoque de ML novedoso no propuesto anteriormente. Este modelo de ML basado en DL considera la geometría 3D de la próstata y las interacciones espaciales entre los datos fisiológicos para controlar el crecimiento tumoral basado en los resultados de pruebas de sangre de PSA específicos del paciente. Es importante destacar que esta metodología desarrollada es independiente de la geometría individual del paciente para garantizar la generalización entre pacientes sin necesidad de reajustes, ofreciendo el potencial para predecir prognósticos de tumores a largo plazo abarcando meses o incluso años con una notable velocidad computacional.

Así, la metodología propuesta se ha validado prediciendo la progresión del tumor de dos pacientes reales, demostrando el potencial del modelo para reproducir con precisión no solo el volumen y la forma clínicos del tumor a lo largo del tiempo, sino también para ofrecer información sobre la agresividad o el riesgo potencial para el paciente dadas sus dinámicas de crecimiento particulares. Para ello, se ha propuesto un método de calibración para el modelo basado en la física para reproducir el crecimiento específico del tumor de un paciente. A través de la modelización mecanicista, se ha demostrado que los parámetros del modelo basado en la física pueden proporcionar información específica del paciente, especialmente en lo que respecta a la agresividad y el riesgo potencial del tumor del paciente. De este modo, el modelo ha identificado un escenario clínico de pacientes con crecimiento tumoral oculto, donde los niveles de PSA en sangre no aumentan a pesar del desarrollo del tumor, constatando el desafío de monitorear la enfermedad únicamente a través del PSA en sangre y permitiendo mediante simulaciones intervenciones terapéuticas más oportunas y personalizadas.

Por tanto, este marco computacional, que integra modelos de aprendizaje automático informados por la física en el DT de la próstata, abre un nuevo camino para predecir la progresión del tumor con DT y explorar los beneficios de la modelización mecanicista y ML. Además, ofrece un enfoque novedoso para crear un posible sustituto computacional de MRIs para monitorizar el pronóstico del cáncer de próstata y reducir la incertidumbre del crecimiento tumoral determinado a partir de pruebas de sangre de PSA. Por lo tanto, este trabajo representa un paso hacia el empleo de la biología computacional para aplicaciones clínicas, ofreciendo predicciones precisas y en tiempo real del pronóstico del tumor específicas del paciente.

7.2 Líneas futuras

Diferentes direcciones de investigación futura han surgido a lo largo de esta tesis tanto para abordar algunas de las limitaciones presentadas en los trabajos como para explorar nuevas oportunidades y caminos abiertos por esta investigación.

Un aspecto que podría beneficiarse de mejoras en los modelos de ABM desarrollados es un refinamiento de la representación de la *Matriz Extracellular* (ECM, de sus siglas en inglés). En las simulaciones, se ha considerado el componente viscoso de la matriz y se ha calculado la fricción de las células con la matriz mediante la ley de Stokes, mientras que se ha despreciado el componente elástico de la misma. Sin embargo, la ECM es un material compuesto que presenta una respuesta reológica viscoelástica. Considerando esto, una representación más precisa de las propiedades mecánicas de la ECM podría mejorar nuestra capacidad para simular el proceso de formación del lumen y comprender mejor sus efectos. Además, incorporar el componente elástico nos permitiría calcular tensiones y entender de forma integral la influencia de las variables ambientales mecánicas involucradas, no solo durante la formación fisiológica, sino también durante su desviación hacia la formación tumoral. Esto podría llevar a una descripción más detallada de las fuerzas mecánicas resultantes de las interacciones célula-ECM y resultar en una mayor heterogeneidad en las fuerzas mecánicas entre las células.

Otro trabajo futuro que se plantea consiste en profundizar en la metodología combinando ABM y ML para la morfogénesis de organoides, presentado en el *Capítulo 4*. En este caso, otro camino a explorar radica en proporcionar a la NN datos físicos. En el modelo desarrollado, a la NN se le proporcionaban métricas de datos que evaluaban el estado de la simulación, facilitando la comparación y correlación con experimentos. Esto permitió la simulación efectiva de los experimentos, ya que la NN controlaba el comportamiento celular en función de las diferencias entre los experimentos y las simulaciones. Sin embargo, otro enfoque implica considerar aspectos físicos y químicos como fuerzas célula-célula, fuerzas célula-matriz o niveles de glucosa, entre otros, para predecir funciones celulares con la NN. En este caso, la función de pérdida integraría la comparación de métricas de datos entre los experimentos y los resultados de simulación. Por lo tanto, con este enfoque, al tener datos suficientes de experimentos para entrenar a la NN y hacer que aprenda múltiples comportamientos, podemos avanzar hacia un ML que integre todos los posibles comportamientos celulares detrás de la morfogénesis de patrones con una interpretación de la causalidad entre la física y el comportamiento celular. Alternativamente, otra perspectiva sobre la combinación de ABM y ML implica el empleo de Convolutional Neural Networks (CNNs) para predecir el comportamiento celular directamente a partir de imágenes experimentales en simulaciones. En este caso, en lugar de proporcionar información basada en la física a la NN, una *Red Neuronal Convolucional* (CNN, de sus siglas en inglés) recibiría la imagen del resultado experimental para guiar el comportamiento celular en el ABM. En este caso, la función de pérdida integraría las diferencias entre las imágenes de entrada y las imágenes de los resultados de simulación del ABM para aproximar los experimentos espacialmente.

Siguiendo esta línea de combinación de modelos discretos con ML, la integración de ABM con esquemas de RL ofrece una gran amplitud de posibilidades

para simular sistemas biológicos. En este enfoque, en el cual los agentes aprenden interactuando con el entorno y otros agentes para maximizar una señal de recompensa, se alinea e integra naturalmente con los ABM. Esta recompensa puede representar el destino celular e incluir la homeostasis, el funcionamiento fisiológico correcto, la supervivencia o, en el caso de las células tumorales, maximizar la proliferación o la invasión. Por lo tanto, este enfoque tiene el potencial de ofrecer nuevas perspectivas para comprender procesos como la migración celular y cómo las señales mecánicas y químicas del entorno la guían, simular la competición de células tumorales y sanas por la obtención de nutrientes como la glucosa u oxígeno, reproducir el crecimiento tumoral o analizar las interacciones mecánicas que regulan los procesos de desarrollo. Por lo tanto, el RL en biología es un campo prometedor aún por investigar.

Centrando la atención en la macroescala a nivel de órgano, otro enfoque para predecir la evolución del cáncer en DTs consistiría en un modelo completamente de ML mediante *Redes Neuronales de Grafos* (Graph Neural Networks (GNNs), de sus siglas en inglés) [345]. En este nuevo modelo, en primer lugar se generaría un grafo a partir del DT, con cada nodo condensando toda la información derivada de la MRI. La GNN luego crearía un embebido del grafo, capturando las interacciones geométricas entre los datos fisiológicos de la MRI. Posteriormente, un modelo de ML con una NN puede hacer predicciones sobre la evolución del tumor considerando la información holística del grafo que representa al DT. Este modelo puede ser entrenado utilizando un conjunto de datos de diagnósticos mediante MRIs para generar el grafo y sus correspondientes MRIs de seguimiento para validar las predicciones. Además, ciertos modelos de GNN también pueden considerar cambios en el grafo con el tiempo, prediciendo no solo el crecimiento del tumor, sino también cambios en la geometría. Por lo tanto, este enfoque promete una mejor predicción del crecimiento del cáncer, aprovechando toda la información condensada en la MRI.

7.3 Contribuciones de la tesis

En esta sección se recogen las contribuciones de esta tesis.

7.3.1 Publicaciones en revistas internacionales revisadas por pares

Las publicaciones en revistas internacionales revisadas por pares logradas durante el período de tesis se enumeran a continuación:

1. Juste-Lanas, Y., Guerrero, P. E., **Camacho-Gomez, D.**, Hervas-Raluy, S., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2022). *Confined Cell Migration*

and Asymmetric Hydraulic Environments to Evaluate the Metastatic Potential of Cancer Cells. ASME. Journal of Biomechanical Engineering, 144(7): 074502. https://doi.org/10.1115/1.4053143.

2. **Camacho-Gomez, D.**, Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2022). *A 3D multi-agent-based model for lumen morphogenesis: the role of the biophysical properties of the extracellular matrix*. Engineering with Computers, 38(5), 4135-4149. https://doi.org/10.1007/s00366-022-01654-1.

3. **Camacho-Gomez, D.**, Sorzabal-Bellido, I., Ortiz-de-Solorzano, C., Garcia-Aznar, J. M., & Gomez-Benito, M. J. (2023). *A hybrid physics-based and data-driven framework for cellular biological systems: Application to the morphogenesis of organoids*. iScience, 26(7). https://doi.org/10.1016/j.isci.2023.107164.

4. **Camacho-Gomez, D.**, Movilla, N., Borau, C., Martin, A., Oñate-Salafranca, C., Pardo, J., Gomez-Benito, M. J., & Garcia-Aznar, J. M. (2024). *An agent-based method to estimate 3D cell migration trajectories from 2D measurements: Quantifying and comparing T vs CAR-T 3D cell migration*. Under review.

5. **Camacho-Gomez, D.**, Borau, C., Garcia-Aznar, J. M., Gomez-Benito, M. J., Girolami, M., & Perez, M. A. (2024). *A physics-informed machine learning digital twin framework for prostate cancer: predicting tumor growth via serum PSA blood tests*. Submitted.

7.3.2 Contribuciones en congresos

Durante la tesis, se han realizado las siguientes comunicaciones en conferencias y reuniones nacionales e internacionales (los títulos subrayados indican los trabajos presentados por el autor).

1. Camacho-Gomez, D., Borau, C., Garcia-Aznar, J. M., Gomez-Benito, M. J., Girolami, M., & Perez, M. A. *A data-driven physics-based model for predicting prostate cancer progression from the PSA blood test*. 9th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS). 3 - 7 junio 2024. Lisboa, Portugal. Participación oral en persona.

2. Gomez-Benito, M. J., Camacho-Gomez, D., Aparicio, R., Garcia-Aznar, J. M. Advancing our understanding of pathological and physiological processes through Computational Mechanobiology. 7th Barcelona VPH Summer School. 7 - 9 junio 2023. Barcelona, España. Coautor de presentación oral en persona.

3. Camacho-Gomez, D., Garcia-Aznar, J. M., Bellido-Sorzabal, I., Ortiz de Solorzano Aurusa, C., & Gomez-Benito, M. J. *A hybrid model of organoids morphogenesis*. International Symposium on Computer Mechanics in Biomechanics and Biomedical Engineering (CMBBE2023). 3 - 5 mayo 2023. Paris, Francia. Participación oral en persona.

4. Camacho-Gomez, D., Guerrero, P. E., Garcia-Aznar, J. M., Gomez-Benito, M. J. *An agent-based model to simulate DIPG migration in microfluidic devices*. Virtual Physiological Human (VPH22). 6 - 9 septiembre 2022. Oporto, Portugal. Participación oral en persona.

5. Camacho-Gomez, D., Perez, A., Borau, C., Gomez-Benito, M. J., Garcia-Aznar, J. M., Perez, M. A. *Simulation of Tumors using Computational Models*. Disfrutar Divulgando Desinteresadamente conferencia. 2022. Zaragoza. Participación oral en persona.

6. Camacho-Gomez, D., Garcia-Aznar, J. M., Gomez-Benito, M. J. *A computational model to study the interplay between cell fluid secretion and matrix stiffness in lumen morphogenesis*. 11th European Solid Mechanics Conference (ESMC22). 4 - 8 julio 2022. Galway, Ireland. Participación oral en persona.

7. Camacho-Gomez, D., Guerrero, P. E., Garcia-Aznar, J. M., Gomez-Benito, M. J. *Hybrid discrete-continuum model to evaluate DIPG cells invasion and matrix degradation*. USACM Thematic Conference on the Role of Mathematical and Computational Modeling in Cancer Research. Participación oral virtual.

8. Camacho-Gomez, D., Garcia-Aznar, J. M., Gomez-Benito, M. J. *Computational model for cell-matrix interaction as a regulator of lumen formation*. X Reunión del Capítulo Español de la Sociedad Europea de Biomecánica (ESB). 25-26 de octubre de 2021, Granada. Participación oral en persona.

9. Camacho-Gomez, D., Garcia-Aznar, J. M., Gomez-Benito, M. J. *A 3D agent-based model to explore DIPG cell invasion in different stiffness matrices*. 6th U.S. National Congress on Computational Mechanics (USNCCM16). 2021. Participación oral virtual.

10. Guerrero, P. E., Camacho-Gomez, D., Mendoza, N., Schuhmacher, A., Gomez-Benito, M. J., Garcia-Aznar, J. M. Mechanobiology of extracellular matrix on diffuse intrinsic pontine glioma (DIPG) tumor growth and migration. American Society of Cell Biology (ASCB). 2021. Coautor de la presentación de póster.

11. Guerrero, P. E., Camacho-Gomez, D., Mendoza, N., Schuhmacher, A., Gomez-Benito, M. J., Garcia-Aznar, J. M. Diffuse Intrinsic Pontine Glioma (DIPG)

tumor growth and migration using 3D biomimetic hydrogels. Organoids in Cancer Research an Agora - EPFL Workshop. 2021. Coautor de la presentación de póster.

12. Goncalves, I., Camacho-Gomez, D., P. E. Guerrero, Gomez-Benito, M. J., Garcia-Aznar, J. M. Mechanobiology of extracellular matrix on the tumor growth. 26th Congress of the European Society of Biomechanics. 2021. Coautor de la presentación oral virtual.

7.3.3 Software de código abierto

Además del desarrollo de los trabajos, se ha contribuido publicando el código de tres de los modelos desarrollados en repositorios abiertos:

- https://github.com/daniel-camacho-gomez/AIOrganoids
- https://github.com/daniel-camacho-gomez/2D_Data-to-3D_Migration
- https://github.com/daniel-camacho-gomez/ProstateNet

7.4 Docencia y supervisiones

El autor ha co-supervisado un trabajo fin de master titulado *Un modelo híbrido basado en física y aprendizaje profundo para la simulación de la migración celular en canales confinados*, un trabajo fin de grado titulado *Estudio de la dinámica celular a través de modelado basado en agentes* y ha supervisado un trabajo de fin de grado titulado *Estudio computacional de la influencia de la anisotropía de la matriz extracelular en la migración celular.*

Además, el autor ha impartido los siguientes cursos:

1. 2022. Precurso "Modelos de agentes en aplicaciones biomédicas." en la XI Reunión del Capítulo Español de la Sociedad Europea de Biomecánica (ESB), Zaragoza.
2. 2022-2024. Curso titulado "Modelos de agentes para la simulación de células" en el seminario interdisciplinar (69723) del Máster en Ingeniería Biomédica de la Universidad de Zaragoza.

7.5 Colaboraciones

Durante el desarrollo de la tesis, el autor ha participado y establecido las siguientes colaboraciones:

Individual and Collective Migration of Immune Cellular Systems (ICoMICS). Este proyecto financiado por la Unión Europea tiene como objetivo desarrollar

un enfoque de modelado capaz de predecir cómo migran las células inmunes terapéuticas, interactúan con el microambiente tumoral y contribuyen a la mejora de los resultados de la inmunoterapia. Los resultados de este trabajo se presentan en el *Capítulo 2*.

Proyecto de Mecanobiología y Análisis de Imágenes en Microfluídica para la Simulación del Crecimiento Tumoral en 3D (TUMOR-ON-CHIP). Este proyecto establece una colaboración entre la Fundación para la Investigación Médica Aplicada (FIMA) de la Universidad de Navarra, expertos en análisis de imágenes y microscopía cuantitativa aplicada al estudio de problemas biomédicos, y la Universidad de Zaragoza. El objetivo es estudiar el crecimiento de organoides tumorales cultivados en dispositivos de microfluídica. Esta colaboración implica un trabajo complementario, donde los modelos computacionales desarrollados por la Universidad de Zaragoza serían validados experimentalmente utilizando las herramientas de microscopía cuantitativa desarrolladas por FIMA. Los resultados de este proyecto conjunto se recogen en el *Capítulo 4*.

Proyecto ProCanAid. Esta iniciativa tiene como objetivo utilizar un gemelo digital de la próstata para ayudar en la detección, diagnóstico y simulación de los efectos y la efectividad de varios tratamientos oncológicos para el cáncer de próstata. El proyecto implica la colaboración entre el Hospital La Fe, que proporciona datos de pacientes, la empresa Quibim, que emplea análisis de imágenes para automatizar la segmentación y extracción de biomarcadores de imágenes, y la Universidad de Zaragoza, que utiliza estos datos para desarrollar modelos computacionales para predecir el crecimiento del cáncer de próstata. El *Capítulo 5* presenta el trabajo realizado dentro de este proyecto.

Además, el autor ha realizado una estancia de investigación de tres meses en el grupo de Estadística Computacional y Aprendizaje Automático del Departamento de Ingeniería de la Universidad de Cambridge, dirigido por el Prof. Mark Girolami. Durante esta estancia, el autor desarrolló el modelo de aprendizaje automático para el proyecto ProCanAid presentado en el *Capítulo 5*.

Adicionalmente, el autor ha contribuido realizando simulaciones computacionales de fluidos en entornos hidráulicos asimétricos en dispositivos microfluídicos en el estudio para evaluar el potencial metastásico de las células cancerosas dentro de canales confinados, con el objetivo de comprender la influencia de los gradientes de presión en la migración celular [49].

Finalmente, el autor también ha participado en el proyecto europeo PRIMAGE. Este proyecto, financiado por la Comisión Europea, cuenta con 16 socios europeos que participan en el consorcio y tiene una duración de implementación de 4 años. Este proyecto propone una plataforma en la nube abierta para apoyar la toma de decisiones en el manejo clínico de dos cánceres pediátricos, el neuroblastoma (NB), el cáncer sólido más frecuente en la primera infancia, y el glioma pontino intrínseco difuso (DIPG), la principal causa de muerte relacionada con tumores cerebrales en niños. La plataforma PRIMAGE implementa los últi-

mos avances en biomarcadores de imágenes in silico y modelado del crecimiento tumoral hacia un diagnóstico, pronóstico y seguimiento de la terapia personalizados.

7.6 Financiación

Esta tesis doctoral ha sido llevada a cabo gracias al apoyo de PRedictive Insilico Multiscale Analytics to support cancer personalized diaGnosis and prognosis, Empowered by imaging biomarkers (PRIMAGE) (G.A. no. 826494) y Procanaid Next Generation EU (Subvención No. PLEC2021-007709). La estancia de investigación en la Universidad de Cambridge también fue respaldada por el programa Erasmus+ gestionado por Campus Iberus y la Fundación Ibercaja-Cai (No IT 5/23).

Appendices

Appendix A

Supplementary Material

What we know is a drop, what we don't know is an ocean.

– **Isaac Newton**

Contents

A.1 Appendix of *Chapter 3 - Lumen morphogenesis*

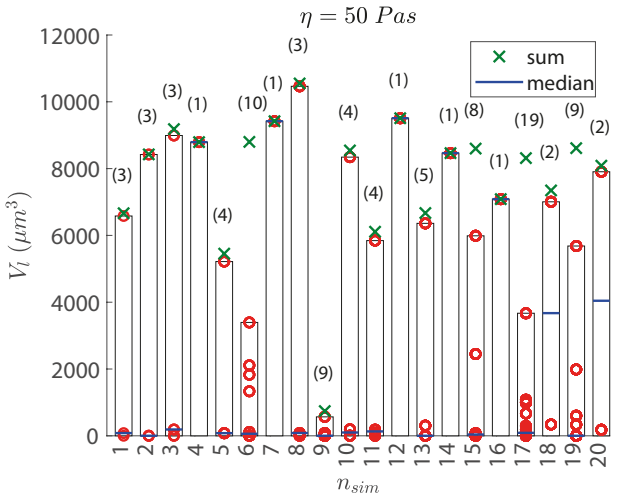

Figure A.1: ***Lumen volume after 7 days for each of the twenty simulations in the intermediate-density matrix*** *($\eta = 50$ Pa s). The numbers in parentheses indicate the number of lumens produced, the red circles represent the discrete volume of each lumen, the green crosses show the sum of the volumes of all lumens in each simulation and the blue line is the median. The bars represent the volume of the largest lumen, which are plotted for the sake of the visualization.*

A.2 Appendix of *Chapter 4 - Organoid morphogenesis*

A.2.1 Murine PDAC cells culture and maintenance

Cystic organoids formed from PDAC93-GFP and solid organoids formed from PM12500-GFP pancreatic ductal adenocarcinoma (PDAC) cells derived from Pdx1-Cre; Kras+/LSL-G12D; Trp53+/LSL-R172H male mice (KPC) were kindly donated by Dr. Mariano Ponz and Dr. Silvestre Vicent. Authentication of cell lines was not required. The cell line used to generate the organoids was a primary line donated by a collaborator, generated from a transgenic mouse model of pancreatic cancer. It is not a commercial line, thus authentication is not possible. The organoid cell line was tested for mycoplasma contamination using MycoAlert® Mycoplasma Detection Kit (Lonza). Cells were thawed from frozen stock in 50 ml Falcon tubes containing 10 ml of DMEM and centrifuged to remove any traces of the cryoprotectant DMSO. Then, the cells were incubated in T75 flasks (TC treated; Nunc EasyFlask, Thermo Scientific) containing 10 ml of DMEM supplemented with 10% FetalClone III (SH30109.02, Cytiva) at 37 °C and 5% CO2. Once cells reached 90%

confluence, cell cultures were passed to a new T75 flask using 0.05% Trypsin-EDTA (25300-096, Gibco) followed by centrifugation and resuspension in $10\,\mathrm{mL}$ of fresh DMEM supplemented with serum and subsequent incubation at 37 °C and 5% CO2.

A.2.2 3D cultures of murine PDAC cells in Matrigel

PDAC93-GFP and PM12500-GFP cells were grown in T75 flasks (TC treated; Nunc EasyFlask, Thermo Scientific) containing $10\,\mathrm{mL}$ of DMEM supplemented with 10% FetalClone III (SH30109.02, Cytiva). Once the cells reached 90% confluence, they were detached using 0.05% trypsin-EDTA (25300-096, Gibco), centrifuged, and resuspended in $1\,\mathrm{mL}$ of fresh DMEM supplemented with serum. Then, the cell suspensions were embedded in $4\,\mathrm{mg/mL}$ growth factor reduced (GFR) Matrigel at a final concentration of $300.000\,\mathrm{cells/mL}$. A total of $20\,\mu\mathrm{L}$ of this cell-hydrogel mixture was added to individual $5\,\mathrm{mm}$ diameter wells of a homemade PDMS device followed by a $15\,\mathrm{min}$ incubation at 37 °C to complete the Matrigel gelation process. Finally, $800\,\mu\mathrm{L}$ of organoid feeding media (a detailed description of the cell culture media used is available in Table A.1) was added to the organoid culture device, and the 3D cell culture was incubated at 37 °C and 5% CO_2 as needed.

Table A.1: *Detailed description of the cell culture media.*

Reagent	Concentration
Advanced DMEM F12	1X, (Base medium)
HEPES	1X
GlutaMax	1X
Penicillin/Streptomycin	1X
A83-01	0.5 µM
mEGF	50 ng/ml
hFGF10	100 ng/ml
Gastrin I	10 nM
mNoggin	100 ng/ml
Y-27632	14 µM
N-acetylcysteine	1.25 mM
Nicotinamide	10 mM
B-27	1X

A.2.3 Cell functions

To reproduce the biological characteristics of tumor pancreatic organoids with lumen, three cell functions are considered within the ABM: proliferation, quiescence, and fluid secretion. Proliferation is modeled by a growth phase, in which the cell progressively increases its volume as a result of DNA replication, and by a mitosis phase, in which the cell divides into two daughter cells. Thus, we track the volume of each cell $V_i(t)$ and calculate the volume growth from [303]:

$$\frac{dV_i(t)}{dt} = \alpha V_i(t), \tag{A.1}$$

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

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

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

A.2.4 Multiphoton Microscopy of PDAC organoids

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

A.2.5 Quantification of 3D PDAC organoid morphology

Nuclei and cytoplasmic GFP segmentation masks were generated using StarDist 3D [346] and the Trainable Weka Segmentation 3D [347] plugin for Fiji, respectively. The StarDist 3D model was trained from scratch for 50 epochs on 45 paired

image patches (patch size: (72,72,32), batch size 1, number of rays: 32, augmentation: true) [348], accelerated using an NVIDIA Quadro P1000 GPU. Afterward, the generated masks were analyzed using a homemade script for Fiji. Briefly, cytoplasmic GFP masks were preprocessed using a GPU-accelerated 3D median filter and binary closing available in the CLIJ2 library [349]). Then, each individual organoid was labeled, and the intersection between the cytoplasmic GFP mask and the nuclear mask was calculated using a logical AND operator. The number of nuclei and morphological descriptors, such as the organoid volume or sphericity, were quantified using MorpholibJ [350]. Statistical analyses and figure generation were performed using MATLAB (MathWorks). The exact number of simulations conducted to test the robustness in Figure 4.4 was indicated in the legend. The R-squared values of the linear regressions in Figure A.5 and Figure A.6 were also specified in the legend.

A.2.6 Learning algorithm

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

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

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

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

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

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

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

– arithmetic crossover: sum of parents:

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

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

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

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

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

$$\text{off. } 10 = \overbrace{\begin{bmatrix}5\\6\\7\\8\end{bmatrix}}^{\text{Parent } B} + mut_{rate} \cdot \begin{bmatrix}rand(-1,1)\\rand(-1,1)\\rand(-1,1)\\rand(-1,1)\end{bmatrix} \odot \overbrace{\begin{bmatrix}5\\6\\7\\8\end{bmatrix}}^{\text{Parent } B} \tag{A.3}$$

A.2.7 Agent-based modeling

Mechanical interactions between agents make them move and change their positions. To describe these interactions we follow our previous work [303]. We illustrate how the cell position $\boldsymbol{x}_c$ and particle position $\boldsymbol{x}_p$ are calculated. Let $\mathbf{N}_c$ be the set of cells $\mathbf{N}_c = \{1, ..., N_c\}$, and let $\mathbf{N}_p$ be the set of particles $\mathbf{N}_p = \{1, ..., N_p\}$. First, the velocity of each *i*-cell $\boldsymbol{v}_{c_i}$ and each *k*-particle $\boldsymbol{v}_{p_k}$ are calculated from the balance of forces:

$$m_{c_i}\frac{d\boldsymbol{v}_{c_i}}{dt} = \sum_{j\in\mathbf{N}_c}\left(\boldsymbol{F}_{c_ic_j}\right) + \sum_{j\in\mathbf{N}_p}\left(\boldsymbol{F}_{c_ip_j}\right) + \boldsymbol{F}_{c_{i_{drag}}} \approx 0, \tag{A.4}$$

$$m_{p_k}\frac{d\boldsymbol{v}_{p_k}}{dt} = \sum_{j\in\mathbf{N}_p}\left(\boldsymbol{F}_{p_kp_j}\right) + \sum_{j\in\mathbf{N}_c}\left(\boldsymbol{F}_{p_kc_j}\right) + \boldsymbol{F}_{p_{k_{drag}}} \approx 0. \tag{A.5}$$

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

$$\boldsymbol{F}_{c_{i_{drag}}} = -6\pi\eta R_{c_i}\boldsymbol{v}_{c_i}, \tag{A.6}$$

$$\boldsymbol{F}_{p_{k_{drag}}} = -6\pi\eta R_p\boldsymbol{v}_{p_k}, \tag{A.7}$$

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

Cell-cell interaction forces are usually modeled as repulsive-attractive forces. The repulsion between cells arises from cell resistance to deformation when their membranes touch, and the attractive forces are the result of the junctions that cells form between themselves through specialized protein complexes ([271]). In the case of particles, the repulsive-attractive forces represent the intermolecular forces in fluids. Accordingly, we modeled the interaction forces $\boldsymbol{F}_{\delta_i\gamma_j}$ (both subindexes δ and γ denote *c* or *p*, depending on whether the *i* and *j* agents are cells (*c*) or particles (*p*)) following [272], as follows:

$$\boldsymbol{F}_{\delta_i\gamma_j} = F_{\delta\gamma}\frac{\boldsymbol{r}_{ij}}{||\boldsymbol{r}_{ij}||} \tag{A.8}$$

where:

$$\boldsymbol{r}_{ij} = \boldsymbol{x}_{\gamma_j} - \boldsymbol{x}_{\delta_i}, \tag{A.9}$$

and:

$$F_{\delta\gamma} = \begin{cases} F_{rep_{\delta\gamma}}\chi(-s)^{3/2}, & s < 0 \text{ (repulsion)}, \\ -F_{adh_{\delta\gamma}}\chi\{(s+s_0)e^{-\lambda(s+s_0)^2} - v_0e^{-\lambda s^2}\}, & s \geq 0 \text{ (adhesion)}. \end{cases} \tag{A.10}$$

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

$$\chi = \frac{R_{\delta_i}}{2}\left(\frac{1}{R_{\delta_i}} + \frac{1}{R_{\gamma_j}}\right), \qquad s = \frac{d - min_{dist}}{R_{\delta_i}}, \tag{A.11}$$

$$x_0 = \sqrt{\frac{1}{2\lambda}}, \qquad v_0 = x_0e^{-\lambda x_0^2}. \tag{A.12}$$

$F_{rep_{\delta\gamma}}$ and $F_{adh_{\delta\gamma}}$ are the strengths of the adhesive and repulsive forces, respectively. $\boldsymbol{r}_{ij}$ is the distance between the centers of the agents, and R_{δ_i} and R_{γ_j} are the radii of the corresponding agents. x_0, v_0 and λ are matching constants, and χ is a geometric correction factor. The value of $min_{dist} = -0.1R_{\delta_i}$ is chosen such that the equilibrium state where the adhesive and repulsive forces are balanced is slightly less than zero, following [272], and $d = ||\boldsymbol{r}_{ij}|| - R_{\delta_i} - R_{\gamma_j}$ is the distance between the agents' surfaces.

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

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

$$\frac{d\boldsymbol{x}_{c_i}(t)}{dt} = \boldsymbol{v}_{c_i}(t) = \frac{1}{6\pi\eta R_{c_i}} \left(\sum_{j \in \mathbf{N}_c} \left(\boldsymbol{F}_{c_i c_j} \right) + \sum_{j \in \mathbf{N}_p} \left(\boldsymbol{F}_{c_i p_j} \right) \right), \tag{A.13}$$

$$\frac{d\boldsymbol{x}_{p_k}(t)}{dt} = \boldsymbol{v}_{p_k}(t) = \frac{1}{6\pi\eta R_{p}} \left(\sum_{j \in \mathbf{N}_p} \left(\boldsymbol{F}_{p_k p_j} \right) + \sum_{j \in \mathbf{N}_c} \left(\boldsymbol{F}_{p_k c_j} \right) \right). \tag{A.14}$$

A.2.8 Cell division

When the volume of the cell reaches twice the value of its initial volume, DNA replication is concluded, and the cell divides. A spatially controlled division is a fundamental condition to maintain the lumen architecture and enhance its growth by enlarging the lumen volume. In this regard, a complex molecularly controlled process regulates the spindle orientation, so mitosis occurs in the plane of the monolayer [260–265]. Here, we distinguish the division between nonpolarized cells and polarized cells. Nonpolarized cells are those that have not yet formed a lumen, and polarized cells are those that belong to a lumen and face it. In the case of nonpolarized cells, the division direction is chosen randomly. Polarized cell division is performed using a random cleavage plane that contains the line that passes through the cell center and the lumen center of mass. The position of the two daughter cells ($\boldsymbol{x}_{daughters}$) are calculated similarly to other models [266, 267] from the center of the parent cell $\boldsymbol{x}_{parent}$ at:

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

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

A.2.9 Fitness function

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

$$\begin{aligned}\text{Fitness}(N_c, V_l, d_{cm}) = 0.5 \cdot 100^{-\frac{(N_c - N_c^t)^2}{2(N_c^t)^2}} + 0.5 \cdot 100^{-\frac{(V_l - V_l^t)^2}{2(V_l^t)^2}} \\ - 250 \left(\frac{e_c}{(V_l^t/10)} \right)^2, \end{aligned} \tag{A.16}$$

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

A.2.10 Eccentricity

In the process of the generation of a cystic organoid, fluid secretion, proliferation, and quiescence must be orchestrated. If cells secrete fluid too early, the preapical patch, which is the closed volume necessary to secrete fluid would not form, so the fluid would leak. Also, if cells secrete too much the lumen would open. Also, if they do not proliferate, they would not create more inner space for the fluid and it would leak. To maintain the structure, there is competition between these processes. Therefore, to make the neural network understand that these cell functions must coordinate to preserve the structure, we introduce a metric called eccentricity that permits us to take into account the structure. This metric refers to the distance between the cells' center of mass and the lumen's center of mass. The lower this value is, the more spherical the organoid is (Figure A.2**a**). Conversely, the higher this value is, the less spherical the organoid is, reaching a certain moment that it opens (Figure A.2**b**). Therefore, the neural network uses this metric to evaluate the state of the cystic structure.

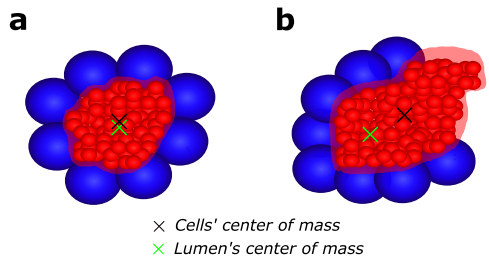

Figure A.2: ***Sketch of the impact of the eccentricity on the structure.*** ***a*** *Case with a low eccentricity.* ***b*** *Case with a high eccentricity, which predicts fluid leakage.*

A.2.11 Organoid seeds

Figure A.3 shows the real 3D cultures of murine pancreatic ductal adenocarcinoma seeds grown in a biomimetic matrix made of Matrigel ($4\,\mathrm{mg/ml}$). Here, we generated a randomly distributed population of organoid seeds ranging from single cells to small clusters of 2-3 cells.

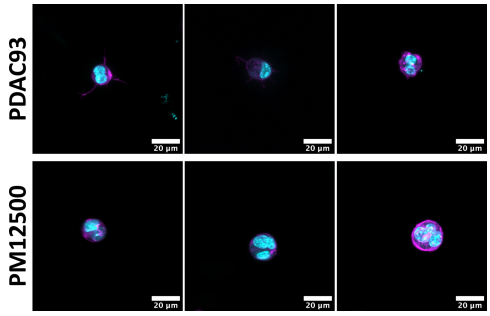

Figure A.3: ***Maximum intensity projection of PDAC93 and PM12500 organoid seeds.*** *Nuclei were stained with DAPI (cyan) and F-actin with phalloidin-TRITC (magenta).*

A.2.12 Simulation parameters for pancreatic tumor cystic organoids

The parameter values used for the simulation of the pancreatic tumor organoids with lumen are presented in Table A.2.

Table A.2: *Parameters of the model for the simulation of pancreatic tumor cystic organoids.*

Parameter	Description	Value	Source
t_{cycle}	Cell cycle time	30 h	Estimated
Δt_q	Quiescence time	6 min	Estimated
Δt_{exo}	Fluid production time	15 min	Estimated
η	Dynamic viscosity of the matrix	20 Pa s	[275]
R_p	Particle radius	2 µm	Estimated
$F_{rep_{cc}}$	Cell-cell repulsive force	-4.80 pN	[272, 303], estimated
$F_{adh_{cc}}$	Cell-cell adhesive force	24 pN	[272, 303], estimated
$F_{rep_{pp}}$	Particle-particle repulsive force	-4.80 pN	[272, 303], estimated
$F_{adh_{pp}}$	Particle-particle adhesive force	24 pN	[272, 303], estimated
$F_{rep_{cp}}$	Cell-particle repulsive force	0.60 pN	[272, 303], estimated
$F_{adh_{cp}}$	Cell-particle adhesive force	0	[272, 303], estimated
λ	Matching constant of the potential function	7	[272, 303]

A.2.13 Characterization of the lumen fluid

We conducted a shear test to validate that our approach represents an actual fluid. In this regard, we used a bulk domain consisting of fluid particles with a radius $R_p = 2$ µm (Figure A.4). The top layer of particles (depicted in black in Figure A.4) was subjected to a velocity u in the X direction, while the bottom layer (depicted in blue in Figure A.4) had fixed displacements in the X and Y directions.

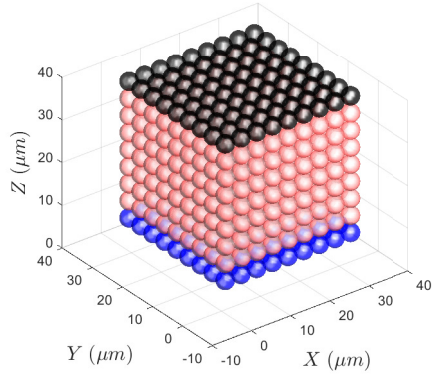

Figure A.4: ***Discretization of the fluid volume with*** $R_p = 2\ \mu m$.

Then, to obtain the steady flow curve to assess the behavior of particles, we set the shear stress experiment with a time duration of $t_{exp} = 30$ min and varied the velocity of the top layer within the values $u = [0.01,\ 0.05,\ 0.1,\ 0.2,\ 0.5]/t_{exp}$ µm/min. Finally, we were able to determine the viscosity of the fluid (η) by measuring the shear stress (τ) as the sum of the reaction forces (F_x) of the particles

forming the bottom XY-plane of the cube of side length l in the X direction $\left(\tau = \frac{\sum F_x}{l^2}\right)$ for each shear rate ($\dot{\gamma} = \partial v / \partial z$).

$$\eta = \frac{\tau}{\frac{\partial v}{\partial z}}. \tag{A.17}$$

We show the linear regression of the shear stress on the deformation rate (Figure A.5). We obtained that the shear stress is linearly proportional to the shear strain, resulting in a constant viscosity of $1.072 \cdot 10^{-3}$ Pa s, with an R-Squared value of 0.9996. Therefore, this analysis confirms that our approach accurately reproduces the expected velocity and shear stress profiles for a Newtonian fluid similar to water ($\sim 10^{-3}$ Pa s) and demonstrates the capability of our method to model real fluids, serving as a minimal model that aims to recapitulate certain features such as mass conservation and momentum exchange. However, it does not rigorously reproduce bulk-scale physical properties.

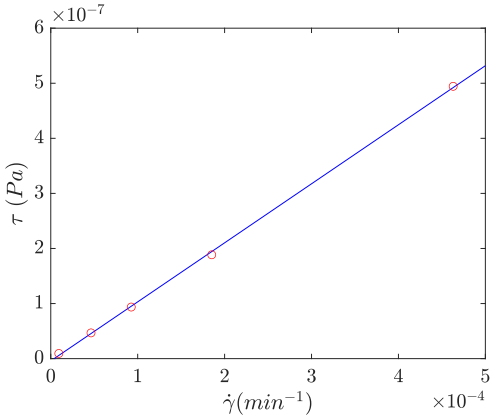

Figure A.5: ***Linear regression of the shear stress on the deformation rate for a particle*** *radius of 2 μm. The calculated viscosity of the fluid is $1.072 \cdot 10^{-3}$ Pa s, with an R-squared value of 0.9996.*

A.2.14 Particle radius influence on the rheological properties of the fluid

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

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

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

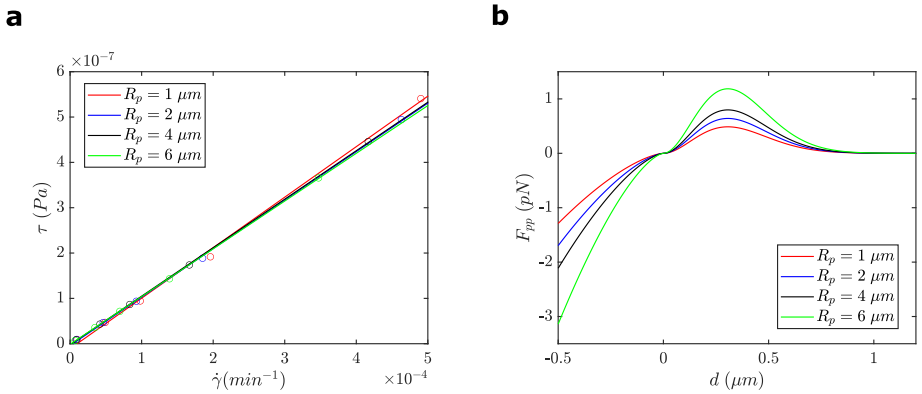

Figure A.6: ***Particle radius influence on the rheological properties of the fluid.*** ***a.*** *Comparison of the linear regression of the shear stress on the deformation rate for a particle radius of* $1\,\mu$m *(*$\eta = 1.112 \cdot 10^{-3}$ Pa s, $R^2 = 0.9979$*),* $2\,\mu$m *(*$\eta = 1.072 \cdot 10^{-3}$ Pa s, $R^2 = 0.9996$*),* $4\,\mu$m *(*$\eta = 1.072 \cdot 10^{-3}$ Pa s, $R^2 = 0.9999$*), and* $6\,\mu$m *(*$\eta = 1.055 \cdot 10^{-3}$ Pa s, $R^2 = 0.9999$*).* ***b.*** *Interacting forces versus the distance between particles for different particle sizes to maintain the same viscosity.*

A.2.15 Particle radius influence on fitness

We analyze the effect of the variation of the particle radius. For this purpose, we simulated an intermediate-size cystic organoid (organoid 2 from Table 4.1 of the *Chapter 4*) with different particle radii to determine how it affects the results. We show the influence of particle radii on the fitness value and computation time while keeping the interacting forces the same as for the case of $R_p = 2\,\mu\text{m}$ (Figure A.7**a**). We found that increasing the particle radius from $1\,\mu\text{m}$ to $2\,\mu\text{m}$ causes an absolute decrease of the fitness value of 0.15% and a relative decrease in the computation of 96.02% (from $1029\,\text{min}$ to $41\,\text{min}$). An increase of the particle radius from $2\,\mu\text{m}$ to $4\,\mu\text{m}$ entails an absolute worsening of the fitness of the approximation of 4.16% and a relative improvement of the computing time of 93.88% (from $41\,\text{min}$ to $3\,\text{min}$). Therefore, this analysis shows that the rheological properties of the fluid are important in the cystic morphogenetic process and require proper interaction to develop the organoid correctly. In particular, making the fluid less viscous (increasing the particle radius) entails a worsening of the simulated organoid. Then, we studied the influence of the particle radii on the fitness value and computation time with the interacting forces obtained in Figure A.6**b** to have the same viscosity as water. In this case (Figure A.7**b**) the fitness is above 0.91 for all particle sizes simulated. However, increasing the particle radius results in a slight decrease in fitness due to a higher increase in lumen volume per fluid secretion compared to smaller particle sizes, leading to a reduced sensitivity in the secretion process. Therefore, a particle radius of $2\,\mu\text{m}$ is optimal for maintaining low computational costs while having sufficient sensitivity in the fluid secretion process.

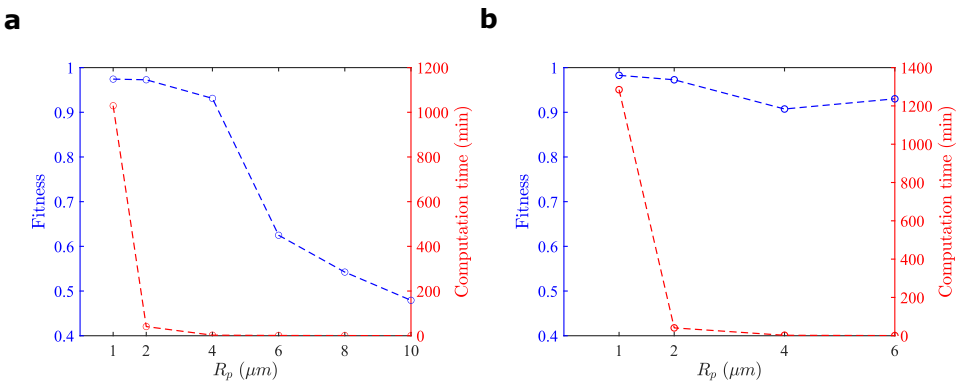

Figure A.7: ***Influence of the particle radius on the fitness value and the computation time. a.*** *The interacting forces for each particle size are the same as for the case of $R_p = 2\,\mu\text{m}$.* ***b.*** *Interacting forces are modified for each particle radius according to Figure A.6**b**.*

A.2.16 Sensitivity analysis

Figure A.8 and Figure A.9 show the sensitivity analysis performed in *Chapter 4*.

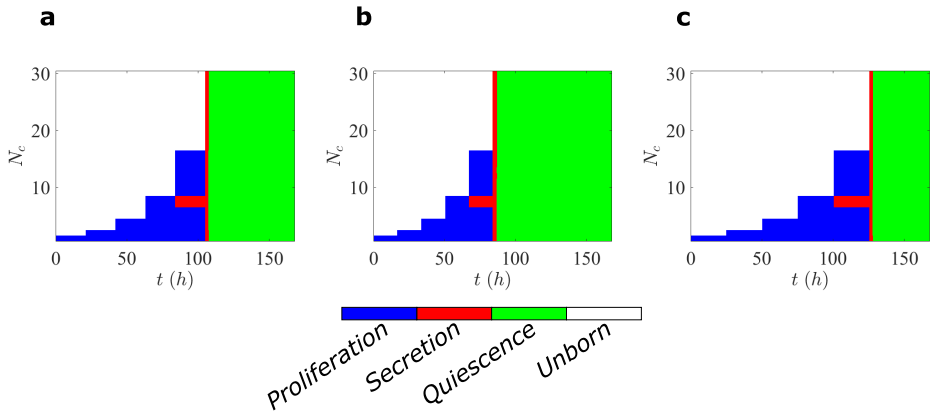

Figure A.8: ***Influence of the cell cycle time on cell coordination.*** ***a.*** $t_{cycle} = 30$ h. ***b.*** $t_{cycle} = 24$ h. ***c.*** $t_{cycle} = 36$ h.

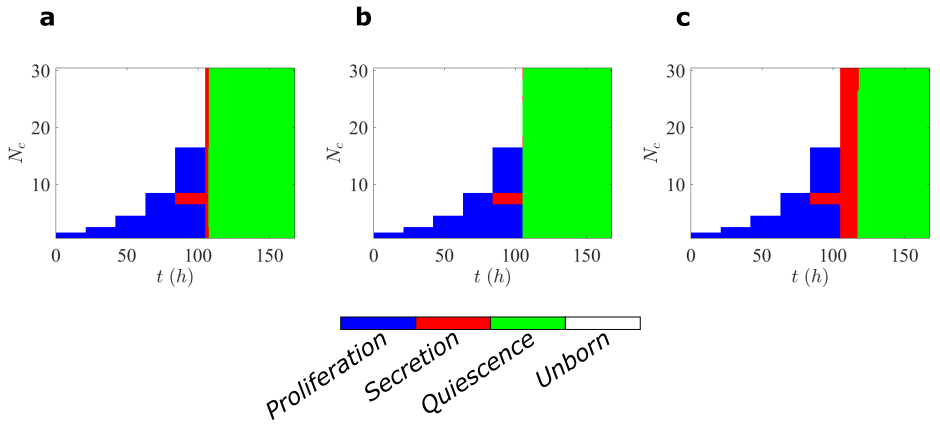

Figure A.9: ***Influence of the fluid production time on cell coordination.*** ***a.*** $\Delta t_{exo} = 15$ min. ***b.*** $\Delta t_{exo} = 6$ min. ***c.*** $\Delta t_{exo} = 60$ min.

A.2.17 Random effects analysis

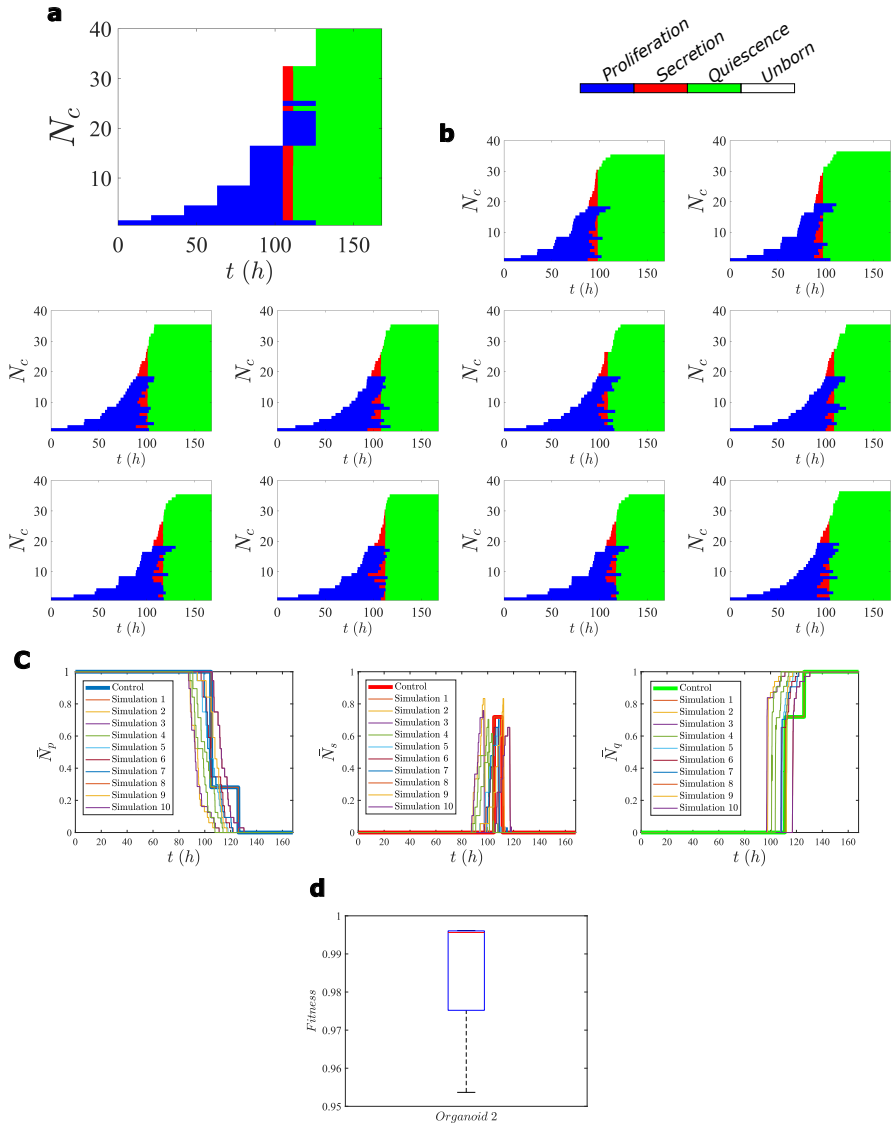

Figure A.10: ***Analysis of random effects of the cell cycle and fluid secretion time parameters.*** ***a.*** *Coordination of cell functions without the variation of parameters.* ***b.*** *Coordination of cell functions of the ten simulations performed with the variation of parameters. Blue color represents proliferation, red color represents secretion, green color represents quiescence, and white color represents an unborn cell.* ***c.*** *Evolution of the normalized number of proliferating cells (left), secreting cells (middle), and quiescent cells (right) without the variation of parameters (control) and with the variation of parameters (simulations 1 to 10).* ***d.*** *Boxplot of the fitness value of the ten simulations performed with the variation of the parameters.*

A.2.18 Application of the framework to solid tumor organoids

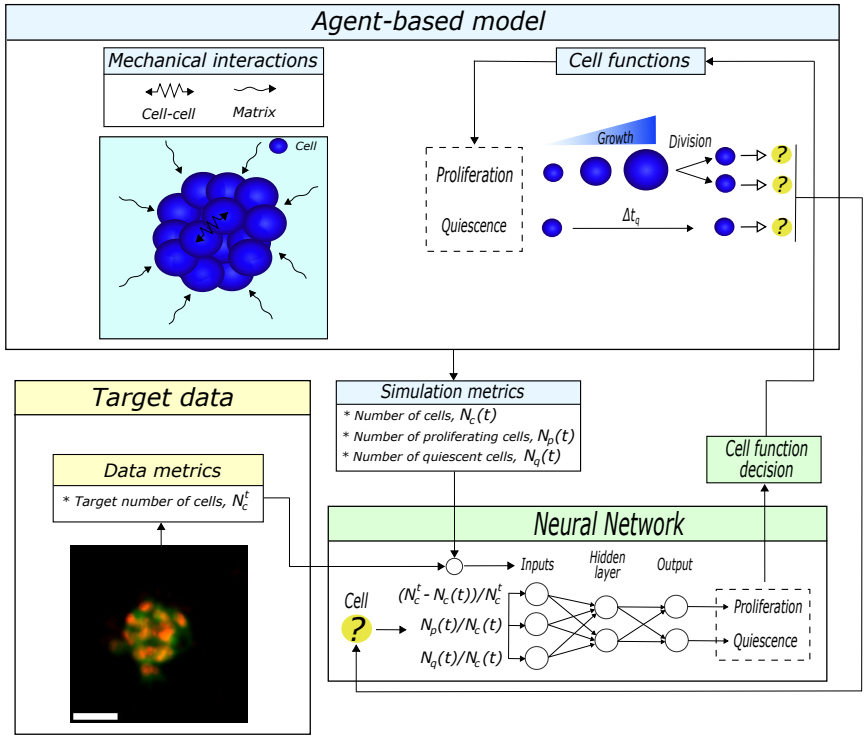

Figure A.11: ***Particularization of the framework for solid organoids.*** *The physics-based model consists in an agent-based model which integrates mechanical interactions and cell functions. Two cell functions are considered; proliferation and quiescence. Δt_q is the minimum period of time that a cell remains quiescent.*

A.2.19 Simulation parameters for pancreatic solid tumor organoids

The parameter values used for the simulation of the pancreatic solid tumor organoids are presented in Table A.3.

Table A.3: ***Parameters of the model for the simulation of pancreatic solid tumor organoids.***

Parameter	Description	Value	Source
t_{cycle}	Cell cycle time	24 h	Estimated
Δt_q	Quiescence time	6 min	Estimated
η	Dynamic viscosity of the matrix	20 Pa s	[275]
$F_{rep_{cc}}$	Cell-cell repulsive force	- 4.80 pN	[272, 303], estimated
$F_{adh_{cc}}$	Cell-cell adhesive force	24 pN	[272, 303], estimated
λ	Matching constant of the potential function	7	[272, 303]

A.2.20 Application of the framework to the evolution of solid organoids

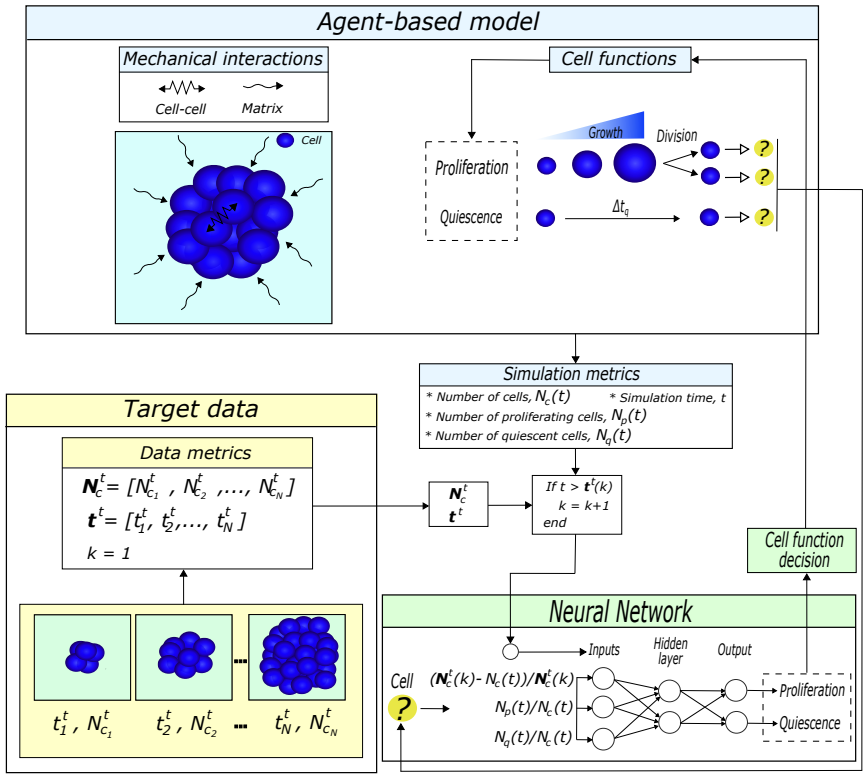

Figure A.12: ***Particularization of the framework for the evolution of solid organoids.*** *Coordination between physics-based and the data-driven models for the simulation of the evolution of solid organoids. In this case, just proliferation and quiescence functions are considered. The target data is specified in different time points.*

A.2.21 Simulation parameters for the evolution of solid organoids

Table A.4: ***Parameters of the model for the simulation of the evolution of solid organoids***

Parameter	Description	Value	Source
t_{cycle}	Cell cycle time	30 h	Estimated
Δt_q	Quiescence time	6 min	Estimated
η	Dynamic viscosity of the matrix	20 Pa s	[275]
$F_{rep_{cc}}$	Cell-cell repulsive force	- 4.80 pN	[272, 303], estimated
$F_{adh_{cc}}$	Cell-cell adhesive force	24 pN	[272, 303], estimated
λ	Matching constant of the potential function	7	[272, 303]

A.2.22 Simulation of the evolution of larger solid organoids

Table A.5: ***Target data for the simulation of the evolution of larger solid organoids.***

	Target number of cells (N_c^t)	
Organoid	7 days	10 days
A	7	500

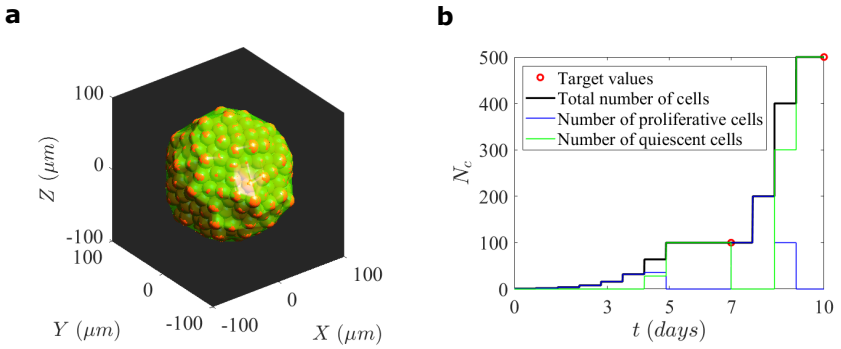

Figure A.13: ***Simulation of the evolution of a large solid tumor organoid.*** ***a.*** *Snapshot at day 10 inserted in a cube of* $200\,\mu\mathrm{m}$ *side. The spheres represent the full cell volume, and the green hull is an estimation of the cell membranes through alpha shapes of the cells with* $\alpha = 2R_c$ ***b.*** *Temporal evolution of the number of cells.*

A.2.23 Cell count analysis of the *in vitro* cysic organoids

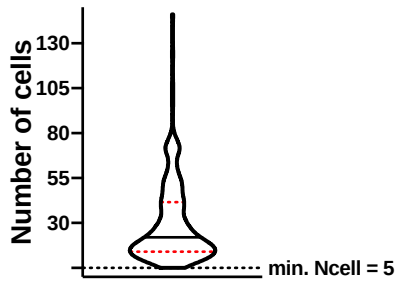

Figure A.14: ***Number of cells for each PADC93 cystic organoid.*** *Black dotted line indicates the minimum number of cells observed on cystic organoids. Solid black line indicates median, dotted red lines indicate 25% and 75% percentiles.*

A.3 Appendix of *Chapter 5 - Prostate cancer*

A.3.1 Physics-based model for prostate cancer

The physics-based model has three main variables, the tissue PSA $P(\boldsymbol{x}, t)$, the serum PSA $P_s(t)$, and the concentration of tumor cells $c_t(\boldsymbol{x}, t)$. First, we introduce a mathematical model to simulate the evolution of the concentration of tumor cells:

$$\frac{\partial c_t(\boldsymbol{x}, t)}{\partial t} = \phi_\theta(\boldsymbol{x}, t)\alpha_t c_t(\boldsymbol{x}, t), \tag{A.18}$$

where $\phi_\theta(\boldsymbol{x}, t)$ is a function approximated by a NN that determines temporarily and spatially the fraction of tumor cells that are proliferating and α_t the tumor growth rate. The initial concentration of tumor cells $c_t(\boldsymbol{x}, t_0)$ is derived from the image data of cellularity $(c(\boldsymbol{x}, t_0))$, the fraction of tumor cells estimated in the patient's biopsy (p_b), and the tumor mask $(T_{mask}(\boldsymbol{x}, t_0))$:

$$c_t(\boldsymbol{x}, t_0) = p_b T_{mask}(\boldsymbol{x}, t_0) c(\boldsymbol{x}, t_0). \tag{A.19}$$

We consider that when the concentration of cells reaches a maximum tissue-carrying capacity equal to 0.8 [339, 340], the tumor mask is expanded to the neighboring voxels. Then, we compute the concentration of tumor cells in the expanded regions. The temporal evolution of the tumor volume $T_{vol}(t)$ can be obtained by integrating tumor mask across the prostate domain (Ω_x):

$$T_{vol}(t) = \int_{\Omega_x} T_{mask}(\boldsymbol{x}, t) d\boldsymbol{x}. \tag{A.20}$$

The tissue PSA increase is a consequence of the PSA leakage of cancer cells. Thus, the tissue PSA $(P(\boldsymbol{x}, t))$ is calculated from:

$$\frac{\partial P(\boldsymbol{x}, t)}{\partial t} = \alpha_p c_t(\boldsymbol{x}, t) - \gamma_p P(\boldsymbol{x}, t) - \frac{m_{w_{gad}}}{m_{w_{PSA}}} k_{trans}(\boldsymbol{x}) \left(P(\boldsymbol{x}, t) - \frac{P_s(t)}{\Omega_{vox}} \right). \tag{A.21}$$

Here, α_p is the tissue PSA production rate by tumor cells, γ_p denotes the natural tissue PSA decay rate, $m_{w_{PSA}}$ represents the molecular weight of PSA [351] $(m_{w_{PSA}} = 26\,\text{Da})$, $m_{w_{gad}}$ denotes the molecular weight of gadolinium [352], and Ω_{vox} the volume of the interchanging region. The gadolinium-based contrast agent used for the MRI scans was Gadoterate Meglumine [353], which has a molecular weight of $753.9\,\text{Da}$. The last term represents the PSA exchange between the tissue and blood in vascularized regions of the prostate, as depicted by the spatial distribution of $k_{trans}(\boldsymbol{x})$. We neglect the diffusive term and assume constant and sufficient blood flow, allowing for a steady concentration of

serum PSA ($P_s(t)$) through the capillaries. The serum PSA's evolution ($P_s(t)$) is determined by integrating the PSA exchange between the tissue and blood:

$$\frac{dP_s(t)}{dt} = \int_{\Omega_x} \frac{m_{w_{gad}}}{m_{w_{PSA}}} k_{trans}(\boldsymbol{x}) \left(P(\boldsymbol{x},t) - \frac{P_s(t)}{\Omega_{vox}} \right) d\boldsymbol{x} - \gamma_s P_s(t) \tag{A.22}$$

with γ_s the natural serum PSA decay rate.

Finally, the temporal scale of the PSA exchange between blood and tissue through $k_{trans}(\boldsymbol{x})$ occurs within seconds, whereas the dynamics of tumor growth typically occur over months. To simulate time frames spanning years and enable follow-ups at any time point, we elaborated a multiscale temporal method to implement this physics-based model (refer to *Multiscale physics-based model* in *Appendix* A.3.2).

A.3.2 Multiscale physics-based model

To solve the physics-based model, we introduce a multiscale model in time, as the time scales of the processes involved in the equations vary significantly. Specifically, the PSA flux between the capillary and the extracellular space operates on a scale of seconds, whereas tumor growth dynamics occur over months. Therefore, to discretize the equations and achieve a stable solution, it is needed a time step of 1×10^{-5} days. However, this renders the simulation of even a single day computationally burdensome. Consequently, we have devised a multiscale model in time to simulate PSA dynamics efficiently. The complete expression for tissue PSA is described as follows:

$$\frac{\partial P(\boldsymbol{x},t)}{\partial t} = \alpha_p c_t(\boldsymbol{x},t) - \gamma_p P(\boldsymbol{x},t) - \frac{m_{w_{gad}}}{m_{w_{PSA}}} k_{trans}(\boldsymbol{x}) \left(P(\boldsymbol{x},t) - \frac{P_s(t)}{\Omega_{vox}} \right). \tag{A.23}$$

where α_p is the tissue PSA production rate by tumor cells, γ_p denotes the natural tissue PSA decay rate, $m_{w_{PSA}}$ represents the molecular weight of PSA ($m_{w_{PSA}} = 26\,\mathrm{Da}$), $m_{w_{gad}}$ denotes the molecular weight of gadolinium, and Ω_{vox} the volume of the interchanging region. The last term represents the PSA exchange between the tissue and blood in vascularized regions of the prostate, as depicted by the spatial distribution of $k_{trans}(\boldsymbol{x})$. Here, we neglect the diffusive term and assume that tissue PSA is already in equilibrium within the prostate, thus eliminating the need for a diffusive term. Additionally, in this equilibrium state, there is no flux between tissue PSA and serum PSA. Therefore, the change in tissue PSA between time steps results solely from the production of PSA by tumor cells and its subsequent decay:

$$\frac{\partial P^*(\boldsymbol{x},t)}{\partial t} = \alpha_p c_t(\boldsymbol{x},t) - \gamma P^*(\boldsymbol{x},t) \rightarrow P^*(\boldsymbol{x},t). \tag{A.24}$$

Then, we calculate the serum PSA that circulates through the blood, equilibrating the flux between the tissue and the blood as a result of variations in tissue PSA due to production or decay:

$$\frac{dP_s(t)}{dt} = \int_{\Omega_x} \frac{m_{w_{gad}}}{m_{w_{PSA}}} k_{trans}(\boldsymbol{x}) \left(P^*(\boldsymbol{x},t) - \frac{P_s(t)}{\Omega_{vox}} \right) d\boldsymbol{x} - \gamma_s P_s(t) \rightarrow P_s(t+\Delta t), \tag{A.25}$$

with Ω_x the prostate domain and γ_s the natural serum PSA decay rate.

Finally, we update the tissue PSA with the value of the serum PSA that balances the flux.

$$\frac{\partial P^*(\boldsymbol{x},t)}{\partial t} = \alpha_p c_t(\boldsymbol{x},t) - \gamma P^*(\boldsymbol{x},t) \rightarrow P^*(\boldsymbol{x},t) \rightarrow P(\boldsymbol{x},t+\Delta t) \tag{A.26}$$

A.3.3 Model calibration

The NN is trained with 20% dropout employing a training and validation dataset. We employed the Python package Optuna [354] for hyperparameter tuning. The loss function used was mean squared error (squared L2 norm). At each epoch, we used a stochastic gradient optimizer (Adam) [355] for the training dataset and then tested over the validation dataset. The training dataset included three different serum PSA values, $\boldsymbol{P}_{s_{train}} = [1.5,\ 3,\ 5]$ ng/mL at 60 days, representing a slow, medium and fast serum PSA increase. The validation dataset comprised three different serum PSA values, $\boldsymbol{P}_{s_{val}} = [1,\ 2,\ 4]$ ng/mL at day 60. We tuned the hyperparameters until minimizing the loss over the validation dataset. The optimized NN hyperparameters resulted in 2 hidden layers of 353 nodes each one, trained with a learning rate of 0.001 and 4585 epochs. The patient-specific parameters of the physics-based model are also optimized using the Optuna framework. In this case, we minimize the error based on the mean squared error (squared L2 norm) between the predicted and the clinical tumor volume and the serum PSA.

A.3.4 Model implementation

The digital twin's geometry generation and discretization were performed in Matlab R2023b. The entire computational model was implemented in Python, utilizing the library PyTorch for the machine learning model. The simulations were performed using an Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz, 32.0 GB RAM, and NVIDIA GeForce GTX 1050 Ti GPU. For Patient A, a full simulation spanning 786 real-world days, with a geometry discretization of $1\ \mathrm{mm}^3$ and 66,691 voxels, takes approximately $25\ \mathrm{s}$ on the GPU and $53\ \mathrm{s}$ on the CPU. For Patient B, a simulation covering 956 real-world days, with a geometry discretization of 1 mm^3 and 52,709 voxels, takes approximately $25\ \mathrm{s}$ on the GPU and $52\ \mathrm{s}$ on the CPU.

Bibliography

[1] I. Abubakar, T. Tillmann, and A. Banerjee. "Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013". In: *Lancet* 385.9963 (2015), pp. 117–171 (↑ 2).

[2] Eurostat. *Causes of death statistics*. Accessed: July 2023. 2020 (↑ 2).

[3] J. Xu, S. L. Murphy, K. D. Kochanek, and E. Arias. "Deaths: Final data for 2019". In: *National Vital Statistics Reports* 70.8 (2021) (↑ 2).

[4] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: *CA: a cancer journal for clinicians* 71.3 (2021), pp. 209–249 (↑ 2).

[5] T. Hofmarcher, P. Lindgren, N. Wilking, and B. Jönsson. "The cost of cancer in Europe 2018". In: *European Journal of Cancer* 129 (2020), pp. 41–49 (↑ 2).

[6] K. R. Yabroff, J. Lund, D. Kepka, and A. Mariotto. "Economic burden of cancer in the United States: estimates, projections, and future research". In: *Cancer epidemiology, biomarkers & prevention* 20.10 (2011), pp. 2006–2014 (↑ 2).

[7] J. Park and K. A. Look. "Health care expenditure burden of cancer care in the United States". In: *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 56 (2019) (↑ 2).

[8] J. Haier and J. Schaefers. "Economic perspective of cancer care and its consequences for vulnerable groups". In: *Cancers* 14.13 (2022), p. 3158 (↑ 2).

[9] A. B. Mariotto, K. Robin Yabroff, Y. Shao, E. J. Feuer, and M. L. Brown. "Projections of the cost of cancer care in the United States: 2010–2020". In: *Journal of the National Cancer Institute* 103.2 (2011), pp. 117–128 (↑ 2).

[10] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal. "Cancer statistics, 2023". In: *Ca Cancer J Clin* 73.1 (2023), pp. 17–48 (↑ 2).

[11] S. Pilleron, E. Soto-Perez-de-Celis, J. Vignat, J. Ferlay, I. Soerjomataram, F. Bray, and D. Sarfati. "Estimated global cancer incidence in the oldest adults in 2018 and projections to 2050". In: *International journal of cancer* 148.3 (2021), pp. 601–608 (↑ 2).

[12] L. Rahib, M. R. Wehner, L. M. Matrisian, and K. T. Nead. "Estimated projection of US cancer incidence and death to 2040". In: *JAMA Network Open* 4.4 (2021), e214708–e214708 (↑ 2).

[13] J. Maddams, M. Utley, and H. Møller. "Projections of cancer prevalence in the United Kingdom, 2010–2040". In: *British journal of cancer* 107.7 (2012), pp. 1195–1202 (↑ 2).

[14] C. B. Blackadar. "Historical review of the causes of cancer". In: *World journal of clinical oncology* 7.1 (2016), p. 54 (↑ 2).

[15] G. Shivashankar, M. Sheetz, and P. Matsudaira. *Mechanobiology*. 2015 (↑ 2).

[16] J. García-Aznar. "Mechanotherapy as an alternative for cancer treatment". In: *Physics of Life Reviews* 47 (2023), pp. 157–158 (↑ 2).

[17] A. D. Theocharis, S. S. Skandalis, C. Gialeli, and N. K. Karamanos. "Extracellular matrix structure". In: *Advanced drug delivery reviews* 97 (2016), pp. 4–27 (↑ 3).

[18] M. Bachmann, S. Kukkurainen, V. P. Hytönen, and B. Wehrle-Haller. "Cell adhesion by integrins". In: *Physiological reviews* 99.4 (2019), pp. 1655–1699 (↑ 3).

[19] M. Barczyk, S. Carracedo, and D. Gullberg. "Integrins". In: *Cell and tissue research* 339 (2010), pp. 269–280 (↑ 3).

[20] A. Hartsock and W. J. Nelson. "Adherens and tight junctions: structure, function and connections to the actin cytoskeleton". In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1778.3 (2008), pp. 660–669 (↑ 3).

[21] D. A. Fletcher and R. D. Mullins. "Cell mechanics and the cytoskeleton". In: *Nature* 463.7280 (2010), pp. 485–492 (↑ 3).

[22] B. Wickstead and K. Gull. "The evolution of the cytoskeleton". In: *Journal of Cell Biology* 194.4 (2011), pp. 513–525 (↑ 3).

[23] T. Svitkina. "The actin cytoskeleton and actin-based motility". In: *Cold Spring Harbor perspectives in biology* 10.1 (2018), a018267 (↑ 3).

[24] J. D. Humphrey, E. R. Dufresne, and M. A. Schwartz. "Mechanotransduction and extracellular matrix homeostasis". In: *Nature reviews Molecular cell biology* 15.12 (2014), pp. 802–812 (↑ 3).

[25] J. M. Tarbell, S. Weinbaum, and R. D. Kamm. "Cellular fluid mechanics and mechanotransduction". In: *Annals of biomedical engineering* 33 (2005), pp. 1719–1723 (↑ 3).

[26] K. A. Gerhold and M. A. Schwartz. "Ion channels in endothelial responses to fluid shear stress". In: *Physiology* 31.5 (2016), pp. 359–369 (↑ 3).

[27] G.-R. Li and X.-L. Deng. "Functional ion channels in stem cells". In: *World journal of stem cells* 3.3 (2011), p. 19 (↑ 3).

[28] W. A. Abou Alaiwi, S. T. Lo, and S. M. Nauli. "Primary cilia: highly sophisticated biological sensors". In: *Sensors* 9.9 (2009), pp. 7003–7020 (↑ 3).

[29] F. Van Roy and G. Berx. "The cell-cell adhesion molecule E-cadherin". In: *Cellular and molecular life sciences* 65 (2008), pp. 3756–3788 (↑ 3).

[30] W. J. Nelson. "Regulation of cell-cell adhesion by the cadherin–catenin complex". In: *Biochemical Society Transactions* 36.2 (2008), pp. 149–155 (↑ 3).

[31] R. Desai, R. Sarpal, N. Ishiyama, M. Pellikka, M. Ikura, and U. Tepass. "Monomeric α-catenin links cadherin to the actin cytoskeleton". In: *Nature cell biology* 15.3 (2013), pp. 261–273 (↑ 3).

[32] X. Trepat, Z. Chen, and K. Jacobson. "Cell migration". In: *Comprehensive Physiology* 2.4 (2012), p. 2369 (↑ 5, 34).

[33] E. Scarpa and R. Mayor. "Collective cell migration in development". In: *Journal of Cell Biology* 212.2 (2016), pp. 143–155 (↑ 5).

[34] P. Mrass and W. Weninger. "Immune cell migration as a means to control immune privilege: lessons from the CNS and tumors". In: *Immunological reviews* 213.1 (2006), pp. 195–212 (↑ 5, 34).

[35] L. Li, Y. He, M. Zhao, and J. Jiang. "Collective cell migration: Implications for wound healing and cancer invasion". In: *Burns & trauma* 1.1 (2013), pp. 2321–3868 (↑ 5, 34).

[36] X. Fu, G. Liu, A. Halim, Y. Ju, Q. Luo, and G. Song. "Mesenchymal stem cell migration and tissue repair". In: *Cells* 8.8 (2019), p. 784 (↑ 5, 34).

[37] D. Pantaloni, C. L. Clainche, and M.-F. Carlier. "Mechanism of actin-based motility". In: *Science* 292.5521 (2001), pp. 1502–1506 (↑ 5).

[38] M. Schaks, G. Giannone, and K. Rottner. "Actin dynamics in cell migration". In: *Essays in biochemistry* 63.5 (2019), pp. 483–495 (↑ 5).

[39] J. V. Small, T. Stradal, E. Vignal, and K. Rottner. "The lamellipodium: where motility begins". In: *Trends in cell biology* 12.3 (2002), pp. 112–120 (↑ 5).

[40] R. Aguilar-Cuenca, A. Juanes-García, and M. Vicente-Manzanares. "Myosin II in mechanotransduction: master and commander of cell migration, morphogenesis, and cancer". In: *Cellular and molecular life sciences* 71 (2014), pp. 479–492 (↑ 5).

[41] P. Friedl, K. Wolf, and J. Lammerding. "Nuclear mechanics during cell migration". In: *Current opinion in cell biology* 23.1 (2011), pp. 55–64 (↑ 5).

[42] P. Agarwal and R. Zaidel-Bar. "Diverse roles of non-muscle myosin II contractility in 3D cell migration". In: *Essays in Biochemistry* 63.5 (2019), pp. 497–508 (↑ 5).

[43] C.-M. Lo, H.-B. Wang, M. Dembo, and Y.-l. Wang. "Cell movement is guided by the rigidity of the substrate". In: *Biophysical journal* 79.1 (2000), pp. 144–152 (↑ 6).

[44] B. C. Isenberg, P. A. DiMilla, M. Walker, S. Kim, and J. Y. Wong. "Vascular smooth muscle cell durotaxis depends on substrate stiffness gradient strength". In: *Biophysical journal* 97.5 (2009), pp. 1313–1322 (↑ 6).

[45] M. Raab, J. Swift, P. D. P. Dingal, P. Shah, J.-W. Shin, and D. E. Discher. "Crawling from soft to stiff matrix polarizes the cytoskeleton and phosphoregulates myosin-II heavy chain". In: *Journal of Cell Biology* 199.4 (2012), pp. 669–683 (↑ 6).

[46] B. J. DuChez, A. D. Doyle, E. K. Dimitriadis, and K. M. Yamada. "Durotaxis by human cancer cells". In: *Biophysical journal* 116.4 (2019), pp. 670–683 (↑ 6).

[47] H. V. Prentice-Mott, C.-H. Chang, L. Mahadevan, T. J. Mitchison, D. Irimia, and J. V. Shah. "Biased migration of confined neutrophil-like cells in asymmetric hydraulic environments". In: *Proceedings of the National Academy of Sciences* 110.52 (2013), pp. 21006–21011 (↑ 6).

[48] H. D. Moreau et al. "Macropinocytosis overcomes directional bias in dendritic cells due to hydraulic resistance and facilitates space exploration". In: *Developmental cell* 49.2 (2019), pp. 171–188 (↑ 6).

[49] Y. Juste-Lanas, P. E. Guerrero, D. Camacho-Gómez, S. Hervás-Raluy, J. M. García-Aznar, and M. J. Gomez-Benito. "Confined cell migration and asymmetric hydraulic environments to evaluate the metastatic potential of cancer cells". In: *Journal of Biomechanical Engineering* 144.7 (2022), p. 074502 (↑ 6, 35, 124, 137).

[50] Y. Li, K. Konstantopoulos, R. Zhao, Y. Mori, and S. X. Sun. "The importance of water and hydraulic pressure in cell dynamics". In: *Journal of cell science* 133.20 (2020), jcs240341 (↑ 6).

[51] A. M. Turing. "The chemical basis of morphogenesis". In: *Bulletin of mathematical biology* 52 (1990), pp. 153–197 (↑ 7).

[52] K. W. Rogers and A. F. Schier. "Morphogen gradients: from generation to interpretation". In: *Annual review of cell and developmental biology* 27 (2011), pp. 377–407 (↑ 7).

[53] D. E. Ingber. "Mechanical control of tissue morphogenesis during embryological development". In: *International Journal of Developmental Biology* 50.2-3 (2003), pp. 255–266 (↑ 7).

[54] P. Patwari and R. T. Lee. "Mechanical control of tissue morphogenesis". In: *Circulation research* 103.3 (2008), pp. 234–243 (↑ 7).

[55] T. Lecuit and P.-F. Lenne. "Cell surface mechanics and the control of cell shape, tissue patterns and morphogenesis". In: *Nature reviews Molecular cell biology* 8.8 (2007), pp. 633–644 (↑ 7).

[56] C. Guillot and T. Lecuit. "Mechanics of epithelial tissue homeostasis and morphogenesis". In: *Science* 340.6137 (2013), pp. 1185–1189 (↑ 7).

[57] D. N. Clarke and A. C. Martin. "Actin-based force generation and cell adhesion in tissue morphogenesis". In: *Current Biology* 31.10 (2021), R667–R680 (↑ 7).

[58] J. M. Sawyer, J. R. Harrell, G. Shemer, J. Sullivan-Brown, M. Roh-Johnson, and B. Goldstein. "Apical constriction: a cell shape change that can drive morphogenesis". In: *Developmental biology* 341.1 (2010), pp. 5–19 (↑ 7).

[59] B. M. Gumbiner. "Cell adhesion: the molecular basis of tissue architecture and morphogenesis". In: *Cell* 84.3 (1996), pp. 345–357 (↑ 7).

[60] B. M. Gumbiner. "Regulation of cadherin-mediated adhesion in morphogenesis". In: *Nature reviews Molecular cell biology* 6.8 (2005), pp. 622–634 (↑ 7).

[61] D. Godt and U. Tepass. "Drosophila oocyte localization is mediated by differential cadherin-based adhesion". In: *Nature* 395.6700 (1998), pp. 387–391 (↑ 7).

[62] S. Wacker, K. Grimm, T. Joos, and R. Winklbauer. "Development and control of tissue separation at gastrulation in Xenopus". In: *Developmental Biology* 224.2 (2000), pp. 428–439 (↑ 7).

[63] B. N. Mason, J. P. Califano, and C. A. Reinhart-King. "Matrix stiffness: a regulator of cellular behavior and tissue formation". In: *Engineering biomaterials for regenerative medicine: novel technologies for clinical applications* (2012), pp. 19–37 (↑ 7).

[64] W. P. Daley and K. M. Yamada. "ECM-modulated cellular dynamics as a driving force for tissue morphogenesis". In: *Current opinion in genetics & development* 23.4 (2013), pp. 408–414 (↑ 7).

[65] M. Daems, H. M. Peacock, and E. A. Jones. "Fluid flow as a driver of embryonic morphogenesis". In: *Development* 147.15 (2020), dev185579 (↑ 7).

[66] P. McMillen and S. A. Holley. "Integration of cell-cell and cell-ECM adhesion in vertebrate morphogenesis". In: *Current opinion in cell biology* 36 (2015), pp. 48–53 (↑ 7).

[67] D. T. Butcher, T. Alliston, and V. M. Weaver. "A tense situation: forcing tumour progression". In: *Nature Reviews Cancer* 9.2 (2009), pp. 108–122 (↑ 8).

[68] M. J. Paszek et al. "Tensional homeostasis and the malignant phenotype". In: *Cancer cell* 8.3 (2005), pp. 241–254 (↑ 8, 50, 51, 53, 59, 61, 69, 79, 116, 128).

[69] P. P. Provenzano, D. R. Inman, K. W. Eliceiri, J. G. Knittel, L. Yan, C. T. Rueden, J. G. White, and P. J. Keely. "Collagen density promotes mammary tumor initiation and progression". In: *BMC medicine* 6.1 (2008), pp. 1–15 (↑ 8).

[70] H. A. Messal, S. Alt, R. M. Ferreira, C. Gribben, V. M.-Y. Wang, C. G. Cotoi, G. Salbreux, and A. Behrens. "Tissue curvature and apicobasal mechanical tension imbalance instruct cancer morphogenesis". In: *Nature* 566.7742 (2019), pp. 126–130 (↑ 8).

[71] J. L. Leight, A. P. Drain, and V. M. Weaver. "Extracellular matrix remodeling and stiffening modulate tumor phenotype and treatment response". In: *Annual Review of Cancer Biology* 1 (2017), pp. 313–334 (↑ 9).

[72] B. Deng, Z. Zhao, W. Kong, C. Han, X. Shen, and C. Zhou. "Biological role of matrix stiffness in tumor growth and treatment". In: *Journal of Translational Medicine* 20.1 (2022), p. 540 (↑ 9).

[73] R. K. Jain, J. D. Martin, and T. Stylianopoulos. "The role of mechanical forces in tumor growth and therapy". In: *Annual review of biomedical engineering* 16 (2014), pp. 321–346 (↑ 9).

[74] M. Kalli and T. Stylianopoulos. "Defining the role of solid stress and matrix stiffness in cancer cell proliferation and metastasis". In: *Frontiers in oncology* 8 (2018), p. 55 (↑ 9).

[75] T. P. Padera, B. R. Stoll, J. B. Tooredman, D. Capen, E. d. Tomaso, and R. K. Jain. "Cancer cells compress intratumour vessels". In: *Nature* 427.6976 (2004), pp. 695–695 (↑ 9).

[76] V. P. Chauhan et al. "Angiotensin inhibition enhances drug delivery and potentiates chemotherapy by decompressing tumour blood vessels". In: *Nature communications* 4.1 (2013), p. 2516 (↑ 9).

[77] B. J. Moeller, R. A. Richardson, and M. W. Dewhirst. "Hypoxia and radiotherapy: opportunities for improved outcomes in cancer treatment". In: *Cancer and Metastasis Reviews* 26 (2007), pp. 241–248 (↑ 9).

[78] M. R. Horsman and J. Overgaard. "The impact of hypoxia and its modification of the outcome of radiotherapy". In: *Journal of radiation research* 57.S1 (2016), pp. i90–i98 (↑ 9).

[79] H. Salavati, C. Debbaut, P. Pullens, and W. Ceelen. "Interstitial fluid pressure as an emerging biomarker in solid tumors". In: *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* (2022), p. 188792 (↑ 9).

[80] S. J. Lunt, A. Fyles, R. P. Hill, and M. Milosevic. "Interstitial fluid pressure in tumors: therapeutic barrier and biomarker of angiogenesis". In: *Future Oncology* 4.6 (2008) (↑ 9).

[81] M. K. Gupta and R.-Y. Qin. "Mechanism and its regulation of tumor-induced angiogenesis". In: *World journal of gastroenterology: WJG* 9.6 (2003), p. 1144 (↑ 9).

[82] P. Carmeliet. "VEGF as a key mediator of angiogenesis in cancer". In: *Oncology* 69.Suppl. 3 (2005), pp. 4–10 (↑ 9).

[83] T. Whiteside. "The tumor microenvironment and its role in promoting tumor growth". In: *Oncogene* 27.45 (2008), pp. 5904–5912 (↑ 9).

[84] F. R. Balkwill, M. Capasso, and T. Hagemann. "The tumor microenvironment at a glance". In: *Journal of cell science* 125.23 (2012), pp. 5591–5596 (↑ 9).

[85] G. P. Gupta and J. Massagué. "Cancer metastasis: building a framework". In: *Cell* 127.4 (2006), pp. 679–695 (↑ 9).

[86] M. Kapałczyńska et al. "2D and 3D cell cultures–a comparison of different types of cancer cell cultures". In: *Archives of Medical Science* 14.4 (2018), pp. 910–919 (↑ 10).

[87] F. Pampaloni, E. G. Reynaud, and E. H. Stelzer. "The third dimension bridges the gap between cell culture and live tissue". In: *Nature reviews Molecular cell biology* 8.10 (2007), pp. 839–845 (↑ 10).

[88] E. Fennema, N. Rivron, J. Rouwkema, C. van Blitterswijk, and J. De Boer. "Spheroid culture as a tool for creating 3D complex tissues". In: *Trends in biotechnology* 31.2 (2013), pp. 108–115 (↑ 10).

[89] M. T. Santini and G. Rainaldi. "Three-dimensional spheroid model in tumor biology". In: *Pathobiology* 67.3 (1999), pp. 148–157 (↑ 10).

[90] L.-B. Weiswald, D. Bellet, and V. Dangles-Marie. "Spherical cancer models in tumor biology". In: *Neoplasia* 17.1 (2015), pp. 1–15 (↑ 10).

[91] M. Huch, J. A. Knoblich, M. P. Lutolf, and A. Martinez-Arias. "The hope and the hype of organoid research". In: *Development* 144.6 (2017), pp. 938–941 (↑ 10).

[92] S. Nath and G. R. Devi. "Three-dimensional culture systems in cancer research: Focus on tumor spheroid model". In: *Pharmacology & therapeutics* 163 (2016), pp. 94–108 (↑ 11).

[93] V. Velasco, S. A. Shariati, and R. Esfandyarpour. "Microtechnology-based methods for organoid models". In: *Microsystems & nanoengineering* 6.1 (2020), p. 76 (↑ 11).

[94] M. Mehling and S. Tay. "Microfluidic cell culture". In: *Current opinion in Biotechnology* 25 (2014), pp. 95–102 (↑ 11).

[95] S. Halldorsson, E. Lucumi, R. Gómez-Sjöberg, and R. M. Fleming. "Advantages and challenges of microfluidic cell culture in polydimethylsiloxane devices". In: *Biosensors and Bioelectronics* 63 (2015), pp. 218–231 (↑ 11).

[96] P. Van Liedekerke, M. Palm, N. Jagiella, and D. Drasdo. "Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results". In: *Computational particle mechanics* 2 (2015), pp. 401–444 (↑ 13, 21, 51).

[97] J. Metzcar, Y. Wang, R. Heiland, and P. Macklin. "A review of cell-based computational modeling in cancer biology". In: *JCO clinical cancer informatics* 2 (2019), pp. 1–13 (↑ 13, 21).

[98] D. G. Mallet and L. G. De Pillis. "A cellular automata model of tumor–immune system interactions". In: *Journal of theoretical biology* 239.3 (2006), pp. 334–350 (↑ 13).

[99] B. M. Rubenstein and L. J. Kaufman. "The role of extracellular matrix in glioma invasion: a cellular Potts model approach". In: *Biophysical journal* 95.12 (2008), pp. 5661–5680 (↑ 14).

[100] J. A. Engelberg, A. Datta, K. E. Mostov, and C. A. Hunt. "MDCK cystogenesis driven by cell stabilization within computational analogues". In: *PLoS Computational Biology* 7.4 (2011), e1002030 (↑ 14, 51, 53, 75).

[101] R. F. van Oers, E. G. Rens, D. J. LaValley, C. A. Reinhart-King, and R. M. Merks. "Mechanical cell-matrix feedback explains pairwise and collective endothelial cell behavior in vitro". In: *PLoS computational biology* 10.8 (2014), e1003774 (↑ 14).

[102] A. Shirinifard, J. S. Gens, B. L. Zaitlen, N. J. Popławski, M. Swat, and J. A. Glazier. "3D multi-cell simulation of tumor growth and angiogenesis". In: *PloS one* 4.10 (2009), e7190 (↑ 14).

[103] M. Scianna, L. Preziosi, and K. Wolf. "A Cellular Potts Model simulating cell migration on and in matrix environments". In: *Mathematical Biosciences & Engineering* 10.1 (2012), pp. 235–261 (↑ 14).

[104] L. Zhang, C. G. Strouthos, Z. Wang, and T. S. Deisboeck. "Simulating brain tumor heterogeneity with a multiscale agent-based model: linking molecular signatures, phenotypes and expansion rate". In: *Mathematical and computer modelling* 49.1-2 (2009), pp. 307–319 (↑ 14).

[105] D. K. Schlüter, I. Ramis-Conde, and M. A. Chaplain. "Computational modeling of single-cell migration: the leading role of extracellular matrix fibers". In: *Biophysical journal* 103.6 (2012), pp. 1141–1151 (↑ 14).

[106] P. Macklin, M. E. Edgerton, A. M. Thompson, and V. Cristini. "Patient-calibrated agent-based modelling of ductal carcinoma in situ (DCIS): from microscopic measurements to macroscopic predictions of clinical progression". In: *Journal of theoretical biology* 301 (2012), pp. 122–140 (↑ 14).

[107] H. Kempf, H. Hatzikirou, M. Bleicher, and M. Meyer-Hermann. "In silico analysis of cell cycle synchronisation effects in radiotherapy of tumour spheroids". In: *PLoS computational biology* 9.11 (2013), e1003295 (↑ 14).

[108] B. S. Gardiner, K. K. Wong, G. R. Joldes, A. J. Rich, C. W. Tan, A. W. Burgess, and D. W. Smith. "Discrete element framework for modelling extracellular matrix, deformable cells and subcellular components". In: *PLoS computational biology* 11.10 (2015), e1004544 (↑ 14, 15).

[109] G. Schaller and M. Meyer-Hermann. "Multicellular tumor spheroid in an off-lattice Voronoi-Delaunay cell model". In: *Physical Review E* 71.5 (2005), p. 051910 (↑ 15).

[110] P. Van Liedekerke et al. "A quantitative high-resolution computational mechanics cell model for growing and regenerating tissues". In: *Biomechanics and modeling in mechanobiology* 19 (2020), pp. 189–220 (↑ 15, 51).

[111] C. M. Glen, M. L. Kemp, and E. O. Voit. "Agent-based modeling of morphogenetic systems: Advantages and challenges". In: *PLoS computational biology* 15.3 (2019), e1006577 (↑ 15, 29, 75).

[112] P. Richmond, R. Chisholm, P. Heywood, M. K. Chimeh, and M. Leach. "FLAME GPU 2: A framework for flexible and performant agent based simulation on GPUs". In: *Software: Practice and Experience* (2023) (↑ 15).

[113] E. A. Sarapata and L. De Pillis. "A comparison and catalog of intrinsic tumor growth models". In: *Bulletin of mathematical biology* 76 (2014), pp. 2010–2024 (↑ 16).

[114] R. Molina-Peña and M. M. Álvarez. "A simple mathematical model based on the cancer stem cell hypothesis suggests kinetic commonalities in solid tumor growth". In: *PloS one* 7.2 (2012), e26233 (↑ 16).

[115] K. P. Wilkie and P. Hahnfeldt. "Modeling the dichotomy of the immune response to cancer: cytotoxic effects and tumor-promoting inflammation". In: *Bulletin of mathematical biology* 79.6 (2017), pp. 1426–1448 (↑ 16).

[116] R. Sachs, L. Hlatky, and P. Hahnfeldt. "Simple ODE models of tumor growth and anti-angiogenic or radiation treatment". In: *Mathematical and Computer Modelling* 33.12-13 (2001), pp. 1297–1305 (↑ 16).

[117] L. DePillis, A. Eladdadi, and A. Radunskaya. "Modeling cancer-immune responses to therapy". In: *Journal of pharmacokinetics and pharmacodynamics* 41 (2014), pp. 461–478 (↑ 16).

[118] R. R. Sarkar and S. Banerjee. "Cancer self remission and tumor stability–a stochastic approach". In: *Mathematical Biosciences* 196.1 (2005), pp. 65–81 (↑ 16).

[119] L. DePillis, H. Savage, and A. Radunskaya. "Mathematical model of colorectal cancer with monoclonal antibody treatments". In: *arXiv preprint arXiv:1312.3023* (2013) (↑ 16).

[120] A. P. Browning, T. D. Lewin, R. E. Baker, P. K. Maini, E. G. Moros, J. Chaudell, H. M. Byrne, and H. Enderling. "Predicting Radiotherapy Patient Outcomes with Real-Time Clinical Data Using Mathematical Modelling". In: *Bulletin of Mathematical Biology* 86.2 (2024), p. 19 (↑ 17).

[121] A.-J. Koch and H. Meinhardt. "Biological pattern formation: from basic mechanisms to complex structures". In: *Reviews of modern physics* 66.4 (1994), p. 1481 (↑ 17).

[122] S. Ferreira Jr, M. Martins, and M. Vilela. "Reaction-diffusion model for the growth of avascular tumor". In: *Physical Review E* 65.2 (2002), p. 021907 (↑ 17).

[123] M. A. Chaplain and G. Lolas. "Mathematical modelling of cancer invasion of tissue: dynamic heterogeneity." In: *Networks Heterog. Media* 1.3 (2006), pp. 399–439 (↑ 17).

[124] J. A. Weis, M. I. Miga, L. R. Arlinghaus, X. Li, A. B. Chakravarthy, V. Abramson, J. Farley, and T. E. Yankeelov. "A mechanically coupled reaction–diffusion model for predicting the response of breast tumors to neoadjuvant chemotherapy". In: *Physics in Medicine & Biology* 58.17 (2013), p. 5851 (↑ 17).

[125] X. Chen, R. M. Summers, and J. Yao. "Kidney tumor growth prediction by coupling reaction–diffusion and biomechanical model". In: *IEEE Transactions on Biomedical Engineering* 60.1 (2012), pp. 169–173 (↑ 17).

[126] P. Ciarletta, D. Ambrosi, G. Maugin, and L. Preziosi. "Mechano-transduction in tumour growth modelling". In: *The European Physical Journal E* 36 (2013), pp. 1–9 (↑ 17).

[127] T. Roose, P. A. Netti, L. L. Munn, Y. Boucher, and R. K. Jain. "Solid stress generated by spheroid growth estimated using a linear poroelasticity model". In: *Microvascular research* 66.3 (2003), pp. 204–212 (↑ 17).

[128] M. Fraldi and A. R. Carotenuto. "Cells competition in tumor growth poroelasticity". In: *Journal of the Mechanics and Physics of Solids* 112 (2018), pp. 345–367 (↑ 17).

[129] G. Lorenzo, M. A. Scott, K. Tew, T. J. Hughes, Y. J. Zhang, L. Liu, G. Vilanova, and H. Gomez. "Tissue-scale, personalized modeling and simulation of prostate cancer growth". In: *Proceedings of the National Academy of Sciences* 113.48 (2016), E7663–E7671 (↑ 17, 100).

[130] J. Xu, G. Vilanova, and H. Gomez. "A mathematical model coupling tumor growth and angiogenesis". In: *PloS one* 11.2 (2016), e0149422 (↑ 17).

[131] G. Vilanova, I. Colominas, and H. Gomez. "A mathematical model of tumour angiogenesis: growth, regression and regrowth". In: *Journal of The Royal Society Interface* 14.126 (2017), p. 20160918 (↑ 17, 18).

[132] K. A. Rejniak and A. R. Anderson. "Hybrid models of tumor growth". In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 3.1 (2011), pp. 115–125 (↑ 19).

[133] Z. Wang, J. D. Butner, R. Kerketta, V. Cristini, and T. S. Deisboeck. "Simulating cancer growth with multiscale agent-based modeling". In: *Seminars in cancer biology*. Vol. 30. Elsevier. 2015, pp. 70–78 (↑ 19).

[134] I. Ramis-Conde, M. A. Chaplain, and A. R. Anderson. "Mathematical modelling of cancer cell invasion of tissue". In: *Mathematical and Computer Modelling* 47.5-6 (2008), pp. 533–545 (↑ 19).

[135] J. Jeon, V. Quaranta, and P. T. Cummings. "An off-lattice hybrid discrete-continuum model of tumor growth and invasion". In: *Biophysical journal* 98.1 (2010), pp. 37–47 (↑ 19).

[136] G. Vilanova, M. Burés, I. Colominas, and H. Gomez. "Computational modelling suggests complex interactions between interstitial flow and tumour angiogenesis". In: *Journal of The Royal Society Interface* 15.146 (2018), pp. 2–0180415 (↑ 19, 50).

[137] J. S. Lowengrub, H. B. Frieboes, F. Jin, Y.-L. Chuang, X. Li, P. Macklin, S. M. Wise, and V. Cristini. "Nonlinear modelling of cancer: bridging the gap between cells and tumours". In: *Nonlinearity* 23.1 (2009), R1 (↑ 19).

[138] H. Hatzikirou et al. "Integrative physical oncology". In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 4.1 (2012), pp. 1–14 (↑ 19).

[139] A. Chauviere, H. Hatzikirou, I. G. Kevrekidis, J. S. Lowengrub, and V. Cristini. "Dynamic density functional theory of solid tumor growth: preliminary models". In: *AIP advances* 2.1 (2012) (↑ 19).

[140] C. Borau et al. "A multiscale orchestrated computational framework to reveal emergent phenomena in neuroblastoma". In: *Computer Methods and Programs in Biomedicine* 241 (2023), p. 107742 (↑ 19, 20, 110).

[141] H. Rocha, R. Almeida, E. Lima, A. Resende, J. Oden, and T. Yankeelov. "A hybrid three-scale model of tumor growth". In: *Mathematical Models and Methods in Applied Sciences* 28.01 (2018), pp. 61–93 (↑ 20).

[142] T. Erol, A. F. Mendi, and D. Doğan. "The digital twin revolution in healthcare". In: *2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT)*. IEEE. 2020, pp. 1–7 (↑ 20).

[143] E. A. Stahlberg et al. "Exploring approaches for predictive cancer patient digital twins: Opportunities for collaboration and innovation". In: *Frontiers in Digital Health* 4 (2022), p. 1007784 (↑ 20).

[144] A. Chaudhuri et al. "Predictive digital twin for optimizing patient-specific radiotherapy regimens under uncertainty in high-grade gliomas". In: *Frontiers in Artificial Intelligence* 6 (2023) (↑ 20).

[145] B. Björnsson et al. "Digital twins to personalize medicine". In: *Genome medicine* 12 (2020), pp. 1–4 (↑ 20).

[146] T. Hernandez-Boussard, P. Macklin, E. J. Greenspan, A. L. Gryshuk, E. Stahlberg, T. Syeda-Mahmood, and I. Shmulevich. "Digital twins for predictive oncology will be a paradigm shift for precision cancer care". In: *Nature medicine* 27.12 (2021), pp. 2065–2066 (↑ 20).

[147] S. Subramanian, A. Ghafouri, K. M. Scheufele, N. Himthani, C. Davatzikos, and G. Biros. "Ensemble inversion for brain tumor growth models with mass effect". In: *IEEE Transactions on Medical Imaging* 42.4 (2022), pp. 982–995 (↑ 21).

[148] A. Platas-López, A. Guerra-Hernández, M. Quiroz-Castellanos, and N. Cruz-Ramirez. "A survey on agent-based modelling assisted by machine learning". In: *Expert Systems* (2023), e13325 (↑ 21, 30).

[149] J. Bergstra and Y. Bengio. "Random search for hyper-parameter optimization." In: *Journal of machine learning research* 13.2 (2012) (↑ 22).

[150] S. Ruiz-Arrebola, A. Tornero-López, D. Guirado, M. Villalobos, and A. Lallena. "An on-lattice agent-based Monte Carlo model simulating the growth kinetics of multicellular tumor spheroids". In: *Physica Medica* 77 (2020), pp. 194–203 (↑ 22).

[151] S. Huang et al. "Validation and application of a fast Monte Carlo algorithm for assessing the clinical impact of approximations in analytical dose calculations for pencil beam scanning proton therapy". In: *Medical physics* 45.12 (2018), pp. 5631–5642 (↑ 22).

[152] H. Badry, L. Oufni, H. Ouabi, and H. Hirayama. "Monte Carlo dose calculation for HDR brachytherapy source using EGS5 code". In: *Radiation Physics and Chemistry* 150 (2018), pp. 76–81 (↑ 22).

[153] A. L. Fielding. "Monte-Carlo techniques for radiotherapy applications I: Introduction and overview of the different Monte-Carlo codes". In: *Journal of Radiotherapy in Practice* 22 (2023), e80 (↑ 22).

[154] S. Hervas-Raluy et al. "Tumour growth: An approach to calibrate parameters of a multiphase porous media model based on in vitro observations of Neuroblastoma spheroid growth in a hydrogel microenvironment". In: *Computers in Biology and Medicine* 159 (2023), p. 106895 (↑ 22).

[155] E. A. Lima, D. Faghihi, R. Philley, J. Yang, J. Virostko, C. M. Phillips, and T. E. Yankeelov. "Bayesian calibration of a stochastic, multiscale agent-based model for predicting in vitro tumor growth". In: *PLoS Computational Biology* 17.11 (2021), e1008845 (↑ 23).

[156] P. Qods, J. Arkat, and Y. Batmani. "Optimal administration strategy in chemotherapy regimens using multi-drug cell-cycle specific tumor growth models". In: *Biomedical Signal Processing and Control* 86 (2023), p. 105221 (↑ 24).

[157] C. Akasiadis et al. "Parallel model exploration for tumor treatment simulations". In: *Computational Intelligence* 38.4 (2022), pp. 1379–1401 (↑ 24).

[158] D. Knopoff, D. Fernández, G. Torres, and C. Turner. "A mathematical method for parameter estimation in a tumor growth model". In: *Computational and Applied Mathematics* 36 (2017), pp. 733–748 (↑ 24).

[159] S. Badillo et al. "An introduction to machine learning". In: *Clinical pharmacology & therapeutics* 107.4 (2020), pp. 871–885 (↑ 25).

[160] J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones. "A guide to machine learning for biologists". In: *Nature reviews Molecular cell biology* 23.1 (2022), pp. 40–55 (↑ 25).

[161] M. Torrisi, G. Pollastri, and Q. Le. "Deep learning methods in protein structure prediction". In: *Computational and Structural Biotechnology Journal* 18 (2020), pp. 1301–1310 (↑ 25).

[162] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni. "Comparing different supervised machine learning algorithms for disease prediction". In: *BMC medical informatics and decision making* 19.1 (2019), pp. 1–16 (↑ 25).

[163] H. Seo, M. Badiei Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, and L. Xing. "Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications". In: *Medical physics* 47.5 (2020), e148–e167 (↑ 25).

[164] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman. "Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities". In: *Information Fusion* 50 (2019), pp. 71–91 (↑ 25).

[165] M. Alber et al. "Integrating machine learning and multiscale modeling — perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences". In: *NPJ digital medicine* 2.1 (2019), p. 115 (↑ 25).

[166] R. E. Baker, J.-M. Pena, J. Jayamohan, and A. Jérusalem. "Mechanistic models versus machine learning, a fight worth fighting for the biological community?" In: *Biology letters* 14.5 (2018), p. 20170660 (↑ 25).

[167] G. C. Peng et al. "Multiscale modeling meets machine learning: What can we learn?" In: *Archives of Computational Methods in Engineering* 28 (2021), pp. 1017–1037 (↑ 25).

[168] Y. Zhu and N. Zabaras. "Bayesian deep convolutional encoder-decoder networks for surrogate modeling and uncertainty quantification". In: *Journal of Computational Physics* 366 (2018), pp. 415–447 (↑ 26).

[169] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris. "Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data". In: *Journal of Computational Physics* 394 (2019), pp. 56–81 (↑ 26).

[170] C. Angione, E. Silverman, and E. Yaneske. "Using machine learning as a surrogate model for agent-based simulations". In: *Plos one* 17.2 (2022), e0263150 (↑ 26).

[171] F. Lamperti, A. Roventini, and A. Sani. "Agent-based model calibration using machine learning surrogates". In: *Journal of Economic Dynamics and Control* 90 (2018), pp. 366–389 (↑ 26).

[172] L. Liang, M. Liu, C. Martin, and W. Sun. "A deep learning approach to estimate stress distribution: a fast and accurate surrogate of finite-element analysis". In: *Journal of The Royal Society Interface* 15.138 (2018), p. 20170844 (↑ 26).

[173] E. Karabelas et al. "Global sensitivity analysis of four chamber heart hemodynamics using surrogate models". In: *IEEE Transactions on Biomedical Engineering* 69.10 (2022), pp. 3216–3223 (↑ 26).

[174] I. M. Gherman, Z. S. Abdallah, W. Pang, T. E. Gorochowski, C. S. Grierson, and L. Marucci. "Bridging the gap between mechanistic biological models and machine learning surrogates". In: *PLoS Computational Biology* 19.4 (2023), e1010988 (↑ 26).

[175] M. Raissi, P. Perdikaris, and G. E. Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *Journal of Computational physics* 378 (2019), pp. 686–707 (↑ 26).

[176] D. N. Tanyu, J. Ning, T. Freudenberg, N. Heilenkötter, A. Rademacher, U. Iben, and P. Maass. "Deep learning methods for partial differential equations and related parameter identification problems". In: *Inverse Problems* 39.10 (2023), p. 103001 (↑ 26).

[177] D. Sainz-DeMena, M. Pérez, and J. M. García-Aznar. "Exploring the potential of Physics-Informed Neural Networks to extract vascularization data from DCE-MRI in the presence of diffusion". In: *Medical Engineering & Physics* 123 (2024), p. 104092 (↑ 27).

[178] A. Zhu. "Accelerating Parameter Inference in Diffusion-Reaction Models of Glioblastoma Using Physics-Informed Neural Networks". In: *SIAM Undergraduate Resources Online* (2022) (↑ 27).

[179] R. Z. Zhang, I. Ezhov, M. Balcerak, A. Zhu, B. Wiestler, B. Menze, and J. Lowengrub. "Personalized Predictions of Glioblastoma Infiltration: Mathematical Models, Physics-Informed Neural Networks and Multimodal Scans". In: *arXiv preprint arXiv:2311.16536* (2023) (↑ 27).

[180] J. N. Kutz. "Machine learning for parameter estimation". In: *Proceedings of the National Academy of Sciences* 120.12 (2023), e2300990120 (↑ 27).

[181] T. Gaskin, G. A. Pavliotis, and M. Girolami. "Neural parameter calibration for large-scale multiagent models". In: *Proceedings of the National Academy of Sciences* 120.7 (2023), e2216415120 (↑ 27).

[182] W. Zhang, A. Valencia, and N.-B. Chang. "Synergistic integration between machine learning and agent-based modeling: A multidisciplinary review". In: *IEEE Transactions on Neural Networks and Learning Systems* 34.5 (2021), pp. 2170–2190 (↑ 27).

[183] N. Sivakumar, C. Mura, and S. M. Peirce. "Innovations in integrating machine learning and agent-based modeling of biomedical systems". In: *Frontiers in systems biology* 2 (2022), p. 959665 (↑ 27).

[184] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018 (↑ 29).

[185] D. Bray. "Intracellular signalling as a parallel distributed process". In: *Journal of theoretical biology* 143.2 (1990), pp. 215–231 (↑ 29).

[186] J. Vohradsky. "Neural model of the genetic network". In: *Journal of Biological Chemistry* 276.39 (2001), pp. 36168–36173 (↑ 29).

[187] P. Gerlee and A. R. Anderson. "An evolutionary hybrid cellular automaton model of solid tumour growth". In: *Journal of theoretical biology* 246.4 (2007), pp. 583–603 (↑ 29).

[188] T. T. Tang, J. A. Zawaski, K. N. Francis, A. A. Qutub, and M. W. Gaber. "Image-based classification of tumor type and growth rate using machine learning: a preclinical study". In: *Scientific reports* 9.1 (2019), p. 12529 (↑ 30).

[189] Ș. Ionescu, C. Delcea, N. Chiriță, and I. Nica. "Exploring the Use of Artificial Intelligence in Agent-Based Modeling Applications: A Bibliometric Study". In: *Algorithms* 17.1 (2024), p. 21 (↑ 30).

[190] M. C. Bischoff, S. Lieb, R. Renkawitz-Pohl, and S. Bogdan. "Filopodia-based contact stimulation of cell migration drives tissue morphogenesis". In: *Nature communications* 12.1 (2021), p. 791 (↑ 34).

[191] T. L. Whiteside. "Immune suppression in cancer: effects on immune cells, mechanisms and future therapeutic intervention". In: *Seminars in cancer biology*. Vol. 16,1. Elsevier. 2006, pp. 3–15 (↑ 34).

[192] S. Stock, M. Schmitt, and L. Sellner. "Optimizing manufacturing protocols of chimeric antigen receptor T cells for improved anticancer immunotherapy". In: *International journal of molecular sciences* 20.24 (2019), p. 6223 (↑ 34).

[193] H. Dana, G. M. Chalbatani, S. A. Jalali, H. R. Mirzaei, S. A. Grupp, E. R. Suarez, C. Rapôso, and T. J. Webster. "CAR-T cells: Early successes in blood cancer and challenges in solid tumors". In: *Acta Pharmaceutica Sinica B* 11.5 (2021), pp. 1129–1147 (↑ 34).

[194] N. Gagelmann, K. Riecken, C. Wolschke, C. Berger, F. A. Ayuk, B. Fehse, and N. Kröger. "Development of CAR-T cell therapies for multiple myeloma". In: *Leukemia* 34.9 (2020), pp. 2317–2332 (↑ 34).

[195] F. Sun et al. "Bispecific BCMA/CD24 CAR-T cells control multiple myeloma growth". In: *Nature communications* 15.1 (2024), p. 615 (↑ 34).

[196] N. Kramer, A. Walzl, C. Unger, M. Rosner, G. Krupitza, M. Hengstschläger, and H. Dolznig. "In vitro cell migration and invasion assays". In: *Mutation Research/Reviews in Mutation Research* 752.1 (2013), pp. 10–24 (↑ 35).

[197] H.-C. Chen. "Boyden chamber assay". In: *Cell migration*. Springer, 2005, pp. 15–22 (↑ 35).

[198] C.-C. Liang, A. Y. Park, and J.-L. Guan. "In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro". In: *Nature protocols* 2.2 (2007), pp. 329–333 (↑ 35).

[199] R. Weigert, M. Sramkova, L. Parente, P. Amornphimoltham, and A. Masedunskas. "Intravital microscopy: a novel tool to study cell biology in living animals". In: *Histochemistry and cell biology* 133.5 (2010), pp. 481–491 (↑ 35).

[200] E. Beerling, L. Ritsma, N. Vrisekoop, P. W. Derksen, and J. van Rheenen. "Intravital microscopy: new insights into metastasis of tumors". In: *Journal of cell science* 124.3 (2011), pp. 299–310 (↑ 35).

[201] E. M. Hillman, C. B. Amoozegar, T. Wang, A. F. McCaslin, M. B. Bouchard, J. Mansfield, and R. M. Levenson. "In vivo optical imaging and dynamic contrast methods for biomedical research". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1955 (2011), pp. 4620–4643 (↑ 35).

[202] H. Yamaguchi, J. Wyckoff, and J. Condeelis. "Cell migration in tumors". In: *Current opinion in cell biology* 17.5 (2005), pp. 559–564 (↑ 35).

[203] F. Entschladen, T. L. Drell IV, K. Lang, K. Masur, D. Palm, P. Bastian, B. Niggemann, and K. S. Zaenker. "Analysis methods of human cell migration". In: *Experimental cell research* 307.2 (2005), pp. 418–426 (↑ 35).

[204] D. Palm, K. Lang, B. Brandt, K. S. Zaenker, and F. Entschladen. "In vitro and in vivo imaging of cell migration: two interdepending methods to unravel metastasis formation". In: *Seminars in Cancer Biology* 15.5 (2005), pp. 396–404 (↑ 35).

[205] Y. Shin, S. Han, J. S. Jeon, K. Yamamoto, I. K. Zervantonakis, R. Sudo, R. D. Kamm, and S. Chung. "Microfluidic assay for simultaneous culture of multiple cell types on surfaces or within hydrogels". In: *Nature protocols* 7.7 (2012), pp. 1247–1259 (↑ 35, 39, 40).

[206] O. Moreno-Arotzena, G. Mendoza, M. Cóndor, T. Rüberg, and J. García-Aznar. "Inducing chemotactic and haptotactic cues in microfluidic devices for three-dimensional in vitro assays". In: *Biomicrofluidics* 8.6 (2014) (↑ 35, 40).

[207] V. Olivares, M. Cóndor, C. Del Amo, J. Asín, C. Borau, and J. M. García-Aznar. "Image-based characterization of 3D collagen networks and the effect of embedded cells". In: *Microscopy and Microanalysis* 25.4 (2019), pp. 971–981 (↑ 35).

[208] M. Anguiano et al. "Characterization of three-dimensional cancer cell migration in mixed collagen-Matrigel scaffolds using microfluidics and image analysis". In: *PloS one* 12.2 (2017), e0171417 (↑ 35).

[209] J. Plou, Y. Juste-Lanas, V. Olivares, C. Del Amo, C. Borau, and J. García-Aznar. "From individual to collective 3D cancer dissemination: roles of collagen concentration and TGF-β". In: *Scientific reports* 8.1 (2018), p. 12723 (↑ 35).

[210] I. K. Zervantonakis, S. K. Hughes-Alford, J. L. Charest, J. S. Condeelis, F. B. Gertler, and R. D. Kamm. "Three-dimensional microfluidic model for tumor cell intravasation and endothelial barrier function". In: *Proceedings of the National Academy of Sciences* 109.34 (2012), pp. 13515–13520 (↑ 35).

[211] Y. Juste-Lanas et al. "3D collagen migration patterns reveal a SMAD3-dependent and TGF-β1-independent mechanism of recruitment for tumour-associated fibroblasts in lung adenocarcinoma". In: *British Journal of Cancer* 128.6 (2023), pp. 967–981 (↑ 35).

[212] O. Moreno-Arotzena, C. Borau, N. Movilla, M. Vicente-Manzanares, and J. García-Aznar. "Fibroblast migration in 3D is controlled by haptotaxis in a non-muscle myosin II-dependent manner". In: *Annals of Biomedical Engineering* 43 (2015), pp. 3025–3039 (↑ 35).

[213] N. Movilla, C. Borau, C. Valero, and J. García-Aznar. "Degradation of extracellular matrix regulates osteoblast migration: A microfluidic-based study". In: *Bone* 107 (2018), pp. 10–17 (↑ 35).

[214] L. Boneschansker, J. Yan, E. Wong, D. M. Briscoe, and D. Irimia. "Microfluidic platform for the quantitative analysis of leukocyte migration signatures". In: *Nature communications* 5.1 (2014), p. 4787 (↑ 35).

[215] H. Cavanagh, D. Kempe, J. K. Mazalo, M. Biro, and R. G. Endres. "T cell morphodynamics reveal periodic shape oscillations in three-dimensional migration". In: *Journal of the Royal Society Interface* 19.190 (2022), 20220 081 (↑ 35).

[216] L. Saucedo-Mora, M. Á. Sanz, F. J. Montáns, and J. M. Benítez. "A simple agent-based hybrid model to simulate the biophysics of glioblastoma multiforme cells and the concomitant evolution of the oxygen field". In: *Computer Methods and Programs in Biomedicine* 246 (2024), p. 108046 (↑ 36).

[217] I. G. Gonçalves and J. M. Garcia-Aznar. "Extracellular matrix density regulates the formation of tumour spheroids through cell migration". In: *PLoS computational biology* 17.2 (2021), e1008764 (↑ 36, 75).

[218] G. Bretti and A. De Gaetano. "An Agent-Based Interpretation of Leukocyte Chemotaxis in Cancer-on-Chip Experiments". In: *Mathematics* 10.8 (2022), p. 1338 (↑ 36).

[219] Q. Peng, F. J. Vermolen, and D. Weihs. "Physical confinement and cell proximity increase cell migration rates and invasiveness: A mathematical model of cancer cell invasion through flexible channels". In: *Journal of the Mechanical Behavior of Biomedical Materials* 142 (2023), p. 105843 (↑ 36).

[220] C. Valero, H. Amaveda, M. Mora, and J. M. García-Aznar. "Combined experimental and computational characterization of crosslinked collagen-based hydrogels". In: *PLoS One* 13.4 (2018), e0195820 (↑ 37).

[221] N. Movilla, I. G. Gonçalves, C. Borau, and J. M. García-Aznar. "A novel integrated experimental and computational approach to unravel fibroblast motility in response to chemical gradients in 3D collagen matrices". In: *Integrative Biology* 14.8-12 (2022), pp. 212–227 (↑ 40).

[222] D. M. Davies et al. "Flexible targeting of ErbB dimers that drive tumorigenesis by using genetically engineered T cells". In: *Molecular medicine* 18.4 (2012), pp. 565–576 (↑ 40).

[223] H. Li et al. "Targeting brain lesions of non-small cell lung cancer by enhancing CCL2-mediated CAR-T cell migration". In: *Nature Communications* 13.1 (2022), p. 2154 (↑ 45).

[224] C. Del Amo, V. Olivares, M. Cóndor, A. Blanco, J. Santolaria, J. Asín, C. Borau, and J. M. García-Aznar. "Matrix architecture plays a pivotal role in 3D osteoblast migration: The effect of interstitial fluid flow". In: *Journal of the mechanical behavior of biomedical materials* 83 (2018), pp. 52–62 (↑ 45).

[225] J. Huth et al. "TimeLapseAnalyzer: multi-target analysis for live-cell imaging and time-lapse microscopy". In: *Computer methods and programs in biomedicine* 104.2 (2011), pp. 227–234 (↑ 46).

[226] Â. Carvalho, T. Esteves, P. Quelhas, and F. J. Monteiro. "MobilityAnalyser: A novel approach for automatic quantification of cell mobility on periodic patterned substrates using brightfield microscopy images". In: *Computer methods and programs in biomedicine* 162 (2018), pp. 61–67 (↑ 46).

[227] A. Datta, D. M. Bryant, and K. E. Mostov. "Molecular regulation of lumen morphogenesis". In: *Current Biology* 21.3 (2011), R126–R136 (↑ 50, 55, 79).

[228] K. Bäumler et al. "Fluid–structure interaction simulations of patient-specific aortic dissection". In: *Biomechanics and modeling in mechanobiology* 19.5 (2020), pp. 1607–1628 (↑ 50).

[229] A. Liu, R. K. Wang, K. L. Thornburg, and S. Rugonyi. "Dynamic variation of hemodynamic shear stress on the walls of developing chick hearts: computational models of the heart outflow tract". In: *Engineering with Computers* 25.1 (2009), pp. 73–86 (↑ 50).

[230] C. A. Figueroa, S. Baek, C. A. Taylor, and J. D. Humphrey. "A computational framework for fluid–solid-growth modeling in cardiovascular simulations". In: *Computer methods in applied mechanics and engineering* 198.45-46 (2009), pp. 3583–3602 (↑ 50).

[231] E. Latorre et al. "Active superelasticity in three-dimensional epithelia of controlled shape". In: *Nature* 563.7730 (2018), pp. 203–208 (↑ 50, 52, 79).

[232] A. Navis and M. Bagnat. "Developing pressures: fluid forces driving morphogenesis". In: *Current opinion in genetics & development* 32 (2015), pp. 24–30 (↑ 50, 52, 79).

[233] S. B. Dahl-Jensen et al. "Deconstructing the principles of ductal network formation in the pancreas". In: *PLoS biology* 16.7 (2018), e2002842 (↑ 50).

[234] S. Checa, M. K. Rausch, A. Petersen, E. Kuhl, and G. N. Duda. "The emergence of extracellular matrix mechanics and cell traction forces as important regulators of cellular self-organization". In: *Biomechanics and modeling in mechanobiology* 14.1 (2015), pp. 1–13 (↑ 50, 79).

[235] A. Rubiano et al. "Viscoelastic properties of human pancreatic tumors and in vitro constructs to mimic mechanical properties". In: *Acta biomaterialia* 67 (2018), pp. 331–340 (↑ 50).

[236] J. Debnath and J. S. Brugge. "Modelling glandular epithelial cancers in three-dimensional cultures". In: *Nature Reviews Cancer* 5.9 (2005), pp. 675–688 (↑ 51, 69).

[237] H. Aram F., K. Alec C., S. Ben Z., N. Bardeesy, and D. Ronald A. "Genetics and biology of pancreatic ductal adenocarcinoma". In: *Genes & Development* 20.10 (2006), pp. 1218–1249 (↑ 51).

[238] B. Cerruti et al. "Polarity, cell division, and out-of-equilibrium dynamics control the growth of epithelial structures". In: *Journal of Cell Biology* 203.2 (2013), pp. 359–372 (↑ 51).

[239] S. E. Boas and R. M. Merks. "Synergy of cell–cell repulsion and vacuolation in a computational model of lumen formation". In: *Journal of The Royal Society Interface* 11.92 (2014), p. 20131049 (↑ 51).

[240] P. van Liedekerke, L. Gannoun, A. Loriot, F. Lemaigre, and D. Drasdo. "Influence of cell mechanics in embryonic bile duct lumen formation: insight from quantitative modeling". Pre-publication. 2021 (↑ 51, 94).

[241] K. A. Rejniak and A. R. Anderson. "A computational study of the development of epithelial acini: I. Sufficient conditions for the formation of a hollow structure". In: *Bulletin of mathematical biology* 70.3 (2008), pp. 677–712 (↑ 51).

[242] K. A. Rejniak and A. R. Anderson. "A computational study of the development of epithelial acini: II. Necessary conditions for structure and lumen stability". In: *Bulletin of mathematical biology* 70.5 (2008), p. 1450 (↑ 51).

[243] S. Dasgupta, K. Gupta, Y. Zhang, V. Viasnoff, and J. Prost. "Physics of lumen growth". In: *Proceedings of the National Academy of Sciences* 115.21 (2018), E4751–E4757 (↑ 51, 79).

[244] C. Duclut, N. Sarkar, J. Prost, and F. Jülicher. "Fluid pumping and active flexoelectricity can promote lumen nucleation in cell assemblies". In: *Proceedings of the National Academy of Sciences* 116.39 (2019), pp. 19264–19273 (↑ 51, 79).

[245] B. C. Thorne, A. M. Bailey, D. W. DeSimone, and S. M. Peirce. "Agent-based modeling of multicell morphogenic processes during development". In: *Birth Defects Research Part C: Embryo Today: Reviews* 81.4 (2007), pp. 344–353 (↑ 51, 75).

[246] D. Drasdo and S. Höhme. "A single-cell-based model of tumor growth in vitro: monolayers and spheroids". In: *Physical biology* 2.3 (2005), pp. 133–147 (↑ 51, 75).

[247] I. Ramis-Conde, D. Drasdo, A. R. Anderson, and M. A. Chaplain. "Modeling the influence of the E-cadherin-β-catenin pathway in cancer cell invasion: a multiscale approach". In: *Biophysical journal* 95.1 (2008), pp. 155–165 (↑ 51).

[248] J. Galle, M. Loeffler, and D. Drasdo. "Modeling the effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations in vitro". In: *Biophysical journal* 88.1 (2005), pp. 62–75 (↑ 51).

[249] A. Ferrari, A. Veligodskiy, U. Berge, M. S. Lucas, and R. Kroschewski. "ROCK-mediated contractility, tight junctions and channels contribute to the conversion of a preapical patch into apical surface during isochoric lumen initiation". In: *Journal of cell science* 121.21 (2008), pp. 3649–3663 (↑ 52).

[250] K. Schafer. "The cell cycle: a review". In: *Veterinary pathology* 35.6 (1998), pp. 461–478 (↑ 53).

[251] A. Tzur, R. Kafri, V. S. LeBleu, G. Lahav, and M. W. Kirschner. "Cell growth and size homeostasis in proliferating animal cells". In: *Science* 325.5937 (2009), pp. 167–171 (↑ 53).

[252] M. Mir, Z. Wang, Z. Shen, M. Bednarz, R. Bashir, I. Golding, S. G. Prasanth, and G. Popescu. "Optical measurement of cycle-dependent cell growth". In: *Proceedings of the National Academy of Sciences* 108.32 (2011), pp. 13124–13129 (↑ 53).

[253] R. Kafri, J. Levy, M. B. Ginzberg, S. Oh, G. Lahav, and M. W. Kirschner. "Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle". In: *Nature* 494.7438 (2013), pp. 480–483 (↑ 53).

[254] M. P. Wescott, M. Rovira, M. Reichert, J. Von Burstin, A. Means, S. D. Leach, and A. K. Rustgi. "Pancreatic ductal morphogenesis and the Pdx1 homeodomain transcription factor". In: *Molecular biology of the cell* 20.22 (2009), pp. 4838–4844 (↑ 53).

[255] P. P. Provenzano, D. R. Inman, K. W. Eliceiri, and P. J. Keely. "Matrix density-induced mechanoregulation of breast cell phenotype, signaling and gene expression through a FAK–ERK linkage". In: *Oncogene* 28.49 (2009), pp. 4326–4343 (↑ 53, 69).

[256] Y.-T. Yeh, S. S. Hur, J. Chang, K.-C. Wang, J.-J. Chiu, Y.-S. Li, and S. Chien. "Matrix stiffness regulates endothelial cell proliferation through septin 9". In: *PloS one* 7.10 (2012), e46889 (↑ 53, 69).

[257] M. A. Wozniak, R. Desai, P. A. Solski, C. J. Der, and P. J. Keely. "ROCK-generated contractility regulates breast epithelial cell differentiation in response to the physical properties of a three-dimensional collagen matrix". In: *The Journal of cell biology* 163.3 (2003), pp. 583–595 (↑ 53, 69).

[258] E. A. Klein et al. "Cell-cycle control by physiological matrix elasticity and in vivo tissue stiffening". In: *Current biology* 19.18 (2009), pp. 1511–1518 (↑ 53, 69).

[259] S. Huang, C. S. Chen, and D. E. Ingber. "Control of cyclin D1, p27Kip1, and cell cycle progression in human capillary endothelial cells by cell shape and cytoskeletal tension". In: *Molecular biology of the cell* 9.11 (1998), 3179–3193 (↑ 53, 69).

[260] Z. Zheng, H. Zhu, Q. Wan, J. Liu, Z. Xiao, D. P. Siderovski, and Q. Du. "LGN regulates mitotic spindle orientation during epithelial morphogenesis". In: *Journal of Cell Biology* 189.2 (2010), pp. 275–288 (↑ 54, 149).

[261] F. Martin-Belmonte, A. Gassama, A. Datta, W. Yu, U. Rescher, V. Gerke, and K. Mostov. "PTEN-mediated apical segregation of phosphoinositides controls epithelial morphogenesis through Cdc42". In: *Cell* 128.2 (2007), pp. 383–397 (↑ 54, 149).

[262] A. B. Jaffe, N. Kaji, J. Durgan, and A. Hall. "Cdc42 controls spindle orientation to position the apical surface during epithelial morphogenesis". In: *The Journal of cell biology* 183.4 (2008), pp. 625–633 (↑ 54, 149).

[263] Y. Qin, W. H. Meisen, Y. Hao, and I. G. Macara. "Tuba, a Cdc42 GEF, is required for polarized spindle orientation during epithelial cyst formation". In: *Journal of Cell Biology* 189.4 (2010), pp. 661–669 (↑ 54, 149).

[264] A. E. Rodriguez-Fraticelli, S. Vergarajauregui, D. J. Eastburn, A. Datta, M. A. Alonso, K. Mostov, and F. Martín-Belmonte. "The Cdc42 GEF Intersectin 2 controls mitotic spindle orientation to form the lumen during epithelial morphogenesis". In: *Journal of Cell Biology* 189.4 (2010), pp. 725–738 (↑ 54, 149).

[265] Y. Hao et al. "Par3 controls epithelial spindle orientation by aPKC-mediated phosphorylation of apical Pins". In: *Current Biology* 20.20 (2010), pp. 1809–1818 (↑ 54, 149).

[266] A. Ghaffarizadeh, R. Heiland, S. H. Friedman, S. M. Mumenthaler, and P. Macklin. "PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems". In: *PLoS computational biology* 14.2 (2018), e1005991 (↑ 54, 59, 149).

[267] S. Suveges, I. Chamseddine, K. A. Rejniak, R. Eftimie, and D. Trucu. "Collective cell migration in a fibrous environment: a hybrid multi-scale modelling approach". In: *Frontiers in applied mathematics and statistics* 7 (2021), p. 34 (↑ 54, 149).

[268] M. V. Blagosklonny and A. B. Pardee. "The restriction point of the cell cycle". In: *Cell cycle* 1.2 (2002), pp. 102–109 (↑ 54).

[269] C.-K. Hu, M. Coughlin, and T. J. Mitchison. "Midbody assembly and its regulation during cytokinesis". In: *Molecular biology of the cell* 23.6 (2012), pp. 1024–1034 (↑ 55).

[270] D. M. Bryant, A. Datta, A. E. Rodriguez-Fraticelli, J. Peränen, F. Martin-Belmonte, and K. E. Mostov. "A molecular network for de novo generation of the apical surface and lumen". In: *Nature cell biology* 12.11 (2010), pp. 1035–1045 (↑ 55).

[271] E. Evans. "Detailed mechanics of membrane-membrane adhesion and separation. II. Discrete kinetically trapped molecular cross-bridges". In: *Biophysical Journal* 48.1 (1985), pp. 185–192 (↑ 57, 148).

[272] E. Palsson. "A 3-D model used to explore how cell adhesion and stiffness affect cell sorting and movement in multicellular systems". In: *Journal of Theoretical Biology* 254.1 (2008), pp. 1–13 (↑ 57, 58, 61, 75, 148, 149, 152, 158, 159).

[273] A. Caiazzo and I. Ramis-Conde. "Multiscale modelling of palisade formation in gliobastoma multiforme". In: *Journal of theoretical biology* 383 (2015), pp. 145–156 (↑ 59).

[274] T. Heck, B. Smeets, S. Vanmaercke, P. Bhattacharya, T. Odenthal, H. Ramon, H. Van Oosterwyck, and P. Van Liedekerke. "Modeling extracellular matrix viscoelasticity using smoothed particle hydrodynamics with improved boundary treatment". In: *Computer Methods in Applied Mechanics and Engineering* 322 (2017), pp. 515–540 (↑ 59).

[275] C. Valero, H. Amaveda, M. Mora, and J. M. García-Aznar. "Combined experimental and computational characterization of crosslinked collagen-based hydrogels". In: *PLoS One* 13.4 (2018), e0195820 (↑ 59, 61, 152, 158, 159).

[276] H. Edelsbrunner and E. P. Mücke. "Three-dimensional alpha shapes". In: *ACM Transactions on Graphics (TOG)* 13.1 (1994), pp. 43–72 (↑ 61, 82).

[277] M. A. Lancaster and J. A. Knoblich. "Organogenesis in a dish: modeling development and disease using organoid technologies". In: *Science* 345.6194 (2014), p. 1247125 (↑ 74).

[278] S. E. Blutt and M. K. Estes. "Organoid Models for Infectious Disease". In: *Annual review of medicine* 73 (2022), pp. 167–182 (↑ 74).

[279] S. Sahu and S. K. Sharan. "Translating embryogenesis to generate organoids: Novel approaches to personalized medicine". In: *Iscience* 23.9 (2020), p. 101485 (↑ 74).

[280] A. Dart. "Organoid diversity". In: *Nature Reviews Cancer* 18.7 (2018), pp. 404–405 (↑ 74).

[281] D. Tuveson and H. Clevers. "Cancer modeling meets human organoid technology". In: *Science* 364.6444 (2019), pp. 952–955 (↑ 74).

[282] T. Ishiguro, H. Ohata, A. Sato, K. Yamawaki, T. Enomoto, and K. Okamoto. "Tumor-derived spheroids: relevance to cancer stem cells and clinical applications". In: *Cancer science* 108.3 (2017), pp. 283–289 (↑ 75).

[283] M. Hofer and M. P. Lutolf. "Engineering organoids". In: *Nature Reviews Materials* 6.5 (2021), pp. 402–420 (↑ 75).

[284] A. Gierer and H. Meinhardt. "A theory of biological pattern formation". In: *Kybernetik* 12.1 (1972), pp. 30–39 (↑ 75).

[285] L. Wolpert. "Positional information and the spatial pattern of cellular differentiation". In: *Journal of theoretical biology* 25.1 (1969), pp. 1–47 (↑ 75).

[286] P. Van Liedekerke, M. Palm, N. Jagiella, and D. Drasdo. "Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results". In: *Computational particle mechanics* 2.4 (2015), pp. 401–444 (↑ 75, 91).

[287] N. Sepúlveda, L. Petitjean, O. Cochet, E. Grasland-Mongrain, P. Silberzan, and V. Hakim. "Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model". In: *PLoS computational biology* 9.3 (2013), e1002944 (↑ 75).

[288] P. Van Liedekerke, P. Ghysels, E. Tijskens, G. Samaey, D. Roose, and H. Ramon. "Mechanisms of soft cellular tissue bruising. A particle based simulation approach". In: *Soft Matter* 7.7 (2011), pp. 3580–3591 (↑ 75).

[289] S. Okuda, Y. Inoue, M. Eiraku, T. Adachi, and Y. Sasai. "Vertex dynamics simulations of viscosity-dependent deformation during tissue morphogenesis". In: *Biomechanics and modeling in mechanobiology* 14.2 (2015), pp. 413–425 (↑ 75).

[290] I. González-Valverde and J. M. García-Aznar. "Mechanical modeling of collective cell migration: An agent-based and continuum material approach". In: *Computer Methods in Applied Mechanics and Engineering* 337 (2018), pp. 246–262 (↑ 75).

[291] P. Van Liedekerke et al. "A quantitative high-resolution computational mechanics cell model for growing and regenerating tissues". In: *Biomechanics and modeling in mechanobiology* 19.1 (2020), pp. 189–220 (↑ 75).

[292] D. Bergman, R. F. Sweis, A. T. Pearson, F. Nazari, and T. L. Jackson. "A global method for fast simulations of molecular dynamics in multiscale agent-based models of biological tissues". In: *iScience* (2022), p. 104387 (↑ 75).

[293] F. J. Montáns, F. Chinesta, R. Gómez-Bombarelli, and J. N. Kutz. "Data-driven modeling and learning in science and engineering". In: *Comptes Rendus Mécanique* 347.11 (2019), pp. 845–855 (↑ 75).

[294] D. An, N. H. Kim, and J.-H. Choi. "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews". In: *Reliability Engineering & System Safety* 133 (2015), pp. 223–236 (↑ 75).

[295] M. Uzkudun, L. Marcon, and J. Sharpe. "Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud". In: *Molecular systems biology* 11.7 (2015), p. 815 (↑ 75).

[296] B. Lambert, A. L. MacLean, A. G. Fletcher, A. N. Combes, M. H. Little, and H. M. Byrne. "Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis". In: *Journal of mathematical biology* 76.7 (2018), pp. 1673–1697 (↑ 75).

[297] B. Lusch, J. N. Kutz, and S. L. Brunton. "Deep learning for universal linear embeddings of nonlinear dynamics". In: *Nature communications* 9.1 (2018), pp. 1–10 (↑ 76).

[298] H. Takeda, Y. Kameo, and T. Adachi. "Continuum modeling for neuronal lamination during cerebral morphogenesis considering cell migration and tissue growth". In: *Computer Methods in Biomechanics and Biomedical Engineering* 24.7 (2021), pp. 799–805 (↑ 76).

[299] A. M. Stein, T. Demuth, D. Mobley, M. Berens, and L. M. Sander. "A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment". In: *Biophysical journal* 92.1 (2007), pp. 356–365 (↑ 76).

[300] J. Sharpe. "Computer modeling in developmental biology: growing today, essential tomorrow". In: *Development* 144.23 (2017), pp. 4214–4225 (↑ 76).

[301] S. Tripathi, D. A. Kessler, and H. Levine. "Biological networks regulating cell fate choice are minimally frustrated". In: *Physical Review Letters* 125.8 (2020), p. 088101 (↑ 76).

[302] S. Tripathi, H. Levine, and M. K. Jolly. "The physics of cellular decision making during epithelial–mesenchymal transition". In: *Annual Review of Biophysics* 49.1 (2020) (↑ 76).

[303] D. Camacho-Gómez, J. M. García-Aznar, and M. J. Gómez-Benito. "A 3D multi-agent-based model for lumen morphogenesis: the role of the biophysical properties of the extracellular matrix". In: *Engineering with Computers* (2022), pp. 1–15 (↑ 79, 94, 145, 147, 152, 158, 159).

[304] B. M. Wilamowski. "Neural network architectures and learning algorithms". In: *IEEE Industrial Electronics Magazine* 3.4 (2009), pp. 56–63 (↑ 91).

[305] S. Ray. "A quick review of machine learning algorithms". In: *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. IEEE. 2019, pp. 35–39 (↑ 91).

[306] K. Fei, J. Zhang, J. Yuan, and P. Xiao. "Present Application and Perspectives of Organoid Imaging Technology". In: *Bioengineering* 9.3 (2022), p. 121 (↑ 91).

[307] C. C.-W. Yu, A. L. Woods, and D. A. Levison. "The assessment of cellular proliferation by immunohistochemistry: a review of currently available methods and their applications". In: *The Histochemical journal* 24.3 (1992), pp. 121–131 (↑ 93).

[308] P. Rawla. "Epidemiology of prostate cancer". In: *World journal of oncology* 10.2 (2019), p. 63 (↑ 98).

[309] L. Wang, B. Lu, M. He, Y. Wang, Z. Wang, and L. Du. "Prostate cancer incidence and mortality: global status and temporal trends in 89 countries from 2000 to 2019". In: *Frontiers in Public Health* 10 (2022), p. 176 (↑ 98).

[310] L. Bubendorf, A. Schöpfer, U. Wagner, G. Sauter, H. Moch, N. Willi, T. C. Gasser, and M. J. Mihatsch. "Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients". In: *Human pathology* 31.5 (2000), pp. 578–583 (↑ 98).

[311] C. J. Logothetis and S.-H. Lin. "Osteoblasts in prostate cancer metastasis to bone". In: *Nature Reviews Cancer* 5.1 (2005), pp. 21–28 (↑ 98).

[312] H. Saitoh, M. Hida, T. Shimbo, K. Nakamura, J. Yamagata, and T. Satoh. "Metastatic patterns of prostatic cancer: correlation between sites and number of organs involved". In: *Cancer* 54.12 (1984), pp. 3078–3084 (↑ 98).

[313] K. Datta, M. Muders, H. Zhang, and D. J. Tindall. "Mechanism of lymph node metastasis in prostate cancer". In: *Future oncology* 6.5 (2010), pp. 823–836 (↑ 98).

[314] E. C. Woodhouse, R. F. Chuaqui, and L. A. Liotta. "General mechanisms of metastasis". In: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 80.S8 (1997), pp. 1529–1537 (↑ 98).

[315] R. T. Gupta, K. A. Mehta, B. Turkbey, and S. Verma. "PI-RADS: Past, present, and future". In: *Journal of Magnetic Resonance Imaging* 52.1 (2020), pp. 33–53 (↑ 98).

[316] L. Egevad, T. Granfors, L. Karlberg, A. Bergh, and P. Stattin. "Prognostic value of the Gleason score in prostate cancer". In: *BJU international* 89.6 (2002), pp. 538–542 (↑ 98).

[317] G. Draisma, R. Postma, F. H. Schröder, T. H. van der Kwast, and H. J. de Koning. "Gleason score, age and screening: modeling dedifferentiation in prostate cancer". In: *International journal of cancer* 119.10 (2006), pp. 2366–2371 (↑ 98).

[318] H. Lilja, D. Ulmert, and A. J. Vickers. "Prostate-specific antigen and prostate cancer: prediction, detection and monitoring". In: *Nature Reviews Cancer* 8.4 (2008), pp. 268–278 (↑ 98).

[319] U.-H. Stenman, J. Leinonen, W.-M. Zhang, and P. Finne. "Prostate-specific antigen". In: *Seminars in cancer biology*. Vol. 9-2. Elsevier. 1999, pp. 83–93 (↑ 98).

[320] M. B. Gretzer and A. W. Partin. "PSA levels and the probability of prostate cancer on biopsy". In: *European Urology Supplements* 1.6 (2002), pp. 21–27 (↑ 99).

[321] I. M. Thompson and D. P. Ankerst. "Prostate-specific antigen in the early detection of prostate cancer". In: *Cmaj* 176.13 (2007), pp. 1853–1858 (↑ 99, 109).

[322] B. Holmström, M. Johansson, A. Bergh, U.-H. Stenman, G. Hallmans, and P. Stattin. "Prostate specific antigen for early detection of prostate cancer: longitudinal study". In: *Bmj* 339 (2009) (↑ 99, 109).

[323] M. Adhyam and A. K. Gupta. "A review on the clinical utility of PSA in cancer prostate". In: *Indian journal of surgical oncology* 3 (2012), pp. 120–129 (↑ 99, 109).

[324] D. Ilic, M. M. Neuberger, M. Djulbegovic, and P. Dahm. "Screening for prostate cancer". In: *Cochrane Database of Systematic Reviews* 1 (2013) (↑ 99).

[325] H. V. Jain, S. K. Clinton, A. Bhinder, and A. Friedman. "Mathematical modeling of prostate cancer progression in response to androgen ablation therapy". In: *Proceedings of the National Academy of Sciences* 108.49 (2011), pp. 19701–19706 (↑ 99).

[326] P. Colli, H. Gomez, G. Lorenzo, G. Marinoschi, A. Reali, and E. Rocca. "Mathematical analysis and simulation study of a phase-field model of prostate cancer growth with chemotherapy and antiangiogenic therapy effects". In: *Mathematical Models and Methods in Applied Sciences* 30.07 (2020), pp. 1253–1295 (↑ 99).

[327] G. Lorenzo, V. M. Pérez-García, A. Mariño, L. A. Pérez-Romasanta, A. Reali, and H. Gomez. "Mechanistic modelling of prostate-specific antigen dynamics shows potential for personalized prediction of radiation therapy outcome". In: *Journal of the Royal Society Interface* 16.157 (2019), p. 20190195 (↑ 99).

[328] M. N. Kamel Boulos and P. Zhang. "Digital twins: from personalised medicine to precision public health". In: *Journal of Personalized Medicine* 11.8 (2021), p. 745 (↑ 99).

[329] A. L. Gimpel, W. J. Stark, R. Heckel, and R. N. Grass. "A digital twin for DNA data storage based on comprehensive quantification of errors and biases". In: *Nature Communications* 14.1 (2023), p. 6026 (↑ 99).

[330] K. Maksymenko, A. K. Clarke, I. Mendez Guerra, S. Deslauriers-Gauthier, and D. Farina. "A myoelectric digital twin for fast and realistic modelling in deep learning". In: *Nature Communications* 14.1 (2023), p. 1600 (↑ 99).

[331] S. Hervas-Raluy, D. Sainz-DeMena, M. J. Gómez-Benito, and J. M. García-Aznar. "Image-based biomarkers for engineering neuroblastoma patient-specific computational models". In: *Engineering with Computers* (2024) (↑ 99).

[332] M. B. Boubaker, M. Haboussi, J.-F. Ganghoffer, and P. Aletti. "Finite element simulation of interactions between pelvic organs: predictive model of the prostate motion in the context of radiotherapy". In: *Journal of biomechanics* 42.12 (2009), pp. 1862–1868 (↑ 100).

[333] G. Lorenzo, T. J. Hughes, P. Dominguez-Frojan, A. Reali, and H. Gomez. "Computer simulations suggest that prostate enlargement due to benign prostatic hyperplasia mechanically impedes prostate cancer growth". In: *Proceedings of the National Academy of Sciences* 116.4 (2019), pp. 1152–1161 (↑ 100).

[334] N. Atuegwu, D. Colvin, M. Loveless, L. Xu, J. Gore, and T. Yankeelov. "Incorporation of diffusion-weighted magnetic resonance imaging data into a simple mathematical model of tumor growth". In: *Physics in Medicine & Biology* 57.1 (2011), p. 225 (↑ 101).

[335] P. S. Tofts. "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging". In: *Journal of magnetic resonance imaging* 7.1 (1997), pp. 91–101 (↑ 101, 110).

[336] D. Sainz-DeMena, W. Ye, M. Á. Pérez, and J. M. García-Aznar. "A finite element based optimization algorithm to include diffusion into the analysis of DCE-MRI". In: *Engineering with Computers* 38.5 (2022), pp. 3849–3865 (↑ 101).

[337] G. L. Pishko, G. W. Astary, T. H. Mareci, and M. Sarntinoranont. "Sensitivity analysis of an image-based solid tumor computational model with heterogeneous vasculature and porosity". In: *Annals of biomedical engineering* 39 (2011), pp. 2360–2373 (↑ 110).

[338] E. LoCastro et al. "Computational modeling of interstitial fluid pressure and velocity in head and neck cancer based on dynamic contrast-enhanced magnetic resonance imaging: feasibility analysis". In: *Tomography* 6.2 (2020), pp. 129–138 (↑ 110).

[339] B. Zelhof, M. Pickles, G. Liney, P. Gibbs, G. Rodrigues, S. Kraus, and L. Turnbull. "Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer". In: *BJU international* 103.7 (2009), pp. 883–888 (↑ 110, 161).

[340] A. Chatterjee, G. Watson, E. Myint, P. Sved, M. McEntee, and R. Bourne. "Changes in epithelium, stroma, and lumen space correlate more strongly with Gleason pattern and are stronger predictors of prostate ADC changes than cellularity metrics". In: *Radiology* 277.3 (2015), pp. 751–762 (↑ 110, 161).

[341] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2 (2021), pp. 203–211 (↑ 110).

[342] A. Jimenez-Pastor et al. "Automated prostate multi-regional segmentation in magnetic resonance using fully convolutional neural networks". In: *European Radiology* 33.7 (2023), pp. 5087–5096 (↑ 110).

[343] D. Sainz-DeMena, J. M. García-Aznar, M. Á. Pérez, and C. Borau. "Im2mesh: A Python Library to Reconstruct 3D Meshes from Scattered Data and 2D Segmentations, Application to Patient-Specific Neuroblastoma Tumour Image Sequences". In: *Applied Sciences* 12.22 (2022), p. 11557 (↑ 110).

[344] V. Noel, M. Ruscone, R. Shuttleworth, and C. K. MacNamara. "PhysiMeSS-A New PhysiCell Addon for Extracellular Matrix Modelling". In: *bioRxiv* (2023), pp. 2023–10 (↑ 119).

[345] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. "Neural message passing for quantum chemistry". In: *International conference on machine learning*. PMLR. 2017, pp. 1263–1272 (↑ 120, 133).

[346] M. Weigert, U. Schmidt, R. Haase, K. Sugawara, and G. Myers. "Star-convex polyhedra for 3d object detection and segmentation in microscopy". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020, pp. 3666–3673 (↑ 145).

[347] I. Arganda-Carreras, V. Kaynig, C. Rueden, K. W. Eliceiri, J. Schindelin, A. Cardona, and H. Sebastian Seung. "Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification". In: *Bioinformatics* 33.15 (2017), pp. 2424–2426 (↑ 145).

[348] M. D. Bloice, C. Stocker, and A. Holzinger. "Augmentor: an image augmentation library for machine learning". In: *arXiv preprint arXiv:1708.04680* (2017) (↑ 146).

[349] R. Haase et al. "CLIJ: GPU-accelerated image processing for everyone". In: *Nature methods* 17.1 (2020), pp. 5–6 (↑ 146).

[350] D. Legland, J. V. Silva, C. Cauty, K. Irina, and J. Floury. "Quantitative image analysis of binary microstructures: Application to the characterisation of dairy systems". In: *European Microscopy Congress 2016: Proceedings*. Wiley Online Library. 2016, pp. 591–592 (↑ 146).

[351] A. Bélanger, H. Van Halbeek, H. C. Graves, K. Grandbois, T. A. Stamey, L. Huang, I. Poppe, and F. Labrie. "Molecular mass and carbohydrate structure of prostate specific antigen: studies for establishment of an international PSA standard". In: *The Prostate* 27.4 (1995), pp. 187–197 (↑ 161).

[352] J. C. Böck, F. Kaufmann, and R. Felix. "Comparison of gadolinium-DTPA and macromolecular gadolinium-DTPA-polylysine for contrast-enhanced pulmonary time-of-flight magnetic resonance angiography". In: *Investigative radiology* 31.10 (1996), pp. 652–657 (↑ 161).

[353] National Center for Biotechnology Information. *PubChem Compound Summary for CID 6918037, Gadoterate Meglumine*. https://pubchem.ncbi.nlm.nih.gov/compound/Gadoterate-Meglumine. Retrieved March 8, 2024. 2024 (↑ 161).

[354] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. "Optuna: A Next-Generation Hyperparameter Optimization Framework". In: *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, pp. 2623–2631 (↑ 163).

[355] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014) (↑ 163).