Multi-scale modeling of tumor development

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Jakob Rosenbauer
aus Stuttgart

Köln, Februar 2021

aus dem Jülich Supercomputing Center (IAS)
des Forschungszentrums Jülich

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstattende:

1. Prof. Dr. Gunnar Schröder
2. Prof. Dr. Alexander Schug

Tag der mündlichen Prüfung:

Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.
Die Dissertation wurde in dieser oder ähnlicher Form noch bei keiner anderen Institution eingereicht. Ferner habe ich bislang keine erfolglosen oder erfolgreichen Promotionsversuche unternommen.

_____

Jakob Rosenbauer

_____

Datum

Abstract

TISSUES are made up of a large number of cells that interact with each other and, through their individual properties, generate emergent behavior on the tissue scale. Those interactions are essential for the formation of embryos and the maintenance of adult organisms. However, malignant cells can lead to unwanted effects that are beyond the control of the organism, for example, the formation of tumors. Interference in those processes is necessary to treat and avoid illnesses. Therefore, a mechanistic understanding of the complex interactions that drive tumor progression is necessary. The formation of cancer is an inherently multi-scale problem since changes on the cellular level enable uncontrolled growth on the tissue scale and experimental data is available from the molecular to organism scale. Theoretical mechanistic models can connect these scales and are great tools to find causal relationships. The transfer of physical models to complex biological problems has proven successful and mechanical interactions between cells have been found to influence the progression and development of tissues and tumors.

In this work, I focus on simulating tissue with the aim of modeling tissue on multiple scales and finding the underlying principles that govern tumor development.

I develop a computational model that can simulate single cells with their geometric three-dimensional shape as well as millions of those cells interacting with each other in a tissue. The tissue simulation model adapts the Potts model for the simulation of cells and is optimized to be used on modern supercomputers. The framework enables large-scale simulations of tissues with high resolution. I demonstrate the ability of the model to simulate large-scale tissue and emergent behavior from single-cell parametrization. Heterogeneous tumor growth is simulated and the effect of different treatment schemes is compared in a qualitative model. The properties of the explicitly modeled single cells generate emergent behavior on the tissue scale consisting of millions of cells.

For quantitative observations, I focus on the mechanical and physical interactions of cells and their influence on tumor growth. I use a two-type tumor model and observe the effects of changes in tumor-associated mechanical cellular properties. Cell adhesion, motility, and stiffness are varied and the effects on the emergent tumor growth are observed. I demonstrate the effects of cellular mechanical properties on the growing tumor and observe a dependency between growth speed and tumor shape.

Cells within a growing tumor can mutate and alter their behavior, therefore tumors consist of multiple cell types that compete over the available resources. Mutations, together with tumor internal competition, lead to an evolutionary behavior that drives the tumor composition towards cell types with optimal properties. I introduce a model of heterogeneous tumor growth and observe the development of the tumor composition with the influence of a dynamic nutrient surrounding. The evolutionary speed shows dependency on the frequency of the fluctuations and the tumor is driven towards a low-adhesion regime.

In this thesis, I show the development of a multi-scale model for computational tissue modeling. The presented models allow tissue modeling on high-performance computers and open up possibilities for multi-scale tissue simulations. I implement models for the simulation of tumor growth and find explicit dependencies of the emergent tumor properties on the mechanical single-cell properties and the tumor surroundings.

Zusammenfassung

Zellgewebe ist ein grundlegender Baustein multizellulären Lebens und besteht aus interagierenden Zellen. Durch ihr Zusammenspiel definieren die Eigenschaften der einzelnen Zellen das Verhalten auf der Gewebeebene. Diese Interaktionen sind essentiell für die Formgebung in Embryos sowie die Instandhaltung von ausgewachsenem Gewebe. Ist jedoch das Verhalten einzelner Zellen gestört, kann dies zu vom Körper nicht kontrollierbaren Verhaltensweisen führen, wie dem Wachstum von Tumoren. Um Krankheiten vorzubeugen oder zu heilen wird in dieses Verhalten von Zellen eingegriffen, dies setzt ein grundlegendes mechanistisches Verständniss der Vorgänge im Gewebe vorraus. Die Entwicklung von Tumoren im Körper ist ein skalenübergreifendes Problem, da Veränderungen in einzelnen Zellen ein unkontrolliertes Wachstum anstoßen, welches auf der Gewebe- und Zellebene reguliert wird. Experimentelle und klinische Daten werden an beiden Enden dieser wechselwirkenden Skalen erhoben. Theoretische Modelle können diese Skalen verbinden und kausale Zusammenhänge zwischen den Ebenen herstellen. Theoretische Konzepte der Physik werden erfolgreich auf biologische Problemstellungen angewandt, so beeinflussen mechanische Eigenschaften und mechanisches Feedback einzelner Zellen das Wachstum von Zellen und Tumoren.

Der Fokus dieser Arbeit liegt auf der Simulation von Gewebe, mit dem Ziel das Wachstum und die Entwicklung von Tumoren unter Berückuchtigung mehrerer Skalen zu modellieren.

Ich entwickle ein Simulationsframework, welches die Dynamik der dreidimensionalen Struktur von Zellen in einem Gewebe, sowie die Interaktion Millionen solcher Zellen simuliert. Das Computermodel basiert auf dem Potts Modell und ist für die Verwendung auf modernen Supercomputern optimiert. Zunächst demonstriere ich die Fähigkeit des Modells große Gewebeteile und emergentes Verhalten aus der Einzelzellparametrisierung zu simulieren. Heterogenes Tumorwachstum wird simuliert und die Wirkung verschiedener Behandlungsschemata in einem qualitativen Modell verglichen.

Um das Verhalten quantitativ zu betrachten konzentriere ich mich auf mechanische und physikalische Wechselwirkungen zwischen Zellen und Gewebe. Ein Zwei-Typ-Tumormodell wird verwendet, um die Auswirkungen von tumorassoziierten Zelleigenschaften auf das Tumorwachstum zu untersuchen. Veränderungen von Zelladhäsion, Motilität und Kompressibilität und ihre Auswirkungen auf die Wachstumsgeschwindigkeit, Form und Invasivität des entstehenden Tumors werden beobachtet.

Zellen in einem Tumor können durch Mutationen Eigenschaftsänderungen erfahren, Tumore bestehen daher aus mehren Zelltypen, welche um die verfügbaren Resuoucen konkurrieren. Dies führt zu einem evolutionären Verhalten, welches Zelltypen mit optimalen Eigenschaften bevorzugt. Ich stelle ein heterogenes Tumormodell vor und beobachte die Entwicklung der Tumorzusammensetzung unter verschiedenen äußeren Einflüssen. Eine Abhänigkeit der Geschwindigkeit der evolutionären Anpassung von der Frequenz von Nährstoffffluktuationen, sowie ein evolutionärer Druck zu Zellen mit niedriger Adhäsion wird beobachtet.

In dieser Arbeit zeige ich die Entwicklung und Anwendung eines Multiskalenmodells für die rechnergestützte Gewebemodellierung. Die vorgestellten Modelle ermöglichen die Modellierung von Gewebe auf Höchleistungsrechnern und eröffnen Möglichkeiten für großskalige Gewebesimulationen. Ich implementiere mehrere Simulationsmodelle für Tumorwachstum und zeige, wie emergente Tumoreigenschaften von den mechanischen Eigenschaften einzelner Zellen abhängen.

Acknowledgements

I want to express my sincere gratitude towards my supervisor, adviser, and mentor during my time as a doctoral student, Alexander Schug. He always left me a large amount of freedom and independence during my research and still guided me through the jungle of the scientific landscape. During my time in his lab(s), he enabled me to go to several international conferences, was supportive in professional and private matters, and always had an open ear.

I thank Gunnar Schröder for being my first supervisor at HHU and facilitating my graduation!

I especially want to thank Marco Berghoff for the years of working together, complementing each other, and writing so so many lines of code. Hopefully to many more years of solving problems together.

Thank you Helmut, for paving the way.

I want to thank my family, especially Anette, Stefan, and Hildburg Rosenbauer, who always had my back and supported me throughout my entire studies.
I thank my partner Jana, for always being there for me and for enduring covid together with a stressed-out graduating boyfriend.

Thank you to the Jülich and Karlsruhe group members Julian, Emanuel, Marie, Fabrizio, Mehari and Oskar. I want to thank James Glazier, Paul Macklin, Jens Elgeti, and Steffen Scholpp for their open minds towards a stubborn Ph.D. student and many great discussions.

x

List of Publications

This thesis includes peer-reviewed articles and submitted manuscripts about the work undertaken during my doctoral project.
Published:

- Berghoff, M.*, **Rosenbauer, J.***, Hoffmann, F. and Schug, A., 2020. *Cells in Silico–introducing a high-performance framework for large-scale tissue modeling.* BMC bioinformatics, 21(1), pp.1-21. **Shared first authors.*

In preparation:

- Chapter 3 is prepared for submission to a peer-reviewed journal, with myself as first author, together with M. Berghoff, J.A. Glazier, and A. Schug.

Preprint:

- **Rosenbauer, J.**, Berghoff, M. and Schug, A., 2020. *Emerging tumor development by simulating single-cell events.* bioRxiv.

The manuscript was published on the preprint server bioRxiv and not yet published in a peer-reviewed journal. The contents are partially reproduced and adapted in Chapter 3.

Other contributions carried out throughout the course of my Ph.D. studies, not included in this thesis:

- Mattes, B., Dang, Y., Greicius, G., Kaufmann, L.T., Prunsche, B., **Rosenbauer, J.**, Stegmaier, J., Mikut, R., Özbek, S., Nienhaus, G.U. and Schug, A., 2018. *Wnt/PCP controls spreading of Wnt/β-catenin signals by cytonemes in vertebrates.* Elife, 7, p.e36953.

- **Rosenbauer, J.**, Zhang, C., Mattes, B., Reinartz, I., Wedgwood, K., Schindler, S., Sinner, C., Scholpp, S. and Schug, A., 2020. *Modeling of Wnt-mediated tissue patterning in vertebrate embryogenesis*. PLOS Computational Biology, 16(6), p.e1007417.

- Cui, H., Wang, X., Wesslowski, J., Tronser, T., **Rosenbauer, J.**, Schug, A., Davidson, G., Popova, A.A. and Levkin, P.A., 2020. *Assembly of Multi-Spheroid Cellular Architectures by Programmable Droplet Merging*. Advanced Materials, p.2006434.

Contents

List of abbreviations

AFM Atomic Force Microscope.

ATP Adenosine Triphosphate.

CPM Cellular Potts model.

CPU Compute Processing Unit.

CRISPR Clustered regularly interspaced short palindromic repeats.

CT Computed Tomography.

DNA Deoxyribonucleic Acid.

ECM Extra-cellular matrix.

GFP Green Fluorescent Protein.

GPU Graphics Processing Unit.

MCS Monte Carlo Sweep.

MERFISH Multiplexed Error-Robust Fluorescence In Situ Hybridization.

MIBI Multiplexed Ion Beam Imaging.

MPI Message Passing Interface.

MRI Magnetic Resonance Imaging.

ODE ordinary differential equation.

PET Positron Emission Tomography.

RNA Ribonucleic Acid.

SBML Systems Biology Markup Language.

TME Tumor Microenvironment.

1 Introduction

ASKING *the right questions is the key ingredient for scientific advances. Without a meaningful problem, the results and findings are seldom groundbreaking and self-explanatory enough to propel our knowledge. Asking good questions requires profound* a priori *knowledge about the subject and system, posing a challenge for interdisciplinary research. The state of the art of all involved disciplines has to be considered, which is increasingly difficult due to the high specialization of individual research directions. In this chapter, I give an overview of tumor research and observe tumor growth from a mechanistic physics perspective. I introduce biological phenomena that are relevant to the modeling of tumor growth. I highlight the complexity of tissue development and the interactions that are involved in different spatial levels from single cells to tissues. I describe physical phenomena of tumor development and how physical models can help with the understanding of biological questions. Different computational and mathematical models are portrayed and compared, based on their complexity and level of detail. Based on this summary of the biological and physical properties of tumor development, I motivate the questions asked throughout this thesis.*

The scientific understanding of the world around us has tremendously increased over the past centuries. From reaching the moon and harvesting nuclear power to the understanding of quanta and relativity, we constantly expanded the horizon of our knowledge. Yet, the inner workings of living organisms, and finally of humans remain understood incompletely. Historically, the understanding and formulation of fundamental rules have advanced physics and lead to the technological progress that we all take advantage of in our daily lives.

Biological discoveries and experimental methods have allowed an increasingly detailed picture of the processes facilitating life. Many of the found theories are, from a physicist's point of view, unsettlingly specialized. More and more physicists and chemists follow Feynman's lead [1] and focus on the description of life and biological processes. The joint research of researchers from many disciplines so far has led to advances in the field of life sciences by the application of theoretical and experimental methods traditionally used in other disciplines. This trend can be observed in the Nobel Prizes of recent years: Nobel Prize in physics for optical tweezers (2018), Nobel Prize in chemistry for CRISPR-Cas9 (2020), and most remarkably the Nobel Prize in chemistry 2017 for Cryo-EM were awarded to three physicists. Through interdisciplinary work and cooperation between different fields, new discoveries can be made. This is especially true for the field of biology and medicine, since nature and finally, our bodies are the result of a billion-year-long optimization process and therefore staggeringly complex. This complexity makes the quantitative description of biology very hard and several urging questions are not solved yet: Why do we age? How can we cure and/or prevent cancer, HIV/AIDS, Alzheimer's disease, and recently Covid-19?
Here, an interdisciplinary approach promises new insights and exciting research in the future.

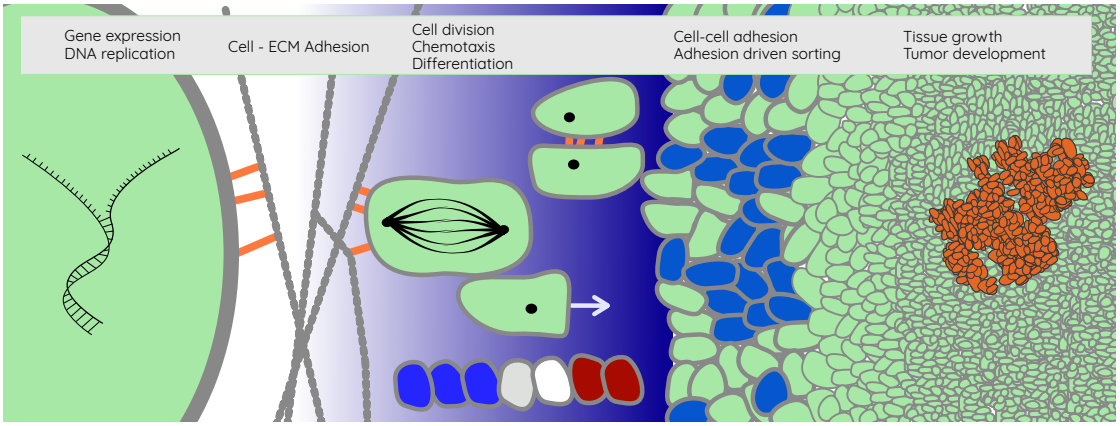

Figure 1: **From sub-cellular processes to tissue scale effects**. Effects that affect tumor growt are highlighted. Elements of cells and cellular interactions are depicted in a schematic representation, starting with the intracellular elements such as the DNA and the cell membrane O(nm), with the level of detail decreasing to the right from cell-Extra-cellular matrix (ECM) interactions over cell-cell interactions to tissue scale effects, such as tumor growth O(mm-cm).

In this thesis, computational modeling techniques are used for the exploration of tumor development, applying physical models to simulate the emergent behavior of single cells on the tissue level. In this chapter, I zoom out from the inner workings of a cell to the interplay of millions of cells that make up tissues, tumors, and entire organisms. For each scale, I focus on the biological and physical effects that determine the behavior on that scale (cf. Fig. 1). Furthermore, I describe the experimental methods and theoretical models that are applied on this scale.

Organisms found in nature are highly complex and diverse, with a much broader range of effects and properties than can be covered in this thesis. Therefore, I focus on cells in vertebrates and will disregard invertebrates, bacteria, plants, and fungi. Furthermore, I focus on mechanical properties and behaviors that are relevant to tumor development. For each scale, I elaborate on a selection of relevant effects for tumor growth and highlight important experimental methods and theoretical models.

1.0.1 Single-cell Interactions

All life we know is built upon cells. Cells are highly variable in their properties and perform a complex multitude of tasks. Here, I highlight some structural and functional elements that are relevant for tumor growth.

A lipid bilayer membrane encases the cell plasma (cytoplasm), which contains functional proteins and RNA. Inside the cell, the nucleus contains the DNA, that hosts all genetic information of the cell. Cells perform various tasks in different parts of the organism and distinct cell types are responsible for those different tasks. All cell types are necessary for a functioning organism, e.g., red blood cells transport oxygen to muscle cells that then can contract.

In most cases, cells have the fascinating ability to identically replicate themselves through cell division (mitosis). Cell division is part of the cell cycle in which the cell performs a sequence of operations to replicate its DNA and divide the cell into two new daughter cells. The cellular life cycle is structured into distinct phases that are necessary for reproduction. Most generally, for

animal cells, there are two major phases of the cell cycle that a cell needs to complete to divide. The cell grows and duplicates its DNA during interphase, followed by the mitotic phase in which the chromosomes and cell constituents are sorted and the cell is divided into two daughter cells.

Two typical pathways can lead to the death of a cell. Apoptosis is programmed cell death that is initiated by the cell itself and from external signal cues and leads to a planned disintegration of the cell. Necrosis, on the other hand, is the unregulated death of a cell without control over the process, e.g., through acute injury of the cell. In necrosis, the membrane ruptures and spills out the contents of the cell into the Extra-cellular matrix (ECM), which can be toxic and lead to inflammation of the surrounding cells.

The behavior of cells is determined by the activated (or expressed) genes at a certain time. The starting point of transcription is defined by transcription factors binding to the promoter DNA, defining the starting point of a gene. The sequence of nucleic acids of an expressed gene is transcribed into messenger RNA (mRNA) inside the nucleus by the enzyme RNA polymerase.

The mRNA then leaves the nucleus and moves into the cytoplasm. There, the mRNA can then be translated by ribosomes into functional proteins, or perform functions on its own, such as inhibiting or expressing other genes and even its original gene. Those interactions between expressed genes, the molecules that are produced, and the cell internal feedback between thousands of genes with different pathways enable the adaptation of the cell to a broad range of external stimuli.

Adhesion is a major interaction of cells with their surroundings. Adhesion molecules populate the exterior of the cell membrane and can bind to structures outside of the cell. Possible binding partners are structures in the ECM, such as collagen fibers and the adhesion molecules of other cells. The mechanics of cell adhesion have been studied extensively and a wide variety of adhesion mechanisms have been identified, such as the maturation of focal adhesions (cell-ECM adhesion) over time.

Cells can remodel their geometry through their cytoskeleton, which consists of a multitude of protein filaments such as actin. Actin is a polymer whose polymerization and degradation can be actively influenced by the cell. This active remodeling of the cytoskeleton enables cells to expand and deform depending on their internal state and external stimuli. Through a combination of remodeling of the cell geometry and adhesion to the surroundings, cells can actively migrate through space. Chemotaxis describes the ability of cells to follow gradients of an external stimulus, such as signaling molecules or nutrients.

Nutrients (such as glucose and oxygen) diffuse from blood vessels through the ECM and are absorbed by cells, where they provide energy for cellular tasks. The cells use glucose and oxygen to generate Adenosine Triphosphate (ATP), which is then used by the molecular machinery to drive cell internal processes. When a cell receives fewer nutrients than are necessary for maintenance of the cell, it can downregulate cell internal processes and stop the cell cycle. Cells can enter this state, termed hypoxia, to sustain for a longer time. Hypoxic cells excrete signals to the surrounding cells and hypoxic signaling can trigger the formation of new blood vessels.

The signaling networks and cellular repair mechanisms, together with the immune system, keep tissues in a healthy state. Mutated or dysfunctional cells are generally discarded. If the changes of a single cell lead to enhanced proliferation and remain undetected, tissue homeostasis

can get off balance and a tumor can form. The bulk mass of a tumor typically displays larger stiffness than the surrounding tissue. This property is widely used in cancer screening for various tumors. While the tumor mass expresses higher stiffness, the individual malignant cells in the tumor have been shown to be significantly less stiff than the cells of the surrounding tissue [2, 3]. The increased bulk stiffness is introduced by an elevated remodeling of the ECM by tumor cells, leading to a denser surrounding polymer network.

A loss of cell-cell adhesion has been found in tumor cells [4, 5]. A related property to cell-cell adhesion is the ability of cells to migrate. Cells within a tumor display enhanced motility, which facilitates the dissemination of the tumor and the formation of metastases. In epithelial tumors, the epithelial-to-mesenchymal transition is a step towards malignancy. Here, cells alter their phenotype from adherent and stationary to low-adhesion and migrating type [6].

Stem cells are special cells that are progenitor cells to more than one possible specialized cell type. Their properties are not completely defined and depending on external cues, stem cells can differentiate into different cell types. Especially in the formation of embryos, all cells start as stem cells and subsequently differentiate into more specialized cells based on external cues, as the development progresses. Stem cells play an important role in tissue renewal, and their impact on tumor growth has gained increasing importance recently [7].

Experimental Methods

For the measurement of the mechanical properties of single cells and cell adhesion, indenting and pulling experiments with an Atomic Force Microscope (AFM) are used. The membrane stiffness of a cell can be determined by indenting the cell membrane with the tip of a cantilever. By quantifying the force needed for indenting to a certain depth, the deformability and cell internal response to deformation can be measured. By attaching adhesion molecules or even entire cells to the tip of the cantilever, the adhesion strength can be measured by attaching the cantilever to a cell and then pulling away. The deflection and time-dependent adhesion strength can be measured in this way [8].

The elastic and viscous properties of single living cells can be determined by micro-pipette experiments [9]. The surface of a cell is aspirated with a glass micropipette with a diameter smaller than the cell. Through a pressure difference, forces can be applied to the cells with high precision and in the range of pN. This allows the measurement of the surface tension, viscosity, and elastic modulus of a cell.

Microfluidic devices can deform cells and together with high-resolution microscopy, mechanical properties can be inferred from the flow conditions and the deformation of the cells [10]. Furthermore, microfluidic devices allow for the sorting of single cells in fluid samples, e.g. blood, into cell-types, that can then be studied individually.

Cells from different individuals contain different DNA, leading to variability in the behavior of the cells. However, DNA can change over the lifetime through mutations. The cells making up tissue can vary by various mutations and lead to inhomogeneous behavior. Mutations are driving causes of diseases, especially in cancer. Therefore, the characterization of the DNA and the identification of mutations can help understand the behavior of tissue. Genetic analysis of the DNA and RNA present in single cells can help reconstruct the developmental trajectory of a tumor

and identify optimal treatment options [11, 12]. Even gene expression and mRNA can be measured by single-cell transcriptomics [13, 14].

Theoretical Models

Models that describe single cells mostly focus on cell internal properties, especially gene expression. The field of systems biology is concerned with the description of cell internal gene networks, their interaction, and the effects that external and internal stimuli have on the cell [15]. In systems biology, the activation of genes is described by coupled differential equations and describes the temporal and response to signals. This formalism helps to understand the regulation of cellular pathways and the complex network dynamics in cells [15]. A convenient way to store and represent a model of systems biology is the Systems Biology Markup Language (SBML) that is incorporated in many tissue models to determine the single-cell phenotype, behavior, and response.

The influence of cell size and internal checkpoints on the cell cycle was successfully modeled in yeast, which represents an organism of relatively low complexity [16]. During cell division, the repair mechanisms of the cells are active and damaged cells are hindered from cell division. Cell internal mechanisms arrest cell division if DNA damage is detected and can induce apoptosis, a main driver of this tumor suppression mechanism is the gene p53 [17]. Mutations of p53 alter the repair mechanisms and can disable the suppressive function. Single genes, such as p53, and their expression can significantly influence the behavior of a cell, cancer systems biology studies the influence and dependencies of genes and external stimuli [18].

Furthermore, models of individual cells have been focused on the mechanical properties of the cell geometry and deformation. Models of red blood cells can characterize the shape and dynamic behavior in flow conditions within capillaries and on surfaces [19, 20]. The cell membrane is explicitly modeled by a dynamically triangulated surface model that interacts with a hydrodynamics model. The forces on the cell can be modeled and different modes of rolling and shape regimes are found.

1.0.2 Many Cell and Cell-ECM Interactions

Tissues are made up of many cells that interact with each other and the ECM. In the human body, there are many different tissues that perform very different tasks. The cells that make up those tissues, as well as the interactions differ greatly. In healthy adult tissues, the rates of cell division and cell death are in equilibrium to achieve a constant size of the tissue. If the tissue is in this homeostatic condition, it remains in a constant size with regular self-renewal. The disturbance of the homeostasis condition leads to a temporally nonstatic behavior and is a sign of illnesses, but can also be necessary for tissue generation.

During tissue formation in developing organisms (embryogenesis), tissues expand and cells that initially are stem cells continuously differentiate into specific cellular fates. The differentiation of the cells is determined by factors like tissue geometry and external signals. Signaling molecules, termed morphogens, can be emitted by some cells and induce fate acquisition in other cells. Some cells, e.g., the cells at the rim of the expanding tissue, start to excrete a signaling molecule

(termed a morphogen) that is transported through the tissue, e.g. by diffusion. The cells within the tissue react to the morphogen concentration by acquiring a cellular fate based on the local morphogen concentration. Often one or more morphogens form gradients throughout the tissue and depending on the local concentration of those compounds, the cellular fate is then decided.

In developing tissues, fate determination can also occur at random, leading to a salt and pepper pattern [21]. A mechanism to separate two populations of mixed cells is adhesion-driven cell sorting, which for example plays a crucial role in the development of the human embryo [22]. Through a higher surface affinity and stronger adhesions between one cell type over the other, the cells dynamically sort into two regions.

Signaling also plays an important role in the developed organism. Here, phenotypic changes and apoptosis can be triggered by external signaling. Collective cellular behavior, such as wound healing and the immune response is orchestrated by cell-to-cell signaling.

Tissue growth through cell division and cell internal processes (e.g., muscle contraction) require an energy source. The energy in the form of glucose and oxygen is circulated through the organism inside the blood that is confined to the blood vessels. The nutrients then diffuse from the blood vessels into the tissue and deliver energy to the cells. An important mechanism that sustains the growth of tissue is angiogenesis, the establishment of new blood vessels in regions that are not sufficiently supplied with nutrients.

Experimental methods

In the last decade, new insights could be found through new microscopy methods in combination with controlling genetic parameters in cells. Methods such as light-sheet and confocal microscopy enable high resolution, spatiotemporally resolved imaging of living cells *in vivo* and *in vitro* [23, 24]. With 3D imaging technologies such as laser confocal microscopes, the spatial distribution of proteins can be tracked with subcell resolution. With those methods, it is possible to observe which pathways are expressed and where the processes around a protein are located within the tissue and even a cell. Proteins can be selectively tagged with fluorescent dyes in living cells. The fluorescent labeling of proteins can be introduced by genetic material (DNA or RNA) into the cell, which codes for the protein of interest and is coupled to a fluorescent protein (such as GFP).

Through traction force microscopy, mechanical forces exerted by single cells or tissue onto the surrounding matrix can be observed *in vitro*. This is achieved by distributing fluorescently labeled particles, often in the form of polymer spheres, in the matrix material (e.g., agar or collagen). Those spheres are then imaged by a 3D imaging method, such as confocal microscopy, at the same time as the tissue develops. The displacement of the spheres over time is correlated to the local force, which can be determined using the elastic properties of the matrix material. This allows for an accurate spatial description of the pulling and pushing forces of a cell or tissue on its surroundings [25].

Developing embryos (e.g., of zebrafish and Drosophila) are excellent models for the observation of developing tissue. During embryogenesis, numerous processes such as tissue growth, expansion, and differentiation occur at the same time. Through the small scale and accessibility of those models as well as their flexibility, many processes such as patterning, wound healing, and tissue expansion can be observed and studied. Additionally, embryogenic systems, especially zebrafish,

are transparent during a long time of development. This transparency allows for capturing the entire organism with light-sheet microscopy and thereby allowing access to the spatio-temporal trajectories of the cells [26].

Theoretical Models

Theoretical models for the description of the patterning of embryonic tissue were developed early, due to the accessibility of single-cell observations in developing embryos, e.g., Drosophila.

An early theory for the emergence of separated segments of tissue was the differential adhesion hypothesis [27]. The differential adhesion hypothesis assumes that similarly to non-mixing liquids, the interfacial energy is minimized, leading to a clustering of cells with equal adhesive strength. Leading to the sorting of cells, cells with lower adhesive strength envelop cells with higher adhesive strength [28]. Simulation models were used to simulate this mechanism in early computational studies [29].

Wolpert described tissue patterning in his French flag analogy, to be dependent on a single morphogen, that is produced in some producing cells [30]. From the cluster of producing cells, the morphogen diffuses into the surrounding tissue, creating a decaying gradient. The cells in the recipient tissue then decide which cell fate to accept based on the local morphogen concentration. In the French flag model, two thresholds determine two boundaries in the tissue and therefore a three-stripe pattern. The length scale of the patterning is determined by the diffusion length of the pattern defining morphogen. Fluctuations in the absolute size of the developing tissue can lead to asymmetries of the pattern and lacking scaling of the subdivisions. More stable patterning is achieved, with a dependency of the fate definition on more than one morphogen. For example, opposing morphogen gradients in the Drosophila neural tube patterning leads to a scalable behavior [31].

A more complex mechanism was described by Turing in his reaction-diffusion model [32]. Here, all cells are the source of two morphogens, morphogen A is self-amplifying and B suppresses the formation of A. This interaction leads to the formation of periodic structures, such as stripes, that initially form due to random initial fluctuations.

1.0.3 Tumor Growth

A cancerous tumor is a collection of cells that grows and expands beyond the control mechanisms of the organism. Thereby, the tumor tissue expands and causes harm to the host's body. Tumors can originate in very different parts of the body (among others breast, lung, skin, prostate, and colon) and therefore grow in very different surroundings and can have drastically different properties, lethalities, as well as treatment options. However, some fundamental properties are shared by all tumors and were brought together by Hanahan and Weinberg in the six hallmarks of cancer [33, 34]. The hallmarks are: maintenance of proliferative signaling, evasion of growth suppressing signaling, resisting apoptosis, enabling unlimited replicative potential, angiogenesis, and activating invasion and metastasis. Those six processes together enable the growth and invasion of a tumor that is then causing harm to the host.

Tumors are initiated by mutations in single cells. Mutations occur frequently in cell division

and in most cases mutations can be repaired, do not change cell properties, or lead to the death of the cell. In some cases, a mutation alters the properties of the cell. If the changed properties provide an advantage of the cell against its surroundings, it can start to grow into a cell mass. However, a single beneficial mutation towards growth is highly unlikely to form cancer, since all six hallmarks of cancer are necessary. Cancer develops over a long time with the accumulation of mutations in cells. Even if a tumor forms, growth often quickly ceases, and a stable dormant tumor forms.

Tumors live in constant competition with the host tissue and the immune system, and the mutation rate of cells in a tumor is elevated. With this elevated mutation rate, tumors mostly do not consist of a single 'tumor cell type', but of many cell types that make up the tumor. Tumors consisting of many cell types are termed heterogeneous tumors. The cell types inside a heterogeneous tumor not only compete for nutrients and space with the surrounding tissue but also with each other. This competition, together with the elevated mutation rate inside the tumor leads to an evolutionary selection of the fittest cell types. The evolutionary process allows for a gradual optimization of tumor properties over time and can drive the development of dormant to growing and malignant tumors. Yet, not all cellular properties can be optimized at the same time and beneficial properties can come with trade-offs. Trade-offs in cellular properties and tumor evolution are topics that recently gain more attention in the understanding of tumor development [35].

Tumor Invasion

In recent years the mechanical properties of tumor cells in comparison to 'normal' cells have been discussed frequently. Many studies observe a link between mechanical properties such as stiffness and fluidity with tumor invasion. Invasive tumor cells are found to be softer than non-cancerous cells [2].

Furthermore, the so-called jamming transition can differentiate between different modes of collective behavior of cells. In a jammed tissue, cells retain their nearest neighbors, whereas tissues in other states can act as an active fluid or gas-like [36, 37]. Those properties of the tissue are determined by the tissue density, the stiffness of the surrounding, as well as the adhesion strength. Depending on the tumor and surrounding properties, different modes of invasion into the surrounding tissue can be observed.

Tumor Microenvironment (TME)

Additionally to the properties of the tumor cells, the environment around the growing tumor largely influences the progression of the tumor. Depending on the tumor site in the organism, the geometry and surroundings differ. For example, in breast cancer, ductal carcinoma form in the milk ducts that consist of epithelial cells and constrain the tumor to the inside of the duct in the initial stages. Whereas glioblastoma, an invasive brain tumor, forms in the more homogeneous brain mass. Different external factors largely influence tumor development. ECM stiffness and the directionality of fibers influence and facilitate the transition to cell types of higher motility and invasiveness.

Through interactions with the immune system, immune cells are accumulated at the tumor site. Stromal cells that accumulate in and around the tumor can help suppress immune responses

to the tumor. Tumors can even induce phenotype changes in fibroblasts to tumor-associated fibroblasts that together with other cells create a pro-tumorigenic environment.

While the individual cells of malignant tumors are usually softer than the surrounding tissue, the tumor mass is usually stiffer than the surrounding. This stiffening of tumors allows for manual haptic detection of some tumors. The stiffening can be associated with an elevated excretion of extracellular matrix proteins by stromal cells (e.g., fibroblasts).

Expanding tumors quickly exceed the size that can be sustained by diffusing nutrients from adjacent blood vessels. Cells in the tumor center can enter a hypoxic state and cease dividing due to a lack of nutrients. With further expansion of the tumor, cells in the center can not sustain, and a necrotic core forms. An escape mechanism of tumors that avoids hypoxia and necrosis is the induction of angiogenesis. Angiogenesis is the formation of new blood vessels and can be triggered by hypoxic signaling pathways. With the formation of new blood vessels and capillaries in hypoxic areas, the tumor can continue to grow and is not limited in size by diffusion length.

Experimental Methods

Traditionally the cultivation of cells was performed on Petri-dishes and was limited to a two-dimensional extent. With the development of tumor spheroids, spherical and three-dimensional structures of cells can be grown and cultured (e.g., in hanging drops) [38]. Spheroids can better capture the cellular structure of a tumor and are used for a large variety of experiments up to testing treatments (e.g. glioblastoma spheroids inside a matrix show invasive behavior [39]).

Pre-clinical testing of treatments, as well as basic research about tumor growth, is conducted largely in the mouse model. Besides experimental data, a growing abundance of patient data exists, since all patients treated for cancer are thoroughly examined. Imaging methods such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Positron Emission Tomography (PET) are used to observe the spatial development of tumors.

Tissue samples of tumor suspects are taken from patients as a routine procedure to identify malignant tumors. Methods for the pathological examination of those tissue samples have evolved into powerful tools. Patient-derived explants sustain the tumor heterogeneity of a patient and can be used to test treatments and develop biomarkers [40]. Furthermore, techniques such as Multiplexed Ion Beam Imaging (MIBI), Multiplexed Error-Robust Fluorescence In Situ Hybridization (MERFISH) and imaging mass cytometry allow for the characterization of gene expression and phenotypes of single cells in pathology samples [41–43]. Working with tissue samples from patients preserves the tumor microenvironment as well as the heterogeneity of the tumor, which is often not the case in other experimental tumor models.

Theoretical Models

Tumor invasion describes the movement of tumor cells into the surrounding tissue and can be influenced by collective cell migration. Collective cell migration is a driving force in many cellular rearrangements and has been extensively studied. Similar to material science, a jamming transition of cells has been proposed, as the cell density and cell-cell adhesion increase, cells become trapped between neighboring cells retain their nearest neighbors, and cell movement ceases [44]. As the density decreases, parts of the tissue become unjammed, and parts of the tissue that are still close

to the jamming transition start to collectively migrate. Multiple modes of tumor invasion have been identified for different tumor types and surroundings. Two phase diagrams have recently been developed for cancer invasion, here cell density and cell-cell adhesion define four states, jammed, active fluid, active nematics, and gas-like behavior [36, 45].

Models that describe tumor growth on the scale of the tumor population, such as the tumor volume or size, are conceptually simple means of describing the dynamics of a growing tumor. An ordinary differential equation (ODE), describes the change of the tumor volume $V(t)$, depending on the current volume as well as external influences. The simplest case describing unconstrained growth is the ODE for exponential growth

$$\frac{d}{dt}V(t) = a \cdot V(t) \tag{1}$$

$$V(t) = V_0 \cdot e^{at}. \tag{2}$$

Here, the increase in tumor volume is directly proportional to the current volume, leading to an exponential increase in tumor volume (see also 2, blue). This model can correctly describe the initial growth phase of a tumor when the growth is unrestricted. For longer times, the tumor is always restricted in size by constraints of the host (e.g., blood and energy supply and space) and growth is mostly driven on the tumor surface.

A more realistic description of tumor dynamics is achieved by assuming an upper bound on tumor size, termed the carrying capacity K. The carrying capacity determines the tumor size which is approached for large t, $\lim_{t\to\infty} V(t) = K$. An implementation of an upper bound can lead to logistic growth with the following ODE, here the growth rate linearly decreases with tumor volume, reads

$$\frac{d}{dt}V(t) = a \cdot V(t)\left(1 - \frac{V(t)}{K}\right), \tag{3}$$

$$V(t) = \frac{K}{1 + \left(\frac{K-V_0}{V_0}\right) \cdot e^{-at}}. \tag{4}$$

Logistic growth is characterized by a symmetric function, with an initial exponential increase which decreases and asymptotically approaches K.

Gompertzian growth is often used to describe the dynamics of tumor growth, as well as the size of entire organisms. The growth rate of the tumor logarithmically decreases with increasing tumor volume

$$\frac{d}{dt}V(t) = aV(t) \cdot ln\left(\frac{K}{V(t)}\right), \tag{5}$$

$$V(t) = Ke^{-ln\left(\frac{K}{V_0}\right)\cdot e^{-at}}. \tag{6}$$

The nonlinear dependency of growth rate and tumor volume leads to an asymmetric graph with a fast increase and a slower approach to K (c.f. Fig. 2, green).

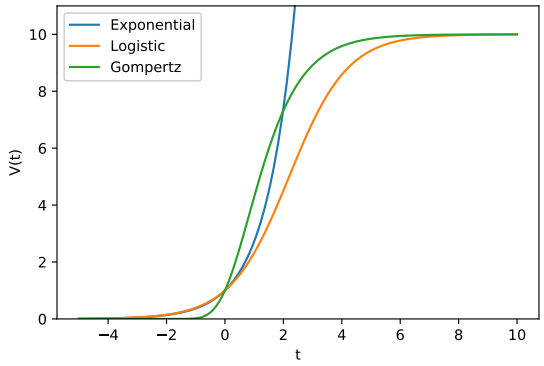

Figure 2: **Graphs of three tumor growth models** Exponential, logistic and gompertzian growth are depicted. With $a = 1$ and $K = 10$.

Those models represent the tumor size at the population scale and have been used (among others) to describe tumor growth and the effects of treatment [46]. Chemotherapy can be modeled in the simplest case by assuming a decrease of tumor volume with the drug concentration $c(t)$. Therefore, the term $-c(t) \cdot V(t)$ has to be added to the ODEs. In other cases, radiation treatment was modeled by adapting the carrying capacity K of the tumor. Simple models like this have been successful in predicting the treatment trajectory and classifying the regime of treatment outcome with clinical data [47].

1.1 Physics-based Models of Cancer

The physical description of tumors and tumor growth aims at understanding the basics of tumor growth and to find basic principles beyond the hallmarks of cancer. By the theoretical description of tumors, causal insights can be gained and tested in experiments and contribute to the development of new treatments, tumor markers, and diagnostics. One way of classifying the mathematical description of tumors distinguishes between two different categories, conceptual and descriptive models [48]. Conceptual models integrate a small set of interactions and assumptions into a set of equations, as exemplarily described above. Those equations can then be solved for different parameter values and initial conditions. Conceptual models, therefore provide quantitative results that depend on the assumptions and parameters. Quantitative results are major drivers of experiments and provide new insights. However, conceptual models struggle to capture the high complexity and strong interdependence between processes in tissue development and especially tumor growth, and stochastic events and single-cell effects are difficult to incorporate. Incorporating many different processes into a conceptual model is often not possible or not solvable. To describe tumor growth and especially its interaction with the immune system and tumor treatment, a large number of effects are necessary.

Descriptive models can incorporate many different effects and mechanisms by numerically deriving outcomes from the initial condition with parametrized effects. Through the continuous growth of computational power, as well as the ability to collect growing amounts of data during recent years, data and computation-heavy models are developed more and more. Machine learning has enabled the automated analysis of large-scale data sets and is even able to identify treatments

and tumor markers [49]. Supercomputers with continuously growing computational capabilities are able to compute increasingly complex models.

Descriptive models typically have a large parameter space due to the complexity of the represented mechanisms, sometimes even with correlated parameters. Here, the challenge is to control the model and to derive meaningful results from parameter variation. Quantitative results are possible by statistically sampling parameter ranges but are often computationally expensive. Qualitative insight, on the other hand, can be obtained more easily and findings may rely on the back-and-forth between experimental data and simulations. However, descriptive models can handle the complexity and multiscale processes that are present in tissues. Together with growing data resources and data-driven model building, descriptive models provide increasingly accurate results and the chance to complement experiments and medicine.

1.1.1 Computational Tissue Models

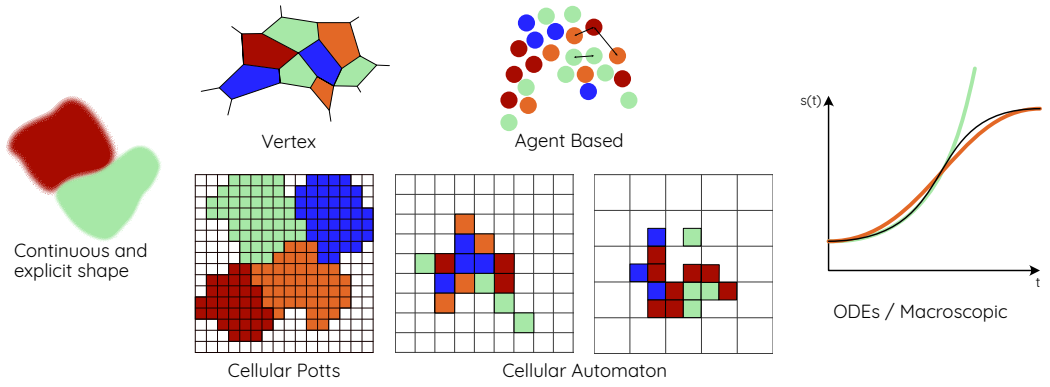

Figure 3: **Computational tissue models of different resolutions.** Models span in resolution from the continuous description of one or few cells (left), over the simulation of multiple cells interacting in a tissue in off- and on-grid models to descriptions of the tumor size on the population scale (right).

Computational modeling plays an important role in a multitude of disciplines and proves especially useful in cases where experiments are complicated and insights are limited to one or a few variables at a time. For example, in molecular dynamics, simulations allow the generation of trajectories that are impossible to observe in real life.

The development of tumors and their progression in the human body is an extremely complex process consisting of many puzzle pieces. Each hallmark or cancer in itself provides a great challenge due to the high complexity of the human body. Understanding the interaction and escape from the immune system alone requires knowledge about the interaction and complex interplay of many cell types within and outside the tumor.

Our understanding of the interactions of each of the hallmarks is steadily rising. However, incorporating all interactions together in one biological model is experimentally extremely challenging due to the high dimensionality and experimental complexity to measure and control single parameters. Experiments are costly, time-intensive, and often highly specific to one ques-

tion or problem. Here, computational models are convenient for the analysis of the effects of single parameters. One is not limited by the experimental feasibility and duration of experiments, and computational models enable sensitivity analysis for each effect. Due to full control over all parameters and initial assumptions, the impact of each model parameter can be quantified.

The objective of this manuscript is to demonstrate the construction and analysis of large-scale tumor models that are enabled by the development of a high-performance infrastructure for tissue simulations [50].

A large arsenal of approaches, tools, and frameworks has been developed to describe the development of tissues. Here we focus on models with spatial components. Different model complexities can be found, ranging from one-dimensional differential equations to complex multidimensional behavior (cf. Fig. 3). Generally, models that describe tissue consisting of individual cells can be classified into two categories, namely on-grid and off-grid models.

Off-grid models describe the positions of individual cells or cell components as particles moving in a continuous space. Examples for these kinds of models are agent-based models describing each cell as one or two particles that interact via potentials, these models have been implemented into [51] and form the base of the PhysiCell framework [52]. An additional form of off-lattice models are vertex models, in which cells are described by an area or volume spanned between vertex points that can move through space. Such a model with underlying fluid dynamics can be found in [53].

In on-grid models, on the other side, the position or geometry of cells is described as discrete points on a lattice. Those implementations are computationally less complex since the interactions and distances are discretized. The most straightforward implementation of such a model is a cellular automaton, in which each lattice point represents one cell, e.g., used in [54]. A more fine-grained model of cells is the cellular Potts model. Here, the cells are described as coherent regions of the grid with the same integer spin and the interaction between the cells is described by a Hamiltonian energy function. This model is implemented in frameworks like CompuCell3D [55], Morpheus [56] and the framework recently introduced and described in Chapter 2, called CellsInSilico [50].

Many of the computational frameworks implement additional functionalities, such as diffusion of compounds through the tissue, system-based approaches for single-cell behavior (e.g., SBML solver), and even interchangeable underlying tissue models [57].

Some discrete models describe the tissue as a continuum and do not differentiate single cells, but cell types (e.g., cancerous and non-cancerous tissues using a phase-field model in [58]), allowing the simulation of larger tissue regions.

All computational models come with advantages and drawbacks [59]. Center-based models enable the direct implementation of forces and physical time, while the spatial shape of cells is not explicitly included and the simulation size is limited. Cellular automaton models enable large cell numbers but are coarse-grained and rule-based with implicit physical parameters and jump-type processes for cellular movements. Cellular Potts models allow a higher resolution of individual cells in 2D and 3D and are easily adaptable, but the model is energy-based and driven by a Monte Carlo simulation, leading to complications, adapting time steps, and physical parameters. Vertex models represent the cellular shape and can easily incorporate mechanical stresses, however, the computational costs are high and only a limited number of cells can be modeled. The mathematical

details of the models are listed below.

For the simulations in Chapters 3 to 5 a simulation framework based on the cellular Potts model is used. The development of the framework is documented in Chapter 2. A Cellular Potts model (CPM) was chosen due to the explicit representation of the cellular geometry, which plays a crucial role in morphogenic developments such as embryogenesis and tumor development.

Off-grid Models

Agent-based Models Agent-based models (abm) describe single cells as one or multiple particles that are allowed to propagate continuously in space. The interactions of individual cells or agents are defined by a set of interaction potentials between the cells. The position of cells is described by the cell center and cells do not have a spatial extent. Therefore, a volume exclusion term can introduce repulsion for distances shorter than the spatial extent of the cell. Additional terms can introduce a preferential movement direction for motility or chemotaxis and attractive terms between cells to model adhesion. The system propagation is performed by evaluating the equation of motion for each agent and moving the agents along the resulting velocities.

In [51], each cell is represented by two points with a repulsive force between the two points. Cell division is induced as soon as the two cells exceed a set distance, and the magnitude of the repulsive force is a determining factor in tissue growth and competition. In abms, the spatial structure of the cells is only implicitly introduced by the interaction potentials.

Vertex Models In vertex models, the cells are modeled by 2D surfaces or 3D volumes that are spanned between the vertices. The vertices structure the space and define the cellular shape. Different from an abm, here the equations of motion are calculated for each vertex. Due to the explicit spatial representation of each cell, the volume and surface of a cell can be defined by an energy function, leading to an inherent volume exclusion of cells.

Vertex models can be ideally used for the simulation of tissues that are exclusively made up of cells with little extracellular matter (e.g., epithelial tissue).

On-grid Models

Simple Model on Irregular Grid In a recent publication [60], we investigated tissue patterning in a quickly expanding tissue. Through a combination of 2D single-cell simulation and *in vivo* experiments, we observe the effects of different modes of transport on a pattern defining morphogen. We find that direct cell-to-cell transport via cell extrusions termed cytonemes [61] facilitates a faster pattern establishment and therefore a faster and more stable patterning compared to diffusive transport. The model is based on an irregular two-dimensional grid of random cell positions. Those positions can be occupied by cells. Cell movement is modeled by the displacement of the cells to different locations on the irregular grid. Cell division and tissue growth by intercalation of overlying cell layers are modeled by introducing new cells within the tissue. To make space for new

cells, the cells are moved in a path to the nearest empty grid space. Cell motility is implemented by a random position switch of nearest neighbors.

A rapidly expanding tissue models the cellular movement during epiboly in zebrafish embryogenesis. During this expansion, the neural plate patterns into three regions, fore, mid, and hindbrain. This patterning is facilitated by Wnt-βcatenin signaling. The morphogen is produced by a set of cells at the leading edge of the tissue and distributed over the length of the expanding tissue. The neural plate progenitor cells receive their cell fate based on the concentration of the received morphogen.

Using the 2D computational model, we compare two transport mechanisms, free diffusion of the morphogen from the producing cells and directed transport via specialized filopodia from the producing to the receiving cells.

We found that the directed transport provides a faster pattern establishment than diffusive transport. Since the patterning in the zebrafish neural plate is a fast process that occurs within two to three hours, a fast establishment of the pattern is important for a stable and size invariant tissue formation.

Without any sorting mechanism, the model produced fuzzy boundaries and mixing between the pattern regions. In a back and forth process between simulations and experiments, we were able to show that Wnt activity contributes to the sorting of cells with similar Wnt activity and therefore stabilizes the pattern.

Cellular Automaton Cellular automaton models are defined on a regular grid (e.g., triangular, rectangular, or hexagonal). Each grid point represents one cell or other media such as ECM or liquid. The system propagation is performed by updating each grid point based on its state and the neighborhood. Volume exclusion can be introduced by limiting the possible number of cells per grid point. For example, cells can divide by introducing new cells into empty neighboring grid points. Cells can move to other points by introducing motility. In [54] a cellular automaton simulation is used to describe the development of tumor heterogeneity and tumor regrowth after treatment. Movements and actions in the model are typically introduced by Monte Carlo simulation steps that evaluate the local energy difference of a change and update based on an acceptance criterion (e.g. Metropolis). Models, termed lattice-gas cellular automata, allow for more than one cell per lattice point, allowing the simulation of large cell numbers.

Cellular automaton simulations provide a simple description of the spatial organization of tissue. Through the on-grid nature and short to medium range interactions, the model scales for large simulations with many cells. However, the spatial structure of the single cells is neglected and the choice of the lattice geometry and neighborhood may introduce spatial dependencies.

Cellular Potts Model The Ising model, which is a well-known model in solid-state physics, describes the interactions of spins in a solid crystal. Each grid point can either inhabit a spin-up or spin-down state and a nearest neighbor potential coupling the nearest neighbors. The Ising model can be used to model the magnetism in materials and depending on the interaction constant can reproduce ferromagnetic and anti-ferromagnetic behavior. A generalization of the Ising model, from binary spin states to integer spin states is called the Potts model. The Potts model is

well studied in solid-state physics and can be modeled by Monte Carlo simulations, which are computationally simple due to the nearest neighbor interactions.

In 1992, Glazier and Graner used a formulation of the Potts model to simulate the interaction of biological cells on a 2D grid [29]. They introduced global volume and surface constraints for each integer spin state, representing one biological cell. This model is called the CPM and can be implemented in two and three dimensions. Temporal dynamics of the system are defined by a Hamiltonian energy function H_{CPM} that consists of multiple energy contributions, which are added to make up the global energy

$$H_{\text{CPM}} = \sum_i E_i = E_V + E_S + E_A + ... \tag{7}$$

The Hamiltonian consists, in a basic case, of three energy terms. One term constraining the volume and surface of each cell and one term that adds an attractive interaction at the cell surface, representing adhesion. The calculations are performed on a grid Ω, which is filled with indices, called cell-ids. Each cell-id is linked to a cell type that specifies the global parameters of the cell. The definition of the individual energy terms is usually realized as a quadratic potential of the surface and volume in three dimensions or the surface and perimeter in two dimensions. Here, the three-dimensional definitions are shown:

$$E_V = \sum_{\text{c}\in\text{cells}} \lambda_V\left(\tau_c\right)\left(V(c) - V_0(\tau_c)\right)^2. \tag{8}$$

The energy term E_V defines a goal Volume for each cell, τ_c represents the cell type of the cell c. V_0 is the goal volume of the cell and $V(c)$ represents the current volume of the cell. The coupling factor λ_V sets the magnitude of the volume energy term and influences the volume compressibility of the cells.

The volume energy term determines the elastic properties of the cell. The for homogeneous isotropic elastic cells, the coupling factor λ_V can be linked to the Young modulus E

$$\lambda_V = \frac{E}{3(1-2\nu)}. \tag{9}$$

With the Poisson ratio ν, usually ≈ 0.5 [59]. Nonelastic, solid cells can therefore not move in the CPM formalism due to the volume energy term together with the locality of the grid update.

While the volume of the cell is usually determined by counting the occupied voxels of a cell, different surface metrics can be applied to extract the cell surfaces from the grid structure, such as the Manhattan metric or the marching cube algorithm. The surface energy is similarly defined

$$E_S = \sum_{\text{c}\in\text{cells}} \lambda_S\left(\tau_c\right)\left(S(c) - S_0(\tau_c)\right)^2. \tag{10}$$

For the surface energy, E_S, S_0 is the goal surface of the cell and $S(c)$ represents the current cell surface. The coupling factor λ_S determines the surface compressibility of the cells

$$E_A = \sum_{i\in\Omega}\sum_{j\in N(i)} A_{\tau_i\tau_j}(1-\delta_{i,j}). \tag{11}$$

For the adhesion energy calculation, the energy of the entire grid Ω is calculated by a sum over each grid point and its cell-id i. The adhesion energy is added for all neighbors $j \in N(i)$ of the current point i. The δ function limits the adhesion to grid points with different cell-ids i.e., different biological cells, since cells usually do not adhere to themselves. A set of special cell-ids is reserved for noninteracting surroundings, such as liquid and solids.

The grid is propagated by a so-called Monte Carlo Sweep (MCS), performing multiple Monte Carlo steps at different positions in the grid. In a Monte Carlo step, the cell-id of a point on the grid can be changed to the cell-id of a nearest neighbor grid point. The energy difference ΔE of this change is calculated using the Hamiltonian energy function. Based on this energy difference, the step is either accepted or rejected, with the probability (p_{accept})

$$p_{\text{accept}} = \begin{cases} 1, & \text{if } \Delta E < 0, \\ \exp(-\Delta E/T), & \text{otherwise.} \end{cases} \tag{12}$$

This acceptance criterion depends on the temperature T and is called the Metropolis criterion. Steps with a negative energy difference are accepted. Steps with a positive energy difference have an exponentially decaying acceptance probability depending on the absolute value of the energy difference and the temperature. For higher temperatures, the probability of accepting energetically higher states rises.

The formalism only depends on the Hamiltonian energy function and the Metropolis temperature, which makes the extension of the model flexible and provides a length scale. However, the model does not inherently map the physical properties of the simulated cells, such as forces or a timescale. Those properties have to be extracted from the simulated data. The system propagation with local nearest-neighbor changes constitutes a Markov-chain process. The grid is statistically uniformly sampled, this allows for the definition of a time step which is usually defined as one Monte-Carlo sweep. However, this does not lead to a physical timestep, the correlation between the Monte-Carlo time base and the physical time has to be inferred from cellular properties and can be challenging [62, 63].

Physical forces can be inferred from the energy function and cellular shapes and generate simulations that are directly comparable to traction force microscopy [64].

1.1.2 Parallel Computing

While the increase of computational speed in computers was initially driven by a rising clock speed, today the majority of the increase of computational power originates from the parallel use of many compute cores. The speed increase is achieved by partitioning the problem into a set of subproblems that are solved as processes on individual compute cores. Communication between the different cores working on one problem ensures consistency across all processes.

In supercomputers, many compute cores (either CPUs or GPUs) are connected on one computer board and share memory in a so-called node. Many of those nodes are connected by a fast network that allows communication between them, parallel software can use compute cores from one or multiple nodes.

For CPUs there are two main approaches of parallelization. Many CPUs read and write on the same physical memory (e.g., with libraries such as OpenMP). Here, the number of available cores is limited by the core number on a physical computer board (called a node) that also hosts the memory. Another parallelization strategy is based on communication between the processes with a message passing interface (mpi). Here, each process has its own memory and the synchronization between processes is performed solely by communication. Parallelization by MPI is not limited to a single node and can therefore scale to a large number of cores but requires more communication between the processes.

Even higher parallelism is employed using GPUs, here thousands of threads with low individual computing power are started and the problem is compartmentalized finely. This approach was initially mainly used for image processing but is used increasingly in general-purpose tasks. The top speed supercomputers of today harness most of their computing power from GPUs.

The development of software that harnesses the full potential of parallel infrastructure presents many challenges. To parallelize software, the problem has to be split into a set of ideally independent subproblems. All dependencies between the subproblems have to be communicated between the respective processes.

For example, particle-based simulations with long-range interactions do not allow for a spatial decomposition, since it would be impossible to calculate long-range interactions from subsets of the space. While grid-based simulations with low range interactions can be parallelized by spatial decomposition and a halo exchange between the sub-grids.

1.2 Challenges and Outlook

The complementing use of modeling and experimental methods has been successfully advanced our understanding of bacterial growth [65]. Unfortunately, experimental access to tumor growth is difficult due to the complex interactions with the microenvironment, immune system, and intra-tumoral heterogeneity. However, mathematical modeling of tumor growth has been successfully used in research and treatment. Theoretical observations of tumor growth and response to treatment have inspired predictions of chemotherapy scheduling that were tested and confirmed in clinical trials [66].

Clinical and experimental methods to observe and analyze tumor growth are continuously advancing, and methods such as patient-derived organoids and single-cell imaging and profiling techniques allow more detailed insights into tumor development. Most recently, spatially resolved transcriptomics has been announced as the method of the year [67]. This growing amount of clinical and experimental data requires an automated analysis, here machine learning methods have shown successes [68]. This large amount of data can be incorporated via model optimization as well as machine learning techniques [47]. Computational models provide the opportunity to link theory, experimental and clinical data, get cross-validation, and influence all fields. Executable tumor models have been proposed for personalized medicine and could incorporate all available data, guide decisions, and optimize therapy [69].

However, several challenges remain to be addressed in the computational modeling of tissue. The use of parallelized software and multiscale modeling approaches are still in their infancy. Scalable computational models are necessary to incorporate large amounts of data on various scales and levels of detail. Large scale simulations can be reached by optimizing models for scalability and supercomputing support for systematic parameter scans. Multi-scale models can reproduce physical and biological properties while enabling causal links between small-scale properties and large-scale effects. Currently, there are no established standards for the reporting of experimental and simulated data as well as computational models. This lack of standardization complicates the comparison of models against each other or even the incorporation of data from different sources. Furthermore, there are no benchmarking datasets to validate and quantify the performance of computational tissue models [70, 71].

In this thesis, I show the development, verification, and application of a physics-based model of large-scale tissue growth and focus on tumor development. First, I present the development of a multiscale tissue model that facilitates large-scale simulations of tissue with single-cell resolution in Chapter 2. Models that span multiple sizes are necessary to account for the variety of observed effects in tumor growth. The model is based on the cellular Potts model and enables the simulation of millions of interacting cells by optimization for supercomputers. The parallelization is realized by a domain decomposition with a halo exchange between the subdomains. The high scalability of the model allows for the simulation of large numbers of cells that are individually spatially resolved. Scalability is necessary to incorporate various spatial scales into the *in silico* experiments and reach sizes that are comparable to the experimental scales of tissue experiments.

Next, in a proof-of-concept approach, a model of tumor growth in a microenvironment and vasculature is developed in Chapter 3. I demonstrate its ability to reproduce experimentally observed behavior, as well as emergent behavior from single cells to large-scale heterogeneous tumors. This model paves the way for further investigation towards balancing model complexity and validating with data. I ask the question: *Can the biologically observed behavior be reproduced on a similar scale O(mm) by parametrizing single cells O(μ m) and observing the emergent behavior with our computational model?*

Following this qualitative model of tumor growth, I focus on the description of different tumor internal effects, especially the influence of the mechanical properties of tumor cells. Reduced model

complexity is necessary to find causal relations between single-cell properties and macroscopic tumor behavior and the characterization of the model. In Chapter 4, I implement a model to observe the effects of three cellular properties that are associated with malignant tumor growth. Cellular motility, stiffness, and adhesion are varied in a set of simulations and the effects on the emergent tumor are compared. The interactions of the growing tumor with surrounding tissue are studied and tumor growth speed and invasiveness are quantified depending on the three parameters. Here, I ask the question: *What are the effects of tumor-associated mechanical cellular properties on the forming tumor?*

In the subsequent Chapter 5, the model complexity is increased and tumor heterogeneity is introduced by enabling phenotype changing mutations in cells. A nutrient-dependent cell cycle, together with spatial constraints leads to a competition of cellular subpopulations over space and nutrients. The evolutionary behavior along two parameters, adhesion, and motility with division-rate trade-off is observed and the effect of a temporally variant nutrient supply is studied. I ask the question: *How does the nutrient environment influence the evolutionary behavior of the single-cell property optima of adhesion and motility?*

The discussion of the results can be found in the respective chapters. Chapter 6, at the end of the thesis, summarizes the results and provides an outlook on future work.

References

[1] E. Schrodinger, "What is life," 1944.

[2] W. Xu, R. Mezencev, B. Kim, L. Wang, J. McDonald, and T. Sulchek, "Cell stiffness is a biomarker of the metastatic potential of ovarian cancer cells," *PloS one*, vol. 7, no. 10, p. e46609, 2012.

[3] Y. L. Han, A. F. Pegoraro, H. Li, K. Li, Y. Yuan, G. Xu, Z. Gu, J. Sun, Y. Hao, S. K. Gupta, *et al.*, "Cell swelling, softening and invasion in a three-dimensional breast cancer model," *Nature physics*, vol. 16, no. 1, pp. 101–108, 2020.

[4] S. N. Abduljauwad *et al.*, "Enhancing cancer cell adhesion with clay nanoparticles for countering metastasis," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[5] M. Janiszewska, M. C. Primi, and T. Izard, "Cell adhesion in cancer: Beyond the migration of single cells," *Journal of Biological Chemistry*, vol. 295, no. 8, pp. 2495–2505, 2020.

[6] M. Yilmaz and G. Christofori, "Mechanisms of motility in metastasizing cells," *Molecular cancer research*, vol. 8, no. 5, pp. 629–642, 2010.

[7] A. Kuşoğlu and Ç. B. Avcı, "Cancer stem cells: A brief review of the current status," *Gene*, vol. 681, pp. 80–85, 2019.

[8] P.-H. Puech, A. Taubenberger, F. Ulrich, M. Krieg, D. J. Muller, and C.-P. Heisenberg, "Measuring cell adhesion forces of primary gastrulating cells from zebrafish using atomic force microscopy," *Journal of cell science*, vol. 118, no. 18, pp. 4199–4206, 2005.

[9] R. M. Hochmuth, "Micropipette aspiration of living cells," *Journal of biomechanics*, vol. 33, no. 1, pp. 15–22, 2000.

[10] F. Reichel, J. Mauer, A. A. Nawaz, G. Gompper, J. Guck, and D. A. Fedosov, "High-throughput microfluidic characterization of erythrocyte shapes and mechanical variability," *Biophysical journal*, vol. 117, no. 1, pp. 14–24, 2019.

[11] A. Saadatpour, S. Lai, G. Guo, and G.-C. Yuan, "Single-cell analysis in cancer genomics," *Trends in Genetics*, vol. 31, no. 10, pp. 576–586, 2015.

[12] X. Ren, B. Kang, and Z. Zhang, "Understanding tumor ecosystems by single-cell sequencing: promises and limitations," *Genome biology*, vol. 19, no. 1, pp. 1–14, 2018.

[13] F. Valdes-Mora, K. Handler, A. M. Law, R. Salomon, S. R. Oakes, C. J. Ormandy, and D. Gallego-Ortega, "Single-cell transcriptomics in cancer immunobiology: the future of precision oncology," *Frontiers in immunology*, vol. 9, p. 2582, 2018.

[14] J. Fan, K. Slowikowski, and F. Zhang, "Single-cell transcriptomics in cancer: Computational challenges and opportunities," *Experimental & Molecular Medicine*, vol. 52, no. 9, pp. 1452–1465, 2020.

[15] U. Alon, *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.

[16] B. Novak and J. J. Tyson, "Modeling the control of dna replication in fission yeast," *Proceedings of the National Academy of Sciences*, vol. 94, no. 17, pp. 9147–9152, 1997.

[17] J. Seo, D. Seong, S. R. Lee, D.-B. Oh, and J. Song, "Post-translational regulation of arf: Perspective in cancer," *Biomolecules*, vol. 10, no. 8, p. 1143, 2020.

[18] W. Du and O. Elemento, "Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies," *Oncogene*, vol. 34, no. 25, pp. 3215–3225, 2015.

[19] A. K. Dasanna, D. A. Fedosov, G. Gompper, and U. S. Schwarz, "State diagram for wall adhesion of red blood cells in shear flow: from crawling to flipping," *Soft matter*, vol. 15, no. 27, pp. 5511–5520, 2019.

[20] H. Noguchi and G. Gompper, "Shape transitions of fluid vesicles and red blood cells in capillary flows," *Proceedings of the National Academy of Sciences*, vol. 102, no. 40, pp. 14159–14164, 2005.

[21] B. Mathew, S. Muñoz-Descalzo, E. Corujo-Simon, C. Schröter, E. H. Stelzer, and S. C. Fischer, "Mouse icm organoids reveal three-dimensional cell fate clustering," *Biophysical journal*, vol. 116, no. 1, pp. 127–141, 2019.

[22] S. B. Nissen, M. Perera, J. M. Gonzalez, S. M. Morgani, M. H. Jensen, K. Sneppen, J. M. Brickman, and A. Trusina, "Four simple rules that are sufficient to generate the mammalian blastocyst," *PLoS biology*, vol. 15, no. 7, p. e2000737, 2017.

[23] J. M. Girkin and M. T. Carvalho, "The light-sheet microscopy revolution," *Journal of Optics*, vol. 20, no. 5, p. 053002, 2018.

[24] J. Jonkman and C. M. Brown, "Any way you slice it—a comparison of confocal microscopy techniques," *Journal of biomolecular techniques: JBT*, vol. 26, no. 2, p. 54, 2015.

[25] A. Ferrari, "Recent technological advancements in traction force microscopy," *Biophysical reviews*, vol. 11, no. 5, pp. 679–681, 2019.

[26] A. Y. Kobitski, J. C. Otte, M. Takamiya, B. Schäfer, J. Mertes, J. Stegmaier, S. Rastegar, F. Rindone, V. Hartmann, R. Stotzka, *et al.*, "An ensemble-averaged, cell density-based digital model of zebrafish embryo development derived from light-sheet microscopy data with single-cell resolution," *Scientific reports*, vol. 5, p. 8601, 2015.

[27] P. L. Townes and J. Holtfreter, "Directed movements and selective adhesion of embryonic amphibian cells," *Journal of experimental zoology*, vol. 128, no. 1, pp. 53–120, 1955.

[28] R. A. Foty and M. S. Steinberg, "The differential adhesion hypothesis: a direct evaluation," *Developmental biology*, vol. 278, no. 1, pp. 255–263, 2005.

[29] F. Graner and J. A. Glazier, "Simulation of biological cell sorting using a two-dimensional extended potts model," *Physical Review Letters*, vol. 69, no. 13, p. 2013, 1992.

[30] L. Wolpert, "Positional information and the spatial pattern of cellular differentiation," *Journal of theoretical biology*, vol. 25, no. 1, pp. 1–47, 1969.

[31] D. Čapek and P. Müller, "Positional information and tissue scaling during development and regeneration," *Development*, vol. 146, no. 24, 2019.

[32] A. Turing, "Philosophical the royal biological transqfctions society sciences," *Phil. Trans. R. Soc. Lond. B*, vol. 237, pp. 37–72, 1952.

[33] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *cell*, vol. 100, no. 1, pp. 57–70, 2000.

[34] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: the next generation," *cell*, vol. 144, no. 5, pp. 646–674, 2011.

[35] J. Hausser and U. Alon, "Tumour heterogeneity and the evolutionary trade-offs of cancer," *Nature Reviews Cancer*, pp. 1–11, 2020.

[36] O. Ilina, P. G. Gritsenko, S. Syga, J. Lippoldt, C. A. La Porta, O. Chepizhko, S. Grosser, M. Vullings, G.-J. Bakker, J. Starruß, *et al.*, "Cell–cell adhesion and 3d matrix confinement determine jamming transitions in breast cancer invasion," *Nature cell biology*, vol. 22, no. 9, pp. 1103–1115, 2020.

[37] K.-J. Streitberger, L. Lilaj, F. Schrank, J. Braun, K.-T. Hoffmann, M. Reiss-Zimmermann, J. A. Käs, and I. Sack, "How tissue fluidity influences brain tumor progression," *Proceedings of the National Academy of Sciences*, vol. 117, no. 1, pp. 128–134, 2020.

[38] H. Cui, X. Wang, J. Wesslowski, T. Tronser, J. Rosenbauer, A. Schug, G. Davidson, A. A. Popova, and P. A. Levkin, "Assembly of multi-spheroid cellular architectures by programmable droplet merging," *Advanced Materials*, p. 2006434, 2020.

[39] M. S. Ozturk, V. K. Lee, H. Zou, R. H. Friedel, X. Intes, and G. Dai, "High-resolution tomographic analysis of in vitro 3d glioblastoma tumor model under long-term drug treatment," *Science Advances*, vol. 6, no. 10, p. eaay7513, 2020.

[40] I. R. Powley, M. Patel, G. Miles, H. Pringle, L. Howells, A. Thomas, C. Kettleborough, J. Bryans, T. Hammonds, M. MacFarlane, *et al.*, "Patient-derived explants (pdes) as a powerful preclinical platform for anti-cancer drug and biomarker discovery," *British journal of cancer*, pp. 1–10, 2020.

[41] H. W. Jackson, J. R. Fischer, V. R. Zanotelli, H. R. Ali, R. Mechera, S. D. Soysal, H. Moch, S. Muenst, Z. Varga, W. P. Weber, *et al.*, "The single-cell pathology landscape of breast cancer," *Nature*, vol. 578, no. 7796, pp. 615–620, 2020.

[42] M. Allam, S. Cai, and A. F. Coskun, "Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics," *NPJ Precision Oncology*, vol. 4, no. 1, pp. 1–14, 2020.

[43] C. Xia, H. P. Babcock, J. R. Moffitt, and X. Zhuang, "Multiplexed detection of rna using merfish and branched dna amplification," *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.

[44] M. Sadati, N. T. Qazvini, R. Krishnan, C. Y. Park, and J. J. Fredberg, "Collective migration and cell jamming," *Differentiation*, vol. 86, no. 3, pp. 121–125, 2013.

[45] W. Kang, J. Ferruzzi, C.-P. Spatarelu, Y. L. Han, Y. Sharma, S. A. Koehler, J. A. Mitchel, J. P. Butler, D. Roblyer, M. H. Zaman, J.-A. Park, M. Guo, Z. Chen, A. F. Pegoraro, and J. J. Fredberg, "Tumor invasion as non-equilibrium phase separation," *bioRxiv*, 2020.

[46] H. Murphy, H. Jaafari, and H. M. Dobrovolny, "Differences in predictions of ode models of tumor growth: a cautionary example," *BMC cancer*, vol. 16, no. 1, pp. 1–10, 2016.

[47] G. D. Grass, J. C. L. Alfonso, E. A. Welsh, K. Ahmed, J. Teer, L. B. Harrison, J. Cleveland, J. Mule, S. Eschrich, H. Enderling, *et al.*, "Harnessing tumor immune ecosystem dynamics to personalize radiotherapy," *bioRxiv*, 2020.

[48] R. A. Beckman, I. Kareva, and F. R. Adler, "How should cancer models be constructed?," *Cancer Control*, vol. 27, no. 1, p. 1073274820962008, 2020.

[49] D. Ho, "Artificial intelligence in cancer therapy," *Science*, vol. 367, no. 6481, pp. 982–983, 2020.

[50] M. Berghoff, J. Rosenbauer, F. Hoffmann, and A. Schug, "Cells in silico–introducing a high-performance framework for large-scale tissue modeling," *BMC bioinformatics*, vol. 21, no. 1, pp. 1–21, 2020.

[51] T. Büscher, N. Ganai, G. Gompper, and J. Elgeti, "Tissue evolution: Mechanical interplay of adhesion, pressure, and heterogeneity," *New Journal of Physics*, vol. 22, no. 3, p. 033048, 2020.

[52] 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," *PLoS computational biology*, vol. 14, no. 2, p. e1005991, 2018.

[53] S. Tanaka, D. Sichau, and D. Iber, "Lbibcell: a cell-based simulation environment for morphogenetic problems," *Bioinformatics*, vol. 31, no. 14, pp. 2340–2347, 2015.

[54] B. Waclaw, I. Bozic, M. E. Pittman, R. H. Hruban, B. Vogelstein, and M. A. Nowak, "A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity," *Nature*, vol. 525, no. 7568, p. 261, 2015.

[55] M. H. Swat, G. L. Thomas, J. M. Belmonte, A. Shirinifard, D. Hmeljak, and J. A. Glazier, "Multi-scale modeling of tissues using compucell3d," in *Methods in cell biology*, vol. 110, pp. 325–366, Elsevier, 2012.

[56] J. Starruß, W. de Back, L. Brusch, and A. Deutsch, "Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology," *Bioinformatics*, vol. 30, no. 9, pp. 1331–1332, 2014.

[57] G. R. Mirams, C. J. Arthurs, M. O. Bernabeu, R. Bordas, J. Cooper, A. Corrias, Y. Davit, S.-J. Dunn, A. G. Fletcher, D. G. Harvey, *et al.*, "Chaste: an open source c++ library for computational physiology and biology," *PLoS Comput Biol*, vol. 9, no. 3, p. e1002970, 2013.

[58] 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," *Proceedings of the National Academy of Sciences*, vol. 113, no. 48, pp. E7663–E7671, 2016.

[59] P. Van Liedekerke, M. Palm, N. Jagiella, and D. Drasdo, "Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results," *Computational particle mechanics*, vol. 2, no. 4, pp. 401–444, 2015.

[60] J. Rosenbauer, C. Zhang, B. Mattes, I. Reinartz, K. Wedgwood, S. Schindler, C. Sinner, S. Scholpp, and A. Schug, "Modeling of wnt-mediated tissue patterning in vertebrate embryogenesis," *PLOS Computational Biology*, vol. 16, no. 6, p. e1007417, 2020.

[61] E. Stanganello, A. I. Hagemann, B. Mattes, C. Sinner, D. Meyen, S. Weber, A. Schug, E. Raz, and S. Scholpp, "Filopodia-based wnt transport during vertebrate tissue patterning," *Nature communications*, vol. 6, p. 5846, 2015.

[62] A. Voss-Böhme, "Multi-scale modeling in morphogenesis: a critical analysis of the cellular potts model," *PloS one*, vol. 7, no. 9, p. e42852, 2012.

[63] J. Metzcar, Y. Wang, R. Heiland, and P. Macklin, "A review of cell-based computational modeling in cancer biology," *JCO clinical cancer informatics*, vol. 2, pp. 1–13, 2019.

[64] E. G. Rens and L. Edelstein-Keshet, "From energy to cellular forces in the cellular potts model: An algorithmic approach," *PLoS computational biology*, vol. 15, no. 12, p. e1007459, 2019.

[65] A. J. Lopatkin and J. J. Collins, "Predictive biology: modelling, understanding and harnessing microbial complexity," *Nature Reviews Microbiology*, pp. 1–14, 2020.

[66] M. T. McKenna, J. A. Weis, A. Brock, V. Quaranta, and T. E. Yankeelov, "Precision medicine with imprecise therapy: computational modeling for chemotherapy in breast cancer," *Translational oncology*, vol. 11, no. 3, pp. 732–742, 2018.

[67] V. Marx, "Method of the year: spatially resolved transcriptomics," *Nature Methods*, vol. 18, no. 1, pp. 9–14, 2021.

[68] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8–17, 2015.

[69] M. A. Clarke and J. Fisher, "Executable cancer models: successes and challenges," *Nature Reviews Cancer*, pp. 1–12, 2020.

[70] A. Fletcher and J. Osborne, "Seven challenges in the multiscale modelling of multicellular tissues," 2020.

[71] P. Macklin, "Key challenges facing data-driven multicellular systems biology," *GigaScience*, vol. 8, no. 10, p. giz127, 2019.

2 Publication 1: Cells in Silico, a high-performance framework for tissue simulations

THE *publication 'Cells in Silico – introducing a high-performance framework for large-scale tissue modeling' w as published on October* 6^{th} *2020 in BMC Bioinformatics. Marco Berghoff and I are the first authors of this paper. The complete paper can be found in Appendix 1 of this thesis.*

2.1 Summary

Here, I describe the development of a parallel implementation of the CPM, by domain decomposition. I demonstrate that the energy calculation of the Hamiltonian energy function can be performed using only nearest-neighbor lattice points. This property of a local energy calculation, together with the nearest-neighbor interactions of the CPM facilitate the parallel execution of subdomains. The parallelization is realized by a domain decomposition of the regular grid into blocks. Each block is then propagated using one CPU core. Around each block, a halo of values from the neighboring blocks is established and updated after each time step. Cell metadata, such as the volume, surface, and signal values of the cells are incrementally updated by data exchange between neighboring blocks. By limiting the communication to nearest-neighbor communications of blocks and avoiding all-to-all communications, high scalability is achieved. The framework is realized using C++ with MPI for the communication between cores. Here, the implementation of the parallel version of the CPM is described in detail and the necessary assumptions are introduced. The parallel simulation model is validated with a non-parallel version of the code. The scaling behavior of the code for up to 45 thousand CPU cores is shown on a modern supercomputer. The usage and user interface of the developed framework is introduced and documented. Finally, I demonstrate its ability to solve biological and physical questions by performing large simulations of tissue dynamics.

The framework is available under an open-source license at `https://gitlab.com/nastja/nastja`. The full publication is available in the appendix of this thesis (Appendix 1).

2.2 Contribution

The development and implementation of the model and the documentation in the article was made in close collaboration between Marco Berghoff and myself, with equal contributions.

3 A generalized model of tumor growth shows emergent behavior from single-cell events

In *this chapter, I introduce a generalized computational model of tumor growth. Proliferative cells with a nutrient-dependent cell cycle are simulated in surrounding tissue and vasculature. The tumor cells proliferate and form a three-dimensional tumor, allowing the observation of emergent behavior from single-cell properties. Tumor heterogeneity is introduced by mutations of single cells and the effect of diffusible treatment compounds (chemotherapy) and global interference (radiation therapy) on tumor development is modeled.*
The tumors form complex three-dimensional shapes in dependency of the vasculature and the influence of varying treatment schemes on tumor heterogeneity is observed. By parametrizing the properties and behavior of single cells, the emergent behavior on a large scale is demonstrated by the simulation of 10^6 *interacting cells.*

Tumor progression is based on many complex effects acting concurrently to facilitate the uncontrolled growth of some cells. Cell-internal processes, e.g., mutations upregulating cell division or chemotherapeutic resistance, have a significant impact on the size, shape, and heterogeneity of the final tumor and strongly affect the treatment response [1] and drug resistance [2]. Unfortunately, while microscopic properties are of high importance, they are clinically poorly accessible. Hence, the treatment protocol and prognosis are inferred based on accessible macroscopic properties such as patient condition, tumor size as visible in MRI scans, and biopsies. Bridging the scales between experimental single-cell findings and clinical data can improve the understanding of cancer as an emerging property of its cellular composition [3]. This link would allow optimization of treatment. Ideally, personalized treatment strategies could be optimized by modeling the outcome of different treatment regimes.

One option for predicting tumor growth is leveraging the exponentially increasing computing capabilities of modern supercomputers. A crucial ingredient is the simulation parametrization, which is fueled by new microscopy techniques and genomic tools that have made immense progress in the observation of cells, tissue, and the temporal evolution of those [4–6] as well as gene expression and mutations [7]. This already has driven the modeling of tissue development and dynamics in the related fields of embryogenesis [8, 9], morphogenesis, tissue dynamics, homeostasis [10–17], and tumor growth [18–23] simulations. These and other tissue modeling approaches paint an increasingly detailed picture, enabling predictive simulations that can be verified by experiments and *vice versa* for a large variety of biological phenomena [24, 25]. Still, the scope of cancer simulations is either the detailed description of individual cells or large numbers of cells as point-like agents or a coarse-grained description of tissue [26]. Nevertheless, both ends of the resolution range are necessary to map the complexity of tumor development, since both the individual cells and the macroscopic environment play a crucial role.

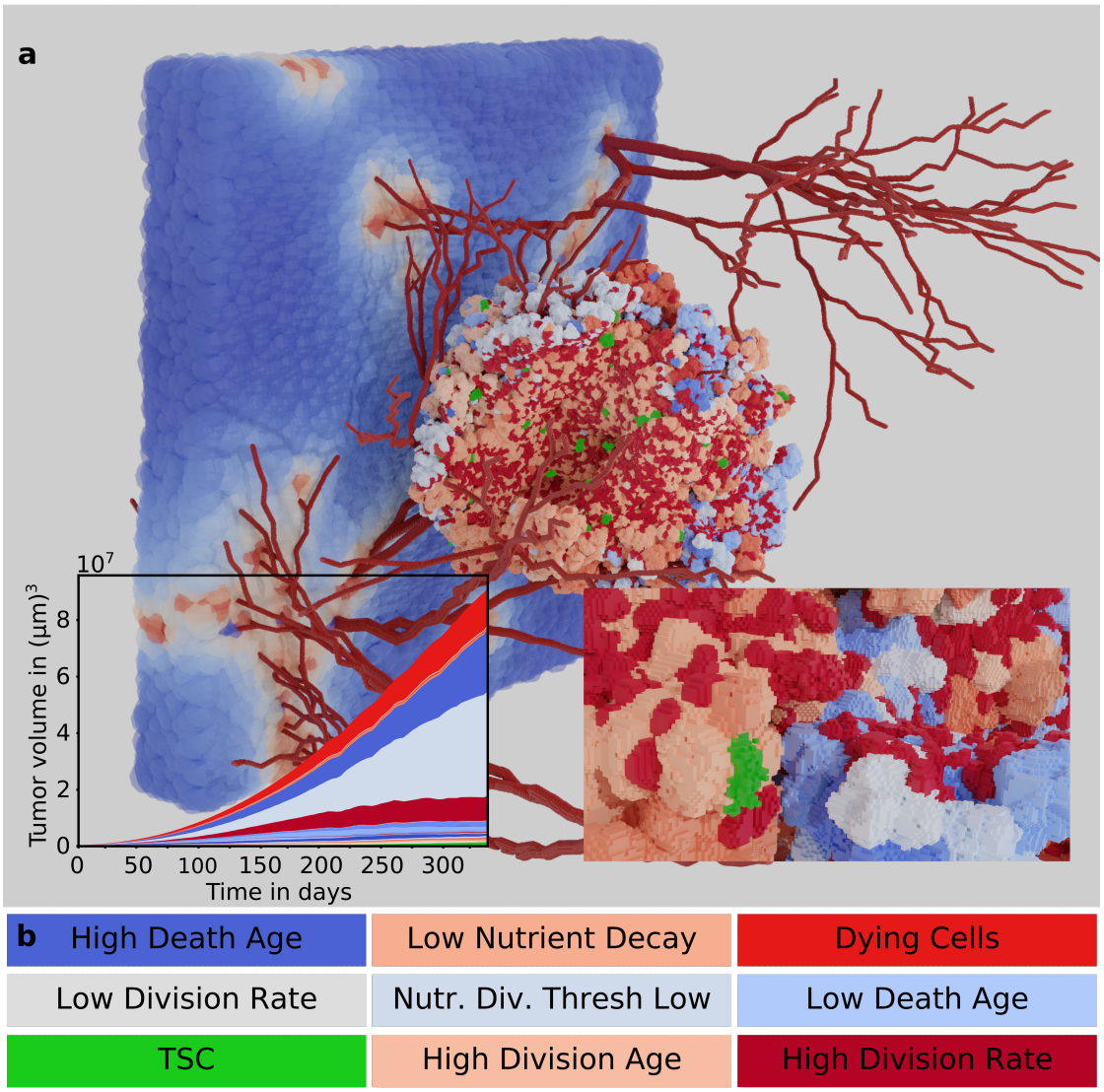

Figure 4: **High-resolution Tumor Simulation:** a) $1\,\mathrm{mm}^3$ tissue simulation with $1\,\mu\mathrm{m}$ resolution of single cells. Blood vessels (dark red) distribute nutrients (blue, white, and red on the sheet in the background representing low, medium, and high concentrations) that facilitate cell divisions and tumor expansion. Heterogeneous tumor growth (colors represent different cell types) results in emergent behavior of nutrient dependant cell-division, -death, and mutations. The inlays show growth of cell types over time (left), a zoom-in on the tumor surface highlights the µm resolution. b) The color-coding of the majorly contributing cell types, each color indicates one cell type with its individual parameter-set, colors are used for the remainder of the figures (for all cell-types see Appendix 3:Figure 20).

Here, we demonstrate the ability of a parallel model with a single-cell resolution to generate emergent large-scale behavior on the tissue scale ($O(\text{mm})$) from single-cell effects. We simulate the growth of a tumor in vascularized homogeneous tissue. Our cancer simulation considers competing single-cell effects leading to emerging tissue scale behavior. We introduce a computational microscope that enables access to time-resolved trajectories of all included cellular properties, going beyond what is accessible in wet-lab experiments. We see that highly proliferative cells in a surrounding tissue form tumors of distinct shapes. The introduction of a nutrition-dependent cell cycle leads to hypoxic and necrotic regions but also requires subcellular resolution to treat nutrient flow and other surface-based cell-to-cell interactions realistically. Tumor heterogeneity is incorporated by the mutation of cells into predefined cell types reflecting driver mutations. The effect of growth-inhibiting diffusible substances (representing a chemotherapy drug) and global damage of tumor cells (representing radiotherapy) is introduced. The impact of different dosing schemes on tumor growth and heterogeneity is compared. We show simulations of tissues up to a clinically relevant size of mm^3 composed of over a million individual geometry-resolved cells over time scales up to a year at a temporal resolution of a minute.

3.1 Methods

Naturally, one has to balance model abstraction with its complexity, parameter availability, and computational cost. Here, we choose to focus to explicitly model the cell geometry of both cancerous and regular tissue, mutations and cancer heterogeneity, nutrient availability from blood vessels (but no angiogenesis), and treatment and resistance development to both chemo- and radiotherapy in the context of the host environment. More specifically, our spatiotemporal multiscale model describes the collective behavior of $O(1\text{ Mio})$ individual cells to simulate a macroscopic tissue of $O(1\,\text{mm}^3)(=1\,000^3$ voxel, cf. Figure 4). The model consists of three layers and is implemented in the CellsInSilico framework [27]. Figure 5 depicts an overview of the layers and model parts that are acting on them.

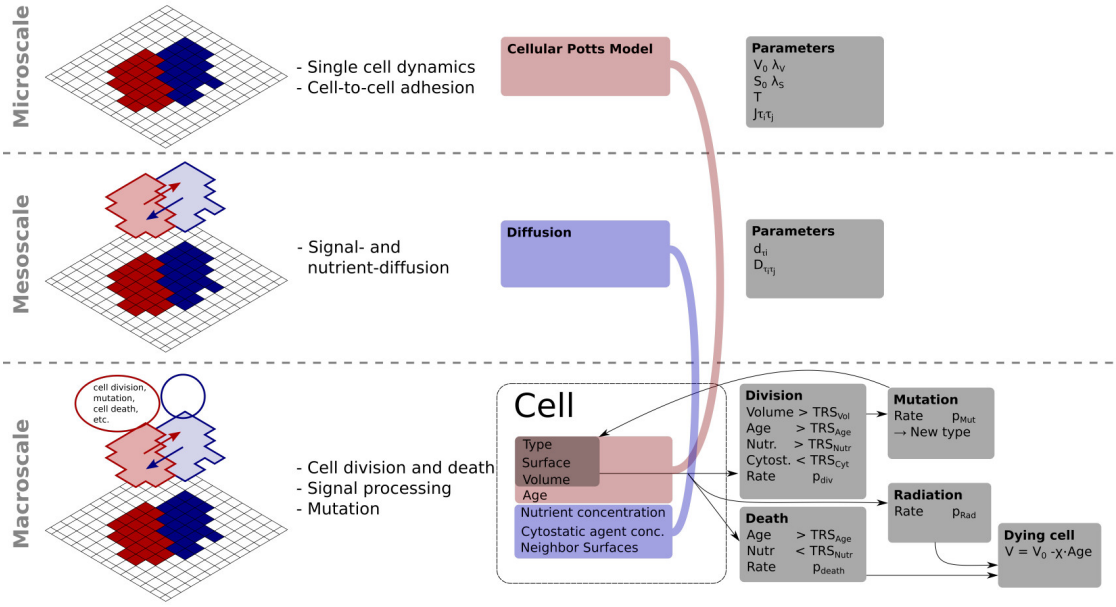

Figure 5: **Simulation model.** The schematic shows an overview of the simulation layers and actions of the agent-based model used in Cells in Silico.

Each layer represents a different length scale. Our model is based on a cellular Potts model acting on the microscopic layer, see Section 3.1.1. The diffusion of signals such as nutrients and drugs through the simulated area is modeled on the mesoscopic layer, which is described in Section 3.1.2. An agent-based model controls cellular events such as cell division and cell death; this macroscopic layer is explained in Section 3.1.3.

The parallel version of our model is implemented using the NAStJA framework [28]. Besides, synchronization steps ensure a consistent state of the entire domain; these are the halo exchange and the local exchange of global cell properties is described in Refs. [27, 29]. Most simulations in this study use a cubic box with an edge length of $320\,\mu\text{m}$, with selected simulations using a box edge length of $1\,000\,\mu\text{m}$. The computational framework, *Cells In Silico*, handles the distribution of computing load resulting in superlinear speedup on large CPU-core numbers $(O(10^5))$ on supercomputers.

For detailed model parameters Appendix 3. The published open-source package of *Cells In Silico* in the NASTjA framework can be found at `https://gitlab.com/nastja/nastja`.

3.1.1 Microscale: Cellular Potts Model and Hamiltonian

The lowest layer is a 3D Cellular Potts model (CPM) layer which models the cells on a grid [13, 30–32]. The CPM is based on a Hamiltonian with local interactions. Modeling of cell-cell adhesion proportional to the cell interfacial surface, nutrient transport, and cell-to-cell signaling properties are explicitly dependent on the shape and surface of each cell, with each cell occupying around 10^3 voxel corresponding to $1\,000\,\mu\text{m}^3$.

The CPM defines the mechanical properties of the cells: compressibility, volume constraints, and adhesive forces. The Hamiltonian energy is defined as in the Introduction, 1.1.1. Three energy terms are introduced, namely surface, volume, and adhesion energy, that make up the Hamiltonian

and represent a basic CPM implementation without additional energy terms. Periodic boundary conditions reflect the behavior of an extended macroscopic tissue.

3.1.2 Mesoscale: Signal and Nutrient Transport

In the intermediate layer, the diffusion of signaling compounds, nutrition, and chemotherapeutic drugs is realized by flux through the cell membrane to the adjoining neighbors of a cell. The simulation considers the transmission and propagation of multiple substances, such as nutrients and drugs. We define a class of signaling, e.g., nutrient contents, of each cell $\sigma_i^{(\varrho)}$, denoting the concentration of signal ϱ in cell i. Those represent either oxygen and glucose as nutrients for the cell, cell-to-cell signaling compounds, or drugs. The diffusion of nutrients can be approximated by flow through the surfaces of the cells. Actions, such as cell division, -death, and -mutations, may depend on these signals and nutrient contents.

Diffusion Diffusion of signals between the cells occurs through the surface of these cells. We determine the shared surface $S_{i,j}$ for each pair of cells i, j with $i \neq j$. The diffusion depends on the type of cells, so we define for each combination of types a diffusion constant $D_{\tau(i),\tau(j)}$, $\tau(i)$ denoting the cell type of cell i. The flux $J_{i,j}^{(\varrho)}$ for a signal ϱ is defined by

$$J_{i,j}^{(\varrho)} = \left(\frac{S_{i,j}}{S_i} + \frac{S_{i,j}}{S_j} \right) D_{\tau(i),\tau(j)} (\sigma_j^{(\varrho)} - \sigma_i^{(\varrho)}),$$

where S_i is the surface from cell i and σ_i is the signal value in cell i and S_j, σ_j from cell j, respectively. The first bracket is the arithmetic mean of the two surface fractions with respect to the common surface. The flux $J_{i,j}$ is subtracted from the signal of one cell and added to the other. Here, we distinguish between cells and fixed signal suppliers, such as blood vessels. For fixed signal supplies, the signal content is kept constant, i.e., the flux is neither subtracted nor added to those cells.

Decay Metabolic processes occur inside the cells. We used a simple model in which the signals are changed relative to their values,

$$\sigma_i^{(\varrho)}\Big|_{t+1} = d_{\tau(i)}^{(\varrho)} \cdot \sigma_i^{(\varrho)}\Big|_t .$$

Where $\sigma_i^{(\varrho)}\Big|_t$ is the signal ϱ in cell i at time t and $d_{\tau(i)}^{(\varrho)}$ is the relative change of the signal ϱ depending on the type of cell i.

3.1.3 Macroscale: Agent-based Model

The top layer is an *agent-based* model that handles cell phenotype parametrization, cell internal signal processing, cell division, cell death, and mutation of cells. The surrounding tissue is initialized as a nondividing and nondying population. Cell division and cell death depend on cell age, nutrient availability, cytostatic drug concentration, as well as division and death rates. On the macroscale, cell attributes such as cell age, signal level, cell type, etc. are used to generate

actions based on these values. The parameters can be linked to cell biological experiments and simulations (cf. Appendix 3;Table 3).

Cell types A cell type is assigned to each cell, which determines the parametrization and phenotype of that cell. The cell type defines the characteristics of the individual cells, i.e., the target volume V_0, the target surface S_0, and the thresholds TRS_{Vol}, TRS_{Age}. In that way, not each cell has to be individually parameterized. Cells that divide usually generate two new cells of the previous cell type. Cell types allow the classification of each cell in the simulation and the tracking of cell type subpopulations. Through the definition of a predefined set of cell types and mutations between those types, the parameter space is controlled, and the accumulation of purely favorable traits in a single cell type is prevented. Each cell type only has a single variation with respect to the initial tumor cells. It is possible to define an arbitrary number of cell types for different use cases. Here we define a set of 27 cell types for heterogeneous tumor growth (see Appendix 3;Figure 20)

Blood vessels and solid We introduced a subset of cells that are not participating in the spatio-temporal propagation via the cellular Potts model. Those cells are solid structures, which can model blood vessels or the extracellular matrix. They are able to participate in cell-to-cell signaling and may act as sources for signals.

Cell division In each time step, each cell is checked for cell division. Whether a cell divides depends on several internal and external factors. Division conditions are:

- Volume above a threshold $V > V_{\text{Div}} = 0.9 \cdot V_0$
- Nutrition above a threshold C_{DivMin}
- Age above a certain threshold $\text{Age}_{\text{DivMin}}$
- Comparing random number $\in [0, 1]$ with the division rate R_{Div}
- Chemotherapy content below a threshold TRS_{Ch}

If all conditions are met, a random plane through the cell center is chosen and the cell is split along that plane. After cell division, the cellular age is set to zero. Post division, both cells expand due to pressure by the volume and the surface energy term. Specific cell types can also be excluded from cell division, such as the surrounding tissue in our simulations.

Cell death Cell death conditions are:

- Nutrition below a threshold C_{Death}
- Age above a certain threshold $\text{Age}_{\text{DeathMin}}$
- Comparing random number $\in [0, 1]$ with death Rate R_{Death}
- If all above conditions are not met comparing random $\in [0, 1]$ with default death Rate $R_{\text{Death}}/1000$ to account for natural cell death

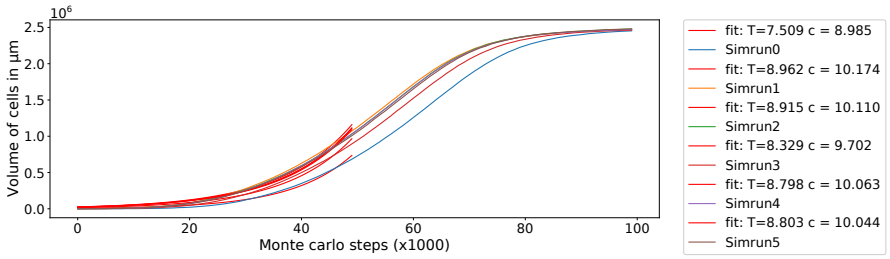

Figure 6: **Growth of a tumor in surrounding tissue without external influences is used to define the timestep**. Multiple simulation runs with different seeds are shown, and the time constant of the growth is fitted.

Cell death is induced by changing the cell type of the cell to a dedicated cell type that describes dying cells. For this cell type, the goal volume in the Hamiltonian is changed over time $V_{0\text{Apop}}(t) = V_0 - \chi \cdot \text{age}$, effectively lowering the volume of the cell to zero voxels. Once the cell reaches $V = 0$, the cell is deleted.

Mutation Mutations are possible events accompanying cell division, assigning a new phenotype to one of the daughter cells. After cell division, the two daughter cells are reinitialized. If a mutation event occurs (mutation rate R_{Mut}), one of the daughter cells is initialized with a randomly chosen cell type. The range of cell types that can be chosen is predefined. The transition matrix between all cell types can be defined so that the transition probabilities between cell types vary. Here, we use a constant transition probability.

Treatments Chemotherapy is implemented as a diffusive drug that suppresses cell division and is distributed via blood vessels. Radiotherapy introduces immediate cell death of a fraction of cells and globally reduces division rates proportionally to accumulated radiation exposure. The model parameters are largely based on experimental measurements (cf. Appendix 3;Tables 3,4).

3.1.4 Time-step

The simulated time of the *in silico* tumor growth is not explicitly defined by the simulation parameters. To correlate the simulated time with real-time, the growth curves of the simulated tumor were compared with tumor growth curves in animal models. Assuming the growth to be of similar speed, the simulated time in Monte Carlo Sweep (MCS) can be linked to real-time in hours. We observe an initial exponential growth of the *in silico* tumor when modeling the free growth of a tumor in an environment with sufficient nutrients and without treatment. We find that a size doubling time of $T = 8\,552\,\text{MonteCarloSweeps(MCS)}$ in our model corresponds to an untreated tumor with an experimentally determined doubling time of $150\,\text{h}$ (see Figure 6). Comparison with the growth observed *in vivo* (cf. Appendix 3;Table 3) we assume exponential growth of the tumor volume $V_{\text{Tumor}}(t) = \exp(\ln(2) \cdot t/T) + c$ and derive $T = 150\,\text{h}$ from the experimental data for a non-treated tumor. Therefore, $1\,\text{MCS}$ equals $1.05\,\text{min}$ and $1\,\text{kMCS} = 0.73\,\text{days}$.

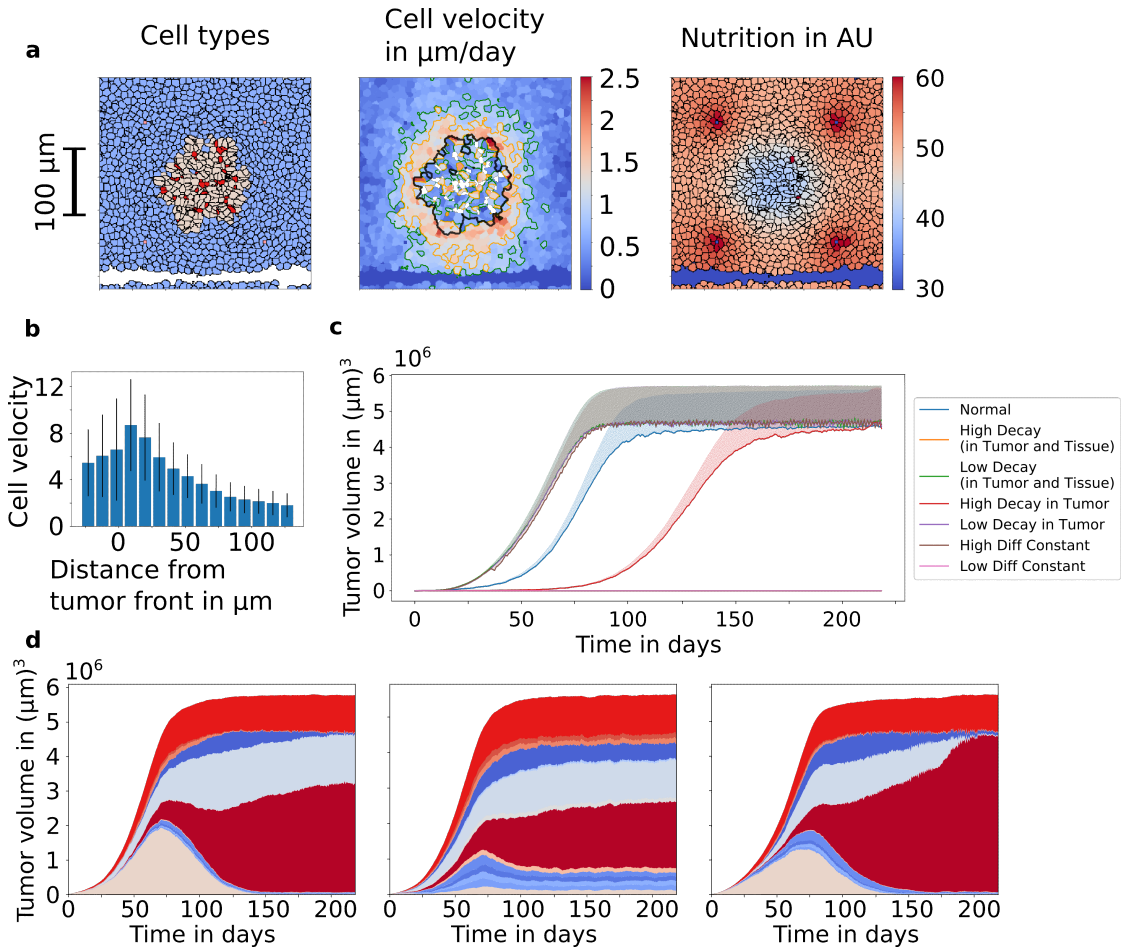

Figure 7: **Buildup of the Model:** a) 2D slices of tumor simulation with a rectangular blood vessel grid. Left to right: Coloring by cell types (Types see Figure 4), coloring by averaged cell velocities (black line indicates tumor outline), coloring by nutrient availability b) Cellular velocity dependent on the distance to the tumor front, negative values are inside the tumor. c) Dependence of the tumor growth rates on metabolism parameters. Parameter variations in the decay of nutrients and diffusion constant in the tissue and tumor change the growth rates of the tumors. The shaded area indicates the volume of dying cells. d) Tumor heterogeneity through mutation after cell division. A variation of the mutation rate results in different tumor compositions. The simulation on the left has a medium, in the center a high mutation rate. On the right, the blood vessel configuration was changed for medium mutation rates. Coloring as in Figure 4.

3.2 Results

3.2.1 Homogeneous Tumor Growth

Simulations of a tumor, consisting of a single-non mutating cell type are performed in a simulation box with an edge length of 320 µm. 35 cells of a proliferative cell type are placed in a simulation box filled with non-proliferative and nondying 'surrounding' cells. The development of the forming tumor is observed over a simulated time of one year with a one-minute time step. Over this period, one can observe homogeneous tumor growth (i.e., composed of a single cell type) into the surrounding tissue. Blood vessels are represented by rigid structures in the simulation, that are symmetrically placed as four straight structures along each axis (cf. Figure 7 a)). Nutrients

distributed by those blood vessels represent a growth-limiting factor for the growing tissue. In the simulations, nutrients diffuse from blood vessels through the tissue, with each cell degrading the local nutrition concentration (cf. Figure 7 a) right). A gradient of nutrient concentration develops originating from the blood vessels.

As time progresses, the tumor grows in the surrounding tissue. The tumor reaches a finite size once the simulated volume is entirely filled by cells. Due to the volume constraints of cell division and the constant number of surrounding cells, the absolute number of cells and thereby the tumor size is limited. Once the maximal number of cells is reached, cell death and proliferation of the tumor cells are in equilibrium. An upregulated metabolism in tumor cells (represented by a higher depletion of nutrients in the cells) leads to a significantly decreased growth rate, while a downregulated metabolism leads to faster growth of the tumor (cf. Figure 7 c)). Tumor cells deplete nutrients at a higher rate leading to the formation of experimentally known intermediate states such as invasive fingers and hypoxic or even necrotic areas in the center of the tumor. Our analysis of cell velocities and cell displacement shows highly mobile or dynamic tumor cells at the boundary of the tumor. In contrast, the cell movements within the tumor and in the surrounding tissue are much lower (cf. Figure 7 b)). Cell density and velocity have been associated with tumor invasion, jamming, and unjamming transitions within a tumor [33].

3.2.2 Heterogeneity

Primary tumors develop over long periods of time, and tumor internal heterogeneity arises from mutations in cells. The limited inflow of nutrients leads to a competition of the cell phenotypes, and the partition into subpopulations indicates the fitness of the individual cell types.
This mutation process continues during the whole lifetime of the tumor. Hence, a tumor does not consist of just one cancerous cell type, but a variety of cellular phenotypes that compete over resources and complicate tumor treatment since the effect of therapies may differ between the types. Therefore, tumor heterogeneity is a crucial factor in planning cancer treatment [34].
We implemented the mutation of cells into our model by introducing the possibility of a change of cell type of one of the daughter cells after division. These mutations are reflecting one or multiple somatic mutations that affect the behavior of the cell and lead to altered behavior. Each of the predefined cell types represents one cellular phenotype and has exactly one parameter up-or down-regulated with respect to the tumor type introduced in 3.2.1. The transition rates are constant from and to every cell type. Cell-to-cell adhesion, cell division age and nutrient thresholds, cell death age and thresholds as well as nutrient uptake, and division and death rates were altered.

We run a set of simulations with different mutation rates and equal transition rates between the predefined tumor cell phenotypes. As visible in Figure 7 d), the final heterogeneity of a tumor strongly depends on the mutation rate. At low (every 200th division) and medium (every 20th division) mutation rates around day 70, the initial tumor cells dominate the tumor mass but get outcompeted with time as the total size of the tumor is stunted by the lack of compressible surrounding tissue. At low mutation rates, cell types with a higher division rate and delayed cell death begin to dominate the tumor after day 70. Medium mutation rates lead to similar yet

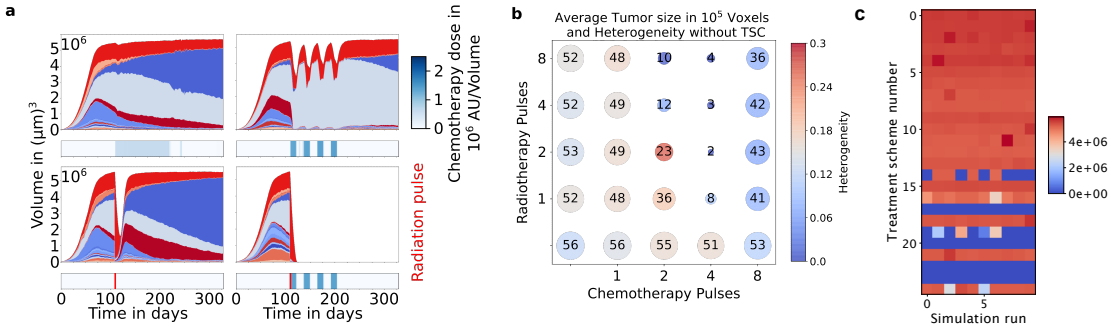

Figure 8: **Treatment Response:** a) Tissue size and composition response to different treatment regimes of constant accumulated doses of chemo- and radiotherapy. Colouring as in Figure 4. The treatment protocol of chemo- (blue) and radiotherapy (red) are depicted below the growth curve. b) Integrated tumor size (bubble size) post-treatment and tumor heterogeneity (bubble color). c) The final tumor size of different treatment schemes (y-) and different simulation runs (x-axis) shows the stability of the simulation outcome. Coloring by tumor size at $t = 325$ days.

accelerated qualitative behavior. To observe the influence of the local environment, we increase the blood vessel density in the simulation, leading to a more rapid preeminence of fast-dividing cells.

3.2.3 Probing Treatment Regimes

Models to optimize chemotherapy dosage have been implemented and convincingly used as early as the 70s [35, 36]. Figure 8 a) depicts the response of simulated heterogeneous tumors to different treatment schemes. We assume detection and onset of treatment of the tumor from day 110 until day 220. The drop in the total tumor size post-treatment until the final size at day 330 strongly depends on the treatment protocol. For all different treatment protocols, the total dose of a therapeutic agent stays constant. It is redistributed into shorter peaks with higher concentrations and within the same time frame between days 110 to 220.

Shorter pulses of chemotherapy show a greater effect than a uniform application, whereas a single strong radiotherapy pulse reduces the tumor size more drastically than multiple weaker pulses. Multiple pulses of radiation and chemotherapy shift the tumor composition towards a more homogeneous tumor by a cumulative adaptation through advantageous cell types surviving. This coincides with studies, which identify tumor heterogeneity as a driving force in treatment resistance [34]. Surprisingly, cell types with increased resistance to chemo- or radiotherapy are less favored than fast-dividing cells in relapse post-treatment. Combinations of therapies using chemo and radiotherapy show a more significant effect on the tumor as one of the methods alone since the growth-inhibiting effect is two-fold. As clearly visible in Figure 8 b), treatment effectivity increases when going from one to four pulses of chemotherapy but then drastically decreases for eight pulses. For radiotherapy, the effect on the tumor also increases when dividing the dose into smaller pulses.

Thus, we can probe the treatment regimes for a given tumor. We can systematically probe the treatment effects of different treatment regimes and combinations and judge the effectivity based on tumor properties.

3.2.4 Stability

To provide predictions for the temporal development of tissue *in vitro* and *in silico* for clinical applications, knowledge about the statistics and robustness of the system development is essential. Variability in the simulation outcome is visible when running the same simulation with different random seeds. This reveals the impact of random and rare events on macroscopic tumor development. The deviation in tumor size and heterogeneity in Figure 8 b) is neglectable, where the treatment weakly impacts the tumor size. Whereas for treatment schemes that drastically reduce the tumor size, the local surrounding and rare events have a more significant impact and lead to greater variability in the subsequent development. Figure 8 c) depicts the stochasticity of the final tumor volume after different treatment protocols and shows increased variability in treatment schemes. In some cases, rare events can toggle between the disappearance or relapse of the tumor post-treatment.

3.2.5 Scaling up

To take full advantage of the supercomputing architecture and parallelization, the simulation of heterogeneous tumor growth described in the previous sections is scaled up from $320\,\mu m$ to $1000\mu m$ size of the simulation box. The increase in the system size leads to an absolute number of cells of ≈ 1.1 million. We introduce an irregular vasculature system (cf. Fig. 4), that distributes nutrients. Due to the large system size and irregular symmetry, complex spatial structures arise as the tumor expands. Vicinity to blood vessels enables faster growth of the tumor, which leads to preferential growth of the tumor along blood vessels.

3.3 Discussion

Here, we highlight the possibility of simulating emerging macroscopic tumor development resulting from microscopically explicit shape-represented single cells. The high computing and data handling complexity can be mitigated via current-day supercomputing capabilities. The model makes it possible to test arbitrary 'what-if' scenarios, unrestricted by experimental constraints, with direct control over all parameters of each individual cell. In our virtual tumors, we observe nutrition-dependent heterogeneous tumor growth. We find emergent behavior of the tumor growth on the mm scale, that arises from the collective interactions of a large number of individual cells with a spatial extent of $\approx 10\mu$m each. The properties and behavior of the individual cells alone enable the behavior that can be observed on a much larger scale.

We can show that different treatment plans strongly influence the final tumor cell type composition. We model cancer therapeutic agents in our system and show agreement with experimentally measured behaviors, reflecting growth curves [37] cf. Appendix 3:Table 4. Each simulation results in a fully spatiotemporally resolved trajectory, which allows tracing even single-cell events. We observe that the tumor growth is mostly driven close to the surface of the tumor and can investigate the changing tumor composition over time. In some simulation regimes, rare events influence not only details of the individual simulation but can influence the macroscopic behavior, such as the resurgence of tumors post-treatment. This improved theoretical accessibility of cancer growth as emerging behavior opens new research avenues. One could envision the application in improving early-stage cancer detection by characterizing detectible early growth pathways. Once parameter sets for specific cancer types have been developed, such simulations could revolutionize clinical treatment via optimized, personalized medicine regimes *in silico*.

References

[1] N. McGranahan and C. Swanton, "Clonal heterogeneity and tumor evolution: past, present, and the future," *Cell*, vol. 168, no. 4, pp. 613–628, 2017.

[2] N. Vasan, J. Baselga, and D. M. Hyman, "A view on drug resistance in cancer," *Nature*, vol. 575, no. 7782, pp. 299–309, 2019.

[3] X. Ren, B. Kang, and Z. Zhang, "Understanding tumor ecosystems by single-cell sequencing: promises and limitations," *Genome biology*, vol. 19, no. 1, pp. 1–14, 2018.

[4] J. M. Girkin and M. T. Carvalho, "The light-sheet microscopy revolution," *Journal of Optics*, vol. 20, no. 5, p. 053002, 2018.

[5] J. Jonkman and C. M. Brown, "Any way you slice it—a comparison of confocal microscopy techniques," *Journal of biomolecular techniques: JBT*, vol. 26, no. 2, p. 54, 2015.

[6] M. Adli, "The crispr tool kit for genome editing and beyond," *Nature communications*, vol. 9, no. 1, p. 1911, 2018.

[7] Z. R. Chalmers, C. F. Connelly, D. Fabrizio, L. Gay, S. M. Ali, R. Ennis, A. Schrock, B. Campbell, A. Shlien, J. Chmielecki, *et al.*, "Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden," *Genome medicine*, vol. 9, no. 1, p. 34, 2017.

[8] E. Stanganello, A. I. Hagemann, B. Mattes, C. Sinner, D. Meyen, S. Weber, A. Schug, E. Raz, and S. Scholpp, "Filopodia-based wnt transport during vertebrate tissue patterning," *Nature communications*, vol. 6, p. 5846, 2015.

[9] B. Mattes, Y. Dang, G. Greicius, L. T. Kaufmann, B. Prunsche, J. Rosenbauer, J. Stegmaier, R. Mikut, S. Özbek, G. U. Nienhaus, *et al.*, "Wnt/pcp controls spreading of wnt/β-catenin signals by cytonemes in vertebrates," *Elife*, vol. 7, p. e36953, 2018.

[10] J. Ranft, M. Basan, J. Elgeti, J.-F. Joanny, J. Prost, and F. Jülicher, "Fluidization of tissues by cell division and apoptosis," *Proceedings of the National Academy of Sciences*, vol. 107, no. 49, pp. 20863–20868, 2010.

[11] T. Hirashima, E. G. Rens, and R. M. Merks, "Cellular potts modeling of complex multicellular behaviors in tissue morphogenesis," *Development, growth & differentiation*, vol. 59, no. 5, pp. 329–339, 2017.

[12] 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, pp. 70–78, Elsevier, 2015.

[13] A. Szabó and R. M. Merks, "Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution," *Frontiers in oncology*, vol. 3, p. 87, 2013.

[14] 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," *PloS one*, vol. 4, no. 10, p. e7190, 2009.

[15] B. M. Rubenstein and L. J. Kaufman, "The role of extracellular matrix in glioma invasion: a cellular potts model approach," *Biophysical journal*, vol. 95, no. 12, pp. 5661–5680, 2008.

[16] N. Podewitz, F. Jülicher, G. Gompper, and J. Elgeti, "Interface dynamics of competing tissues," *New Journal of Physics*, vol. 18, no. 8, p. 083020, 2016.

[17] T. Büscher, N. Ganai, G. Gompper, and J. Elgeti, "Tissue evolution: Mechanical interplay of adhesion, pressure, and heterogeneity," *New Journal of Physics*, vol. 22, no. 3, p. 033048, 2020.

[18] 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," *Proceedings of the National Academy of Sciences*, vol. 113, no. 48, pp. E7663–E7671, 2016.

[19] W. M. Iwasaki and H. Innan, "Simulation framework for generating intratumor heterogeneity patterns in a cancer cell population," *PloS one*, vol. 12, no. 9, p. e0184229, 2017.

[20] L. M. Loew and J. C. Schaff, "The virtual cell: a software environment for computational cell biology," *TRENDS in Biotechnology*, vol. 19, no. 10, pp. 401–406, 2001.

[21] M. H. Swat, G. L. Thomas, J. M. Belmonte, A. Shirinifard, D. Hmeljak, and J. A. Glazier, "Multi-scale modeling of tissues using compucell3d," in *Methods in cell biology*, vol. 110, pp. 325–366, Elsevier, 2012.

[22] P. M. Altrock, L. L. Liu, and F. Michor, "The mathematics of cancer: integrating quantitative models," *Nature Reviews Cancer*, vol. 15, no. 12, p. 730, 2015.

[23] P. Macklin, S. McDougall, A. R. Anderson, M. A. Chaplain, V. Cristini, and J. Lowengrub, "Multiscale modelling and nonlinear simulation of vascular tumour growth," *Journal of mathematical biology*, vol. 58, no. 4-5, pp. 765–798, 2009.

[24] B. Waclaw, I. Bozic, M. E. Pittman, R. H. Hruban, B. Vogelstein, and M. A. Nowak, "A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity," *Nature*, vol. 525, no. 7568, p. 261, 2015.

[25] A. Karolak, D. A. Markov, L. J. McCawley, and K. A. Rejniak, "Towards personalized computational oncology: from spatial models of tumour spheroids, to organoids, to tissues," *Journal of The Royal Society Interface*, vol. 15, no. 138, p. 20170703, 2018.

[26] 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," *PLoS computational biology*, vol. 14, no. 2, p. e1005991, 2018.

[27] M. Berghoff, J. Rosenbauer, F. Hoffmann, and A. Schug, "Cells in silico–introducing a high-performance framework for large-scale tissue modeling," *BMC bioinformatics*, vol. 21, no. 1, pp. 1–21, 2020.

[28] M. Berghoff, I. Kondov, and J. Hötzer, "Massively parallel stencil code solver with autonomous adaptive block distribution," *IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 10, pp. 2282–2296, 2018.

[29] M. Berghoff and I. Kondov, "Non-collective scalable global network based on local communications," in *2018 IEEE/ACM 9th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems (scalA)*, pp. 25–32, IEEE, 2018.

[30] F. Graner and J. A. Glazier, "Simulation of biological cell sorting using a two-dimensional extended potts model," *Physical Review Letters*, vol. 69, no. 13, p. 2013, 1992.

[31] N. Chen, J. A. Glazier, J. A. Izaguirre, and M. S. Alber, "A parallel implementation of the cellular potts model for simulation of cell-based morphogenesis," *Computer physics communications*, vol. 176, no. 11-12, pp. 670–681, 2007.

[32] A. Tomeu, A. Gámez, and A. Salguero, "A parallel implementation for cellular potts model with software transactional memory," in *International Conference on Practical Applications of Computational Biology & Bioinformatics*, pp. 53–60, Springer, 2019.

[33] L. Oswald, S. Grosser, D. M. Smith, and J. A. Käs, "Jamming transitions in cancer," *Journal of physics D: Applied physics*, vol. 50, no. 48, p. 483001, 2017.

[34] I. Dagogo-Jack and A. T. Shaw, "Tumour heterogeneity and resistance to cancer therapies," *Nature reviews Clinical oncology*, vol. 15, no. 2, p. 81, 2018.

[35] K. Bahrami and M. Kim, "Optimal control of multiplicative control systems arising from cancer therapy," *IEEE Transactions on Automatic Control*, vol. 20, no. 4, pp. 537–542, 1975.

[36] G. W. Swan, "Role of optimal control theory in cancer chemotherapy," *Mathematical biosciences*, vol. 101, no. 2, pp. 237–284, 1990.

[37] D. Iliopoulos, H. A. Hirsch, and K. Struhl, "Metformin decreases the dose of chemotherapy for prolonging tumor remission in mouse xenografts involving multiple cancer cell types," *Cancer research*, vol. 71, no. 9, pp. 3196–3201, 2011.

4 A computational two-type tumor model, bench-marking physical tumor properties

MECHANICAL *properties of tumor cells impact the behavior of developing tumors and influence their emergent properties. In this chapter, I introduce a computational model of tumor growth that consists of two competing cell types. The invasion of the tumor cell type into the surrounding tissue is modeled and the effects of tumor-associated mechanical cell properties are observed. Three properties of the expanding cell type are altered: cell motility, stiffness, and cell-cell adhesion. The effects of the three properties on the tumor growth rate, tumor shape, and invasiveness are observed. Low adhesion leads to an elevated growth rate, while lower growth rates increase the mixing of the cell types.*

Growing tumor tissue expands into and against the surrounding tissue. For this expansion, the tumor competes with the surrounding healthy tissue over the available resources such as space and nutrients. The interactions of a tumor with its microenvironment largely influence tumor development [1]. Tissue vascularization, interactions with the immune system, Extra-cellular matrix (ECM) and external signaling strongly influence the development of a tumor and add largely to the complexity and variability of tumor treatment. Beyond the interactions with the tumor environment, the properties of the individual cells that make up the tumor play a large role in the tumor development and define its growth. Through their properties, the individual cells define the emergent behavior of the macroscopic tumor.
Mechanical properties of the expanding tissue increasingly gain attention for their effect on tumor growth. Cell-cell adhesion, mechanical confinement, and cell density can drive phase transitions from jammed and active to gas-like states [2]. Motility forces can influence the geometry of interfaces [3] and cell-cell adhesion and growth force influence tumor heterogeneity and evolution [4]. The cellular properties of cells in malignant tumors were characterized by several mechanical properties. *Loss of adhesion*, *high motility* and *low cellular stiffness* have been attributed to tumor cells [5–7].

Here, we focus on the mechanical effects of each of those three single-cell parameters and how they influence the emergent behavior on the tumor level.

The properties may be linked to other cellular parameters through signaling pathways (e.g., β catenin regulates both cell division and cell-cell adhesion by e-cadherin [8]) and those parameters might be also influencing tumor growth. Here we independently vary the three parameters, while keeping all other parameters constant, to be able to isolate their respective effects.

We use *Cells in Silico* (CiS), a recently published framework for simulating tissue with the Cellular Potts model (CPM) on supercomputers [9]. We model the competition of two cell types in a simulation box. We vary three parameters, namely adhesion, motility, and stiffness to see the effect of each mechanism on tumor growth. For each parameter, 5 different values are chosen which results in a phase space of $5 \times 5 \times 5 = 125$ simulations.

We observe the emergent behavior of each tumor. The size development is observed and different models for the growth curves are fitted. We compare the time constants of fitting Gompertz, logistic, and exponential growth. Furthermore, we exploit the access to all cellular parameters and

observe the velocity correlation of tumor cells, the shape and geometry of the tumor, and quantify the invasiveness.

Hypothesis: The three single-cell properties, adhesion, stiffness, and motility are major driving forces of tumor growth. All three are important to drive tumor expansion and invasion. Here we explore the isolated and combined effects of each parameter on tumor growth.

4.1 Methods

In this manuscript, we focus on a computational model of tumor growth. Two cell types compete over space, with one having an increased proliferative potential. This tumor cell type is placed in a solid tissue consisting of the healthy cell type. In the simulations, the tumor tissue expands and takes over the available space. The *in silico* modeling of this two-type system gives us the possibility to screen and scan the impact of different cell-cell and cell-tumor interactions. We incorporate three properties into the model that are often associated with tumor cells by changing the adhesive properties, cell stiffness, and cell motility. These parameters are varied in the tumor cells and the effects on the emergent tumor are observed. The simulation is performed in a simulation box of $200times200 \times 200$ voxels with each cell occupying ≈500 voxels. The box is filled with cells, leading to ≈24 000 interacting cells in one simulation.

Adhesion: Loss of adhesion or downregulation of cell-cell adhesion is often linked to tumor spreading and expansion [5]. Cell-adhesions are mediated by adhesion molecules on the cell membrane (e.g. cadherins). Those molecules can bind to each other and thereby mechanically link cells to other cells and the surroundings. In the computational model, adhesion is accounted for by the binding energy between cell surfaces. The coupling of different cells of the same cell type j is determined by a factor J_{jj}, that couples the shared surface to energy. The factor between different cell types i and j, is determined by the smaller internal coupling $J_{ij} = \min(J_{jj}, J_{ii})$, since we assume a limited number of adhesion molecules on the surface, that limit the adhesion strength. The adhesion of the surrounding healthy tissue $J_{hh} = 25$, $h-$ healthy cells, while the tumor adhesion values are changed from $J_{tt} = 0 - 80$, $t-$ tumor cells.

Stiffness: Malignancy of tumors is often associated with soft tumor cells [6]. The stiffness references in this case the deformability of the individual tumor cells. While the tumor mass typically is harder than the surrounding tissue, this is due to the stiffening of the ECM that is often initiated by tumors [10]. In our model, the varied parameter to determine the influence of stiffness on tumor growth is the prefactor of the surface energy λ_S. High values of λ_S lead to low surface deformability and therefore to stiffer cells. The parameter is varied from $\lambda_{S,\text{tumor}} = 1..10.25$ for tumor cells, while in the surrounding tissue cells $\lambda_{S,\text{tissue}} = 5.325$

Motility: Motility describes the ability of cells to actively move in space. Motility can be directed (e.g., chemotaxis) and nondirected (e.g., random walk). Here, we focus on a random movement of the cells. Random motility is implemented as the coupling of each cell to a potential along a

randomly chosen 3D unit vector. The direction of the vector is randomly reassigned each 100 MCS. The coupling of each cell to that potential is called λ_M and will be referred to as the amount of motility for a cell.

Cell Division and Death: Division of a cell is possible once the cell exceeds 0.9 of its goal volume V_0 and exceeds an age of 2000 MCS. If those conditions are met, the cell divides with a probability of 0.03 for tumor cells and 0.005 for surrounding cells in each time step. Cells die similarly with a probability of 0.001 after the age of 8000 MCS for tumor cells and a probability of 0.0005 after 4000 MCS for surrounding cells. Since the simulation box is filled by cells, the number of cells is limited by the space and the volume constraint of cell division.

Growth Models: Three different growth models with a time constant τ are evaluated in the analysis of the growth curves of the *in-silico* tumors. The functions are:

Logistic growth

$$\phi(t) = \frac{a}{1 + e^{-\tau \cdot (t-c)}} \tag{13}$$

Gompertzian growth

$$\phi(t) = a \cdot e^{-b \cdot e^{-\tau \cdot t}} \tag{14}$$

Exponential growth

$$\phi(t) = a \cdot e^{-\tau \cdot (t-c)} \tag{15}$$

Table 1: Cell parameters used in the simulations.

Parameter	**Symbol**	**Value / Range**
Tumor		
Volume compressibility	λ_V	7.5
Surface compressibility (=stiffness)	λ_S	1.0, 3.3125, 5.625, 7.9375, 10.25
Motility	λ_M	0, 25, 50, 75, 100
Cell-cell adhesion	J	0, 20, 40, 60, 80
Division condition		(volume $\geq 0.9 \cdot V_0$) & (age > 2000) & (rnd() ≤ 0.03)
Death condition		(age > 4000) & (rnd() ≤ 0.001)
Surrounding		
Volume compressibility	λ_V	7.5
Surface compressibility (=stiffness)	λ_S	5.625
Motility	λ_M	0
Cell-cell adhesion	J	25
Division condition		(volume $\geq 0.9 \cdot V_0$) & (age > 4000) & (rnd() ≤ 0.005)
Death condition		(age > 8000) & (rnd() ≤ 0.0005)
General		
Temperature		50
Simulation box		200 voxels^3
Cell volume	V_0	500
Cell surface (marching cubes)	S_0	400

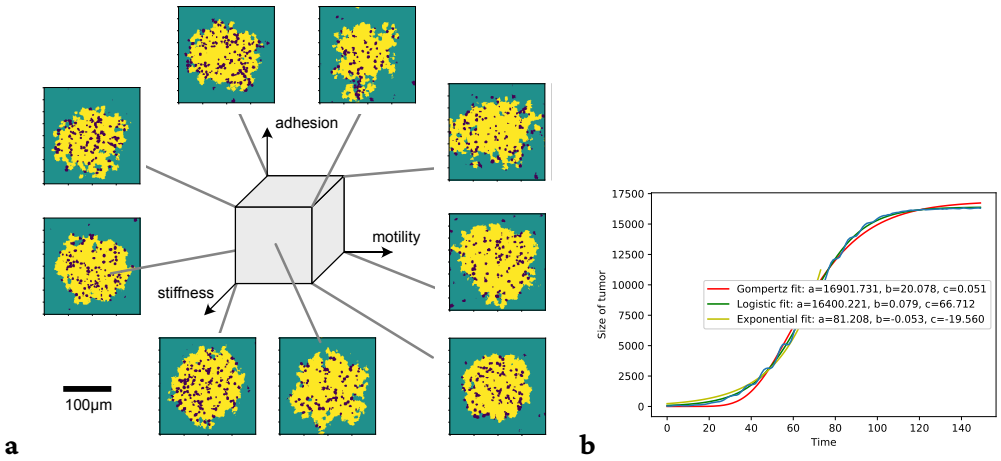

Figure 9: **Overview over the observed phase space, tumor topologies, and temporal trajectory. a** 2D slices through the simulations at the minimum and maximum of all three varied parameters. Colors indicate the different cell types, green is the surrounding cells, yellow is tumor cells and purple is dying cells. The images are taken as soon as the tumor exceeds 2000 cells. **b** The growth curve of the tumor with all minimal values (0,0,0). Different growth models are fitted to the curve and the parameters are indicated.

4.2 Results

The simulations are initialized with a cluster of 32 cells of tumor type and 14 000 cells of surrounding type. Subsequently, the tumor spreads by the elevated division rate of the tumor cells, compared to the surrounding cells. The tumor grows at first with an exponential increase in cell numbers. The fast growth slows down as soon as the size of the tumor approaches the size of the simulation box and continues to decrease until the surrounding tissue is completely displaced by tumor cells, cf Fig. 9 b. The growth curve resembles logistic or Gompertzian growth. 125 parameter combinations were simulated, five values each of the three parameters adhesion, motility, and stiffness were implemented. Therefore, the simulations span a $5 \times 5 \times 5$ cube in the parameter space, representing all parameter value combinations. Examining 2D slices of the 3D simulation for equal numbers of tumor cells, but different parameter combinations show different topologies of the tumor shape (cf. Fig. 9 a). The tumor shapes range from cohesive outlines to a rugged appearance with a further spread.

To quantitatively describe the above-mentioned properties, several tumor properties are fitted to the data and projected on the $5 \times 5 \times 5$ matrix in the parameter space.

Growth speed:

First, different tumor growth models are fitted to the growth curve of the tumor. Gompertzian, logistic and exponential growth are fitted (cf. Fig. 9 b for a single simulation). The time constant of logistic growth is shown in Fig. 10 a, projected on the parameter space. A higher time constant indicates faster growth of the tumor. We find:

With decreasing adhesion, the tumor grows systematically faster. A low adhesion allows for a faster dissociation of the tumor cells from each other, leading to lower friction inside the tumor and therefore a faster propagation of low pressure following cell death.

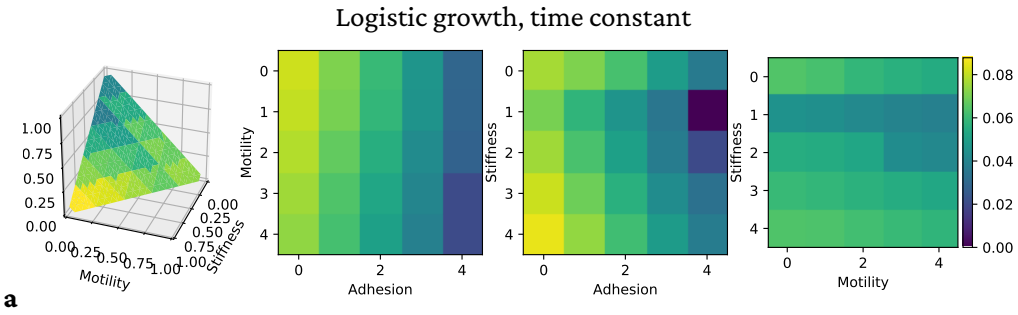

Figure 10: **Logistic growth constants projected on the parameter space.** Parameter projections on the 125 simulations on a 5x5x5 cube. Left image shows a cut through the 4,4,4 plane. The other images depict the mean value of the cube along each axis. **a** Time constant of logistic growth.

Motility only has a small influence on the growth rate, with slightly faster growth for low motility values. We see a counter-intuitive behavior since cell motility is often associated with faster tumor spreading.

Stiffness changes lead to a lowering and a subsequent rise in growth speed. The influence of cell stiffness is less expressed than the influence of cell adhesion, but it shows a nonlinear behavior. As the cell stiffness increases from 1 to 10, the growth speed first decreases and then increases for values larger than 3.

Roughness:

In Fig. 11 the fraction of the 3D surface and volume is plotted for all simulations. The fraction (here called roughness) increases with rising motility and rising adhesion and shows an inverse dependency on growth speed (cf. Fig. 10 a) The dependency on stiffness shows the same but inverse behavior as observed for the growth speed, namely an initial increase, followed by a decrease.

To quantify the observed dependency between roughness and growth speed, the properties are plotted against each other in Fig. 11 c. Non-linear dependency can be observed. The roughness drops steeply for low values of the growth speed. Once the growth speed exceeds 0.028, the steepness decreases drastically and the roughness decrease is slower.

Extent:

The spatial extent of the tumor is measured for each simulation once the number of tumor cells exceeds 2000. The extent is determined along the three main axes and the mean of the three values is shown. Qualitatively, a similar behavior to the tumor surface roughness can be observed (cf Fig. 11 d). When comparing the extent with the tumor roughness in Fig. 11 d, a positive nonlinear dependency can be observed.

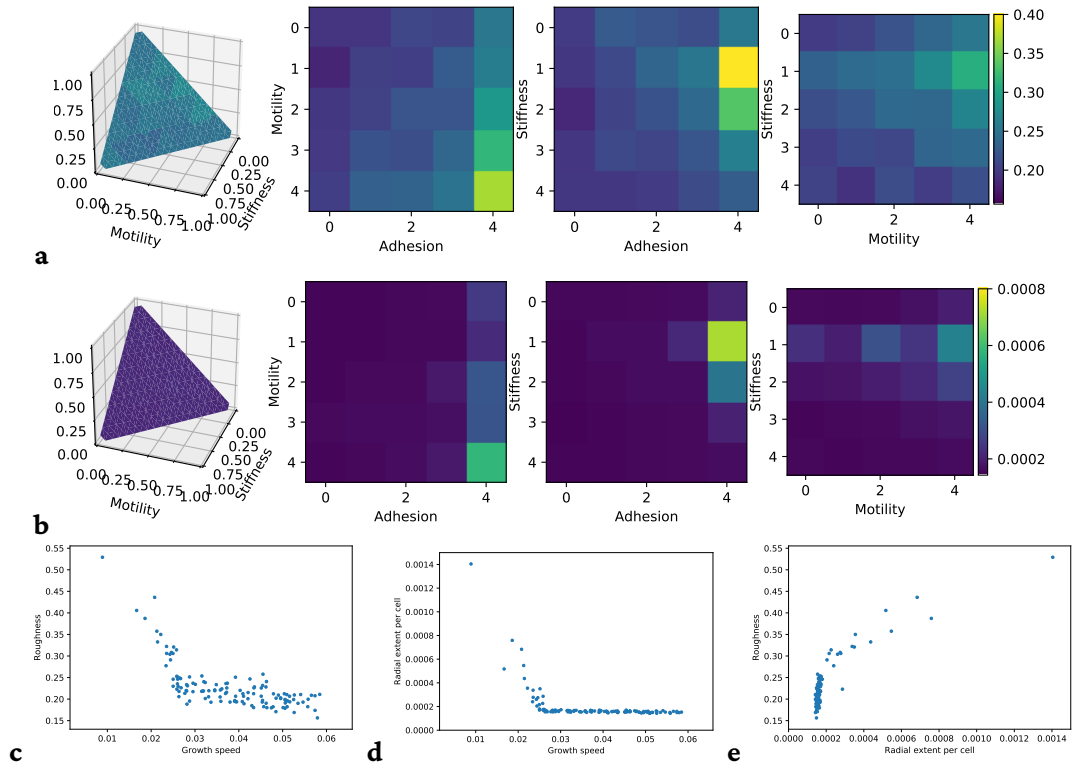

Figure 11: **Tumor roughness and the dependence on tumor growth rate.** **a** Ratio of tumor surface to tumor volume, projected on a cube. The roughness is measured as soon as the tumor exceeds 2000 cells. **b** Mean spatial tumor extent, projected on a cube. The extent is measured as soon as the tumor exceeds 2000 cells. **c** The roughness is plotted against the time constant of fitted logistic growth (plotted in 10). **d** The mean spatial extent of the tumor is plotted against the time constant of fitted logistic growth **e** The roughness is plotted against the mean spatial extent of the tumor.

4.3 Discussion

Tumor roughness and shape are strongly influenced by growth speed. High adhesion, high motility, and very soft or very stiff cells produce a fast-growing tumor that is spherical and smooth. While low adhesion, low motility, medium stiffness cells lead to a slower-growing tumor with a rough surface.

We define tumor invasiveness by the penetration of tumor cells into the surrounding tissue, since a larger roughness of the tumor leads to an increased mixing of the tumor and surrounding cells. While a smooth tumor surface represents a low invasion, high surface roughness and a nonspherical irregular shape characterize a more invasive growth. Both measured parameters, spatial extent, and surface roughness contribute to the invasiveness of a tumor. We find higher invasiveness in tumors that are mechanically driven to a lower growth speed. Both invasiveness and growth speed provide advantages for the tumor. We find a trade-off between growth speed and invasiveness that results from the mechanical properties and is independent of external signaling and division rates (nonlinear, as seen in Fig. 11 c and d). We find a negative correlation between growth speed and adhesion strength and a positive correlation between adhesion and invasiveness. We find a nonlinear behavior of the tumor properties depending on the cell stiffness. Motility and invasiveness are positively correlated, while the growth speed is only minorly influenced by motility.

The mechanical properties of single cells define the emergent behavior of growing tumors and have a large impact on growth speed and invasiveness, independent of their division rates.

References

[1] A. Nagelkerke, J. Bussink, A. E. Rowan, and P. N. Span, "The mechanical microenvironment in cancer: How physics affects tumours," in *Seminars in cancer biology*, vol. 35, pp. 62–70, Elsevier, 2015.

[2] O. Ilina, P. G. Gritsenko, S. Syga, J. Lippoldt, C. A. La Porta, O. Chepizhko, S. Grosser, M. Vullings, G.-J. Bakker, J. Starruß, *et al.*, "Cell–cell adhesion and 3d matrix confinement determine jamming transitions in breast cancer invasion," *Nature cell biology*, vol. 22, no. 9, pp. 1103–1115, 2020.

[3] T. Büscher, A. L. Diez, G. Gompper, and J. Elgeti, "Instability and fingering of interfaces in growing tissue," *arXiv preprint arXiv:2003.04601*, 2020.

[4] T. Büscher, N. Ganai, G. Gompper, and J. Elgeti, "Tissue evolution: Mechanical interplay of adhesion, pressure, and heterogeneity," *New Journal of Physics*, vol. 22, no. 3, p. 033048, 2020.

[5] U. Cavallaro and G. Christofori, "Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough," *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, vol. 1552, no. 1, pp. 39–45, 2001.

[6] M. Yilmaz and G. Christofori, "Mechanisms of motility in metastasizing cells," *Molecular cancer research*, vol. 8, no. 5, pp. 629–642, 2010.

[7] W. Xu, R. Mezencev, B. Kim, L. Wang, J. McDonald, and T. Sulchek, "Cell stiffness is a biomarker of the metastatic potential of ovarian cancer cells," *PloS one*, vol. 7, no. 10, p. e46609, 2012.

[8] A. Stockinger, A. Eger, J. Wolf, H. Beug, and R. Foisner, "E-cadherin regulates cell growth by modulating proliferation-dependent β-catenin transcriptional activity," *The Journal of cell biology*, vol. 154, no. 6, pp. 1185–1196, 2001.

[9] M. Berghoff, J. Rosenbauer, F. Hoffmann, and A. Schug, "Cells in silico–introducing a high-performance framework for large-scale tissue modeling," *BMC bioinformatics*, vol. 21, no. 1, pp. 1–21, 2020.

[10] H. T. Nia, L. L. Munn, and R. K. Jain, "Physical traits of cancer," *Science*, vol. 370, no. 6516, 2020.

5 Mechanical influences on *in silico* tumor evolution

CELLS *in a growing tumor compete over the available resources, which are space and nutrients. Through mutations, cells change their behavior and properties during tumor growth. Competition over space and nutrients, together with changes in cell properties leads to an evolutionary behavior of the tumor composition towards the fittest cell types. What cell properties are selected for in this survival of the fittest? A 3D spheroid tumor model is developed and its evolutionary behavior is observed. Mechanical properties alone are able to drive the tumor towards low adhesion which is associated with tumor invasion. A dynamically changing nutrient surrounding representing the fluctuating blood supply through blood vessel collapse and angiogenesis is introduced. I highlight a dependency of the evolutionary speed on the frequency of the fluctuations and identify a frequency domain in which the evolutionary speed is significantly increased over a tumor with a constant nutrient supply. The findings suggest that the mechanically induced fluctuations can accelerate tumor evolution.*

Note: This chapter is prepared to be submitted to a peer-reviewed journal, with the authors: Jakob Rosenbauer, Marco Berghoff, James A. Glazier, and Alexander Schug. I am the first author and designed the model, performed the simulations and analysis, wrote the manuscript, and assembled the figures.

The emergence and development of tumors in humans still presents a significant challenge to medicine and is one of the leading causes of death. Knowledge about the development of tumors is continuously expanding. Especially, the large variety of tumor types, differences between patients and tumors consisting of many types of tumor cells (termed tumor heterogeneity) present challenges. Here, a tumor in surrounding solid tissue and the development of tumor properties is observed over time in a computational model. The development of a tumor can be described as an evolutionary system in which cell types present species competing over resources [1, 2]. The evolution of a tumor is driven by mutations. These mutations change the properties (e.g., motility or cell-cell adhesion) of cells, which then can provide advantages or disadvantages for further growth. Competition over resources and space leads to the selection of the most advantageous tumor cell types, whose subpopulation will overtake the rest. However, cells can not optimize all parameters at the same time, several trade-offs in cellular properties in cancer have been described and studied [3, 4]. We assume a limited energy budget for each cell and introduce a trade-off of proliferation against motility for cell types.

While the properties and parameters of a cell are accessible in experiments, the fitness of a cell is very hard to determine experimentally. The fitness determines the ability of a cell to reproduce in its surroundings and multiply, therefore cells with high fitness are selected for in evolution. To determine the fitness of cells experimentally, the lineage of cells has to be tracked over time, making it necessary to track each individual cell over a long time, which is currently not possible in *in vivo* tumors. Models of the fitness landscape based on genetic changes and driver mutations have been introduced [5], yet phenotypic changes in cells can also drive mutation [6, 7]. Naturally,

an ensemble of cells with different cell types evolves towards the cell type of highest fitness. The fitness is not an intrinsic property of the cell but results from the interplay of cells with their surroundings. Finding the fitness of cell types depending on their parameters and surroundings allows for directed interventions to change the course of evolution. Taking control of the evolution of a tumor could serve as a tool to stop the spreading and further growth of a tumor and aid treatment.

The mechanical interactions of the tumor with its environment and between the tumor cells influence the trajectory of the tumor [8]. Cells with low cell–cell adhesion mechanically sort to the outside of tumors, providing a higher nutrient supply and therefore evolutionary advantages. The nutrition levels on the surface of tumors are higher than in the center of the tumor, leading to an induced advantage of low adhesion cells. Therefore, we will test this prediction and observe whether those cells have an evolutionary advantage in an evolving heterogeneous tumor. Cell motility describes the active movement of cells, we predict that random cellular movements should provide no significant evolutionary advantage in a constant surrounding. Phenotypical changes towards higher motility cells and the epithelial to mesenchymal transition are associated with tumor invasion [9–11]. In a dynamically changing nutrient environment, cells with higher motility can dynamically occupy the most advantageous positions. Enhanced motility is known to be a driver of the formation of metastases [12], therefore a dynamic nutrient field could induce metastases.

In this work, we focus on two variables in cellular properties, namely *cell–cell adhesion* and *motility*. Increased motility is a landmark for the development of metastases [12]. We hypothesize that mechanical and geometric constraints alone are sufficient to drive the evolution of a tumor towards high motility.

The computational modeling of tumor development and heterogeneity has evolved and was successfully applied, e.g., a cellular automaton model was used to observe the influence of cell dispersal and turnover on tumor heterogeneity [13]. The use of computational models in clinical applications and personalized treatment plans are gaining momentum [14, 15]. Work on tumor internal evolution has been recently done by Büscher et al. [16] where they compared adhesion and 'growth strength' in an evolving tumor and found that in some cases a mixture of different cell types emerges as a stable state. The importance of cell–cell adhesion for tumor invasion has been recently highlighted by Ilina et al. [17]. In Ref. [18] the authors showed with a mathematical model using the evolutionary game theory that tumor heterogeneity and the optimal cell phenotype depends on the microenvironment and position within the tumor. The 'go-or-grow' hypothesis describes reversible phenotype changes between motile and proliferative phenotypes in cancer. This hypothesis was proposed and tested with positive [19] and negative correlations [20] for different cancer cell lines. Trade-offs between two or more cell properties have been studied by Gallaher et al. [4]. Here, we are not focusing explicitly on the 'go-or-grow' hypothesis, since we assume each cell to retain its phenotype throughout its lifetime and changes of cell properties can only occur at mutation events during cell division.

Cell adhesion and cellular motility have been recently shown to strongly influence the invasion behavior of breast cancer and to differentiate between solid-like, fluid-like, and gas-like behavior [17]. Confinement by the extracellular matrix and cellular motility were recently investigated

systematically and a phase space of tumor invasion modes was proposed [21].

The intricate surrounding around a developing tumor is important and strongly influences its progression together with the intrinsic properties of the tumor cells. An invading tumor is expanding in volume and requires nutrients that can only diffuse through a finite length of tissue. Angiogenesis, the growth of new blood vessels, is initiated for a better nutrient supply. This leads to a quick volume increase of the tumor, which builds up pressure and solid stresses within the tumor that can collapse blood vessels [22]. This interplay of formation and collapse of blood vessels can lead to a fluctuating availability of nutrients for the cells in the tumor.

During the development of tumors, single-cell effects are of major importance, with mutations initially occurring in a single cell. Stochasticity and rare events are non-negligible during tissue development. As we showed in [23], single-cell effects are important for the patterning of tissue. Cell migration is facilitated by many concurrent processes and represents a multiscale process, therefore requiring multiscale modeling [24]. To capture the complexity of tumor growth, computational models with single-cell resolution enable the incorporation of single-cell behavior that would be difficult to capture using continuous ODE models. Since evolution is a stochastic process requiring many iterations to find a stochastically meaningful result and the requirement of large-scale tissues makes the use of high-performance computing necessary.

The recently developed framework CellsInSilico [25] is used for the simulations, it is based on the cellular Potts model (cpm) [26] that has been established for the simulation of tumor growth [27]. Through its parallelization and optimization for supercomputers, the framework enables large-scale three-dimensional simulations and high numbers of simulations, which is necessary to generate sufficient statistics to study the proposed problems.
We implement a discrete evolution of two independent parameters, namely cell-cell adhesion and cell motility. Cells can alter and change their parameters stepwise, with a small probability at each division. Parameters are changed incrementally, which implies a continuous evolution and neglects possible mutations with larger effects. For cell motility, we introduce a trade-off on division rates since we assume cells have a limited energy contingent. We simulate a spheroid tumor in an external gradient of nutrients. The tumor is surrounded by a population of non-mutating host tissue. We observe the evolution of the tumor composition along both free parameters. Different mechanisms that couple cell division rates to nutrient availability are compared. We compare the effect of temporally changing nutrient availability.

Table 2: Parameters varied during simulations

Parameter	Range	Dependency
Cell–cell adhesion	[0...2T] (no repulsion)	Independent
Motility	[0...2T]	Trade-off motility vs. division rate
Division rate	$R = -k(\text{Motility} - 130)$ $k \in$ [1e-4...1e-6]	
Metropolis Temperature T	55 (Constant)	
Motility recalculation Time t_{motility}	100 MCS (50,200,400)	
System size	50,100,400 $\mu m^3 \approx$ voxels^3	
Coupling to central potential	-70	

5.1 Methods

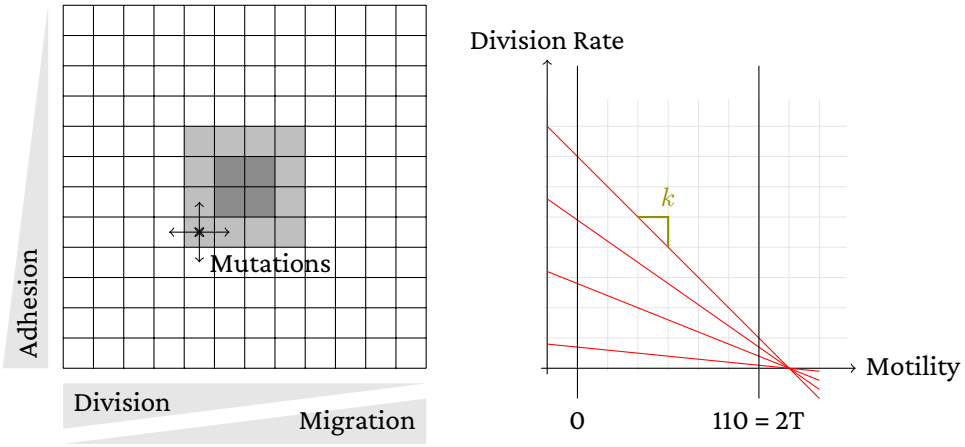

Figure 12: **Cell type parametrization.** Cell types span a 12×12-matrix. Linear trade-off model between division rate and motility.

Energy functions are:

- Volume
- Surface
- Adhesion
- Random motility
- Central potential

Adhesion

The cell–cell adhesion is proportional to the contact area between cells and independent of the duration of the adhesion. The strength is not limited or quantized by focal adhesion but only determined by the adhesion parameter between the cell types and the shared area.

Nutrient
The nutrient availability of a cell is determined by its location in 3D space. The position of a cell is defined as the center of mass of its spatial extent. The function is a radially linear decay within a sphere, in the center of the simulation box. The center of the nutrient well can be temporally constant or moving, to represent constant or dynamic tumor environments. The nutrient represents a growth-limiting factor for the cells.
Central Potential
To avoid all cells accumulating in the outer regions with constant high nutrient availability, a potential is introduced. This potential leads to all tumor cells experiencing a constant force towards the center point of the simulation. This point is also the center of the nutrient well, with the lowest availability. The potential leads to an increase in pressure at the center of the tumor.
Random motility
Motility is implemented by assigning a preferential direction of movement to each cell. This direction is defined by a potential along a vector. The three-dimensional direction of this vector is randomly reassigned in a regular interval of 100 Monte Carlo sweeps. The cells are coupled to this potential by a constant force that is determined by the coupling of the energy term to the potential. This coupling constant varies for different cell types and is referred to as the motility strength in this manuscript.
Cell division and death
To divide, cells need to exceed the age of 2 kMCS and their volumes have to exceed 90% of their goal volume. Similarly, cells can divide once they exceed the age of 4 kMCS.

There are three different cases for dependency of cell division and death on nutrient availability:

1. No dependence: Constant division probability of division () and death ()
2. Thresholds:
    - Division: Constant rate (≈ 0.005) if nutrient exceeds 15, otherwise no divisions
    - Death: If nutrient is below 25 higher rate (0.01) otherwise lower rate (0.00001)
3. Linear dependence:
    - Division: Rate linearly increases from 0 to ≈ 0.005 with increasing nutrient availability
    - Death: Rate linearly decreases from 0.001 to 0 with increasing nutrient availability

Division rates vary from cell-type to cell-type since they are determined by the division rate - motility trade-off.

The tissue that surrounds the tumor does not participate in cell death or division. The cells that make up the surroundings, therefore participate in the entire simulation and act as a medium that the tumor cells at the tumor edge can interact with and redistribute forces and pressure.

Evolution speed and spread: The speed of evolution is calculated by tracking the center of mass of the distribution of cell types in the phenotype space. The spread is measured by the extent of the distribution.

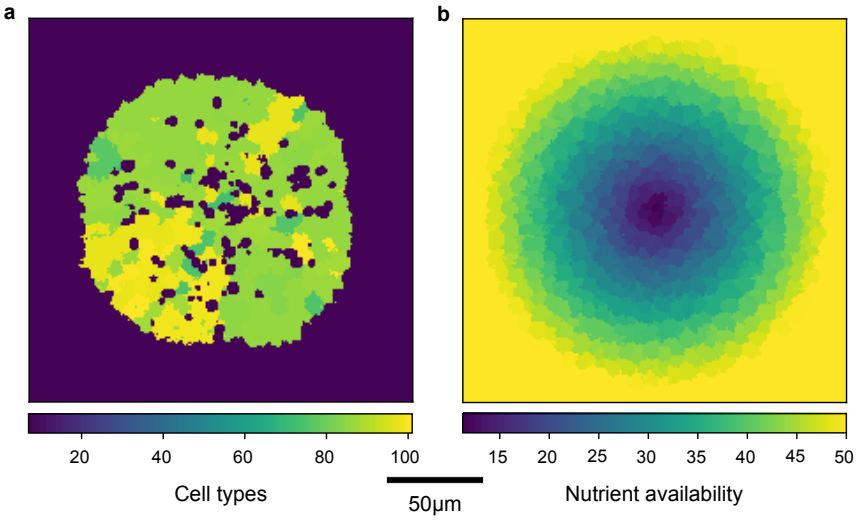

Figure 13: **Geometry of tumor simulations. a** Spatial organization of the cell types, 2D slice through the center of the 3D simulation. The color map indicates the different cell types, that are linearly numbered. **b** coloring by nutrient availability. Nutrient availability in arbitrary units (AI).

5.2 Results

5.2.1 Constant Environment

First, we investigate the dynamics of the system without the dependency of cell division and death on nutrient availability. Division and death are determined by the respective rates and age and volume constraints. A spheroid tumor develops from an initial tumor seed, consisting of ≈ 3700 tumor cells (cf. Fig. 13) in a surrounding of nondividing and nondying cells. The tumor grows until the space provided by the 3D simulation box is used up by the surrounding tissue and the tumor. Cell division is limited to cells above a threshold volume ($V_{\text{THRS}} = 0.9 \cdot V_0$) and cells are only compressible to a finite extent, therefore the absolute number of tumor cells is limited and the cells compete over the available space. In the emerging steady state, the tumor size remains constant with cells dying and dividing at equal rates. The cell division rate is higher than the death rate (cf. Fig. 14 a-c) and the absolute number of cells is geometrically constrained, thereby the effective division rate is limited by the death rate.

Observing the statistical occurrence of cell events in relation to the radial distance to the tumor center, we find that cell deaths are located in the tumor center while cell divisions are located at the tumor margin (see Fig. 14 a). This leads to an inward movement of cells (as seen by the stripes in Fig. 13). There is a buildup of pressure inside the tumor, this is facilitated by the inwards movement of the cells and the central potential. This behavior is well studied in experiments and computational models [28].

The simulation is started with a single cell type in the center of the phenotype space. Through mutations during tumor growth, more cell types are introduced which leads to a distribution in the phenotype space around the initial cell type. With the progressing simulation, the distribution moves in the phase space towards an evolutionary favorable position. The center of mass, spread,

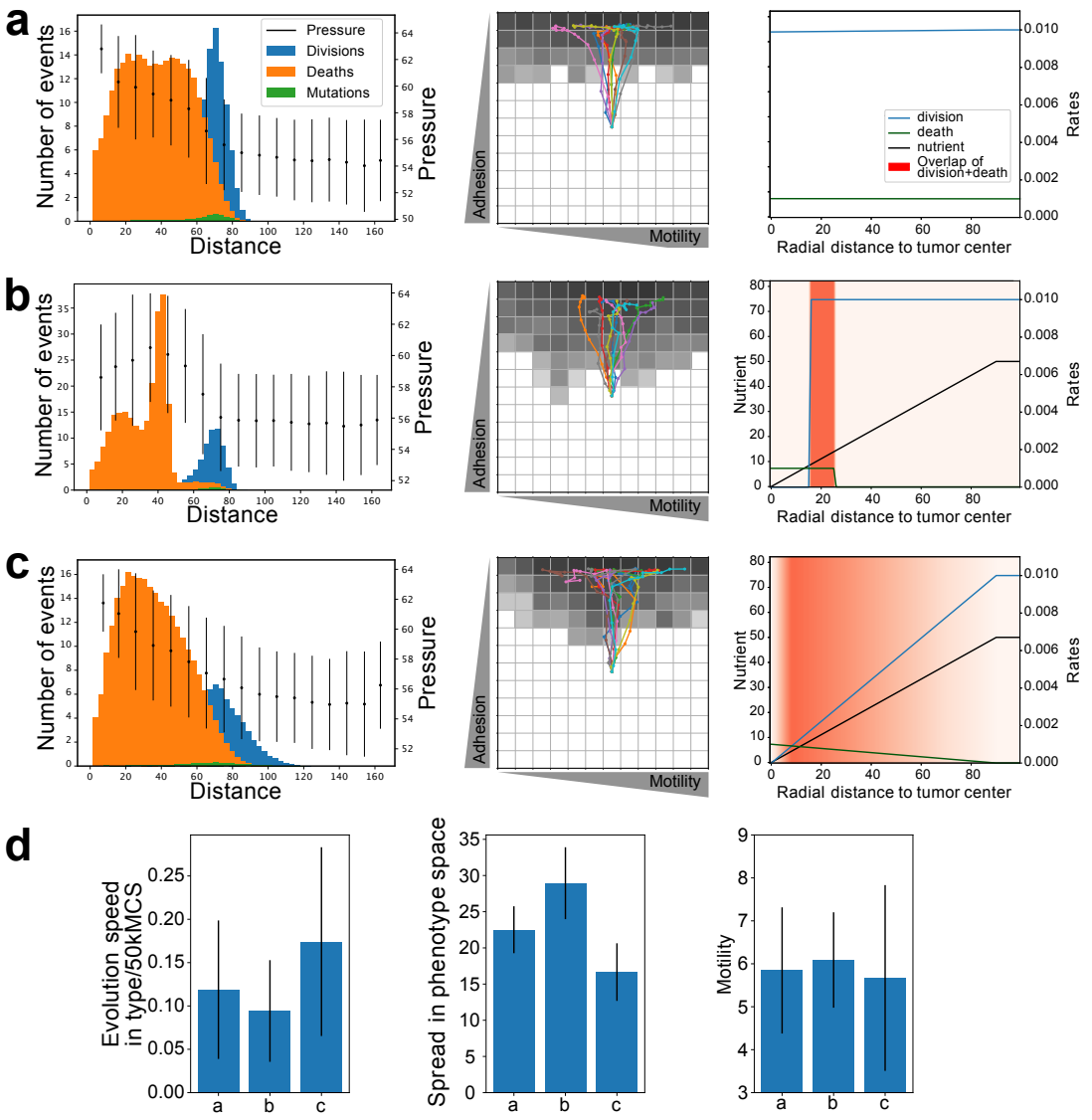

Figure 14: **Spheroid tumor growth for different nutrient dependency mechanisms.** In a-c, on the left, the radially averaged location of the events with respect to the center of the spheroid on the left. The number of events is normalized with $1/r^2$, the non-normalized plots can be found in Appendix 4. Simulation size is $200\,\text{voxel}^3$. The center plots show the trajectories of the centroid of the phase space occupation. The points on trajectories indicate temporally equidistant points. The shading shows the average distribution in the phenotype space at the endpoint at $t = 1500\,\text{kMCS}$ of 15 simulations. The right plots depict the spatial dependency of the division and death rates for a cell and the dependence on nutrient availability, depending on the distance from the tumor center. The red shading indicates areas of overlap of dividing and dying cells, here evolutionary pressure is most effective. **a** without the dependency of cell division and death on nutrient surrounding, **b** division and death regulated by thresholds on the constant nutrient surroundings and **c** linear dependence of the rates on the nutrient. **d** Macroscopic tumor properties for the respective nutrient availability cases of **a**, **b** and **c**. The black lines describe the standard deviation and the bars the mean values of 15 simulations.

and movement speed of this distribution are analyzed and compared. With the model introduced so far, we find that the tumor evolves towards the low adhesion regime. The individual trajectories start with a high directionality towards low adhesion, the trajectories then develop without directional preference along the motility axis (see Fig. 14 a-c).

Nutrient Dependency

We investigate how nutrient dependency of cell division and death influences the tumor properties in our model. By introducing a dependency of cell division and death on nutrient availability, cells compete over space and nutrients. Nutrients are introduced as a growth-limiting factor (see Methods), representing, e.g., oxygen or glucose. In an *in vivo* tumor, nutrients diffuse into the tissue and are internally degraded and used up. Here, we simplify his dynamic nutrient surrounding and it is implemented by introducing nutrient availability solely dependent on the position of the cell. The nutrient availability of cells is linearly increasing from the tumor center up to a maximal value and stays constant for further distances, as pictured in Fig. 13 b. We introduce two different dependencies of the cell division rate and death rate on the nutrient availability, as seen in Fig. 14 a-c, right (see also Methods). Firstly, a threshold-based dependency (TBD) that introduces a constant division probability if a cell exceeds a certain nutrient value, and a step decrease in death rate above another threshold. Secondly, we introduce a linear rate dependency (LRD). Here, cells linearly adapt the division and death rates depending on the local nutrient concentration.

Macroscopically, the introduction of a nutrient dependency affects the evolution speed of the tumor (see Fig. 14 d). TBD decreases the evolutionary speed of the system while LRD accelerates the evolution. Enhanced directionality is visible in the different sizes of the spread in the phenotype space. A smaller spread is achieved through a more pronounced directionality of evolution. TBD increases the spread while LRD lowers it. This increase in the evolutionary speed of LRD over TBD can be explained by the larger regions, in which cell divisions and cell deaths occur. In these critical regions (tagged red in Fig. 14 a-c, right), the competition between different cell types is most pronounced since cells that can stay in this area or escape outside will survive, while cells that are pushed to the inside will die. Despite the different evolutionary speeds and spreads, the evolution is highly directional towards low adhesion for both mechanisms and no nutrient dependency. The composition center of mass in the phenotype space is only insignificantly changed by the introduction of nutrient dependencies of cell death and division on the migration axis. Interestingly, the introduction of nutrient dependency in a constant environment does not introduce significant evolutionary queues towards higher motility or higher division rate for the parameters chosen here. The nutrient dependency of cell death and division leads to a shift of cell deaths towards the inside of the tumor. TBD introduces a drop in pressure at the center of the tumor (see Fig. 14 b).

The mechanism of nutrient dependency with LRD shows a larger selective pressure on the tumor and therefore leads to a higher speed of evolution. Furthermore, a linear dependency is biologically more reasonable, since cells do not binarily up- or down-regulate cell division in most cases, but adapt continuously [29]. Hence, LRD is used in the subsequent manuscript if not explicitly stated otherwise.

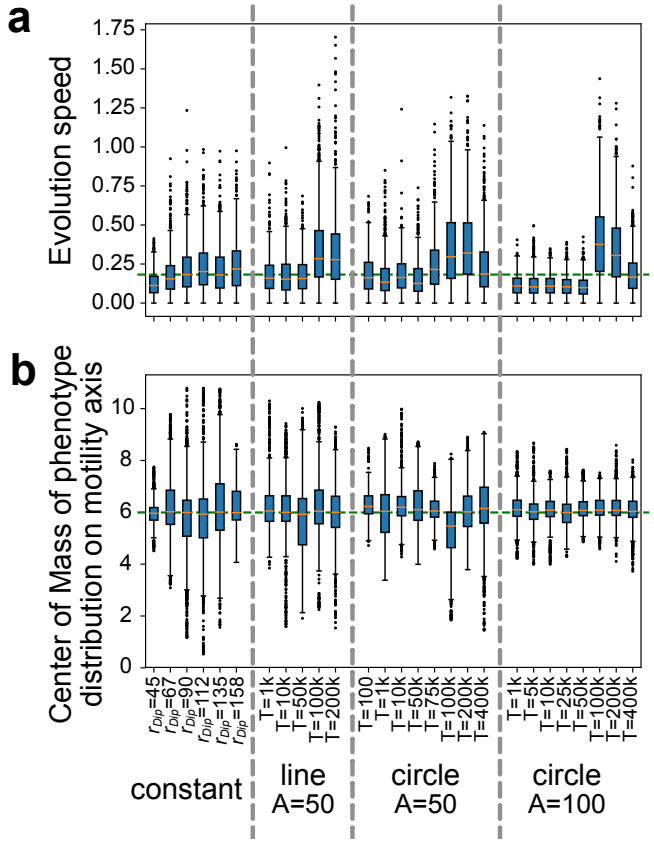

Figure 15: **Macroscopic tumor properties for linear dependency of division rates to nutrients.** **a** (top) Evolution speed in conformation space. Values are averaged between $1500 - 3000$ kMCS. In the first group (left), the simulations are performed with constant nutrient availability and a varying radius of the dip in nutrient availability (r_{dip}). In the following groups, the nutrient availability is dynamic, and the dip moves on a circle or a line in the x,y-plane with radius A and period T. In the dynamic cases is $r_{\text{dip}} = 90$. The green dashed line indicates the value of the reference simulation with constant surroundings. **b** The average center of mass in the conformation space on the `motility–1/division-rate' axis. Cases are distinguished and data is collected as in a).

Gradient steepness

Next, we observe how the gradient steepness which is linear to the gradient extent influences evolution. Evolutionary speed increases with decreasing steepness of the nutrient gradient(see Fig. 15 a, left). The speed is increased because the extent of the critical area, in which cells both die and divide expands together with the size of the gradient. Therefore, a shallower gradient leads to a larger volume with high selective pressure, which accelerates the evolution. The spread of the phenotype ensemble decreases with shallower gradients. With small nutrient dips, the number of cells inside the gradient is smaller, leading to a larger portion of cells that are outside the gradient. Cells outside the gradient find optimal conditions for growth and therefore do not experience evolutionary pressure. This leads to a loss of directionality of evolution and therefore to an expansion in all directions. A shallower gradient introduces a higher variability in the evolution of the motility/division axis. The median value is not significantly changed. For all subsequent simulations, a $r_{\text{dip}} = 90$ is used, and this simulation is indicated by the dashed line in Fig. 15 and 16.

5.2.2 Dynamic Environment

To examine the effects of a dynamic surrounding, a temporal dependency on nutrient availability was introduced. The center of the negative source was shifted in a periodical manner in one or two dimensions and different extents. All other parameters of the simulation remain the same. Three different possibilities are compared. One, the center of the dip moves on a straight line along the x-axis, following a sine function with amplitude $A = 50$ and period T. Two, the center follows a circle in the XY-plane with an amplitude A and period T, here two different amplitudes $A = 50$ and $A = 100$ are compared.

Global tumor effects

Evolution speed: For small values of the movement period ($1\,\text{kMCS} \leq T \leq 50\,\text{kMCS}$) the speed is significantly decreased, compared to the constant surrounding case (see Fig. 15 a). By further increasing the period, an increase in evolution speed can be observed for periods between $T = 75\,\text{kMCS}$ and $T = 200\,\text{kMCS}$. For further increasing values, the evolution speed remains high and decreases for a high amplitude circle case. This increase in evolution speed can be seen independently of the amplitude and motion type of the dynamics at the same frequencies.

Composition center: The period T of the dynamics does not seem to significantly affect the location of the evolutionary optimum. For the circular dynamics with a high amplitude (circle, $A = 100$), the variability of the optimum is reduced, visible through a smaller standard deviation of the result. The change on the motility axis when introducing a dynamic surrounding is small and can not be conclusively confirmed by only observing the center of mass in the phenotype space (see Fig. 15 b).

Spread: For small amplitudes ($A = 50$) the spread is increased over the constant surrounding case. The changing condition at each position presents temporally changing nutrient availability and therefore different evolutionary optima. This leads to a broadening in the phenotype space.

For simulations in which the amplitude of the circle is larger than the nutrient dip ($A = 100, r_{\text{dip}} = 90$), the spread is elevated significantly more (see Appendix 4 Fig. 28, right). Here, the spread of the distribution in the phenotype space is doubled, compared to the constant case for small values of the movement period ($1\,\text{kMCS} \leq T \leq 50\,\text{kMCS}$). For larger values of T the spread decreases (cf. Appendix 4 Fig. 28). The tumor center is never affected by the nutrient drop, and therefore always provides optimal conditions. This 'save spot' in the tumor center is responsible for the increased spread since the evolutionary direction is lost in there.

Local cell effects

In Fig. 16 the dynamics and temporal behavior of individual cells are observed. Tumor cells statistically move from the spheroid boundary inwards towards the tumor center during their lifetime (cf. Appendix 4 Fig. 33 a). In the constant case, the cells move inwards quicker and die earlier than in the dynamic case. Comparing the nutrient availability of individual cells over their

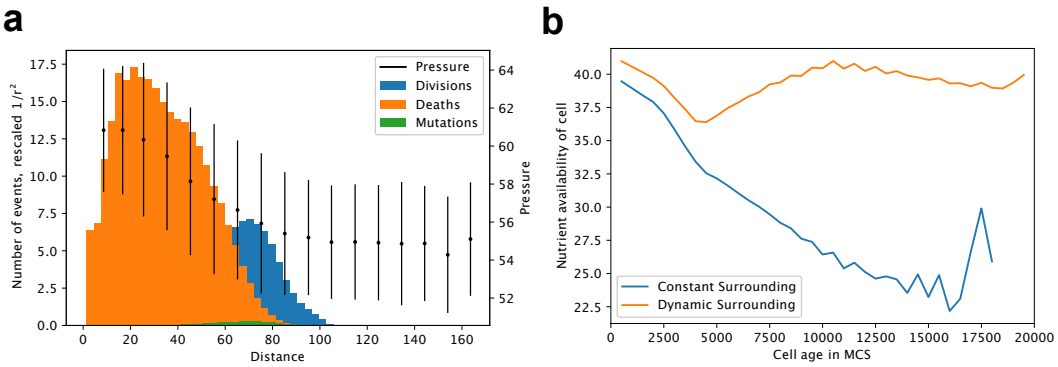

Figure 16: **Single-cell properties in dynamic surrounding.** **a** Normalized distribution of the locations of cell division, deaths and pressure for dynamic surrounding (line, $A = 50, T = 100$ kMCS). **b** Nutrient availability of the cells in relation to the cells age (time after cell division), for constant surrounding and dynamic surrounding (line, $A = 50, T = 100$ kMCS).

lifetime, a static decline proportional to the distance is visible for the constant case (cf. Fig. 16 b). This is reasonable since the nutrient availability is directly coupled to the position. Looking at the dynamic case, the nutrient availability first decreases, but then increases. This is due to the fact that cells divide in a region in which other cells die, cell death is introduced by a shortage of nutrients. After division, cells move inward towards the tumor center and the nutrient drop continues to move. Here, preferably cells survive that quickly exit the nutrient drop, or divide at the 'rising edge' of the moving gradient. We hypothesize that this is the main driver of the change in evolutionary optimum on the motility scale since fast-moving cells have a statistical advantage in this case.

We define the lifetime of a cell as the time between the last division and cell death. The mean lifetime and the extent of the lifetime distribution significantly increase with the introduction of a dynamic surrounding that enables a 'save spot' (cf. Appendix 4 Fig. 34 a). Here, the nutrient dip moves around the tumor in a circle that is larger than the tumor. The nutrient availability in the tumor center is therefore always optimal, providing zero evolutionary pressure and therefore a 'natural reserve' on the population scale. We observe the development of the mean lifetime of the tumor cells over time, see Appendix 4 Fig. 28 e. Overall, an increase in the mean lifetime is visible for both dynamic and constant nutrient surroundings. This increase in lifetime can be linked to the adaptation of the cellular properties to the surroundings and is a result of the evolution of the system.

Fitness evolution

We measure the fitness of each cell by tracing the lineage for the following eight generations and counting the descendants. The acquired numbers are then projected on the cell types and then on the parameter space spanned by the adhesion and motility parameters. Like this, the fitness optimum can be tracked over time. The behavior on the adhesion axis has been identified clearly, all tumors developed towards the low adhesion regime. The motility parameter was less conclusive by observing the center of mass of the cell type ensemble (cf. Fig. 15).

The fitness of the cell types is averaged along the adhesion axis. The resulting distribution

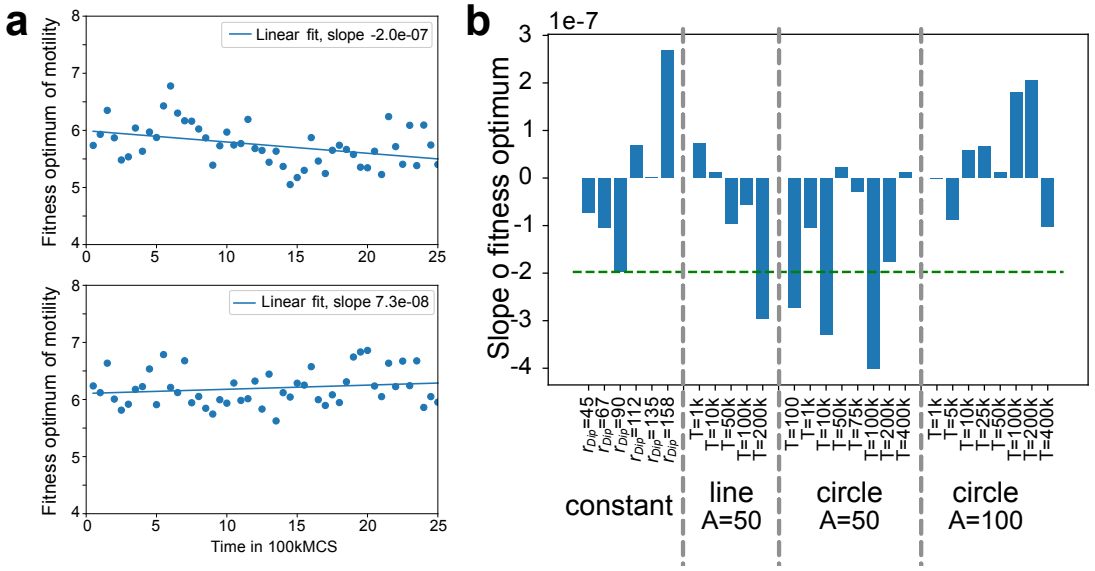

Figure 17: **Fitness evolution** The number of descendants in the next 8 generations, is defined as the fitness of a cell. **a** The fitness of all cells is determined at different times and projected on the cell types and their parameters. A Gaussian function is fitted to the resulting distribution and the central value is determined to find the fitness optimum on the parameter axis. Here, the fitness optimum of the motility parameter with division rate trade-off is plotted over time. An average of 15 simulations is pictured. Two simulations are compared, top: constant surrounding, bottom: dynamic on a line, $A = 50, T = 1$ kMCS. **b** The slope of the fitness development is plotted for each simulation. A positive slope leads to a development towards high motility, whereas a negative slope leads to a development towards low motility and high division rates.

shows the fitness in relation to the motility and division rate. This distribution has a clear maximum that is determined by fitting a Gaussian distribution to the data. The top value of this distribution is then plotted over time for different simulations in Fig. 17 a. We assumed to find a clear trend towards high motility cells in a dynamic surrounding, however, the behavior can not be decisively found and may be lost in fluctuations c.f. Fig. 17 b. A trend is visible when comparing the dynamic case of a circular motion with a radius of 100 to the dynamic case that shifts the fitness optimum towards more motile cells at the cost of a lower division rate.

5.3 Discussion

We present a computational model of a spheroid tumor in surrounding tissue. Mutation of cells is enabled by a change of phenotype at cell divisions. Two parameters can be changed during a mutation, cell–cell adhesion and the motility of the cell. We introduce a division rate trade-off for motility. The system is allowed to evolve freely and the tumor composition is tracked in parameter space over time.

We find that the mechanical and geometrical properties of the system are sufficient to drive the ensemble towards low-adhesion cell types. This mechanical effect in tissue evolution has been described in [16]. Mechanical properties alone drive proliferation at the tumor edge and cell death in the center.

We introduce a dependency of cell divisions and deaths on nutrient availability, which is linearly decreasing towards the tumor center. Using a linear dependence on nutrient availability for proliferation and inverse linear dependence for cell death leads to a higher evolutionary speed than a threshold-based dependency.

In *in vivo* tumors, solid stresses through tissue displacement are built up that are able to compress and block blood vessels [22]. This can lead to fluctuating and nutrient availability in tumors. We investigate how fluctuating nutrient availability influences tumor evolution.

We find that a temporally variable nutrient surrounding introduces a larger life span for cells. Especially a 'save spot' enables much longer lifetimes and a broader evolutionary spread

We find a significant dependency of dynamic nutrient surroundings on the evolutionary speed in phenotype space. The speed shows a frequency dependency, with a lower evolutionary speed for fast fluctuations followed by an increase over the constant case for lower frequencies. A critical time scale exists for the fluctuation of nutrient availability that provides a distinct peak in evolution speed, which we find to be between $T = 100\,\mathrm{kMCS}$ and $T = 200\,\mathrm{kMCS}$.

The fitness of cells can be determined by lineage tracing and the fitness is linked to the cell types and their parameters. With this, the effect of parameter values of a cell on its fitness can be determined. While a clear preference in fitness is visible for low adhesion cells, no clear change in the preferential direction of evolution can be identified along the motility axis. A trend towards higher motility is visible for large radii of nutrient fluctuations. We predicted that dynamic nutrient availability influences the fitness optimum of tumor evolution, this could not be conclusively be answered and has to be explored further by extending the range of possible motility and introducing different trade-offs.

Experimental work on spheroid tumors could provide verification of the results found here. The single-cell motility of cells grown in different spheroid cultures could be measured and compared [30]. The nutrient surrounding of the growing spheroid culture can be varied from a constant availability to a periodically changing nutrient concentration in the surrounding media. We expect to find cells with higher motility in the latter setup.

References

[1] L. M. Merlo, J. W. Pepper, B. J. Reid, and C. C. Maley, "Cancer as an evolutionary and ecological process," *Nature reviews cancer*, vol. 6, no. 12, pp. 924–935, 2006.

[2] N. McGranahan and C. Swanton, "Clonal heterogeneity and tumor evolution: past, present, and the future," *Cell*, vol. 168, no. 4, pp. 613–628, 2017.

[3] J. Hausser and U. Alon, "Tumour heterogeneity and the evolutionary trade-offs of cancer," *Nature Reviews Cancer*, pp. 1–11, 2020.

[4] J. A. Gallaher, J. S. Brown, and A. R. Anderson, "The impact of proliferation-migration tradeoffs on phenotypic evolution in cancer," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[5] R. Diaz-Uriarte and C. Vasallo, "Every which way? on predicting tumor evolution using cancer progression models," *PLoS computational biology*, vol. 15, no. 8, p. e1007246, 2019.

[6] Z. Bódi, Z. Farkas, D. Nevozhay, D. Kalapis, V. Lázár, B. Csörgő, Á. Nyerges, B. Szamecz, G. Fekete, B. Papp, *et al.*, "Phenotypic heterogeneity promotes adaptive evolution," *PLoS biology*, vol. 15, no. 5, p. e2000644, 2017.

[7] M. K. Jolly, P. Kulkarni, K. Weninger, J. Orban, and H. Levine, "Phenotypic plasticity, bet-hedging, and androgen independence in prostate cancer: Role of non-genetic heterogeneity," *Frontiers in oncology*, vol. 8, p. 50, 2018.

[8] A. Nagelkerke, J. Bussink, A. E. Rowan, and P. N. Span, "The mechanical microenvironment in cancer: How physics affects tumours," in *Seminars in cancer biology*, vol. 35, pp. 62–70, Elsevier, 2015.

[9] F. Bocci, L. Gearhart-Serna, M. Boareto, M. Ribeiro, E. Ben-Jacob, G. R. Devi, H. Levine, J. N. Onuchic, and M. K. Jolly, "Toward understanding cancer stem cell heterogeneity in the tumor microenvironment," *Proceedings of the National Academy of Sciences*, vol. 116, no. 1, pp. 148–157, 2019.

[10] S. Tripathi, P. Chakraborty, H. Levine, and M. K. Jolly, "A mechanism for epithelial-mesenchymal heterogeneity in a population of cancer cells," *PLoS computational biology*, vol. 16, no. 2, p. e1007619, 2020.

[11] M. Yilmaz and G. Christofori, "Mechanisms of motility in metastasizing cells," *Molecular cancer research*, vol. 8, no. 5, pp. 629–642, 2010.

[12] C. H. Stuelten, C. A. Parent, and D. J. Montell, "Cell motility in cancer invasion and metastasis: insights from simple model organisms," *Nature reviews Cancer*, vol. 18, no. 5, pp. 296–312, 2018.

[13] B. Waclaw, I. Bozic, M. E. Pittman, R. H. Hruban, B. Vogelstein, and M. A. Nowak, "A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity," *Nature*, vol. 525, no. 7568, p. 261, 2015.

[14] M. A. Clarke and J. Fisher, "Executable cancer models: successes and challenges," *Nature Reviews Cancer*, pp. 1–12, 2020.

[15] G. D. Grass, J. C. L. Alfonso, E. A. Welsh, K. Ahmed, J. Teer, L. B. Harrison, J. Cleveland, J. Mule, S. Eschrich, H. Enderling, *et al.*, "Harnessing tumor immune ecosystem dynamics to personalize radiotherapy," *bioRxiv*, 2020.

[16] T. Büscher, N. Ganai, G. Gompper, and J. Elgeti, "Tissue evolution: Mechanical interplay of adhesion, pressure, and heterogeneity," *New Journal of Physics*, vol. 22, no. 3, p. 033048, 2020.

[17] O. Ilina, P. G. Gritsenko, S. Syga, J. Lippoldt, C. A. La Porta, O. Chepizhko, S. Grosser, M. Vullings, G.-J. Bakker, J. Starruß, *et al.*, "Cell–cell adhesion and 3d matrix confinement determine jamming transitions in breast cancer invasion," *Nature cell biology*, vol. 22, no. 9, pp. 1103–1115, 2020.

[18] M. C. Lloyd, J. J. Cunningham, M. M. Bui, R. J. Gillies, J. S. Brown, and R. A. Gatenby, "Darwinian dynamics of intratumoral heterogeneity: not solely random mutations but also variable environmental selection forces," *Cancer research*, vol. 76, no. 11, pp. 3136–3144, 2016.

[19] A. Corcoran and R. F. Del Maestro, "Testing the "go or grow" hypothesis in human medulloblastoma cell lines in two and three dimensions," *Neurosurgery*, vol. 53, no. 1, pp. 174–185, 2003.

[20] T. Garay, É. Juhász, E. Molnár, M. Eisenbauer, A. Czirók, B. Dekan, V. László, M. A. Hoda, B. Döme, J. Tímár, *et al.*, "Cell migration or cytokinesis and proliferation?–revisiting the "go or grow" hypothesis in cancer cells in vitro," *Experimental cell research*, vol. 319, no. 20, pp. 3094–3103, 2013.

[21] W. Kang, J. Ferruzzi, C.-P. Spatarelu, Y. L. Han, Y. Sharma, S. A. Koehler, J. A. Mitchel, J. P. Butler, D. Roblyer, M. H. Zaman, J.-A. Park, M. Guo, Z. Chen, A. F. Pegoraro, and J. J. Fredberg, "Tumor invasion as non-equilibrium phase separation," *bioRxiv*, 2020.

[22] H. T. Nia, L. L. Munn, and R. K. Jain, "Physical traits of cancer," *Science*, vol. 370, no. 6516, 2020.

[23] J. Rosenbauer, C. Zhang, B. Mattes, I. Reinartz, K. Wedgwood, S. Schindler, C. Sinner, S. Scholpp, and A. Schug, "Modeling of wnt-mediated tissue patterning in vertebrate embryogenesis," *PLOS Computational Biology*, vol. 16, no. 6, p. e1007417, 2020.

[24] A. Buttenschön and L. Edelstein-Keshet, "Bridging from single to collective cell migration: A review of models and links to experiments," Nov. 2020.

[25] M. Berghoff, J. Rosenbauer, F. Hoffmann, and A. Schug, "Cells in silico–introducing a high-performance framework for large-scale tissue modeling," *BMC bioinformatics*, vol. 21, no. 1, pp. 1–21, 2020.

[26] F. Graner and J. A. Glazier, "Simulation of biological cell sorting using a two-dimensional extended potts model," *Physical Review Letters*, vol. 69, no. 13, p. 2013, 1992.

[27] A. Szabó and R. M. Merks, "Cellular potts modeling of tumor growth, tumor invasion, and tumor evolution," *Frontiers in oncology*, vol. 3, p. 87, 2013.

[28] J. A. Bull, F. Mech, T. Quaiser, S. L. Waters, and H. M. Byrne, "Mathematical modelling reveals cellular dynamics within tumour spheroids," *PLoS computational biology*, vol. 16, no. 8, p. e1007961, 2020.

[29] K. J. Flynn, D. O. Skibinski, and C. Lindemann, "Effects of growth rate, cell size, motion, and elemental stoichiometry on nutrient transport kinetics," *PLoS computational biology*, vol. 14, no. 4, p. e1006118, 2018.

[30] S. Galarza, H. Kim, N. Atay, S. R. Peyton, and J. M. Munson, "2d or 3d? how cell motility measurements are conserved across dimensions in vitro and translate in vivo," *Bioengineering & translational medicine*, vol. 5, no. 1, p. e10148, 2020.

6 Conclusions and Outlook

DESPITE spectacular successes in the understanding and treatment of cancer, tumors still present a leading cause of death. A large arsenal of mechanisms operates in unison to prevent illnesses in the human body. Intra and inter-cellular signaling detects dysfunctions and malignancies and activates cellular pathways that try to inhibit the disturbance. Tissues are in homeostasis, a state of constant self-renewal, that repairs defects and in case of disturbances, drives the tissue back towards homeostasis. The immune system actively scans the body for intruders and removes unknown and unwanted matter and cells. The defense mechanisms against external disturbances have been optimized as long as life exists and therefore display an immense intricacy. The disease agents, however, evolve alongside and display equally elaborate mechanisms to avoid the control of the body. Tumor development is, unlike viruses or bacteria, a condition that can arise within an organism without external influences. Growing tumors bypass the control mechanisms and cancer cells feign to be healthy. This requires the tumor cells to outwit the cell's internal control mechanisms, to grow against the surrounding tissue, to restructure the surroundings for sustained growth, and to evade the immune system. Those mechanisms act on different scales, from the sub-single cell to the tissue and organism scale. Each of those evaded control mechanisms constitutes a possible means of intervention for tumor treatments.

Tumor growth is a multi-scale problem with causes and effects in a range of length and time-scales from O($s, \mu m$) to O(years,cm). The scales depend on each other and together facilitate the growth and spread of tumors. Treatment and diagnosis focus on one scale at a time, leading to experimental and clinical data as well as behavioral models on each scale. The connection between the scales, however, is difficult to assess experimentally. Here, theoretical and computational modeling can help to understand the interplay between the different aspects. Computational modeling can connect these scales and improve our understanding of the relevant interactions. Models are always only an incomplete representation of the real-life situation and have been described as *"All models are wrong, but some are useful."*. Reflecting that models can help gain intuition, guide insights, and causal relationships by simplifying the view, as recently discussed by Enderling and Wolkenhauer [1]. Model building requires us to define and reevaluate the assumptions we make about the system and reflect a minimal set of assumptions that generate the behavior. A multi-scale model enables the detection of dependencies between single-cell properties and macroscopic tissue behavior.

In this thesis, I work towards the development and application of a physics-based multi-scale model of tissue development to drive our understanding of tumor growth.

A variety of different computational models is available, all with drawbacks and advantages. Here, I show the implementation of a cellular Potts model that incorporates the 3D cellular shape and makes use of supercomputing architectures. Short-range interactions facilitate parallelization of the model by a decomposition of the grid into communicating sub-grids. Cellular Potts models tend to generate grid based artifacts, this drawback is solved here by introducing an alternative surface metric employing the Marching Cube algorithm. I demonstrate the scalability on a supercomputer up to 50,000 cores and simulations of millions of individually resolved cells.

Next, the ability of the previously developed model to reproduce biologically observed behavior is validated by simulating $1.5 \cdot 10^6$ individually resolved interacting cells. A generalized model of tumor growth is introduced and the growth of heterogeneous tumors in surrounding healthy tissue and vasculature is modeled. By incorporating simple treatment models, the effects of global interventions on tumor size and heterogeneity are observed. The single-cell properties generate emergent behavior on the tissue scale (O(mm)) from single-cell parameters (O(µm)).

Down-regulated adhesion, increased motility, and lowered cellular stiffness are single-cell properties that have been associated with malignant tumors [2–5]. I ask how tumor-associated mechanical properties of single cells affect the dynamics of a forming tumor. Adhesion, motility, and stiffness are systematically scanned in a two-type model. In the simulations, a tumor cell population displaces a healthy cell population and the emergent tumor properties are observed for different cellular properties. Single-cell mechanical properties influence the invasiveness of a tumor and low-adhesion cells are most beneficial for tumor growth speed, while low growth rates facilitate the mixing of the cell types. Those properties are similarly found in malignant clinical tumors.

The adaption of a tumor to its surrounding and the optimization of cellular properties in a growing tumor depends on a large variety cell internal and surrounding of factors, next I ask: How does tumor growth select for the most advantageous cellular traits? The mutation rate of intra-tumoral cells is elevated, leading to heterogeneity in cellular properties within a tumor. Limited space and nutrients introduce competition between the cell types, leading to an evolutionary behavior of the cellular phenotypes. These evolutionary dynamics select for the fittest cell types in the current surrounding, fitness is a combined property of the tumor cells, their surroundings, and nutrient supply. In fast-growing tumors, the nutrient availability is unstable and irregular due to the pressure-induced collapse of blood vessels and reestablishment of new blood vessels through angiogenesis. Mutations that alter cell properties are introduced into a computational model, leading to the development of heterogeneous tumors. Through competition over space and nutrients, an evolutionary behavior between different cell types is observed in the tumor. I observe an evolutionary pressure towards low adhesion cells. A dynamically changing nutrient surrounding influences the evolutionary speed with a frequency-dependent increase. The temporal behavior of nutrient availability plays an important role in tumor evolution, and I determine a frequency regime in which a dynamic nutrient environment accelerates the evolutionary trajectory. Irregular availability of nutrients could therefore lead to accelerated tumor invasion. Through full access to the time-resolved trajectories of all model parameters, many different properties can be analyzed at the same time and be compared to experiments. The fitness of individual cells can be measured by lineage tracing, allowing access to parameters that are inaccessible experimentally.

Despite the recent successes in tissue modeling, challenges remain. Most prominently, there is a need for the incorporation of the large amount of heterogeneous data that is available from different sources. Here, comparability metrics and model optimization strategies need to be developed. Moreover, the physical characterization of tumor growth, invasion, and the formation of metastases in dependency of the cellular properties can help to identify means of intervention. Overall, the modeling of tissue is still a small area of research. However, the recent increase in available experimental data and computing power promises a bright future by bringing together

different aspects and scales of cancer research.

Many experimental and clinical measurements are taken on a single scale (e.g. tissue scale) and hard to transfer to other scales (e.g. cell or organism) which complicates the finding of causal dependencies. Computational models enable the analysis of causality by incorporating a multi-scale approach. As seen in recent works, simulations can help to extend the parameter space that is accessible in experiments, fill in missing regions, and even find new regimes [6, 7]. Agent-based modeling was used to identify critical parameters for the growth trajectories of treated tumors [8]. Computational models allow hypothesis testing with *in silico* experiments and fill the gap between different clinical and experimental methods. Mechanistic models can investigate the causality of correlative findings by deriving trajectories from a set of well-controlled assumptions. Together with the recent advances in spatial omics techniques, immune therapeutics, and genetic engineering capabilities, computational modeling of tissue will change our view on cancer. As we understand more and more about tumor development in its entirety. By combining the strengths of clinicians, experimentalists, and modelers, our understanding of tumor growth will be pushed further. Theoretical models as the one presented in this thesis enable a more mechanistic view on tumor growth and can connect different theories and experimental findings, thereby presenting tools to unite the different approaches to cancer research.

References

[1] H. Enderling and O. Wolkenhauer, "Are all models wrong?," *Computational and Systems Oncology*, vol. 1, no. 1, p. e1008, 2020.

[2] M. Janiszewska, M. C. Primi, and T. Izard, "Cell adhesion in cancer: Beyond the migration of single cells," *Journal of Biological Chemistry*, vol. 295, no. 8, pp. 2495–2505, 2020.

[3] M. Yilmaz and G. Christofori, "Mechanisms of motility in metastasizing cells," *Molecular cancer research*, vol. 8, no. 5, pp. 629–642, 2010.

[4] J. E. Kim, D. S. Reynolds, M. H. Zaman, and M. Mak, "Characterization of the mechanical properties of cancer cells in 3d matrices in response to collagen concentration and cytoskeletal inhibitors," *Integrative Biology*, vol. 10, no. 4, pp. 232–241, 2018.

[5] V. Swaminathan, K. Mythreye, E. T. O'Brien, A. Berchuck, G. C. Blobe, and R. Superfine, "Mechanical stiffness grades metastatic potential in patient tumor cells and in cancer cell lines," *Cancer research*, vol. 71, no. 15, pp. 5075–5080, 2011.

[6] O. Ilina, P. G. Gritsenko, S. Syga, J. Lippoldt, C. A. La Porta, O. Chepizhko, S. Grosser, M. Vullings, G.-J. Bakker, J. Starruß, *et al.*, "Cell–cell adhesion and 3d matrix confinement determine jamming transitions in breast cancer invasion," *Nature cell biology*, vol. 22, no. 9, pp. 1103–1115, 2020.

[7] W. Kang, J. Ferruzzi, C.-P. Spatarelu, Y. L. Han, Y. Sharma, S. A. Koehler, J. A. Mitchel, J. P. Butler, D. Roblyer, M. H. Zaman, J.-A. Park, M. Guo, Z. Chen, A. F. Pegoraro, and J. J. Fredberg, "Tumor invasion as non-equilibrium phase separation," *bioRxiv*, 2020.

[8] G. D. Grass, J. C. L. Alfonso, E. A. Welsh, K. Ahmed, J. Teer, L. B. Harrison, J. Cleveland, J. Mule, S. Eschrich, H. Enderling, *et al.*, "Harnessing tumor immune ecosystem dynamics to personalize radiotherapy," *bioRxiv*, 2020.

Appendices

7 Appendix 1: Embedded Publication 1

Berghoff, M.*, **Rosenbauer, J.***, Hoffmann, F. and Schug, A., 2020. *Cells in Silico–introducing a high-performance framework for large-scale tissue modeling.* BMC bioinformatics, 21(1), pp.1-21.
**Shared first authors.*

Copy permissions

Full Article

Berghoff *et al. BMC Bioinformatics* (2020) 21:436
https://doi.org/10.1186/s12859-020-03728-7

BMC Bioinformatics

SOFTWARE Open Access

Cells in Silico – introducing a high-performance framework for large-scale tissue modeling

Marco Berghoff[1†], Jakob Rosenbauer[2†], Felix Hoffmann[1] and Alexander Schug[2,3*]

*Correspondence:
al.schug@fz-juelich.de
†Marco Berghoff and Jakob Rosenbauer contributed equally to this work.
[2]John von Neumann Institute for Computing, Jülich Supercomputer Centre, Forschungszentrum Jülich, 52428 Jülich, Germany
[3]Faculty of Biology, University of Duisburg-Essen, 45141 Essen, Germany
Full list of author information is available at the end of the article

Abstract

Background: Discoveries in cellular dynamics and tissue development constantly reshape our understanding of fundamental biological processes such as embryogenesis, wound-healing, and tumorigenesis. High-quality microscopy data and ever-improving understanding of single-cell effects rapidly accelerate new discoveries. Still, many computational models either describe few cells highly detailed or larger cell ensembles and tissues more coarsely. Here, we connect these two scales in a joint theoretical model.

Results: We developed a highly parallel version of the cellular Potts model that can be flexibly applied and provides an agent-based model driving cellular events. The model can be modular extended to a multi-model simulation on both scales. Based on the NAStJA framework, a scaling implementation running efficiently on high-performance computing systems was realized. We demonstrate independence of bias in our approach as well as excellent scaling behavior.

Conclusions: Our model scales approximately linear beyond 10,000 cores and thus enables the simulation of large-scale three-dimensional tissues only confined by available computational resources. The strict modular design allows arbitrary models to be configured flexibly and enables applications in a wide range of research questions. Cells in Silico (CiS) can be easily molded to different model assumptions and help push computational scientists to expand their simulations to a new area in tissue simulations. As an example we highlight a 1000^3 voxel-sized cancerous tissue simulation at sub-cellular resolution.

Keywords: Tissue growth, Massively parallel, Cellular Potts model

Background

The mathematical description of organisms dates back to the beginning of the 20th century [1]. Since then, the theoretical understanding of biology has grown steadily, showing a more and more complex picture. With the emergence of computational models in physics, biophysicists started to adapt those models to describe biological processes [2]. An early development describing tissue development and cell–cell interactions was the

so-called cellular Potts model (CPM) by Graner and Glazier '92 [3]. This model derives from the Potts model and describes cells as connected areas on a grid. They were able to replicate known biological phenomena, such as adhesion driven cell sorting or tissue-growth. From then on, experimental insight into tissue on the cellular level as well as the power of computers has grown steadily, while the size and extent of cell-based tissue simulation have not proportionally evolved. Here, we present a modular framework for supercomputers to accommodate large-scale simulations of tissue with sub-single cell resolution.

Related work

There are several attempts to parallelize the CPM. Scianna and Preziosi [4] give an overview over advantages and disadvantages. Different methods were applied; for example, shared memory approaches set a lock to the memory that is accessed from parallel processes. Tomeu et al. [5] introduce a lock-free approach: the stencils compute concurrently, the write-back is only allowed if there are no other changes on the specific data, else an unroll is done. Some authors replace the random sampling of the field in the Monte Carlo, with a random walker that is simpler to parallelize. Gusatto et al. [6] used a mutex for shared memory and Cercato et al. [7] used a distributed memory version. Those implementations provide a maximum speedup of 5.4 for 12 cores and a decreasing speedup for increasing core numbers.

Another method that works for shared and distributed memory is a checkerboard method introduced by Chen et al. [8, 9] Here, the distributed sub-domains are split into $2 \times 2 \times 2$ parts, and only one part is active so that there is no overlap with other processes. For this model, a trade-off between accuracy and speed has been observed. If the sub-domain part is changed with a high frequency, a lot of communication is done compared to the runtime. On the other side, if it changes with a low frequency, cell movements stick to sub-domain boundaries. Tapia and D'Souza [10, 11] use this method to implement a single Graphics Processing Unit (GPU) version. Yu and Yang [12] use OpenCL to execute their model on GPUs and multi-core Compute Processing Units (CPUs).

He et al. [13] present a hybrid parallel version, where the CPM is calculated in shared memory, while additional partial differential equations use distributed memory methods.

Implementation

Cells in Silico (CiS), was implemented into the NAStJA (Neoteric Autonomous Stencil code for Jolly Algorithms) framework [14, 15]. Implementing the parallel CPM into the framework imposed several challenges, such as quasi global cell-state information and isotropic sampling of the field. To incorporate all necessary prerequisites, the framework was vastly extended to provide all the required infrastructure for large scale tissue simulations with the CPM.

NAStJA framework

The NAStJA framework is a modular, flexible framework for massively parallel stencil code applications. It uses the Message Passing Interface (MPI) to communicate between processes. The entire simulation domain is decomposed into small blocks, and these blocks are distributed to the different MPI ranks, see Fig. 1.

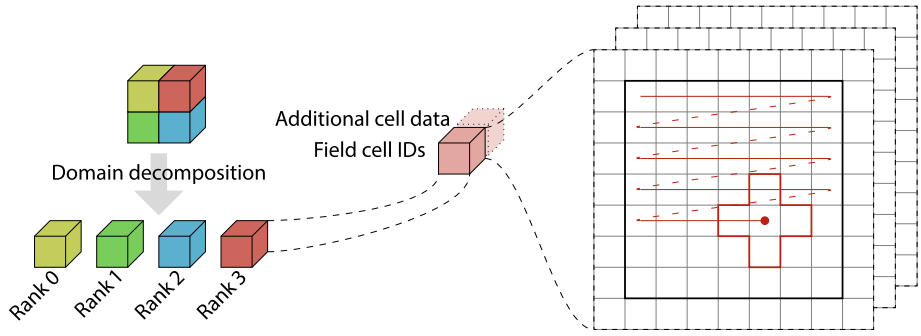

Fig. 1 The domain is decomposed and distributed to MPI ranks. So each rank holds one block. Each block contains a field with the cellID and additional cell data. A field is a three-dimensional array on which the compute actions are performed

Blocks represent a skeleton of the geometry, i.e., the size and position of the parallel entity in the global domain. Stencil codes act on regular grids, called fields. The blocks can hold one or more fields. The data inside the fields are located in voxels. Each voxel contains a data value. For the calculation, a stencil containing the calculation rules is applied to the voxels of the field. The calculation rule determines the data access pattern of the stencil. For n dimensions with m neighbor accesses, the stencil is denoted DnCm. In three dimensions, usual access patterns are the D3C7, i.e., the central voxel plus the first six neighboring voxels, or the D3C27 with 26 neighboring voxels, i.e., the full $3 \times 3 \times 3$ surrounding of a voxel. The neighbors are accessed read-only. Writing is always done at the central voxel of the stencil. For a consistent parallel calculation, the field in each block is enlarged by a halo layer, which overlaps with the neighboring blocks' fields. In order to keep the data in the halo up-to-date, each time-step is proceeded by a halo exchange.

After an initialization phase, NAStJA continuously runs the calculation loop. On a time-step base, a sequence of actions is executed independently on each MPI rank. Figure 2 provides an overview of the actions used in CiS. Details to the actions will be given in later sections.

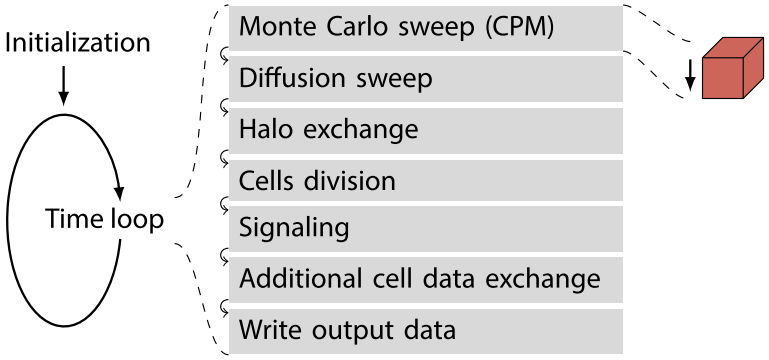

Fig. 2 The actions in one time loop are structures in modules. Additional modules can be added depending on the simulated system

Actions that iterate a stencil over the field are called sweeps. After all sweeps and actions, synchronization steps such as the halo exchange as well as output actions are executed. All sweeps and actions are implemented in a modular fashion, thereby enabling a quick alteration of the code structure by exchanging the modules or changing their order.

Additionally, NAStJA offers an interface for reading in configurations to parametrize simulations. Get-functions are implemented that read the data for a certain config-key from a JSON config file. This allows users to easily parametrize the simulation by changing the config file without modifying and recompiling the code.

Parallel cellular Potts model

In the last section, we reviewed the framework, the basic structure with blocks and fields, and the flexibility of the actions. This section first describes our implementation of the CPM and then its parallelization.

Each voxel in the field holds an integer value that denotes a cell identifier (cellID). Voxels that contain the same cellID belong to an individual biological cell. In addition to this spatial cell description, each cell has a set of Additional Cell Data (ACD), e.g., the cell volume V, the cell surface S, the cell age θ, and the cell type τ (cf. Table 1).

Cell types

A cell type is assigned to each cell, determining the parametrization and phenotype of that cell. The cell type defines the characteristics of the individual cells, i.e., the target volume V_0 and the target surface S_0 (cf. Table 2). This allows a parametrization of a set of cells instead of specifying the parameters individually. We introduced a subset of cell types that are not participating in the spatio-temporal propagation via the CPM. Those cells are fixed structures that can model blood vessels or the extracellular matrix, termed *solid* in our framework. The particular cell type *liquid* denotes the surroundings of the cells. It acts as a place holder for the growth of cells and describes the medium into which the cells grow.

Hamiltonian

The CPM was introduced by Glazier and Graner [3] in 1992 to simulate adhesion driven cell sorting. It is based on a Potts model that describes integer spin states on a regular

Table 1 Cell Properties (dynamically change during simulation)

Global variables (kept up to date in all blocks)	
cellID	The Value in the field that identify the cell
Volumes	The cell Volume
Surface	The cell surface (side counting or marching cubes)
Birth	Time of initialization of the cell
Type	The cell type
Center of mass	Center of mass of the cell.
Signal vector	Signal content of each signal within the cell
Temporary variables (block internal)	
Cell neighbor surfaces	Shared surfaces with neighboring cells
ΔVolume	Volume change during a MCS
ΔSurface	Surface change during a MCS
ΔSignal	Signal changes during a MCS

Table 2 Cell type properties (set by config file)

Parameter	Description
V_0	Target volume
S_0	Target surface
λ_V	Volume coupling factor
λ_S	Surface coupling factor
A_{ij}	Adhesion coupling matrix
Size change	constant rate of change of V_0 and S_0
Diffusion matrix	Diffusion constant matrix
Signal decay	Signal decay per time-step (relative and absolute)
Constant signal	Has constant signal
Start signal	Cells of this type are initiated with the signal content
Division	
Rate	Division rate
Age	Minimum division age
Signal thresholds	Minimum and maximum signal value
Mutation matrix	Probability to mutate to another type
Cell death	
Apoptotic cell type	Cell type of the apoptotic cells
Rate	Cell death rate
Age	Minimum cell age
Signal thresholds	Minimum and maximum signal value

lattice, in both two and three dimensions. The temporal propagation of the system is performed by Monte Carlo Sweeps (MCSs) over the field. Nearest neighbor interactions are evaluated by energy functions and are accepted with the Metropolis criterion. A Hamiltonian energy function defines the system energy, denoted as a sum of energy contributions E_i, weighted with λ_i. It reads,

$$
\begin{aligned}
H_{\mathrm{CPM}} &= \sum_i \lambda_i E_i \\
&= \underbrace{\lambda_{\mathrm{v}} \sum_{\varsigma \in \mathcal{C}} \left(\nu(\varsigma) - V(\tau(\varsigma))\right)^2}_{\text{Cell volumes}} + \underbrace{\lambda_{\mathrm{s}} \sum_{\varsigma \in \mathcal{C}} \left(s(\varsigma) - S(\tau(\varsigma))\right)^2}_{\text{Cell surfaces}} \\
&\quad + \underbrace{\sum_{i \in \Omega} \sum_{j \in N(i)} A_{\tau(\varsigma i), \tau(\varsigma j)} \left(1 - \delta\left(\varsigma_i, \varsigma_j\right)\right)}_{\text{Cell-to-cell adhesion}} + \ldots,
\end{aligned}
$$

where $\mathcal{C}$ is the set of all cells, Ω is the whole domain, and $N(i)$ are the neighbors of voxel i. Further, ς_i is the corresponding cell at voxel i, and ς_j is the corresponding cell at the neighboring voxel. Cell-to-cell adhesion is modeled by an energy contribution that is proportional to the shared surface of different cells. A is the adhesion coefficient matrix giving the adhesion between two cells of types $\tau(\varsigma_i), \tau(\varsigma_j)$, δ is the Kronecker delta. $\nu(\varsigma)$ is the volume of cell ς, $V(\tau(\varsigma))$ is the target volume of the cell type, λ_{v} is a coupling term regulating the strength of the volume constraint. $s(\varsigma_i)$ is the surface of cell ς, $S(\tau(\varsigma_i))$ is the target surface of the cell type, λ_{s} is a coupling term adjusting the strength of the surface constraint. The ellipsis indicates that the energy can be extended with various energy contributions.

The system propagation in the CPM is based on nearest-neighbor interactions. The cellID of a voxel can be changed to the cellID of a randomly chosen nearest neighbor voxel. Then, the energy difference ΔE of this local confirmation change is calculated via the change of the Hamiltonian energy function. Changes with negative energy differences are accepted, and positive energy differences have an exponentially decaying acceptance probability

$$p_{\text{accept}} = \begin{cases} 1, & \text{if } \Delta E < 0, \\ \exp(-\Delta E/T), & \text{otherwise.} \end{cases}$$

This is the Metropolis acceptance criterion with temperature T.

Energy calculations

The modularity of CiS allows adding various energy functions to the Hamiltonian. Each energy function gets the stencil and the direction of the neighbor as input parameters and returns the energy difference ΔE and local change of the surface and volume (ΔS and ΔV). In the function, the magnitude of the energy is determined by the internal cell states, as well as the corresponding coupling terms λ.

Surface Calculation The calculation of the surface of objects on a cubic grid is not unique. Depending on the chosen surface metric, dependencies may occur that prefer some spatial directions, leading to anisotropies in the emerging structures. Traditionally, a Manhattan metric is used to calculate the surface in the CPM. With this metric, the distance d between two points $\boldsymbol{a}, \boldsymbol{b}$ is defined by the sum of the absolute differences of their coordinates, $d(\boldsymbol{a}, \boldsymbol{b}) = \sum_i \mid a_i - b_i \mid$. In two dimensions, this corresponds to counting edges of pixels and in the three-dimensional to counting surfaces of voxels. With this metric, a unit circle has the same surface as a unit square. Likewise, in three dimensions, an ideal sphere of diameter a corresponds to a cube of edge length a after minimizing the surface. Particularly in the three-dimensional case, cell clusters tend to assume a cubic shape, when using the Manhattan distance for the surface calculations, introducing a non-isotropic grid dependence in the model. In order to ensure a more isotropic sampling of the field and to diminish grid artifacts, we use the marching cubes algorithm [16, 17]. The centers of eight adjacent voxels form the edges for the cube of the marching cube algorithm. Then, we distinguish between all edges that have the cellID that surface is calculated and all other cellIDs. Technically, we calculate the iso-surface for 0.5 by set the corners of the calculated cellID to 1 and all others to 0. The surfaces of both algorithms are presented in Fig. 3.

Volume Calculation The volume calculation is either done by counting the voxel or using the marching cube algorithm to approximate the volume.

Adhesion Calculation The adhesion energy difference is calculated by using the different surface metrics (side counting or marching cubes) to determine the change in shared surfaces between cells. The energy difference is determined by weighting the surface difference with the adhesion coupling matrix A.

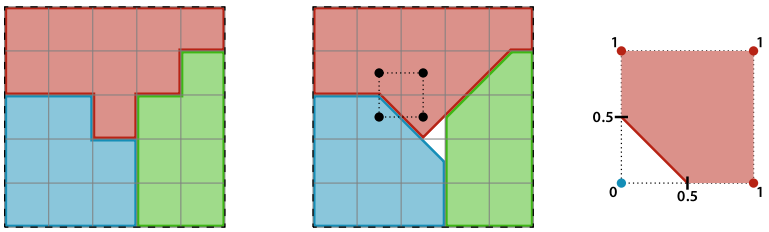

Fig. 3 Manhattan surface calculation (left) and a two-dimensional representation of the marching cube surface calculation (middle). With a surface of the red cell 8 using side counting (6.24, marching cubes), blue and green cells 6 (5.12). The marching cubes are shifted at denoted by the black rectangle, i.e., each voxel contributes to four marching cubes in 2D and eight in 3D space. The right side shows a detailed version of one marching cube, determined the surface for the red cell. The edges get the value 1 when it lies inside the red cell, 0 otherwise. The surface then is the 0.5 iso-line

Parallelism

While being propagated in parallel, the entire field has to be consistent. A stencil is needed for the calculation, which writes at the central position while reading from the neighboring voxels. For neighboring voxels located outside of the current block, a copy of the data from neighboring blocks is available due to the halo exchange. The halo data is constant during a Monte Carlo Sweep (MCS), consisting of a certain amount of Monte Carlo steps. To keep the halo data consistent with the neighbor block's data, the neighbor must not change the values read by the stencil. Therefore, it must be ensured that the neighboring voxels that are read have not been changed. This strategy is essential for all voxels in the halo. Since each voxel requires a uniform chance of sampling, we extend this condition to the entire field. Consequently, all read values within an MCS are from the previous time-step. Hence, the field data read within an MCS is independent of the access order.

To ensure the separation of read and written data, we introduce voxel-wise disjoint subsets similar to the black and white squares on a checkerboard. These subsets are regularly distributed over the entire domain, and only one subset is set active, i.e., only these voxels can be changed during system propagation. Note, the stencil can read all other voxels for the calculations. On the one hand, this ensures a uniform access pattern by not handling the boundary separately. On the other hand, it ensures that a stencil with a white center only reads from black fields. This satisfies the prerequisites described above.

During one MCS, the cell properties, such as surface and volume, stay constant. All changes in those properties are accumulated to delta storages, e.g., ΔS and ΔV. After one MCS, a synchronization step exchanges the halo and the deltas. The subsequent MCS acts on another active subset.

Checkerboards

The stencil size determines the number of necessary disjoint subsets. Figure 4 (left) shows a two-dimensional representation of the checkerboard for the D3C7-stencil.

Here, we need two subsets since the stencil only accesses the six nearest neighbors. To stay in the analogy, we denote each subset as a color of the checkerboard. For a given voxel, e.g., the red dot on the white voxel, the direct neighbor voxels are read but are not allowed to change. Diagonal neighbor voxels and the next-nearest neighbor voxels can change, so here a stencil calculation can be performed. In three dimensions, the layer in the front and the back are shifted by color.

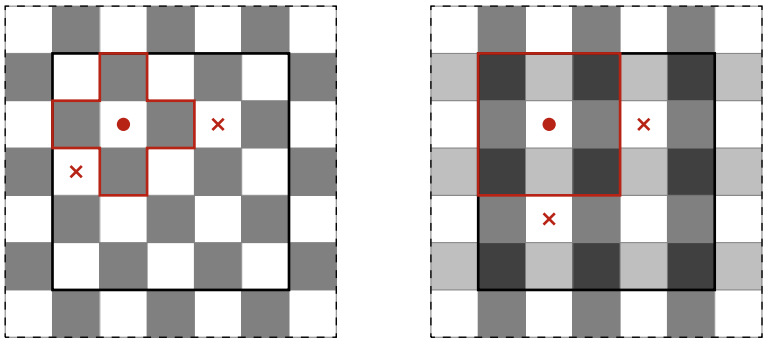

Fig. 4 A two-dimension representation of the D3C7- or a D3C27-stencil (A 3D stencil with 7 or 27 active voxels, respectively) with the two-colored (left) and eight-colored (right) checkerboard. The white subset is active and hence read and write accessible, while the other colors are read-only. The red dot represents the actual stencil position, and the red surrounding marks the input voxels. The red crosses mark the next possible stencil position in x- and y-directions

Figure 4 (right) presents the eight-colored checkerboard for the D3C27-stencil. The diagonals are used by the stencil itself so that the next stencil can only act on the voxel's next-nearest neighbors. In three dimensions, the layer in the front and the back use four different colors.

To achieve a uniform probability across the whole field, two or eight MCSs are required for the two- or eight-colored checkerboard, respectively.

Quality of pseudo-random numbers

Pseudo-random number generators in parallel applications can produce unintentional patterns [18]. This happens when the sequences overlap in different ranks, and the parallel entities use the same numbers. We use a standard generator based on the Mersenne Twister algorithm. Per MPI rank, one generator is used and initialized based on the MPI rank, so each generator starts on a different position in the random number sequence. We use the generator for all random numbers, e.g., random access and energy acceptance. Depending on the local domain data, a varying amount of random numbers is generated per Monte Carlo step. Therefore blocks with overlapping random number sequences, which is statistically extremely unlikely, do not correlate since the random numbers are used for various purposes.

Visitor pattern

We introduce a linear random access pattern. Therefore, the field on the active color is accessed in a linear walk. Since the volumes and surfaces are only updated after a complete MCS, large changes in volume or surface in one sweep produce unwanted behavior since the stored value strongly differs from the actual value. To restrict this discrepancy only a subset of all possible positions is sampled to avoid overshooting the changed parameters. Instead of randomly sample the whole field, we go linearly through the field while skipping a random amount of voxels until the end of the field is reached. By using this linear access pattern cached data can be reused and cache misses can be avoided. We use preliminary virtual voxels to ensure that the first voxel in the sweep is also selected uniformly.

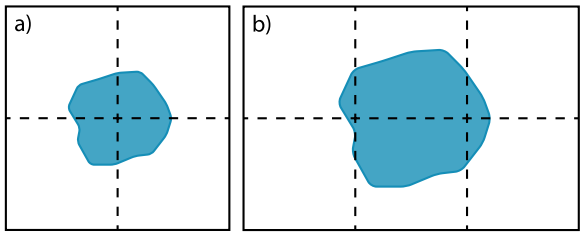

Fig. 5 (**a**) Valid and (**b**) invalid distribution of an cell (blue) over blocks (dashed rectangles). The invalid object distribution overlaps three blocks in x-dimension. The size of the cell is larger than the size of one box

Localize global information

The halo exchange ensures the consistency of the field data. Additionally, it must also be ensured that the ACD is updated after each MCS. Each block containing a part of a cell must have up-to-date ACD for that cell. A global exchange using collective MPI functions does not scale very well. However, introducing minor prerequisites allows an exchange of the ACD to all requiring blocks with local communications. If we limit the exchange to the first 26 neighbors, one cell may only stretch beyond the block boundaries on one side per dimension. Consequently, the cell size must be smaller than the block size per dimension, as shown in Fig. 5a. This can be guaranteed if the size of the cells is limited or the block size is large enough.

If a cell is illegitimate large and overlaps three blocks, as shown in Fig. 5b, a ACD exchange will not update consistently in all blocks. The changes in the left blocks do not reach the right blocks and vice versa.

The exchange is performed after each MCS as shown in Fig. 6.

A message is created that stores the ACD and additional exchanged information. For example, for each cell in a block the values of volume and surface as well as the delta volume and delta surface are stored this message package. Then, the package is sent to all 26 neighbors, received and unpacked. As soon as the data has been received from all neighbors, it can be processed. The sent deltas are accumulated to calculate the absolute values. Since cells move, they can newly enter into blocks. So, in addition to the changes of volume and surface, their absolute values must be transferred, such that the newly entered blocks can calculate the current volume and surface from the changes. The amount of transferred data depends on the number of different cells and the number of different

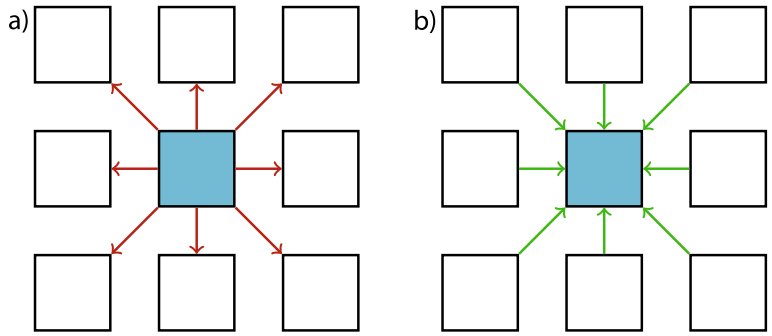

Fig. 6 Local neighbor exchange in two dimensions. The center block (blue) (**a**) sends to eight neighboring blocks and (**b**) receives from the same eight neighbors

types of values. Details of this exchange can be found in Ref. [19]. Each block holds the ACD for all cells, which are inside the block or in the halo. ACD for cells that are no longer in the block or halo is removed.

Cell events

Single-cell events have to be processed simultaneously on all blocks that hold a part of the specific cell. Therefore, single-cell events are split into two steps. The first step is the *determination step*, where events are detected and determined. This only happens in one block, namely the block containing the center of mass of the cell. The event is not executed immediately. The instruction is propagated to all adjacent blocks via the ACD exchange. In the following time-step, the *execution step* is processed in all blocks containing the cell. Here, the single-cell event is then executed consistently across all involved blocks.

Sanity action

As described above, the efficient parallel calculation requires a restriction of the cell size. Since we cannot have absolute control over every cell via the stochastic process, some single voxels may detach from a cell. Some references prevent this non-biological behavior directly in the energy calculation [20]. Here, we detect single voxels of a cellID without direct contact and replace them with liquid. In the rare event that several connected voxels detach, the identification of a segment is complex and cannot be calculated locally. In this case, we have two options, (i) we delete all voxels outside a predefined radius around the cell center, or (ii) we ignore the voxel detachment as long as it does not violate the requirements. I.e., if the cell and the voxel segment are moving away from each other, the condition that a cell can only go beyond one block boundary per dimension can be violated, which in turn leads to inconsistencies. This is detected within the ACD exchange, and the premature death of the cell will be inaugurated. If an ACD exchange receives ACD for one cell from opposite sides, then a so-called Message of Death (MoD) will be generated. This MoD is sent for two time-steps to all 26 neighbors, stored and forwarded to the neighbors of the neighbors. And finally, the cell is deleted simultaneously from all 125 neighbor and next-nearest neighbor blocks. This ensures that the cell is deleted from all blocks in which it can occur, and resolve the inconsistencies.

Agent-based cell actions

In addition to the system propagation described by the CPM, CIS provides several modules that allow multi-model simulation of more complex systems. These modules are using NAStJA's action system and are implemented as actions acting on the cell objects directly. Cell attributes such as the cell age, the signal level, cell type, etc. are append to the ACD. Actions act depending on these values.

Signal and nutrient transport

The simulation considers the transmission and propagation of multiple substances, such as nutrients and drugs. We define a class of signaling, e.g., nutrient contents, of each cell $\sigma_i^{(\varrho)}$, denoting the concentration of signal ϱ in cell i. Those represent nutrients for the cell (e.g., oxygen, glucose, or an effective nutrient concentration), cell-to-cell signaling compounds, or arbitrary drugs. Diffusion is approximated by a flow through the cells' surfaces.

Diffusion The diffusion of signals between the cells occurs through the surface of these cells. We determine the shared surface $S_{i,j}$ for each pair of cells i, j with $i \neq j$. The shared surfaces of cells are determined in a sweep over the field that locally saves the neighbors as well as the respective shared surface of each cell. The diffusion depends on the type of cells, so we define for each combination of types a diffusion constant $D_{\tau(i),\tau(j)}$, $\tau(i)$ denoting the cell type of cell i. The flux $J_{i,j}^{(\varrho)}$ for a signal ϱ is defined by

$$J_{i,j}^{(\varrho)} = \left(\frac{S_{i,j}}{S_i} + \frac{S_{i,j}}{S_j}\right) D_{\tau(i),\tau(j)} \left(\sigma_j^{(\varrho)} - \sigma_i^{(\varrho)}\right),$$

where S_i is the surface from cell i and σ_i is the signal value in cell i and S_j, σ_j from cell j, respectively. This is the arithmetic mean of the two surface fractions with respect to the common surface. The flux $J_{i,j}$ is subtracted from the delta signal of one cell and added to the other. Here, we distinguish between cells and fixed signal suppliers, such as blood vessels. For fixed signal suppliers, the signal content is kept constant, i.e., the flux is neither subtracted nor added for those cells. In order to keep the signal contents of all cells up to date, the delta signals are communicated with the ACD exchange to all neighboring blocks.

Decay Metabolic processes take place inside the cells, those as well as other signal depleting processes are described by the signal decay. In our model, the signals are changed relative to their value,

$$\sigma_i^{(\varrho)}\Big|_{t+1} = d_{\tau(i)}^{(\varrho)} \cdot \sigma_i^{(\varrho)}\Big|_t,$$

where $\sigma_i^{(\varrho)}\Big|_t$ is the signal ϱ in cell i at time t and $d_{\tau(i)}^{(\varrho)}$ is the relative change of the signal ϱ depending on the type of cell i.

Division and mutation

Cell division is a fundamental property of tissue development. During a cell division, one mother cell splits into two daughter cells. Those daughter cells usually inherit the properties of the mother cell, but in special cases such as asymmetric cell division and mutations, the properties can differ.

In each time-step, each cell is checked for cell-division. Whether a cell divides depends on several internal and external factors. Division happen with the division rate R_{Div} when the following conditions are fulfilled:

- Volume above a threshold $V > V_{\text{Div}} = 0.9 \cdot V_0$.
- Nutrition above a threshold C_{DivMin}.
- Age above a certain threshold $\text{Age}_{\text{DivMin}}$.

Then, a random plane through the cell center as well as a new cell type (see "§Mutation" section) is chosen. To ensure synchronous execution, this decision is then communicated to all neighboring blocks as described in "§Cell events" section. In the next time-step, the cell is split along the previously determined plane. The cell is split, and the two arising cells are reinitialized while measuring surface and volume. One keeps the cellID of the mother cell while the other receives a new cellID. After a cell division, the cellular age is set to zero for both daughter cells. Post division, both cells expand enforced by the volume and the surface energy term. Specific cell types can also be excluded from cell-division.

Mutation Mutation can occur with a rate of R_{Mut}, which is defined per cell type. If a mutation is accepted, one of the daughter cells is assigned a randomly selected cell type. A transition matrix between all cell types can be defined so that the transition probabilities between cell types vary. If no mutation occurs, the new cell inherits the type of the mother cell.

Cell death

The cell death is implemented with a death rate of R_{Death}, when the following conditions are fulfilled:

- Nutrition below a threshold C_{Death}.
- Age above a certain threshold $\text{Age}_{\text{DeathMin}}$.

Furthermore, cells dying with a reduced death rate $R_{\text{Death}}/1000$ to account for natural cell death. To ensure simultaneous execution of a cell death across all blocks, the death decision is communicated to all neighboring blocks as described in "§Cell events" section. Cell death is induced by changing the cell type to a dedicated cell type that describes dying cells. For this cell type the target volume in the Hamiltonian is changed over time $V_{0\text{Apop}}(t) = V_0 - \chi \cdot \text{age}$, effectively lowering the volume of the cell to zero voxels with a linear temporal dependence on the factor χ, that can be set for each cell type. Once the cell reaches $V = 0$, the cell and its ACD are deleted.

Output

NAStJA provides several input and output methods. In the following, we present the relevant writers for CIS. A writer is an action that can prepare, collect, and write out simulation data. The time resolution of the output can be chosen so that every n MCS an output frame is created.

CellInfo

The CellInfo writer outputs the ACD data of all cells to a comma-separated values (CSV) file per frame. The first line is a header that describes the parameter in each column. Each other row contains the data of one cell, e.g., cell type, age, volume, surface, center of mass, signals. Technically, each worker process creates the output of all cells that center of mass is inside its blocks. A master process collects this and writes it to a file.

Parallel VTK

The field data containing the cellIDs is written to a file in parallel using MPI-IO. Resulting in a single binary VTK image (VTI) file per frame. The file contains the whole simulation domain stored in a regular grid similar to a three-dimensional (3D) image. Each value is represented by a 32 bit or a 64 bit integer value, depending on the expected number of total cells. While CSV is a simple text file format, it can be easily read and processed. The VTI files can be read with the Visualization Toolkit (VTK) that provides a python binding and is supported by visualization software like ParaView. These file formats (.csv and .vti) together provide maximum interchangeability with other tools. Furthermore, we developed the NAStJA viewer. A fast and lightweight, quasi 3D visualization software that natively supports the combination of VTI and CSV files, as demonstrated in Fig. 7. It is freely available under https://gitlab.com/nastja/viewer.

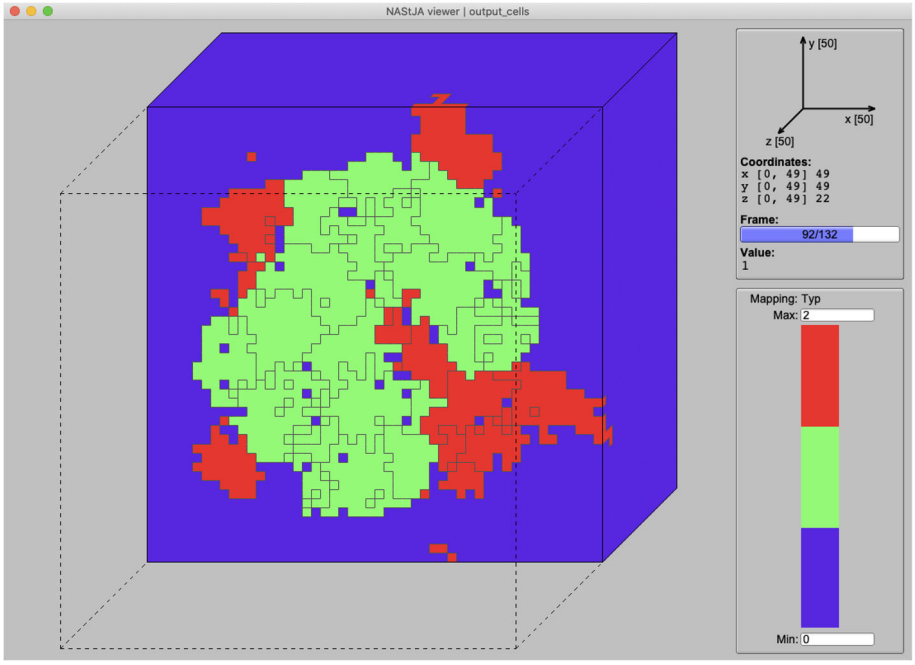

Fig. 7 The NAStJA viewer visualizes a slice of the domain. Black outlines mark cells, and the color represents a post-processing value denoting the neighboring to specific cell types

Results and discussion

In this section, we first show that the changes to the traditional model necessary for parallelization do not change the behavior of the model. Then we show the parallel performance and usage.

Statistical analysis

We verified that the execution in our parallelized framework does not distort the model behavior and does not depend on the chosen subdivision. We run 60 simulations with two cells distributed to $2 \times 2 \times 2$ blocks. We use cubic blocks, such that a block size of 100 refers to a cubic block with an edge length of 100 voxels without the halo. In the following, we write for cubic blocks shortly 100 ⧈. The one cell is set to the center of one block and the other cell is set to the edge of all blocks, i.e., to the center of the whole domain. The cells have $V_0 = 1\,000$ and $S_0 = 1\,400$. The marching cube algorithm is used for the surface calculations. Figure 8a shows the average over all simulations of the fluctuation in volume and surface over 250 000 MCSs, on the left side, the center cell and on the right the edge cell.

The behavior does not differ depending on the position within the subdivided field, i.e., a cell overlapping two blocks does not experience any directional bias. The temporal variation in surface and volume is statistically around 3%, this is due to the thermal fluctuations introduced by the metropolis criterion as well as a minor contribution of the delayed update of volumes and surfaces. Figure 8b shows the average position of cell center. Note, the cell centers are represented by an integer value denoting a specific voxel. The center of the cell statistically moves around the original position to a very small extent

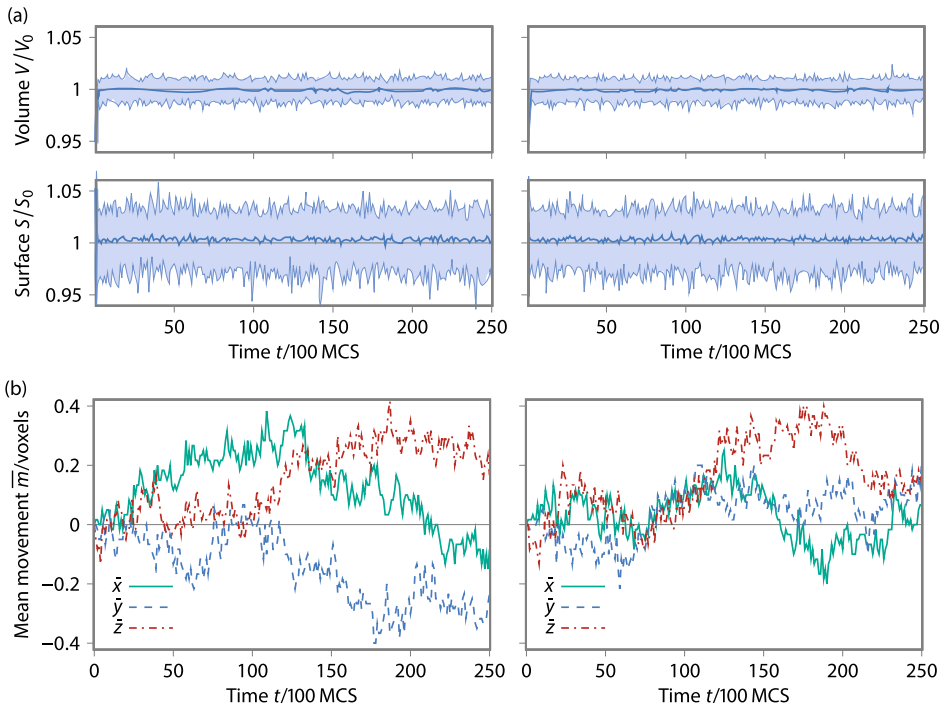

Fig. 8 (**a**) Ratio of the surface and volume to the goal values (V_0 and S_0) respectively (averaged over 60 simulations, shaded area indicates standard deviation). In the left plots the cell lies in the center of a block, on the right, the cell overlaps all eight blocks. (**b**) Movement of the cell center in the absence of an external potential. Averaged over 60 simulations

(0.4 voxels) in comparison to the extent of the cell (10 voxels). The movement of the cells also does not depend on the position within the subdivided field. These results confirm the strategy of the parallelization is valid.

Performance and scaling

We use a single node (kasper) and the high-performance computing systems ForHLR II at the Karlsruhe Institute of Technology (fh2) and JUWELS at the Jülich Supercomputing Centre to perform the performance and scaling tests. The single node has two quad-core Intel Xeon processors E5-2623 v3 with Haswell architecture running at a base frequency of 3 GHz, and have 4 × 256 KB of level 2 cache, and 10 MB of shared level 3 cache. The node has 54 GB main memory.

The ForHLR II has 1152 20-way Intel Xeon compute nodes [21]. Each of these nodes contains two deca-core Intel Xeon processors E5-2660 v3 with Haswell architecture running at a base frequency of 2.6 GHz, and have 10 × 256 KB of level 2 cache, and 25 MB of shared level 3 cache. Each node has 64 GB main memory, and an FDR adapter to connect to the InfiniBand 4X EDR interconnect. In total, 256 nodes can be used, which are connected by a quasi fat-tree topology, with a bandwidth ratio of 10:11 between the switches and leaf switches. The leaf switches connect 23 nodes. The implementation of Open MPI in version 3.1 is used.

JUWELS has 2271 48-way Intel Xeon compute nodes [22]. Each of these nodes contains two 24-core Intel Xeon Platinum 8168 with Skylake architecture running at a base frequency of 2.7 GHz, and have 24 × 1 MB of level 2 cache, and 24 × 1.375 MB of level

3 cache. Each node has 96 GB main memory, and an InfiniBand 4X EDR interconnect. ParaStation MPI in version 5.4 is used.

Node-level We run single-core CPM simulations including boundary condition (halo exchange) and the ACD exchange with sending and receiving on the same core. Since the cores in one processor have a shared level 3 cache and we want to avoid the related effects, we run a single-core application on each core simultaneously. Two different access patterns are used, a random access to the active checkerboard color and our linear access using a random jump width. Both methods use a mean voxel step width of 40. We vary the block size and compare the performance of the code in Fig. 9.

As a metric we use the number of million Monte Carlo step attempts per second (MMCs/s). The pure MCS (execution time of CPM-sweep) performance reaches the maximum of 37.8 MMCs/s for 20 🗗 blocks with a voxel step width of 40 (yellow line) It decreases until 100 🗗 and stays more or less constant for larger block sizes. The peak is clearly defined by the level 2 cache that has a maximum capacity of an equivalent 32 🗗 block, but is not exclusively usable by this data. Even if the whole block does not fit into the cache, we can profit from the property of the stencil to access only three layers of the field, the three layers are cached by the level 3 cache, and the data can be reused. The level 3 cache has a capacity of an equivalent of a 68 🗗 block, so that until this size no access to the main memory is needed. This describes the slope change in the curve at a block size of 60 🗗. The random access pattern can not benefit so much from caching the field and reaches only a peak performance of 4.6 MMCs/s.

The total time-step including exchange and cleaning stages, reaches a peak-performance of 11.4 MMCs/s at a block size of 40 🗗 for the linear access pattern, and 3.5 MMCs/s for the random access pattern, respectively. Here, we see that the overlap of the calculation and a nearly constant management overhead shifts the peak to a larger block size.

Scaling For testing the parallel scaling and efficiency, we use weak scaling. The simulation is initialized as a densely filled area of cells with a volume of 512 voxels each. The MCS used a mean step-width of five, with the eight-colored checkerboard. Signal diffusion is enabled. For each core we use one block, the size is varied from 20 🗗 to

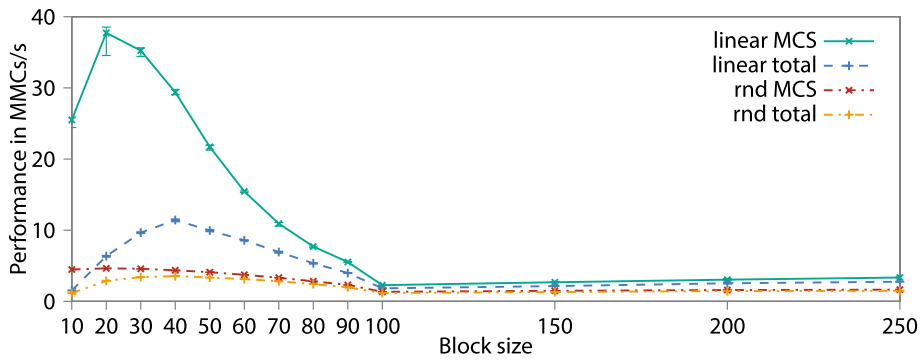

Fig. 9 Cache usage and performance of the pure MCS and the total time-step for a linear access pattern with random step width (linear) and a pure random access pattern (rnd)

100 🗊. Each simulation runs three-times on 1 to 256 nodes on fh2 and 1024 nodes on JUWELS. The largest simulations are containing approximately 100 million individual cells. Figure 10 shows the scaling performance and efficiency for up to 49 152 cores on JUWELS.

The efficiency $\eta = T_1/T_n$, where T_1 is the reference time for one node and T_n is the time for n nodes. On JUWELS (10b) two ranges can be recognized. First the small blocks (20 🗊–40 🗊) which show a high performance in the beginning and slow down with many cores, 128 nodes for 20 🗊 or 256 nodes for 30 🗊 and 40 🗊. The large blocks (60 🗊–100 🗊) do not reach the maximum performance, but do not drop down much and reach an efficiency of 60% on 49k cores. Up to 128 nodes (6 144 cores), a parallel efficiency of over 85% is reached for the small blocks. For more nodes, the communication overhead becomes significant compared to the calculation time for the small blocks. The gap between the small and large blocks reflects the influence of the cache examined in the previous section. Note that the block size 50 🗊 benefits from the cache for one node, but not for two and more nodes. The efficiency based on two nodes would show a scaling similar to the larger blocks.

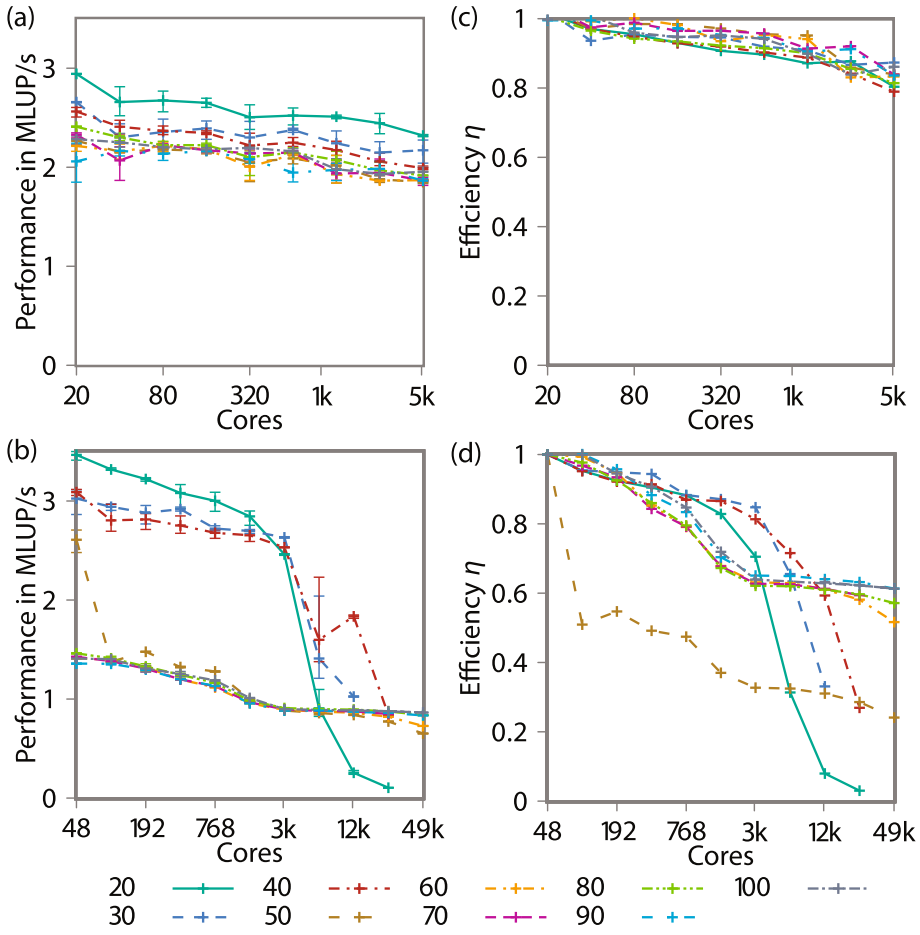

Fig. 10 Weak-scaling performance per core on fh2 (**a**) and JUWELS (**b**) in MLUP/s, i.e., scaled by the number of voxels per block. Efficiency of the scaling on fh2 (**c**) and JUWELS (**d**) of the CPM simulations, for the entire time-step, including halo- and ACD-exchange. The error bars denote the slowest and fastest run

On the fh2, the performance per core lies between the small and large blocks on JUWELS. Compared to all blocks, only the 20 shows a slightly better performance. The efficiency on 256 nodes (5 120 cores) is 80% – 90%.

User interface

The model parameters can be flexibly specified through a JSON config file. Multiple field initialization functions allow the placement of cells in the field. The placement of single cells at defined positions, as well as sets of cells in predefined shapes (cubes and spheres), are possible. Figure 11 list a example config file. The result is presented in Fig. 7.

```
{
  "Application": "Cells",
  "CellsInSilico": {
    "energyfunctions": ["Volume00", "Surface01", "Adhesion00"],
    "liquid": 0,
    "checkerboard": "01",
    "temperature": 50.0,
    "surface" : {
      "default": [0, 500, 500],
      "lambda": [0.0,2.0,2.0],
      "sizechange": [0, 0, 0]
    },
    "volume" : {
      "default": [0, 400, 400],
      "lambda": [0.0,4.0,4.0],
      "sizechange": [0, 0, 0]
    },
    "adhesion" : {
    "map": [[0.0,  0.0,  0.0],
            [0.0, 50.0,  5.0],
            [0.0,  5.0, 10.0]]
    }
  },
  "Filling": {
    "cells": [
      {
        "pattern":  "voronoi",
        "shape":    "sphere",
        "box": [    [15, 15, 15],[35, 35, 35]],
        "celltype": [ 0, 50, 50],
        "center":   [ 25,25, 25],
        "count":     100,
        "radius":    10
      }
    ],
    "initialoutput": true
  },
  "Geometry": {
    "blockcount": [1,1,1],
    "blocksize": [50,50,50]
  },
  "Settings": {
    "randomseed": 42,
    "timesteps": 1000000
  },
  "Writers": {
    "CellInfo": {
      "steps": 2000,
      "writer": "CellInfo"
    },
    "ParallelCells": {
      "field": "cells",
      "outputtype": "UInt32",
      "steps": 2000,
      "writer": "ParallelVtkImage"
    }
  },
  "WriteActions": ["ParallelCells", "CellInfo"]
}
```

Fig. 11 Minimal example of a config for a cell simulation

Since the actions, sweeps, and energy functions are modularly designed, it is possible to add specialized and additional code to the system by simply adding a function in a C++ file. Each function is documented in the code and input and output specifications are specified in the doxygen documentation of the code.

Currently, CiS is under heavy development to introduce new Features to discover new effects. Since NAStJA already provides a GPU infrastructure, we plan to provide a multi GPU implementation.

Without bias on the subdivision and grid, the model enables larger simulated volumes than other implementation of the CPM. These large scales allow us to study, among many other applications, emergent behavior of single-cell shapes to macroscopic tissues as well as tissue scale effects. Figure 12 shows a simple tumor model [23], in which a tumor seed grows through cell divisions into a large tumor.

Cells can acquire new types at cell division, corresponding to mutations in cells that yield new phenotypes. Cell division and cell death, depending on the availability of nutrients, which are distributed from a set of stationary blood vessels. Through the parametrization of a set of phenotypes, tumor development and the emergence and evolution of heterogeneity can be observed and tracked. The simulation yields a trajectory of each individual cell through time, giving access to all properties. Here, the effect of single-cell properties can be observed at a tissue scale. Tissue scale effects, such as fingering growth and a necrotic core, are also observable. Simulations are performed on a field consisting of $1\,000 \times 1\,000 \times 1\,000$ voxels and including 10^6 cells with 1 000 voxels

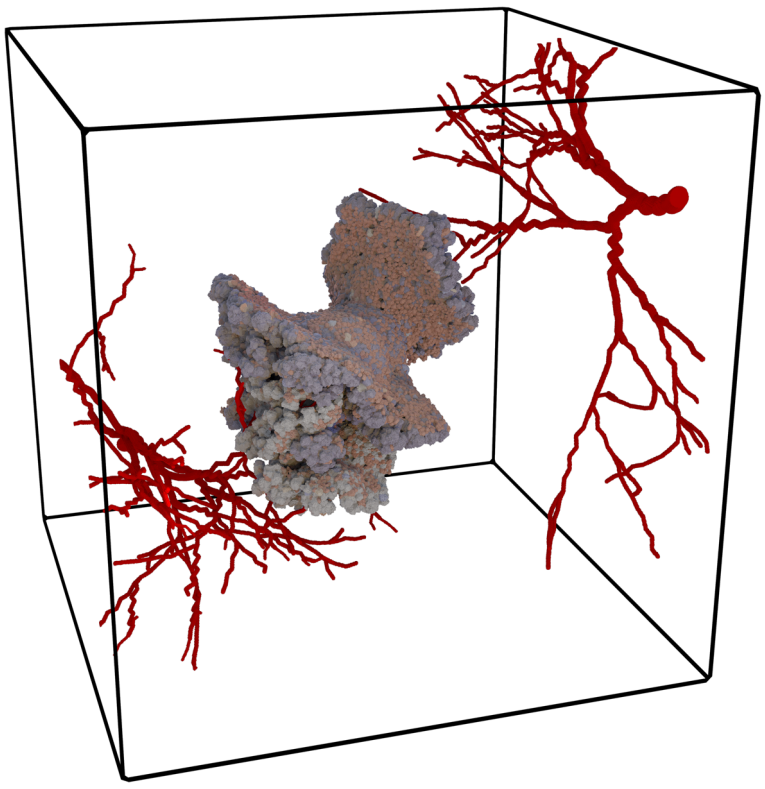

Fig. 12 Evolution of a tumor using CiS. Starting from a few tumor cells, the cancer grows along the nutrient-supplying blood vessels. For a clearer view, the surrounding healthy cells are not visible

each. The calculation was distributed to 1 000 CPU-cores and simulated for 24 h on the high-performance computing (HPC) system JUWELS [22]. The reached simulated time is around nine months in real-time.

Conclusions

We introduce a parallel implementation of the cellular Potts model (CPM) and demonstrate that the parallelization is bias-free. Around the CPM we developed the extendable parallel simulation framework Cells in Silico (CiS) for the simulation of tissue growth. CiS provides a user-friendly environment to implement new models. It provides an excellent scalability on supercomputers with a parallel efficiency of up to 90% on a small machine (fh2) with 5 120 cores. Larger simulations show an efficiency of up to over 60% for 49 152 cores, the lower efficiency is explained by an additional layer of switches, which is required for large simulations. The demonstrated performance per core for a full time-step is between 1 and 3.5 MLUP/s depending on the block size and the number of cores. With this scaling behavior, CiS enables large-scale tissue simulations up to some mm's and millions of interacting cells, while providing a geometric shape resolution of individual cells. Additionally to the geometry resolved cell simulations, we provide an agent-based model running in parallel performing single-cell events such as cell divisions, mutations, and cell death. Signal and nutrient transport through the tissue are enabled by a diffusion and signaling module that interacts with the cell geometries as well as the agent-based model to determine cell events and behavior. The model setup is designed to be user friendly by setting up simulations through a single configuration file that specifies model behavior and initialization. The simulation output is formatted in transferable easy to access data formats for broad compatibility and the use of standard tools. The entire framework is designed to have a fundamentally modular structure for easy model assembly and quick extension. The model is freely available to everyone under an open-source license.

It enables the use of the framework in a wide range of scientific applications opening up new areas for computational research by connecting the scales between single-cell data and tissue data in a single model. This model can be applied in simulations of tumor evolution and heterogeneity simulations, developmental biology, such as tissue patterning. The large scale of the simulations will enable new simulation of epithelial tissue, such as wound healing. In the future, the model can complement wet-lab experiments and testing through enabling large-scale simulations comparable to experimental and medical imaging methods.

With this contribution, we enable a new scale of tissue simulations that connect single-cell data with tissue scale measurements. We lift the barriers for large-scale simulations to a point where the upper bound is determined not by the model but by the parametrization and our imagination. This paves the way to bridge the scales between microbiological findings and medical images.

Availability and requirements

- Project name: NAStJA – Cells in Silico
- Project repository: https://gitlab.com/nastja/nastja
- Project home page: https://nastja.gitlab.io
- Operating systems: Linux, Mac OS

- Programming language: C++
- Other requirements: MPI, CMake
- License: Mozilla Public License, version 2.0 (MPL2.0)
- Any restrictions to use by non-academics: none

Abbreviations
2D: Two-dimensional; 3D: Three-dimensional; ACD: Additional cell data; cellID: Cell identifier; CiS: Cells in silico; CPM: Cellular Potts model; CPU: Compute processing unit; CSV: Comma-separated values; EDR: Enhanced data rate; FDR: Fourteen data rate; GPU: Graphics processing unit; HPC: High-performance computing; JSON: JavaScript object notation; MCS: Monte Carlo sweep; MLUP/s: Million lattice updates per second; MMCs/s: Million Monte Carlo step attempts per second; MoD: Message of death; MPI: Message passing interface; NAStJA: Neoteric autonomous Stencil code for jolly algorithms; OpenCL: Open computing language; VTI: VTK image; VTK: Visualization toolkit

Acknowledgments
The authors thank Steffen Scholpp and Nicolai Pfisterer for helpful discussions. This work was partly performed on the supercomputer ForHLR funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

Authors' contributions
MB leads the development of NAStJA. MB, JR, and AS designed the method and conceived the project. MB and JR designed examples and wrote code. MB and JR wrote the manuscript. FH proceeds simulations and create the statistical analysis plots. AS helped to finalize the manuscript. All authors read and approved the final manuscript.

Funding
MB, JR, and AS recognize support from the Helmholtz Impuls- und Vernetzungsfond. The funding body played no roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials
The source-code and examples are available on the gitlab repository.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Steinbuch Centre for Computing, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany. [2]John von Neumann Institute for Computing, Jülich Supercomputer Centre, Forschungszentrum Jülich, 52428 Jülich, Germany. [3]Faculty of Biology, University of Duisburg-Essen, 45141 Essen, Germany.

Received: 3 July 2020 Accepted: 31 August 2020
Published online: 06 October 2020

References
1. Schrodinger E. What is life? The physical aspect of the living cell. Cambridge University Press; 1944.
2. Green JB, Sharpe J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. Development. 2015;142(7):1203–11.
3. Graner F, Glazier JA. Simulation of biological cell sorting using a two-dimensional extended Potts model. Phys Rev Lett. 1992;69(13):2013.
4. Scianna M, Preziosi L. Multiscale developments of the cellular Potts model. Multiscale Model Simul. 2012;10(2):342–82.
5. Tomeu Hardasmal AJ, Salguero Hidalgo AG. A lock free approach to parallelize the cellular Potts model: Application to ductal carcinoma in situ. J Integr Bioinforma. 2020;1(ahead-of-print):. http://hdl.handle.net/10498/23293. https://doi.org/10.1515/jib-2019-0070.
6. Gusatto É, Mombach JC, Cercato FP, Cavalheiro GH. An efficient parallel algorithm to evolve simulations of the cellular Potts model. Parallel Process Lett. 2005;15(01n02):199–208.
7. Cercato FP, Mombach JC, Cavalheiro GG. High performance simulations of the cellular Potts model. In: 20th International Symposium on High-Performance Computing in an Advanced Collaborative Environment (HPCS'06). Piscataway: IEEE; 2006. p. 28. https://doi.org/10.1109/HPCS.2006.28.
8. Chen N, Glazier JA, Alber MS. A parallel implementation of the cellular Potts model for simulation of cell-based morphogenesis. In: International Conference on Cellular Automata. Berlin: Springer; 2006. p. 58–67. https://doi.org/10.1007/11861201_10.

9. Chen N, Glazier JA, Izaguirre JA, Alber MS. A parallel implementation of the Cellular Potts Model for simulation of cell-based morphogenesis. Comput Phys Commun. 2007;176(11-12):670–81.
10. Tapia JJ, D'Souza R. Data-parallel algorithms for large-scale real-time simulation of the cellular Potts model on graphics processing units. In: 2009 IEEE International Conference on Systems, Man and Cybernetics. Piscataway: IEEE; 2009. p. 1411–8. https://doi.org/10.1109/ICSMC.2009.5346282.
11. Tapia JJ, D'Souza RM. Parallelizing the cellular Potts model on graphics processing units. Comput Phys Commun. 2011;182(4):857–65.
12. Yu C, Yang B. Parallelizing the cellular potts model on gpu and multi-core cpu: An opencl cross-platform study. In: 2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE). Piscataway: IEEE; 2014. p. 117–22. https://doi.org/10.1109/JCSSE.2014.6841853.
13. He K, Jiang Y, Dong S. A hybrid parallel framework for the cellular Potts model simulations. In: 2009 15th International Conference on Parallel and Distributed Systems. Piscataway: IEEE; 2009. p. 624–31. https://doi.org/10.1109/ICPADS.2009.131.
14. Berghoff M, Kondov I, Hötzer J. Massively parallel stencil code solver with autonomous adaptive block distribution. IEEE Trans Parallel Distrib Syst. 2018;29(10):2282–96.
15. Berghoff M, Rosenbauer J, Pfisterer N. The NAStJA Framework: Zenodo; 2020. https://doi.org/10.5281/zenodo.3740079.
16. Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. ACM Siggraph Comput Graph. 1987;21:163–9. ACM.
17. Newman TS, Yi H. A survey of the marching cubes algorithm. Comput Graphics. 2006;30(5):854–79.
18. Matsumoto M, Wada I, Kuramoto A, Ashihara H. Common defects in initialization of pseudorandom number generators. ACM Trans Model Comput Simul (TOMACS). 2007;17(4):15.
19. Berghoff M, Kondov I. Non-collective scalable global network based on local communications. In: 2018 IEEE/ACM 9th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems (scalA). Piscataway: IEEE; 2018. p. 25–32. https://doi.org/10.1109/ScalA.2018.00007.
20. Durand M, Guesnet E. An efficient cellular Potts model algorithm that forbids cell fragmentation. Comput Phys Commun. 2016;208:54–63.
21. Steinbuch Centre for Computing. Forschungshochleistungsrechner ForHLR II. Website. https://www.scc.kit.edu/dienste/forhlr2.php. Accessed 14 Sept 2020.
22. Jülich Supercomputing Centre. JUWELS: Modular Tier-0/1 supercomputer at the Jülich supercomputing centre. J Large-Scale Res Facil. 2019;5:135. https://doi.org/10.17815/jlsrf-5-171.
23. Rosenbauer J, Berghoff M, Schug A. Emerging tumor development by simulating single-cell events. bioRxiv. 2020. https://doi.org/10.1101/2020.08.24.264150. https://www.biorxiv.org/content/early/2020/08/24/2020.08.24.264150.full.pdf.

Publisher's Note

8 Appendix 2: Supporting Information, Generalized Model

8.1 Simulations

General simulation parameter can be found in Table 3.

8.1.1 Initial Simulations

Treatment of one cell type: Homogeneous tumor growth in a non-dividing surrounding tissue. Simulations are conducted in a box with an edge size of 320 µm. Variation of treatment resistivity and resulting tumor response of a two-pulse treatment with radio- and chemotherapy, respectively.

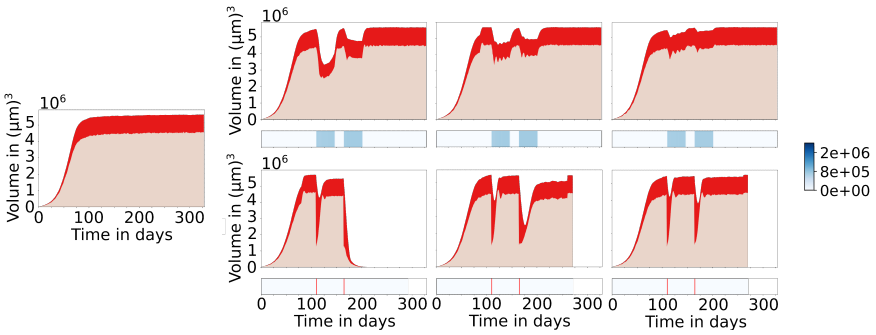

Figure 18: Tumor size response of a single cell type tumor of varying resistivities. From left to right: No Treatment, low, medium and high resistance against treatment. Chemotherapy schemes on the top row and radiation therapy schemes on the bottom row.

Treatment of resistivity heterogeneous cell type: Heterogeneous tumor growth in a surrounding tissue with a box edge size of 320 µm. Variation of treatment resistivity and resulting tumor response of a two-pulse treatment with radio- and chemotherapy, respectively.

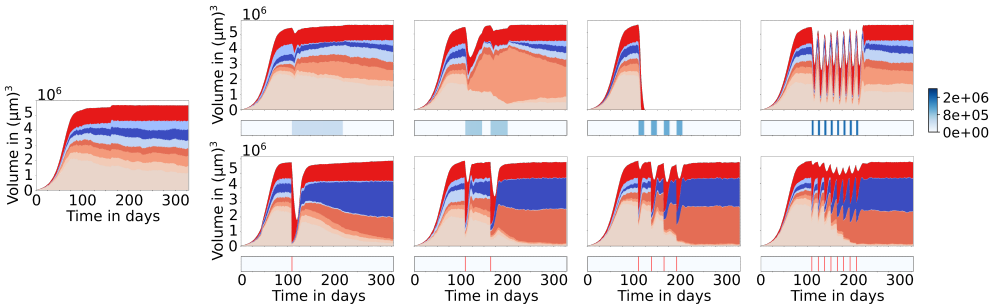

Figure 19: Treatment response to a heterogeneous tumor. The cell types are a different subset of cells than used for heterogeneity simulations in the main text, cell types only differ in treatment resistivity. Chemotherapy schemes on the top row and radiation therapy schemes on the bottom row.

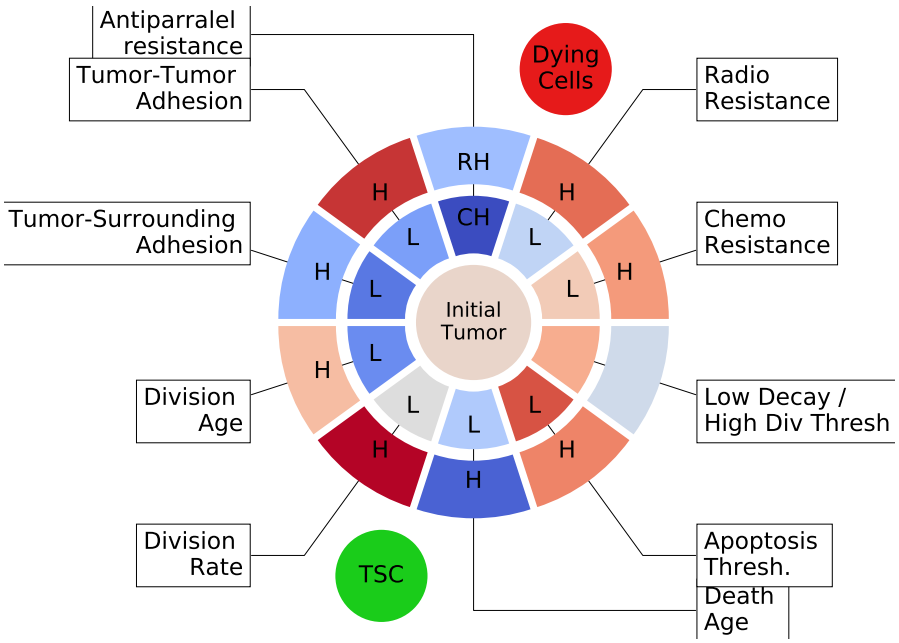

Figure 20: Colour coding of all cell types. The inner circle holds cell types with down-regulated parameters, while the outer circle holds up-regulated parameters.

8.1.2 Heterogeneity

We measure heterogeneity as:

$$h := 1 - \frac{\#T}{2(\#T-1)} \sum_{t \in \mathrm{T}} \left| \frac{\#A_t}{\#C} - \frac{1}{\#T} \right|, \quad \text{with } A_t = \{c \in C \mid \sigma(c) = t\} \tag{16}$$

T are all cell types C are all cells A_t are number of cells of type t, # denotes number.

8.1.3 Up-scaling

An interesting question is how the tumor development and treatment response depend on the tumor age and size. We take advantage of the scalability of our simulation framework and scale up the simulations by a factor of 27 so that the simulation covers $1\ \mathrm{mm}^3$ of tissue. The surrounding tissue is initialized to be densely vascularized.
In the tumor growth we see due to the bigger size of the tumors, more complex structures arising. The tumor grows into the direction of the closest blood vessels and even divides up into smaller compartments. Holes in the tumor emerge, and apoptotic regions are repopulated with the surrounding tissue.

8.1.4 Tumor Stem Cells

Similarly, tumor stem cells (TSC) are implemented as cells with a slow cell cycle. Both mutations and TSC lead to tumor heterogeneity and affect treatment response and tumor progression and

rejuvenation.
Tumor stem cells (TSC) are specialized cells within a tumor which through asymmetric cell divisions and a slower cell cycle, produce cancerous cells [1] and contribute to tumor rejuvenation as well as treatment resistance [2].
In our model, TSCs do not significantly change tumor size, growth rates, and final composition on unrestricted growth (see Figure 21 e)). The subpopulation of TSCs grows at a longer time scale due to a slower cell cycle (see Figure 21 e)) and are localized in small clusters but evenly distributed around the tumor (see Figure 21 d)).
The immediate treatment response is seemingly unchanged. Where subsequent chemotherapy in the heterogeneous tumor was able to suppress regrowth, TSCs can facilitate a subpopulation of cells to remain as a reservoir and trigger regrowth. TSCs lead to an increase in heterogeneity and boost the sub-populations of cell types that were suppressed. Figure 4 a) shows that TSCs can impact the final treatment outcome negatively by increasing tumor size as well as heterogeneity, especially in cases where treatment was successful in tumors without TSCs, see Figure 3 b).
In the heatmap in Figure 21 c), the type of a cell is compared to the types of its nearest neighbors. Diagonal elements are the major contributions, representing coherent clusters of cells of the same type. We find that TSCs lead to smaller cell clusters and a stronger mixing of cell types throughout the tumor, visible by increased off-diagonal elements. This is consistent with biologically observed behavior since tumors can regrow from a small number of remaining cells that are below the detection limit. [6]
Tumor stem cells introduce a source of treatment resistivity and are capable of facilitating a relapse of the tumor after seemingly successful treatment. Fundamental differences in the treatment response between tumors with and without TSCs were highlighted in this work, such as elevated intra-tumor heterogeneity and mixing. The stemness of a tumor has been experimentally associated with enhanced heterogeneity and treatment resistance [11, 12].

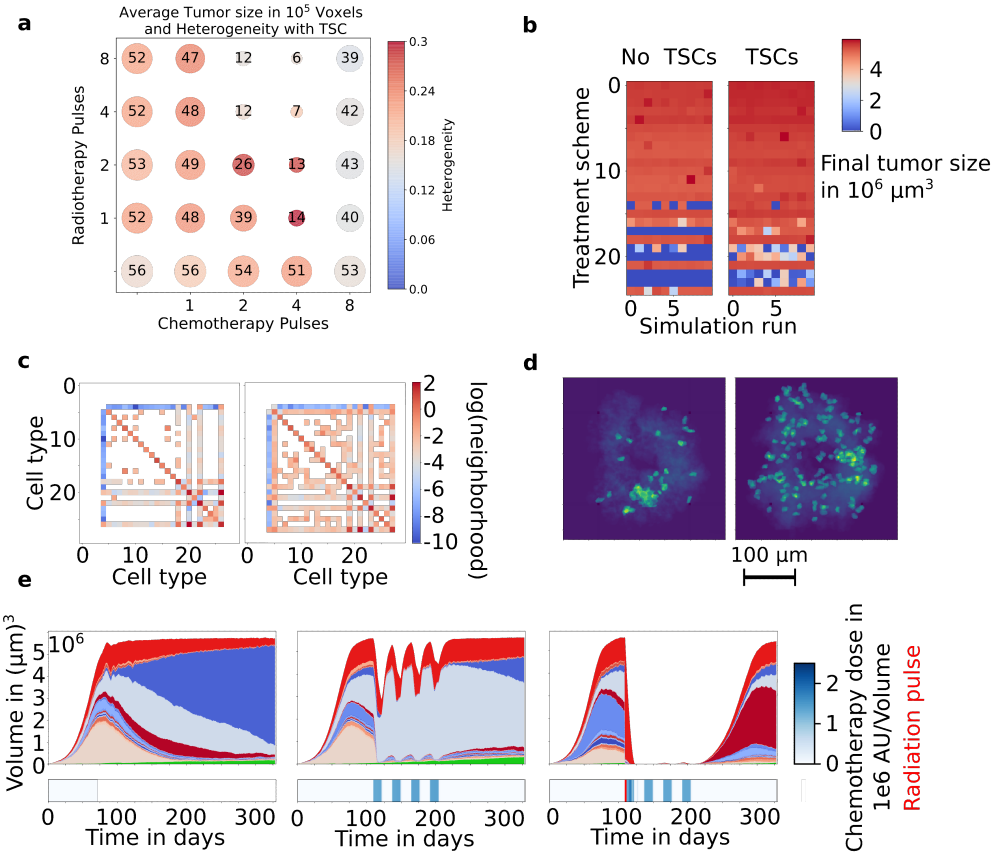

Figure 21: **Tumor Stem Cells:** a) Effect on final tumor size after treatment and tumor heterogeneity, errors as in Figure 8. b) The final tumor size of different treatment schemes (y-) and different simulation runs (x-axis) shows the stability of the simulation outcome. Coloring by tumor size at $t = 325$ days. c) Heatmap of the frequency of local neighborhood of cell types without (left) and with (right) TSCs. d) Localization of the tumor stem (in green) cells at $t = 73$ days (left) and $t = 292$ days (right), integrated over the entire volume. e) Response to different treatment schemes of chemo- and radiotherapy (Note regrowth bottom right). Colouring as in Figure 4.

Table 3: Model parameters that were used in the simulations.

Parameter	Symbol	Value	Source
Grid spacing		$1\,\text{voxel} = 1\,\mu\text{m}$	[3]
Cell Target Volume	V_0	$1\,000\,\text{voxel} = 2\,425\,\mu\text{m}^3$	Enterocyte volume $\approx 1\,400\,\mu\text{m}^3$; [16] HeLa Cells $2\,425\,\mu\text{m}^3$ Range $1\,198 \ldots 4\,290\,\mu\text{m}^3$
Cell Surface	S_0	$520\,\text{voxel}$	Ideal sphere: $S = (4\pi)^{\frac{1}{3}} (3V_0)^{\frac{2}{3}} = 483.6 \rightarrow S_0 = 1.075S = 520$
Surface compressibility	λ_S	1	[13]: $\lambda_S = 0.4$; [14]: $\lambda_S = 1$
Volume compressibility	λ_V	3	[10]: $\lambda_V = 2 \ldots 5$; [15]: $\lambda_V = 0 \ldots 20$; [13]: $\lambda_V = 20$
Adhesion	J_{ij}	$0 \ldots 100$	[13]: $(J = 0 \ldots 8)$; [14]: $(J = 1 \ldots 30)$ [10]: $(J = 0 \ldots 8)$; [15]: $(J = 0 \ldots 99)$
Temperature	T	50	[13] $(T = 60)$
Division rate	R_{Div}	0.001	
Division minimum age	$\text{Age}_{\text{DivMin}}$	2 000	
Minimal division volume	V_{Div}	$0.9 \cdot V_0$	
Apoptosis minimum age	$\text{Age}_{\text{DeathMin}}$	8 000	
Apoptosis rate	R_{Death}	0.02	
Size change of apoptotic cells $V_{0\text{Apop}}(t) = V_0 - \chi \cdot \text{age}$	χ	-0.05	
Division threshold Nutrition	C_{DivMin}	13	
Apoptosis threshold Nutrition	C_{Death}	12	
Division threshold *Chemo*		11	
Mutation rate	R_{Mut}	0.05	
Signal decay O^2		0.995	
Signal decay *Chemo*		0.995	
Diffusion constant O^2	D_{O^2}	0.1	
Diffusion constant *Chemo*	D_{Chemo}	0.1	

Table 4: Fitted tumor growth time constants from various *in vivo* experiments, mostly on mouse models. Time constants T (in hours) were fitted for tumors without and with treatment (chemotherapy and/or radiotherapy). Through manual literature mining, experimental parameters of growth rates of tumors were extracted. Exponential growth was fitted to control tumor growth and to treated tumor growth.

Paper	Treatment	Dose	Celltype	Model	T
[5]	Radiation	6 Gy	1483 HNSCC (Head and Neck cancer)	nude m.	201.22
[5]	Radiation	6 Gy fractioned doses day $0, 4, 10$	HT29 (human colon cancer)	nude m.	466.38
[5]	None		1483 HNSCC (Head and Neck cancer)	nude m.	116.19
[5]	None		HT29 (human colon cancer)	nude m.	254.82
[5]	Chemo, Radi	6 Gy, Celecoxib 250 mg/kg/day	1483 HNSCC (Head and Neck cancer)	nude m.	431.56
[5]	Chemo, Radi	6 Gy, 2B5 10 mg/kg $3\times$ per week	1483 HNSCC (Head and Neck cancer)	nude m.	258.60
[5]	Chemo, Radi	6 Gy fractioned doses day $0, 4, 10$, Celecoxib 250 mg/kg/day	HT29 (human colon cancer)	nude m.	599.50
[5]	Chemo	Celecoxib 250 mg/kg/day	1483 HNSCC (Head and Neck cancer)	nude m.	134.47
[5]	Chemo	2B5 10 mg/kg $3\times$ per week	1483 HNSCC (Head and Neck cancer)	nude m.	111.96
[5]	Chemo	Celecoxib 250 mg/kg/day	HT29 (human colon cancer)	nude m.	307.00
[8]	None		ER-Src (mamary gland)	mouse	121.70
[8]	None		BT-474 (Human Breast cancer)	mouse	127.87
[8]	Chemo	DOX 4 mg/kg	ER-Src (mamary gland)	mouse	173.38
[8]	Chemo	DOX 4 mg/kg	BT-474 (Human Breast cancer)	mouse	98.92
[4]	Radiation	8 Gy	Osteosarcoma (Bone tumor)	mouse	392.78
[4]	Radiation	3×3.6 Gy	Osteosarcoma (Bone tumor)	mouse	258.68
[4]	None		Osteosarcoma (Bone tumor)	mouse	126.22
[4]	Chemo, Radi	Caelyx 8 mg/kg, 8 Gy	Osteosarcoma (Bone tumor)	mouse	442.00
[4]	Chemo, Radi	Caelyx 8 mg/kg, 3×3.6 Gy	Osteosarcoma (Bone tumor)	mouse	359.40
[4]	Chemo	Caelyx 8 mg/kg	Osteosarcoma (Bone tumor)	mouse	210.28
[7]	Radiation	8 Gy	S180 (Sarcoma)	mouse	46.83
[7]	None		EL4 (lymphoma)	mouse	97.15
[7]	None		S180 (Sarcoma)	mouse	133.56
[9]	Radiation	5 Gy	HT29 (human colon cancer)	nude m.	556.39
[9]	Radiation	9×2 Gy	HT29 (human colon cancer)	nude m.	390.94
[9]	None		HT29 (human colon cancer)	nude m.	218.98
[9]	Chemo, Radi	5 Gy, Sunitinib 10 mg/kg/day	HT29 (human colon cancer)	nude m.	435.34
[9]	Chemo, Radi	9×2 Gy, Sunitinib 10 mg/kg/day	HT29 (human colon cancer)	nude m.	656.19
[9]	Chemo	Sunitinib 10 mg/kg/day	HT29 (human colon cancer)	nude m.	193.50

References

[1] E. Batlle and H. Clevers. Cancer stem cells revisited. *Nature medicine*, 23(10):1124, 2017.

[2] F. Bocci, L. Gearhart-Serna, M. Boareto, M. Ribeiro, E. Ben-Jacob, G. R. Devi, H. Levine, J. N. Onuchic, and M. K. Jolly. Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. *Proceedings of the National Academy of Sciences*, 116(1):148–157, 2019.

[3] R. J. Buschmann and D. J. Manke. Morphometric analysis of the membranes and organelles of small intestinal enterocytes. i. fasted hamster. *Journal of ultrastructure research*, 76(1):1–14, 1981.

[4] C. d. L. Davies, L. M. Lundstrøm, J. Frengen, L. Eikenes, Ø. S. Bruland, O. Kaalhus, M. H. Hjelstuen, and C. Brekken. Radiation improves the distribution and uptake of liposomal doxorubicin (caelyx) in human osteosarcoma xenografts. *Cancer research*, 64(2):547–553, 2004.

[5] T. W. Davis, J. M. O'Neal, M. D. Pagel, B. S. Zweifel, P. P. Mehta, D. M. Heuvelman, and J. L. Masferrer. Synergy between celecoxib and radiotherapy results from inhibition of cyclooxygenase-2-derived prostaglandin e2, a survival factor for tumor and associated vasculature. *Cancer research*, 64(1):279–285, 2004.

[6] G. V. Echeverria, Z. Ge, S. Seth, X. Zhang, S. Jeter-Jones, X. Zhou, S. Cai, Y. Tu, A. McCoy, M. Peoples, et al. Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Science translational medicine*, 11(488):eaav0936, 2019.

[7] M.-f. Guo, Y. Zhao, R. Tian, L. Li, L. Guo, F. Xu, Y.-m. Liu, Y.-b. He, S. Bai, and J. Wang. In vivo 99m tc-hynic-annexin v imaging of early tumor apoptosis in mice after single dose irradiation. *Journal of experimental & clinical cancer research*, 28(1):136, 2009.

[8] D. Iliopoulos, H. A. Hirsch, and K. Struhl. Metformin decreases the dose of chemotherapy for prolonging tumor remission in mouse xenografts involving multiple cancer cell types. *Cancer research*, 71(9):3196–3201, 2011.

[9] E. A. Kleibeuker, E. Fokas, P. D. Allen, V. Kersemans, A. W. Griffioen, J. Beech, J. H. Im, S. C. Smart, K. C. Castricum, J. van den Berg, et al. Low dose angiostatic treatment counteracts radiotherapy-induced tumor perfusion and enhances the anti-tumor effect. *Oncotarget*, 7(47):76613, 2016.

[10] J. F. Li and J. Lowengrub. The effects of cell compressibility, motility and contact inhibition on the growth of tumor cell clusters using the cellular potts model. *Journal of theoretical biology*, 343:79–91, 2014.

[11] N. K. Lytle, A. G. Barber, and T. Reya. Stem cell fate in cancer growth, progression and therapy resistance. *Nature Reviews Cancer*, 18(11):669–680, 2018.

[12] A. Miranda, P. T. Hamilton, A. W. Zhang, S. Pattnaik, E. Becht, A. Mezheyeuski, J. Bruun, P. Micke, A. de Reynies, and B. H. Nelson. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proceedings of the National Academy of Sciences*, 116(18):9020–9029, 2019.

[13] N. J. Poplawski, A. Shirinifard, U. Agero, J. S. Gens, M. Swat, and J. A. Glazier. Front instabilities and invasiveness of simulated 3d avascular tumors. *PloS one*, 5(5):e10641, 2010.

[14] B. M. Rubenstein and L. J. Kaufman. The role of extracellular matrix in glioma invasion: a cellular potts model approach. *Biophysical journal*, 95(12):5661–5680, 2008.

[15] A. Szabó and R. M. Merks. Blood vessel tortuosity selects against evolution of aggressive tumor cells in confined tissue environments: A modeling approach. *PLoS computational biology*, 13(7):e1005635, 2017.

[16] L. Zhao, C. Kroenke, J. Song, D. Piwnica-Worms, J. Ackerman, and J. Neil. Intracellular water-specific mr of microbead-adherent cells: the hela cell intracellular water exchange lifetime. *NMR in Biomedicine*, 21(2):159–164, 2008.

9 Appendix 3: Supporting Information, Two-type Model

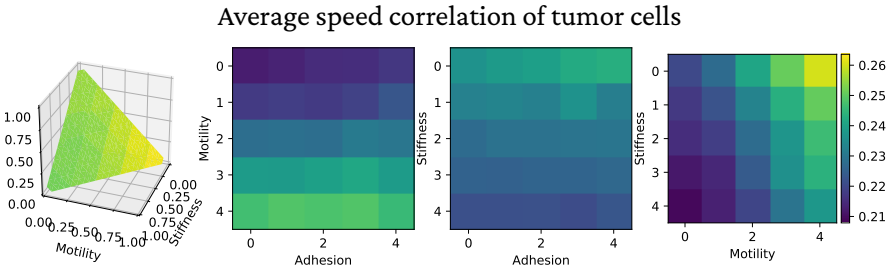

Figure 22: Parameter projections on the 125 simulations on a 5x5x5 cube. Left image shows a cut through the 1,1,1 plane. The other images depict the mean value of the cube along each axis. Speed directional correlation of tumor cells.

An additional tumor property we observe it the directional correlation of the movement of the tumor cells. While adhesion does not affect the correlation, motility and adhesion do have a strong effect. Low stiffness increases the directional correlation of the cells inside the tumor. A continuous rise of correlation is visible with decreasing stiffness. A positive dependency of directional correlation and cell motility is visible.

10 Appendix 4: Supporting Information, Evolution Model

10.1 Transition Rates

In order to be able to observe biases in the evolutionary development in the phenotype space, it is necessary to know the dynamics of the unbiased system. Here, we performed simulations on a 1D / 2D array. The content of the array represents the abundance of a phenotype. To model the process in the simulation, the occupancy of each phenotype doubles at every time step, spreading to the nearest neighbor phenotypes with the rate of the central phenotype. The system is propagated by increasing the value of the 2/4 nearest neighbors by the rate times the content of the current phenotype and the phenotype by (1-numNeihbors*rate)*value
Using a constant rate to populate each nearest neighbor across the phase space, leads to a symmetric spread, with a tendency of the center of mass to center in the field. For reduced rim rates border effects of pileup create not desirable offsets to the extreme values.
A constant rate enables the detection of an evolutionary bias, if the center of mass in the phenotype space deviates from the center. A constant rate as indicated in e) is used in the simulations.

for all nearest neighbors:

$v_{\text{Nei}} += v_{\text{Cur}} * rate(\text{cur})$

for current value:

$v_{\text{Cur}} += (1 - \text{numNeighbors} * rate(\text{cur})) * v_{\text{Cur}}$

At the borders only a smaller number of nearest neighbors is used.

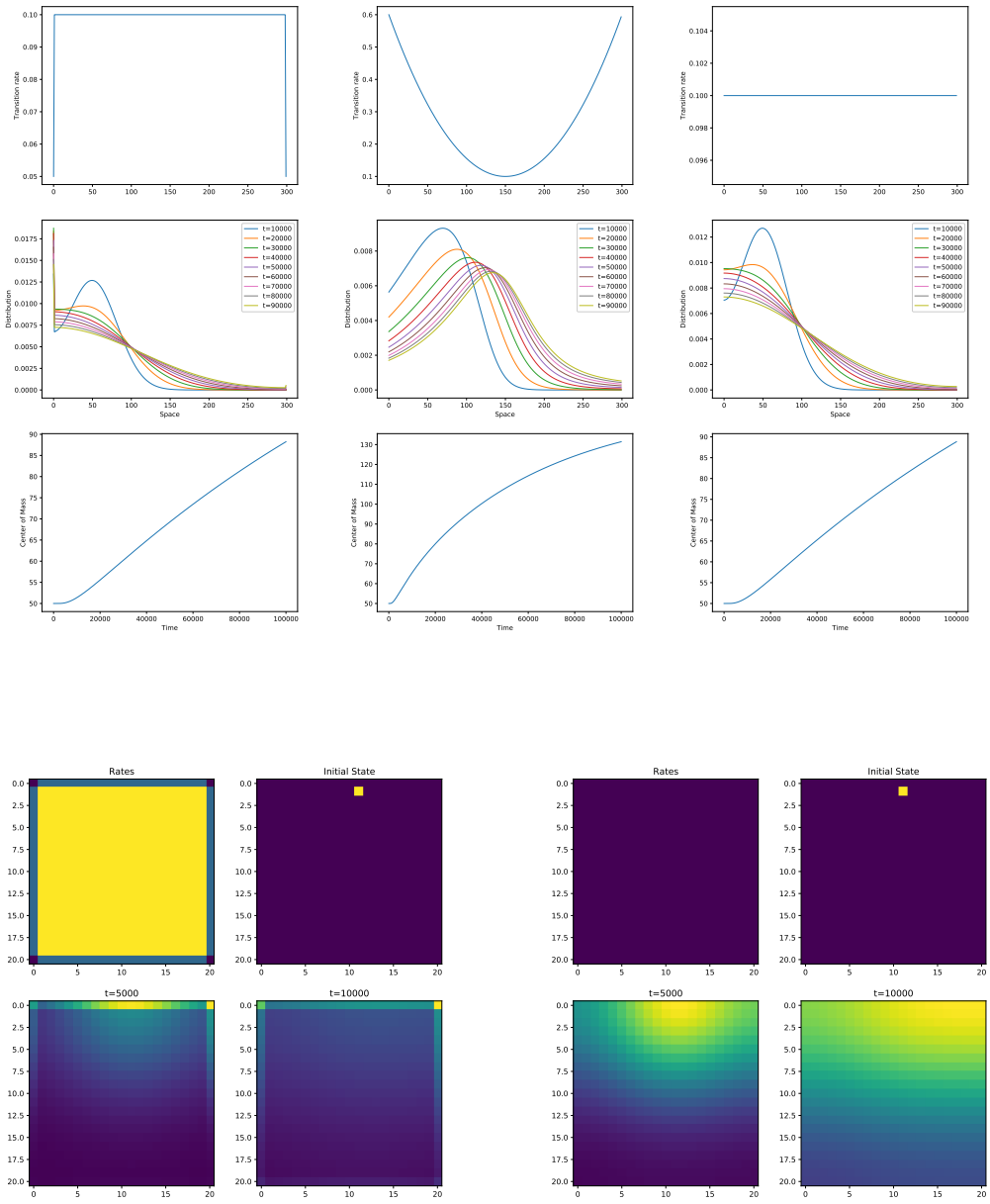

Figure 23: Rate distributions and their influence on the ensemble development. a)-c) 1D Simulations, with a edge reduced, quadratic and constant rate distribution. d) 2D Simulation of an ensemble development with reduced edge rates. e) 2D simulation with constant rates

10.2 Scan mutation rate

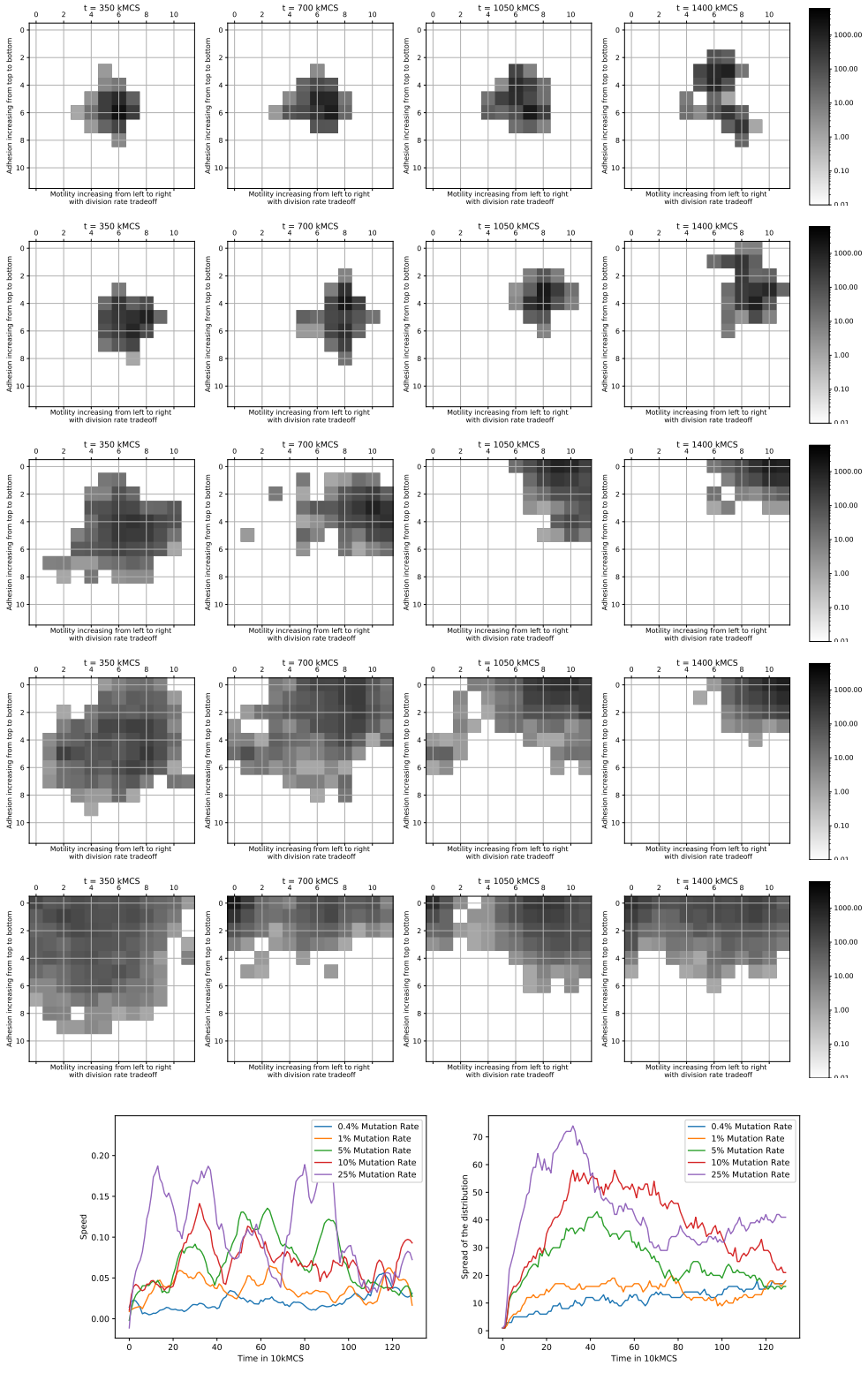

Figure 24: Scan of mutation rate, 0.4 , 1 , 5 , 10 , 25 percent probability of mutation, Potential coupling = 70

10.3 Scan Tradeoff coupling

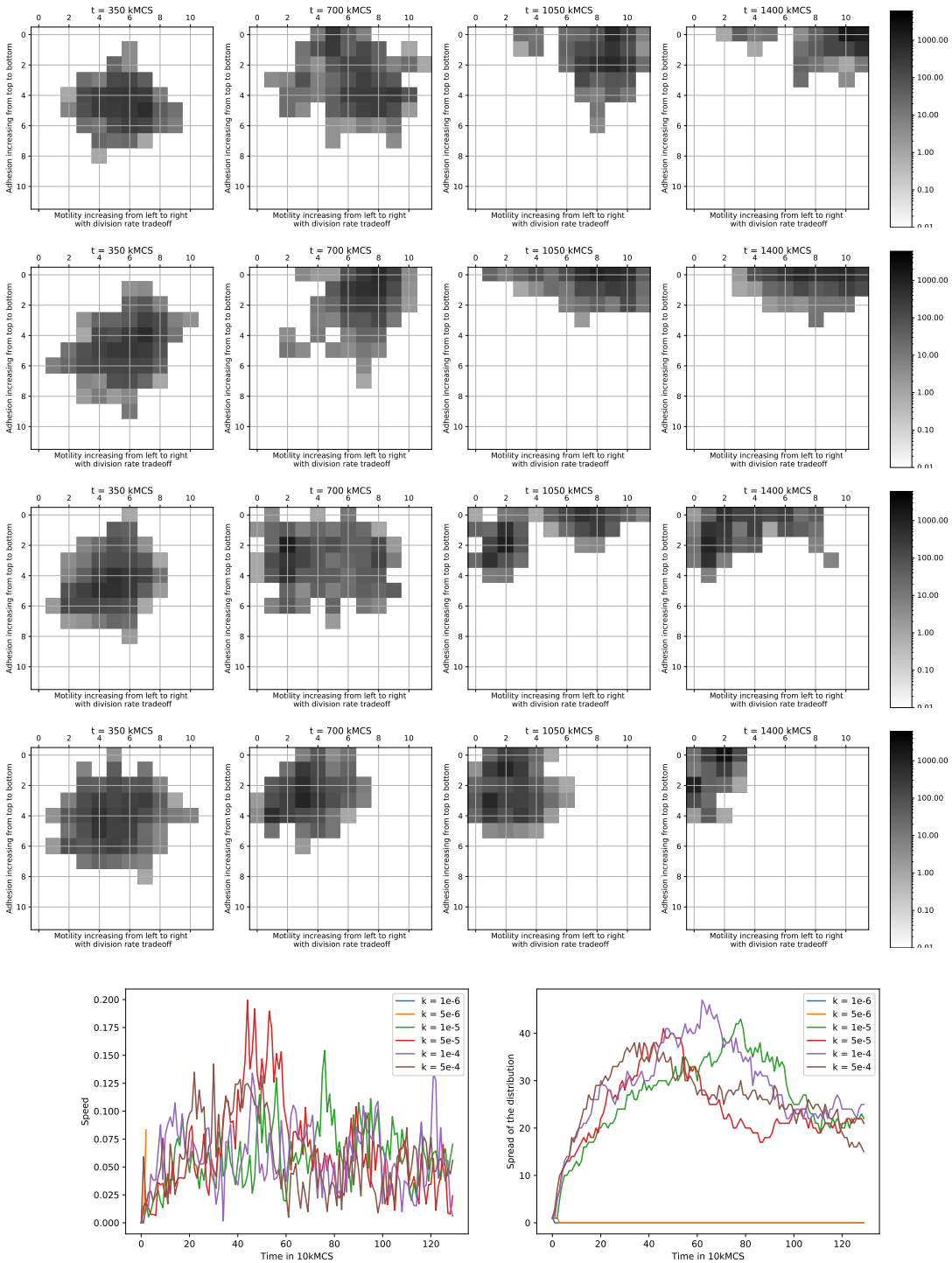

Figure 25: Tradeoff function parameter k, $1 \cdot 10^{-6}$, $5 \cdot 10^{-6}$ did produce no tumor. Displayed are the compositions for k= $1 \cdot 10^{-5}$,$5 \cdot 10^{-5}$,$1 \cdot 10^{-4}$,$5 \cdot 10^{-4}$,$1 \cdot 10^{-3}$

10.4 Scan Central potential coupling

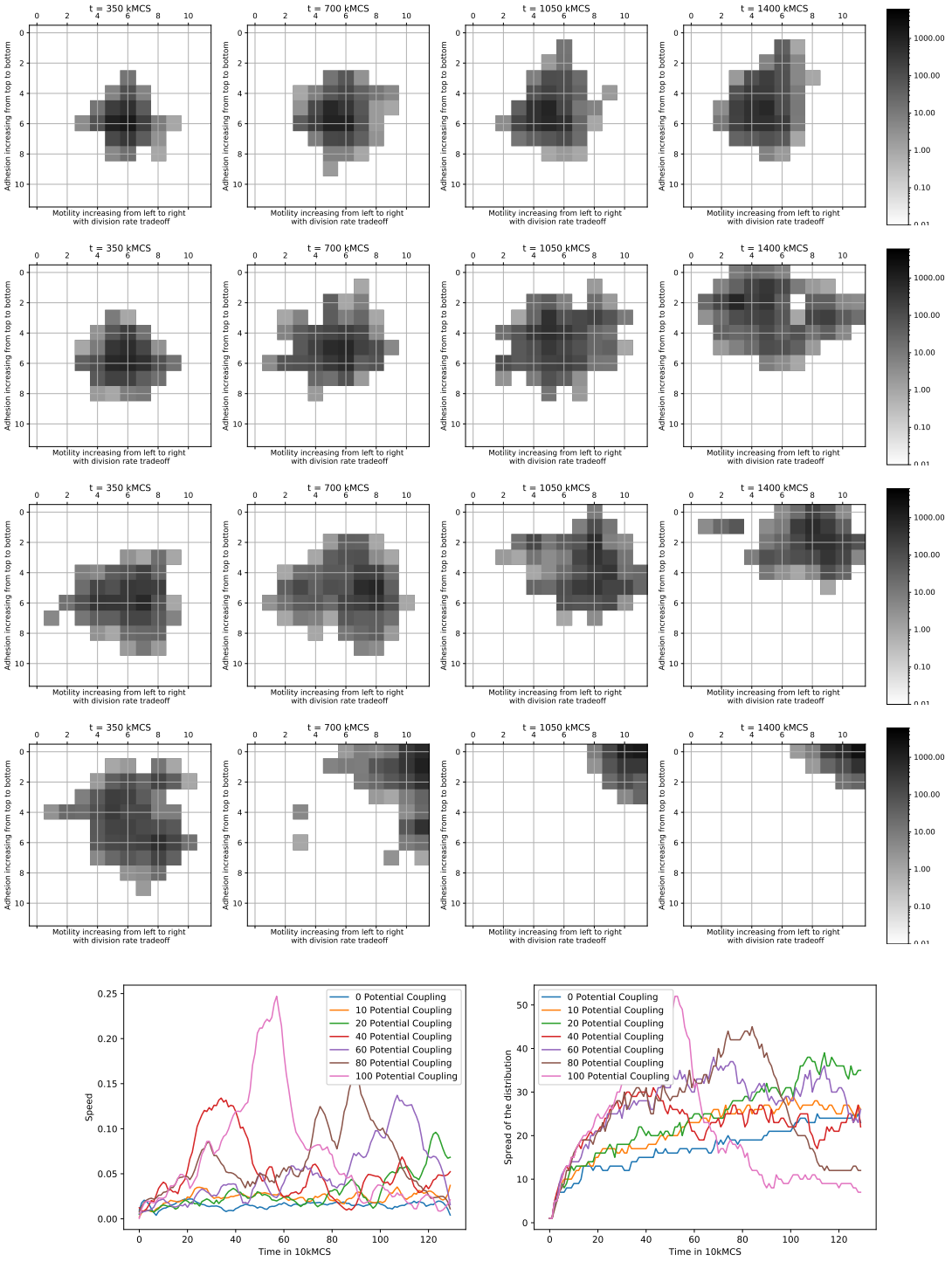

Figure 26: Potential coupling of 0, 20, 60, 100, constant mutation range (4%)

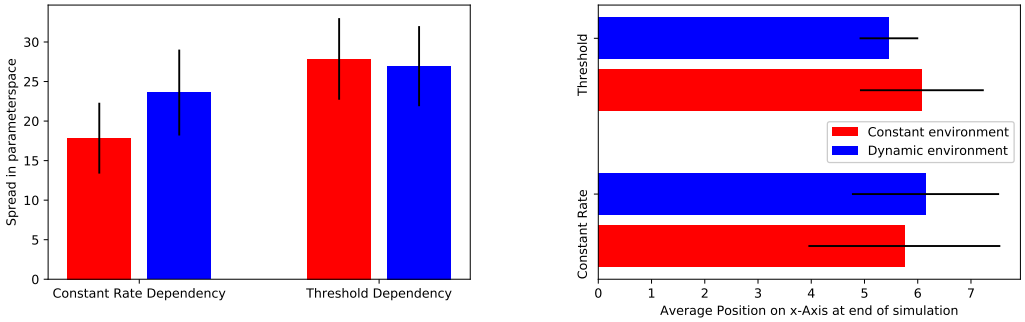

Figure 27: a) Spread in conformation space of all simulations, representing the extent of the phenotypic composition of a tumor. Four different cases are distinguished and the simulations are sorted to each: threshold based or gradual dependence of the division / death rates on the nutrient environment and for each of those cases, constant or dynamic nutrient surroundings. b) The average center of mass in the conformation space on the x (division-rate / migration) axis, at the end of the simulations (t=2200kMCS). Cases are distinguished as in a). Black lines indicate the standard deviation.

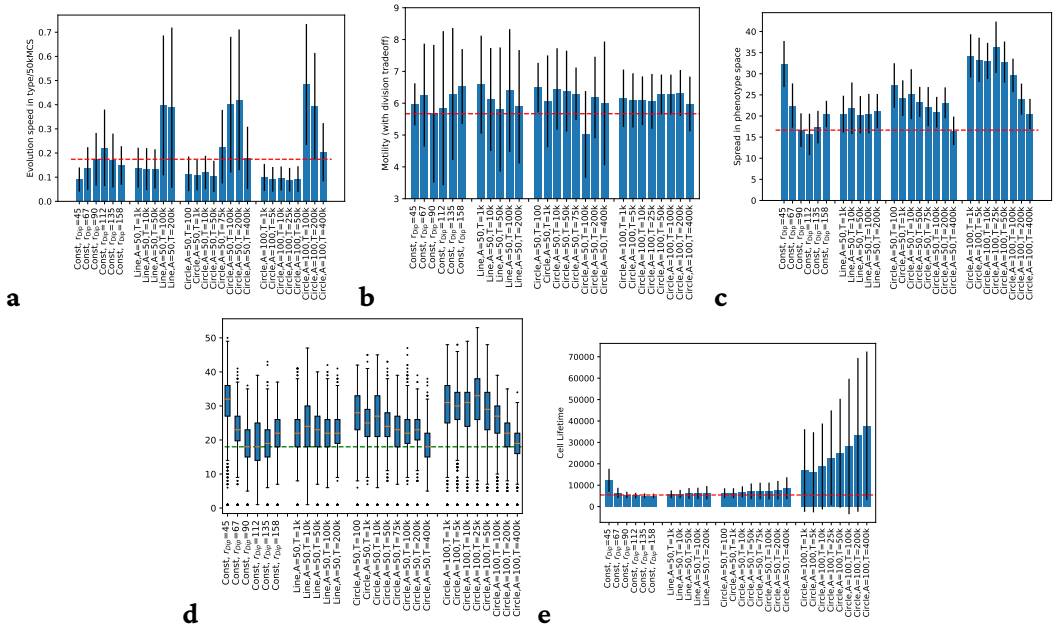

Figure 28: a) Evolution speed in conformation space. Values are averaged between 1500-3000kMCS. Black lines indicate the standard deviation. b) The average center of mass in the conformation space on the 'motility , 1/division-rate' axis. Cases are distinguished and data is collected in a). c) Spread in conformation space of all simulations, representing the extent of the phenotypic composition of a tumor. Cases are distinguished and data is collected in a). c) Spread in conformation space visualized in a box plot. e) Mean lifetime.

10.5 Threshold based parameter variations

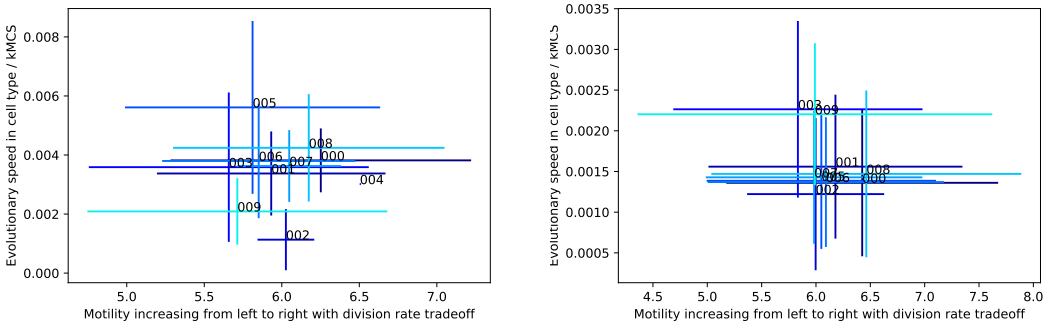

Figure 29: Quantification of evolution speed and direction (on the motility axis). Mean of 15 Simulations and indicating standard deviation. Labels explained below. Times are t = 1000kMCS and t = 2500kMCS

In order to determine the speed of the evolution of the tumor, the centroid (i.e. the center of mass) of the distribution as seen in Fig. 31. That centroid is then tracked over time and the speed through the phenotype space is determined. This speed reflects the speed of the development of the ensemble of tumor types within the tumor. Since all tumor develop into the direction of low adhesion, the evolution speed is plotted against the position of the centroid in the motility-division rate axis. To study the effects, the individual parameters have on the ensemble development, the parameters were up and down regulated in individual sets of simulations. The altered parameters in those runs are:
000: 3 % mutation rate
001: 'Standard Simulation' with 1% mutation rate and 200 system size
002: Low mutation rate
003: 10 % mutation rate
004: 400 voxel system size
005: 100 voxel system size
006: Recalculation time 50
007: Recalculation time 200
008: Recalculation time 400
009: Division independent of nutrient
Without nutrient dependency of cell division and death the evolutionary speed is elevated, similarly to an increased mutation rate of 10 %. All other simulations fall in a similar regime of evolutionary speed and final value of motility.

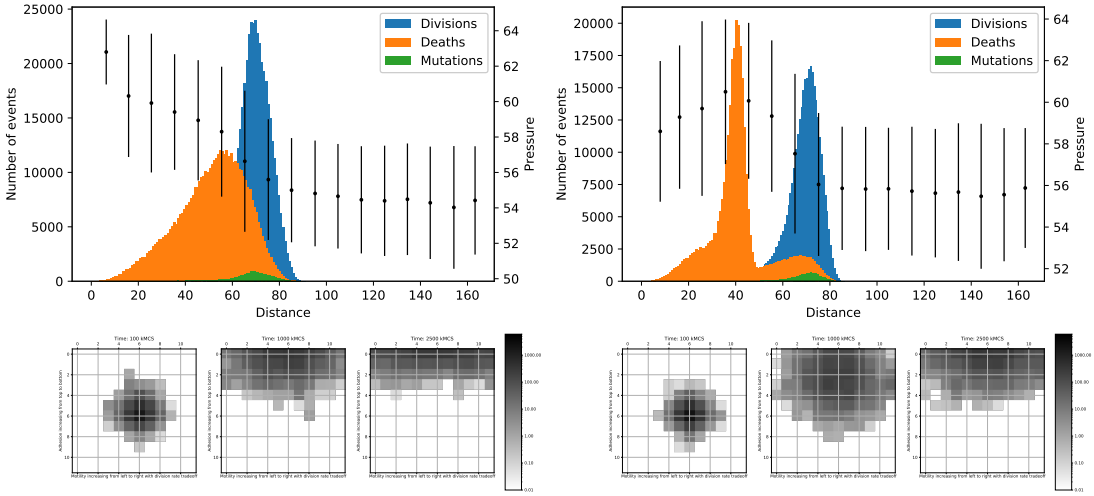

Figure 30: Non-renormalized plots of Fig. 2

10.5.1 System size

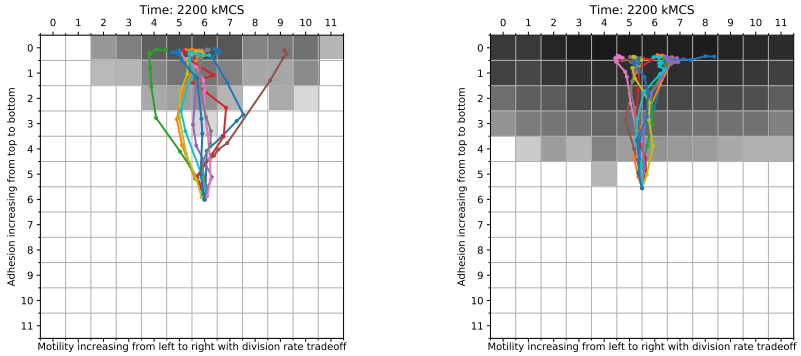

Figure 31: Scaling simulation size of a simulation with threshold dependency on nutrients. a) System size 100voxel^3 and $\approx$ 500 tumor cells b) System size 400voxel^3 and $\approx$ 28 000 tumor cells

By scaling the system size, the statistics in the tumor evolution change, since all phenotypic changes are introduced by discrete single cell events. The system size was reduced and enlarged by a factor of two. The evolution seems to not be significantly changed by the alterations of the system size. We follow that the simulations are in a regime of sufficient division and mutation events to allow sampling of the phase space.

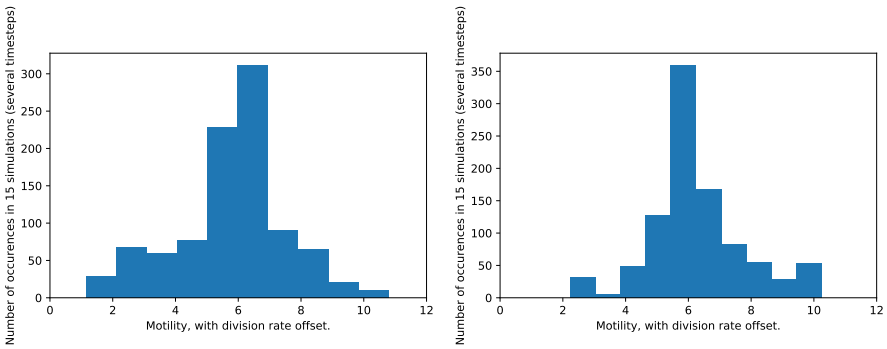

Figure 32: Distribution of x-positions of different simulations. Left, constant surrounding and right dynamic surrounding (line, A=50, T=100kMCS)

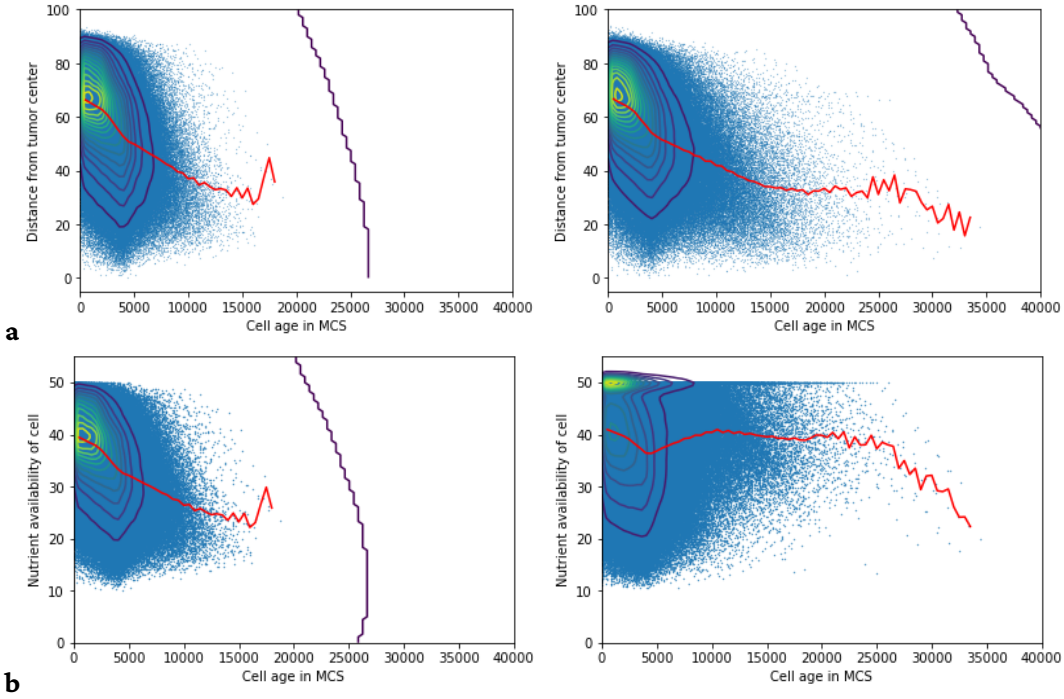

Figure 33: **Single cell properties in dynamic surrounding.** **a** Radial distance of the cells to the spheroid center in relation to the cells age, for constant surrounding (left) and dynamic surrounding (right, line, A = 50, T=100kMCS) **b** Nutrient availability of the cells in relation to the cells age, for constant surrounding (left) and dynamic surrounding (right, line, A = 50, T=100kMCS). Dots indicate single cell states, the red line describes the mean position and nutrient availability

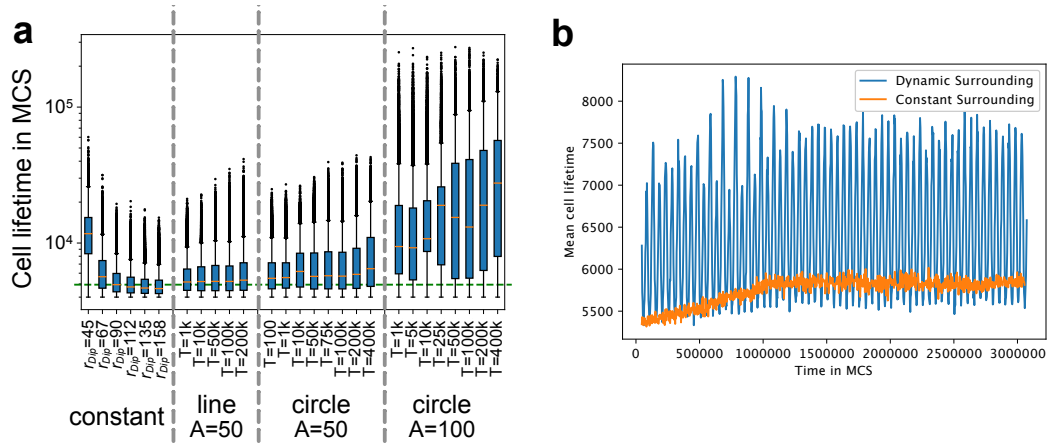

Figure 34: **Single cell properties in dynamic surrounding. a** Mean time between the last cell division and the death of cells, averaged between 150-300kMCS **d** Mean time between the last cell division and the death of all tumor cells is depicted in relation to the simulation time. The constant nutrient surrounding (orange) and the dynamic surrounding (blue, line, A = 50, T=100kMCS) are shown. The fluctuations of the dynamic surrounding lead to a periodic behavior of the mean cell lifetime.