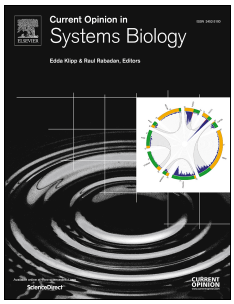
Journal Pre-proof



Systems Biology at the giga-scale: large multi-scale models of complex, heterogeneous multicellular systems

Arnau Montagud, Miguel Ponce de León, Alfonso Valencia

PII: S2452-3100(21)00079-2

DOI: https://doi.org/10.1016/j.coisb.2021.100385

Reference: COISB 100385

To appear in: *Current Opinion in Systems Biology*

Received Date: 21 May 2021

Revised Date: 16 August 2021

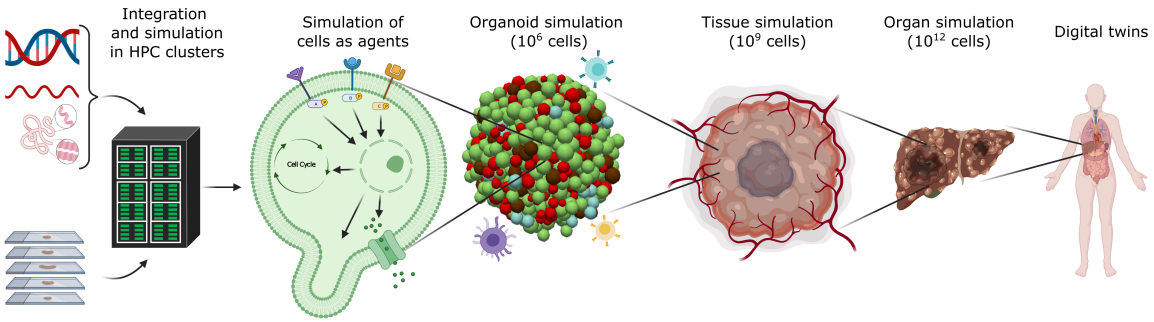
Accepted Date: 15 September 2021

Please cite this article as: Montagud A, Ponce de León M, Valencia A, Systems Biology at the giga-scale: large multi-scale models of complex, heterogeneous multicellular systems, *Current Opinion in Systems Biology*, https://doi.org/10.1016/j.coisb.2021.100385.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 The Author(s). Published by Elsevier Ltd.

Journal Pre-proof



Journal Pre-proof

Journal Pre-proof

Systems Biology at the giga-scale: large multi-scale models of complex, heterogeneous multicellular systems

Arnau Montagud[1,*], Miguel Ponce de León[1], Alfonso Valencia[1,2,*]

1, Barcelona Supercomputing Center (BSC), Barcelona, Spain

2, Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

*: corresponding authors: arnau.montagud@bsc.es, alfonso.valencia@bsc.es

Abstract:

Agent-based modelling has proven its usefulness in several biomedical projects by explaining and uncovering mechanisms in diseases. Nevertheless, the scenarios addressed in these models usually consider a small number of cells, lack cell-specific characterisation and dynamic interactions and have a simplistic environment description. Tools that enable scalable, real-sized simulations of biological systems that require complex set-ups are needed to have simulations closer to biomedical scenarios that can capture cell-to-cell heterogeneity and system-wide emerging properties. To deliver simulations at the giga-scale (10^9 cells), different tools have implemented technologies to run in high-performance computing clusters. We hereby review these efforts and detail the main areas of improvement the field needs to focus on to have simulations that are a step closer to having digital twins.

Keywords:

Agent-based simulations, large-scale simulations, high-performance computing, model exploration, cancer biology

Highlights:

- Agent-based modelling aims to explain, model and predict mechanisms in diseases.
- Large-scale agent-based modelling is needed to address real-sized, complex scenarios such as simulating tumours.
- The scaling-up of such models needs high-performance infrastructures and methods to optimise runtime and distribute memory among computing nodes.
- Such multi-scale models are complex and need tools such as model exploration frameworks to help set-up parameters and provide new hypotheses.
- Bringing together the systems biology and HPC communities is key for the development of the field.

Journal Pre-proof

Introduction:

Modelling has helped researchers gain unprecedented insights into biological mechanisms [1]. With the rise of high-throughput technologies that led to the so-called omic revolution, the use of modelling in systems biology has become an integral part of the field [2,3]. Bridging from intracellular mechanisms to tissue-level phenotype is still a problem to be solved, and multi-scale agent-based modelling is well suited to address it. Agent-based modelling is an approach to computing the potential system-level consequences of the behaviours of sets of individual agents [4]. In this framework, also known as individual-cell, individual-based or cell-based modelling, each agent has a set of behavioural rules that are relative to the environment and to other agents' behaviours. Agent-based models have been ubiquitous in a wide range of scientific disciplines, such as economics, engineering, epidemiology, urban transportation, ecology, and biology [5].

Agent-based models have a balanced combination that suits well systems biology modelling of tissues by, first, combining the study of cell populations' behaviours that have a direct correspondence to heterogeneous cellular systems encompassing a diversity of cell-cell interactions. Second, they have a cell-level granularity that allows studying genotypic and phenotypic variations at the single-cell level. Third, they can include intracellular models that can capture gene-level changes and, fourth, they explicitly describe the environment that can have a direct correspondence to laboratory experiments (Figure 1). Agent-based models have proven their usefulness in biomedicine by studying the selective pressure of the environment on tumour morphology [6], migrating cells [7–9] and their interaction with macrophages to suppress immune responses [10–12], cancer growth and how immunogenicity enable tumour cells at the outskirts to escape immune attack [13], or how dynamic regimes can counter the tumour cells' resistance to TNF [14]. Recently, some of these approaches have been directed at the study of COVID-19 [15].

Agent-based includes different modelling frameworks with different focuses

Different agent-based modelling approaches exist to model cell populations with different underlying assumptions and constraints. Still, they are usually grouped into two families: lattice-based (or on-lattice) and lattice-free (or off-lattice) (Figure 2A). For complete reviews, see references [16,17].

Lattice-based agent-based models are the ones where the environment is defined by a grid or lattice, which the agents will follow or be constrained by it. This type of modelling can be categorised by three different subtypes depending on their spatial resolution:

- Cellular automata, in which there is one agent in each lattice site (e.g., NetLogo [18], Chaste [19], and others [20–22]).
- Cellular Potts model, in which one agent can occupy several lattice sites [23] (e.g., CompuCell3D [24], Morpheus [25], EPISIM [26], and others [27]).
- A third type in which a lattice site can be occupied by more than one agent (e.g., COMETS [28,29], and others [30,31]). Lattice-gas cellular automata is a special such automaton in which the velocity of the agents is also tracked (e.g., NetLogo).

Journal Pre-proof

Lattice-free agent-based models are the ones where the environment is not defined by a grid or lattice. This type of modelling can be categorised into two main families of models, depending on whether the model's focus is the cells' boundaries or the cells' volume:

- Vertex models are boundary-tracking models where cells are modelled as polygons or polyhedra, and the focus of the model is to study the different forces applied to the vertices (e.g., Chaste, Tyssue [32], and reference [33]). They can also use Voronoi tessellation (e.g., Chaste and reference [34]).
- Overlapping spheres or centre-based models (CBM) are models where the agents are represented by spheres that can overlap among them (e.g., Chaste, Biocellion [35], CellSys [36], PhysiCell [37], Timothy [38]).

When modelling in the context of biomedical research, one of the goals is to find molecular mechanisms and interactions that may explain the observed disease phenotypes. Agent-based modelling is general and flexible enough to be the basis for constructing multi-scale models of cell populations that include intracellular models of different processes updated at different time scales when reacting to environmental conditions. This intracellular model can be a system of differential equations (as in Chaste, CompuCell3D, and EPISIM), neural networks [20], metabolic models of bacteria using dynamic FBA in BacArena [22] or COMETS [29] or Boolean models simulated stochastically using PhysiBoSS [14] (Figure 2B).

These CBM frameworks consider different time scales for simulating different processes such as molecules' diffusion, cell movement and mechanical interactions as well as cellular phenotypes (e.g., proliferation, apoptosis, migration, attachment) (Figure 2C) following a type 3 multi-scale model taxonomy from reference [39]. In addition, they consider different types of cells (e.g., immune cells, blood vessels and cancer tissue) and dynamic environments. Still, they usually take into account up to several million cells, which is far from the goal of having large multi-scale simulations of complex, heterogeneous tumours.

There are several alternatives to agent-based modelling. One of them, rule-based modelling, is a flexible space-free approach used in a wide range of fields, including biology, physics, chemistry, and computer science [40]. Notably, it has been successful in biochemistry and cell biology projects by simulating agents not as cells but as proteins and molecular complexes in well-mixed spatial compartments [41]. This approach has been combined with detailed spatial representations (e.g., the integration of BioNetGen [42] with Smoldyn [43] or VCell [44]) and is well suited to simulate a large number of biochemical reactions [45]. Interestingly, there have been efforts to use rule-based modelling to simulate cells and tissues [46].

Another alternative to agent-based modelling, hybrid discrete-continuum modelling, is useful when aiming to model large multi-cellular systems as continuum models without explicitly defining individual cells, their environment or their intrinsic scales [47]. Even though continuum models generally have fewer parameters and are less explicit than CBM, they can be useful to uncover insights into biological mechanisms [48].

These methods represent different approaches to biomedical problems and correspond to different assumptions and modelling decisions. The rest of this review will focus on CBM, a type of agent-based modelling, as we consider that they have a balanced combination of variables and assumptions that suits well tissue and cancer modelling scenarios making it

Journal Pre-proof

possible to model genes, pathways, cells and environment explicitly. In this case, as the models include genes and proteins, we can directly integrate available omics data on them and, as the cells are simulated individually, they can deliver biologically realistic simulations that have a direct correlation to clinical images.

Motivation for a distributed simulation at the giga-scale

Historically, there have been a number of efforts to build realistic surrogates of human tissue with increasing importance in personalised medicine [49]. The development of humanised patient-derived xenografts [50] and digital twins [51] are just two recent examples in biomedicine. To bring these models closer to clinical scenarios, e.g., capable of simulating the evolution of clinically detectable tumours, we still need more powerful large-scale modelling tools able to produce more realistic, real-sized simulations [52].

Giga-scale simulations should be able to address questions that current agent-based, as well as other multi-scale models hardly, can, such as those related to the interaction between processes and mechanisms modelled at different time and spatial scales [14,53,54]. A prototypical example of a very complex giga-scale simulation would be a realistic cancer response to different chemo- and immune therapies, able to explore the influence of tumour shape and environment interactions. This type of model would provide "virtual cuts" images comparable to digital pathology images of real tumours. The set-up needed to perform these simulations requires the knowledge of signalling pathways' activities, population-level biophysics, as well as the effect of the extracellular matrix and immune system on the diffusive properties of different molecules. Most of these mechanisms have been studied separately, their mechanisms have been documented in different situations, and a number of useful models build by experts are available [9,11,55,56], but the explanations of system-wide emerging properties require addressing the challenging problem of putting all of them together. The TNF dynamic regimes and cells' resistance [14] represents an initial example in this direction.

With the current state-of-the-art methods, we are unable to study the aforementioned prototypical example: a real-sized invasive tumour of an irregular shape that is being targeted by drugs arriving from a distant, single blood vessel surrounded by a complex extracellular matrix with patches of different densities that affect drug diffusion and movement of the cells. A new complex set-up will be required to test the shielding effect that the shape of the tumour might have against drugs and immune cells. Also, in regards to the invasive potential of the cells, current simulations are insufficient to describe the heterogeneous biological scenarios [57]. With this simulation, it will be possible to test the effect of an extracellular matrix with different densities together with the cellular heterogeneity on the different invasion phenotypes as described in clinical images. In addition, models that include studies of tumour evolution and genetic drifts might be able to capture behaviours related to the appearance of heterogeneity in otherwise clonal populations that might be missed in smaller simulations [58].

Another system-wide emerging property that could be studied with these giga-scale simulations is drug resistance against cancer therapies. It has been described that population-level dynamics such as competition and cell-cell variability can play a key role in the evolution of drug resistance [59]. In addition, the environment architecture has been described to affect the cells' response to drugs: conventional 2D-cultured cell line screens

Journal Pre-proof

fail in clinical studies [60] as traditional cell cultures do not recapitulate the heterogeneity and intrinsic drug sensitivity of the original tumour [61].

In our opinion, there are three main bottlenecks to achieve such a real-sized tumour simulation: biological knowledge, data availability and technical performance of the tools. Our knowledge of biological processes of different time and spatial scales and, importantly, their interactions is far from perfect. In fact, one of the goals of modelling is to help to close this knowledge gap by mapping what is known and providing hypotheses to explore the more relevant elements of what is still unknown.

In addition, the new technologies provide a unique opportunity to extend and improve these comprehensive models. In this sense, the new developments in the field of single-cell omics provide a unique opportunity to fill models with real data [62], connecting cell heterogeneity with population dynamics [63].

Finally, modelling these complex scenarios needs substantial computing resources that far exceed current desktop and university clusters. Particularly important are the limitations on the use and management of the memory for the data structures [38,64]. Thus, the scale-up of the simulations is only achievable by optimal, full-scale use of parallel systems in HPC systems [38,54]. While the access to additional biological knowledge and heterogeneous data is continuously evolving, the HPC remains the key enabling technology for the giga-scale simulations.

Paving the way for a distributed simulation at the giga-scale

In the past years, efforts have been directed at having large-scale CBM simulations by distributing the computation following a hybrid model combining MPI and OpenMP. We hereby present several centre-based agent-based modelling frameworks that have the potential to reach the billion cells' milestone.

Timothy [38,65] is an open-source, MPI-based tool that has proven to be able to simulate biological models of up to 0.3 billion cells in cancer projects [66] as well as large-scale models with nuclear-cytoplasmic oscillations of NF-κB [53].

Biocellion [35] is a flexible, discrete agent-based simulation framework that has been used to model a wide range of multi-cellular biological models, such as cell sorting, microbial patterning and a bacterial system in soil aggregate of over 26 million cells. Even though it's freely available for academic use, Biocellion's closed source can deter potential users.

Chaste [19] is an open-source, general-purpose simulation package for modelling soft tissues and discrete cell populations that can be used with MPI. This tool allows using different modelling frameworks on a given problem, enabling users to select the most appropriate one for their research and to better understand the limitations of each one of them. Chaste has been used for a wide range of projects, such as intestinal [67] or colonic crypt [68] studies.

FLAME [69] is an open-source, generic formal framework for agent-based modelling that allows parallelisation using MPI, and developers can use it to create models and tools. Implementing a Flame simulation is done by using finite-state automata with memory [70]. This tool has been adapted to be used with distributed GPUs using the OpenCL standard

[71]. Examples of uses of FLAME range from bioreactor studies [72] to immunogenic studies [73] or also epidermis modelling [74].

To study biofilm-scale bacterial population, CellModeller [75] was developed as a rigid-body method that includes models of biophysics, growth of rod-shaped cells and intra- and inter-cellular signalling using ODEs. CellModeller uses OpenCL cross-platform framework that enables the implementation of parallel software on both GPU and CPU architectures.

BioDynaMo [76] is an open-source simulation tool fully parallelised, able to offload computation to hardware accelerators and load-balance agents and their environment on available nodes. Being general-purpose, it allows simulating models from various fields by being extensible and modular, showcasing its use in neurite growth, tumour spheroid and epidemiology [76].

PhysiCell-X (Saxena et al., in preparation) is our current effort to include MPI in PhysiCell, an open-source, flexible, and lattice-free agent-based framework for multi-scale simulation of multi-cellular systems that currently only support shared-memory parallelisation. The main advantage of PhysiCell is its lightweight, very efficient and self-contained framework. Additionally, PhysiCell can be expanded using add-ons, such as PhysiBoSS [14], allowing the integration of individual Boolean models for the signalling networks embedded into each agent. PhysiCell-X expands our efforts to include distributed-memory in the solver that manages the chemicals' diffusion of PhysiCell: BioFVM-X [64].

We have hereby focused on examples of distributed CBM software. Still, other agent-based tools use distributed memory and have the potential of enabling giga-scale simulations that are not CBM [21,25,26,77–80].

Wish list for large-scale multi-scale modelling tools

Many problems addressed by modelling in biology can consider populations of cells ranging from 10^6 to 10^{11} [81], and CBM tools from the last section allow to reach only the lower bound of this range. To have simulations at the giga-scale (10^9 cells), we need novel tools to scale up efficiently by incorporating technologies from biology and computer science (Figure 3).

Nevertheless, having CBM simulations with large numbers of cells is not the only problem, as the underlying models need to be more realistic and complex. We need novel tools that can simulate real-size tumours of billions of cells, where there are explicit descriptions of intracellular mechanisms (i.e., signalling pathways, metabolism, cell cycle, cell division, etc.), each cell has its idiosyncratic properties without systematic regularity, where this idiosyncrasy changes in time, e.g., by the effect of selective pressure, and where we can study the clonal heterogeneities of cells and thus the evolution of cell strains. Such a tool should include, among others, an explicit simulation of the interactions among the different cells types, different cell shapes that better capture the tensions around a cell, and a complete description of a complex 3D environment with blood vessels (ideally with clinical-image-level detail) and cocktails of chemicals (e.g., hormones, metabolites, oxygen, etc.) that allows for the cells to modify it and be modified by it.

These complex simulations should be facilitated by enabling technologies in high-performance computing (HPC) clusters that allow having distributed memory and

parallelisation, analysis and integration of high-throughput personal molecular data in the models and, model exploration and optimisation of the models' parameters.

There are already tools that perform some of the goals described in this wish list, such as those in the previous section, but no one can do it all, let alone at a large scale.

Model exploration is needed to fit and to find optimised solutions

Agent-based models such as CBM are complex objects that can exhibit nonlinear dynamics and emergent behaviours [13]. Thus, the effect of the input parameters and the model dynamics can only be addressed by running simulations where stochasticity is an intrinsic component of the model. Furthermore, complex multi-scale CBM usually have a large number of parameters that should be specified. In the case of multi-cellular models, such parameters may involve, among others, rates of cell processes (e.g., division, death), kinetic constants of the biochemical processes (e.g., reactions, transports), mechanical constants (e.g., cell-cell interactions, migration velocities) as well as physicochemical properties of the environment (e.g., diffusion constants, extracellular matrix properties). In many cases, a model can have so many parameters that the data required for its calibration and validation is currently not available [82,83].

The analysis of complex CBM requires running large numbers of simulations applying efficient strategies to explore the parameter spaces [84,85]. This general problem is commonly referred to as model exploration (ME), a multifaceted iterative process that enables an adaptive exploration of the model's parameter space. ME strategies are commonly applied to calibration, optimisation and exploration of models and usually rely on metaheuristics (e.g., evolutionary algorithm, simulated annealing) to optimise "black box" functions or machine learning approaches to characterise the parameter space [86,87].

Model calibration is the task of estimating the unknown parameters using experimental data and was successfully used in a tumoural CBM by comparing simulated growth dynamics and spatio-temporal data extracted from experimental image analysis [88]. In model optimisation, also called simulation-based optimisation or optimisation via simulation [89], simulations are run to predict the values of the parameters that allow achieving the desired goal. For instance, model optimisation was used with a genetic algorithm to identify the set of biomedical interventions needed to minimise the number of cancer cells in a CBM of tumoural immunosurveillance [90] or to predict the drug dose schedule that minimises the tumour size while avoiding the emergence of resistant cells [91].

As discussed before, the complexity in CBM can come from a large number of cells or by having a more complex environment. As the complexity of multi-scale CBM increases, the simulations become computationally-intensive tasks, and thus ME can become challenging. In such scenarios, ME strategies should be implemented as distributed workflows to be executed in HPC clusters [13,85] in a "many-task computing" paradigm [92]. There are several frameworks for distributed ME on HPC clusters [93], including EMEWS [94], OpenMOLE [95] and ECJ [96]. Nonetheless, besides the great improvement in having HPC-ready ME strategies, the calibration and estimation of many model parameters will require novel high-throughput experimental data at the single-cell level [62,97].

Journal Pre-proof

Alternatively, there have been efforts to leverage the trove of molecular omics data available to parametrise multi-scale models by using machine learning methodologies. To accomplish this, machine learning and multi-scale modelling can complement each other on the parameter level, e.g. by identifying parameter values [98], and on the system level, e.g. by identifying system dynamics [99]. Machine learning can be useful by providing tools towards preventing overfitting or quantifying uncertainty while exploring huge parameter spaces (for a review, see reference [100]).

Perspectives

Modelling diseases has the potential of being an enabling technology in helping deliver truly personalised and precise medicine to the patients. For instance, leveraging single-cell resolution clinical images [101,102] and using personalised intracellular models with patients' data are initial steps in achieving patient-tailored drug regimes [91,103,104]. Nevertheless, researchers need to provide tools that allow for more realistic and complex simulations that include, among others, dynamically responsive environments and simulations' scales of clinically detectable tumour sizes, i.e. from the giga-scale up.

We have reviewed efforts directed at having such CBM tools that aim to model biological phenotypes in HPC clusters. Previously, we have described two improvements that would increase the relevance of these giga-scale simulations: the biological knowledge on the interaction of mechanisms at different time and spatial scales and the experimental data that could provide single-cell heterogeneity [55]. Nevertheless, some additional challenges need to be solved to have simulations that bring the field closer to having digital twins [51] (Figure 1 and 3).

One of them is to extend the cell-level CBM simulations to reach molecular dynamics and tissue simulations. There have been efforts in using CBM to reach down to molecular dynamics [54], to study the effect of molecular crowding on a virtual cytoplasm [105] or to focus on mRNA export mechanisms [106]. Likewise, CBM have been used to reach up to tissue modelling to study muscle regeneration [107] or combining finite elements with agents to study glioma invasion and vascularity [108]. These extensions that integrate models with very different granularities have their limits and challenges, and to address them, the field could learn from lessons from the whole-cell modelling community [83,97,109–111].

Another challenge is to use tools that deal with the complexity of the model set-up, identified as one of the main obstacles for agent-based modelling use [54]. In this line, ME methods have been useful to set-up, fit and explore models [90]. In addition, the use of common languages and standards formats, such as SBML [112], MultiCellDS [113], or OpenABL [114], could increase the complexity of the models by encapsulating its minor details and ease the comparison, reproducibility and benchmarking of models [115].

Finally, all these efforts need to be scalable and efficient to take full advantage of HPC clusters; thus, we need to bring closer the systems biology and HPC communities, following recent efforts such as the PerMEdCoE centre of excellence (https://permedcoe.eu/).

Acknowledgements:

The authors acknowledge the reviewers for their comments and suggestions that helped improve and clarify this manuscript.

Journal Pre-proof

This work has been partially supported by the European Commission under the INFORE project (H2020-ICT-825070) and the PerMedCoE project (H2020-ICT-951773).

References' comments:

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

Macal & North, 2010, [5] *: Fundamental primer on agent-based modelling that presents the main concepts and some applications across disciplines.

Spatarelu et al., 2019, [9] *: Review on experimental and computational methods collective cell migration in cancer and how agent-based methods have helped gain insights into mechanisms.

Reticker-Flynn et al., 2020, [11] *: Review on the field of cancer systems immunology where the authors discuss the advances done by agent-based modelling and address the future avenues of study.

Norton et al., 2019, [12] *: Review on agent-based modelling of the immune microenvironment surrounding tumours and how different modelling approaches are best suited for different spatial scales.

Metzcar et al., 2019, [16] **: Comprehensive review on agent-based modelling, their different types and examples of its use in cancer biology.

Cooper et al., 2020, [19] *: Chaste is a simulation platform that allows users to test their models in three different agent-based modelling approaches (e.g., cellular automata, vertex-based, CBM).

Cytowski et al., 2014, [38] *: Timothy allows scaling up simulations using dynamic and asymmetric domain decomposition, based on Peano-Hilbert space-filling curves, with the potential to reach the giga-scale.

Topol 2019, [49] *: Overview of the potential and limitations of different technologies, such as AI, Big Data and HPC, that have recently started to be used in biomedicine.

Björnsson et al., [51] *: The authors introduce the concept of the digital twin applied to personalised medicine and the tools and data that needs to be included to achieve it.

Nguyen et al., 2021, [63] *: Complete benchmark of different network inference methods developed for single-cell data. These single-cell networks could then be converted to individual cells' models to be used in multi-scale simulations.

Ozik et al., 2018, [94] **: Presentation of a model exploration framework compatible with HPC clusters that helps in setting-up, fitting and exploring multi-scale models.

Journal Pre-proof

References:

1. Jacob F, Monod J: **Genetic regulatory mechanisms in the synthesis of proteins**. *J Mol Biol* 1961, **3**:318–356.
2. Ideker T, Galitski T, Hood L: **A new approach to decoding life: systems biology**. *Annu Rev Genomics Hum Genet* 2001, **2**:343–372.
3. Kitano H: **Computational systems biology.** *Nature* 2002, **420**:206–10.
4. North MJ, Macal CM: *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford University Press; 2007.
5. Macal CM, North MJ: **Tutorial on agent-based modelling and simulation**. *J Simul* 2010, **4**:151–162.
6. Anderson ARA, Weaver AM, Cummings PT, Quaranta V: **Tumor Morphology and Phenotypic Evolution Driven by Selective Pressure from the Microenvironment**. *Cell* 2006, **127**:905–915.
7. Wise SM, Lowengrub JS, Frieboes HB, Cristini V: **Three-dimensional multispecies nonlinear tumor growth—I: Model and numerical method**. *J Theor Biol* 2008, **253**:524–543.
8. Frieboes HB, Jin F, Chuang Y-L, Wise SM, Lowengrub JS, Cristini V: **Three-dimensional multispecies nonlinear tumor growth—II: Tumor invasion and angiogenesis**. *J Theor Biol* 2010, **264**:1254–1278.
9. Spatarelu C-P, Zhang H, Nguyen DT, Han X, Liu R, Guo Q, Notbohm J, Fan J, Liu L, Chen Z: **Biomechanics of Collective Cell Migration in Cancer Progression: Experimental and Computational Methods**. *ACS Biomater Sci Eng* 2019, **5**:3766–3787.
10. Gatenbee C, West J, Baker AM, Guljar N, Jones L, Graham TA, Robertson-Tessi M, Anderson ARA: **Macrophage-mediated immunoediting drives ductal carcinoma evolution: Space is the game changer**. *bioRxiv* 2019, doi:10.1101/594598.
11. Reticker-Flynn NE, Engleman EG: **Cancer systems immunology**. *eLife* 2020, **9**:e53839.
12. Norton K-A, Gong C, Jamalian S, Popel AS: **Multiscale Agent-Based and Hybrid Modeling of the Tumor Immune Microenvironment**. *Processes* 2019, **7**:37.
13. Ozik J, Collier N, Wozniak JM, Macal C, Cockrell C, Friedman SH, Ghaffarizadeh A, Heiland R, An G, Macklin P: **High-throughput cancer hypothesis testing with an integrated PhysiCell-EMEWS workflow**. *BMC Bioinformatics* 2018, **19**:483.
14. Letort G, Montagud A, Stoll G, Heiland R, Barillot E, Macklin P, Zinovyev A, Calzone L: **PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling**. *Bioinformatics* 2019, doi:10.1093/bioinformatics/bty766.
15. Getz M, Wang Y, An G, Becker A, Cockrell C, Collier N, Craig M, Davis CL, Faeder J, Versypt ANF, et al.: **Rapid community-driven development of a SARS-CoV-2 tissue simulator**. *bioRxiv* 2020, doi:10.1101/2020.04.02.019075.
16. Metzcar J, Wang Y, Heiland R, Macklin P: **A Review of Cell-Based Computational Modeling in Cancer Biology**. *JCO Clin Cancer Inform* 2019, doi:10.1200/CCI.18.00069.
17. Osborne JM, Fletcher AG, Pitt-Francis JM, Maini PK, Gavaghan DJ: **Comparing individual-based approaches to modelling the self-organization of multicellular tissues**. *PLOS Comput Biol* 2017, **13**:e1005387.
18. Tisue S, Wilensky U: **Netlogo: A simple environment for modeling complexity**. In *International conference on complex systems*. . Boston, MA; 2004:16–21.
19. Cooper F, Baker R, Bernabeu M, Bordas R, Bowler L, Bueno-Orovio A, Byrne H, Carapella V, Cardone-Noott L, Cooper J, et al.: **Chaste: Cancer, Heart and Soft Tissue Environment**. *J Open Source Softw* 2020, **5**:1848.
20. Gerlee P, Anderson ARA: **A hybrid cellular automaton model of clonal evolution in cancer: The emergence of the glycolytic phenotype**. *J Theor Biol* 2008, **250**:705–722.

Journal Pre-proof

21. Salguero AG, Capel MI, Tomeu AJ: **Parallel Cellular Automaton Tumor Growth Model**. In *Practical Applications of Computational Biology and Bioinformatics, 12th International Conference*. Edited by Fdez-Riverola F, Mohamad MS, Rocha M, De Paz JF, González P. Springer International Publishing; 2019:175–182.
22. Bauer E, Zimmermann J, Baldini F, Thiele I, Kaleta C: **BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities**. *PLOS Comput Biol* 2017, **13**:e1005544.
23. Anderson A, Chaplain M, Rejniak K: *Single-cell-based models in biology and medicine*. Springer Science & Business Media; 2007.
24. Swat MH, Thomas GL, Belmonte JM, Shirinifard A, Hmeljak D, Glazier JA: **Multi-Scale Modeling of Tissues Using CompuCell3D**. In *Methods in Cell Biology*. . Elsevier; 2012:325–366.
25. Starruß J, Back W de, Brusch L, Deutsch A: **Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology**. *Bioinformatics* 2014, **30**:1331–1332.
26. Sütterlin T, Kolb C, Dickhaus H, Jäger D, Grabe N: **Bridging the scales: semantic integration of quantitative SBML in graphical multi-cellular models and simulations with EPISIM and COPASI**. *Bioinformatics* 2013, **29**:223–229.
27. Norton K-A, Wallace T, Pandey NB, Popel AS: **An agent-based model of triple-negative breast cancer: the interplay between chemokine receptor CCR5 expression, cancer stem cells, and hypoxia**. *BMC Syst Biol* 2017, **11**:68.
28. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N, Mehta P, et al.: **Metabolic Resource Allocation in Individual Microbes Determines Ecosystem Interactions and Spatial Dynamics**. *Cell Rep* 2014, **7**:1104–1115.
29. Dukovski I, Bajić D, Chacón JM, Quintin M, Vila JC, Sulheim S, Pacheco AR, Bernstein DB, Rieh WJ, Korolev KS, et al.: **Computation Of Microbial Ecosystems in Time and Space (COMETS): An open source collaborative platform for modeling ecosystems metabolism**. *ArXiv200901734 Q-Bio* 2020,
30. Ruan X, Zhou J, Tu H, Jin Z, Shi X: **An improved cellular automaton with axis information for microscopic traffic simulation**. *Transp Res Part C Emerg Technol* 2017, **78**:63–77.
31. Radszuweit M, Block M, Hengstler JG, Schöll E, Drasdo D: **Comparing the growth kinetics of cell populations in two and three dimensions**. *Phys Rev E* 2009, **79**:051907.
32. Theis S, Suzanne M, Gay G: **Tyssue: an epithelium simulation library**. *J Open Source Softw* 2021, **6**:2973.
33. Fletcher AG, Osterfield M, Baker RE, Shvartsman SY: **Vertex Models of Epithelial Morphogenesis**. *Biophys J* 2014, **106**:2291–2304.
34. González- Valverde I, García- Aznar JM: **A hybrid computational model to explore the topological characteristics of epithelial tissues**. *Int J Numer Methods Biomed Eng* 2017, **33**:e2877.
35. Kang S, Kahan S, McDermott J, Flann N, Shmulevich I: **Biocellion: accelerating computer simulation of multicellular biological system models**. *Bioinforma Oxf Engl* 2014, **30**:3101–3108.
36. Hoehme S, Drasdo D: **A cell-based simulation software for multi-cellular systems**. *Bioinformatics* 2010, **26**:2641–2642.
37. Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P: **PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems**. *PLOS Comput Biol* 2018, **14**:e1005991.
38. Cytowski M, Szymanska Z: **Large-Scale Parallel Simulations of 3D Cell Colony Dynamics**. *Comput Sci Eng* 2014, **16**:86–95.
39. Walpole J, Papin JA, Peirce SM: **Multiscale computational models of complex biological systems**. *Annu Rev Biomed Eng* 2013, **15**:137–154.
40. Chylek LA, Harris LA, Faeder JR, Hlavacek WS: **Modeling for (physical) biologists:**

an introduction to the rule-based approach. *Phys Biol* 2015, **12**:045007.

41. Boutillier P, Maasha M, Li X, Medina-Abarca HF, Krivine J, Feret J, Cristescu I, Forbes AG, Fontana W: **The Kappa platform for rule-based modeling**. *Bioinformatics* 2018, **34**:i583–i592.
42. Harris LA, Hogg JS, Tapia J-J, Sekar JAP, Gupta S, Korsunsky I, Arora A, Barua D, Sheehan RP, Faeder JR: **BioNetGen 2.2: advances in rule-based modeling**. *Bioinformatics* 2016, **32**:3366–3368.
43. Andrews SS: **Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface**. *Bioinformatics* 2017, **33**:710–717.
44. Blinov ML, Schaff JC, Vasilescu D, Moraru II, Bloom JE, Loew LM: **Compartmental and Spatial Rule-Based Modeling with Virtual Cell**. *Biophys J* 2017, **113**:1365–1372.
45. Santibáñez R, Garrido D, Martin AJM: **Atlas: automatic modeling of regulation of bacterial gene expression and metabolism using rule-based languages**. *Bioinformatics* 2020, **36**:5473–5480.
46. Maus C, Rybacki S, Uhrmacher AM: **Rule-based multi-level modeling of cell biological systems**. *BMC Syst Biol* 2011, **5**:166.
47. Lowengrub JS, Frieboes HB, Jin F, Chuang Y-L, Li X, Macklin P, Wise SM, Cristini V: **Nonlinear modelling of cancer: bridging the gap between cells and tumours**. *Nonlinearity* 2010, **23**:R1–R9.
48. Sciumè G, Santagiuliana R, Ferrari M, Decuzzi P, Schrefler BA: **A tumor growth model with deformable ECM**. *Phys Biol* 2014, **11**:065004.
49. Topol EJ: **High-performance medicine: the convergence of human and artificial intelligence**. *Nat Med* 2019, **25**:44.
50. Zhao Y, Shuen TWH, Toh TB, Chan XY, Liu M, Tan SY, Fan Y, Yang H, Lyer SG, Bonney GK, et al.: **Development of a new patient-derived xenograft humanised mouse model to study human-specific tumour microenvironment and immunotherapy**. *Gut* 2018, **67**:1845–1854.
51. Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, Jörnsten R, Lee EJ, Li X, Lilja S, et al.: **Digital twins to personalize medicine**. *Genome Med* 2019, **12**:4.
52. Jiao Y, Torquato S: **Emergent Behaviors from a Cellular Automaton Model for Invasive Tumor Growth in Heterogeneous Microenvironments**. *PLOS Comput Biol* 2011, **7**:e1002314.
53. Szymańska Z, Cytowski M, Mitchell E, Macnamara CK, Chaplain MAJ: **Computational Modelling of Cancer Development and Growth: Modelling at Multiple Scales and Multiscale Modelling**. *Bull Math Biol* 2018, **80**:1366–1403.
54. Soheilypour M, Mofrad MRK: **Agent-Based Modeling in Molecular Systems Biology**. *BioEssays* 2018, **40**:1800020.
55. Cohen DPA, Martignetti L, Robine S, Barillot E, Zinovyev A, Calzone L: **Mathematical Modelling of Molecular Pathways Enabling Tumour Cell Invasion and Migration**. *PLoS Comput Biol* 2015, **11**:e1004571.
56. Calzone L, Tournier L, Fourquet S, Thieffry D, Zhivotovsky B, Barillot E, Zinovyev A: **Mathematical modelling of cell-fate decision in response to death receptor engagement**. *PLoS Comput Biol* 2010, **6**:e1000702.
57. Friedl P, Alexander S: **Cancer Invasion and the Microenvironment: Plasticity and Reciprocity**. *Cell* 2011, **147**:992–1009.
58. Venkatesan S, Swanton C: **Tumor Evolutionary Principles: How Intratumor Heterogeneity Influences Cancer Treatment and Outcome**. *Am Soc Clin Oncol Educ Book* 2016, doi:10.1200/EDBK_158930.
59. Kim E, Kim J-Y, Smith MA, Haura EB, Anderson ARA: **Cell signaling heterogeneity is modulated by both cell-intrinsic and -extrinsic mechanisms: An integrated approach to understanding targeted therapy**. *PLoS Biol* 2018, **16**:e2002930.
60. Horvath P, Aulner N, Bickle M, Davies AM, Nery ED, Ebner D, Montoya MC, Östling

Journal Pre-proof

P, Pietiäinen V, Price LS, et al.: **Screening out irrelevant cell-based models of disease**. *Nat Rev Drug Discov* 2016, **15**:751–769.

61. Jabs J, Zickgraf FM, Park J, Wagner S, Jiang X, Jechow K, Kleinheinz K, Toprak UH, Schneider MA, Meister M, et al.: **Screening drug effects in patient-derived cancer cells links organoid responses to genome alterations**. *Mol Syst Biol* 2017, **13**:955.
62. Kulkarni A, Anderson AG, Merullo DP, Konopka G: **Beyond bulk: a review of single cell transcriptomics methodologies and applications**. *Curr Opin Biotechnol* 2019, **58**:129–136.
63. Nguyen H, Tran D, Tran B, Pehlivan B, Nguyen T: **A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data**. *Brief Bioinform* 2021, **22**.
64. Saxena G, Ponce De Leon M, Montagud A, Vicente Dorca D, Valencia A: **BioFVM-X: An MPI+OpenMP 3-D Simulator for Biological Systems**. In *Computational Methods in Systems Biology*. . Springer International Publishing; 2021:5–18. In press.
65. Cytowski M, Szymanska Z: **Large-Scale Parallel Simulations of 3D Cell Colony Dynamics: The Cellular Environment**. *Comput Sci Eng* 2015, **17**:44–48.
66. Cytowski M, Szymańska Z, Umiński P, Andrejczuk G, Raszkowski K: **Implementation of an Agent-Based Parallel Tissue Modelling Framework for the Intel MIC Architecture**. *Sci Program* 2017, **2017**:e8721612.
67. Dunn S-J, Näthke IS, Osborne JM: **Computational Models Reveal a Passive Mechanism for Cell Migration in the Crypt**. *PLOS ONE* 2013, **8**:e80516.
68. Dunn S-J, Fletcher AG, Chapman SJ, Gavaghan DJ, Osborne JM: **Modelling the role of the basement membrane beneath a growing epithelial monolayer**. *J Theor Biol* 2012, **298**:82–91.
69. Coakley S, Gheorghe M, Holcombe M, Chin S, Worth D, Greenough C: **Exploitation of High Performance Computing in the FLAME Agent-Based Simulation Framework**. In *2012 IEEE 14th International Conference on High Performance Computing and Communication 2012 IEEE 9th International Conference on Embedded Software and Systems*. . 2012:538–545.
70. Coakley S, Smallwood R, Holcombe M: **Using x-machines as a formal basis for describing agents in agent-based modelling**. *Simul Ser* 2006, **38**:33.
71. Richmond P, Walker D, Coakley S, Romano D: **High performance cellular level agent-based simulation with FLAME for the GPU**. *Brief Bioinform* 2010, **11**:334–347.
72. Kaul H, Cui Z, Ventikos Y: **A Multi-Paradigm Modeling Framework to Simulate Dynamic Reciprocity in a Bioreactor**. *PLOS ONE* 2013, **8**:e59671.
73. Kabiri Chimeh M, Heywood P, Pennisi M, Pappalardo F, Richmond P: **Parallelisation strategies for agent based simulation of immune systems**. *BMC Bioinformatics* 2019, **20**:579.
74. Li X, Upadhyay AK, Bullock AJ, Dicolandrea T, Xu J, Binder RL, Robinson MK, Finlay DR, Mills KJ, Bascom CC, et al.: **Skin Stem Cell Hypotheses and Long Term Clone Survival – Explored Using Agent-based Modelling**. *Sci Rep* 2013, **3**:1904.
75. Rudge TJ, Steiner PJ, Phillips A, Haseloff J: **Computational Modeling of Synthetic Microbial Biofilms**. *ACS Synth Biol* 2012, **1**:345–352.
76. Breitwieser L, Hesam A, Montigny J de, Vavourakis V, Iosif A, Jennings J, Kaiser M, Manca M, Meglio AD, Al-Ars Z, et al.: **BioDynaMo: a general platform for scalable agent-based simulation**. *bioRxiv* 2021, doi:10.1101/2020.06.08.139949.
77. Berghoff M, Rosenbauer J, Hoffmann F, Schug A: **Cells in Silico – introducing a high-performance framework for large-scale tissue modeling**. *BMC Bioinformatics* 2020, **21**:436.
78. Collier N, North M: **Repast HPC: A Platform for Large-Scale Agent-Based Modeling**. In *Large-Scale Computing*. . John Wiley & Sons, Ltd; 2011:81–109.
79. Rousset A, Herrmann B, Lang C, Philippe L: **A Survey on Parallel and Distributed Multi-Agent Systems**. In *Euro-Par 2014: Parallel Processing Workshops*. Edited by Lopes L, Žilinskas J, Costan A, Cascella RG, Kecskemeti G, Jeannot E, Cannataro M,

Journal Pre-proof

Ricci L, Benkner S, Petit S, et al. Springer International Publishing; 2014:371–382.

80. Cordasco G, De Chiara R, Mancuso A, Mazzeo D, Scarano V, Spagnuolo C: **Bringing together efficiency and effectiveness in distributed simulations: The experience with D-Mason**. *SIMULATION* 2013, **89**:1236–1253.

81. Byrne H, Drasdo D: **Individual-based and continuum models of growing cell populations: a comparison**. *J Math Biol* 2008, **58**:657.

82. Aguilar B, Gibbs DL, Reiss DJ, McConnell M, Danziger SA, Dervan A, Trotter M, Bassett D, Hershberg R, Ratushny AV, et al.: **A generalizable data-driven multicellular model of pancreatic ductal adenocarcinoma**. *GigaScience* 2020, **9**.

83. Szigeti B, Roth YD, Sekar JAP, Goldberg AP, Pochiraju SC, Karr JR: **A blueprint for human whole-cell modeling**. *Curr Opin Syst Biol* 2018, **7**:8–15.

84. Ozik J, Collier NT, Wozniak JM, Spagnuolo C: **From desktop to large-scale model exploration with SWIFT/T**. *Proc Winter Simul Conf Winter Simul Conf* 2016, **2016**:206–220.

85. Jagiella N, Rickert D, Theis FJ, Hasenauer J: **Parallelization and High-Performance Computing Enables Automated Statistical Inference of Multi-scale Models**. *Cell Syst* 2017, **4**:194-206.e9.

86. Rodriguez-Fernandez M, Egea JA, Banga JR: **Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems**. *BMC Bioinformatics* 2006, **7**:483.

87. Audet C: **A Survey on Direct Search Methods for Blackbox Optimization and Their Applications**. In *Mathematics Without Boundaries*. Edited by Pardalos PM, Rassias TM. Springer New York; 2014:31–56.

88. Jagiella N, Müller B, Müller M, Vignon-Clementel IE, Drasdo D: **Inferring Growth Control Mechanisms in Growing Multi-cellular Spheroids of NSCLC Cells from Spatial-Temporal Image Data**. *PLOS Comput Biol* 2016, **12**:e1004412.

89. Tekin E, Sabuncuoglu I: **Simulation optimization: A comprehensive review on theory and applications**. *IIE Trans* 2004, **36**:1067–1081.

90. Ozik J, Collier N, Heiland R, An G, Macklin P: **Learning-accelerated discovery of immune-tumour interactions**. *Mol Syst Des Eng* 2019, **4**:747–760.

91. Akasiadis C, Ponce-de-Leon M, Montagud A, Michelioudakis E, Atsidakou A, Alevizos E, Artikis A, Valencia A, Paliouras G: **Parallel Model Exploration for Tumor Treatment Simulations**. *ArXiv210314132 Cs Q-Bio* 2021,

92. Raicu I, Foster IT, Yong Zhao: **Many-task computing for grids and supercomputers**. In *2008 Workshop on Many-Task Computing on Grids and Supercomputers*. . 2008:1–11.

93. Carillo M, Cordasco G, Serrapica F, Scarano V, Spagnuolo C, Szufel P: **Distributed simulation optimization and parameter exploration framework for the cloud**. *Simul Model Pract Theory* 2018, **83**:108–123.

94. Ozik J, Collier NT, Wozniak JM, Macal CM, An G: **Extreme-Scale Dynamic Exploration of a Distributed Agent-Based Model With the EMEWS Framework**. *IEEE Trans Comput Soc Syst* 2018, **5**:884–895.

95. Reuillon R, Leclaire M, Rey-Coyrehourcq S: **OpenMOLE, a workflow engine specifically tailored for the distributed exploration of simulation models**. *Future Gener Comput Syst* 2013, **29**:1981–1990.

96. Scott EO, Luke S: **ECJ at 20: toward a general metaheuristics toolkit**. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. . ACM; 2019:1391–1398.

97. Babtie AC, Stumpf MPH: **How to deal with parameters for whole-cell modelling**. *J R Soc Interface* 2017, **14**:20170237.

98. Perdikaris P, Karniadakis GE: **Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond**. *J R Soc Interface* 2016, **13**:20151107.

99. Sahli Costabal F, Perdikaris P, Kuhl E, Hurtado DE: **Multi-fidelity classification using Gaussian processes: Accelerating the prediction of large-scale**

computational models. *Comput Methods Appl Mech Eng* 2019, **357**:112602.

100. Alber M, Tepole AB, Cannon WR, De S, Dura-Bernal S, Garikipati K, Karniadakis G, Lytton WW, Perdikaris P, Petzold L, et al.: **Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences**. *Npj Digit Med* 2019, **2**:1–11.
101. Kiemen A, Braxton AM, Grahn MP, Han KS, Babu JM, Reichel R, Amoa F, Hong S-M, Cornish TC, Thompson ED, et al.: **In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution**. *bioRxiv* 2020, doi:10.1101/2020.12.08.416909.
102. Lomakin A, Svedlund J, Strell C, Gataric M, Shmatko A, Park JS, Ju YS, Dentro S, Kleshchevnikov V, Vaskivskyi V, et al.: **Spatial genomics maps the structure, character and evolution of cancer clones**. *bioRxiv* 2021, doi:10.1101/2021.04.16.439912.
103. Béal J, Montagud A, Traynard P, Barillot E, Calzone L: **Personalization of logical models with multi-omics data allows clinical stratification of patients**. *Front Physiol* 2019, **9**:1965.
104. Liu A, Trairatphisan P, Gjerga E, Didangelos A, Barratt J, Saez-Rodriguez J: **From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL**. *Npj Syst Biol Appl* 2019, **5**:1–10.
105. Ridgway D, Broderick G, Lopez-Campistrous A, Ru'aini M, Winter P, Hamilton M, Boulanger P, Kovalenko A, Ellison MJ: **Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm**. *Biophys J* 2008, **94**:3748–3759.
106. Azimi M, Bulat E, Weis K, Mofrad MRK: **An agent-based model for mRNA export through the nuclear pore complex**. *Mol Biol Cell* 2014, **25**:3643–3653.
107. Westman AM, Peirce SM, Christ GJ, Blemker SS: **Agent-based model provides insight into the mechanisms behind failed regeneration following volumetric muscle loss injury**. *PLOS Comput Biol* 2021, **17**:e1008937.
108. de Montigny J, Iosif A, Breitwieser L, Manca M, Bauer R, Vavourakis V: **An in silico hybrid continuum-/agent-based procedure to modelling cancer development: Interrogating the interplay amongst glioma invasion, vascularity and necrosis**. *Methods* 2021, **185**:94–104.
109. Goldberg AP, Szigeti B, Chew YH, Sekar JA, Roth YD, Karr JR: **Emerging whole-cell modeling principles and methods**. *Curr Opin Biotechnol* 2018, **51**:97–102.
110. Stumpf MPH: **Statistical and computational challenges for whole cell modelling**. *Curr Opin Syst Biol* 2021, doi:10.1016/j.coisb.2021.04.005.
111. Karr JR, Takahashi K, Funahashi A: **The principles of whole-cell modeling**. *Curr Opin Microbiol* 2015, **27**:18–24.
112. Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, Bergmann FT, Finney A, Gillespie CS, Helikar T, et al.: **SBML Level 3: an extensible format for the exchange and reuse of biological models**. *Mol Syst Biol* 2020, **16**:e9110.
113. Friedman SH, Anderson ARA, Bortz DM, Fletcher AG, Frieboes HB, Ghaffarizadeh A, Grimes DR, Hawkins-Daarud A, Hoehme S, Juarez EF, et al.: **MultiCellDS: a community-developed standard for curating microenvironment-dependent multicellular data**. *bioRxiv* 2016, doi:10.1101/090456.
114. Cosenza B, Popov N, Juurlink B, Richmond P, Chimeh MK, Spagnuolo C, Cordasco G, Scarano V: **OpenABL: A Domain-Specific Language for Parallel and Distributed Agent-Based Simulations**. In *Euro-Par 2018: Parallel Processing*. Edited by Aldinucci M, Padovani L, Torquati M. Springer International Publishing; 2018:505–518.
115. Cosenza B, Popov N, Juurlink B, Richmond P, Chimeh MK, Spagnuolo C, Cordasco G, Scarano V: **Easy and efficient agent-based simulations with the OpenABL language and compiler**. *Future Gener Comput Syst* 2021, **116**:61–75.

Figure legends:

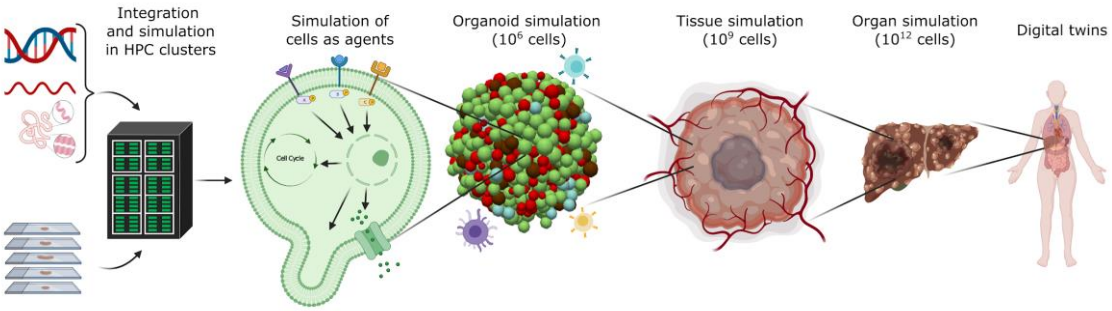
Figure 1: Overview of the motivation to build large-scale CBM in biomedicine. We make use of HPC clusters to integrate different types of omics and clinical data in agent-based models that are simulated with a focus on the cells' behaviours. These agents are studied as cells' populations where their emerging properties can be mechanistically inspected. The scale of these simulations currently ranges on the 10^6 cells but needs to reach 10^{12} cells and have complex environments if they are to simulate realistic, real-size tumours. These giga-scale simulations need to be enabled by HPC-optimised tools. Created with BioRender.com.

Figure 2: Multi-scale agent-based models in systems biology. A) Different agent-based approaches can be used for the modelling and simulation of cell populations, including, from top left to bottom right, lattice-based cellular automata, cellular Potts model, overlapping spheres and vertex models (see [17] for a review). B) Different models of intracellular models (e.g., metabolism, cell signalling) can be embedded into individual cell agents allowing to describe the problem at several different spatial scales and to consider heterogeneous populations. C) Schema of the multi-scale nature of the simulation with time intervals at which different processes are updated during a simulation. These are the diffusion of molecules and biochemical reactions (queried every 0.01 minute in [37]), the mechanical interactions (queried every 0.1 minute) and different cellular processes like commitment to cell division or apoptosis (queried every 6 minutes).

Figure 3: Technologies and properties that will enable having large-scale agent-based simulations. The field needs advances in four different areas to have simulations at the giga- scale: personalised intracellular mechanisms need to be included in the CBM (purple), these simulations need to allow for the description of a complex 3D environment (grey), the agents and the environment need to have dynamical biophysical properties (green), and all these tools need to be optimised for HPC clusters (blue). Created with BioRender.com.

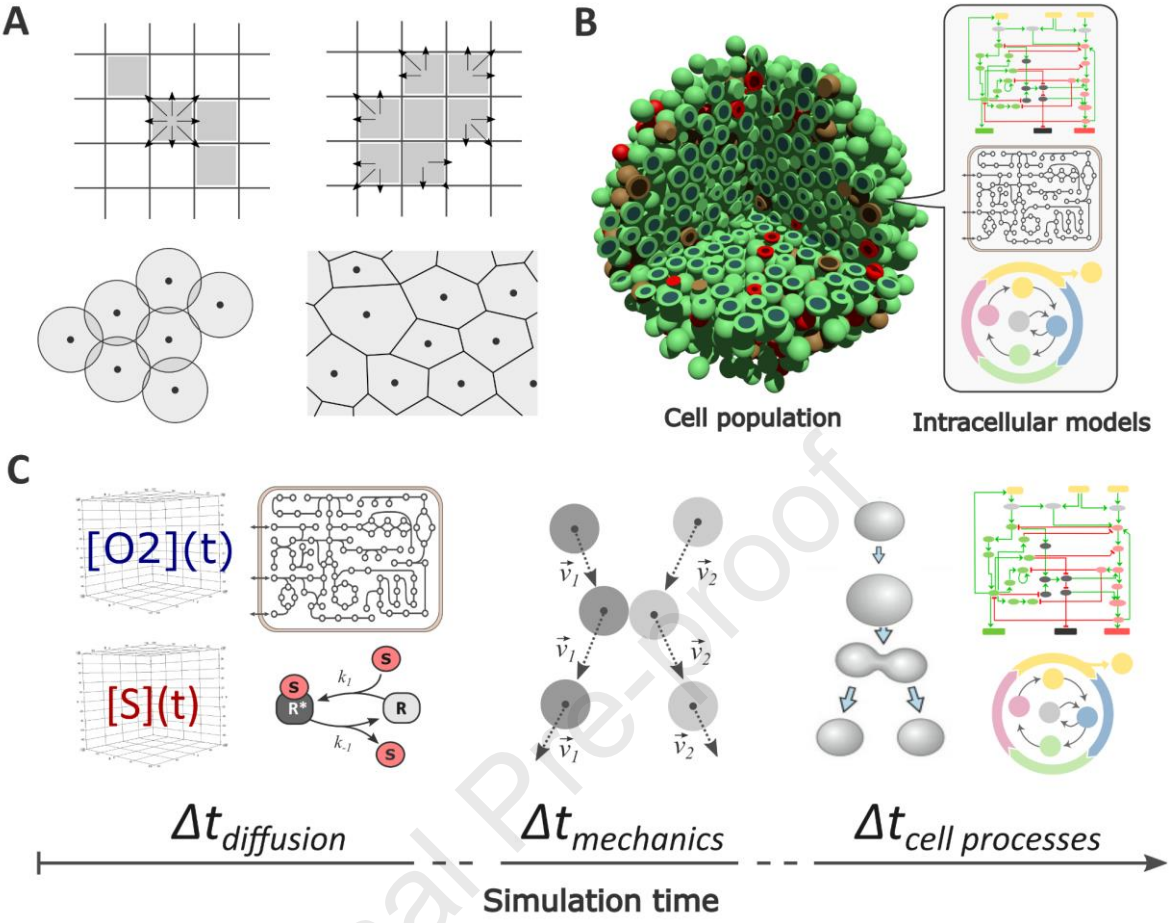
Journal Pre-proof

Figure 1:



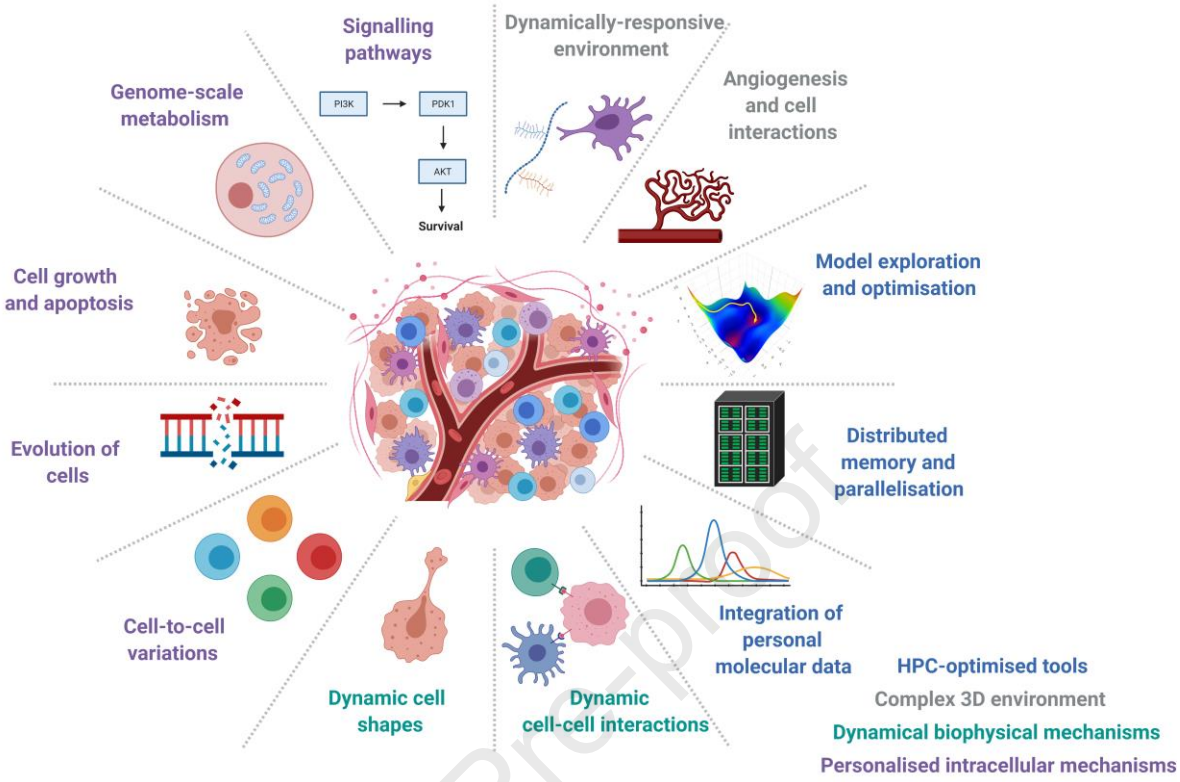
Journal Pre-proof

Figure 2:



Journal Pre-proof

Figure 3:



Journal Pre-proof

References' comments:

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

Macal & North, 2010, [5] *: Fundamental primer on agent-based modelling that presents the main concepts and some applications across disciplines.

Spatarelu et al., 2019, [9] *: Review on experimental and computational methods collective cell migration in cancer and how agent-based methods have helped gain insights into mechanisms.

Reticker-Flynn et al., 2020, [11] *: Review on the field of cancer systems immunology where the authors discuss the advances done by agent-based modelling and address the future avenues of study.

Norton et al., 2019, [12] *: Review on agent-based modelling of the immune microenvironment surrounding tumours and how different modelling approaches are best suited for different spatial scales.

Metzcar et al., 2019, [16] **: Comprehensive review on agent-based modelling, their different types and examples of its use in cancer biology.

Cooper et al., 2020, [19] *: Chaste is a simulation platform that allows users to test their models in three different agent-based modelling approaches (e.g., cellular automata, vertex-based, CBM).

Cytowski et al., 2014, [38] *: Timothy allows scaling up simulations using dynamic and asymmetric domain decomposition, based on Peano-Hilbert space-filling curves, with the potential to reach the giga-scale.

Topol 2019, [49] *: Overview of the potential and limitations of different technologies, such as AI, Big Data and HPC, that have recently started to be used in biomedicine.

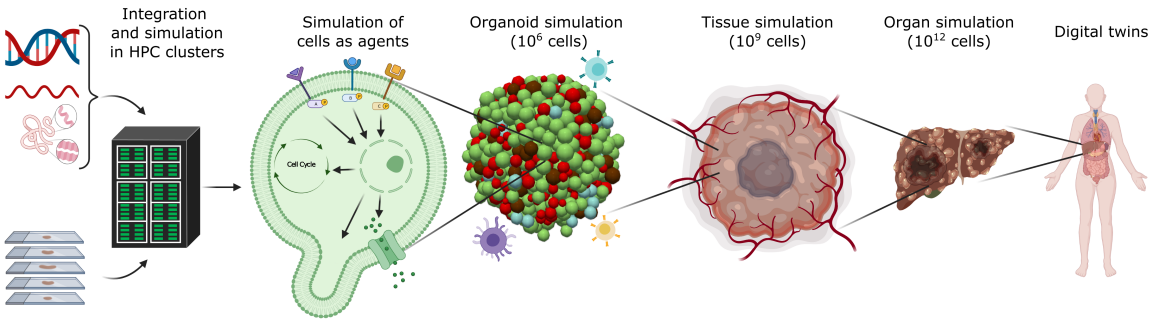
Björnsson et al., [51] *: The authors introduce the concept of the digital twin applied to personalised medicine and the tools and data that needs to be included to achieve it.

Nguyen et al., 2021, [63] *: Complete benchmark of different network inference methods developed for single-cell data. These single-cell networks could then be converted to individual cells' models to be used in multi-scale simulations.

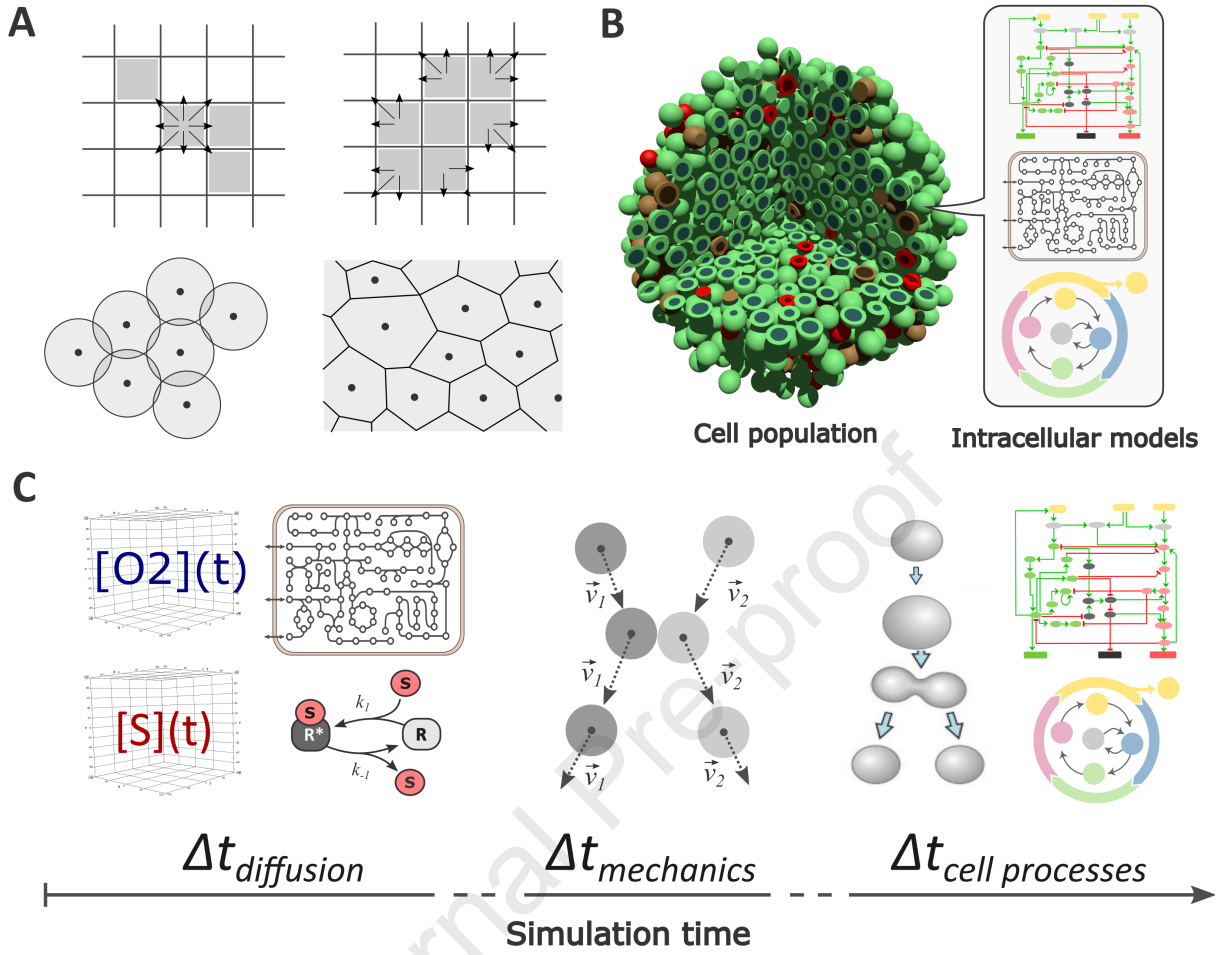
Ozik et al., 2018, [94] **: Presentation of a model exploration framework compatible with HPC clusters that helps in setting-up, fitting and exploring multi-scale models.

Journal Pre-proof

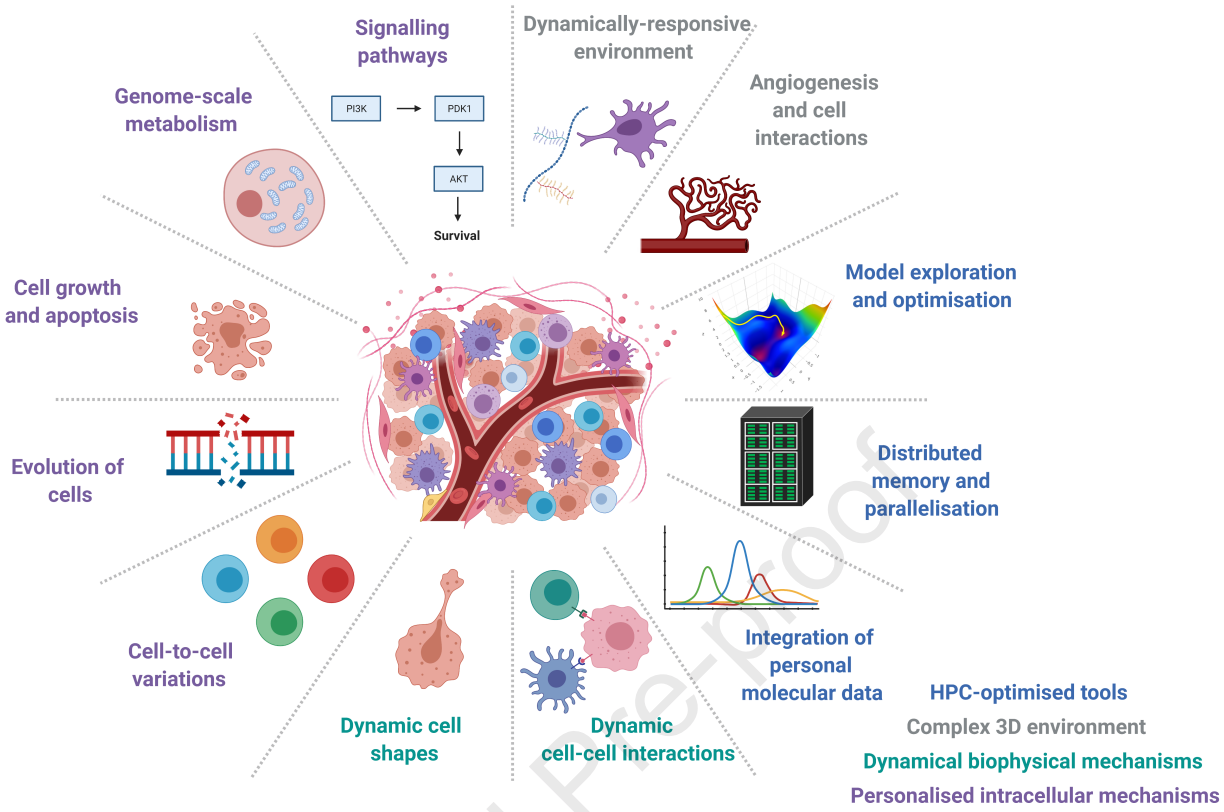
Journal Pre-proof



Journal Pre-proof



Journal Pre-proof



Journal Pre-proof

Journal Pre-proof

Highlights:

- Agent-based modelling aims to explain, model and predict mechanisms in diseases.
- Large-scale agent-based modelling is needed to address real-sized, complex scenarios such as simulating tumours.
- The scaling-up of such models needs high-performance infrastructures and methods to optimise runtime and distribute memory among computing nodes.
- Such multi-scale models are complex and need tools such as model exploration frameworks to help set-up parameters and provide new hypotheses.
- Bringing together the systems biology and HPC communities is key for the development of the field.

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof