Tatra Mt. Math. Publ. **93** (2026), 1–56
DOI: 10.2478/tmmp-2026-0011

TATRA MOUNTAINS
Mathematical Publications

MATHEMATICAL ONCOLOGY: MODELS, METHODS, AND CLINICAL APPLICATIONS

Peter Zigman

Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovakia

Abstract. Mathematical oncology integrates mechanistic modeling, experimental biology, clinical medicine, and data science to improve cancer detection, treatment planning, and understanding of tumor dynamics. This work provides a systematic review of the field, tracing its historical development from early population-level growth models and the log-kill hypothesis to contemporary spatial, evolutionary, multiscale, and data-driven approaches, supported by a large-scale bibliometric analysis of the discipline's emergence and thematic clusters. The review surveys mechanistic descriptions of tumor growth, the tumor microenvironment, clonal evolution, and metastasis, before examining models of treatment response, optimal control of therapy, and evolutionary game theory as frameworks for understanding and counteracting drug resistance. Multiscale modeling approaches for coupling molecular, cellular, and tissue-level processes are discussed alongside data-driven methods, including genomics and bioinformatics, artificial intelligence, and mechanistic learning, which combines knowledge-driven mathematical models with data-driven machine learning. Particular attention is given to digital twins and their prospective clinical validation, and to the practical requirements for translating mathematical models into clinical implementation, including model validation, FAIR data principles, and shared benchmarking infrastructure. The review concludes by outlining future directions for the field, emphasizing that sustained interdisciplinary collaboration across mathematics, biology, data science, and clinical medicine, rather than any single modeling paradigm, will determine the extent to which mathematical oncology translates into improved patient care.

2020 Mathematics Subject Classification: 00-02, 62-02, 68-02, 92-02.
Keywords: mathematical oncology, tumor growth modeling, tumor microenvironment, evolutionary game theory, multiscale modeling, digital twins, machine learning in oncology.
Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project 17I04-04-V05-00114 (AI-ONKO).

Contents

1. Introduction to mathematical oncology

Mathematical oncology connects mechanistic modeling, experimental biology, clinical medicine, and data science. The integration of mathematical techniques into oncology enables more precise cancer detection, treatment planning, and a better understanding of tumor dynamics. The field encompasses models and methodologies focused on biological questions related to tumor growth, treatment efficacy, and patient outcomes. Its goal is to formulate testable hypotheses and to translate available data into decisions with clearly expressed uncertainty [37].

1.1. Scope of the field and historical development

From population-level growth models, the field has expanded toward spatial, evolutionary, multiscale, and data-driven approaches. Its roots reach back to simple yet clinically influential models from the 1960s – most notably the Skipper, Schabel, and Wilcox log-kill hypothesis, according to which cytotoxic treatment kills a constant fraction (not a constant number) of tumor cells regardless of tumor size, which directly shaped the dosing regimens of modern chemotherapy (Section 8) [70]. A large-scale bibliometric analysis of more than 19 000 publications from 1951–2024 shows that the term “mathematical oncology" itself has gained usage gradually, only distinguishing itself as a field separate from the related areas of systems biology and pharmacokinetics/pharmacodynamics in recent decades, alongside a shift in research focus from basic tumor biology toward therapeutically oriented modeling [61]. This historical development and its thematic clusters are discussed in more detail in Section 2. One of the foundational community resources is the portal `https://mathematical-oncology.org/`, which gathers information on research, events, and professional initiatives [51].

1.2. Why mathematics works in oncology

A model separates assumptions from consequences, making it possible to compare mechanisms and to design experiments or treatment strategies. Simulating tumor growth and its response to therapeutic interventions allows treatment regimens to be compared before their clinical use. Models can also investigate tumor aggressiveness, the emergence of resistance, and possibilities for treatment personalization. Bibliometric findings show that research oriented in this way rests on interdisciplinary and international collaboration [62].

1.3. Interdisciplinarity

1.3.1. Mathematics and statistics

Mathematics provides the language for system dynamics, optimization, and reasoning under uncertainty. Statistical modeling is essential for processing large

sets of molecular and clinical data, for parameter estimation, and for evaluating clinical trials. It should not be forgotten, that domain knowledge is more than necessary for the correct interpretation of results and their safe use in clinical practice.

1.3.2. Molecular and cell biology

Biological knowledge determines the mechanisms, state variables, and relationships that a model must capture – from genomic and transcriptomic data characterizing individual cells and clones (Section 12) to the biophysical and signaling processes of the tumor microenvironment (Section 5).

1.3.3. Clinical medicine

The clinical question determines the model's objective, the relevant endpoints, and the limits of its safe use. Without a clearly formulated clinical context – the target population, the decision the model is meant to support, and the manner of its validation (Section 15) – even a mathematically correct model remains clinically unusable.

1.3.4. Computer science and data science

Computational methods enable the simulation of complex systems, the integration of heterogeneous data, and the development of predictive tools, from classical machine learning to foundation models (Section 13) and digital twins continuously updated with data from a specific patient (Section 14).

1.4. Fundamental research questions

- How does a tumor grow, invade, and metastasize? (Sections 4, 5, and 7)
- How do heterogeneity and drug resistance arise? (Sections 6 and 10)
- How can therapy be optimized and personalized? (Sections 8, 9, and 14)
- How can a model be validated and safely translated into clinical practice? (Section 15)

1.5. Structure of the document

The following sections proceed from an overview of the field and a classification of models (Sections 2–3), through the mechanistic modeling of biological processes – tumor growth, the microenvironment, evolution, and metastasis (Sections 4–7), through the modeling of treatment and its optimization, including evolutionary game theory (Sections 8–10) and multiscale integration (Section 11), to data-driven methods, digital twins, and clinical implementation (Sections 12–15), with an outlook on the future direction of the field in the concluding Section 16.

2. Bibliometric review

This section maps the development of publication activity, dominant themes, and collaboration within the mathematical oncology community. The starting point is an analysis of 75 years of research on mathematical methods in oncology [61], complemented by a bibliometric study of interdisciplinarity, internationalization, and thematic evolution [62].

2.1. Data collection and cleaning methodology

The databases, search query, analysis period, inclusion criteria, and deduplication method must be reported in such a way that the analysis is reproducible. The analyzed publications were divided into two corpora. The first comprised approximately 1,500 manually curated papers collected through the newsletter *This Week in Mathematical Oncology* (TWIMO). The second corpus contained approximately 19,000 documents from the years 1951–2024, retrieved from Scopus using relevant keywords [61]. Analysis of these sets provided an overview of the topics and subareas examined.

2.2. Publication and citation trends

The number of publications in the Scopus corpus shows a markedly increasing trend over the period studied, particularly after the year 2000; see Figure 1. The figure was taken from the original version of the document and is based on the bibliometric analysis of [61].

The most frequently represented journals in the analyzed corpus were also identified. Table 1 preserves the H-index and quartile reported in the original version of the document according to the SCImago portal for 2025 [69].

TABLE 1. Most frequently represented journals in the Scopus corpus and bibliometric indicators reported in the original document for 2025.

Journal	Quartile	H-index
Journal of Theoretical Biology	Q2	184
PLOS ONE	Q1	500
Bulletin of Mathematical Biology	Q1	106
Scientific Reports	Q1	382
PLOS Computational Biology	Q1	237
Mathematical Biosciences	Q1	118
Medical Physics	Q1	229
Cancer Research	Q1	522
Physics in Medicine and Biology	Q2	231
Mathematical Biosciences and Engineering	Q2	68

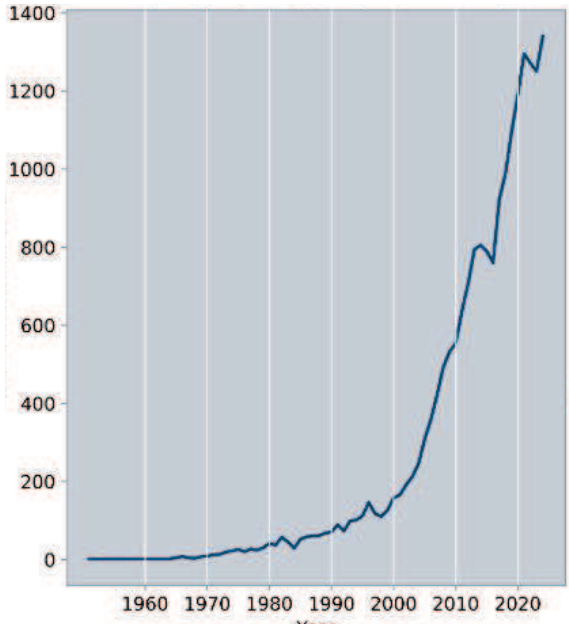

FIGURE 1. Number of publications in the field of mathematical research in oncology published annually in the years 1951–2024. Source: [61].

2.3. Leading research groups and authors

An analysis of bibliographic linkage (co-citation) among the 20 most represented authors identified six clearly separated collaboration clusters in the query-based corpus, which also correspond to the field's main thematic directions [61]:

- *Evolutionary mathematical oncology*, around Anderson, Gatenby, Brown, and Basanta (Sections 6 and 10).
- *Linking models with quantitative imaging*, around Yankeelov, Rockne, Swanson, and Pérez-García (Sections 4 and 14).
- *Fundamental applications of mathematics in oncology*, particularly microenvironment modeling, around Maini, Byrne, and Chaplain (Section 5).
- *Radiotherapy and cancer cell stemness*, spanning fractionation and dose-response modeling in radiotherapy (Section 8) together with the population dynamics of stem-like tumor cells.
- *Multiscale modeling*, integrating processes from the genetic to the whole-body level (Section 11).

- *Optimal control of treatment* (Section 9), where the most cited works include the long-standing research of Ledzewicz and Schättler.

Of the 20 most represented institutions, 13 are based in the USA (e.g., Moffitt Cancer Center, MD Anderson Cancer Center, Memorial Sloan-Kettering, National Cancer Institute) and 4 in the United Kingdom (Oxford, Imperial College London, University College London, University of Dundee), confirming a strong geographic concentration of the field alongside its declared international collaboration [61].

2.4. Thematic clusters and keyword evolution

The six collaboration clusters described above also correspond to six main thematic areas into which most of the analyzed publications can be classified: imaging-linked mathematical oncology, evolutionary mathematical oncology, fundamental applications of mathematics in tumor biology, radiotherapy and stemness, multiscale modeling, and optimal control. The temporal evolution of keywords across the corpus shows a systematic shift in emphasis from questions of basic tumor biology (growth, angiogenesis, apoptosis) toward therapeutically oriented modeling, with the frequency of the term “mathematical oncology" itself increasing over the decades, while the frequency of the related but distinct terms systems biology" and “pharmacokinetics/pharmacodynamics" relatively declines – which the authors interpret as bibliometric evidence of the gradual emergence of mathematical oncology as an independent field rather than a subarea of these older disciplines. Publications in the corpus receive on average 28.1 citations (median 10), with as many as 63 % of them published in first-quartile (Q1) journals compared with an expected approximately 30 % under a random distribution, suggesting an above-average citation impact of this interdisciplinary research [61].

2.5. Conferences, societies, and initiatives

The Society for Mathematical Biology, TWIMO, and the Mathematical Oncology Roadmap are assessed separately. The TWIMO newsletter (This Week in Mathematical Oncology), which provided the smaller, manually curated corpus for this section (above), was launched in 2017 with the aim of gathering current publications and events in the field into a single regular source; approximately 45 % of its corpus overlaps with the independently assembled query-based Scopus corpus, providing a useful cross-check of the coverage of both data collections [61]. Formal consolidation of the field is also represented by the so-called Mathematical Oncology Roadmap – a joint document by dozens of leading researchers (including Rockne, Anderson, and Swanson), which formulates the personalization of medicine through mathematics, modeling, and simulation as its dominant theme: the use of patient-specific clinical data for earlier detection, prediction

of treatment response, the design of adaptive treatment plans that overcome resistance (Section 6), and the establishment of field-wide standards for sharing model outputs (Section 15) [64]. The bibliometric results should also be interpreted as a picture of collaboration among mathematics, biology, and clinical medicine, not merely as a ranking of institutions or authors [62].

2.6. Limitations of the bibliometric analysis

Database coverage, keyword selection, and language and citation bias must be separated from the interpretation of scientific significance. At the same time, the difference between the manually curated TWIMO corpus and the more extensive Scopus results shows that the findings depend on the literature selection strategy.

3. Classification of mathematical models

Models can be classified according to their treatment of randomness, space, scale, the representation of individual cells, and the manner in which data are used. A given study may belong to several categories at once.

3.1. Deterministic models

3.1.1. Ordinary differential equations

Ordinary differential equations (ODEs) describe the temporal evolution of quantities aggregated over an entire population or compartment (e.g. the total number of tumor cells) without explicit spatial structure. This class of models is the simplest to analyze and calibrate and forms the basis of growth laws (Section 4), compartmental treatment models (Sections 8 and 9), and tumor–immune interactions (Section 5).

3.1.2. Partial differential equations

Partial differential equations (PDEs) extend the ODE description with explicit spatial dependence, typically through a reaction–diffusion term. This class is indispensable wherever spatial gradients (of oxygen, drug, tumor cells) themselves determine the dynamics of the system, for example in tumor invasion (Section 4) or in modeling the microenvironment (Section 5).

3.1.3. Delay equations

Delay differential equations (DDEs) introduce an explicit time lag between a stimulus and its response, for example between recognition of the tumor by the immune system and the onset of the effector response, or between drug

administration and its biological effect. A delay can qualitatively change the dynamics of a system (e.g. inducing oscillations that the corresponding delay-free system does not exhibit), and is therefore particularly relevant in models of tumor–immune interaction and pharmacodynamics (Section 8), where the delay in response is biologically meaningful.

3.1.4. Integro-differential equations

Integro-differential equations describe populations structured by a continuous internal variable (e.g. cell age, clone size, or level of resistance), such that the rate of change of the density at a given point depends on the integral (sum) of contributions across the entire distribution of that variable. This formalism is natural for modeling heterogeneous, continually branching tumor populations (Section 6), where discrete cell types are replaced by a continuous distribution of phenotypes or resistance levels.

3.2. Stochastic Models

3.2.1. Markov chains and branching processes

When the number of relevant cells is small (e.g. at the onset of tumor expansion or upon the emergence of a resistant clone), the randomness of individual division, mutation, and death events can no longer be neglected, and the deterministic ODE description loses validity. Markov chains and multitype branching processes model precisely this discrete, stochastic dynamics of small populations and form the basis of the models of clonal expansion and resistance emergence discussed in Section 6.

3.2.2. Birth-death models and the Gillespie algorithm

Birth–death models describe a population as a collection of discrete birth (division) and death events occurring at given rates; if these rates are state-dependent, the resulting process is a continuous-time Markov process. Gillespie derived an algorithm for the exact stochastic simulation of such systems, which, rather than discretizing time, generates a sequence of actual times of the individual reactions (events), and which remains to this day the standard tool for simulating stochastic biochemical and population systems alike [33]. In the oncological context, this apparatus complements the analytical branching processes (above) with the ability to simulate more complex models of clonal dynamics that are not analytically tractable.

3.3. Agent-based and hybrid models

3.3.1. Cellular automata

A cellular automaton represents tissue as a grid of discrete spatial cells, each of which is, at a given time, in one of a finite number of states (e.g. healthy, tumor, necrotic), and this state is updated according to local rules depending on the states of neighboring cells. Moreira and Deutsch, in a critical review of this class of models in oncology, show that even simple local rules can reproduce the complex, experimentally observed patterns of growth of both avascular and vascularized tumors, invasion, and angiogenesis (Section 5), while also highlighting the limitations of this class of models, in particular the sensitivity of the resulting pattern to the choice of grid and neighborhood [55].

3.3.2. Agent-based models

Agent-based models (ABMs) generalize cellular automata in that individual cells (agents) are not bound to a fixed grid, and their behavioral rules can be substantially richer – dependent on an internal intracellular network, the agent's history, or continuous microenvironmental fields. In practice, this class of models is most often combined with a PDE description of diffusive fields into hybrid PDE–ABM models, discussed in detail in Section 11.

3.4. Network models

Gene regulatory, protein interaction, and signaling networks describe the relationships among molecular components. Unlike the preceding classes of models, which describe the dynamics of cells or tissue, network models typically describe the dynamics or topology of relationships *within* a single cell (a regulatory network) or among organs at the level of the whole organism (e.g. a network of anatomically permissible routes of metastatic spread, Section 7); in both cases, nodes represent entities (genes, proteins, organs) and edges represent their mutual relationships or interactions.

3.5. Data-driven models

This category includes statistical learning, Bayesian methods, and machine and deep learning. Unlike the mechanistic classes of models described above, whose structure derives from an assumed biological mechanism, data-driven models derive their structure (or at least their parameters) predominantly from observed data without the need to explicitly formulate a mechanism; their role in modern oncology, from classical machine learning to foundation models, is discussed in detail in Section 13.

3.6. Mechanistic versus data-driven approaches

The choice of model must correspond to the question at hand, the available data, parameter identifiability, and the required degree of interpretability. This trade-off is not merely theoretical: Section 16, drawing on the concept of *mechanistic learning* [53], demonstrates that mechanistic and data-driven approaches are not mutually exclusive alternatives, but can be systematically combined (sequentially, in parallel, extrinsically, or intrinsically) so as to harness the strengths of both – mechanistic interpretability and data-driven flexibility – simultaneously.

4. Modeling tumor growth

Growth laws form the basis for estimating tumor kinetics as well as for modeling treatment effects. For a population of tumor cells or tumor volume $N(t)$, the following models are most commonly compared.

4.1. Exponential model

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN. \tag{1}$$

The relative growth rate $\dot{N}/N = r$ is constant and independent of population size. The model describes early stages of growth well, when the availability of nutrients and space is not limiting, but it systematically overestimates tumor size over longer time horizons, since it contains no saturation mechanism.

4.2. Logistic model

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN\left(1 - \frac{N}{K}\right). \tag{2}$$

The carrying capacity K represents the maximum sustainable population size determined by microenvironmental constraints (Section 5). The relative growth rate decreases linearly with N/K, so the slowdown in growth manifests approximately symmetrically around the inflection point $N = K/2$.

4.3. Gompertzian model

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN\ln\left(\frac{K}{N}\right). \tag{3}$$

Unlike the logistic model, the relative growth rate $\dot{N}/N = r\ln(K/N)$ decreases logarithmically, not linearly, with distance from the carrying capacity, leading to a markedly earlier and asymmetric slowdown in growth. The Gompertzian model is empirically the most commonly used description of solid tumor growth;

however, its parameters r and K do not have an unambiguous mechanistic interpretation and must be understood as effective, data-fitted quantities.

4.4. Bertalanffy and West model

The classical von Bertalanffy model derives the growth of mass N from a balance between anabolism, proportional to the surface area of the cell population, and catabolism, proportional to its volume:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = aN^{2/3} - bN, \tag{4}$$

where the exponent 2/3 follows from the geometric assumption that nutrient uptake is limited by surface area. West, Brown, and Enquist generalized this approach on the basis of the theory of metabolic scaling and the fractal geometry of transport networks (e.g., vascular supply), according to which metabolic rate scales with mass as $B \propto N^{3/4}$ instead of the classical surface rule; the resulting model

$$\frac{\mathrm{d}N}{\mathrm{d}t} = aN^{3/4} - bN \tag{5}$$

after suitable normalization of variables predicts a single universal growth curve common to a wide range of animal species [80]. In an oncological context, this approach provides a mechanistic alternative to purely phenomenological models (exponential, logistic, Gompertzian), since the growth exponent is derived from physiological principles of nutrient and energy transport rather than merely fitted to data.

4.5. Mechanistic growth models

The preceding models describe only the overall tumor size without spatial structure. Mechanistic models, by contrast, explicitly represent the density of tumor cells $c(\boldsymbol{x}, t)$ in space and its spread through diffusion and proliferation. A basic example is the reaction-diffusion, or Fisher–Kolmogorov–Petrovsky–Piskunov, equation

$$\frac{\partial c}{\partial t} = \nabla \cdot (D \nabla c) + \rho c \left(1 - \frac{c}{K}\right), \tag{6}$$

where the diffusion coefficient D describes cell migration (invasion) and ρ their local proliferation. Swanson, Alvord, and Murray used this equation to model glioma growth and invasion, choosing different diffusion coefficients in white and gray matter of the brain according to anatomically precise tissue segmentation; the simulations thereby reproduced clinically observed, diffusely bounded tumor shapes and indicated submicroscopic invasion undetectable by CT or MRI [73]. This type of model combines the advantages of a mechanistic framework (Section 3) with a direct link to patient imaging data and forms the basis for the spatial extensions discussed in Section 5.

4.6. Comparison of models

Each model is evaluated according to the biological interpretation of its parameters, identifiability, advantages, limitations, and suitable areas of application. A systematic comparison of eight classical growth models – exponential, exponential-linear, power-law, Gompertzian, logistic, generalized logistic, von Bertalanffy, and a model with dynamic carrying capacity – against experimental tumor growth data in mice (a Lewis lung carcinoma and a xenografted breast carcinoma) evaluated both their descriptive power and their ability to forecast future growth. For the breast data, the Gompertzian and exponential-linear models best captured the observed dynamics, and the exponential-linear model additionally achieved the highest predictive power, with prediction scores of at least 80 % extending as far as 12 days into the future. For the lung data, the Gompertzian and power-law models provided the most parsimonious and identifiable description, yet none of the models – including these two – could predict beyond the very next measurement with a success rate of at least 70 % [10]. This dissociation between descriptive or parsimonious fit and predictive accuracy is a general warning against assuming that a model's ability to describe past growth implies an ability to forecast future growth, and is particularly relevant for growth models intended for clinical prediction (Section 15).

5. Modeling the tumor microenvironment

The microenvironment influences proliferation, invasion, metabolism, and treatment response alike. Spatial gradients are often described by reaction-diffusion equations.

5.1. Angiogenesis

A tumor exceeding the diffusive range of oxygen (on the order of hundreds of micrometers) requires its own vascular supply for further growth. Anderson and Chaplain described the formation of the capillary network as the response of endothelial cells to a chemical gradient of angiogenic factors (TAF) released by the tumor, modeling cell migration as a combination of random diffusion, chemotaxis toward TAF, and haptotaxis toward fibronectin in the extracellular matrix, in both a continuous and a discrete (cellular) formulation [5]. At a different level of description, Hahnfeldt and colleagues, instead of an explicit vascular network, introduced a carrying-capacity variable $K(t)$ representing the effective vascular supply attainable by the tumor, and coupled it to the tumor volume $N(t)$ through the system

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\lambda N \ln\left(\frac{N}{K}\right), \qquad \frac{\mathrm{d}K}{\mathrm{d}t} = bN - dN^{2/3}K, \tag{7}$$

where b is the rate of angiogenic stimulation and d the rate of vascular degradation proportional to the tumor surface area; the model thus links Gompertzian growth (Section 4) with a dynamically changing carrying capacity tied to angiogenesis, and reproduces the observed growth kinetics of both untreated and antiangiogenically treated tumors [35].

5.2. Hypoxia, oxygen, and nutrients

Oxygen concentration decreases with increasing distance from the nearest blood vessel, creating spatially heterogeneous zones of normoxia, hypoxia, and necrosis. A synthetic review of mathematical approaches to cancer dynamics emphasizes that hypoxia does not act merely as a growth-limiting factor, but actively alters the cell cycle, induces selection of more aggressive phenotypes, and modulates treatment response, with these phenomena modeled across scales ranging from intracellular signaling (e.g., the HIF-1 pathway) to tissue-level reaction-diffusion models of oxygen [12]. Hypoxic regions thus represent a natural link between the angiogenesis described above and the multiscale models discussed in Section 11.

5.3. Extracellular matrix

The extracellular matrix (ECM) is not merely a passive tissue scaffold, but an active component that directs the direction of tumor cell migration while simultaneously being remodeled by those cells. In his hybrid model of solid tumor invasion, Anderson tracks four interconnected variables – tumor cells, host tissue (ECM), matrix-degrading enzymes, and oxygen – whereby cells produce ECM-degrading enzymes and migrate along the emerging geometric gradient via haptotaxis; the resulting shape of the invasive front (compact, fingered, or fragmented) depends critically on the strength of cell-cell adhesion [4]. This approach demonstrates that invasion morphology is an emergent property of local cell-matrix interactions, rather than merely a parameter freely fitted to the observed tumor shape.

5.4. Mechanical forces and growth-induced stress

Growth of a cell population within a confined, elastic environment generates mechanical (so-called solid) stress, which in turn suppresses proliferation and can lead to the collapse of blood and lymphatic vessels within the tumor. Roose and colleagues modeled the growth of multicellular spheroids in a linearly poroelastic medium, in which both the tumor and the surrounding tissue behave as a mixture of a solid component and interstitial fluid; the model predicted that accumulating solid stress slows and eventually halts spheroid growth, consistent with experimental observations of shrinking daughter cells during mitosis in a

spatially confined environment [65]. Mechanical stress thus represents a complementary, biophysical regulatory mechanism alongside the chemical signals (oxygen, nutrients, growth factors) discussed in the preceding sections.

5.5. Immune and stromal cells

Immune cells of the microenvironment can recognize and eliminate the tumor, but their effect depends on the ratio between the killing rate and the tumor growth rate, which can lead to dormancy, equilibrium, or escape from immune surveillance (Section 6). Kuznetsov and colleagues formulated this interaction as a system of ordinary differential equations for populations of effector cells and tumor cells, analogous to predator-prey models, which they calibrated against experimental data on the growth and regression of BCL1 B-cell lymphoma in the mouse spleen; bifurcation analysis of the model reproduced phenomena observed both clinically and experimentally – immunostimulation of tumor growth under a weak response, tumor escape from surveillance ("sneaking through") under an intermediate response, and the emergence of a dormant state under a strong response [45]. This type of model forms the basis for more recent extensions incorporating immunotherapy and checkpoint inhibitors (Section 8).

5.6. Spatial PDE models

The general form for the concentration c is

$$\frac{\partial c}{\partial t} = D\nabla^2 c + R(c, \boldsymbol{x}, t), \tag{8}$$

where D is the diffusion coefficient and R encompasses local production and consumption.

6. Tumor evolution

A tumor is an evolutionary system in which mutation, selection, drift, and ecological interactions change the representation of clones in time and space.

6.1. Darwinian and clonal evolution

The conceptual foundation of modern oncology as an evolutionary discipline was laid by Nowell, who proposed that most neoplasms arise from a single cell and that tumor progression is the result of progressively acquired genetic variability within the original clone, which enables the sequential selection of increasingly aggressive subpopulations [57]. This formulation transfers the Darwinian principles of variation, heredity, and selection to the level of somatic cells: a tumor is not a static collection of cells with a uniform genetic makeup, but a dynamically evolving population in which genetic instability is itself subject to selection. It was precisely this genetic and phenotypic individuality of individual tumors that

Nowell already identified as the reason why effective treatment in one patient can be thwarted by the emergence of a resistant subpopulation.

6.2. Mutations and branching processes

A quantitative framework for clonal evolution is provided by multitype branching processes, in which the tumor is modeled as a population of cells, where each cell, in discrete or continuous time, divides, mutates into a new (typically more aggressive) type, or dies. Bozic and colleagues used this approach to model the accumulation of so-called driver mutations (which slightly increase the rate of clonal expansion) and passenger mutations (selectively neutral) during tumor progression, and showed that even small differences in the advantage conferred by driver mutations lead to enormous variability in growth rate and in the time required to reach a clinically detectable tumor size, which is consistent with the observed heterogeneity in latency periods among patients [11]. The branching process thus provides a mechanistic explanation for a phenomenon that a purely deterministic growth model (Section 4) would fail to capture.

6.3. Intra- and intertumor heterogeneity

While the previous sections describe evolution as a sequence of successive clonal expansions, sequencing of multiple spatially separated samples from the same tumor has revealed that evolution can be markedly branched. Gerlinger and colleagues, using this method (exome sequencing, analysis of chromosomal aberrations and ploidy across multiple regions of primary renal carcinoma and its metastases), found that 63 to 69 % of all somatic mutations were not detectable across all regions of the same tumor, a consequence of branched rather than linear evolution [30]. This intratumor heterogeneity has a direct practical consequence: a single-site tumor biopsy systematically underestimates the mutational burden and can lead to an incorrect choice of targeted therapy, because undetected subclones may carry resistant variants already present before treatment begins.

6.4. Emergence and spread of resistance

A key question for clinical practice is whether resistant cells arise only as a response to treatment, or already exist in the tumor before treatment begins. Iwasa, Nowak, and Michor modeled an exponentially growing population of tumor cells, in which sensitive cells randomly mutate into resistant ones, using a branching process that, in the limiting case without cell death, reduces to the classical Luria–Delbrück model of fluctuations; the model makes it possible to estimate the probability that a resistant clone already exists at the time of diagnosis, depending on tumor size, mutation rate, and the fitness difference between sensitive and resistant cells [41]. This theory of pre-existing resistance is the fundamental assumption underlying the concept of adaptive therapy (Sections 8

and 10), which seeks, instead of eliminating the sensitive clone, to maintain its competitive pressure on the resistant subpopulation.

6.5. Population dynamics

The branching processes described above model evolution as a sequence of discrete mutational events in an exponentially growing population; in practice, however, the same mathematical structure is supplemented by density-dependent phenomena – competition for space and nutrients, ecological interactions among clones (Section 10), and the spatial structure of the population (Section 5) – which modify the simplifying assumptions of purely exponential clonal expansion. Combining branching processes with explicit population dynamics makes it possible, alongside the question of *whether* a resistant clone will arise, to also examine the question of *when* and under what ecological conditions it will prevail against both resistant and sensitive competitors.

6.6. Phylogenetic reconstruction of the tumor

Sequencing data from one or more biopsies of the same tumor make it possible to retrospectively reconstruct the order in which subclones arose, using methods adapted from species phylogenetics, modified for the particularities of somatic evolution (e.g., the need to deconvolute mixed, rather than pure, clonal sequencing samples). A review of this rapidly developing field traces its development from early oncogenetic trees, through methods comparing multiple spatial regions of the same tumor, to current approaches working with single-cell data, emphasizing that despite borrowing the apparatus of species phylogenetics, the distinctiveness of somatic evolution must be taken into account – in particular the substantial contribution of structural variation as a mechanism of change, which classical phylogenetic methods did not originally take into account [66]. The reconstructed phylogenetic tree of the tumor thus provides direct input for identifying clonal (early, shared) versus subclonal (late, private) mutations relevant to treatment selection (Section 15).

7. Metastasis

Modeling the metastatic cascade combines local invasion, intravasation, transport of circulating tumor cells, extravasation, and colonization of a distant organ.

7.1. The metastatic cascade

Metastasis proceeds through a sequence of biologically distinct yet interdependent steps. The cascade begins with local invasion of the surrounding tissue (Section 5) and epithelial-mesenchymal transition, continues with intravasation

into the blood or lymphatic circulation and survival in circulation, and ends with extravasation into distant tissue and colonization, i.e., the formation of a macroscopic secondary lesion [26]. A review of the molecular principles of metastasis emphasizes that the success of the entire cascade is limited primarily by its final steps. Colonization of the distant organ is, statistically, the least efficient step, since the majority of cells that leave the primary tumor fail to complete it – yet it is also the most clinically relevant one [26]. This fact motivates mathematical models that, rather than describing each step of the cascade in detail, aggregate the entire process into a few measurable rate constants, introduced below.

7.2. Circulating tumor cells

Circulating tumor cells (CTCs) represent a directly measurable, yet extremely short-lived, intermediate stage of the metastatic cascade. Hamza and colleagues developed a method of continuous blood exchange between tumor-bearing and healthy mice, which made it possible to simultaneously estimate the rate of CTC release from the primary tumor and the rate of their clearance from the circulation; they measured a half-life of circulating tumor cells of only 40 to 260 seconds across models of small-cell and non-small-cell lung carcinoma and pancreatic ductal adenocarcinoma, i.e., one to two orders of magnitude shorter than the half-life of circulating leukemic cells [36]. The CTC generation and clearance rates calibrated in this way enter as a parameter into the compartmental models of metastatic spread described below, and also provide a mechanistic explanation for why only a minute fraction of released cells is able to establish a distant colony.

7.3. Dormancy and colonization

A cell that successfully extravasates does not necessarily transform immediately into a growing metastasis in the target tissue – it often first enters a long-term non-proliferating, dormant state. Ghajar and colleagues showed, using a model of breast carcinoma metastasis, that the fate of a disseminated cell is determined by the type of vascular microenvironment in which it settles: stable, mature microvasculature keeps cells quiescent through the production of thrombospondin-1, whereas newly formed sprouts of angiogenic vessels (rich in TGF-β and periostin) instead reactivate dormant cells toward proliferation [31]. The transition between dormancy and colonization is thus governed by local signals from the perivascular niche, not solely by intrinsic properties of the tumor cell, which makes it a potential therapeutic target independent of the genetics of the primary tumor.

7.4. Organotropism and the pre-metastatic niche

Different tumor types systematically prefer distinct target organs for metastasis (organotropism), which cannot be explained by the anatomy of the circulatory system alone. The concept of the perivascular niche discussed above is extended

in this context by the notion of the *pre-metastatic niche* – a microenvironment in a distant organ that is prepared to receive tumor cells even before their arrival, through signals sent out by the primary tumor [31, 26]. The combination of anatomical flows (Section 5) and organ-specific niche readiness determines the observed, statistically non-random distribution of metastases among organs, which is modeled by the networks discussed in the final section of this section.

7.5. ODE and compartmental models

Instead of explicitly tracking individual cells, the metastatic process can be described at the level of entire colonies. Iwata, Kawasaki, and Shigesada introduced a dynamic model of the size distribution of multiple metastatic colonies, in which the state of the system is described by the density of colonies according to their size and the time of their formation, whereby new colonies arise through colonization at a rate proportional to the size of existing tumors (both primary and metastatic), and each colony subsequently grows independently, typically following a Gompertzian function (Section 4); for Gompertzian growth, the model has an explicit solution that converges to an asymptotically stable distribution of colony sizes [42]. This type of model makes it possible to work backward from the observed distribution of metastasis sizes at the time of diagnosis to estimate the parameters of the metastatic colonization rate, which are not directly measurable.

7.6. Stochastic models of metastasis formation

The formation of an individual metastasis is a rare, stochastic event, and is therefore naturally modeled at the population level as a random process. Gerlee and Johansson formulated metastatic spread as a stochastic process on a network, in which nodes represent organs and edges represent only the anatomically permissible spreading routes (e.g., determined by the direction of blood and lymphatic flow), thereby restricting the space of possible models to biologically realistic scenarios. By fitting the model to clinical data on the incidence of metastases in tongue and ovarian tumors, they showed that incidence can be described by a small number of biologically interpretable parameters, which reveal site-specific relative rates of spread to individual organs [29].

7.7. Spatial and network models of spread

The network formulation described above also serves as a spatial model of spread: the position of a node (organ) in the network of anatomically permissible routes determines the probability and order of its colonization, independent of its physical distance from the primary tumor as measured directly in the body's space. A model of this type also enables individualized prediction, since from the location and stage of the primary tumor it is possible to derive, for a specific patient, a probability distribution of the organs most at risk of future metastasis

[29]. The network model formulated in this way complements the compartmental models of colony growth (above) with a spatial, topological component of the metastatic process, which is also relevant for planning a patient's imaging follow-up (Section 15).

8. Modeling treatment

Treatment models link pharmacokinetics, pharmacodynamics, the biological mechanism, and the clinical dosing regimen.

8.1. Chemotherapy

8.1.1. Maximum tolerated dose

The classical maximum tolerated dose (MTD) strategy seeks to reduce the tumor population as quickly as possible, with dosing limited only by host toxicity. From the perspective of evolutionary dynamics (Section 6), however, aggressively killing sensitive cells simultaneously removes their competitive pressure on resistant subclones, which can then expand freely once the sensitive population has disappeared [47]. This trade-off between immediate reduction of tumor burden and the long-term risk of selecting for resistance is a central theme of optimal control of treatment (Section 9).

8.1.2. Metronomic therapy

Metronomic dosing replaces short cycles of maximum dose with lower, but far more frequently administered, doses. A review of mathematical models of MTD and metronomic protocols shows that a lower, continuous dose can, in addition to its direct cytotoxic effect, also act antiangiogenically and keep the sensitive subpopulation large enough to competitively suppress resistant cells, thereby altering not only the intensity but also the qualitative character of the treatment response compared with MTD [47]. A formal derivation of the conditions under which such a regimen is optimal is the subject of optimal control theory (Section 9).

8.2. Immunotherapy

8.2.1. T cells and checkpoint inhibitors

Effector T cells recognize and kill tumor cells, but their activity is dampened by inhibitory checkpoints (e.g., binding of PD-L1 on the tumor cell receptor to PD-1 on the T cell), which are the targets of checkpoint inhibitors. De Pillis, Radunskaya, and Wiseman formulated and experimentally validated a system of ordinary differential equations for the interaction of the tumor with NK cells,

$CD8^+$ T lymphocytes, and circulating lymphocytes, which additionally incorporates chemotherapy as a supplementary input; the model was calibrated and tested against experimental tumor growth data in mice and reproduces phenomena also observed clinically, such as immune-mediated elimination, equilibrium, and tumor escape [19]. This predator–prey-type framework extends earlier models of tumor–immune interaction (Section 5) with additional immune subpopulations and serves as a basis for modeling checkpoint inhibitors, which act on the parameters governing contact inhibition between T cells and tumor cells.

8.2.2. CAR-T therapy

CAR-T cells are genetically engineered T lymphocytes bearing a chimeric antigen receptor targeted at a specific tumor antigen, whose population dynamics are modeled analogously to models of tumor–immune interaction, but with explicit distinction between effector and memory CAR-T cells. The CARTmath model by Barros and colleagues describes this dynamic using a system of three coupled ordinary differential equations for tumor cells, effector CAR-T cells, and memory CAR-T cells: effector cells proliferate, kill tumor cells, partly differentiate into memory cells, and partly die off or are suppressed by tumor-induced immunosuppressive mechanisms, while memory cells rapidly redifferentiate back into effector cells upon renewed encounter with the antigen [7]. The model was calibrated on data from preclinical models of hematological malignancies in immunodeficient mice and implemented as a publicly available computational tool, allowing its direct use for simulating alternative CAR-T dosing schedules.

8.3. Radiotherapy

8.3.1. Linear-quadratic model

For a dose d, the surviving fraction of cells is

$$S(d) = \exp(-\alpha d - \beta d^2). \tag{9}$$

The linear term αd represents cell death caused by a single lesion resulting from a DNA double-strand break induced by a single ionizing particle, while the quadratic term βd^2 describes death resulting from the accumulation of two separately sublethal damages; the ratio α/β thus characterizes how sensitively the tissue responds to the size of the individual dose fraction. The model is clinically reliable up to approximately 8–10 Gy per fraction; outside this range its validity has not been sufficiently verified experimentally.

8.3.2. Fractionation and biologically effective dose

Comparing different fractionation schemes (number of fractions, dose per fraction, total treatment time) requires a common measure of biological effect. Fowler, building on the linear-quadratic model, derived the biologically effective dose (BED)

$$\mathrm{BED} = nd\left(1 + \frac{d}{\alpha/\beta}\right), \tag{10}$$

where n is the number of fractions and d the dose per fraction; BED enables conversion between different fractionation schemes so that the resulting biological effect on a given tissue remains equivalent, and it has become a standard tool for comparing unconventional (hypofractionated or hyperfractionated) protocols with conventional treatment [27].

8.4. Antiangiogenic therapy

Antiangiogenic treatment does not target tumor cells directly, but rather the vascular supply described by the dynamic carrying-capacity model (Section 5) [35]. Ergun, Camphausen, and Wein incorporated this model into an optimal control problem combining radiotherapy with antiangiogenic inhibitors and showed that the optimal antiangiogenic monotherapy consists of a large initial dose that rapidly achieves the target ratio of tumor volume to supporting vasculature volume (approximately 20:1), followed by a continuously increasing maintenance dose; in combination therapy, by contrast, the antiangiogenic component is dosed more intensively only in the later part of the radiotherapy fractionation schedule [24]. This result illustrates the general principle that the order and timing of combined modalities can be just as important as their total dose.

8.5. Combination therapy

The individual treatment modalities described above act on distinct biological targets (tumor cells, vascular supply, the immune system), and their simultaneous application therefore need not be merely additive. De Pillis, Gu, and Radunskaya extended the tumor–immune interaction model (above) to include the combined action of vaccine immunotherapy and chemotherapy, and by analyzing the system's equilibrium points, their stability, and bifurcations, showed that there exist parameter regions in which neither chemotherapy alone nor immunotherapy alone can control the tumor, whereas their combination leads to complete tumor elimination [18]. This result is a mathematical expression of synergy: the effect of the combination cannot be explained by the sum of the effects of its individual components, and its existence or extent must therefore be verified separately for each pair of modalities, not assumed automatically.

8.6. Adaptive therapy

Adaptive therapy departs from the maximum-tolerated-dose paradigm described above by using the ongoing tumor response, rather than a fixed schedule, to decide when to give or withhold treatment. Gatenby, Silva, Gillies, and Frieden argued that treating a tumor to the point of profound shrinkage empties the ecological niche occupied by drug-sensitive cells, allowing any pre-existing or newly arising resistant clone to expand free of competition, whereas maintaining a substantial sensitive population instead preserves this competitive suppression and can delay progression, even at the cost of a larger residual tumor [28]. Mathematically, the sensitive and resistant subpopulations are described by competition models analogous to those of Section 6 and Section 10, with the drug dose entering as a control variable that continuously modulates the sensitive population's growth or death rate rather than as a fixed, time-indexed schedule. In its simplest clinical form this becomes a feedback rule: treatment is withdrawn once the tumor burden falls to a specified fraction of its pretreatment value and resumed once it returns to (or exceeds) that baseline – the dosing algorithm used in the prospective trial of adaptive androgen-deprivation therapy for metastatic castration-resistant prostate cancer discussed in Section 10 [84]. Subsequent theoretical work has generalized this single-drug feedback rule to multidrug regimens, in which both the schedule and the identity of the administered drug are continuously adapted to the estimated clonal composition of the tumor [81], and has shown that the benefit of adaptive dosing depends not only on the fitness cost of resistance but also on the rate of cell turnover within the tumor, which governs how quickly withdrawing the drug is reflected in a shift of clonal composition [71]. This dependence on turnover, rather than on the cost of resistance alone, helps explain why adaptive protocols succeed in some tumor types yet fail in others, and directly motivates the individualized, model-informed dose adjustment discussed in the context of digital twins in Section 14.

9. Optimal control

Optimal control theory formalizes the trade-off between antitumor efficacy, toxicity, resistance, and the costs of treatment [67].

In the work of Ledzewicz and Schättler, the dosing of antitumor drugs is formulated as an optimal control problem. The model describes the tumor and the relevant components of the microenvironment, the dose or combination of doses serves as the control variable, and the objective functional represents, for example, tumor burden, toxicity, or drug consumption [68]. The basic apparatus of optimal control theory for biological models is systematically treated by [48].

9.1. Formulation of the control problem

For the state x, control u, and dynamics $\dot{x} = f(x, u, t)$, the following functional is minimized

$$J(u) = \Phi(x(T)) + \int_0^T L(x(t), u(t), t)\,\mathrm{d}t. \tag{11}$$

9.2. Pontryagin's maximum principle

Direct minimization of the functional $J(u)$ over the infinite-dimensional space of admissible controls $u(t)$ is generally intractable. Pontryagin's maximum principle converts this problem into a finite-dimensional, pointwise optimization: if $u^*(t)$ is the optimal control and $x^*(t)$ the corresponding optimal trajectory, there exists an adjoint (costate) variable $\lambda(t)$ such that $u^*(t)$ maximizes (or minimizes, depending on the sign convention) at every time t the Hamiltonian defined below over the set of admissible control values [48]. The principle thus replaces the search for an optimal function of time with the search for an optimal value at each individual instant of time, which is a substantially simpler task both computationally and analytically.

9.3. The Hamiltonian and adjoint equations

For the dynamics $\dot{x} = f(x, u, t)$ and the functional $J(u)$ from the previous section, the Hamiltonian is introduced

$$H(x, u, \lambda, t) = L(x, u, t) + \lambda^{\top} f(x, u, t). \tag{12}$$

The optimal trajectory, control, and adjoint variable satisfy the system of first-order necessary conditions:

$$\dot{x}^* = \frac{\partial H}{\partial \lambda} = f(x^*, u^*, t), \qquad x^*(0) = x_0, \tag{13}$$

$$\dot{\lambda} = -\frac{\partial H}{\partial x}, \qquad \lambda(T) = \left.\frac{\partial \Phi}{\partial x}\right|_{x^*(T)}, \tag{14}$$

$$u^* = \arg\min_{u \in U} H(x^*, u, \lambda, t). \tag{15}$$

The state equation is integrated forward from the given initial condition, whereas the adjoint equation is integrated backward from the terminal (transversality) condition, which gives rise to the typical two-point boundary value problem that must be solved numerically, as described in Numerical optimization below.

9.4. Bang–bang and singular control

If the Hamiltonian depends linearly on the control u, its minimization over $u \in [0, u_{\max}]$ does not lead to an interior extremum but to a solution on the boundary of the admissible set – the optimal control thus takes on only the

extreme values ($u = 0$ or $u = u_{\max}$) and switches between them (so-called bang–bang control) according to the sign of the switching function (the coefficient of u in the Hamiltonian). Exceptionally, if this switching function vanishes over a time interval of positive length, the control is not uniquely determined by the minimum condition, and one speaks of singular control, which must be derived from the condition that higher-order time derivatives of the switching function vanish. Ledzewicz and Schättler showed for models of combined antiangiogenic therapy and chemotherapy that optimal protocols exhibit exactly this bang–bang/singular structure, with the order and duration of the individual segments depending on the model parameters (e.g., the rate of vascular degradation and tumor sensitivity to the drug) [46]. From a clinical standpoint, singular control corresponds to a continuously varying dose tied to the patient's state, whereas bang–bang segments correspond to periods of maximal dose or complete treatment interruption.

9.5. State and dose constraints

A realistic formulation of a treatment protocol must, in addition to the bound on the control $u \in [0, u_{\max}]$ (e.g., the maximum daily dose), also take into account constraints on the state of the system itself – for example, an upper admissible bound on cumulative toxicity or a lower admissible bound on the number of healthy cells or leukocytes. Such state constraints alter the structure of the necessary conditions, since the adjoint variable may contain an additional singular (measure-valued) contribution on the boundary segment of the trajectory; compartmental chemotherapy models with an explicit toxicity constraint therefore require a more careful formulation than problems with a constraint on the control alone [74].

9.6. Numerical optimization

The two-point boundary value problem arising from the system of state and adjoint equations is rarely solved analytically in practice. The standard approach is the iterative forward–backward sweep method: for an estimated control, the state equation is integrated forward, then, along this trajectory, the adjoint equation is integrated backward, the control is updated from the minimum condition of the Hamiltonian, and the procedure is repeated until convergence [48]. This method is computationally inexpensive and straightforward to implement; however, its convergence is not guaranteed in general and must be verified for specific models.

9.7. Examples of oncological treatment regimens

The apparatus described in this section has been applied to a broad range of oncological models: compartmental models of chemotherapy for sensitive and

resistant subpopulations, combined antiangiogenic and cytotoxic therapy, and metronomic regimens derived directly from the optimality condition rather than from empirical selection [68]. A common finding across these works is that the optimal control almost never corresponds to either a pure maximum-tolerated-dose strategy or a purely constant metronomic dose, but rather to a combination of bang–bang and singular segments, the order of which depends on whether the goal is to minimize the final tumor burden, delay the onset of resistance, or both simultaneously. This conclusion directly connects the formal apparatus of optimal control theory with the biological discussion of adaptive therapy in Section 8 and with evolutionary game theory in Section 10.

10. Evolutionary game theory

Evolutionary game theory describes frequency-dependent selection among cellular phenotypes and makes it possible to examine how therapy alters the ecological interactions within a tumor.

10.1. Payoff matrix and biological interpretation

Rather than providing an explicit mechanistic description of every molecular pathway, evolutionary game theory condenses the outcome of the interaction between two phenotypes into a single payoff (fitness) matrix: if a cell playing strategy i encounters a cell playing strategy j, its reproductive fitness changes by the value a_{ij}. Basanta, Hatzikirou, and Deutsch formalized in this way the trade-off between the invasive (*Go*) and proliferative (*Grow*) phenotypes of a tumor cell through two parameters – the benefit of proliferation at the original site and the cost associated with motility – and showed for which values of these parameters the invasive phenotype comes to dominate, coexists, or disappears from the population [8]. The advantage of this representation is that even an incompletely understood biological mechanism (e.g., the precise molecular costs of cell motility) can be replaced by a few empirically measurable fitness parameters, without the need to model intracellular signaling explicitly.

10.2. Hawk–Dove game

The Hawk–Dove game originally describes a conflict over a limited resource between an aggressive (hawk) and a yielding (dove) strategy, in which – unlike the prisoner's dilemma discussed below – no strategy is dominant and the equilibrium state is mixed or polymorphic [52]. Tomlinson and Bodmer applied this game to the production of angiogenic growth factors by tumor cells: a cell producing the factor bears a cost, but the benefit of locally increased vascularization is shared with neighboring cells regardless of whether they also produce

the factor; the analysis showed that, for realistic values of the cost and benefit, a stable polymorphism arises in which producing and non-producing cells coexist in a steady ratio [76]. They formulated an analogous game for the production of auto and paracrine factors that protect against apoptosis.

10.3. Prisoner's dilemma

In the prisoner's dilemma, it is more advantageous for each player to defect regardless of the opponent's choice, even though mutual cooperation would yield a higher joint outcome – defection is thus a dominant, yet collectively inefficient, strategy. Axelrod, Axelrod, and Pienta modeled the competition between healthy and tumor cells for space and resources in tissue precisely within this framework: healthy cells that maintain tissue homeostasis (e.g., by regulating the local environment for the benefit of their surroundings) represent cooperating players, while tumor cells that disregard this regulation in favor of their own proliferation represent defectors, and they modeled the dynamics of the competition for dominance within the population using a stochastic Moran process [6]. This framework explains why carcinogenesis can be viewed as a local takeover of control by defectors over cooperating tissue, rather than merely the sum of independent genetic events in individual cells.

10.4. Replicator dynamics

For the proportion of strategy x_i, the following holds:

$$\dot{x}_i = x_i\big(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\big). \tag{16}$$

The fitness $f_i(\boldsymbol{x})$ of strategy i depends on the current composition of the population $\boldsymbol{x} = (x_1, \ldots, x_n)$ through the payoff matrix introduced above, while $\bar{f}(\boldsymbol{x}) = \sum_i x_i f_i(\boldsymbol{x})$ is the average fitness of the population. The proportion of a strategy thus grows precisely when its fitness is above average – the equation thereby converts the static payoff matrix into continuous, frequency-dependent dynamics of the composition of the tumor population, formally analogous to replicator dynamics in population genetics and ecology.

10.5. Evolutionarily stable strategy

A strategy (or possibly a mixed strategy, i.e., a probability distribution over several phenotypes) is evolutionarily stable if a population that plays it almost exclusively resists invasion by any small group of mutants playing a different strategy – this is an equilibrium concept stricter than the classical Nash equilibrium, since it also requires resistance to perturbation, not merely the absence of a unilateral advantage from deviation [52]. The equilibrium points of the replicator dynamics described above that are simultaneously evolutionarily stable strategies represent candidates for the long-term observable composition of the

tumor population – but only under the assumption that no external intervention (e.g., treatment) acts on the system to change the payoff matrix.

10.6. Applications in adaptive therapy

Adaptive therapy (Section 8) makes practical use of the fact that treatment changes the payoff matrix between sensitive and resistant cells: in the absence of the drug, resistance typically carries a fitness cost, so sensitive cells competitively suppress resistant ones, whereas during treatment this balance of advantage is reversed. Zhang and colleagues tested this idea in the first prospective clinical trial of adaptive therapy in patients with metastatic castration-resistant prostate cancer. Instead of continuous administration of abiraterone, the dose was interrupted based on continuously measured PSA levels, so as to maintain a stable, but not zero, tumor burden and thereby preserve the competitive pressure of sensitive cells on the resistant subpopulation. Compared with a historical control under continuous dosing, the adaptive protocol extended the time to disease progression while simultaneously reducing the cumulative drug dose [84]. This study represents one of the few examples in which the prediction of an evolutionary game-theoretic model was directly tested in a clinical trial (Section 15), rather than only on experimental or simulated data.

11. Multiscale modeling

Multiscale models link processes at the level of genes, cells, tissue, organ, and patient. A key question is which information is transmitted between scales, and with what uncertainty.

11.1. Temporal and spatial scales

Processes relevant to oncology occur at spatial scales ranging from nanometers (molecular signaling) through micrometers (the cell), millimeters and centimeters (tissue, organ), up to the whole body, and at temporal scales ranging from seconds (biochemical reactions) through hours and days (the cell cycle, Section 4) up to months and years (disease progression, Section 6). Deisboeck, Wang, Macklin, and Cristini, in their review of multiscale cancer modeling, emphasize that precisely this multi-order-of-magnitude disproportion between scales is the main reason why a single mathematical formalism (e.g., a purely ODE-based or purely PDE-based model) cannot adequately describe the full range of phenomena, and that it motivates the combination of the different types of models described in this section [20].

11.2. Linking genes, cells, and tissue

A multiscale model must specify which quantities are transmitted from the finer scale to the coarser one (e.g., from gene expression to the rate of cell division) and, conversely, which signals from the tissue scale feed back to affect intracellular states (e.g., hypoxia altering gene expression, Section 5). Deisboeck and colleagues point out that this bidirectional coupling between scales – not merely a one-way, “bottom-up" transfer – is essential for reproducing phenomena such as the phenotypic plasticity of tumor cells in response to the local microenvironment [20]. In practice, this coupling is realized by linking a model of the intracellular regulatory network (Section 3) with a cell- or tissue-level model, where the output of the former consists of parameters or rules that feed into the latter.

11.3. Hybrid PDE–ABM models

The hybrid approach combines a continuous, PDE-based representation of diffusively spreading quantities (oxygen, nutrients, drug, diffusible signaling molecules) with a discrete, agent-based representation of individual cells, whose behavior (division, migration, death) depends on the local value of the continuous fields. Anderson and Chaplain used this combination to model capillary network branching in angiogenesis, where the movement of individual endothelial cells is discrete but guided by a continuous gradient of angiogenic factors and fibronectin [5]. Similarly, Anderson later coupled discrete tumor cells with continuous fields of extracellular matrix, degrading enzymes, and oxygen when modeling invasion (Section 5) [4]. Norton and colleagues, in their review of agent-based and hybrid models of the tumor immune microenvironment, emphasize that this very combination makes it possible to capture the discrete, stochastic nature of individual-cell decision-making (e.g., the choice between proliferation and migration, Section 10) without losing the continuous, diffusion-driven context of the microenvironment [56].

11.4. Linking ODE and agent-based systems

In addition to the spatially explicit hybrid models described above, many applications couple an agent-based model of a cell population with a non-spatial system of ordinary differential equations that describes rapidly equilibrating quantities (e.g., the pharmacokinetics of a drug in the blood or the systemic immune response, Section 8) without requiring their explicit spatial representation. Norton and colleagues document this strategy in models of the response to immunotherapy, in which the local, spatially resolved interaction of tumor and immune cells is represented by an agent-based model, while the systemic dynamics of the drug or cytokines enters as a shared, spatially homogeneous input governed by a system of ODEs [56]. The choice of which model component

requires spatial structure and which can be aggregated without loss of accuracy is a key modeling decision that affects computational cost, as discussed in the following section.

11.5. Calibration and computational cost

Agent-based and hybrid models have computational cost that grows with the number of simulated cells, which makes calibration to experimental or clinical data (typically requiring thousands of model evaluations at different parameter values) computationally demanding or even infeasible. Lima and colleagues addressed this problem for a model of in vitro tumor growth by grouping individual cells into “coarse-grained" agents representing entire cell clusters rather than individual cells, thereby reducing computation time by 93 to 97 % in Bayesian calibration to time-lapse microscopy data while maintaining predictive accuracy within 3 % of the original, fully resolved agent-based model [49]. This approach illustrates a general strategy for reducing the computational cost of multiscale models – reducing the degrees of freedom where fine cellular structure is not essential for the given predictive task – rather than uniformly simplifying the model.

11.6. Verification of multiscale simulations

Verification (agreement of the implementation with the model’s mathematical specification), validation (agreement of the model with reality), and uncertainty quantification (VVUQ) are more demanding for multiscale models than for single-scale models, because errors and uncertainties from individual scales can accumulate or mask one another across the couplings between them. Kemkar and colleagues proposed a protocol for verifiable oncological digital twins at the tissue level that explicitly specifies the VVUQ steps required before a multiscale tissue model can be regarded as a trustworthy basis for clinical decision-making (Section 15), including the systematic propagation of uncertainty across the model’s individual coupled scales [43]. This requirement is a direct extension, to the specific tissue scale of modeling, of the discussion of prospective validation of digital twins in Section 16.

12. Genomics and bioinformatics

Molecular data make it possible to characterize clones, regulatory programs, and the microenvironment, but they require careful quality control and correction for technical biases. Genomics and bioinformatics form the data foundation on which both evolutionary models of the tumor (Section 6) and models of the microenvironment (Section 5) rest, as well as the data-driven and hybrid approaches discussed in Sections 13, 14, and 16. This section is therefore not

merely a catalogue of laboratory methods, but also describes how individual types of molecular data enter as parameters, constraints, or calibration targets into the mathematical models presented in the preceding sections.

Genetic and genomic analyses are among the decisive sources of information for personalized treatment. Their role is not merely to describe individual mutations, but also to identify reproducible gene signatures and biologically interpretable patient groups. An example is the COMBING method, which combines clustering of oncological data (a data-driven approach, Section 3) with the mathematical and biological identification of new gene signatures [9].

The value of genomics increases substantially when integrated with other -omic layers. The Molecular Twin platform combined 6 363 features from six analytes (clinical and surgical-pathological data, DNA, RNA, tissue and plasma proteomics, lipidomics, and computational pathology) in 74 patients with resectable pancreatic ductal adenocarcinoma; the resulting multi-omic model achieved an accuracy of 0.85 (95 % CI 0.73–0.96) and outperformed every individual -omic model, confirming the hierarchical complementarity of the individual molecular layers in predicting survival [58]. This complementarity is the reason why the following sections describe the individual -omic and bioinformatic methods separately, but always with an emphasis on how their outputs can be combined.

12.1. DNA sequencing and somatic variants

Whole-genome or whole-exome sequencing of a tumor relative to the corresponding normal tissue makes it possible to identify somatic variants – point mutations, small insertions and deletions, as well as large-scale structural rearrangements – arising exclusively in tumor cells. The Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium, combining ICGC and TCGA data, analyzed 2 658 whole tumor genomes and their corresponding normal tissue across 38 tumor types and found that the average tumor contains 4 to 5 driver mutations when both coding and non-coding genomic elements are jointly considered, with no driver mutation identified in approximately 5 % of cases; this suggests that the catalogue of driver events responsible for tumor development is still incomplete [40]. The distinction between driver and passenger mutations identified in this way is a direct input for branching processes modeling clonal expansion (Section 6) [11], which presuppose precisely this division of mutations into those that alter cell fitness and those that are selectively neutral.

12.2. RNA-seq

Transcriptome sequencing (RNA-seq) measures the expression level of individual genes and their splice variants with substantially higher resolution and dynamic range than older microarray platforms [79]. In oncology, RNA-seq provides two types of information: a quantitative expression profile, which serves

as input for pathway analysis (below) and for classifiers of molecular subtypes, and it also enables the detection of events visible only at the transcript level – for example, gene fusions or alternative splice isoforms, which DNA sequencing alone cannot reveal. The PCAWG consortium exploited precisely this complementarity when it combined whole-genome data with transcriptomic data from the same tumor to characterize the effects of somatic mutations on splicing, expression level, fusion genes, and promoter activity [40].

12.3. Mutational signatures

Different mutagenic processes (e.g., exposure to tobacco smoke, defective DNA repair, homologous recombination deficiency) leave characteristic, statistically distinguishable patterns of substitutions and other variants in the tumor genome, known as mutational signatures. Alexandrov and colleagues analyzed 84 729 690 somatic mutations from 4 645 whole-genome and 19 184 exome sequences and identified 49 single-base substitution signatures, 11 doublet-base substitution signatures, 4 clustered substitution signatures, and 17 small insertion and deletion signatures; the large scale of the data made it possible to separate even mutually overlapping signatures and decompose them into components corresponding to distinct, yet related, mechanisms of DNA damage, repair, and replication [3]. A mutational signature present in a tumor thus carries information not only about the history of mutagenic exposure, but often also about the functional state of DNA repair pathways (e.g., a homologous recombination deficiency signature as an indicator of sensitivity to PARP inhibitors), and therefore represents a more biologically interpretable input for patient classification than a raw list of mutations.

12.4. Pathway and regulatory network analysis

A list of statistically significantly differentially expressed genes is, by itself, difficult to interpret biologically, especially when the individual genes overlap only partially between independent studies of the same disease. Instead of evaluating individual genes, Subramanian and colleagues proposed the Gene Set Enrichment Analysis (GSEA) method, which assesses whether predefined groups of functionally or regulatorily related genes (e.g., an entire signaling pathway) are concentrated at one end of a ranked list of genes ordered by correlation with phenotype, rather than testing the statistical significance of genes individually. Using lung cancer survival data, the authors showed that while gene-level analysis finds only limited agreement between two independent studies, GSEA at the level of entire pathways reveals many shared biological processes [72]. This shift from the level of the individual gene to the level of the pathway or regulatory network is also a natural link to the network models described in Section 3,

where nodes represent genes or proteins and edges represent their regulatory or interaction relationships.

12.5. Single-cell sequencing

Bulk sequencing measures the average expression or mutational profile across an entire tumor sample and thus obscures the heterogeneity among individual cells that Section 6 identifies as a key driving force of progression and resistance. Single-cell RNA sequencing (scRNA-seq) removes this limitation by measuring expression in each cell separately. Tirosh and colleagues profiled 4 645 individual cells from 19 patients with metastatic melanoma in this way and revealed that tumor cells within the same tumor occupy two distinct transcriptional states (high vs. low expression of the transcription factor MITF, with reciprocally increased expression of the kinase AXL in the latter state), with the low-MITF state enriched after treatment, suggesting its association with treatment resistance [75]. This type of finding – the existence of discrete, treatment-selected transcriptional states within a single tumor – is direct experimental support for the models of intratumor heterogeneity and phenotypic plasticity discussed in Section 6, as well as for the classification of cells in calibrating agent-based models of the microenvironment (Section 11).

12.6. Spatial transcriptomics

Single-cell sequencing, as described above, captures cellular heterogeneity, but dissociating tissue into individual cells loses information about their original spatial location. Spatial transcriptomics preserves this information: methods based on next-generation sequencing (e.g., 10X Visium, Slide-seq, Stereo-seq) achieve a spatial resolution on the order of 0.2 to 55 µm, while methods based on fluorescence in situ hybridization (MERFISH, seqFISH+) achieve subcellular resolution; combined with spatial proteomics (CODEX, t-CycIF, IMC, MIBI), this makes it possible to directly map the compartmentalization of tumor and immune cells within the tumor microenvironment [39]. This class of data represents a natural bridge between genomics (this section) and modeling of the tumor microenvironment (Section 5): rather than an abstract reaction-diffusion field $c(\boldsymbol{x}, t)$, it provides a directly observed, spatially resolved molecular characterization of tissue that can be used to calibrate or validate spatially explicit mechanistic models. An overview of the possibilities for integrating spatial transcriptomics with foundation models and other -omic layers is discussed in more detail in Section 16.

12.7. Open databases

Key resources include TCGA, TCIA, cBioPortal, and GEO; for each use, the version, access date, and cohort selection criteria should be recorded. cBioPortal

is an illustrative example of how an open database lowers the barrier between raw -omic data and their interpretation: the portal aggregates mutational, copy-number, expression, and clinical data from dozens of large studies into a unified interface that enables interactive browsing of genetic alterations across samples, genes, and pathways, as well as linking them to clinical outcomes, without requiring bioinformatic expertise [13]. Data aggregated in this way are also a direct input for the open databases cited in Section 16 in the context of FAIR principles [82], since the value of an individual database grows precisely with its interoperability with other resources (TCIA for imaging data, GEO for expression data outside TCGA).

A fragmentation similar to that seen in genomic data also affects open imaging data. A large-scale review of more than 1000 publicly available medical imaging datasets found that they are unevenly distributed across modalities and tasks, and that datasets combining imaging with genomic data are rare in the public domain [21]. As a partial solution, the authors propose metadata-driven fusion, which merges small datasets according to harmonized ontologies without the need to share raw data – an approach analogous to the cataloguing and harmonization of metadata in the genomic databases discussed above.

13. Artificial intelligence in oncology

Artificial intelligence methods complement mechanistic models for imaging, molecular, and clinical data. Evaluation must separate internal validation from genuine generalization to a new site or population. This section builds on the classification of data-driven models from Section 3 and on the genomic and imaging data from Section 12, which form the input for the methods described below; it is connected to digital twins (Section 14) and clinical implementation (Section 15) by the shared question of whether a model is not only accurate but also trustworthy and clinically usable.

Statistical and computational models make it possible to process large volumes of patient data and also support drug discovery. Combining mathematical methods with artificial intelligence can accelerate the identification of therapeutic candidates and improve prediction of treatment response, although the quality of the conclusions still depends on the data, the validation design, and biological interpretation [1].

13.1. Classical machine learning

13.1.1. Support vector machines

A support vector machine (SVM) seeks a hyperplane that separates two classes with maximum margin; by nonlinearly mapping the input features into a higher dimension (the so-called kernel trick), the same principle can also be applied to linearly non-separable data [17]. Thanks to good generalization with a relatively small number of training samples – a typical situation for oncological cohorts on the order of tens to hundreds of patients – SVM remains a widely used method today for classification based on genomic or radiomic features (Section 12), although on large and heterogeneous omics data it is generally outperformed in accuracy by the ensemble learning methods discussed below.

13.1.2. Random forest and XGBoost

Random forests repeatedly prove in practice to be a robust choice for tabular and mixed omics data. For example, the Molecular Twin multi-omics platform for pancreatic ductal adenocarcinoma used seven types of classifiers over combinations of clinical, genomic, transcriptomic, proteomic, lipidomic, and computational pathology features, and the random forest was, across both individual and combined analytes, most frequently the best-performing model [58].

13.2. Deep learning

13.2.1. Convolutional neural networks

Convolutional neural networks (CNNs) learn a hierarchy of local image features directly from pixels, without the need for hand-crafted features. Esteva and colleagues trained a single CNN end-to-end on 129 450 clinical images of skin lesions (2 032 different diagnoses) and achieved classification of malignant versus benign lesions at the level of board-certified dermatologists, thereby demonstrating that deep learning can achieve expert-level diagnostic accuracy directly from the image, without an intermediate step of explicit segmentation or feature extraction [25]. This 2017 result became a reference point for later architectures (including the vision transformers discussed below) in image-based oncological diagnostics.

13.2.2. Vision transformers

The Vision Transformer (ViT) replaces convolutional layers with an attention mechanism originally developed for text processing: the image is divided into a grid of fixed patches (e.g., 16×16 pixels), which are linearly embedded and

processed as a sequence, analogous to a sequence of words. Dosovitskiy and colleagues showed that such a purely transformer-based model, given a sufficient scale of training data, achieves results on classification tasks comparable to or better than convolutional networks, with lower computational requirements for training [22]. In digital pathology and radiology (Section 12), the ViT architecture today underlies many foundation models (below), since the attention mechanism naturally captures even distant spatial relationships in whole-slide images, which are harder for locally constrained convolutional kernels to access.

13.2.3. Graph neural networks

Graph neural networks (GNNs) represent data as nodes and edges (e.g., a drug molecule, a protein–protein interaction network, or cells in spatial transcriptomics, Section 12) and learn representations by aggregating information from neighboring nodes. Gogoshin and Rodin, in their review of GNNs in oncology research, document their growing application in predicting drug sensitivity by combining the graph of a drug's molecular structure with gene features (expression, mutations), in predicting the synergy of drug combinations, and in repositioning existing drugs, while also highlighting the lack of independent and comprehensive frameworks for benchmarking different GNN architectures across studies [34]. The ability of GNNs to naturally integrate structural, genomic, and pharmacological information within a single model makes them a suitable tool precisely for the network models described in Section 3.

13.3. Foundation models

The development of foundation models for medical imaging is limited by the scale and fragmentation of publicly available data. A review of more than 1000 freely available imaging datasets (2000 – 2025) found that 2D images make up 84.6 % of the corpus (pathology 18.7 %, X-ray 17.1 %, CT 13.1 %, MRI 10.4 %), with image generation tasks accounting for as much as 59 % of recently published images compared to 24.1 % for classification and 14.4 % for segmentation; even the largest public resources (e.g., AbdomenAtlas, CT-RATE) remain orders of magnitude smaller than the corpora used to train large language models. As a partial solution to this fragmentation, metadata-driven fusion (MDFP) has been proposed, which merges small datasets into larger, coherent corpora based on metadata harmonization according to ontologies (e.g., UMLS, MeSH) without the need to share the image data itself, thereby circumventing some of the licensing and privacy-protection obstacles [21].

13.3.1. Language and multimodal models

Multimodal large language models connect perception and multi-step reasoning over heterogeneous inputs [85]. In oncology, their potential role is the integration of textual, imaging, and molecular data for clinical decision support; however, such use requires separate clinical validation.

13.3.2. AI assistants

Alongside foundation models for imaging and molecular data, language models are increasingly being tested in clinical practice as direct assistants in oncological decision-making – summarizing records, extracting structured data from free text, or directly supporting a diagnostic or treatment decision. A systematic review and meta-analysis of 56 studies across 15 tumor types found an average overall accuracy of such tools of 76.2 %, but an average diagnostic accuracy of only 67.4 %, with most evaluations relying on automated metrics without human raters, and almost none systematically addressing safety, potential harm, or the comprehensibility of the output for the patient [38]. This gap between accuracy in summarization and lower reliability in diagnostic reasoning is a concrete illustration of the more general warning from the next section of this section – that a model's technical accuracy on a benchmark does not guarantee its trustworthiness in real clinical decision-making.

13.4. Explainable artificial intelligence

The more complex a model is (deep networks, tree ensembles), the less transparent the relationship between its inputs and its output, which is problematic in clinical decision-making regardless of the accuracy achieved. Lundberg and Lee unified several previously independent explainability methods into a single framework, SHAP (SHapley Additive exPlanations), which assigns to each input feature a contribution to a given prediction derived from cooperative game theory (Shapley values), and proved that within the class of so-called additive feature-importance methods this is the unique solution satisfying the desired theoretical properties (local accuracy, consistency) [50]. In the oncological context, such methods make it possible, for example, to verify whether a classifier based on radiomic or genomic features (Section 12) actually uses biologically meaningful features or relies on artifacts of the data collection – a necessary but not sufficient step toward the external validation discussed below.

13.5. Federated learning

Training a robust model requires data from many sites, but such data are typically locked within individual institutions due to privacy and regulatory constraints (Section 16). Federated learning addresses this problem by training

the model locally on each site's data, with only the model parameter updates shared between sites, not the raw patient data. Rieke and colleagues, in their review of this approach for digital health, emphasize that even such shared model updates can, under certain circumstances, reveal information about the original data, and federated learning is therefore typically complemented by additional privacy-protection techniques (e.g., differential privacy or secure aggregation) rather than replaced by them [63]. This approach is a direct alternative to the centralized data aggregation in the open databases described in Section 12, for situations where centralizing the data itself is not possible for legal or practical reasons.

13.6. Digital pathology

In terms of open data, digital pathology is currently the best-covered imaging modality. Across 117 public datasets, drawn from grand challenges (e.g., MICCAI, ISBI), open data platforms (e.g., TCGA, TCIA), and independent research collections, roughly 2.22 million images are available – predominantly tiles extracted from whole-slide images stained with hematoxylin and eosin, complemented by a smaller number (approximately 67 000) of full whole-slide images. Quilt-1M and PatchCamelyon are the two largest such collections, contributing 1,000,000 and 328,000 images, respectively. Some 82 % of these datasets carry high-quality supervised annotations, yet the sheer cost of manually annotating gigapixel-scale whole-slide images has pushed the field toward self-supervised pretraining on much larger, unlabeled archives. This shift underlies the largest current foundation models for pathology: UNI and Prov-GigaPath are trained on private collections of, respectively, over 100 000 and over 171 000 whole-slide images, well beyond what is publicly available [21].

13.7. Bias, uncertainty, and external validation

Verifying individual steps of reasoning is a separate research problem. The Hard2Verify benchmark examines verification of open-ended mathematical problems [59]; it is not an oncological clinical benchmark, but it illustrates the need to distinguish a correct result from a trustworthy process. In clinical applications, calibration, robustness, generalization, and impact on decision-making must additionally be verified.

Another source of bias is the very structure of the available training data. Datasets that combine imaging with genomic data and clinical text records are extremely rare in the public domain, and most existing collections lack contextual and temporal information about disease course, which limits the models' ability to generalize to real clinical workflows [21]. An example of good validation practice, by contrast, is the Molecular Twin platform, whose predictive models were independently validated on four external cohorts (TCGA, two Johns

Hopkins University cohorts, and Massachusetts General Hospital) rather than relying solely on internal cross-validation [58].

14. Digital twins

A patient's digital twin is a continuously updated computational model designed for prognosis and for comparing possible interventions. It differs from a one-off simulation in being tied to a specific patient's longitudinal data. This section links mechanistic models of tumor growth and the microenvironment (Sections 4–5) with molecular and imaging data (Section 12) and artificial intelligence methods (Section 13) into a single, patient-specific framework; its extensions toward prospective clinical validation and foundation models are discussed in more detail in Section 16.

14.1. Definition and architecture

An example of a computational patient representation oriented toward outcome prediction is the Molecular Twin platform, which integrates an individual patient's clinical, genomic, transcriptomic, proteomic, lipidomic, and computational pathology data from pancreatic ductal adenocarcinoma into a single multidimensional survival prediction model [58]. This is an architecture oriented more toward the statistical prediction of clinical outcome than toward the mechanistic simulation of physiological processes, illustrating the broader spectrum of approaches referred to as digital twins. At the opposite end of this spectrum lie mechanistic digital twins, which, instead of (or alongside) statistical prediction, explicitly simulate physiological processes – for example, tumor growth and invasion using reaction-diffusion models (Section 4) – and whose parameters are continuously personalized and updated with data from a specific patient using the procedures described in the following two sections.

14.2. Parameter personalization

The mechanistic tumor growth model (Section 4) contains parameters (proliferation rate, invasion diffusion coefficient) that, in their general form, describe a population of patients, whereas a digital twin requires their values for a specific individual. Yankeelov and colleagues showed how these parameters can be estimated by solving an inverse problem: from quantitative imaging data (e.g., diffusion-weighted MRI, which indirectly measures cell density, or perfusion imaging, which measures blood supply), local values of the quantities entering the model are estimated on a voxel-by-voxel basis, thereby calibrating the general growth or reaction-diffusion model to a specific patient instead of using population-averaged parameters [83]. The model personalized in this way

can subsequently be used to predict the further course of disease or response to treatment in the same patient, which is a prerequisite for the virtual comparison of treatment scenarios described below.

14.3. Assimilation of longitudinal data

The personalization described above typically takes place as a one-time process, based on data measured at a single time point. A true digital twin, however, should be continuously updated as new measurements accumulate during treatment, which requires a formal apparatus for data assimilation. Kostelich and colleagues applied a local ensemble transform Kalman filter (a method originally developed for numerical weather prediction) to mathematical models of glioblastoma, which accounts for both the uncertainty in model parameters and the noise in MRI measurements, and showed that a model updated in this way can accurately track the growth of a synthetic tumor over six sixty-day prediction-update cycles even in the presence of systematic model error [44]. Filtering of this type formally resolves the trade-off between confidence in the (imperfect) mechanistic model and confidence in the (noisy) measurement, rather than simply recomputing the model from scratch after each new measurement.

14.4. Uncertainty quantification

A personalized and continuously updated model is clinically usable only if its uncertainty is quantified. Otherwise, it is impossible to distinguish a confident but incorrect prediction from a prediction with an appropriately wide confidence interval. Uncertainty quantification for agent-based and hybrid models (Section 11) is a computationally demanding task. One approach reduces the model's degrees of freedom prior to Bayesian calibration [49]. At the level of an entire tissue-scale digital twin, an alternative is a systematic verification, validation, and uncertainty quantification (VVUQ) protocol that explicitly tracks the propagation of uncertainty across the model's individual coupled scales before the model is considered a trustworthy basis for clinical decision-making [43]. Without this step, the digital twin remains merely a technically functional simulation, not a tool supporting a specific clinical decision.

14.5. Virtual comparison of treatment scenarios

One of the main clinical motivations for a digital twin is the possibility of comparing multiple treatment approaches *in silico* on the same (personalized) patient model before deciding which one to actually administer. In this context, Moingeon and colleagues distinguish virtual patients from a digital twin tied to a specific, real patient. A virtual patient is a model parameterization that generates physiologically plausible outputs, derived from the statistical distribution of characteristics of a real population rather than from any single individual.

Together, both approaches form the basis for so-called in silico clinical trials, i.e., a simulated comparison of treatment arms before (or alongside) an actual clinical trial (Section 15) [54]. A virtual patient cohort thus makes it possible to estimate the difference between treatment strategies (e.g., an alternative radiotherapy fractionation schedule, Section 8) before such a comparison could be ethically or practically carried out in real patients.

14.6. Technical, ethical, and clinical limitations

Alongside the technical challenges described above (personalization, data assimilation, uncertainty quantification), the use of digital twins also brings limitations that are not purely mathematical. Drummond and Coulet, in a review focused on digital twins in chronic diseases of children, systematically distinguish four categories of obstacles. Technical obstacles include, in particular, the difficulty of continuously collecting high-quality data in pediatric patients. Ethical obstacles concern the need to preserve the child's place in decisions about their own care. Legal obstacles arise from the simultaneous requirement to guarantee both the system's safety and equal access to it. Social obstacles include the risk of weakening direct human contact between patient and physician, and the broader risk of normalizing continuous monitoring [23]. Although this classification arose in the context of pediatric chronic diseases, the same four categories of obstacles – technical, ethical, legal, and social – are directly relevant also to the barriers to prospective clinical validation of oncological digital twins discussed in Section 16 [32].

15. Clinical implementation

A clinically usable model must have a predetermined purpose, target population, output, and the decision it is meant to support. Model performance is only one part of the evidence for its usefulness. This section therefore closes the transition from the model (Sections 4–11), through data and machine learning methods (Sections 12–13), and the digital twin (Section 14), to the question of exactly what must be demonstrated before such a model is used in the care of an individual patient.

15.1. Personalized and precision oncology

Parsimonious models with a substantially reduced number of features can achieve accuracy comparable to full multi-omic models, which lowers testing costs and opens a path toward the democratization of precision oncology. In pancreatic ductal adenocarcinoma, a parsimonious model using 589 of the original 6 363 multi-omic features achieved the same accuracy (0.85) as the full model,

while the plasma protein alone was a better preoperative predictor of survival (accuracy 0.75; PPV 0.80) than the routinely used biomarker CA 19-9 (accuracy 0.59; PPV 0.53) [58]. Personalizing treatment on the basis of such models is, at the same time, the clinical expression of the personalization of digital twin parameters described in Section 14 – in both cases, the task is to derive individual, patient-specific quantities from a general population model.

15.2. Design of clinical trials

Statistical models help evaluate the efficacy and safety of new therapies and quantify differences in response among patient populations. The analytic plan, endpoints, and subgroup analyses must therefore be specified before the results are interpreted. The molecular stratification of patients described in the previous section also makes it possible to depart from the classical one-disease–one-treatment design. So-called *basket* trials test a single targeted therapy across multiple histological tumor types that share a common molecular alteration. *Umbrella* trials, conversely, assign patients within a single tumor type to different treatments according to their individual molecular profile. Both types can be combined into an adaptive *master protocol* with a continuously changing number of treatment arms [60]. These designs are both statistically and logistically more demanding than classical trials with a fixed number of arms. However, they better reflect the reality that molecularly defined patient subgroups (Section 12) may be too small to be separately enrolled in a classical, histologically defined trial. An analogous, evolutionarily motivated adaptation of trial design to a continuously measured patient response is also illustrated by the prospective trial of adaptive therapy for prostate cancer discussed in Section 10 [84].

15.3. Verification, calibration, and validation of models

The technical verification of an algorithm, the calibration of its outputs, and clinical validation are distinct tasks, and conflating them is a common source of overstated claims about a model's readiness for practice. Verification asks whether the software correctly solves the mathematical model it claims to implement – a question of numerical and programming correctness, independent of whether the model itself is biologically or clinically meaningful (Section 11). Calibration fits the model's free parameters to a training dataset so that its outputs match observations in that same dataset; a well-calibrated model is, by construction, accurate on the data used to calibrate it, which says little about its behavior elsewhere. Clinical validation asks a different question again: whether the model's predictions hold on data it has not seen and, ultimately, whether acting on those predictions leads to better patient outcomes than the available alternative. The distinction between internal validation (e.g., cross-validation or a held-out split of the same cohort) and external validation (an independent

cohort, ideally from a different site, population, or time period) is therefore critical: a model can be well calibrated, and even validate well internally, while still failing on external data because of shifts in patient population, measurement protocol, or disease prevalence. The reliability of a mathematical or AI model on open, benchmark-style tasks does not by itself demonstrate the safety of its use in a clinical setting; only external validation, and ultimately a demonstration that acting on the model's predictions improves clinical decisions – the subject of the following section – can do so.

An example of proper external validation is the Molecular Twin platform, whose predictive models of survival in pancreatic cancer, after being trained on a pilot cohort, were independently validated on four external datasets (TCGA, two Johns Hopkins University cohorts, and a Massachusetts General Hospital cohort), with the predictive value of the plasma protein confirmed across all cohorts [58]. Verification and validation are treated jointly with uncertainty quantification for multiscale and agent-based models under the VVUQ framework introduced in Section 11, since for such models errors in numerical implementation, in calibration, and in the propagation of parameter uncertainty are difficult to disentangle from one another [43]. The regulatory and reporting standards that formalize the distinction between training, calibration, and validation data are addressed later in this section [77, 16].

15.4. Clinical utility and decision thresholds

A model with high discriminative ability (e.g., AUC) need not be clinically useful if its predictions do not lead to better decisions than simple default strategies (treat everyone, or treat no one). Vickers and Elkin responded to this shortcoming with the method of decision curve analysis. For a given range of clinically reasonable decision thresholds (e.g., what probability of malignancy a physician is willing to accept before recommending a biopsy), the method calculates the net benefit of the model compared with these default strategies. Unlike full decision analysis, it requires only the dataset on which the model is tested [78]. Decision curve analysis thus supplements statistical accuracy metrics with the question of whether the model is useful precisely at the decision threshold relevant to the specific clinical situation. This is a stricter requirement than mere statistical significance of discrimination.

15.5. Regulatory science

Software that provides a diagnostic, prognostic, or treatment-related output based on patient data is regulated as a medical device (Software as a Medical Device, SaMD). Machine-learning-based models place particular demands on this regulatory framework, owing to their ability to change after deployment. In 2021, the FDA, together with Health Canada and the UK's MHRA, therefore

formulated ten Good Machine Learning Practice (GMLP) principles [77]. In addition to requirements for the quality of training and validation data and for transparency toward users, these principles also introduce the concept of a predetermined change control plan. This is a predefined scope and manner in which the model may be updated after deployment, without the need for a new regulatory approval application for each modification. This framework directly targets models with continuously updated parameters, such as the digital twins described in Section 14. For such models, a one-time approval of a static version does not correspond to their actual, continuously changing behavior.

15.6. Reproducibility and software management

A predictive model described only in words, without available code and a precise specification of the data on which it was trained and validated, cannot be independently reproduced or fairly compared with other approaches. A response to this problem is the TRIPOD+AI guideline, which updates the original TRIPOD reporting standard so that it uniformly covers predictive models based on both classical regression and machine learning, and which explicitly requires the disclosure of source code, a precise description of data preprocessing, and a clear distinction between the training, validation, and test sets [16]. Consistent adherence to such a standard is a practical precondition for the external validation discussed above: without exact reproducibility of the original model, it is impossible to distinguish whether a given failure on external data reflects a genuine lack of generalization or merely an inconsistent reimplementation.

15.7. FAIR data, privacy, and ethics

A survey of more than 1000 publicly available medical imaging datasets points out that restrictive licensing and privacy regulations (e.g., HIPAA, GDPR) fragment collaboration and hinder the sharing of data even in anonymized or synthetic form [21]. Harmonizing metadata according to standardized ontologies can partially mitigate this problem, since it enables data to be merged and searched without the need to share the actual image files. Federated learning, described in Section 13, represents a complementary, technical solution to the same problem – instead of harmonizing metadata for sharing the data itself, it allows sharing only the updates of a model trained locally at each site.

16. Future directions

The future of the field will depend on the trustworthy integration of biological mechanisms, multimodal data, and clinical decision-making.

16.1. AI and mechanistic models

The synthesis of mechanistic models and machine learning is formalized in mathematical oncology under the term *mechanistic learning* – a synergistic combination of knowledge-driven mathematical modeling with data-driven machine or deep learning. A comprehensive review of this field distinguishes four modes of integration: *sequential*, in which data-driven and knowledge-driven models are applied successively, with the output of one feeding into the other; *parallel*, in which both approaches address the same objective independently and their results are subsequently compared or combined; *extrinsic*, in which learning is applied to process the outputs of the mechanistic model or to separate sub-aspects of the same problem; and *intrinsic*, in which biological or physical knowledge is embedded directly into the architecture or training process of the learning algorithm, for example in the form of physics-informed neural networks [53]. An example of an earlier, more broadly biomedically oriented formulation of the same principle is the integration of machine learning with multiscale modeling, which emphasizes that machine learning alone disregards physical and biological laws and may lead to non-physical solutions, whereas mechanistic models without data calibration generalize poorly to new situations [2]. The review of mechanistic learning in oncology also highlights three main obstacles to broader clinical adoption: limited availability and heterogeneity of clinical data, the requirement for model transparency and interpretability necessary for clinical decision-making, and the complexity of integrating multimodal inputs across genomics, imaging, and clinical records (Sections 12 and 13) [53].

16.2. Digital twins in prospective studies

Despite the extensive conceptual development described in Section 14, prospective clinical validation of oncological digital twins remains rare. A recent umbrella review of reviews on digital twins in oncology identifies fragmentation and heterogeneity of clinical data, the lack of verification of agreement between predictions and real-world outcomes across diverse clinical settings, biological and interindividual tumor heterogeneity, and the high computational demands of the models as the main barriers to clinical adoption [32]. The transition from retrospective calibration to prospective use therefore requires not only a more accurate mechanistic model core (Sections 4–11), but also a predefined clinical objective, standardized collection of longitudinal data, and independent validation on external cohorts, analogous to the approach of the Molecular Twin platform (Section 15) [58]. Without such validation, the gap between a technically functional model and a clinically useful digital twin remains unresolved.

16.3. Multi-omics and spatial data

Whereas the genomic approaches described in Section 12 rely largely on bulk or single-cell sequencing without spatial context, spatial multi-omics enables molecular information to be mapped directly onto the histological structure of tumor tissue. Spatial transcriptomics includes methods based on next-generation sequencing (e.g., 10X Visium, Slide-seq, Stereo-seq) with a resolution of approximately 0.22 to 55 µm, and methods based on fluorescence in situ hybridization (MERFISH, seqFISH+), which achieve subcellular resolution; spatial proteomics (e.g., CODEX, t-CycIF, IMC, MIBI) simultaneously detects 40 to 100 proteins in the same tissue section. Combining these data with single-cell RNA sequencing makes it possible to map cell types back onto their spatial context and to reveal the compartmentalization of tumor and immune cells, markers of T-cell exhaustion, and prognostically significant spatial patterns in breast, colon, and liver carcinoma [39]. The spatially resolved microenvironment characteristics obtained in this way (Section 5) represent a natural input for calibrating spatially explicit mechanistic tumor growth models and for multimodal foundation models.

16.4. Foundation models for precision oncology

The development of foundation models in oncology builds on the discussion in Section 13. An example is the UNI model, a self-supervised foundation model for digital pathology pretrained on more than 100 million image tiles from over 100 000 hematoxylin-and-eosin-stained whole-slide images (>77 TB of data) spanning 20 major tissue types, which, after fine-tuning, outperforms specialized models in classification, segmentation, and subtype estimation tasks across multiple oncological diagnoses [14]. Such models function as general-purpose feature extractors, which can subsequently be combined with clinical, genomic, or proteomic data in the spirit of multi-omics predictive platforms (Section 15) or fine-tuned for narrowly specific clinical tasks. Their scalability is, however, directly limited by the scope and fragmentation of publicly available imaging data (Section 13), which motivates the need for shared data infrastructures discussed in the following sections.

16.5. Open science and FAIR principles

Systematic data sharing and reuse is a precondition for both training and independent validation of the models described in the preceding sections. The FAIR principles – Findable, Accessible, Interoperable, and Reusable – formulate concrete, measurable requirements for the management of scientific data, with an emphasis on their machine-actionability, that is, the ability of computational systems to find, retrieve, link, and use data with minimal human intervention [82]. In the oncological context, these principles directly underpin the functioning of

the open databases mentioned in Section 12 (TCGA, cBioPortal, GEO) as well as imaging archives (Section 13). Fulfilling the FAIR principles is not, however, equivalent to unrestricted open access: restrictive licensing and privacy protection legislation (e.g., HIPAA, GDPR) are compatible with the FAIR principles even for data that remain under controlled access, provided that the metadata, access conditions, and anonymization procedures are transparently documented (Section 15) [21].

16.6. Standards, benchmarks, and shared infrastructure

A practical precondition for fulfilling the FAIR principles is the existence of shared technical infrastructure. The Cancer Imaging Archive (TCIA) illustrates how a publicly funded repository can standardize the format, annotations, and access conditions for large collections of oncological imaging data – already in its first year of operation it assembled 23 collections comprising approximately 3.3 million images [15]. A similar function for molecular data is fulfilled by the TCGA, cBioPortal, and GEO databases (Section 12). Without comparable standards for model benchmarking – that is, agreed-upon reference datasets, metrics, and external validation protocols (Section 15) – however, data availability alone does not guarantee comparability of results across institutions. The need to distinguish a technically correct model output from its trustworthiness in open, uncontrolled tasks has also been emphasized outside of oncology [59]; in oncological practice, an analogous requirement for shared reference frameworks applies, with the difference that the final criterion is not merely the correctness of the output, but its clinical utility.

16.7. Conclusion

The systematic development of mathematical oncology requires not only more accurate models, but also transparent assumptions, high-quality validation, and close collaboration with experimental and clinical teams. None of the modeling paradigms surveyed in this book – population growth laws, spatial and multiscale mechanistic models, evolutionary and game-theoretic descriptions of resistance, optimal control of treatment, or data-driven and foundation models – is sufficient on its own; their value lies in how they are combined and validated for a specific clinical question. Realizing this potential will depend less on any single mathematical or computational advance than on sustained collaboration across mathematics, biology, data science, and clinical medicine, guided throughout by the same standard: that a model's contribution to patient care, not merely its technical sophistication, is what ultimately counts.

REFERENCES

[1] AGRAWAL, A.—KUMAR, R.—MAHESHWARI, S.—SHARMA, M.—BATRA, C. M.—SHARMA, S.—GAUR, V.: *Mathematical and artificial intelligence techniques in modern drug discovery: A review*, Drug Dev. Res. **87** (2026), e70212.

[2] ALBER, M.—TEPOLE, A. B.—CANNON, W. R.—DE, S.—DURA-BERNAL, S.—GARIKIPATI, K.—KARNIADAKIS, G.—LYTTON, W. W.—PERDIKARIS, P.—PETZOLD, L.—KUHL, E.: *Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences*, npj Digit. Med. **2** (2019), Art. 115.

[3] ALEXANDROV, L. B.—KIM, J.—HARADHVALA, N. J.—HUANG, M. N.—NG, A. W. T.—WU, Y.—BOOT, A.—COVINGTON, K. R.—GORDENIN, D. A.—BERGSTROM, E. N. et al.: *The repertoire of mutational signatures in human cancer*, Nature **578** (2020), 94–101.

[4] ANDERSON, A. R. A.: *A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion*, Math. Med. Biol. **22** (2005), 163–186.

[5] ANDERSON, A. R. A.—CHAPLAIN, M. A. J.: *Continuous and discrete mathematical models of tumor-induced angiogenesis*, Bull. Math. Biol. **60** (1998), 857–899.

[6] AXELROD, R.—AXELROD, D. E.—PIENTA, K. J.: *Evolution of cooperation among tumor cells*, Proc. Natl. Acad. Sci. USA **103** (2006), 13474–13479.

[7] BARROS, L. R. C.—PAIXÃO, E. A.—VALLI, A. M. P.—NAOZUKA, G. T.—FASSONI, A. C.—ALMEIDA, R. C.: *CARTmath—A mathematical model of CAR-T immunotherapy in preclinical studies of hematological cancers*, Cancers **13** (2021), Art. 2941.

[8] BASANTA, D.—HATZIKIROU, H.—DEUTSCH, A.: *Studying the emergence of invasiveness in tumours using game theory*, Eur. Phys. J. B **63** (2008), 393–397.

[9] BATTISTELLA, E.—VAKALOPOULOU, M.—SUN, R.—ESTIENNE, T.—LEROUSSEAU, M.—NIKOLAEV, S.—ALVAREZ ANDRES, E.—CARRÉ, A.—NIYOTEKA, S.—ROBERT, C.—PARAGIOS, N.—DEUTSCH, E.: *COMBING: Clustering in oncology for mathematical and biological identification of novel gene signatures*, IEEE/ACM Trans. Comput. Biol. Bioinform. **19** (2022), 3317–3331.

[10] BENZEKRY, S.—LAMONT, C.—BEHESHTI, A.—TRACZ, A.—EBOS, J. M. L.—HLATKY, L.—HAHNFELDT, P.: *Classical mathematical models for description and prediction of experimental tumor growth*, PLoS Comput. Biol. **10** (2014), e1003800.

[11] BOZIC, I.—ANTAL, T.—OHTSUKI, H.—CARTER, H.—KIM, D.—CHEN, S.—KARCHIN, R.—KINZLER, K. W.—VOGELSTEIN, B.—NOWAK, M. A.: *Accumulation of driver and passenger mutations during tumor progression*, Proc. Natl. Acad. Sci. U.S.A. **107** (2010), 18545–18550.

[12] BYRNE, H.—ALARCÓN, T.—OWEN, M. R.—WEBB, S. D.—MAINI, P. K.: *Modelling aspects of cancer dynamics: A review*, Philos. Trans. R. Soc. A **364** (2006), 1563–1578.

[13] CERAMI, E.—GAO, J.—DOGRUSOZ, U.—GROSS, B. E.—SUMER, S. O.—AKSOY, B. A.—JACOBSEN, A.—BYRNE, C. J.—HEUER, M. L.—LARSSON, E. et al.: *The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data*, Cancer Discov. **2** (2012), 401–404.

[14] CHEN, R. J.—DING, T.—LU, M. Y.—WILLIAMSON, D. F. K. et al.: *Towards a general-purpose foundation model for computational pathology*, Nat. Med. **30** (2024), 850–862.

[15] CLARK, K.—VENDT, B.—SMITH, K.—FREYMANN, J.—KIRBY, J. et al.: *The cancer imaging archive (TCIA): Maintaining and operating a public information repository*, J. Digit. Imaging **26** (2013), 1045–1057.

[16] COLLINS, G. S.—MOONS, K. G. M.—DHIMAN, P.—RILEY, R. D.—BEAM, A. L.—VAN CALSTER, B. et al.: *TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods*, BMJ **385** (2024), e078378.

[17] CORTES, C.—VAPNIK, V.: *Support-vector networks*, Mach. Learn. **20** (1995), 273–297.

[18] DE PILLIS, L. G.—GU, W.—RADUNSKAYA, A. E.: *Mixed immunotherapy and chemotherapy of tumors: modeling, applications and biological interpretations*, J. Theor. Biol. **238** (2006), 841–862.

[19] DE PILLIS, L. G.—RADUNSKAYA, A. E.—WISEMAN, C. L.: *A validated mathematical model of cell-mediated immune response to tumor growth*, Cancer Res. **65** (2005), 7950–7958.

[20] DEISBOECK, T. S.—WANG, Z.—MACKLIN, P.—CRISTINI, V.: *Multiscale Cancer Modeling*, Annu. Rev. Biomed. Eng. **13** (2011), 127–155.

[21] DENG, Z.—TANG, C.—HUANG, Z.—LIN, J.—CHEN, Y.—NING, J.—MA, C.—LIU, J.—LI, W.—ZHU, Y. et al.: *Project Imaging-X: A Survey of 1000+ Open-Access Medical Imaging Datasets for Foundation Model Development*, 2026. `https://arxiv.org/abs/2603.27460`.

[22] DOSOVITSKIY, A.—BEYER, L.—KOLESNIKOV, A.—WEISSENBORN, D.—ZHAI, X.—UNTERTHINER, T.—DEHGHANI, M.—MINDERER, M.—HEIGOLD, G.—GELLY, S.—USZKOREIT, J.—HOULSBY, N.: *An image is worth* 16×16 *words: Transformers for image recognition at scale*. In: International Conference on Learning Representations *(ICLR)*, 2021. `https://arxiv.org/abs/2010.11929`.

[23] DRUMMOND, D.—COULET, A.: *Technical, ethical, legal, and societal challenges with digital twin systems for the management of chronic diseases in children and young people*, J. Med. Internet Res. **24** (2022), e39698.

[24] ERGUN, A.—CAMPHAUSEN, K.—WEIN, L. M.: *Optimal scheduling of radiotherapy and angiogenic inhibitors*, Bull. Math. Biol. **65** (2003), 407–424.

[25] ESTEVA, A.—KUPREL, B.—NOVOA, R. A.—KO, J.—SWETTER, S. M.—BLAU, H. M.—THRUN, S.: *Dermatologist-level classification of skin cancer with deep neural networks*, Nature **542** (2017), 115–118.

[26] FARES, J.—FARES, M. Y.—KHACHFE, H. H.—SALHAB, H. A.—FARES, Y.: *Molecular principles of metastasis: a hallmark of cancer revisited*, Signal Transduct. Target. Ther. **5** (2020), Art. 28.

[27] FOWLER, J. F.: *The linear-quadratic formula and progress in fractionated radiotherapy*, Br. J. Radiol. **62** (1989), 679–694.

[28] GATENBY, R. A.—SILVA, A. S.—GILLIES, R. J.—FRIEDEN, B. R.: *Adaptive therapy*, Cancer Res. **69** (2009), 4894–4903.

[29] GERLEE, P.—JOHANSSON, M.: *Inferring rates of metastatic dissemination using stochastic network models*, PLoS Comput. Biol. **15** (2019), e1006868.

[30] GERLINGER, M.—ROWAN, A. J.—HORSWELL, S.—LARKIN, J.—ENDESFELDER, D.—GRONROOS, E.—MARTINEZ, P.—MATTHEWS, N.—STEWART, A.—TARPEY, P. et al.: *Intratumor heterogeneity and branched evolution revealed by multiregion sequencing*, N. Engl. J. Med. **366** (2012), 883–892.

[31] GHAJAR, C. M.—PEINADO, H.—MORI, H.—MATEI, I. R.—EVASON, K. J.—BRAZIER, H.—ALMEIDA, D.—KOLLER, A.—HAJJAR, K. A.—STAINIER, D. Y. R.—CHEN, E. I.—LYDEN, D.—BISSELL, M. J.: *The perivascular niche regulates breast tumour dormancy*, Nat. Cell Biol. **15** (2013), 807–817.

[32] GIANSANTI, D.—MORELLI, S.: *Exploring the Potential of digital twins in cancer treatment: A narrative review of reviews*, J. Clin. Med. **14** (2025), Art. 3574.

[33] GILLESPIE, D. T.: *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. **81** (1977), 2340–2361.

[34] GOGOSHIN, G.—RODIN, A. S.: *Graph neural networks in cancer and oncology research: Emerging and future trends*, Cancers **15** (2023), Art. 5858.

[35] HAHNFELDT, P.—PANIGRAHY, D.—FOLKMAN, J.—HLATKY, L.: *Tumor development under angiogenic signaling: A dynamical theory of tumor growth, treatment response, and postvascular dormancy*, Cancer Res. **59** (1999), 4770–4775.

[36] HAMZA, B.—MILLER, A. B.—MEIER, L.—STOCKSLAGER, M.—NG, S. R. et al.: *Measuring kinetics and metastatic propensity of CTCs by blood exchange between mice*, Nat. Commun. **12** (2021), Art. 5680.

[37] HANAHAN, D.: *Hallmarks of cancer: New dimensions*, Cancer Discov. **12** (2022), 31–46.

[38] HAO, Y.—QIU, Z.—HOLMES, J.—LÖCKENHOFF, C. E.—LIU, W.—GHASSEMI, M.—KALANTARI, S.: *Large language model integrations in cancer decision-making: a systematic review and meta-analysis*, npj Digit. Med. **8** (2025), Art. 450.

[39] HSIEH, W.-C.—BUDIARTO, B. R.—WANG, Y.-F.—LIN, C.-Y.—GWO, M.-C.—SO, D. K.—TZENG, Y.-S.—CHEN, S.-Y.: *Spatial multi-omics analyses of the tumor immune microenvironment*, J. Biomed. Sci. **29** (2022), Art. 96.

[40] ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium: *Pan-cancer analysis of whole genomes*, Nature **578** (2020), 82–93.

[41] IWASA, Y.—NOWAK, M. A.—MICHOR, F.: *Evolution of resistance during clonal expansion*, Genetics **172** (2006), 2557–2566.

[42] IWATA, K.—KAWASAKI, K.—SHIGESADA, N.: *A dynamical model for the growth and size distribution of multiple metastatic tumors*, J. Theor. Biol.**203** (2000), 177–186.

[43] KEMKAR, S.—TAO, M.—GHOSH, A.—STAMATAKOS, G.—GRAF, N.—POOREY, K.—BALAKRISHNAN, U.—TRASK, N.—RADHAKRISHNAN, R.: *Towards verifiable cancer digital twins: tissue level modeling protocol for precision medicine*, Front. Physiol. **15** (2024), Art. 1473125.

[44] KOSTELICH, E. J.—KUANG, Y.—MCDANIEL, J. M.—MOORE, N. Z.—MARTIROSYAN, N. L.—PREUL, M. C.: *Accurate state estimation from uncertain data and models: an application of data assimilation to mathematical models of human brain tumors*, Biol. Direct **6** (2011), Art. 64.

[45] KUZNETSOV, V. A.—MAKALKIN, I. A.—TAYLOR, M. A.—PERELSON, A. S.: *Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis*, Bull. Math. Biol. **56** (1994), 295–321.

[46] LEDZEWICZ, U.—MAURER, H.—SCHÄTTLER, H.: *Optimal and suboptimal protocols for a mathematical model for tumor anti-angiogenesis in combination with chemotherapy*, Math. Biosci. Eng. **8** (2011), 307–323.

[47] LEDZEWICZ, U.—SCHÄTTLER, H.: *A review of optimal chemotherapy protocols: From MTD towards metronomic therapy*, Math. Model. Nat. Phenom. **9** (2014), 131–152.

[48] LENHART, S.—WORKMAN, J. T.: *Optimal Control Applied to Biological Models*. Chapman and Hall/CRC, 2007.

[49] LIMA, E. A. B. F.—FAGHIHI, D.—PHILLEY, R.—YANG, J.—VIROSTKO, J.—PHILLIPS, C. M.—YANKEELOV, T. E.: *Bayesian calibration of a stochastic, multiscale agent-based model for predicting in vitro tumor growth*, PLoS Comput. Biol. **17** (2021), e1008845.

[50] LUNDBERG, S. M.—LEE, S.-I.: *A Unified Approach to Interpreting Model Predictions*. In: *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774.

[51] Mathematical Oncology Community Resource: *Mathematical Oncology*, `https://mathematical-oncology.org/`, Accessed 8 April 2026.

[52] MAYNARD SMITH, J.—PRICE, G. R.: *The logic of animal conflict*, Nature **246** (1973), 15–18.

[53] METZCAR, J.—JUTZELER, C. R.—MACKLIN, P.—KÖHN-LUQUE, A.—BRÜNINGK, S. C.: *A review of mechanistic learning in mathematical oncology*, Front. Immunol. **15** (2024), Art. 1363144.

[54] MOINGEON, P.—CHENEL, M.—ROUSSEAU, C.—VOISIN, E.—GUEDJ, M.: *Virtual patients, digital twins and causal disease models: Paving the ground for in silico clinical trials*, Drug Discov. Today **28** (2023), Art. 103605.

[55] MOREIRA, J.—DEUTSCH, A.: *Cellular automaton models of tumor development: A critical review*, Adv. Complex Syst. **5** (2002), 247–268.

[56] NORTON, K.-A.—GONG, C.—JAMALIAN, S.—POPEL, A. S.: *Multiscale agent-based and hybrid modeling of the tumor immune microenvironment*, Processes **7** (2019), Art. 37.

[57] NOWELL, P. C.: *The clonal evolution of tumor cell populations*, Science **194** (1976), 23–28.

[58] OSIPOV, A.—NIKOLIC, O.—GERTYCH, A.—PARKER, S.—HENDIFAR, A.—SINGH, P.—FILIPPOVA, D.—DAGLIYAN, G.—FERRONE, C. R.—ZHENG, L.—MOORE, J. H.—TOURTELLOTTE, W.—VAN EYK, J. E.—THEODORESCU, D.: *The Molecular Twin artificial-intelligence platform integrates multi-omic data to predict outcomes for pancreatic adenocarcinoma patients*, Nat. Cancer **5** (2024), 299–314.

[59] PANDIT, S.—XU, A.—NGUYEN, X.-P.—MING, Y.—XIONG, C.—JOTY, S.: *Hard2Verify: A step-level verification benchmark for open-ended frontier math*, `https://arxiv.org/abs/2510.13744`.

[60] PARK, J. J. H.—HSU, G.—SIDEN, E. G.—THORLUND, K.—MILLS, E. J.: *An overview of precision oncology basket and umbrella trials for clinicians*, CA Cancer J. Clin. **70** (2020), 125–137.

[61] PRADELLI, F.—STROBL, M.—MARZBAN, S.—DE KERMENGUY, F.—BARNETT, A.—GANESAN, K.—LORENZO, G.—HORMUTH, D. A.—HAMIS, S.—BHASKAR, D.—ANDERSON, A. R. A.—WEST, J.: *75 Years of Mathematical Oncology*, `https://www.biorxiv.org/content/10.64898/2026.01.13.699306v1`.

[62] PUGH, K.—GYLLINGBERG, L.—STRATIEV, S.—HAMIS, S.: *A bibliometric study on mathematical oncology: Interdisciplinarity, internationality, collaboration and trending topics*, Bull. Math. Biol. **87** (2025), Art. 174.

[63] RIEKE, N.—HANCOX, J.—LI, W.—MILLETARÌ, F.—ROTH, H. R.—ALBARQOUNI, S. et al.: *The future of digital health with federated learning*, npj Digit. Med. **3** (2020), Art. 119.

[64] ROCKNE, R. C.—HAWKINS-DAARUD, A.—SWANSON, K. R.—SLUKA, J. P.—GLAZIER, J. A.—MACKLIN, P.—HORMUTH, D. A. et al.: *The 2019 mathematical oncology roadmap*, Phys. Biol. **16** (2019), Art. 041005.

[65] ROOSE, T.—NETTI, P. A.—MUNN, L. L.—BOUCHER, Y.—JAIN, R. K.: *Solid stress generated by spheroid growth estimated using a linear poroelasticity model*, Microvasc. Res. **66** (2003), 204–212.

[66] SCHWARTZ, R.—SCHÄFFER, A. A.: *The evolution of tumour phylogenetics: principles and practice*, Nat. Rev. Genet. **18** (2017), 213–229.

[67] SCHÄTTLER, H.—LEDZEWICZ, U.: *Geometric Optimal Control: Theory, Methods and Examples*. Interdiscip. Appl. Math. vol. 38, Springer, 2012.

[68] SCHÄTTLER, H.—LEDZEWICZ, U.: *Optimal Control for Mathematical Models of Cancer Therapies: An Application of Geometric Methods*. Interdiscip. Appl. Math., vol. 42, Springer, 2015.

[69] SCIMAGO. *SJR—SCImago Journal & Country Rank*. `https://www.scimagojr.com/`, 2026. Accessed 22 April 2026.

[70] SKIPPER, H. E.—SCHABEL, F. M.—WILCOX, W. S.: *Experimental evaluation of potential anticancer agents. XIII. On the criteria and kinetics associated with "Curability" of experimental leukemia*, Cancer Chemotherapy Reports **35** (1964), 1–111.

[71] STROBL, M. A. R.—WEST, J.—VIOSSAT, Y.—DAMAGHI, M.—ROBERTSON-TESSI, M.—BROWN, J. S.—GATENBY, R. A.—MAINI, P. K.—ANDERSON, A. R. A.: *Turnover modulates the need for a cost of resistance in adaptive therapy*, Cancer Res. **81** (2021), 1135–1147.

[72] SUBRAMANIAN, A.—TAMAYO, P.—MOOTHA, V. K.—MUKHERJEE, S.—EBERT, B. L.—GILLETTE, M. A.—PAULOVICH, A.—POMEROY, S. L.—GOLUB, T. R.—LANDER, E. S.—MESIROV, J. P.: *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles*, Proc. Natl. Acad. Sci. U.S.A. **102** (2005), 15545–15550.

[73] SWANSON, K. R.—ALVORD, E. C.—MURRAY, J. D.: *A quantitative model for differential motility of gliomas in grey and white matter*, Cell Proliferation **33** (2000), 317–329.

[74] ŚWIERNIAK, A.—LEDZEWICZ, U.—SCHÄTTLER, H.: *Optimal control for a class of compartmental models in cancer chemotherapy*, Int. J. Appl. Math. Comput. Sci. **13** (2003), 357–368.

[75] TIROSH, I.—IZAR, B.—PRAKADAN, S. M.—WADSWORTH, M. H.—TREACY, D.—TROMBETTA, J. J.—ROTEM, A.—RODMAN, C.—LIAN, C.—MURPHY, G. et al.: *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq*, Science **352** (2016), 189–196.

[76] TOMLINSON, I. P.—BODMER, W. F.: *Modelling the consequences of interactions between tumour cells*, Br. J. Cancer **75** (1997), 157–160.

[77] U.S. Food and Drug Administration and Health Canada and Medicines and Healthcare products Regulatory Agency: *Good Machine Learning Practice for Medical Device Development: Guiding Principles*, Technical report, U.S. Food and Drug Administration, 2021.

[78] VICKERS, A. J.—ELKIN, E. B.: *Decision Curve Analysis: A Novel Method for Evaluating Prediction Models*, Medical Decision Making **26** (2006), 565–574.

[79] WANG, Z.—GERSTEIN, M.—SNYDER, M.: *RNA-Seq: a revolutionary tool for transcriptomics*, Nature Reviews Genetics **10** (2009), 57–63.

[80] WEST, G. B.—BROWN, J. H.—ENQUIST, B. J.: *A general model for ontogenetic growth*, Nature **413** (2001), 628–631.

[81] WEST, J.—YOU, L.—ZHANG, J.—GATENBY, R. A.—BROWN, J. S.—NEWTON, P. K.—ANDERSON, A. R. A.: *Towards multidrug adaptive therapy*, Cancer Res. **80** (2020), 1578–1589.

[82] WILKINSON, M. D.—DUMONTIER, M.—AALBERSBERG, I. J.—APPLETON, G.—AXTON, M.—BAAK, A.—BLOMBERG, N. et al.: *The FAIR guiding principles for scientific data management and stewardship*, Scientific Data **3** (2016), Art. 160018.

[83] YANKEELOV, T. E.—ATUEGWU, N. C.—DEANE, N. G.—GORE, J. C.: *Modeling tumor growth and treatment response based on quantitative imaging data*, Integrative Biology **2** (2010), 338–345.

[84] ZHANG, J.—CUNNINGHAM, J. J.—BROWN, J. S.—GATENBY, R. A.: *Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer*, Nat. Commun. **8** (2017), Art. 1816.

[85] ZHU, W.—CHEN, A.—SONG, Y.—CHEN, K.—ZHU, C.—CHEN, Z.—ZHAO, T.: *From perception to reasoning: Deep thinking empowers multimodal large language models*, `https://arxiv.org/abs/2511.12861`.

Received May 30, 2026
Revised June 28, 2026
Accepted June 30, 2026

Mathematical Institute
Slovak Academy of Sciences
Štefánikova 49
814 73 Bratislava
SLOVAKIA

E-mail: zigman@mat.savba.sk