Executable Cancer Models: Successes and Challenges

Matthew A. Clarke[1] and Jasmin Fisher[1,2†]

[1]Department of Biochemistry, University of Cambridge, Cambridge, UK

[2]UCL Cancer Institute, University College London, London, UK

†Correspondence should be addressed to: jasmin.fisher@ucl.ac.uk

Abstract

Making decisions on how best to treat cancer patients requires the integration of different datasets including genomic profiles, tumour histopathology, radiological images, proteomic analysis, and more. This wealth of biological information calls for novel strategies to integrate such information in a meaningful, predictive and experimentally verifiable way. In this Perspective we explain how executable computational models meet this need. Such models provide a means for comprehensive data integration, can be experimentally validated, are readily interpreted both biologically and clinically, and have the potential to predict effective therapies for different cancer types and subtypes. We explain what executable models are, how they can be used to represent the dynamic biological behaviours inherent in cancer and demonstrate how such models, when coupled with automated reasoning, facilitate our understanding of the mechanisms by which oncogenic signalling pathways regulate tumours. We explore how executable models have impacted the field of cancer research and argue that extending them to represent a tumour in a specific patient (that is, an avatar) will pave the way for improved personalised treatments and precision medicine. Finally, we highlight some of the ongoing challenges in developing executable models, and stress that effective cross-disciplinary efforts are key to forward progress in the field.

[H1] Introduction

Cancers are highly complex and diverse diseases driven by genetic mutations, copy number aberrations and epigenetics that disrupt a multitude of signalling pathways that change phenotypes such as cell proliferation, survival or death, repair, movement and invasion, and operate across different cell types[1]. A cancer cannot be considered as a static organism, it is much more like an evolving species. Cancer development is noisy[2], accounting in part for genetic variation, both between the same cancer type in different individuals, and in the tumour cell populations of an individual tumour[3,4]. Yet, like other biological systems, cancer development is robust[5], due to functional and signalling redundancy, intrinsic error correction and internal self-organization. This can, for example, give rise to drug resistance as a tumour reacts by re-routing signals to restore homeostasis[6] in response to targeted drugs. Even when oncologists attempt to control resistance using intermittent therapies, patients can relapse due to growth of resistant cancer clones[7]. It is therefore clear that cancer can only be understood as a holistic, dynamic, reactive and evolving system.

Despite the enormous success of 21st century molecular biology, a coherent explanation of mechanisms underyling cancer development is still lacking in most cases. While genomic, transcriptomic and proteomic approaches contribute to our depth of understanding of cancer, they cannot be directly translated into better treatments. We need to understand how genomic, transcriptomic, proteomic, cell signalling and microenvironmental factors interact as a system to influence cancer development, progression and resistance to therapy.

A further challenge to laboratory investigations of cancer is that experimental interventions may themselves disrupt the phenomena being studied. Conventional in vitro models of cancers utilise cell lines in reductive 2-dimensional (2D) cultures that distort cell morphology, alter signalling pathways and fail to recapitulate resistance mechanisms[8]. In consequence, agents showing therapeutic promise in vitro have often turned out to lack clinical efficacy[9–13]. While more clinically-relevant 3D patient-derived organoid models have been developed, they too are not perfect genomic representations of their tumours of origin[14]. In fact, while key phenotypic and genetic characteristics are captured, many major patient-derived organoid cohorts reported to date show differences in somatic mutations from matched primary tumours[15–19].

In vivo models too are insufficient. For example, genetically engineered mouse models (GEMMs) and patient-derived xenografts (PDXs), while more physiologically relevant than in vitro models, also experience genomic evolution both at the tumour level and the host level[20]. Furthermore, the xenograft environment is quite distinct from the original patient environment with respect to its metabolism, physiology, cellular interactions, microenvironment and absence of a functioning immune system[21].

In summary, we do not have the means to do experiments on cancer at the whole system level without either using inferior systems (e.g., mouse versus human), compromising the system (e.g., by knocking out genes) or ignoring key features of the system (e.g., immune system response). More inclusive methods are required that recapitulate and integrate the cellular, genomic, microenvironmental and spatial features of cancers to be able to understand and overcome their multifaceted resistance mechanisms. We believe that, despite the innate complexity of biological systems, underlying commonalities and principles exist and govern their behaviour[22] since such systems, and especially cancer tumours, give rise to reproducible phenotypes and are capable of making identifiable binary decisions[2,5]. The question is, how can we uncover these principles, determine how they operate and understand how to exploit them to achieve a desired biological change? By discovering and codifying commonalities and principles in biological systems as discrete instructions i.e. "executable computational models" (Box 1), and understanding how these models interact, we can hope to elucidate biological function and predict the effect of perturbations such as genetic mutations or drugs on specific cancer types to provide more precise and effective disease treatments[23]. While a computational model cannot capture all the features observed in a real system, it makes explicit what can and cannot be included. Moreover, a computational model can be

interrogated in more depth than a real system allowing deeper exploration and the characterisation of its strengths and weakness.

Executable computational models provide a means to mathematically represent mechanistic explanations of biological behaviours at the genetic, molecular and cellular level alongside one another, and specify how these behaviours evolve and interact. In essence, executable models treat biological systems as networks of programmed interacting biological "machines", giving a mechanistic view of the whole system. By describing such biological machines as executable computer programs **[G]**, we gain access to computer science concepts such as abstraction **[G]**, modularity **[G]**, compositionality **[G]**, and concurrency **[G]** to enable creation of multi-scale models of complex biological systems. Furthermore, by adopting computer science techniques for verifying of the behaviour of complex software and hardware systems, we can reason about the dynamic behaviour, properties and responses of biological systems "in silico" as a complement to classical experimental approaches. Over the last two decades, this "Executable Biology"[24] approach has led to new biological insights and a methodology complementary to other computational approaches such as mathematical modelling, bioinformatics and machine learning (Box 2, Table 1). In this Perspective we explore the executable modelling approach to study cancer, the results achieved and future prospects.

[H1] Executable computational models

The basic underpinning of an executable computational model of a biological system is a state transition system that defines how, given certain events (e.g., molecular signals), one system state is transformed into another. The components of a state machine **[G]** represent biological entities - cells, proteins, or genes. In a state transition system, components react to events signalled from other components by changing the state of the system. State transitions are scalable and may be assembled together to generate larger scale and/or finer grained models representing complex, dynamic biological behaviours[24–27]. As the name suggests, such a model can also be treated as an executable computer program – i.e., it can be run as a simulation and can be analysed to predict properties such as stability and termination (i.e., ceasing to evolve). There are various types of executable models in use today for describing biological systems; the most common are outlined in Box 3.

Executable computational modelling is different from quantitative mathematical approaches (Box 1 and Box 2) such as ordinary differential equations (ODEs)[28]. ODEs are limited to describing the behaviour of a biological system with reference to time and observable properties such as diffusion and reaction rates of molecules – i.e., they are continuous or analogue models **[G]**. By contrast, executable computational models are qualitative and expressed in terms of the relative ordering of "events"[29]. In the simplest case an event may correspond to an observed phenomena, but more generally may be defined as a point reached in terms of a logical and temporal relationship between events – i.e., statements such as "A and B followed by C while D is occurring"[30–32]. Such statements naturally lead to an understanding of the dynamics of a system in terms of cause and effect. They are a digital model of the system. Additionally, executable models allow different types of information to be combined to describe the same biological phenomena at different levels of detail and for these levels to be related to one another. For example, executable models can represent the molecular level or, at a higher level of abstraction, the cellular or tissue level, and then express how events at one level correlate with events at another[25–27]. These are often referred to as "multi-scale models" (Box 2).

Executable models are ideal for testing **[G]** and comparing hypotheses about the mechanisms underpinning biological behaviours. Using formal verification **[G]** methods derived from software and hardware engineering, executable models can identify discrepancies between hypothesized models of biological mechanisms and actual experimental observations, identifying missing behaviours in the models and so informing subsequent experiments to further flesh them out. In simple cases, direct execution of a model can be used to ascertain whether a proposed mechanism conforms to the data (Figure 1). However, in a complex system exhibiting noise or

stochastic behaviours, each execution may yield a different outcome, potentially making it impossible to exhaustively check by execution that the full distribution of outcomes conforms to the data. Fortunately, thanks to a formal verification technique called model checking [G] [32,33], it is possible to systematically analyse all of the many possible outcomes of an executable model against a specification [G] without having to execute them one by one. Thus, having model checking confirm that all possible outcomes of the executable model agree with the experimental data (i.e., the specification), and that all experimental outcomes are reproduced, we know the model describes a potential mechanism fitting the data. If the model does not fully match the data, then either the model is wrong or it is incomplete, and it needs to be improved. A further benefit of model checking is that it identifies possible executions that do not fit the data in cases of a mismatch, thereby suggesting additional, targeted experiments to confirm or invalidate the mechanistic hypothesis represented by the model (Figure 1).

By an iterative process of model creation and refinement, interleaved with experimental validation of predictions derived from the model, executable models can deliver grounded, formalised, mechanistic explanations of complex biological systems such as cancer. Hence, it can be hoped that combining experimental data with multi-scale executable models will offer a unique opportunity to better understand the determinants of cancer development and treatment, including the mechanisms of therapy resistance. We can further hope that by tailoring the initial state of an executable cancer model to an individual patient's genetics, we can predict effective and efficient individual patient treatments, delivering on the vision of personalised cancer therapies.

[H1] Delivering mechanistic insights

The first step in building an executable model of a biological system is to identify the interactions driving the behaviour of interest. This can be done bottom-up by deriving the model from multiple datasets across the literature[34,35], or top-down from a single dataset, which can work particularly well for transcriptomic data[36,37]. However, top-down network inference methods such as ARACNE (algorithm for the reconstruction of accurate cellular networks) [38] can result in false positives as it is difficult to disentangle co-expression observed in such datasets from the co-regulation, which cannot be directly measured[39]. Top-down approaches also require that the dataset must cover all the interactions of interest. By contrast, bottom-up approaches have the advantage of being able to combine different datasets and methodologies, which can compensate for the difficulty of measuring different parts of the system. Unfortunately, the literature is often incomplete and may lack information about how the interactions change in different contexts. This means that regardless of the method of construction, executable models must be thoroughly verified against known behaviours. A helpful aspect of model checking is that it can accommodate "non-determinism [G]" in parts of a system, allowing models to represent poorly understood interactions in a qualitative way.

[H2] *Identifying new modes of interaction between signalling pathways*
To test the validity of an executable model it can be run as a simulation, and the phenotype of a cell determined by the integration of a wide array of signals from across the model (Figure 2). This phenotype can then be compared to experimental observations. While mathematical models give quantitative predictions, they require kinetic information for each component, which limits the range of behaviours and scale of systems against which they can be tested. By taking a coarser level of abstraction[40] [G], qualitative executable models are able to capture a broader set of properties for characterising systems. For example, qualitative modelling has been used to discover the key functional modules of signalling in the ERBB family of receptor tyrosine kinases[41]. As this model does not require precise information about kinetic constants, it can model a larger part of the ERBB system than comparable ODE-based models[42]. Similarly, a qualitative model of liver cancer derived from a static network representation found new interactions not known in the literature, and also cases of hypothesized interactions which, when removed from the model, improved its fidelity[43].

[H2] Deciphering the sequence of molecular events

Executable models allow the exploration of the sequence of molecular events in a cell: many events must occur in a precise sequence for a cell to function, from the short scale ordering of events in the cell cycle,[44,45] to more long-term orderings that occur during development[46]. To understand these temporal relationships a dynamic model is required. This is an area in which current computer science approaches to modelling interaction do not carry over well to biology. For example, in order to model the synchronisation between the epidermal growth factor receptor (EGFR) and Notch signalling during the process of cell fate determination in *C. elegans*, a new approach called "Bounded Asynchrony" for describing the timing of molecular interactions[47] was required to understand how different cells within the developing tissue synchronise their signalling[25–27].

In the case of a system that must be very tightly controlled, such as cell development, it is paramount that the model does not allow the healthy cell to deviate from the developmental path. This is difficult to test in a continuous model as the parameter space is infinite, but in a discrete model [G], model checking and formal verification make it possible to test against a high-level specification of all possible stable states of a system, and/or all possible sequences of events. This was demonstrated in a hybrid model combining a discrete signalling network model of *C. elegans* germ cells with a continuous model of their physical interactions[48]. As germ-cells develop, they need to pass through distinct stages at the correct time in the correct order. This process is controlled by different microenvironments to which the cells are exposed as they develop, in particular signalling from Notch and Ras. These signals cue the transition between different developmental phases. To ensure correct development, the germ-cell must be robust to perturbations as these environments change. To check that the model reproduced this robustness model-checking was conducted to find all the possible stable states of the model, for all possible initial conditions, without the need for exhaustive simulation. While the network may have multiple possible stable states, the sequence of signals that the cell receives enforces a stable and linear fate progression in the model. Furthermore, differences in the forces on mitotic cells encourages mixing, and so breaks up regions of clonal dominance. In the germline, clonal dominance must be avoided to maintain genetic diversity, while similar mechanisms are required to reduce the risk of cancer in other tissues. An example of such a mechanism can be seen in the structure of the intestinal crypt. As cells are constantly being pushed out of the niche, any clone that gains an oncogenic mutation that does not immediately improve proliferative fitness is outcompeted by healthy cells, lowering the chance of a tumourigenic clone acquiring dominance[49].

Being able to predict all possible outcomes of perturbations and how the sequence of molecular events and perturbations affects which stable state of a network is reached, will be important for predicting effective cancer therapies. For example, this opens up new opportunities for network reprogramming[50]. This is important for understanding the specification of lineages[51] and therefore for lineage reprogramming[52]. Zañudo et al. explored reprogramming for cancer therapy using a Boolean network model. A Boolean network model of leukemia shows how reprogramming can be of use for cancer therapy[53]. By analysing the stable states in the Boolean network model and how they changed under perturbation, the authors showed how this can reveal sequences of changes that convert a cell from a stably cancerous phenotype to a stably beneficial phenotype, such as apoptosis. Similarly, a Boolean model of transforming growth factor-β (TGFβ) signalling in hepatocellular carcinoma suggested a specific sequence of events needed for Epithelial–mesenchymal transition (EMT)[54], a behaviour that may contribute to tumour progression and metastasis[55]. Further Boolean network models of oncogenic signalling have shown that, even when such reprogramming is not the goal, the timing of interventions can affect their therapeutic impact[56,57].

[H2] Tracing cancer evolution

As a cancer emerges and evolves, each mutation will change the possible stable states of the gene regulatory network. Some cancer types appear to acquire mutations in specific orders[58,59]. This raises the question of whether certain sequences have a benefit over others because they produce stable states associated with higher fitness. Machine learning models have shown to reproduce and predict such order in lung and colon cancers[60], while Boolean network models have been used to investigate

the effect of each mutation in the canonical route to colorectal cancer[61].

It is possible to further model potential alternative orderings of acquired mutations compared to those that are observed clinically to learn why certain orderings of mutations are more prevalent. For example, a mathematical model can show how the order of acquisition of different cell characteristics, such as replication and mutation rates, affects the rate of development of cancer[62]. By using an ODE model, techniques such as sensitivity analysis can be used to determine the relative contributions of each characteristic to the differences between these paths to cancer.

It is possible to further examine this problem at the level of a genetic regulatory network[63]. Here, a method is implemented to compute all the reachable stable states from any subset of initial states in a Qualitative Network, an extension of Boolean networks in which nodes can take multiple finite values beyond simply "on" or "off". This method is used to trace changes in the available stable states as cancer evolves from tumorigenesis onward and investigated all possible orderings of mutations. By tying the states of the network to overall cell phenotypes, they showed which sequence of mutations produced the fastest increase in proliferation while minimizing apoptosis. This suggests which sequences of mutations are likely to be the most commonly observed as a cancer develops, and which sequences are less common, as clones that attempt to follow these routes will be less fit. Reconstructing common patterns of mutation is complex, and for early founding mutations more so[64]. But using executable models to understand the patterns of the order of acquisition of mutations could help to produce more effective therapies by identifying likely early, founder mutations. These founder mutations are likely to be present in all the cells of even a heterogeneous tumour and so may be more effective treatment targets[65–67].

[H2] Discovering novel therapeutic strategies

One of the key uses of a verified executable model of cancer is to accelerate the search for and testing of novel therapies (Figure 2). This can save time and expense by guiding experiments to those areas most likely to be effective and thereby overcome the slowdown in drug discovery currently attributed to the exhaustive nature of purely experimental approaches[11,12]. One can either search for perturbations that shift the network from one stable state to another[50,53], or that introduce new stable states that are beneficial to the patients[68].

Targeted therapies are highly effective, but quickly lead to resistance[69,70]. To combat this, combinations of drugs are often used, as it is less likely that a tumour will contain cells with mutations rendering it resistant to multiple drugs at once[71,72]. An executable network model allows the easy simulation and comparison of the effects of multiple pairwise combinations, and their comparison to monotherapies, to find additive, possibly even synergistic, drug combinations[34,73,74]. Executable models allow for the testing of more complex drug combinations, which would otherwise be cost and time prohibitive in a laboratory. For example, digital systems design methods were first used to generate a Boolean representation of the growth factor signalling networks for the insulin-like growth factor 1 receptor (IGF1R) and ERBB families[75]. The authors then found and classified potential faults that would lead to aberrant dynamics, including cancer, analogously to how an electrical digital system might be assessed for potential faults. Finally, they tested 64 possible combinations of 6 drugs for their capability to reduce proliferation in the case of such faults, and identified the most effective combinations, as well as those potential faults that cannot be treated with any combination of drugs. Although only tested here on a small scale, such an approach could be used to cut down on the number of potential combinations to be tested and focus *in vitro* and *in vivo* validation only on the most promising combinations. However, such brute force searches may not always be possible in very large networks, or for more than pairwise combinations of a large number of drugs.

A key question is how different cancers will respond differently to therapy, for example when targeting the p53 transcription factor network[68]. The authors constructed Boolean network models of p53 signalling, with unique networks resulting from rewiring caused by different mutations and copy number alterations in various genes. Considering 83 cancer cell lines derived from 14 tissues, they found 45 different versions of the p53 network. For each network, the stable states, or attractors [G] , were tied back to cellular behaviour based on the stable state of certain nodes in the network. For example, stable Caspase activity was tied to cell death. The authors then simulated how

these attractors change when one or a combination of nodes is perturbed (to simulate drug therapy) and predicted which single and combination therapies should be the most effective depending on the genotype, validating their results experimentally for 8 cancer cell lines. Attempting to cluster cancers on genotype alone, rather than by the attractors, did not correlate with the simulated drug responses, showing the importance of dynamic models to correctly predict the effects of therapy and choose the best treatments in a personalised manner.

[H2] Determining mechanisms of drug resistance

Executable cancer models offer the opportunity to predict resistance mechanisms, which in turn may lead to discovery of new methods to pre-empt or overcome resistance. For example, Boolean network models were constructed separately for three breast cancer cell lines with different levels of resistance to common inhibitors of ERBB family members, namely sensitive to both trastuzumab (which targets ERBB2) and pertuzumab (which inhibits dimerization of ERBB2 with other ERBB family members), trastuzumab resistant, or pertuzumab resistant[76]. The choice of a Boolean framework allowed the inference of these cell-specific networks without the need for measurement or estimation of kinetic constants. The authors then compared these models for mutations and novel interactions within the gene regulatory network that could underpin potential drug resistance mechanisms and found unique feedback loops in the trastuzumab resistant cell line, providing a potential mechanism for resistance. Others have attempted to predict new resistance mechanisms. For example Zañudo et al.[77] developed a Boolean model of oestrogen receptor-positive breast cancer based on literature data, while Silverbush et al.[73] developed a qualitative network model of acute myeloid leukaemia using proteomics data from cell lines with different levels of resistance to PIM kinase inhibitors. By simulating the effect of single or double gene mutations under the effects of PI3K inhibition[77] or PIM inhibition[73], these models suggested novel resistance mechanisms to the inhibitors. Silverbush et al. used this understanding to predict and validate that AKT1/2 inhibition can overcome resistance to PIM inhibition, while Zañudo et al. further model the degree to which PI3K inhibitors might affect the survival of the cell[77].

Executable network models can also address another layer of complexity in that tumours are largely heterogeneous, with different clones having different mutations[4,78]. This is important as drugs may have different effects on different clones within a tumour[65,79]. Being able to choose the most effective therapy therefore necessitates a mechanistic understanding of how these clones compete and cooperate. Qualitative network modelling has been used to investigate how clones with normal and dysregulated MYC expression cooperate in breast cancer[80]. Using this model, the authors predicted the difference in the effects of single and pairwise combination therapies on different MYC clones within the tumour. They further investigated how these effects are changed by the cooperation of the clones in a heterogeneous tumour compared to pure clonal tumours. The authors then predicted the most effective combination therapy and validated its effectiveness in a novel mouse model of such heterogeneous tumours.

These examples illustrate how executable and mathematical models could impact cancer treatment through the guidance of experiments and drug discovery in academia. In the pharmaceutical industry, there is an increased need for pharmacokinetic/pharmacodynamic (PKPD) models that can simulate systems level interactions, such as in the case of the interactions across pathways in response to combination therapy. In order to bridge this gap the field of Quantitative Systems Pharmacology (QSP) is emerging. The challenges at the interface of pharmacology and systems biology and the emerging methods to overcome them are reviewed by van Hasselt et al.[81] and Kirouac et al.[82].

[H1] Clinical applications

Executable computational models have great potential for impact in the clinic as tools to aid prognosis and to bring the planning of treatments to a personalised level. As the area of executable modelling is still fairly new, it will take time for such methods to clear the regulatory hurdles and be accepted into clinical practice. However, we can see how things are likely to develop by looking at

some of the earlier statistical, machine learning and mathematical models, which have blazed a trail that executable models can follow. Importantly, a key benefit of executable models over machine learning approaches is that an executable model affords a mechanistic interpretation, which can be reviewed for biological plausibility and fit to clinical expertise. This is critical if prognostic and treatment planning tools derived from computational models are to be acceptable for use by clinicians.

[H2] Improved prognostic markers

Presently disease prognosis is dependent on low-throughput tests such as cytogenetics[83,84]. High-throughput genomic and transcriptomic technologies offer the potential to increase the accuracy of such tests and to stratify patients into more useful prognostic categories using statistical models. For example, the OncoTypeDX test[85] offers a 21-gene recurrence score assay that can identify patients who are not likely to require chemotherapy, sparing them painful and stressful treatment. There has been a decrease in the use of adjuvant chemotherapy since the production of the test[86] and this is likely to decrease further as the accuracy of the test becomes better understood[87].

While such statistical models can be immensely useful in predicting what will happen during the progression of a cancer, they are of limited use in elucidating how and why the disease develops differently in different patients. By contrast, mechanistic models can be used to find new treatments that can be applied in the clinic. As an example, a model of the competition between drug-resistant and drug-sensitive cells within a cancer using evolutionary game theory was used to develop a new treatment strategy for metastatic castration-resistant prostate cancer based on the predictions of the model[88]. The authors used cycles of dosing to reduce the bulk of the tumour, but not eliminate the sensitive cells, as these then competed with the resistant cells during drug holidays. In this manner the cancer could be managed as a chronic disease. They report that an ongoing pilot clinical trial[88,89] showed significant benefit compared to those treated with standard-of-care methods. Critically, their modelling of previously attempted trials of intermittent dosing strategies[90–93], suggest that such non-adaptive therapies are no better than continuous treatment at maximum therapeutic dose. Similarly, dosing strategies can be improved based on mathematical modelling of evolution in prostate cancer[94]. Further such models will be necessary to find the correct dosing schedules to apply the broader adaptive therapy **[G]** paradigm[95], as every cancer type is likely to have unique dynamics. This can be aided by applying these principles to executable models, which, as shown in the prior sections, can flexibly model different backgrounds and adapt to incorporate resistance mechanisms as they emerge.

[H2] Personalised treatment regimes

Modelling can also be useful once treatment begins. One of the main limiting factors in personalised cancer medicine today is the inability to plan treatment choices based on a patient's response over time, whether due to changing sensitivity to specific therapies as the cancer develops resistance or in managing dosage and timing to give the patient reasonable quality of life during treatment. Mathematical models are being developed to make radiotherapy more personalised and responsive by evaluating not only variables such as tumour volume or lymph node involvement, but the rate of change over time, and using this to inform ongoing treatment[96]. By the same token it is expected that executable models can also be personalised using such biomarkers and their dynamics over time. Such models would then allow more responsive therapy as they can be quickly updated to respond to new data, allowing changes to the model to be made based on measurements taken during treatment, and in turn allowing changes to be made to therapy to adapt to unforeseen tumour behaviour. However, executable models may be able to work well even in cases where it is impractical to gain the level of detailed measurement a mathematical model requires, as they operate at a higher level of abstraction.

[H2] Clinical trials

The maturity of executable modelling is now such that results from using these techniques are reaching clinical trial. Network-based approaches dominate, although these are so far mostly

restricted to analysis of the topology of static networks[97]. The current focus is trying to find the causes of aberrant gene expression by predicting which areas of a network are affected by perturbations[98]. For example, network inference with ARACNE[38] and topology analysis to find master regulator transcription factors with MARINa (master regulator inference algorithm) MARINa[99] was used to suggest Janus kinase (JAK) as a drug target in breast cancer[100], and followed up with an ongoing clinical trial[101]. A further trial based on targets identified by ARACNE, MARINa and similar techniques for gastroenteropancreatic neuroendocrine tumors is also underway[102,103].

[H1] Challenges and outlook

From the forgoing, it is evident that executable computational models are contributing to the understanding of cancer and there is emerging evidence that use of such models in clinical settings could potentially improve cancer treatment. The key questions are therefore whether current executable modelling approaches are sufficient or whether they need further development to be fully practical, and how they can be made more widely accessible to both the research and the clinical communities? It is evident that practical executable computational modelling calls for broad expertise spanning cell biology, biochemistry, theoretical computer science, software engineering, bioinformatics, data science, and clinical research. It is unreasonable to expect to find all these skills in any one individual and so work in this field is inevitably interdisciplinary and therefore brings with it both the opportunities of cross-fertilization between different disciplines and the challenges of working across organizational and funding boundaries. Particularly in an academic context, how can members of an interdisciplinary team communicate the value of their contributions back to their own discipline and how do they balance innovating in their own domain against applying existing knowledge to benefit others in the team? From our own experience we have found that interdisciplinary work can stimulate new developments in one's own field.

[H2] From a reductionist to holistic view

Biological sciences are going through a cultural shift, moving away from the conventional reductionist approach of investigating single molecules at depth towards an approach based on identifying patterns of behaviour (i.e., models) to elucidate a mechanistic understanding of biological phenomena as a coherent whole (rather than as a collection of parts). This shift is not intuitive and requires a leap of faith that biological mechanisms can be explained without needing to understand every detail at every level, i.e., that the principle of abstraction is as applicable to biology as it is to mathematics, physics or computer science. It is inevitable that in order to handle biological complexity we must come up with abstract models that will necessarily make simplifying assumptions about the systems we study. Importantly though, abstraction must be accompanied by real world validation – for a model to be justifiable it must be shown to fit experimental data. Indeed, there is a co-dependency between experiments and models – inconsistency between data and a model requires the model to be improved to fit, and behaviours predicted by models need to be validated experimentally.

[H2] Levels of abstraction

Abstraction is key to modelling complex systems. While all computational models are necessarily based on some form of abstraction, this can vary from low-level differential equations through mathematical logics to behaviour expressed as computation, i.e., computer programs. There is no "magic bullet" level of abstraction suitable for all purposes. Biology can learn from the approach taken by computer engineers to design complex hardware systems: a microprocessor cannot be understood in terms of the behaviour of its constituent transistors and logic gates – the detail is overwhelming. Higher level abstractions such as "instruction set architecture [G]" are used to explain the internal operation of the electronics in a form that is meaningful to a computer programmer. Similarly, the behaviour of a cell cannot be understood entirely by studying its chemistry at a molecular level. Dealing with the complexity of biological systems similarly needs a structured toolkit of biological abstractions.

In formal terms such a toolkit is effectively a graph (in the mathematical sense of nodes and edges) of modelling approaches encompassing different forms of abstraction with arcs expressing equivalences between entities appearing in each form of abstraction[40]. So, a key question to be answered is what are the best forms of abstraction and modelling methods to meet the needs of biologists? This question has many of the aspects of discussing the design of a computer programming language – it has to have sufficient expressive power to capture a wide range of biological knowledge yet remain accessible to practitioners and not just of relevance to computational biologists. In this respect we see use of computer program-like approaches being particularly attractive as many biologists have some training in programming or have people in their research group who have such training and if models are executable as programs they can be run as simulations to look for unknown emergent behaviours. But as computer scientists have found, resolving the tensions between scope of application, expressive power and accessibility is a thorny problem.

[H2] Spatial modelling

Spatial aspects such as the extent and structure of a tumour also impact its evolution and reaction to treatments, and so discrete, continuous and mixed methods to model this have been developed (Box 2). Extending an existing computational model by adding a spatial dimension brings the task of integrating continuous physical properties of a biological system with discrete cell signalling all in one modelling framework. Such hybrid models are challenging on several levels. First, they need to combine very different forms of abstraction (e.g., Boolean networks and ODEs) into one model and allow the different abstractions to talk to one another. Secondly, the analysis of hybrid models usually cannot go beyond simulation: sadly, model checking is not an option here as it is only applicable to models with a finite number of states. Innovative methods for the construction and analysis of hybrid computational models are very much needed. However, combining a discrete representation of cells with continuous modelling of the microenvironment has been shown to yield key benefits over a purely continuous formalism[104], so there are rewards for overcoming these challenges.

[H2] Automated model construction

Executable modelling as described so far is essentially a manual process. Automatic processes are attractive as they are potentially less labour intensive and more scalable to multiple patients and diseases. So, an important question for computational modelling is whether there is the possibility of learning models from data? One possible approach that can be borrowed from computer science is "program synthesis [G]", which is a method to produce programs automatically from high-level specifications relating "inputs" to "outputs". In the context of biological modelling, synthesis would correspond to splitting data into inputs (e.g., biomarkers, treatments) and outputs (e.g., tumour type, clinical outcomes) and from these automatically constructing "behaviours" that tie them together. We anticipate such an approach could accelerate model development but would still need some level of engagement with the biologist to steer the synthesis algorithm [G] to biologically plausible mechanisms.

A further challenge with program synthesis is the issue of scalability. On the one hand, there has been success in automating the construction of computational models from mutation experimental data[105] and high-throughput gene expression measurements[36]. Current combinatorial synthesis methods include the Single Cell Network Synthesis tool[106], which reconstructs Boolean network models from single-cell gene expression data, as demonstrated by a model of the regulatory network controlling haematopoitic development[36]. Additionally, the Reasoning Engine for Interaction Networks tool[107], is a method to narrow down a set of possible Boolean network models to those that reproduce a set of known experimental behaviours, for example to model the choice between self-renewal and pluripotency in mouse embryonic stem cells. On the other hand, these tools do not scale up sufficiently to describe mechanisms as complex as those found in cancer. There is ongoing exploration on how to use machine learning algorithms in combination with program synthesis to reduce the search space [G] and provide more general solutions that can be applied to a wider range of datasets[108].

[H2] Model verification and predictive power

The use of formal verification methods allows validation of all possible behaviours of an executable model, and so confirmation that a model is able to reproduce a specification of all known behaviours of a system. This is only as good as the knowledge of the ground truth, and so the clinical utility of these models will depend on the ability to improve the measurement of patient status with emerging technologies such as next generation sequencing and proteomics. However, while the number of predictions these models make has long outstripped out capacity to validate them experimentally, new techniques such as CRISPR-Cas screening[109] mean that we can now test, *in vitro*, changes to gene regulatory and signalling networks at a similar scale as we can model. This will allow more thorough validation of these models and better understanding of their trustworthiness and evaluation of the reliability of their predictions, which will be vital to safely integrating them into clinical use.

[H2] Digital clinical advisor for improved patient outcomes

In addition to their contribution to the science of cancer, there are already examples of computational and mathematical models being used in clinical practice to improve decision making. Because of their ability to represent treatments such as drugs or radiotherapy as perturbations to the default evolution of the cancer, computational models have the potential to integrate combinations of therapies into an overall treatment plan that gives the best patient outcomes. The obvious next step is to take this forward using patient-specific data to create a personalised executable model, derived from learned population models of different types of cancer (Figure 3). One can think of such a model as an extension of a personal electronic health record including images, omics data, treatments and outcomes, with a digital avatar with the ability to run a forward prognosis of the patient's susceptibility to different forms of cancer, enabling early detection and, when detected, predict personalised treatments. As the disease progresses, the model can be rerun on new data allowing the treatment to be optimised as it proceeds. This leads naturally to the concept of a "clinical advisor" app, that interacts with a patient "digital avatar" and allows cancer to be treated as a chronic rather than acute condition, with hopefully better survivability and quality of life for patients.

[H1] Concluding remarks

Executable computational modelling is now established as an important tool for understanding complex biological systems, sitting alongside in vivo and in vitro experimentation, mathematical modelling and data science. Executable models bring particular benefits to the exploration of dynamic, evolving systems such as those manifested in cancer and have led to substantial steps forward in our understanding of the underlying mechanisms of this insidious disease. Executable computational modelling is being exploited to speed up the discovery of new therapies and treatment regimes. The current frontier is to bring such modelling into the clinical context through tools that allow patient-specific management of cancer through early prediction and adaptive therapies. The hope is that by such means cancer can be changed from being an acute medical condition to one that is regarded as a manageable chronic condition with significantly improved patient outcomes. From the perspective of those who worked in the area of computational modelling since the earliest days, the approach has reached a level of maturity and capability to make it a key methodology inthe framework of modern biology and act as an enabler for new clinical approaches to the treatment and management of cancer patients.

References

1.   Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

2.   Siegal, M. L. & Bergman, A. Waddington's canalization revisited: Developmental stability and evolution. *Proc. Natl. Acad. Sci.* **99**, 10528–10532 (2002).

3.   Gerlinger, M. *et al.* Cancer: evolution within a lifetime. *Annu. Rev. Genet.* **48**, 215–236 (2014).

4.   Tabassum, D. P. & Polyak, K. Tumorigenesis: it takes a village. *Nat. Rev. Cancer* **15**, 473–483 (2015).

5.   Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–4 (2001).

6.   Echeverria, G. V. *et al.* Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Sci. Transl. Med.* **11**, eaav0936 (2019).

7.   Morken, J. D., Packer, A., Everett, R. A., Nagy, J. D. & Kuang, Y. Mechanisms of Resistance to Intermittent Androgen Deprivation in Patients with Prostate Cancer Identified by a Novel Computational Method. *Cancer Res.* **74**, 3673–3683 (2014).

8.   Ben-David, U., Beroukhim, R. & Golub, T. R. Genomic evolution of cancer models: perils and opportunities. *Nat. Rev. Cancer* **19**, 97–109 (2019).

9.   Morgan, P. *et al.* Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nat. Rev. Drug Discov.* **17**, 167–181 (2018).

10.   Cook, D. *et al.* Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13**, 419–431 (2014).

11.   Pammolli, F., Magazzini, L. & Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discov.* **10**, 428–438 (2011).

12.   Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012).

13.   Gleeson, M. P., Hersey, A., Montanari, D. & Overington, J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Discov.* **10**, 197–208 (2011).

14.   Roerink, S. F. *et al.* Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* **556**, 457–462 (2018).

15.   Beshiri, M. L. *et al.* A PDX/Organoid biobank of advanced prostate cancers captures genomic and phenotypic heterogeneity for disease modeling and therapeutic screening. *Clin. Cancer Res.* **24**, 4332–4345 (2018).

16.   Van De Wetering, M. *et al.* Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* **161**, 933–945 (2015).

17.   Sachs, N. *et al.* A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. *Cell* **172**, 373-386.e10 (2018).

18.   Vlachogiannis, G. *et al.* Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* **359**, 920–926 (2018).

19.   Li, X. *et al.* Organoid cultures recapitulate esophageal adenocarcinoma heterogeneity providing a model for clonality studies and precision therapeutics. *Nat. Commun.* **9**, 1–13 (2018).

20.   Kersten, K., Visser, K. E., Miltenburg, M. H. & Jonkers, J. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol. Med.* **9**, 137–153 (2017).

21.   Villacorta-Martin, C., Craig, A. J. & Villanueva, A. Divergent evolutionary trajectories in transplanted tumor models. *Nat. Genet.* **49**, 1565–1566 (2017).

22.   Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).

23.   Di Ventura, B., Lemerle, C., Michalodimitrakis, K. & Serrano, L. From in vivo to in silico biology and back. *Nature* **443**, 527–533 (2006).

24.   Fisher, J. & Henzinger, T. A. Executable cell biology. *Nat. Biotechnol.* **25**, 1239–49 (2007).

25.   Fisher, J., Piterman, N., Hubbard, E. J. A., Stern, M. J. & Harel, D. Computational insights into Caenorhabditis elegans vulval development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1951–1956 (2005).

26.   Nusser-Stein, S. *et al.* Cell-cycle regulation of NOTCH signaling during C. elegans vulval development. *Mol. Syst. Biol.* **8**, 1–14 (2012).

27.   Fisher, J., Piterman, N., Hajnal, A. & Henzinger, T. A. Predictive modeling of signaling crosstalk during C. elegans vulval development. *PLoS Comput. Biol.* **3**, e92 (2007).

28.   Kirouac, D. C. *et al.* Computational Modeling of ERBB2-Amplified Breast Cancer Identifies Combined ErbB2/3 Blockade as Superior to the Combination of MEK and AKT Inhibitors. *Sci. Signal.* **6**, ra68–ra68 (2013).

29.   Moignard, V. *et al.* Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* **15**, 363–72 (2013).

30.   Ahmed, Z. *et al.* Bringing LTL Model Checking to Biologists. in *Verification, Model Checking, and Abstract Interpretation* (eds. Bouajjani, A. & Monniaux, D.) 1–13 (Springer International Publishing, 2017).

31.   Hall, B. A., Piterman, N. & Fisher, J. Linear Temporal Logic for Biologists in BMA. in *Computational Methods in Systems Biology* (eds. Bartocci, E., Lio, P. & Paoletti, N.) **9859**, 348–350 (Springer International Publishing, 2016).

32.   Fisher, J. & Piterman, N. Model Checking in Biology. in *A Systems Theoretic Approach to Systems and Synthetic Biology I: Models and System Characterizations* (eds. Kulkarni, V. V, Stan, G.-B. & Raman, K.) 255–279 (Springer Netherlands, 2014).

33.   Clarke Jr, E. M., Grumberg, O. & Peled, D. A. *Model checking*. (MIT press, 1999).

34.   Chuang, R. *et al.* Drug Target Optimization in Chronic Myeloid Leukemia Using Innovative Computational Platform. *Sci. Rep.* **5**, 8190 (2015).

35.   Duarte, N. C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci.* **104**, 1777–1782 (2007).

36.   Moignard, V. *et al.* Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33**, 269–76 (2015).

37.   Jolly, M. K. & Levine, H. Computational systems biology of epithelial-hybrid-mesenchymal transitions. *Curr. Opin. Syst. Biol.* **3**, 1–6 (2017).

38.   Margolin, A. A. *et al.* ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* **7**, S7 (2006).

39.   Markowetz, F. & Spang, R. Inferring cellular networks--a review. *BMC Bioinformatics* **8 Suppl 6**, S5 (2007).

40.   Fisher, J., Piterman, N. & Vardi, M. Y. The Only Way Is Up. in *FM 2011: Formal Methods: 17th International Symposium on Formal Methods, Limerick, Ireland, June 20-24, 2011. Proceedings* (eds. Butler, M. & Schulte, W.) 3–11 (Springer Berlin Heidelberg, 2011).

41.   Samaga, R., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K. & Klamt, S. The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput. Biol.* **5**, e1000438 (2009).

42.   Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D. & Muüller, G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* **20**, 370–375 (2002).

43.   Saez-Rodriguez, J. *et al.* Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.* **5**, 1–19 (2009).

44.   Tyson, J. J., Chen, K. C. & Novak, B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–31 (2003).

45.   Paterson, Y. Z. *et al.* A toolbox for discrete modelling of cell signalling dynamics. *Integr. Biol.* **10**, 370–382 (2018).

46.   Pina, C. *et al.* Inferring rules of lineage commitment in haematopoiesis. *Nat. Cell Biol.* **14**, 287–294 (2012).

47.   Fisher, J., Henzinger, T. A., Mateescu, M. & Piterman, N. Bounded asynchrony: Concurrency for modeling cell-cell interactions. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **5054 LNBI**, 17–32 (2008).

48.   Hall, B. A., Piterman, N., Hajnal, A. & Fisher, J. Emergent stem cell homeostasis in the C. elegans germline is revealed by hybrid modeling. *Biophys. J.* **109**, 428–38 (2015).

49.   Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).

50.   Mandon, H., Haar, S. & Paulevé, L. Temporal Reprogramming of Boolean Networks. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **10545 LNBI**, 179–195 (2017).

51.   Iwasaki, H. *et al.* The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. *Genes Dev.* **20**, 3010–3021 (2006).

52.   Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).

53.   Zañudo, J. G. T. & Albert, R. Cell Fate Reprogramming by Control of Intracellular Network Dynamics. *PLoS Comput. Biol.* **11**, 1–24 (2015).

54.   Steinway, S. N. *et al.* Network modeling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Res.* **74**, 5963–5977 (2014).

55.   Nieto, M. A., Huang, R. Y.-J., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).

56.   Vahedi, G., Faryabi, B., Chamberland, J. F., Datta, A. & Dougherty, E. R. Sampling-rate-dependent probabilistic Boolean networks. *J. Theor. Biol.* **261**, 540–547 (2009).

57.   Sizek, H., Hamel, A., Deritei, D., Campbell, S. & Regan, E. R. Boolean model of growth signaling, cell cycle and apoptosis predicts the molecular mechanism of aberrant cell cycle progression driven by hyperactive PI3K. *PLoS Comput. Biol.* **15**, 1–43 (2019).

58.   Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319**, 525–32 (1988).

59.   Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).

60.   Auslander, N., Wolf, Y. I. & Koonin, E. V. In silico learning of tumor evolution through mutational time series. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9501–9510 (2019).

61.   Fumiã, H. F. & Martins, M. L. Boolean Network Model for Cancer Pathways: Predicting Carcinogenesis and Targeted Therapy Outcomes. *PLoS One* **8**, (2013).

62.   Spencer, S. L., Berryman, M. J., García, J. A. & Abbott, D. An ordinary differential equation model for the multistep transformation to cancer. *J. Theor. Biol.* **231**, 515–524 (2004).

63.   Clarke, M. A., Woodhouse, S., Piterman, N., Hall, B. A. & Fisher, J. Using State Space Exploration to Determine How Gene Regulatory Networks Constrain Mutation Order in Cancer Evolution. in *Automated Reasoning for Systems Biology and Medicine* 133–153 (Springer, Cham, 2019).

64.   Caravagna, G. *et al.* Detecting repeated cancer evolution from multi-region tumor sequencing data. *Nat. Methods* **15**, 707–714 (2018).

65.   Swanton, C. Intratumor heterogeneity: Evolution through space and time. *Cancer Res.* **72**, 4875–4882 (2012).

66.   Sun, Q.-Y. *et al.* Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). *Leukemia* **31**, 1–10 (2017).

67.   Lee, J. K. *et al.* Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nat. Genet.* **49**, 594–599 (2017).

68.   Choi, M., Shi, J., Zhu, Y., Yang, R. & Cho, K.-H. Network dynamics-based cancer panel stratification for systemic prediction of anticancer drug response. *Nat. Commun.* **8**, 1940 (2017).

69.   Chong, C. R. & Jänne, P. A. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat. Med.* **19**, 1389–400 (2013).

70.   Wagle, N. *et al.* Dissecting therapeutic resistance to RAF inhibition in melanoma by

tumor genomic profiling. *J. Clin. Oncol.* **29**, 3085–3096 (2011).

71.   Bozic, I. *et al.* Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife* **2**, e00747 (2013).

72.   Al-Lazikani, B., Banerji, U. & Workman, P. Combinatorial drug therapy for cancer in the post-genomic era. *Nat. Biotechnol.* **30**, 679–692 (2012).

73.   Silverbush, D. *et al.* Cell-Specific Computational Modeling of the PIM Pathway in Acute Myeloid Leukemia. *Cancer Res.* **77**, 827–838 (2017).

74.   Shorthouse, D. *et al.* Exploring the role of stromal osmoregulation in cancer and disease using executable modelling. *Nat. Commun.* **9**, 3011 (2018).

75.   Layek, R., Datta, A., Bittner, M. & Dougherty, E. R. Cancer therapy design based on pathway logic. *Bioinformatics* **27**, 548–555 (2011).

76.   von der Heyde, S. *et al.* Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines. *BMC Syst. Biol.* **8**, 75 (2014).

77.   Zañudo, J. G. T., Scaltriti, M. & Albert, R. A network modeling approach to elucidate drug resistance mechanisms and predict combinatorial drug treatments in breast cancer. *Cancer Converg.* **1**, 5 (2017).

78.   McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15–26 (2015).

79.   Bhang, H. E. C. *et al.* Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21**, 440–448 (2015).

80.   Kreuzaler, P. *et al.* Heterogeneity of Myc expression in breast cancer exposes pharmacological vulnerabilities revealed through executable mechanistic modeling. *Proc. Natl. Acad. Sci.* **116**, 22399–22408 (2019).

81.   Van Hasselt, J. G. C. & Van Der Graaf, P. H. Towards integrative systems pharmacology models in oncology drug development. *Drug Discov. Today Technol.* **15**, 1–8 (2015).

82.   Kirouac, D. C. & Onsum, M. D. Using network biology to bridge pharmacokinetics and pharmacodynamics in oncology. *CPT pharmacometrics Syst. Pharmacol.* **2**, e71 (2013).

83.   Grimwade, D. *et al.* The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* **92**, 2322–33 (1998).

84.   Liersch, R., Müller-Tidow, C., Berdel, W. E. & Krug, U. Prognostic factors for acute myeloid leukaemia in adults - biological significance and clinical use. *Br. J. Haematol.* **165**, 17–38 (2014).

85.   Solin, L. J. *et al.* A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. *J. Natl. Cancer Inst.* **105**, 701–710 (2013).

86.   Kurian, A. W. *et al.* Recent Trends in Chemotherapy Use and Oncologists' Treatment Recommendations for Early-Stage Breast Cancer. *J. Natl. Cancer Inst.* **110**, 493–500 (2018).

87.   Sparano, J. A. *et al.* Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N. Engl. J. Med.* **379**, 111–121 (2018).

88.   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**, 1816 (2017).

89.   US National Library of Medicine. *ClinicalTrials.gov* (2015). Available at: https://clinicaltrials.gov/ct2/show/NCT02415621.

90.   Crook, J. M. *et al.* Intermittent Androgen Suppression for Rising PSA Level after Radiotherapy. *N. Engl. J. Med.* **367**, 895–903 (2012).

91.   US National Library of Medicine. *ClinicalTrials.gov* (2003). Available at: https://clinicaltrials.gov/ct2/show/NCT00003653.

92.   Hussain, M. *et al.* Intermittent vs continuous androgen deprivation in mPC-Intergroup data (Intermittent versus Continuous Androgen Deprivation in Prostate Cancer. *N. Engl. J. Med.* **368**, 1314–1325 (2013).

93.   US National Library of Medicine. *ClinicalTrials.gov* (2003). Available at:

https://clinicaltrials.gov/ct2/show/NCT00002651.

94.   West, J. B. *et al.* Multidrug Cancer Therapy in Metastatic Castrate-Resistant Prostate Cancer: An Evolution-Based Strategy. *Clin. Cancer Res.* **25**, 4413–4421 (2019).

95.   Gatenby, R. A., Silva, A. S., Gillies, R. J. & Frieden, B. R. Adaptive therapy. *Cancer Res.* **69**, 4894–903 (2009).

96.   Enderling, H., Alfonso, J. C. L., Moros, E., Caudell, J. J. & Harrison, L. B. Integrating Mathematical Modeling into the Roadmap for Personalized Adaptive Radiation Therapy. *Trends in Cancer* **5**, 467–474 (2019).

97.   Sidders, B. *et al.* Network-Based Drug Discovery: Coupling Network Pharmacology with Phenotypic Screening for Neuronal Excitability. *J. Mol. Biol.* **430**, 3005–3015 (2018).

98.   Chindelevitch, L. *et al.* Causal reasoning on biological networks: Interpreting transcriptional changes. *Bioinformatics* **28**, 1114–1121 (2012).

99.   Lefebvre, C. *et al.* A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).

100. Rodriguez-Barrueco, R. *et al.* Inhibition of the autocrine IL-6–JAK2–STAT3–calprotectin axis as targeted therapy for HR−/HER2+ breast cancers. *Genes Dev.* **29**, 1631–1648 (2015).

101. US National Library of Medicine. *ClinicalTrials.gov* (2014). Available at: https://clinicaltrials.gov/ct2/show/NCT02066532.

102. US National Library of Medicine. *ClinicalTrials.gov* (2014). Available at: https://clinicaltrials.gov/ct2/show/NCT03211988.

103. Alvarez, M. J. *et al.* A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat. Genet.* **50**, 979–989 (2018).

104. Anderson, A. R. A., Weaver, A. M., Cummings, P. T. & Quaranta, V. Tumor Morphology and Phenotypic Evolution Driven by Selective Pressure from the Microenvironment. *Cell* **127**, 905–915 (2006).

105. Koksal, A. S. *et al.* Synthesis of biological models from mutation experiments. in *Principles of Programming Languages (POPL)* 469 (ACM Press, 2013).

106. Woodhouse, S., Piterman, N., Wintersteiger, C. M., Göttgens, B. & Fisher, J. SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC Syst. Biol.* **12**, 59 (2018).

107. Dunn, S.-J. S.-J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. G. Defining an essential transcription factor program for naïve pluripotency. *Science* **344**, 1156–60 (2014).

108. Fisher, J. & Woodhouse, S. Program synthesis meets deep learning for decoding regulatory networks. *Curr. Opin. Syst. Biol.* **4**, 64–70 (2017).

109. Fellmann, C., Gowen, B. G., Lin, P. C., Doudna, J. A. & Corn, J. E. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat. Rev. Drug Discov.* **16**, 89–100 (2017).

110. Materi, W. & Wishart, D. S. Computational systems biology in drug discovery and development: methods and applications. *Drug Discov. Today* **12**, 295–303 (2007).

111. Baker, R. E., Peña, J. M., Jayamohan, J. & Jérusalem, A. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol. Lett.* **14**, 1–4 (2018).

112. Johnston, M. D., Edwards, C. M., Bodmer, W. F., Maini, P. K. & Chapman, S. J. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 4008–4013 (2007).

113. Du, W. *et al.* Effective combination therapies for B-cell lymphoma predicted by a virtual disease model. *Cancer Res.* **77**, 1818–1830 (2017).

114. Aldridge, B. B., Burke, J. M., Lauffenburger, D. A. & Sorger, P. K. Physicochemical modelling of cell signalling pathways. *Nat. Cell Biol.* **8**, 1195–203 (2006).

115. Csajka, C. & Verotta, D. Pharmacokinetic-pharmacodynamic modelling: History and perspectives. *J. Pharmacokinet. Pharmacodyn.* **33**, 227–279 (2006).

116. Gevaert, O., De Smet, F., Timmerman, D., Moreau, Y. & De Moor, B. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian

networks. *Bioinformatics* **22**, 184–190 (2006).

117. Cruz-Ramírez, N., Acosta-Mesa, H. G., Carrillo-Calvet, H., Alonso Nava-Fernández, L. & Barrientos-Martínez, R. E. Diagnosis of breast cancer using Bayesian networks: A case study. *Comput. Biol. Med.* **37**, 1553–1564 (2007).

118. Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J. & Jarvis, E. D. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* **20**, 3594–3603 (2004).

119. Gerstung, M., Baudis, M., Moch, H. & Beerenwinkel, N. Quantifying cancer progression with conjunctive Bayesian networks. *Bioinformatics* **25**, 2809–2815 (2009).

120. Von Heydebreck, A., Gunawan, B. & Füzesi, L. Maximum likelihood estimation of oncogenetic tree models. *Biostatistics* **5**, 545–556 (2004).

121. Beerenwinkel, N., Schwarz, R. F., Gerstung, M. & Markowetz, F. Cancer evolution: Mathematical models and computational inference. *Syst. Biol.* **64**, e1–e25 (2015).

122. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–52 (2012).

123. Rueda, O. M. *et al.* Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature* **567**, 399–404 (2019).

124. Lee, A. *et al.* BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet. Med.* **21**, 1708–1718 (2019).

125. Geman, D., Ochs, M., Price, N. D., Tomasetti, C. & Younes, L. An argument for mechanism-based statistical inference in cancer. *Hum. Genet.* **134**, 479–495 (2015).

126. Halasz, M., Kholodenko, B. N., Kolch, W. & Santra, T. Integrating network reconstruction with mechanistic modeling to predict cancer therapies. *Sci. Signal.* **9**, ra114 (2016).

127. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **13**, 8–17 (2015).

128. Ching, T. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **15**, 20170387 (2018).

129. Ma, J. *et al.* Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **15**, 290–298 (2018).

130. Karolak, A., Markov, D. A., McCawley, L. J. & Rejniak, K. A. Towards personalized computational oncology: From spatial models of tumour spheroids, to organoids, to tissues. *J. R. Soc. Interface* **15**, (2018).

131. Metzcar, J., Wang, Y., Heiland, R. & Macklin, P. A Review of Cell-Based Computational Modeling in Cancer Biology. *JCO Clin. cancer informatics* **3**, 1–13 (2019).

132. Macklin, P., Edgerton, M. E., Cristini, V. & Lowengrub, J. Discrete cell modeling. in *Multiscale Modeling of Cancer* 88–122 (Cambridge University Press, 2009).

133. Giatili, S. G. & Stamatakos, G. S. A detailed numerical treatment of the boundary conditions imposed by the skull on a diffusion-reaction model of glioma tumor growth. Clinical validation aspects. *Appl. Math. Comput.* **218**, 8779–8799 (2012).

134. Anderson, A. R. A. *et al.* Microenvironmental independence associated with tumor progression. *Cancer Res.* **69**, 8797–8806 (2009).

135. Rejniak, K. A. & Anderson, A. R. A. Hybrid models of tumor growth. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **3**, 115–125 (2011).

136. Osborne, J. M. *et al.* A hybrid approach to multi-scale modelling of cancer. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **368**, 5013–5028 (2010).

137. Powathil, G. G., Gordon, K. E., Hill, L. A. & Chaplain, M. A. J. Modelling the effects of cell-cycle heterogeneity on the response of a solid tumour to chemotherapy: Biological insights from a hybrid multiscale cellular automaton model. *J. Theor. Biol.* **308**, 1–19 (2012).

138. Hofree, M., Shen, J. P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nat. Methods* **10**, 1108–1115 (2013).

139. Kim, J. W. *et al.* Decomposing Oncogenic Transcriptional Signatures to Generate Maps of Divergent Cellular States. *Cell Syst.* **5**, 105-118.e9 (2017).

140. Ashcroft, P., Michor, F. & Galla, T. Stochastic Tunneling and Metastable States During the Somatic Evolution of Cancer. *Genetics* **199**, 1213–1228 (2015).

141. Stites, E. C., Trampont, P. C., Ma, Z. & Ravichandran, K. S. Network Analysis of Oncogenic Ras Activation in Cancer. *Science* **318**, 463–467 (2007).

142. Akhmetzhanov, A. R. *et al.* Modelling bistable tumour population dynamics to design effective treatment strategies. *J. Theor. Biol.* **474**, 88–102 (2019).

Acknowledgments

J.F. conceived the ideas for this article and structured the manuscript. J.F. and M.A.C. researched data for the article, wrote, reviewed and edited the manuscript. J.F. conceived the figures. J.F. and M.A.C. designed the figures.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Cancer thanks P. Tamayo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Glossary

Discrete model: a model with a countable number of states. States represent remembered history of the system modeled. Discrete models are to be contrasted with continuous models.

Continuous model: a model with an infinite number of states, which may also be called analogue models.

State machine: an abstract model of a discrete system. A state machine can only be in exactly one of a countable (often finite) number of states at any given time. The machine can change from one state to another in response to some external (input) events (input signals); the change from one state to another is called a (state) transition and may give rise to external (output) signals.

Algorithm: a step-by-step sequence of basic operations required to produce a desired result in a discrete system.

Computer programs: collections of instructions that perform a specific algorithm when executed by a computer.

Specification: a set of known behaviours which a model must be able to produce in the correct circumstances in order to be considered valid.

Formal verification: a method to prove or disprove the correctness of computer programs with respect to a certain formal specification or property, by treating the program as a mathematical structure and proving theorems about it.

Testing: the process of checking the consistency of a program with a given specification by comparing inputs to outputs across multiple runs of the program.

Model checking: a means of checking whether a program meets a given specification by the used of automated theorem proving (a branch of mathematical logic dealing with proving mathematical theorems by computer programs).

Program synthesis: a technique that automatically constructs a computer program that satisfies a given high-level specification.

Abstraction: a model at a certain level of description, simplifying lower-level details in a principled way preserving key properties of the system behaviour.

Levels of abstraction: a hierarchy of abstractions where higher levels of abstraction are placed at the top and more detailed concepts underneath.

Concurrency: the parallel execution of multiple interacting computer programs.

Non-determinism: the abstraction of a complex behaviour showing more than one possible output (e.g., phenotype) for a given input (e.g., genotype).

Modularity: the focus on keeping components of a model or a computer program in discrete units allowing them to be flexibly put together in different combinations.

Compositionality: the ability to combine separate component models into a larger overall system model.

Instruction set architecture: an abstract description of a computer processer at the level required by programmers, including those writing compilers for high-level programming languages (e.g., Python).

Attractors: states toward which a system tends to evolve from a wide variety of starting conditions. A system in an attractor state remains stable even if subject to perturbation.

Adaptive Therapy: the application of cancer treatment in a manner which quickly responds to changes in the disease rather than following a fixed protocol, with the goal of managing the cancer and maintaining a limited tumour burden, rather than attempting to totally eliminate the disease[95].

Search Space: the set of possible states or solutions through which an algorithm must search to find the optimum solution to some problem.

Phase Plane: a 2-dimensional visualisation of the behaviour of a system of differential equations where each axis shows one variable of the equations. Often used to aid the visualisation of the long-term behaviour of these systems. Higher dimensional visualisations are refered to as a phase space. For example, the Lotka-Volterra equations model the change in predator and prey populations over time. Plotting the number of predators or prey against time would show oscillatory behaviour, but plotting prey vs predator would reveal a closed loop in the phase plane, revealing the balanced tradeoff in the number of predators and prey.

Logic Gates: a representation of a Boolean logical operation. Combines binary inputs to produce a binary output based on operations such as AND, OR or XOR (exclusive OR).

Table 1 Comparison of Models

Model	Discrete vs continuous	Input data	Construction	Outputs	Key insights	Mechanistic explanation	Reusability and extensibility	Strengths	Challenges	Examples of success
Network	Discrete	Quantitative (e.g. gene expression data), qualitative (e.g. transcription factor targets) or both	Can be inferred from one dataset (top-down) or built up from multiple datasets (bottom up)	Stable states of the system (attractors), attractor basins and ordering of events	Ordering and connection between molecular events; identifying all possible states of the system	Yes	Easy	Flexible in terms of data to be built from, yield precise knowledge of all possible behaviours	Automatic network inference and synthesis	43,68,80,138
Statistical	Continuous	Quantitative (e.g. gene expression data) or qualitative (e.g. tumour images of whether or not cancer regression occurs)	Inferred from data (top down); high volumes needed for machine learning approaches	Classification, clustering and regression	Structure of high-dimensional and high-volume data	No	Difficult	Able to extract patterns from large volumes of high-dimensional data	Model interpretability and selection	38,122,129,139,140
ODE	Continuous	Quantitative (kinetic parameter measurements)	Construction of models from different hypothesised processes (bottom up); estimation of parameters	Stability analysis, relationship between variables (phase planes [G]) and effects of changes to parameters (sensitivity analysis)	Precise prediction of future behaviour of continuous systems	Yes	Possible	Precise predictions and change over time; ability to investigate stability and sensitivity	Parameter measurement or estimation, analysis of non-linear systems; model identifiability	42,88,104,141,142

Figure Legends

Figure 1. Computational modelling iterative cycle. The first step is to analyse experimental data to suggest a mechanistic hypothesis that is then formalised as a computational model. The second step is to compare rigorously the computational model to the experimental data, in a formal way (i.e., using formal verification). If the model fails to reproduce the data, the model needs refinement until it fully conforms with the data. The third step is to use the computational model to make predictions of behaviours outside the experimental data. These predictions are then checked experimentally and if validated are added to the experimental dataset.

Figure 2. Exploring cancer therapies using a computational model. A schematic for a computational model of a signalling network (model) in a healthy cell (beige cells) vs a cancer cell (blue cells) and the output cell phenotypes produced by the model (output). Numerical values in purple boxes to the right of nodes in the signalling network represent level of activity for proteins (e.g., 0 represents low or no activity, 1 represents moderate activity and 2 represents high activity). Values to the right of the phenotypic outputs (e.g., proliferation) can represent absence (0), moderate level (1) or high level (2) of cell behaviours, or the presence (1) or absence (0) of a cell property (e.g., growth arrest) – all examples of cellular phenotypic outputs. The difference between a healthy cell and a cancer cell is simulated by setting the initial level of relevant nodes (e.g. to 2 for an activating mutation in a gene encoding a relevant protein, to 0 for inhibition by a drug) and read by the phenotypic outputs predicted by the model. The top row of cells illustrates the inputs and outputs of simulating untreated cells (healthy vs cancer). The bottom row of cells shows the simulation of drug treatments, and their effect on the phenotypic output as predicted by the computational model. As illustrated, if the treatment is of low toxicity to healthy cells, the phenotypic output is less than the effect on the cancer cells. If the treatment is effective when applied to the cancer cells there will be a corresponding phenotypic output indicating a decrease in disease state.

Figure 3. A possible digital avatar for a cancer patient. Schematic showing how different datasets could be integrated into a computational model leading to a digital avatar that could advise clinical decision making. Computational models of different types of cancer would be built based on the integration of diverse biomedical datasets. This library of computational cancer models could then be personalised with individual patient data to identify personalised therapies. As these models are applied across many different patients, their successes and failures could be used to further improve the models by leveraging machine learning techniques to identify principal features linking patient data to cancer stage and evolution. This would then feedback into ever more accurate predictions of the most effective therapies. Cloud repository allows such datasets to   be aggregated across clinical populations and subject to techniques such as deep learning.

Box 1. Computational and Mathematical Models

A computational model is a discrete model which prescribes a sequence of steps or instructions to be executed by an idealised computer that defines the semantics of the model. Since the model is computational it can be either directly executed by a real computer or transformed algorithmically into a suitable form for execution.

A mathematical model is a continuous model using mathematical equations to show the relationship between quantities and how they change over time. There may be many different algorithms to solve the equations, either analytical methods for deriving exact solutions or, e.g., numerical methods for deriving approximate solutions. The application of such methods to biology is reviewed by Materi et al.[110]

Computational modelling versus data science

A comprehensive understanding of cancer is necessarily grounded by analysis and integration of pre-clinical data (i.e., genomic, transcriptomic, proteomic, metabolomic and phenotypic cellular data) with data more commonly measured in the clinic (e.g., radiomics, histopathology, liquid biopsies). Such data integration collectively analyses the different datasets and builds an overarching data model for the aggregated data. Machine learning approaches can then be applied to look for patterns in the aggregated data that might provide opportunities for improved diagnostics or treatment pathways based on experience to date. While not denying the value of such insights, machine learning is a 'black box' approach and essentially a form of interpolation. It is not able to directly generate mechanistic insights nor is it able to predict outcomes outside of those captured in the data used for training[111]. By contrast, computational modelling, by definition, gives a mechanistic explanation of the data from which it is derived and is able to be predictive of the result of perturbing a system beyond the scope of that originating data.

Box 2. Mathematical models in cancer biology

[H1] ODE and PDE models

Systems of ordinary differential equations (ODEs) and partial differential equations (PDEs) are widely used in physics and engineering to predict how interdependent variables change over time. These often cannot be solved analytically, and so solutions must be estimated numerically. However, there are a wide array of frameworks for analysis of how the system evolves (stability analysis and phase planes) and how sensitive the behavior of the system is to changes to different parameters (sensitivity analysis). It is challenging to accurately measure or fit parameters, especially in non-linear cases. They can be used to model the population of cells in a tumour[94,112], or chemical reactions, such as in a regulatory pathway[42,113,114]. An important branch of these models is pharmacokinetic/dynamic (PKPD) models for modelling the effects of drugs[115].

[H1] Probabilistic Models

Probabilistic models concern the probability that random variables will take certain values. A common use in cancer modelling is probabilistic graphical models, which are graphs that expresses the conditional dependence between variables. Bayesian networks are often used for cancer prognosis[116,117]. Bayesian networks also lend themselves to network inference[118] or modelling cancer evolution[119,120] with further examples in the review by Beerenwinkel et al.[121].

[H1] Statistical models

Statistical models can be used to extract patterns or find parameters to describe large datasets. Patterns within such datasets can be used to correlate groups of particular characteristics with specific outcomes, for example finding cancer prognostic parameters[122]. These descriptive models can also be a first step towards a predictive model[123,124] and mechanistic models[125] using for example network inference[39,126].

[H1] Machine learning

Machine learning methods, such as Artificial Neural Networks or Random Forests, attempt to generalize from a limited training set. Taking advantage of modern computing resources, machine learning methods are able to extract structure from high-volume and high-dimensional datasets. Common applications include classification, clustering and fitting regression models. There has been particular interest in the use of machine learning to leverage the increasing amounts of patient data, such as sequencing and expression data, to improve cancer prognostics[127,128]. These models are often opaque and trade-off explanatory power for predictive accuracy. However, there are various attempts to ameliorate this drawback and extract mechanistic hypotheses from such models[129].

[H1] Spatial models

Malignant tumours are heterogenous shifting ecosystems of cells. There is therefore a need for models which can account for signalling between clones and cell types, and the spatial structure that affects signalling and the microenvironment[130,131]. This may be accomplished in a discrete[132], continuous[133] or hybrid[104,134] models.

[H1] Hybrid and multiscale models

Biological problems cross many different scales and are influenced by processes happening in different regimes. These may be suited to different modelling techniques, yet all contribute to one phenomenon. Therefore, hybrid models that can combine techniques and scales are proving useful[135,136]. Such hybrid and multiscale models include for example: the effects of heterogeneity in cell cycle stage within a tumour on treatment effectiveness[137] and the effects of a harsh microenvironment on tumour invasion and evolution[104], which compares the hybrid model to a prior continuous model, and shows how the hybrid model reveals new biologically important behaviours.

Box 3. Types of computational models

[H1] State machines

State machines are discrete models that define the behaviour of a system over time in response to external signals or events. At each such event, the machine transitions from one state to another, where a state represents remembered knowledge of the history of the system. State machines can be combined to describe how different parts of a larger system communicate: for example, state transitions in one part generating events that trigger transitions elsewhere in the system. The state of a system at any point in time is determined by the states of its component parts. For example, the state of a cell can be represented by the states of various genes and proteins, each having its own reaction to the presence or absence of certain other molecules. Changes in the overall state of the cell are determined by the interdependent state changes of its component parts.

[H1] Boolean networks

Boolean networks represent state machines as directed graphs and are widely used as discrete models of biological networks such as gene regulatory networks. The nodes of such graphs typically represent expression status of genes and the directed edges represent the actions of genes on one another. The state of each node is a Boolean variable (0 or 1) determined by the sum of its inputs.

[H1] Qualitative networks

Qualitative networks are extensions to Boolean networks. These graph models allow the state of each node to have more than Boolean values (e.g., off, low, medium, high). Furthermore, these models have target functions that provide "weight" to the edges in a form of logic gates **[G]** (e.g., AND, OR, MIN, MAX).

[H1] Hybrid models

Hybrid models aim to bridge the gap between mathematical models and computational models by combining the two in a single framework, with variables that span discrete and continuous domains.

Discrete variables are state machines with input events triggered by values of continuous variables. Continuous variables are governed by differential equations, which depend on discrete states represented as numbers. In effect, the discrete part of such models is an executable control mechanism that drives a continuous physical system.

Table of contents summary
This Perspective discusses how executable computational models, integrating various datasets derived from pre-clinical models and cancer patients, can be used to represent the dynamic biological behaviours inherent in cancer. It argues that these models might be used as patient avatars to improve personalised treatments.