

<https://doi.org/10.1038/s41746-024-01402-3>

# Challenges and opportunities for digital twins in precision medicine from a complex systems perspective

Check for updates

Manlio De Domenico<sup>1,2,3,4</sup> ✉, Luca Allegrì<sup>1</sup>, Guido Caldarelli<sup>5,6,7</sup>, Valeria d'Andrea<sup>1,4</sup>, Barbara Di Camillo<sup>2,8,9</sup>, Luis M. Rocha<sup>10,11</sup>, Jordan Rozum<sup>10</sup>, Riccardo Sbarbati<sup>1,4</sup> & Francesco Zambelli<sup>1,4</sup>

Digital twins (DTs) in precision medicine are increasingly viable, propelled by extensive data collection and advancements in artificial intelligence (AI), alongside traditional biomedical methodologies. We argue that including mechanistic simulations that produce behavior based on explicitly defined biological hypotheses and multiscale mechanisms is beneficial. It enables the exploration of diverse therapeutic strategies and supports dynamic clinical decision-making through insights from network science, quantitative biology, and digital medicine.

Precision medicine aims to deliver diagnostic, prognostic, and therapeutic strategies specifically tailored to individuals by explicitly accounting for their genetic information, lifestyle, and environment<sup>1</sup>, which are organized in a network structure<sup>2</sup>. The success of this approach relies on at least two fundamental and non-trivial assumptions: The first is that it is possible to predict the response of a patient to a specific treatment by means of computational, cellular, and organism-based models with reasonable accuracy. The second is that it is possible to use heterogeneous data sources (multiomics, electronic health records, individual and social behavior, and so forth) to build massive databases with enough statistics to stratify a population and characterize the distinctive features of clinical interest<sup>3</sup>.

It is not surprising that the field of precision medicine is growing<sup>4,5</sup>, attracting the interest of both national health systems for investments<sup>6</sup>, and scholars, who span a wide range of disciplines, from molecular biology to computer science, medicine, physics, and engineering. Nevertheless, precision medicine, with its revolutionary promises, is usually associated with clinical genomics<sup>7</sup> and multiomics<sup>8</sup>, with a strong focus on the idea that combining heterogeneous, multi-scale sources of data will lead to timely predictions about individual medical outcomes. Recently, attention has shifted to the possibility of integrating such molecular data with traditional<sup>9–11</sup> and non-traditional<sup>12</sup> data sources of clinical relevance into a multiscale predictive modeling methodology. This leads to the creation of a *digital twin*, and it allows for testing therapeutic strategies in-silico with the ultimate goal of maximizing successful treatments and outcomes in vivo.

**Defining digital twins in precision medicine** The first pioneering precursors of digital twins for personalized medicine came out in the early

2000s and proposed the idea that models of the human body for specific patients could improve clinical practices. They also pointed out challenges that are still valid, such as the need for a structure to handle multiple-source data integration<sup>13</sup> and the importance of having solid mathematical models that can describe the system at the desired level of precision<sup>14</sup>. In recent years, medical digital twins have experienced a huge increase in interest, with the birth of many programs devoted to them<sup>15,16</sup>. Two of the most significant successes in the field are the “artificial pancreas” of the ARCHIMEDES program on diabetes<sup>17</sup> and the mechanistic models of the heart used for cardiovascular disease monitoring and prevention<sup>18,19</sup>. Recent research emphasizes the potential of having a comprehensive model of the human body that could reflect the possible consequences of a perturbation, such as a viral infection<sup>20</sup> or taking a drug<sup>21,22</sup>, on a specific patient. To actually implement these models, now a decade from the first proposals<sup>23</sup>, network and complex systems are starting to be considered<sup>15</sup>. Meanwhile, approaches based on artificial intelligence (AI) and machine learning have been widely adopted despite some limitations and critical aspects<sup>24</sup>.

Given the currently broad spectrum of definitions and applications, it is important to set the operational definition that we adopt throughout this paper Box 1, Box 2.

Broadly speaking, a digital twin exchanges data with its real-world counterpart, synchronizing inputs and outputs; they operate synergistically, with the digital twin informing, controlling, aiding, and augmenting the original system.

Indeed, a digital twin is a virtual replica of a physical system, object, or process. It is designed to reproduce the behavior, conditions, and responses

<sup>1</sup>Department of Physics and Astronomy “Galileo Galilei”, University of Padua, Padua, Italy. <sup>2</sup>Padua Center for Network Medicine, University of Padua, Padua, Italy. <sup>3</sup>Padua Neuroscience Center, University of Padua, Padua, Italy. <sup>4</sup>Istituto Nazionale di Fisica Nucleare, sez. di Padua, Italy. <sup>5</sup>DSMN and ECLT Ca’ Foscari University of Venice, Venezia, Italy. <sup>6</sup>Institute of Complex Systems (ISC) CNR unit Sapienza University, Rome, Italy. <sup>7</sup>London Institute for Mathematical Sciences, Royal Institution, London, UK. <sup>8</sup>Department of Information Engineering, University of Padua, Padua, Italy. <sup>9</sup>Department of Comparative Biomedicine and Food Science, University of Padua, Padua, Italy. <sup>10</sup>School of Systems Science and Industrial Eng., Binghamton University, Binghamton, NY, USA. <sup>11</sup>Universidade Católica Portuguesa, Católica Biomedical Research Centre, Lisbon, Portugal. ✉e-mail: [manlio.dedomenico@unipd.it](mailto:manlio.dedomenico@unipd.it)

## Box 1 | Definition of digital twins in precision medicine

A digital twin is an in-silico framework that replicates a biological cell, sub-system, organ, or a whole organism, with a transparent predictive

model of their relevant causal mechanisms and response to interventions.

## Box 2 | Operational definitions of context-specific terms

**Precision medicine:** Also referred to as personalized medicine, this refers to a medical approach that tailors healthcare to the individual characteristics of each patient, based on their genetic makeup, environment, lifestyle, and other personal data.

**Network model:** A mathematical model to represent a complex system as a set of entities (nodes) connected pairwise by links (edges) to indicate association or interaction. Nodes and edges can be of very different natures, such as, respectively, proteins and physical interaction between them for protein-protein interaction networks, neurons and synapses for neural networks, species and predation interactions for ecological networks.

**Organoid:** A self-organized 3D tissue that is typically derived from stem cells (pluripotent, fetal, or adult) and mimics the key functional, structural and biological complexity of an organ<sup>146</sup>.

**Exposome:** External factors (environmental, social, etc.) that an individual is subjected to throughout life that may be related to their state of health.

**Multiscale models:** Mathematical models that describe interactions within and between components of a system with spatial or temporal scales that differ by orders of magnitude. For example, in gene expression, the folding and unfolding processes of the chromatin typically occurs over hours to days, while transcription occurs over minutes.

**Agent-based model:** A computational simulation that represents a system as a collection of autonomous proxies (e.g., representing cells or individuals) interacting within a defined environment, each with its own characteristics, rules, and behaviors that collectively give rise to complex system dynamics.

**Boolean automata:** An agent in a discrete dynamical model that can take one of two states-ON or OFF. It determines its next state based on the values of pre-specified input variables, which are combined using a fixed Boolean logic function.

**Boolean networks:** A collection of interconnected Boolean automata in which each automata's (node's) inputs correspond to the states of other automata (parent nodes or regulators).

of its real-world counterpart in real time or near-real time. However, while a digital twin excels at reproducing these behaviors, it does not inherently explain them. The situation is similar to that of the maps, as noted by Borges<sup>25</sup>; if we want to build the best possible twin of a system we have to replicate the system itself, which does not necessarily advance our understanding. Explanation requires understanding why processes and phenomena happen. Although a digital twin can show what is happening, it does not provide underlying reasons or insights unless specific analysis tools or models are integrated. For example, it may reproduce a failure in the system under analysis, but it does not automatically explain the root cause unless additional diagnostics or analytics are applied. Unfortunately, causes can be intertwined in a network of other networks. Accordingly, these objects are best described and understood as complex systems. Indeed, explanation often involves understanding cause-and-effect relationships, system interdependencies, or emergent properties that may not be obvious from reproducing real-time behavior.

Interest in digital twins has also exploded well beyond medicine due to increasing access to memory, computational power, and massive data gathering. Digital twins have also been applied to cities<sup>26</sup>, primarily to simulate the intricate infrastructural configurations and products<sup>27</sup>, by leveraging contemporary technologies such as data analytics, IoT-driven physical modeling<sup>28</sup>, machine learning, and AI. For cities<sup>29</sup>, it may be far more efficient to consider the emerging behavior arising from the intricate web of relationships, processes, and correlations that characterize the complex adaptive system<sup>30–32</sup> than to produce a mere copy of it.

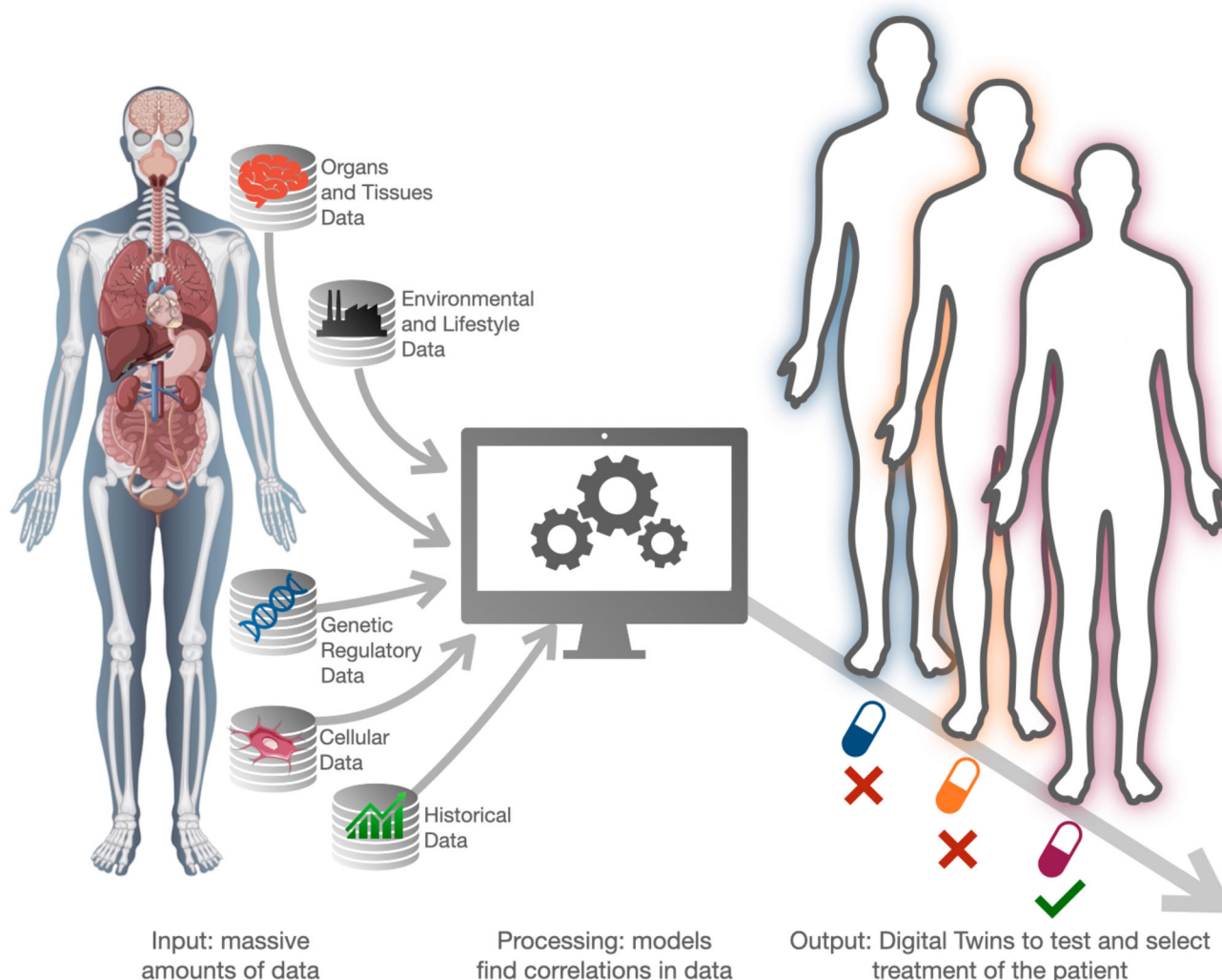
**Open challenges** We consider current methodological advances and challenges in these other fields to highlight the existing challenges in precision medicine. Despite promising opportunities to create a digital copy of every individual, there are some caveats to address to allow for personalized analysis and testing of individual-specific therapeutic strategies<sup>15,21,22</sup>.

On the one hand, if digital twins must be designed to be perfect replicas of individuals, then the amount of data required vastly exceeds our present, and even future, possibilities. The gigantic number of intervening functional

units, from biomolecules to cells, makes any analytical or computational approach impossible. Even in the ideal case that a perfectly functioning computational framework was technologically accessible, the nonlinear dynamics of interacting biological units lead to emergent phenomena that cannot be simply simulated or predicted, which is a hallmark feature of complex systems<sup>33,34</sup>. Because of this, recent advances in predictive biology are based on building models of increasing complexity to reproduce only the most salient characteristics of complex biological processes in engineered and natural populations<sup>35</sup>.

On the other hand, human patients have their own dynamical response to internal dysfunctions or differentiated coupling to the environment. This includes individual histories of host-microbiome and host-pathogen interactions, which might jeopardize any predictive model. Even more widely, the full individual exposome includes all past exposure to specific multi-scale environmental factors, such as diet and reactions to stressful biochemical or social conditions<sup>36,37</sup>. While the causal mechanisms in multiomic regulation can be partially reconstructed and accounted for, the full individual exposome is almost impossible to replicate or reproduce with a digital twin.

We have made great strides in capturing the exposome via the collection of new types of data from sources such as mobile devices<sup>38</sup> and social media<sup>12</sup>. However, even in the most ideal cases, unknown factors such as the level of disease progression and unmeasured lifestyle changes can lead to a broad set of distinct outcomes that make the design of digital twins very sensitive to the quantity and accuracy of input data. This technology may struggle to adapt and accurately predict these dynamic changes, which would lead to sub-optimal personalized treatment recommendations. These potential issues can dramatically hinder the purpose of digital twins, which might suggest that only methods based on advanced statistical data analysis, such as machine learning, are viable. This is not the case, however, because such methods provide predictive models that (i) do not easily generalize to situations and conditions for which they have not been trained, and (ii) might recommend clinically sub-optimal solutions when they retrieve multiple outcomes that they have ranked similarly (Fig. 1).



**Fig. 1 | Precision medicine standard approach for digital twins.** The framework relies on using large-scale heterogeneous data sources (pre-clinical, clinical, environmental, lifestyle, etc.). This massive database can be used in sophisticated computational models (such as deep learning), while relying solely on statistical data

analysis to construct a series of digitalized instances—the digital twins—of a patient, which can then be used to test one or more therapeutic strategies for clinical decision-making. Human body design by Freepck and osteocytes from Servier Medical Art (smart.servier.com).

Therefore, a more comprehensive approach based on methods that capture the essential features of complex interconnected and interdependent systems<sup>39,40</sup> at many scales is needed. This approach must (i) reduce the dimensionality of the problem of interest by identifying the key biological, clinical, and environmental variables needed for an adequate description on short time scales; (ii) characterize the conditions under which a complex adaptive system like the human body (or even a cell line) can be simulated by a digital twin in terms of separated components or sub-systems; and (iii) provide a transparent computational framework for testing actionable intervention strategies that are based on what-if scenarios and clinically relevant, model-informed, data-driven, and evidence-based questions.

In short, this calls for a more holistic and quantitative approach based on the complex adaptive nature of every patient rather than a mere replica of their salient aspects for statistical analysis.

### Multiscale modeling in health and disease: from genes to systems

To design effective digital twins, accounting for the multiscale nature of biological systems is of paramount importance. Recent progress in the study of complex systems, especially those with interconnected and interdependent structure, dynamics, and function, provide promising ground for figuring out and illustrating how diverse functional units and sub-systems interact at

different scales. Indeed, in addition to extracting multiscale molecular details from large omics datasets (e.g., transcriptomic, genomic, metabolomic, and microbiomic), we can now extract large-scale human behavior data of biomedical relevance from social media, mobile devices, and electronic health records, including new patient-stratification principles and unknown disease correlations<sup>9,11,12,38,41–43</sup>. Accordingly, the holistic integration and analysis of such multiscale data sources constitutes a novel opportunity to further improve personalization by including the exposome in the study of multilevel human complexity in disease<sup>42,44</sup>. This can be used to inform more accurate models for predictive purposes in biomedicine<sup>35,43,45–47</sup>.

### Processes at the intracellular scale

At the smallest scale, gene regulatory networks are systems of interacting genes and their regulatory elements within a cell that control the level of gene expression. In these networks, the nodes usually represent genes and the edges usually represent regulatory interactions between them. They describe the timing, spatial distribution, and intensity of gene expression, thereby orchestrating various cellular processes such as development, differentiation, and response to environmental stimuli<sup>48–51</sup>. A protein-protein interaction (PPI) network captures distinct types of interactions (e.g., physical contacts) between proteins in a cell. In PPI networks, nodes represent individual proteins, and edges encode interactions between them, which can

be transient, as in signal transduction, or more stable, as in protein-complex formation<sup>52,53</sup>. PPI networks provide insights into cellular processes, functional associations, and the modular organization of proteins, and analyzing the structure and dynamics of PPI networks helps uncover the underlying principles of cellular organization and function<sup>46,54–60</sup>. Metabolic networks<sup>43,61,62</sup> map out the biochemical reactions that occur within an organism, detailing how individual metabolites are synthesized, degraded, and interconverted<sup>63</sup>. These networks are either composed of nodes, representing metabolites, and edges, indicating the enzymatic reactions facilitating the transformation from one metabolite to another, or bipartite networks, where nodes are chemical species on one side and reactions on the other. In the latter representation, the web of metabolic interactions is intricately woven, while in the former, it is more straightforward. Beyond individual reactions, these networks highlight the interconnected nature of metabolic pathways, which reveal redundancies, feedback loops, and regulatory mechanisms that maintain cellular homeostasis.

Intracellular networks have time-evolving states that describe which genes are active, which proteins are present (or phosphorylated, oxidized, ubiquitinated, etc.), the concentrations of metabolites, and so on. State evolution is often studied using ordinary differential equation (ODE) models, which can be fit to match experimental state and kinetic data<sup>64</sup>. In many cases, the available data is insufficient to fully constrain the parameters of an ODE model. Also, it is often the case that the underlying biological dynamics is of a threshold nature<sup>65</sup>. In these cases, a discrete causal model, such as a Boolean network (or multistate automata network, more generally), may be appropriate<sup>66,67</sup>. In a Boolean network, the state of each node in the intracellular network is binarized: a gene is either active or inactive, and the active form of a protein is either above some unspecified threshold of abundance or below it. The binarized states change in time according to logical (Boolean) update functions; that is, each network node is an automaton<sup>68</sup>. The causal effect of various interventions (e.g., drugs) can be evaluated by manipulating the states of individual nodes and observing the resulting dynamics.

Because Boolean automata can be grouped to model variables with more than two states, the approach is widely applicable for modeling cellular components with various levels of activation, such as the proportion of cells that enter apoptosis in breast cancer cell lines<sup>69</sup>. Indeed, a common application of these models is in studying the effects of combinatorial drug interventions, particularly in the context of cancer<sup>69,70</sup>. To serve as a component of a digital twin, Boolean networks must reconcile their discrete time steps with physical time. This is often done by updating node states asynchronously according to tunable node transition rates, essentially treating the dynamics as a continuous Markov process<sup>71</sup>. This approach has been applied, for example, to suggest personalized drug therapies for prostate cancer patients using personalized Boolean network models<sup>72</sup>. One important advantage of Boolean or multistate automata networks is that they allow the precise characterization of polyadic relationships<sup>65,68</sup>. In other words, they are a type of higher-order network that can capture multivariate associations and interactions beyond the pairwise relations afforded by graphs in typical network science praxis<sup>73,74</sup>. In addition, this discrete dynamics approach has been used to infer important dynamical pathways in multilayer networks, which is another case of a higher-order network<sup>40</sup>. For instance, tying molecular factors (from multiomics, brain, and retinal imaging data) to clinical phenotype (from patient data) in multiple sclerosis<sup>47</sup>. This is an exciting avenue that allows complex regulatory dynamics to be studied on static multilayer networks obtained from heterogeneous data sources. Thereby, each node can integrate incoming signals differently, which goes well beyond the typical analysis via spreading or information dynamics on networks.

It is important to note that automata network models can typically be greatly simplified by reducing dynamically redundant interactions<sup>68</sup>, which are due to the ubiquity of canalized dynamics in biology<sup>75,76</sup>. This results in scalable causal models capable of uncovering actionable interventions, conditioned on different input assumptions, in a transparent manner<sup>65,68,74</sup>. Boolean networks are especially amenable to causal analysis because they

can be converted to simplified causal representations (according to Boolean minimization criteria)<sup>65,68,77,78</sup>. They stand in stark contrast to the black-box predictions of traditional machine learning methods and tallying the outputs of Monte Carlo simulations of large dynamical models (including non-simplified Boolean Networks).

Thus, automata network models—whose parameters can be inferred and validated from perturbation experiment, multiomics, and exposome data—are ideal components to consider for the top level of digital twins. They synthesize the large-scale underlying data into simplified, explainable, causal networks amenable to investigating actionable interventions. Indeed, these features show how this modeling approach directly responds to the needs of the digital twin approach identified in the introduction: dimensionality reduction, scalable modularity, and transparency.

### Processes at the whole-cell scale

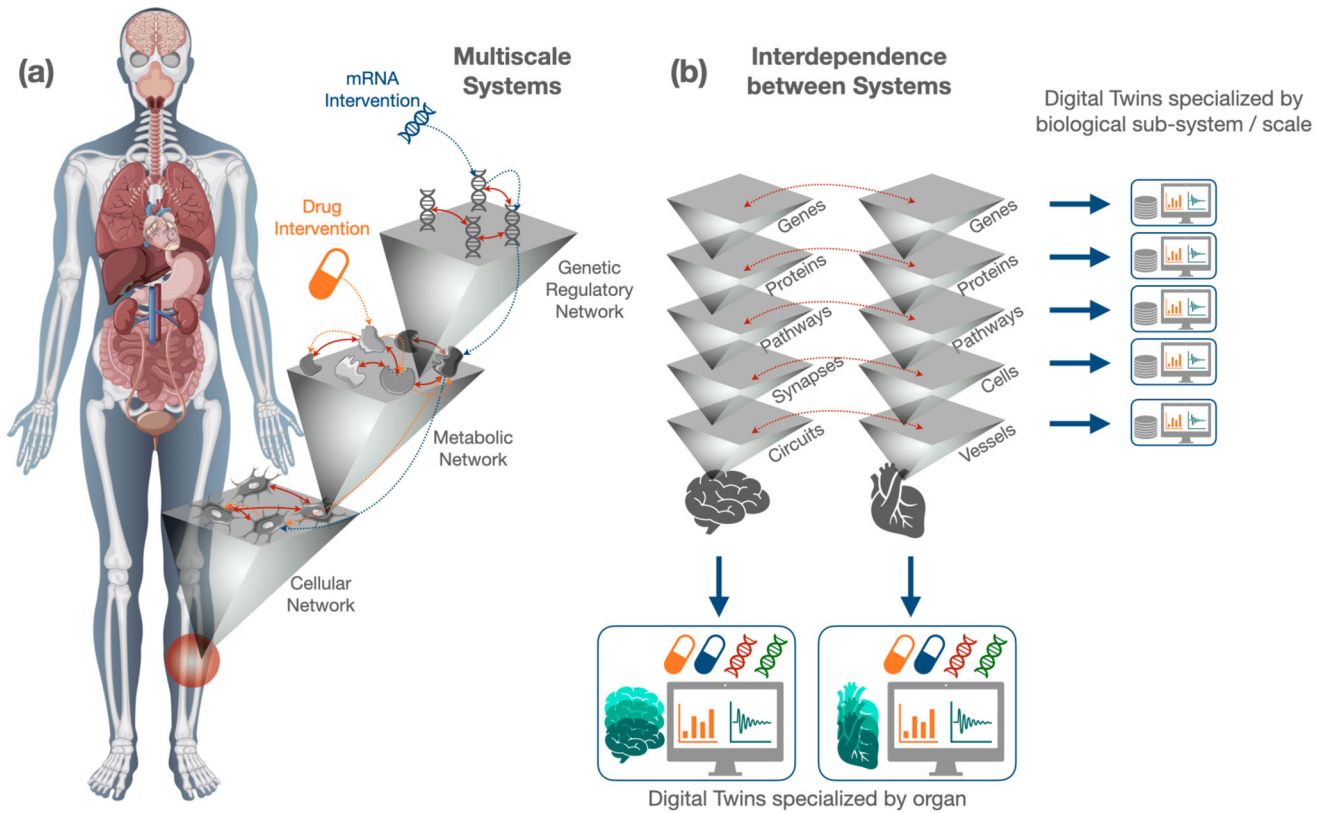
Whether discrete or continuous, the dynamics of intracellular networks can be coupled with each other and with physical processes to produce whole-cell models. These models attempt to describe the whole genome, proteome, and metabolome of a cell over the course of its life cycle in a fine-grained dynamical model<sup>79</sup>, as was first demonstrated in the human pathogen *Mycoplasma genitalium*<sup>80</sup>. More recent efforts have been focused on identifying minimal genomes<sup>81</sup> or modeling organisms with larger genomes, such as *E. Coli*<sup>82</sup>. Currently, the biomedical application of such detailed models is limited by the enormous effort required to construct them. Fortunately, to build a medically relevant digital twin, it is often the case that only specific processes need to be incorporated. Narrowing the focus of the model at the cellular level makes model construction and personalization more feasible, lowers computational barriers, and facilitates embedding these models into multicellular models, as in<sup>83</sup>.

An interesting focus arises from single cell data analysis. Even cells of the same type exhibit different system state and expression profiles in tissues, which are complex multi-agent systems made-up of multiple sub-populations of cells. They are spatially and temporally organized, able to communicate and interact with each other and orchestrate self-assembly and response to stimuli as a whole. This is fundamental in many biological contexts, such as early embryonic development and tumor etiology, where different cells are characterized by distinctive genetic mutations or expression profiles. These differences are regulated by cell-to-cell communication and underlie complex dynamic responses characterizing healthy and pathological tissue development<sup>84</sup>.

An example of how the interaction between cells can be modeled to describe emergent behaviors is the study of the interaction dynamics between immune and tumor cells in human cancer using agent-based models. By coupling a discrete agent-based model with a continuous partial-differential-equation-based model, these models capture essential components of the tumor microenvironment and are able to reproduce its main characteristics. Each tumor is characterized by a specific and unique tumor microenvironment, which emphasizes the need for specialized and personalized studies of each cancer scenario. Recently, a model of colon cancer has been proposed that can be informed with patient transcriptomic data<sup>85</sup>. It would be interesting to extend this model by informing it through methods that infer cellular communication<sup>86–89</sup>, which has the advantage of characterizing the tumor environment more specifically by defining the probability of an agent's action in response to received communication.

### Processes at the intercellular scale and beyond

At the tissue scale, systems that must be considered include neural, cardiovascular, and respiratory, but also their dysfunctions, such as cancer. Whether investigating the intricate networks within the human brain or the simpler wiring maps of organisms like *C. elegans* or *Drosophila Melanogaster*, the objective remains consistent: to elucidate the interconnections and organization of neurons and regions at the mesoscale. Several studies have focused on examining neural connections within an organism's brain, commonly referred to as the connectome<sup>90</sup>. Functional imaging techniques are utilized to explore the relationship among activities in specific brain



**Fig. 2 | Multiscale and network modeling for digital twins.** **a** Once the potential source of a dysfunction is identified, multiple biological systems might be involved across different spatial and temporal scales. Treatments based on, e.g., mRNA therapy or classical drugs, usually target biological units at a specific scale. However, treatment effects might propagate to other units at the same scale or across scales due to the presence of interactions and interdependencies among biomolecules and functional sub-systems, such as cells. **b** A multiscale illustration of the

interdependent sub-systems related to the function of distinct organs. Each scale can be simulated by a specialized digital twin, or multiscale integration can be simulated by a more complex, but still specialized, digital twin. The effects of distinct treatments can be analyzed on several distinct instances of the digital twin using a model-informed and data-driven search. Human body design by Freepik and osteocytes from Servier Medical Art (smart.servier.com).

areas<sup>91</sup>. The analytical and computational tools from network theory allow us to build maps of structural and functional connections, revealing characteristics of complex networks—such as small-world topology, highly connected hubs, and modularity—that manifest at both the whole-brain and cellular scales<sup>92–94</sup>. The human brain is an emblematic case study for the design of ambitious computational models toward developing digital twins. Nevertheless, despite the aforementioned significant advancements, the explicit goal of building a realistic computer simulation of the brain within a few years has not met expectations<sup>95</sup>.

Overall, sub-systems are part of a broader complex, adaptive, interdependent system of systems which are organized in hierarchies of increasing complexity with modular organization<sup>96</sup>. This is a fact well recognized for at least half a century and summarized in the Jacob’s statement that “every object that biology studies is a system of systems”<sup>97</sup>. Sub-systems exchange information (e.g., in terms of electrical, chemical, and electrochemical signals) to regulate each other and operate out of equilibrium<sup>98–100</sup>. Consequently, considering sub-systems in isolation from one another provides an incomplete representation of each sub-system and leads to inaccurate models and predictions of biological processes. A partial solution to this problem comes from the statistical physics of multilayer systems, which allows each scale to be described by a *level* of organization, and each level to be characterized by multiple context *layers*<sup>40,101</sup>. Levels can be interdependent<sup>39,102</sup> while also being characterized by different contexts. In the case of biological systems<sup>103</sup>, this is reflected in the distinct types of interactions among the same set of biomolecules or the distinct channels available for cell-cell communication, and also in the interdependence between distinct systems such as the cardiovascular and nervous systems

(Fig. 2). This web of interconnections and interdependencies involving diverse and heterogeneous functional biological units across scales plays a pivotal role in human health, and it is plausible to associate their dysfunction with disease states<sup>104,105</sup>.

While gathering, ingesting, analyzing, and accessing in real-time all the necessary multiscale data for multilayer network models will remain a challenge for the foreseeable future, data science already provides scalable methods for federating and visualizing such heterogeneous data<sup>106</sup>. One exciting approach is the construction of knowledge graphs to represent biomedical concepts and relationships extracted from human-annotated databases or automated pipelines<sup>107</sup>. These have been used, for instance, to design epilepsy patient self-management tools that link disease factors from molecular and pharmacological databases, clinical practice and epidemiology from electronic health records, and even exposome factors extracted from social media<sup>108</sup>.

Successful use cases of multilayer networks also include applications from personalized multiomics, where information from mRNA, miRNA, and DNA methylation has been integrated into a multilayer structure to unravel the peculiar microscopic and mesoscopic organization in Chronic Obstructive Pulmonary Disease<sup>109,110</sup>, rare diseases such as Congenital Myasthenic Syndromes<sup>43</sup> and Medulloblastoma<sup>111</sup>. Multilayer networks provided fertile ground for characterizing diseases while accounting for multidimensional factors such as molecular interactions and symptoms<sup>112</sup>, for building integrated models of signaling and regulatory networks<sup>113</sup>, and for analyzing functional organization in the human brain for subjects affected by psychiatric disorders such as Schizophrenia<sup>114</sup> and neurodegenerative diseases such as Alzheimer’s<sup>115–117</sup>.

## Challenges in multiscale modeling

### Limitations in data-driven approaches to digital twins in precision medicine

Multiscale modeling of biological systems presents formidable challenges, primarily due to the intricate and redundant networks of interactions and interdependent processes taking place. They unfold across different scales, from molecular (microscopic) to organismic (macroscopic) levels. These systems are characterized by dynamic processes that operate far from equilibrium; they exchange various types of signals—for example, chemical, electrochemical, and more – thereby creating a complex ecosystem of interlinked dynamical processes<sup>35</sup>. Such complexity poses significant difficulties in developing models that are both consequential and coherent, while avoiding extremes like reductionism, which assumes that sufficient computational power can simulate an entire organism, or oversimplification, which relies excessively on abundant data to sidestep the need for intricacy.

Moreover, biological systems are inherently adaptive, adjusting dynamically to environmental changes<sup>118</sup>. This adaptiveness is crucial for accurately simulating the impact of external factors such as therapeutic interventions or changes in environmental conditions, like pollution<sup>119</sup> or alterations in food sources<sup>120,121</sup>. Responses to these changes start at the cellular level, influencing gene expression, post-translational modifications to proteins, metabolic fluxes, and so on, leading to sometimes irreversible epigenetic changes<sup>43,122,123</sup>. These responses ultimately extend to organs, higher biological systems, and overall phenotype response through complex signaling pathways. Such adaptive complexity, represented in now available genomic, epigenomic, and other multiomic data<sup>123</sup>, cannot be accounted for by statistical methods only. It must be integrated into causal models to accurately reflect the biological response to external stimuli within the spatial and temporal scale of interest and to generate causal hypotheses<sup>124</sup>.

In the broader context of precision medicine, integrating digital twins that reflect these multiscale, multiomics and adaptive features poses even more challenges. The models often employed are predominantly phenomenological, focusing more on observed phenomena rather than the underlying mechanisms. This approach results in a significant gap in mechanistic understanding, which is essential for bridging various biological scales effectively. Cities face similar multiscale integration challenges<sup>29</sup> and require a similar framework to address the complex interplay of different components within the living system, which potentially guides the development of more effective biomedical models. Accordingly, a crucial preliminary step is establishing reliable and widely accepted standards for defining and measuring relevant quantities, along with their associated errors. The medical field often struggles with standardization due to the significant impact of individual physician expertise on diagnosis and treatment, which explains the challenges in creating uniform procedures. Digital twins can help overcome these obstacles by providing a shared framework of quantifiable parameters and standardized measurement practices. However, even with carefully defined protocols, some key properties may remain elusive. In such cases, network reconstruction techniques can assist by inferring the structure of biological networks from incomplete or noisy data. Because medical networks are rarely fully observable, reconstructing missing links or predicting unobserved connections becomes essential for a comprehensive system description<sup>125–129</sup>.

### Toward hypothesis-driven generative models

By leveraging properties of network topology, such as degree distributions and community structures within statistical ensemble frameworks, we can enhance the accuracy of our network predictions, ultimately improving our mechanistic understanding across biological scales. Note, however, that standalone data-driven models heavily rely on the completeness and quality of data, while missing, conflicting, or poor-quality data can lead to inaccurate predictions. In the context of a digital twin, hypothesis-driven models could leverage prior knowledge or established theories to fill gaps, even in the absence of complete data, thus providing a clear advantage. Genetic-algorithm inference of qualitative models provides a computationally efficient way to explore multiple cellular contexts and to ultimately reconcile

apparently contradictory experimental measurements without being overly sensitive to small changes in experimental design<sup>130–132</sup>.

We use the term “hypothesis-driven generative models” to refer to mechanistic simulations that produce behavior based on explicitly defined biological hypotheses and mechanisms. These models simulate biological processes by incorporating known or hypothesized interactions and pathways, which enables the generation of predictive and explanatory behaviors grounded in mechanistic understanding of the intervening physical, chemical, and biological processes. This approach differs from data-driven generative AI models, such as those based on variational autoencoders or generative adversarial networks, which learn patterns from data without necessarily incorporating underlying biological mechanisms.

By critically analyzing these challenges through the lens of complexity science, we can better understand and possibly overcome the hurdles in creating cohesive and predictive multiscale models that are crucial for the future of biomedical research and therapeutic development. In the case of interconnected systems at a given scale, we can introduce a suitable object, named multilayer adjacency tensor  $M_{j\beta}^{i\alpha}(t)$ , to operationally encode all the interactions at time  $t$  between a biological unit  $i$  (e.g., a single protein or a protein complex) in a layer  $\alpha$  (e.g., a class of biological processes or a pathway) and another biological unit  $j$  (e.g., another protein or a metabolite). The framework is so general that it can also include potential cross-layer structural interactions. In fact, due to the high number of interacting units (such as biomolecules, cells, etc.), biological modeling often assumes such deterministic processes, for example, that reactions occur at constant rates, compartmental interactions are fully-mixed, or mean-field approximations apply. Therefore, even at some good level of approximation, the dynamics of some quantity  $\mathbf{x}(t)$  of interest, for example, the concentration of metabolites or the population of some species such as cancer cells or bacteria, might be described by multilayer differential equations<sup>40,133</sup> like

$$\frac{\partial x_{j\beta}(t)}{\partial t} = f_{j\beta}(x_{j\beta}, t) + \sum_i \sum_{\alpha} g_{j\beta} \left[ M_{j\beta}^{i\alpha}(t), x_{i\alpha}(t), x_{j\beta}(t), t \right], \quad (1)$$

where  $f_{j\beta}(\cdot)$  is a function only of the variable  $x_{j\beta}(t)$  corresponding to a specific unit  $j$  in a specific context or layer  $\beta$ , and  $g_{j\beta}(\cdot)$  is a function that accounts for the interactions between pairs of units, that is, for the effects due to the intervening networks.

It is remarkable how such a simplified deterministic framework can model different phenomena of medical interest. Examples include responses to clinical treatment, cell activity stimulation, or even protein production, which are triggered by specific factors, such as basic chemical reactions, pH levels, drug concentration, or specific mRNA targets<sup>134,135</sup>. In light of these simple arguments, it might be tempting to rely only on such deterministic approaches—based on sets of differential equations, such as Eq. (1), or on agent-based modeling – to predict the behavior of a therapeutic intervention. After all, if we have systematic cause-effect relations linking interventions to biological and clinical outcomes, it would be enough to calibrate our models on the specific features of a patient to determine their response to treatments and potentially cure a disease.

However, in complex and variable environments such as a living organism, adaptiveness, randomness, and biological noise might affect the model outcomes. Still, adaptiveness can be reflected by such simplified models. If we indicate with  $u_{j\beta}(t)$  some external input signal or control applied to a biological system, and with  $\Theta$  the set of parameters that dynamically change based on the system’s states or external inputs, then a more general model at a given scale could be formalized as

$$\begin{aligned} \frac{\partial x_{j\beta}(t)}{\partial t} &= f_{j\beta}(x_{j\beta}, t) + \sum_i \sum_{\alpha} g_{j\beta} \left[ M_{j\beta}^{i\alpha}(t), x_{i\alpha}(t), x_{j\beta}(t), \Theta, u_{j\beta}(t), t \right] \\ \frac{\partial M_{j\beta}^{i\alpha}(t)}{\partial t} &= \ell(M_{j\beta}^{i\alpha}(t), x_{i\alpha}(t), x_{j\beta}(t), \Theta, u_{j\beta}(t), t) \\ \frac{\partial \Theta(t)}{\partial t} &= h(x_{i\alpha}(t), x_{j\beta}(t), \Theta, u_{j\beta}(t), t), \end{aligned} \quad (2)$$

which is much more complicated than Eq. (1), but it can be managed from a computational point of view. The integration of randomness and biological noise, which remains a challenge in mechanistic models, is further discussed in the following section.

### Limitations in multiscale deterministic modeling

Noise may be inherent to one or more aspects of the involved systems, for example, in the form of biochemical and electrochemical variability. Noise may also be linked to specific mechanisms altered by internal or external perturbations, such as virus-host interactions, environmental changes, and so on. Accordingly, when to include the effects of noise depends on the scale and impact of the biological process being modeled. For instance, including DNA replication errors for the analysis of short-term effects of a therapeutic drug might not add relevant biological or clinical insights, but would add undesirable complexity to the model. Another emblematic case is the use of discretized structures, such as networks, to model processes that are manifestly continuous (e.g., in space). Under such conditions, using complex networks introduces a level of sophistication that is not necessary to gain insights about a biological process.

Noise sources introduce an additional level of stochasticity that cannot be easily taken into account by statistical models, even the most sophisticated ones based on machine learning. Nevertheless, what is usually assumed to be a bug might be a feature: for other complex systems in nature, stochasticity is indeed structured and can lead to self-organized behaviors and processes<sup>136–138</sup>. The theory of nonlinear dynamical systems and the statistical physics of complex networks provide suitable theoretical and computational frameworks to model such complex biological phenomena<sup>100</sup>. They should be considered essential ingredients in designing reliable digital twins, either specialized or not, for any living organism.

The most important obstacle to describing realistic biological systems is how to incorporate multiple dynamic processes across the multiple intervening scales. The difficulty is primarily due to the diverse nature of the laws governing these processes at each scale. One significant technical challenge is effectively bridging these scales. This involves not just scaling up or down the processes, but also ensuring that interactions between scales are accurately captured. This might involve developing intermediate models or using scale-bridging techniques like homogenization or coarse-graining, which themselves can introduce approximation errors or require simplifications that might affect model accuracy. While some models are based on fundamental laws—such as reaction-diffusion processes for chemical networks—other models are genuinely phenomenological. Reconciling the dynamics of such different natures is challenging, since the latter class of models might not be suitable to capture novel phenomenology. This problem can be solved only partially by developing more fundamental models because biological processes are characterized by emergent phenomena that cannot be directly deduced even with full knowledge of their units and interactions<sup>30–33</sup>. To overcome this problem, we must simultaneously account for the evolution of the system according to dynamics similar to those in Eq. (2) and the fact that the underlying mechanisms can change while satisfying the constraints imposed by physics and chemistry, which requires meta-dynamical models<sup>139</sup>.

Additionally, multiscale models often require extensive parameterization, which can be difficult when experimental data are scarce at certain scales. Validating these models across all scales can be exceptionally challenging. This happens especially when direct observations or experiments at certain scales are not feasible or when they provide, at best, indirect measurements (such as correlations) about the phenomenon of interest that requires an adequate inferential framework<sup>127</sup>.

Furthermore, models should be able to propagate perturbations from one scale to another to realistically mimic the behavior of a living organism. As previously discussed, the possibility that a perturbation at the lowest scale (e.g., a random mutation or an mRNA intervention) can alter biological processes at larger scales is a mandatory feature for any reliable design of a digital twin.

### Discussion and outlook

Innovative approaches for model integration within digital twins have huge transformative potential in precision medicine by enabling a synergy between generative modeling, advanced AI and machine learning techniques, and traditional biomedical insights. The fusion of these techniques, rather than the choice of a specific one, is expected to facilitate the development of new frameworks for multiscale modeling. This is pivotal in capturing the intricate dynamics of pathogenesis in humans. Through these frameworks, the overarching goal is to resolve the challenges we have identified and significantly enhance the accuracy and clinical relevance of digital twins beyond inductive modeling via advanced statistics.

### From black to transparent boxes

On the one hand, the integration of mechanistic models into digital twins also addresses the challenges of parameter indeterminacy and overfitting, which are prevalent in systems characterized by vast parameter spaces. Digital twins constrain these spaces by, for example, the coarse-grained dynamics afforded by multiscale automata network models that synthesize large-scale data about biological mechanisms. In so doing, digital twins not only gain in robustness and explainability but also offer a more reliable foundation for the simulation of therapeutic outcomes, thereby increasing their utility in clinical practice. This activity can bring about a mechanistic clinical decision support system, that is, a type of decision support tool in healthcare that uses mechanistic models to assist clinicians in making medical decisions. Mechanistic models are based on an understanding of the underlying biological, physiological, or physical mechanisms that govern the behavior of systems in the human body. These models use known principles, from biochemistry, physics, population dynamics, and systems theory to predict outcomes or provide interpretations for clinical data. Mechanistic models are built on a theoretical understanding of how biological systems function. They often use mathematical equations to represent physiological processes such as blood flow, metabolism, or drug dynamics. Of course, these methods need to be tailored to each patient, and for this purpose, the methods of machine learning and analysis of each patient's past behavior are of the utmost importance<sup>40</sup>.

On the other hand, it is also worth discussing what is missing in current technologies and techniques developed for the same aim. For instance, a critical advantage of digital twins over state-of-the-art non-computational models, such as organoids<sup>141</sup>, is their ability to simulate complex, interdependent processes across multiple biological scales effectively. They also provide explanatory and causal understanding and control at relatively small costs. Indeed, we should consider the full spectrum of digital twins. They are not restricted to whole organisms, but can also be used to model cell lines, sub-systems and organs. This makes them an exciting alternative or complement to organoids. It is certainly feasible to compare, for instance, digital twins of cancer cell lines<sup>142,143</sup> with organoids synthesized for the same cell lines<sup>144</sup>. As discussed next, in this type of scenario, digital twins provide various advantages that need to be considered.

### The advantages of mechanistic digital twins

Organoids can be engineered, using the power of modern synthetic biology<sup>145</sup>, to recapitulate features of the function and response of complex biological mechanisms of the corresponding *in vivo* target, but they have important limitations. For one, reproducibility is a major bottleneck<sup>146</sup>. However, digital twins can excel at this, especially if built under an open-source framework. Additionally, organoids do not yet capture the entire physiological repertoire of cell types, or even the behavior that is relevant for a particular disease. This means, for instance, that the response to drugs or other interventions needs to be studied for organoids *per se*, separately from the *in vivo* target. They also have a relatively limited range of heterogeneity in responses, but a broader range is needed to develop truly personalized digital twins<sup>146</sup>. Finally, while organoids are more direct analogs of biomolecular mechanisms, they cannot incorporate simultaneously the multiple scales and historical information about patients, including the microbiome

and exposome, which are major factors in complex conditions such as cancer, depression, and many chronic diseases.

This is where the comprehensive multiscale network- and data-driven digital twin approach is particularly crucial. Many complex diseases unfold across various multiomic sub-systems and exposome histories. Modular computational architectures that can synthesize and integrate multiple subsystems as separate network layers or agent-based models are well within the realm of possibility. They might require a robust non-specialized digital twin, effectively integrating different specialized ones, to accommodate complex interactions and interdependency of biological and exposome processes. While non-specialized digital twins may not allow individual patient precision, the approach could still increase precision for specific cohorts within the whole population. For diseases with more circumscribed features, however, specialized digital twins might offer precise intervention strategies and outcome predictions that could perform as well or better than those based on organoids.

Another remarkable advantage of digital twins is that they allow a scenario-based modeling approach for actionable interventions—akin to strategies routinely used in epidemic modeling for policy decision-making<sup>147</sup>—that enhances their applicability and safety in clinical settings. This method avoids the standard pitfalls of an oracle-like predictive model by allowing for exploration via direct simulation of multiple clinical scenarios, thereby providing a robust tool for decision support in personalized medicine. It requires the crucial integration of massive data sets about disease or treatment progressions. This provides reliable statistical samples that can be stratified to approximate the characterizing features of a patient, and to validate the output of models. Therefore, the expected model-informed and data-driven output would not be a unique therapeutic strategy or an intervention, but a whole spectrum of alternatives with the advantages and disadvantages of each plausible strategy outlined to inform human decision-making.

### Explainability and falsifiability via complex systems science

Digital twins have the potential to revolutionize a wide range of clinical applications, including risk stratification, diagnostics, monitoring, prevention, prognostics, and treatment selection. By moving beyond traditional statistical methods that seek latent space classifications, our mechanistic modeling approach builds on the underlying biological processes, offering deeper insights into disease mechanisms. This not only aids in identifying new therapeutic targets but also enables the safe comparison of multiple therapies through simulation. Moreover, digital twins can complement in vitro and in vivo experimentation by prioritizing targets and predicting efficacy before laboratory testing. Trust in model predictions is enhanced through transparency and explainability, as mechanistic models provide both confidence levels and detailed explanatory mechanisms. Fully addressing all these aspects is critical, and the community has already started to do so<sup>10,148</sup>. While acknowledging their importance as critical areas for future research and adoption of the methodology, our complementary focus here is to show that the complex systems methodology can help integrate multiscale, multiomics, and exposome data to design hypothesis-driven digital twins based on data-driven generative models.

Received: 10 May 2024; Accepted: 16 December 2024;

Published online: 17 January 2025

### References

- Mirnezami, R., Nicholson, J. & Darzi, A. Preparing for precision medicine. *N. Engl. J. Med.* **366**, 489–491 (2012).
- Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
- König, I. R., Fuchs, O., Hansen, G., von Mutius, E. & Kopp, M. V. What is precision medicine? *Eur. Respiratory J.* **50**, 1700391 (2017).
- Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
- Lu, C. Y., Terry, V. & Thomas, D. M. Precision medicine: affording the successes of science. *NPJ Precis. Oncol.* **7**, 3 (2023).
- Wong, E. et al. The singapore national precision medicine strategy. *Nat. Genet.* **55**, 178–186 (2023).
- Ashley, E. A. Towards precision medicine. *Nat. Rev. Genet.* **17**, 507–522 (2016).
- Chen, R. & Snyder, M. Promise of personalized omics to precision medicine. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **5**, 73–82 (2013).
- Hoffman, J. M., Flynn, A. J., Juskewitch, J. E. & Freimuth, R. R. Biomedical data science and informatics challenges to implementing pharmacogenomics with electronic health records. *Annu. Rev. Biomed. Data Sci.* **3**, 289–314 (2020).
- Venkatesh, K. P., Brito, G. & Kamel Boulos, M. N. Health digital twins in life science and health care innovation. *Annu. Rev. Pharmacol. Toxicol.* **64**, 159–170 (2024).
- Jensen, P. B., Jensen, L. J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* **13**, 395–405 (2012).
- Correia, R. B., Wood, I. B., Bollen, J. & Rocha, L. M. Mining social media data for biomedical signals and health-related behavior. *Annu. Rev. Biomed. Data Sci.* **3**, 433–458 (2020).
- Hucka, M. et al. The systems biology markup language (sbml): A medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
- Hunter, P. J. & Borg, T. K. Integration from proteins to organs: the physiome project. *Nat. Rev. Mol. Cell Biol.* **4**, 237–243 (2003).
- Björnsson, B. et al. Digital twins to personalize medicine. *Genome Med.* **12**, 1–4 (2020).
- Erol, T., Mendi, A. F. & Doğan, D. The digital twin revolution in healthcare. In *2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT)*, 1–7 (IEEE, 2020).
- Eddy, D. M. & Schlessinger, L. Validation of the archimedes diabetes model. *Diab. Care* **26**, 3102–3110 (2003).
- Corral-Acero, J. et al. The ‘digital twin’ to enable the vision of precision cardiology. *Eur. Heart J.* **41**, 4556–4564 (2020).
- Coorey, G., Figtree, G. A., Fletcher, D. F. & Redfern, J. The health digital twin: advancing precision cardiovascular medicine. *Nat. Rev. Cardiol.* **18**, 803–804 (2021).
- Laubenbacher, R., Sluka, J. P. & Glazier, J. A. Using digital twins in viral infection. *Science* **371**, 1105–1106 (2021).
- Laubenbacher, R., Mehrad, B., Shmulevich, I. & Trayanova, N. Digital twins in medicine. *Nat. Comput. Sci.* **4**, 184–191 (2024).
- Kamel Boulos, M. N. & Zhang, P. Digital twins: From personalised medicine to precision public health. *J. Personalized Med.* **11**, 745 (2021).
- Milanesi, L., Romano, P., Castellani, G., Remondini, D. & Liò, P. Trends in modeling biomedical complex systems. *BMC Bioinform.* **10**, 1–13 (2009).
- Matheny, M. E., Whicher, D. & Thadaneys Israni, S. Artificial intelligence in health care: A report from the national academy of medicine. *JAMA* **323**, 509 (2020).
- Borges, J. L. A universal history of infamy. 1935. *Trans. Norman Thomas di Giovanni. New York: Penguin* (1975).
- Batty, M. *Inventing future cities* (The MIT Press, Boston, USA, 2018).
- Grieves, M. W. Product lifecycle management: the new paradigm for enterprises. *Int. J. Prod. Dev.* **2**, 71–84 (2005).
- Niederer, S. A., Sacks, M. S., Girolami, M. & Willcox, K. Scaling digital twins from the artisanal to the industrial. *Nat. Comput. Sci.* **1**, 313–320 (2021).
- Caldarelli, G. et al. The role of complexity for digital twins of cities. *Nat. Comput. Sci.* **3**, 374–381 (2023).

30. Gell-Mann, M. Complex adaptive systems. In *Santa Fe Institute Studies In The Sciences of Complexity-proceedings Volume-*, Vol. 19, 17–17 (Addison-Wesley Publishing Co, 1994).
31. Holland, J. H. Complex adaptive systems and spontaneous emergence. In *Complexity and Industrial Clusters: Dynamics and Models in Theory and Practice*, 25–34 (Springer, 2002).
32. Holland, J. H. Studying complex adaptive systems. *J. Syst. Sci. Complex.* **19**, 1–8 (2006).
33. Artime, O. & De Domenico, M. From the origin of life to pandemics: emergent phenomena in complex systems. *Philosophical Transactions of the Royal Society A*. **380**, <https://doi.org/10.1098/rsta.2020.0410> (2022).
34. Rosen, R. On complex systems. *Eur. J. Oper. Res.* **30**, 129–134 (1987).
35. Lopatkin, A. J. & Collins, J. J. Predictive biology: Modelling, understanding and harnessing microbial complexity. *Nat. Rev. Microbiol.* **18**, 507–520 (2020).
36. Vermeulen, R., Schyamski, E. L., Barabási, A.-L. & Miller, G. W. The exposome and health: Where chemistry meets biology. *Science* **367**, 392–396 (2020).
37. Vineis, P. et al. What is new in the exposome? *Environ. Int.* **143**, 105887 (2020).
38. Steinhubl, S. R., Muse, E. D. & Topol, E. J. The emerging field of mobile health. *Sci. Transl. Med.* **7**, 283rv3–283rv3 (2015).
39. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).
40. De Domenico, M. More is different in real-world multilayer networks. *Nat. Phys.* **19**, 1247–1262 (2023).
41. Sánchez-Valle, J. et al. Prevalence and differences in the co-administration of drugs known to interact: an analysis of three distinct and large populations. *BMC Med.* **22**, 166 (2024).
42. Dolley, S. Big data's role in precision public health. *Front. Public Health* **6**, 297813 (2018).
43. Núñez-Carpintero, I. et al. Rare disease research workflow using multilayer networks elucidates the molecular determinants of severity in congenital myasthenic syndromes. *Nat. Commun.* **15**, 1227 (2024).
44. Pescosolido, B. A. et al. The social symbiome framework1: Linking genes-to-global cultures in public health using network science. In *Handbook of applied system science*, 25–48 (Routledge, 2016).
45. Walpole, J., Papin, J. A. & Peirce, S. M. Multiscale computational models of complex biological systems. *Annu. Rev. Biomed. Eng.* **15**, 137–154 (2013).
46. Correia, R. B. et al. The conserved transcriptional program of metazoan male germ cells uncovers ancient origins of human infertility. *eLife* In Press. Biorxiv: 2022.03.02.482557 (2024).
47. Kennedy, K. E. et al. Multiscale networks in multiple sclerosis. *PLOS Comput. Biol.* **20**, e1010980 (2024).
48. Thattai, M. & Van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci.* **98**, 8614–8619 (2001).
49. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci.* **102**, 4936–4942 (2005).
50. Davidson, E. & Levin, M. Gene regulatory networks. *Proc. Natl. Acad. Sci.* **102**, 4935–4935 (2005).
51. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770–780 (2008).
52. Vazquez, A., Flammini, A., Maritan, A. & Vespignani, A. Global protein function prediction from protein-protein interaction networks. *Nat. Biotechnol.* **21**, 697–700 (2003).
53. Szklarczyk, D. et al. String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
54. Han, J.-D. J. et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93 (2004).
55. Han, J.-D. J., Dupuy, D., Bertin, N., Cusick, M. E. & Vidal, M. Effect of sampling on topology predictions of protein–protein interaction networks. *Nat. Biotechnol.* **23**, 839–844 (2005).
56. Stelzl, U. et al. A human protein–protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957–968 (2005).
57. Deeds, E. J., Ashenberg, O. & Shakhnovich, E. I. A simple physical model for scaling in protein–protein interaction networks. *Proc. Natl. Acad. Sci.* **103**, 311–316 (2006).
58. Stelzl, U. & Wanker, E. E. The value of high quality protein–protein interaction networks for systems biology. *Curr. Opin. Chem. Biol.* **10**, 551–558 (2006).
59. Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T. & Müller, T. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. *Bioinformatics* **24**, i223–i231 (2008).
60. Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein–protein interaction networks. *Nat. methods* **9**, 471–472 (2012).
61. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
62. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proc. Natl. Acad. Sci.* **110**, 2460–2465 (2013).
63. Guimera, R. & Nunes Amaral, L. A. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
64. Mackey, M. C., Santillán, M., Tyrán-Kamińska, M. & Zeron, E. S. Simple Mathematical Models of Gene Regulatory Dynamics. Lecture Notes on Mathematical Modelling in the Life Sciences (Springer International Publishing, Cham, 2016).
65. Gates, A. J., Brattig Correia, R., Wang, X. & Rocha, L. M. The effective graph reveals redundancy, canalization, and control pathways in biochemical regulation and signaling. *Proc. Natl. Acad. Sci.* **118**, e2022598118 (2021).
66. Schwab, J. D., Kühlwein, S. D., Ikononi, N., Kühl, M. & Kestler, H. A. Concepts in Boolean network modeling: What do they all mean? *Comput. Struct. Biotechnol. J.* **18**, 571–582 (2020).
67. Rozum, J. C., Campbell, C., Newby, E., Nasrollahi, F. S. F. & Albert, R. Boolean Networks as Predictive Models of Emergent Biological Behaviors. *Elements in the Structure and Dynamics of Complex Networks* (2024).
68. Marques-Pita, M. & Rocha, L. M. Canalization and control in automata networks: body segmentation in drosophila melanogaster. *PLoS one* **8**, e55946 (2013).
69. Gómez Tejeda Zañudo, J. et al. Cell line-specific network models of ER+ breast cancer identify potential pi3ka inhibitor resistance mechanisms and drug combinations. *Cancer Res.* **81**, 4603–4617 (2021).
70. Folkesson, E., Sakshaug, B. C., Hoel, A. D., Klinkenberg, G., & Flobak, Å. Synergistic effects of complex drug combinations in colorectal cancer cells predicted by logical modelling. *Front. Syst. Biol.* **3**, 1112831 (2023).
71. Calzone, L., Noël, V., Barillot, E., Kroemer, G. & Stoll, G. Modeling signaling pathways in biology with MaBoSS: From one single cell to a dynamic population of heterogeneous interacting cells. *Comput. Struct. Biotechnol. J.* **20**, 5661–5671 (2022).
72. Montagud, A. et al. Patient-specific Boolean models of signalling networks guide personalised treatments. *eLife* **11**, e72626 (2022).
73. Klamt, S., Haus, U.-U. & Theis, F. Hypergraphs and cellular networks. *PLoS Comput. Biol.* **5**, e1000385 (2009).
74. Rocha, L. M. On the feasibility of dynamical analysis of network models of biochemical regulation. *Bioinformatics* **38**, 3674–3675 (2022).
75. Manicka, S., Marques-Pita, M. & Rocha, L. M. Effective connectivity determines the critical dynamics of biochemical networks. *J. R. Soc. Interface* **19**, 20210659 (2022).

76. Costa, F. X., Rozum, J. C., Marcus, A. M. & Rocha, L. M. Effective connectivity and bias entropy improve prediction of dynamical regime in automata networks. *Entropy* **25**, 374 (2023).
77. Rozum, J. C., Deritei, D., Park, K. H., Gómez Tejeda Zañudo, J. & Albert, R. Pystablemotifs: Python library for attractor identification and control in Boolean networks. *Bioinformatics* **38**, 1465–1466 (2022).
78. Beneš, N. et al. AEON.py: Python library for attractor analysis in asynchronous Boolean networks. *Bioinformatics* **38**, 4978–4980 (2022).
79. Bhat, N. G. & Balaji, S. Whole-cell modeling and simulation: A brief survey. *N. Gener. Comput.* **38**, 259–281 (2020).
80. Karr, J. R. et al. A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012).
81. Rees-Garbutt, J. et al. Designing minimal genomes using whole-cell models. *Nat. Commun.* **11**, 836 (2020).
82. Sun, G., Ahn-Horst, T. A. & Covert, M. W. The E. coli whole-cell modeling project. *EcoSal* **9**, eESP-0001–2020 (2021).
83. Ponce-de-Leon, M. et al. PhysiBoSS 2.0: A sustainable integration of stochastic Boolean and agent-based modelling frameworks. *npj Syst. Biol. Appl.* **9**, 1–12 (2023).
84. Shalek, A. K. et al. Single-cell rna-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
85. Cesaro, G. et al. Mast: a hybrid multi-agent spatio-temporal model of tumor microenvironment informed using a data-driven approach. *Bioinforma. Adv.* **2**, vbac092 (2022).
86. Baruzzo, G., Cesaro, G. & Di Camillo, B. Identify, quantify and characterize cellular communication from single-cell rna sequencing data with scseqcomm. *Bioinformatics* **38**, 1920–1929 (2022).
87. Kumar, M. P. et al. Analysis of single-cell rna-seq identifies cell-cell communication associated with tumor characteristics. *Cell Rep.* **25**, 1458–1468 (2018).
88. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624 (2017).
89. Zhang, Q. et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* **179**, 829–845 (2019).
90. Sporns, O., Tononi, G. & Kötter, R. The Human Connectome: A Structural Description of the Human Brain. *PLoS Comput. Biol.* **1**, e42 (2005).
91. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
92. Sporns, O., Tononi, G. & Edelman, G. M. Theoretical neuroanatomy and the connectivity of the cerebral cortex. *Behav. Brain Res.* **135**, 69–74 (2002).
93. Sporns, O., Chialvo, D. R., Kaiser, M. & Hilgetag, C. C. Organization, development and function of complex brain networks. *Trends Cogn. Sci.* **8**, 418–425 (2004).
94. Chialvo, D. R. Emergent complex neural dynamics. *Nat. Phys.* **6**, 744–750 (2010).
95. Siva, N. What happened to the Human Brain Project? *Lancet* **402**, 1408–1409 (2023).
96. Wagner, G. P., Pavlicev, M. & Cheverud, J. M. The road to modularity. *Nat. Rev. Genet.* **8**, 921–931 (2007).
97. Jacob, F. *The logic of living systems: A history of heredity* (Allen Lane. First edition, 1974).
98. Haken, H. Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. *Rev. Mod. Phys.* **47**, 67 (1975).
99. Ritort, F. Nonequilibrium fluctuations in small systems: From physics to biology. *Adv. Chem. Phys.* **137**, 31 (2008).
100. Fang, X., Kruse, K., Lu, T. & Wang, J. Nonequilibrium physics in biology. *Rev. Mod. Phys.* **91**, 045004 (2019).
101. De Domenico, M. et al. Mathematical formulation of multilayer networks. *Phys. Rev. X* **3**, 041022 (2013).
102. Gao, J., Buldyrev, S. V., Stanley, H. E. & Havlin, S. Networks formed from interdependent networks. *Nat. Phys.* **8**, 40–48 (2012).
103. De Domenico, M. Multilayer network modeling of integrated biological systems. comment on “network science of biological systems at different scales: A review” by gosak et al. *Phys. Life Rev.* **24**, 149–152 (2018).
104. Greene, J. A. & Loscalzo, J. Putting the patient back together—social medicine, network medicine, and the limits of reductionism (2017).
105. Halu, A., De Domenico, M., Arenas, A. & Sharma, A. The multiplex network of human diseases. *NPJ Syst. Biol. Appl.* **5**, 15 (2019).
106. Manz, T. et al. Viv: multiscale visualization of high-resolution multiplexed bioimaging data on the web. *Nat. Methods* **19**, 515–516 (2022).
107. Nicholson, D. N. & Greene, C. S. Constructing knowledge graphs and their biomedical applications. *Comput. Struct. Biotechnol. J.* **18**, 1414–1428 (2020).
108. Correia, R. B. et al. myaura: Personalized health library for epilepsy management via knowledge graph sparsification and visualization. *arXiv preprint arXiv:2405.05229* (2024).
109. Olvera Ocaña, N. et al. Lung tissue multi-layer network in COPD. *Eur. Respir. J.* **60**, 2469 (2022).
110. Olvera, N. et al. Lung tissue multi-layer network analysis uncovers the molecular heterogeneity of COPD. *Am. J. Respir. Crit. Care Med.* <https://doi.org/10.1164/rccm.202303-0500OC> (2024).
111. Núñez-Carpintero, I., Petrizzelli, M., Zinovyev, A., Cirillo, D. & Valencia, A. The multilayer community structure of medulloblastoma. *iScience* **24**, <https://doi.org/10.1016/j.isci.2021.102365> (2021).
112. Verstraete, N. et al. Covmulnet19, integrating proteins, diseases, drugs, and symptoms: A network medicine approach to covid-19. *Netw. Syst. Med.* **3**, 130–141 (2020).
113. Martini, L. et al. Detecting and dissecting signaling crosstalk via the multilayer network integration of signaling and regulatory interactions. *Nucleic Acids Res.* **52**, e5–e5 (2024).
114. De Domenico, M., Sasai, S. & Arenas, A. Mapping multiplex hubs in human functional brain networks. *Front. Neurosci.* **10**, <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2016.00326> (2016).
115. Canal-Garcia, A. et al. Dynamic multilayer functional connectivity detects preclinical and clinical Alzheimer’s disease. *Cereb. Cortex* **34**, bhad542 (2024).
116. El-Sappagh, S., Alonso, J. M., Islam, S. M. R., Sultan, A. M. & Kwak, K. S. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer’s disease. *Sci. Rep.* **11**, 2660 (2021).
117. Guillon, J. et al. Disrupted core-periphery structure of multimodal brain networks in Alzheimer’s disease. *Netw. Neurosci. (Camb., Mass.)* **3**, 635–652 (2019).
118. Freeman, W. J., Kozma, R. & Werbos, P. J. Biocomplexity: adaptive behavior in complex stochastic dynamical systems. *Biosystems* **59**, 109–123 (2001).
119. Fuller, R. et al. Pollution and health: a progress update. *Lancet Planet. Health* **6**, e535–e547 (2022).
120. van der Laan, L. J., Bosker, T. & Peijnenburg, W. J. Deciphering potential implications of dietary microplastics for human health. *Nat. Rev. Gastroenterol. Hepatol.* **20**, 340–341 (2023).
121. Al Mamun, A., Prasetya, T. A. E., Dewi, I. R. & Ahmad, M. Microplastics in human food chains: Food becoming a threat to health safety. *Sci. Total Environ.* **858**, 159834 (2023).
122. Parreno, V. et al. Transient loss of polycomb components induces an epigenetic cancer fate. *Nature* **629**, 688–696 (2024).
123. Kreitmaier, P., Katsoula, G. & Zeggini, E. Insights from multi-omics integration in complex disease primary tissues. *Trends Genet.* **39**, 46–58 (2023).

124. Dugourd, A. et al. Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *Mol. Syst. Biol.* **17**, e9730 (2021).
125. Peixoto, T. P. Network reconstruction and community detection from dynamics. *Phys. Rev. Lett.* **123**, 128301 (2019).
126. Raimondo, S. & De Domenico, M. Measuring topological descriptors of complex networks under uncertainty. *Phys. Rev. E* **103**, 022311 (2021).
127. Peel, L., Peixoto, T. P. & De Domenico, M. Statistical inference links data and theory in network science. *Nat. Commun.* **13**, 6794 (2022).
128. Cimini, G. et al. The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**, 58–71 (2019).
129. Peixoto, T. P. Network reconstruction via the minimum description length principle. *arXiv:2405.01015* (2024).
130. Park, K. H., Rozum, J. C. & Albert, R. From Years to Hours: Accelerating Model Refinement (2023).
131. Flobakk, Å. Systems medicine: From modeling systems perturbations to predicting drug synergies. Doctoral thesis, NTNU (2016).
132. Li, R. et al. Inferring gene regulatory networks using transcriptional profiles as dynamical attractors. *PLOS Comput. Biol.* **19**, e1010991 (2023).
133. De Domenico, M., Granell, C., Porter, M. A. & Arenas, A. The physics of spreading processes in multilayer networks. *Nat. Phys.* **12**, 901–906 (2016).
134. Kowalski, P. S., Rudra, A., Miao, L. & Anderson, D. G. Delivering the messenger: Advances in technologies for therapeutic mma delivery. *Mol. Ther.* **27**, 710–728 (2019).
135. Rohner, E., Yang, R., Foo, K. S., Goedel, A. & Chien, K. R. Unlocking the promise of mna therapeutics. *Nat. Biotechnol.* **40**, 1586–1600 (2022).
136. Nicolis, G. & Prigogine, I. Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations. *Berichte der Bunsengesellschaft für physikalische Chemie* (1977).
137. Heylighen, F. et al. The science of self-organization and adaptivity. *Encycl. Life support Syst.* **5**, 253–280 (2001).
138. Camazine, S. et al. Self-organization in biological systems. In *Self-Organization in Biological Systems* (Princeton university press, 2020).
139. Bagley, R. J. et al. Modeling adaptive biological systems. *Biosystems* **23**, 113–137 (1989).
140. Auconi, P. et al. The validity of machine learning procedures in orthodontics: what is still missing? *J. Personalized Med.* **12**, 957 (2022).
141. Corrò, C., Novellasdemunt, L. & Li, V. S. A brief history of organoids. *Am. J. Physiol.-Cell Physiol.* **319**, C151–C165 (2020).
142. Mollica, L. et al. Digital twins: a new paradigm in oncology in the era of big data. *ESMO Real. World Data Digital Oncol.* **5**, 100056 (2024).
143. Filippo, M. D. et al. Single-cell digital twins for cancer preclinical investigation. *Metabolic Flux Analysis in Eukaryotic Cells: Methods and Protocols* 331–343 (2020).
144. Yan, H. H. et al. A comprehensive human gastric cancer organoid biobank captures tumor subtype heterogeneity and enables therapeutic screening. *Cell Stem Cell* **23**, 882–897 (2018).
145. Hofer, M. & Lutolf, M. P. Engineering organoids. *Nat. Rev. Mater.* **6**, 402–420 (2021).
146. Zhao, Z. et al. Organoids. *Nat. Rev. Methods Prim.* **2**, 94 (2022).
147. Hofman, J. M. et al. Integrating explanation and prediction in computational social science. *Nature* **595**, 181–188 (2021).
148. Venkatesh, K. P., Raza, M. M. & Kvedar, J. C. Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation. *NPJ Digital Med.* **5**, 150 (2022).

### Acknowledgements

M.D.D., V.dA. and F.Z. acknowledge partial financial support from the INFN grant “LINCOLN”. L.M.R. was partially funded by the NIH National Library of Medicine Program Grant 01LM011945-01 and the Fundação para a Ciência e a Tecnologia Grant 2022.09122.PTDC (<https://doi.org/10.54499/2022.09122.PTDC>).

### Author contributions

M.D.D. designed the research. M.D.D., L.A., G.C., V.dA., B.dC., L.M.R., J.R., R.S. and F.Z. wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Manlio De Domenico.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025