Taylor & Francis
Taylor & Francis Group

Expert Opinion on Drug Discovery

ISSN: 1746-0441 (Print) 1746-045X (Online) Journal homepage: www.tandfonline.com/journals/iedc20

AI-powered virtual cells, digital twins and agentic AI are reshaping multi-omics, spatial biology and drug discovery

Xixi Li , Nicolo Fabila , Xuan Liu , Christine R. Keenan , Robyn M. Brown , Phong B. H. Nguyen & Michael P. Menden

To cite this article: Xixi Li , Nicolo Fabila , Xuan Liu , Christine R. Keenan , Robyn M. Brown , Phong B. H. Nguyen & Michael P. Menden (17 Nov 2026): AI-powered virtual cells, digital twins and agentic AI are reshaping multi-omics, spatial biology and drug discovery, Expert Opinion on Drug Discovery, DOI: 10.1080/17460441.2026.2763719

To link to this article: https://doi.org/10.1080/17460441.2026.2763719

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 17 Nov 2026.

Submit your article to this journal

Article views: 125

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=iedc20

EXPERT OPINION ON DRUG DISCOVERY
https://doi.org/10.1080/17460441.2026.2763719

Taylor & Francis
by informa

REVIEW

OPEN ACCESS Check for updates

AI-powered virtual cells, digital twins and agentic AI are reshaping multi-omics, spatial biology and drug discovery

Xixi Li[a*], Nicolo Fabila[a*], Xuan Liu[a*], Christine R. Keenan[a], Robyn M. Brown[a], Phong B. H. Nguyen[a,b] and Michael P. Menden[a,b]

[a]Department of Biochemistry and Pharmacology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Australia; [b]Institute of Computational Biology, Computational Health Centre, Helmholtz Munich, Munich, Germany

ABSTRACT

Introduction: Predicting individual disease trajectories and treatment responses remains challenging in drug discovery. Molecular profiling provides static snapshots that miss cellular dynamics, spatial context and adaptive responses. The convergence of single-cell and spatial omics with generative AI is driving a paradigm shift toward AI-powered virtual cells that model cellular states, interactions and treatment responses.

Areas covered: This review introduces a nested framework of six capabilities for AI-powered virtual cells, spanning population-derived models, digital twins and agentic twins. Virtual cells generate synthetic cellular states and predict responses from population data. Digital twins link models to specific biological counterparts, enabling longitudinal updating and prospective prediction. Agentic twins add spatial context, biologically constrained interactions and actions, and environmental feedback, enabling personalized, dynamic and interactive representations of cellular ecosystems. This narrative review is based on literature acquired from PubMed, Web of Science, bioRxiv and arXiv searches up to October 2026.

Expert opinion: AI-powered virtual cells could support drug discovery as foundation models that transfer biological knowledge across tasks, including data-limited settings. Their value depends on faithful extrapolation beyond training data and prospective validation against biological observations. Coupling these models to repeatedly measured experimental systems could enable personalized prediction, iterative updating and prioritization of perturbations, targets and therapeutic strategies.

ARTICLE HISTORY
Received 9 September 2026
Accepted 10 November 2026

KEYWORDS
Agentic twin cell; digital twin cell; drug discovery; generative artificial intelligence; single-cell omics; spatial omics; virtual cell

1. Introduction

Over the past decade, single-cell technologies have scaled extraordinarily, and spatial omics now resolve molecular states within native tissue architecture. The CZ CELLxGENE Discover Census aggregates profiles from more than 93 million cells [1]; HISSTA comprises over 28 million *in situ* profiled cells across 16 human tissues [2]. Yet, our ability to observe cellular states has advanced faster than our ability to predict their behavior. Responses to perturbation, including proliferation, differentiation, migration, secretion and apoptosis, remain determined largely through empirical experimentation rather than computational prediction [3].

This predictive limitation directly affects drug discovery. Targets are often nominated from differential expression, disease-linked genetic variants and pathway enrichment. These associations identify biological relevance but do not predict whether modulating a target will produce the desired cellular response. This disconnect may contribute to translational attrition, with lack of efficacy accounting for an estimated 40–59% of clinical drug development failures [4].

Closing this gap requires more than increasingly detailed cellular atlases. Population-derived models can generalize poorly to unseen cell types, perturbations and donors [5] and cannot fully capture individual biology. Destructive single-cell assays provide terminal snapshots; dissociation also removes the tissue context through which perturbation effects propagate [6,7]. Predictive models must therefore faithfully extrapolate beyond reference data, using generative artificial intelligence (AI) to make high-fidelity predictions in unseen biological contexts, personalize predictions, track changes over time and resolve spatial interactions.

Three nested computational paradigms are emerging to address these challenges. Population-trained generative AI models establish virtual cells by generating synthetic cellular states and predicting responses [3,8–10]. Digital twin cells extend these models through persistent anchoring to specific biological counterparts and longitudinal updating [11–13]. Agentic twin cells further incorporate context perception, constrained actions and environmental feedback [14,15]. Unlike conventional agent-based models, they retain a persistent link to their biological

CONTACT Phong B. H. Nguyen phong.nguyen.1@unimelb.edu.au; Michael P. Menden michael.menden@unimelb.edu.au Department of Biochemistry and Pharmacology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, 30 Flemington Road, Parkville, Victoria 3052, Australia; Institute of Computational Biology, Computational Health Center, Helmholtz Munich, Ingolstädter Landstraße 1, Neuherberg, Munich 85764, Germany
*Equal contribution.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17460441.2026.2763719

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Article highlights

- Single-cell and spatial omics provide the scale and tissue context for AI models to learn transferable cellular representations and predict unmeasured responses.
- Our nested framework distinguishes population-derived virtual cells from digital twins through counterpart anchoring and longitudinal updating, and from agentic twins through spatial interaction, action and feedback.
- A digital twin of a cell maintains a persistent, testable link to a real biological counterpart, enabling longitudinal updating and validation against observed states.
- Spatial grounding connects cellular interactions with environmental feedback, providing a foundation for biologically meaningful agency.
- An agentic twin of a cell combines spatially grounded interaction and action with persistent counterpart anchoring and longitudinal recalibration, enabling models that both respond to and influence their biological environment.

counterpart. This review defines this progression from population-derived virtual cells to digital and agentic twins, establishes their requirements for falsifiability, and examines emerging applications in drug discovery. We searched PubMed, Web of Science, bioRxiv and arXiv from inception to 20 October 2026, selecting articles relevant to the proposed capabilities, enabling experimental platforms, validation requirements and drug-discovery applications. We synthesize representative evidence across these areas to identify research priorities.

2. A capability ladder for virtual cells

Virtual-cell models are difficult to compare when models designed for different tasks are treated as interchangeable. We therefore define six nested capabilities, each building on the previous one (Figure 1). C1 (representation) uses generative AI to learn cellular representations and generate synthetic states that capture biological variation [16,17]. C2 (prediction) builds on these representations to predict unmeasured cellular states and responses [8,9,18,19]. C3 (anchoring) establishes a persistent link to an identified biological counterpart through counterpart-specific observations [3]. C4 (simulation) evolves this anchored state over time and under alternative interventions, recalibrating predictions as new observations become available [13]. C5 (spatial interaction) places modeled cells within their environment and couples them through signaling or contact, allowing cells to influence one another [20–23]. Finally, C6 (closed-loop action and adaptation) generates biologically constrained actions that modify the environment and feed back to shape subsequent cellular states and responses [15,20,24].

A model is assigned to a class only when it demonstrates every capability that defines that class and all capabilities below it. Population-derived virtual cells are defined by C1–C2, digital twin cells additionally require C3 counterpart anchoring and C4 longitudinal simulation and recalibration, and agentic twin cells further incorporate C5 multicellular interaction and C6 closed-loop action and adaptation.

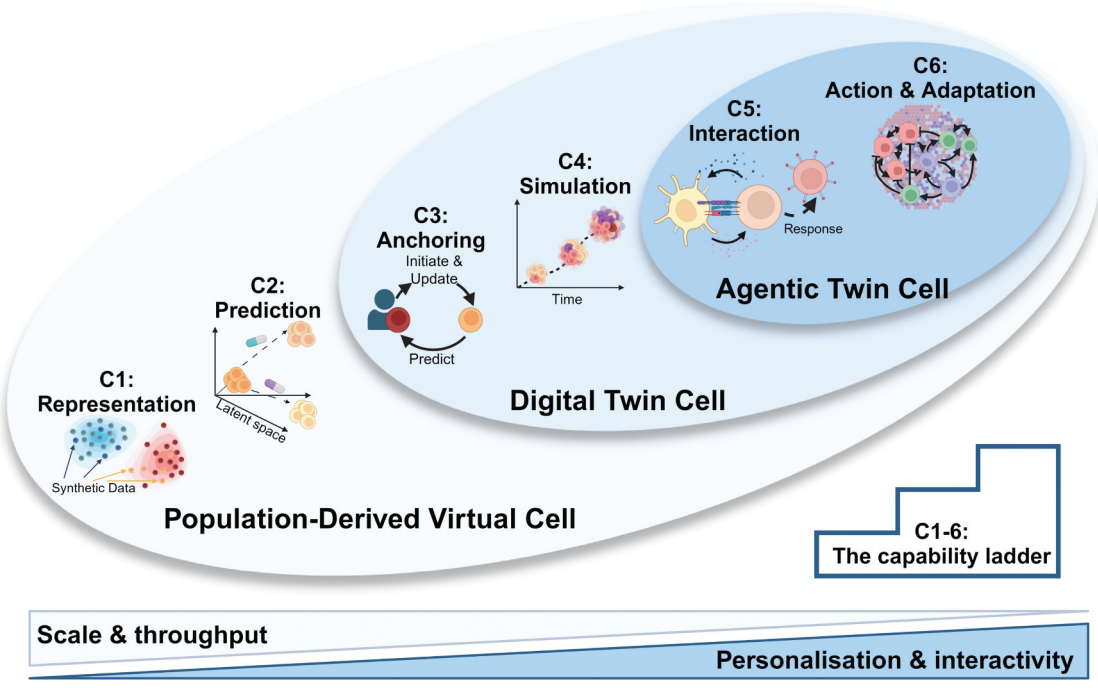


Figure 1. The capability ladder of virtual-cell models. Population-derived virtual cells support representation and prediction (C1–C2); digital twin cells extend these capabilities through counterpart anchoring, longitudinal simulation and recalibration (C3–C4); and agentic twin cells further incorporate multicellular interaction and closed-loop action and adaptation (C5–C6). A model is assigned to a class only when it demonstrates every capability that defines that class and all capabilities below it. Two boundary cases lie outside the main ladder: spatially informed foundation models incorporate spatial context without a persistent biological counterpart, and conventional agent-based models implement interaction and action without counterpart anchoring. Created in BioRender. Li, X. (2026). https://BioRender.com/qo2xoo2.

Individual capabilities can nevertheless appear outside this order, and these cases define the boundaries between classes. A foundation model that incorporates spatial context [20] but has no persistent biological counterpart remains a spatially informed virtual cell rather than an agentic twin. Similarly, conventional agent-based models implement cellular interaction and action [25,26] but lack anchoring to a measured counterpart and therefore are not agentic twins. A model calibrated only to a patient's baseline is personalized but lacks C4; it becomes a digital twin only when subsequent observations are used to evaluate and update its predictions [13].

To illustrate the progression, consider a patient-derived colorectal cancer organoid line exposed to a candidate drug. A virtual cell pretrained on population data, such as single-cell chemical screens across cancer cell lines [27], can predict how comparable cells respond, including to compounds absent from training [28], but carries no information about this patient. A digital twin cell estimates the state of this line at baseline (C3), forecasts its response at a later timepoint (C4), and is evaluated and updated when the line is profiled again, with live imaging between molecular timepoints providing an additional nondestructive update channel [29] (Figure 2). An agentic twin cell adds co-cultured fibroblasts or immune cells whose interactions with tumor cells are modeled (C5) and acted upon (C6), asking whether surrounding cells buffer or amplify the drug response. Each step therefore adds a data requirement: repeated measurement of a single counterpart for the digital twin, and measurement of cell–cell interactions under perturbation for the agentic twin.

2.1. Population-level virtual cells

Population-level virtual cells form the computational foundation of this progression, learning from population data to generate biologically plausible cellular states. They establish C1–C2 through structured representation and prediction of unmeasured states. By learning transferable priors from large cellular datasets, these models support state inter-/extrapolation, perturbation prediction, data integration and denoising. They operate across population-derived distributions without persistent linkage to an identified biological counterpart, with generalization to unseen biological contexts remaining a key challenge.

Modern virtual-cell approaches emerged from probabilistic latent-variable models such as scVI, which established generative modeling of single-cell transcriptomes [16]. Foundation models have since expanded the scale, scope and transferability of cellular representations [17,30,31]. Geneformer supports network inference and *in silico* perturbation [30], scGPT multiomic integration and perturbation prediction [17], SATURN cross-species generalization [31], and Nicheformer incorporation of spatial context [20]. Together, these approaches span C1 representation learning and C2 prediction, increasingly enabling the generation of unmeasured cellular states across biological contexts.

Generative perturbation models, often trained on pooled clustered regularly interspaced short palindromic repeats (CRISPR) screens [32], have progressively incorporated richer inductive biases, including latent vector arithmetic [8], disentangled perturbation, dose and cellular-context embeddings [9,28], optimal-transport maps between control and perturbed states [18,33], gene–gene graphs for multigene perturbations [19], and multi-scale architectures spanning cellular contexts [10]. Related generative approaches extend to synthetic patient cohorts and control arms [34], broadening their application across biological scales.

Population-level virtual cells already enable prediction across diverse cellular contexts, although performance can decline for unseen cell types, perturbations and donors [5]. Current models predominantly predict transcriptomic endpoints, with cellular behaviors and continuous trajectories

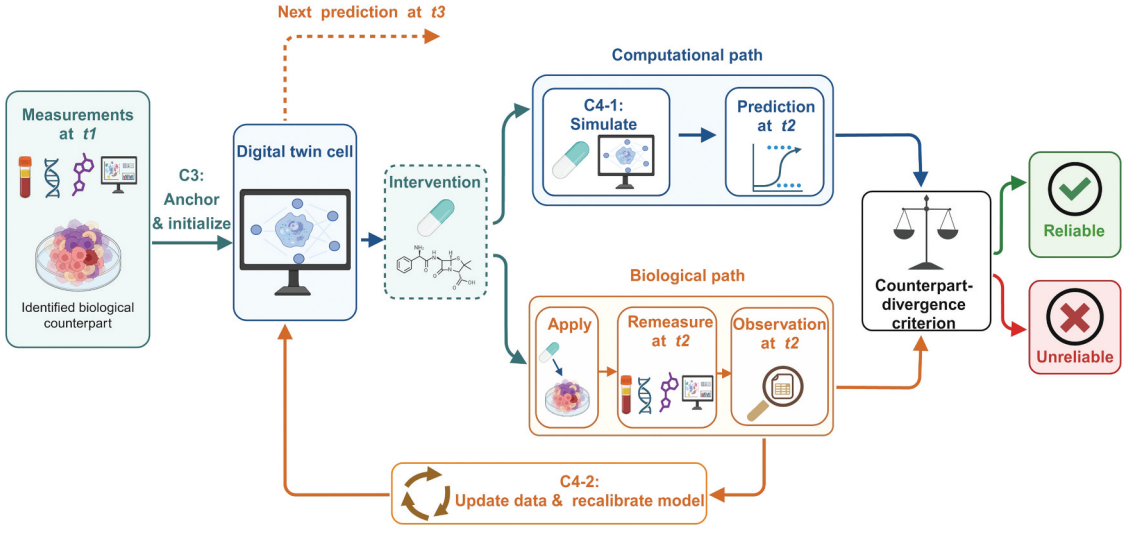


Figure 2. Longitudinal evaluation and update cycle of a digital twin cell. Baseline measurements at *t1* anchor and initialize the model (C3). The twin then simulates the future state, with or without intervention (C4-1), which is compared with the biological state observed at *t2*. A counterpart-divergence criterion evaluates predictive reliability before the new observation updates the model (C4-2), preventing temporal data leakage and initiating the next prediction cycle toward *t3*. Created in BioRender. Li, X. (2026). https://BioRender.com/jni41rx.

representing important next frontiers [5,35]. As these models scale, their greatest potential may lie as transferable foundation models that encode broad biological knowledge, extrapolate beyond observed data and adapt to specific datasets, perturbations and tasks, including settings with limited training data.

2.2. Digital twin cells: from population prediction to longitudinal twinning

Digital twin cells extend population-level virtual cells by establishing a persistent link to an identified biological counterpart. Originating in engineering, digital-twin concepts are increasingly applied to drug discovery and clinical research [11–13]. At the cellular scale, a digital twin cell is a specimen-specific computational surrogate that extends representation and prediction (C1–C2) through counterpart-specific state estimation (C3) and longitudinal simulation and updating (C4) [13]. The counterpart may be a defined cell population, patient-derived organoid, biopsy-derived system or tissue niche. Although most single-cell assays provide destructive snapshots, dynamic molecular signals, such as splicing dynamics and transcription-factor activity, can reveal the direction and magnitude of cellular state transitions and inform prediction of future states [6,7].

C3 and C4 connect the current state of a biological counterpart with its future trajectory. C3 estimates the current state from partial observations, providing the starting point for C4 to predict how that state will evolve over time or under intervention. Together, they establish three requirements for a digital twin cell: a defined biological counterpart, a longitudinal update channel and a falsifiable criterion for predictive reliability [13,36].

The biological counterpart and its boundaries must first be specified. Initialization may integrate multimodal molecular, imaging and clinical data [6], but must retain the variables relevant to the intended application. A longitudinal update channel then connects observations and predictions through a measure–simulate–remeasure–evaluate–recalibrate loop (Figure 2) [13]. A prediction at *t1* is evaluated against the observation at *t2*, which informs the next model state and prediction at *t3*.

Finally, this relationship must remain falsifiable. Predefined outcomes should quantify divergence from observed biology and determine whether the twin remains an adequate surrogate. Together, these elements establish a persistent and testable relationship between model and counterpart. Verification, validation and uncertainty-quantification frameworks remain immature [36], making rigorous evaluation essential.

Translating these requirements into practice places specific demands on study design. At least two longitudinal timepoints are needed to anchor the twin and prospectively evaluate its predictions, although repeated sampling enables iterative evaluation and updating. Such designs are readily achievable in patient-derived organoids and *ex vivo* tissues [11,29], but more challenging in patients. Prospective evaluation should also quantify predictive uncertainty, particularly when candidate therapies have similar projected effects. Model updating may capture transient variation, noise or batch effects, making it important to distinguish improved fit from improved prediction [36].

Patient-derived organoids are currently the most tractable biological counterparts for cellular digital twins. Organoid research now integrates microfluidics, gene editing, imaging, multi-omic profiling and AI, supports living biobanks, and is applied to disease modeling, drug screening and toxicology, although standardization and batch-to-batch consistency remain unresolved [37]. Their value as counterparts depends on whether they reflect the patient: in a living biobank from heavily pretreated patients with metastatic colorectal and gastro-esophageal cancer enrolled in phase 1/2 trials, organoids closely resembled the original tumors, and comparison with clinical outcomes suggested that organoid drug responses can recapitulate patient responses [38]. Organoids also support repeated measurement. Deep-learning platforms track the growth and morphology of single organoids over time without destroying them, providing the longitudinal update channel that destructive omics assays lack [29]. Replicate wells from a single line further allow terminal profiling of some wells while others continue in culture, enabling predictions to be evaluated at several timepoints from one counterpart. Organ-on-a-chip devices add fluid flow, multiple cell types and inter-organ coupling, and have reproduced human clinical responses to drugs, toxins and radiation [39], providing the multicellular context required for C5 and C6. Because variation between organoid batches can be mistaken for biological change, twins built on organoid systems should model batch as an explicit source of variation [37].

The biological scope of the twin should ultimately match its intended endpoint [13]. The chosen counterpart determines which observations inform the model, which trajectories can be predicted and how predictions are evaluated. Defining this scope is therefore central to developing digital twins with meaningful and reliable predictive capabilities.

2.3. Agentic twin cells: from prediction to interactive simulation

An agentic twin cell integrates counterpart-specific digital twinning, multicellular interaction and biologically constrained action. It inherits representation, prediction, counterpart anchoring and longitudinal simulation (C1–C4), while adding context-dependent interaction (C5) and closed-loop action (C6). C5 enables represented cells to perceive and respond to their local environment, allowing signals from one population or compartment to influence another. C6 translates cellular state and context into biologically constrained actions that modify the environment and shape subsequent behavior (Figure 3). Together, these capabilities enable interactive simulation of multicellular systems.

Conventional cellular agent-based models provide established computational foundations for these capabilities. PhysiCell couples cellular phenotypes to a dynamic microenvironment, whereas PhysiBoSS links intracellular signaling networks to cell-level agents [25,26]. Agentic twin cells combine such interactive simulations with counterpart anchoring and longitudinal updating: agents and their environment are

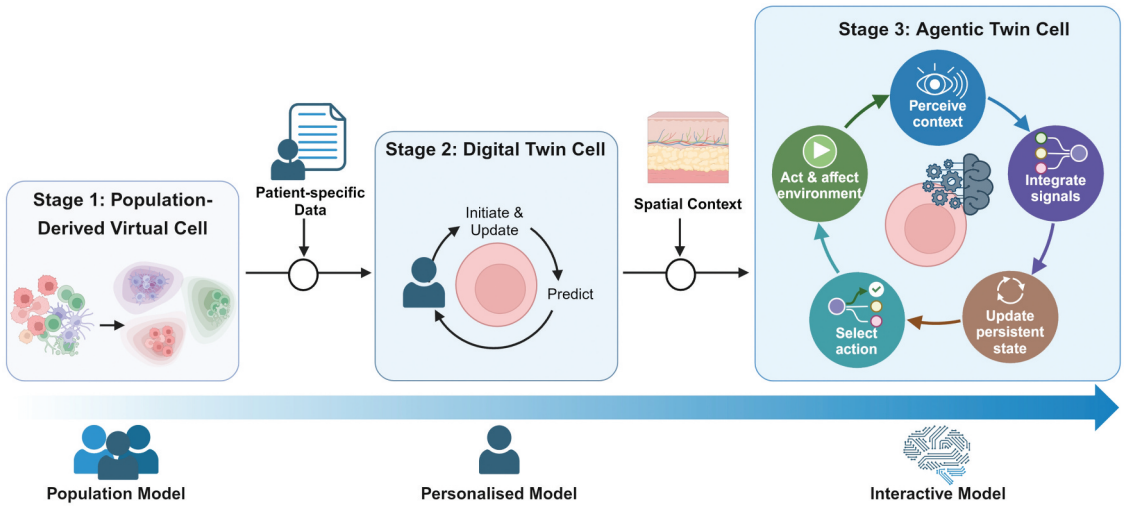


Figure 3. Data inputs and feedback loops that distinguish each virtual-cell model class. Each panel shows only what its model class adds. Virtual cell: population-scale data from which plausible cellular states are generated, without a specific biological counterpart. Digital twin cell: a counterpart-specific time series and the prediction–recalibration (update) loop that links the model to it. Agentic twin cell: spatial context and the behavioral loop of perception, state updating, action selection and environmental feedback, which operates alongside the model-level loop inherited from the digital twin cell. Created in BioRender. Li, X. (2026). https://BioRender.com/v3r6ij7.

initialized from a defined biological counterpart, predictions are prospectively evaluated against subsequent observations, and the model is updated accordingly. This couples the interactive capabilities of agent-based modeling with the persistent biological correspondence of a digital twin.

Two feedback loops connect these components. The behavioral loop operates within the simulation, where cellular actions modify the environment and thereby shape subsequent cellular inputs. The model-level loop connects simulation with experiment, where new observations evaluate counterpart-specific predictions before updating the model. Together, these loops couple interactive multicellular dynamics with prospective, falsifiable twinning.

Implementation can flexibly combine learned and mechanistic modules according to the biological question. Foundation or world models can encode state transitions, spatial encoders represent neighborhoods and gradients, mechanistic or neural-symbolic modules impose biological and physical constraints, and reinforcement-learning or other policy formulations map states and percepts to actions [15,40–42]. Large language models may further support perception, knowledge integration and policy generation [43]. These modular architectures provide a flexible foundation for implementing agentic twin cells across biological systems.

2.3.1. Spatial grounding, interaction and action

Spatial and physiological context provides the environment in which agentic twin cells interact and act. In tissues, this context includes architecture, neighboring cell identities and states, ligand and metabolite gradients, extracellular matrix and physical constraints [7,44]. Spatial omics and imaging can therefore ground cellular states within native microenvironments, including epidermal–dermal–immune compartments in inflammatory skin disease and malignant–stromal–immune niches in tumor ecosystems [45–48]. For circulating or migratory cells, context may instead comprise soluble signals, vascular transitions and transient cellular contacts. The representation should ultimately reflect the biology of the system being modeled.

C5 translates this context into interactions between represented cells. Ligand–receptor inference identifies potential signaling between cell types [21,44,49], whereas spatially informed approaches incorporate distance, local neighborhoods and signal propagation to identify which cells are positioned to influence one another [22,23,50]. Spatial perturbation models and spatially resolved genetic screens increasingly provide data to test whether perturbing one cellular component produces the predicted downstream responses in others [51–54]. These approaches provide the interaction channels through which cellular outputs can alter the subsequent inputs of neighboring cells.

Computationally inferred interactions represent hypotheses rather than evidence of functional signaling. Ligand–receptor methods score co-expression against curated interaction resources, and a systematic comparison of 16 resources and seven methods showed that both choices strongly influence the predicted interactions, although predictions were broadly consistent with spatial colocalisation and receptor protein abundance [55]. Evidence for an interaction can therefore be graded at four levels: (i) co-expression of ligand and receptor in sender and receiver populations [21,49]; (ii) spatial proximity within the range of the signal [22,23]; (iii) a downstream response in receiver cells, such as induction of known target genes of the inferred ligand [56]; and (iv) perturbation of the ligand, receptor or sender population that alters the receiver as predicted [51–54]. Only the fourth level establishes functional communication. Multicellular simulations become more reliable when they admit only interactions supported at level

(ii) or above and impose biological constraints, including distance-dependent decay of secreted signals, the requirement for physical contact in contact-dependent signaling, and receptor protein abundance as a check that transcripts alone cannot provide [22,23,55].

C6 extends interaction to biologically constrained action and feedback. Actions may include differentiation, apoptosis, secretion, migration or proliferation, represented as discrete transitions, continuous quantities or probability distributions over feasible behaviors [15,25,26]. The distinction between C5 and C6 is therefore functional: C5 determines how a cell perceives and is influenced by its context, whereas C6 determines how it acts on that context and thereby changes what happens next. In inflamed skin, for example, immune-derived interleukin-17 (IL-17) and tumor necrosis factor (TNF) provide contextual signals that alter keratinocyte states (C5) [46]. Keratinocytes may subsequently increase chemokine secretion (C6), modifying immune-cell recruitment and the local cytokine environment, which in turn alters subsequent cellular inputs [45,46]. This closes the behavioral feedback loop.

Implementing this loop requires persistent cellular state and biologically constrained action policies. Persistent state allows previous exposures and actions to influence subsequent responses, while a policy maps cellular state and contextual inputs to feasible actions under regulatory, signaling, metabolic and physical constraints. Here, a policy denotes a state-dependent mapping from biological inputs to cellular behaviors, for which human-interpretable multicellular grammars provide one inspectable implementation [15]. Environmental updating then propagates the consequences of secretion, migration, proliferation or death to the same cell and its neighbors, coupling successive simulation steps [25,26].

Agentic twin cells thereby integrate two complementary feedback processes: the behavioral loop between cells and their environment, and the model-level loop between simulation and biological observation inherited from digital twinning. Therapeutic optimization can operate above these cellular dynamics, comparing predicted outcomes across drugs, combinations or schedules without assigning therapeutic objectives to the simulated cells themselves.

2.3.2. Emerging evidence, validation and translational opportunities

Emerging approaches provide many of the building blocks required for agentic twin cells (Table 1). CellAgentChat models cell–cell communication and *in silico* perturbation using agent-based representations [24], interpretable multicellular grammars encode inspectable behavioral rules [15], and scMultiSim generates controlled multi-omic and spatial data under predefined regulatory and interaction networks [57]. Spatial communication methods and perturbation assays further provide mechanisms and experimental evidence for context-dependent interaction (C5) [21–23,44,50–54]. Together, these approaches establish a growing toolkit for modeling cellular perception, interaction and action, creating the foundation for their integration with counterpart anchoring and longitudinal updating. To our knowledge, no published system yet integrates all six capabilities with prospective validation against a biological counterpart. Agentic twin cells therefore remain conceptual as a complete model class, although each of their components has clear precedent (Table 1).

Validating this integration requires testing both biological correspondence and behavioral feedback. Counterpart divergence evaluates whether prospective predictions remain aligned with subsequent biological observations. Complementarily, we propose counterfactual coupling to quantify action–environment dependence within the simulation. Removing the action-to-environment channel while holding other inputs fixed reveals how reciprocal interactions influence predicted outcomes. Its interpretation should be grounded in perturbation data, temporal validation and comparison with simpler baselines, as the magnitude of coupling alone does not establish biological validity. Together, counterpart divergence and counterfactual coupling provide complementary tests of model-level correspondence and within-simulation feedback.

The translational opportunity lies in modeling treatment responses as multicellular processes rather than isolated cellular outcomes. Beyond forecasting target inhibition in a patient-derived system, agentic twin cells could test whether neighboring cells compensate, resistant states propagate through tissue or treatment sequence reshapes the multicellular response. Near-term applications are particularly compelling in organoids, *ex vivo* tissues and repeatedly sampled disease systems, where predictions can be experimentally tested [11,29]. Ultimately, population-level virtual cells provide transferable biological priors, digital twin cells anchor these models to an identified counterpart, and agentic twin cells extend them through context-dependent interaction, action and feedback.

3. Biological and biomedical applications

The capability framework translates into four major biological and translational applications: modeling cell differentiation, reconstructing disease trajectories, predicting perturbation responses and designing therapeutic interventions (Figure 4). Increasing capabilities extend these applications from population-level prediction toward personalized and interactive simulation of multicellular systems. Inflammatory skin disease and tumor ecosystems provide complementary examples, representing an accessible tissue shaped by epithelial–stromal–immune feedback and a heterogeneous system characterized by clonal evolution and adaptive resistance.

3.1. Where is the system now, and where will it go?

Understanding where a cellular system is now and where it will go requires linking state estimation with trajectory prediction. Population-derived fate landscapes and trajectory models can identify candidate states and transitions [16,17,58], while longitudinal observations are needed to determine whether these trajectories predict how an identified biological system evolves over time. Time-resolved lineage data, longitudinal imaging and repeated molecular measurements provide complementary routes for such evaluation. Donor-specific

Table 1. Capability framework, current evidence and validation gaps for virtual-cell models.

Capability	Operational requirement	Representative current evidence	Maturity[a]	Principal validation gap
C1 Representation	Generative representation of cellular states that captures biological variation beyond technical effects	scVI [16], Geneformer [30], scGPT [17], SATURN [31], Nicheformer [20]	Established. Multiple foundation models and independent benchmarks [16,17,20,30,31,68]; zero-shot transfer remains limited [67]	Transferability of learned representations across biological contexts and downstream tasks
C2 Prediction	Generation of an unmeasured cellular state under specified conditions	scGen [8], CPA and chemCPA [9,28], CellOT and conditional Monge gap [18,33], GEARS [19], STATE [10]	Established within represented contexts; emerging for unseen cell types, perturbations and donors. Independent benchmarks show limited generalization and no consistent gain over simple baselines [5,69,70]	Robust generalization to unseen cell types, perturbations and donors
C3 Anchoring	Persistent linkage to a defined biological counterpart and state estimation from counterpart-specific observations	Patient-derived organoid systems [29]; patient- and tumor-level twin frameworks [11,34]; *in vivo*-calibrated bone-metastasis twin [63]	Emerging. Patient- and tumor-level frameworks; no prospective cell-level state estimation to date [11,34,63]	Prospective counterpart-specific state estimation from partial and multimodal observations
C4 Simulation	Longitudinal simulation and updating, with prospective evaluation against subsequent observations	Longitudinal calibration [63]; iterative update frameworks [13,36]	Emerging. Longitudinal calibration in an *in vivo* system; update frameworks proposed [13,36,63]	Prospective evaluation against later timepoints and predefined counterpart-divergence criteria
C5 Spatial Interaction	Context-dependent channels through which cellular outputs influence subsequent inputs of neighboring or interacting cells	Co-expression, level (i): CellChat and CellPhoneDB [21,49]; spatial proximity, level (ii): COMMOT [22], CytoSignal [23], CellNEST [50]; receiver response, level (iii): NicheNet [56]; perturbation, level (iv): SpatialProp [51], Perturb-DBiT [52], Perturb-Multimodal [53], Perturb-FISH [54]	Established for inference; emerging for validated propagation. Mature inference methods [21–23,49,50,56]; spatial perturbation screens now test propagation [51–54]	Experimental validation that predicted interactions transmit perturbation effects across cells and compartments
C6 Closed-loop action and adaptation	Biologically constrained actions that modify the environment and feed back to shape subsequent cellular states and responses	PhysiCell [25], PhysiBoSS [26], CellAgentChat [24], interpretable multicellular grammars [15], scMultiSim [57]	Established in simulation; conceptual when anchored to a counterpart. Agent-based simulators and multicellular grammars [15,24–26]	Validation that simulated actions and environmental feedback reproduce observed multicellular dynamics

[a]Established: predictive performance on held-out data reported in several independent studies or independent benchmarks. Emerging: demonstrated in proof-of-concept studies in specific systems, without prospective counterpart-specific validation. Conceptual: components exist but have not been integrated or tested together. Levels (i)–(iv) for C5 follow the evidence grading in Section 2.3.1.

organoid and induced pluripotent stem-cell systems are particularly valuable when genetic background influences differentiation outcomes [59]. At the multicellular level, trajectories may additionally depend on lateral inhibition, competition for niche signals and density-dependent feedback, requiring interactions between cells and their environment to be modeled [15,60].

Inflammatory skin disease illustrates how these capabilities could support prediction of disease trajectories. Serial biopsies, imaging and molecular measurements could establish whether an individual lesion is progressing toward resolution, persistence or relapse [45,46,61]. In tumors, longitudinal measurements could similarly track changes in malignant-cell states, clonal composition and microenvironmental organization during treatment [47,48]. In both settings, prospective predictions can be compared with subsequent observations to evaluate predictive reliability and update the model as the biological system evolves.

3.2. What happens after perturbation, and does the tissue compensate?

Perturbations rarely act on cells in isolation. A genetic, pharmacological or environmental intervention first alters the targeted cellular state, but its consequences can propagate through signaling, cellular interactions and changes in the local microenvironment. Predicting these effects remains challenging across unseen perturbations, donors and biological contexts [3,5,35]. Counterpart-specific measurements can personalize the predicted response, while spatial information provides the context needed to determine how perturbation effects propagate across cellular neighborhoods.

This propagation can reshape the response beyond the directly targeted cells. Spatial perturbation assays and spatially resolved genetic screens increasingly enable these effects to be traced across interacting cellular compartments [51–54]. In inflammatory skin disease, reducing keratinocyte inflammatory responses may weaken immune recruitment and cytokine

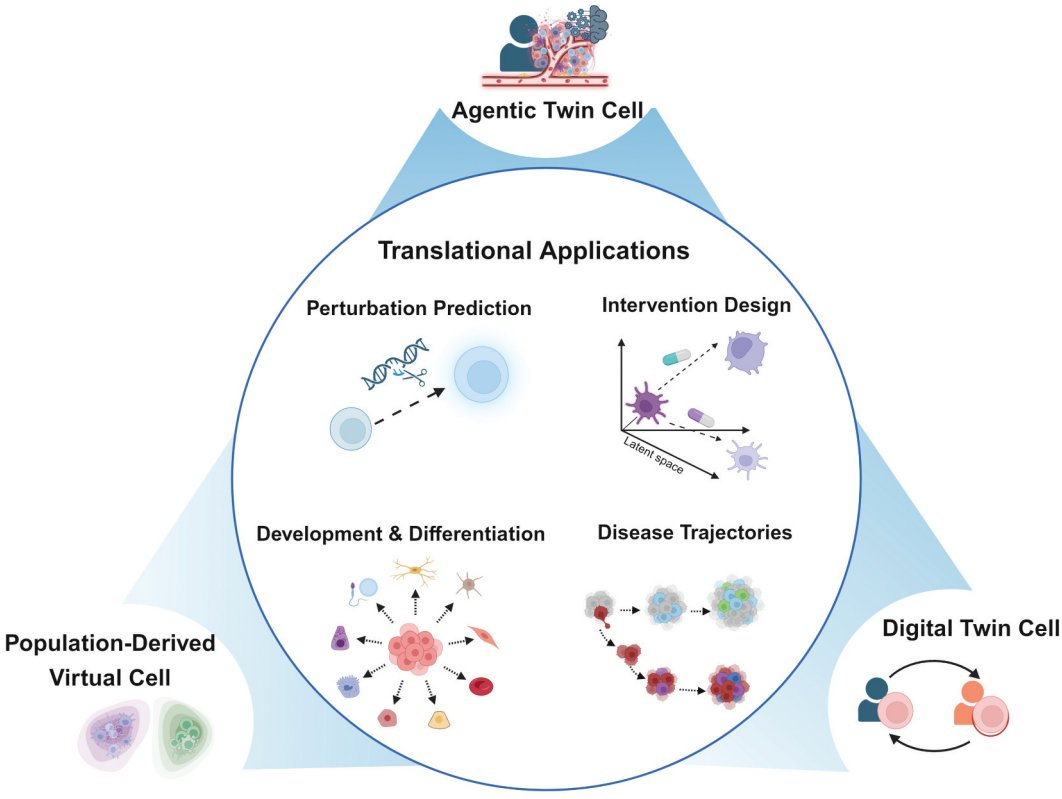


Figure 4. Translational applications of virtual-cell models. Virtual-cell models support four major translational applications: cell differentiation, disease trajectories, perturbation prediction and intervention design. Created in BioRender. Li, X. (2026). https://BioRender.com/x1tseul.

feedback, shifting the tissue toward lesion resolution [45,46,51]. In tumors, treatment can reconfigure malignant–immune–stromal interactions and create selective pressures that favor resistant clones [47,48,62].

The resulting treatment response is therefore a property of the multicellular system rather than the targeted cell alone. Predictive models should capture not only the immediate effect of an intervention, but also how that effect propagates through tissue and triggers compensatory or adaptive responses. Such interaction-aware simulations could reveal whether an initially effective perturbation disrupts a pathogenic circuit or is ultimately buffered by the surrounding tissue.

3.3. Which intervention should be tested?

Once disease trajectories and perturbation responses can be predicted, the next question is which intervention is most likely to redirect the system toward a desired biological state. Candidate targets, doses, combinations and treatment sequences can be compared by their predicted effects on both the targeted cells and the wider multicellular system. Therapeutic objectives should remain separate from the cellular model itself, with an optimization layer ranking interventions according to predefined biological endpoints.

This creates an iterative cycle between prediction and experiment. Mechanistically distinct interventions can be simulated, prioritized according to their predicted outcomes and uncertainty, and experimentally tested to refine subsequent predictions [11,13,36]. In inflammatory skin disease, this could prioritize interventions that disrupt pathogenic epidermal–immune feedback while preserving barrier function. In tumors, it could identify treatment sequences that constrain resistant clones while maintaining favorable immune and stromal responses [46,47].

The near-term opportunity is therefore not autonomous treatment selection, but increasingly informed prioritization of what to test next. By integrating counterpart-specific trajectories with multicellular responses and uncertainty, virtual-cell models could focus experimental effort on interventions most likely to produce durable system-level responses, while retaining human oversight over therapeutic decisions.

3.4. Where virtual-cell models enter drug discovery

Published examples already reach four stages of drug discovery, although nearly all operate at C1–C2 (Table 2). For target identification, Geneformer fine-tuned with limited patient data identified candidate therapeutic targets for cardiomyopathy through *in silico* perturbation [30]. For compound prioritization, sci-Plex profiled approximately 650,000 single cells from three cancer cell lines exposed to 188 compounds [27], and models trained on such screens predict responses to doses, combinations and compounds absent from training [9,28]. For drug-response prediction, CellOT predicted single-cell responses in held-out patients, including lupus samples exposed to interferon-β and glioblastoma samples treated with panobinostat [18]. For treatment optimization, agent-

Table 2. Drug-discovery stages, published examples and the capabilities they demonstrate.

Stage	Question	Published example	Capability demonstrated	Contribution of higher capabilities
Target identification	Which gene, if modulated, shifts diseased cells toward a healthy state?	Geneformer *in silico* perturbation in cardiomyopathy [30]	C1–C2	Patient-specific target ranking (C3); tissue compensation (C5)
Compound prioritization	Which compounds produce the desired cellular response?	sci-Plex chemical screens [27]; CPA and chemCPA [9,28]	C2	Ranking within a single patient-derived line (C3–C4)
Drug-response prediction	How will this patient's cells respond?	CellOT in held-out patients [18]; organoid drug screens [38]	C2; experimental counterpart	Updating as resistance emerges (C4)
Treatment optimization	Which dose, combination or sequence?	Bone-metastasis agent-based twins [63]	C5–C6 in simulation, calibrated to *in vivo* data	Specimen-level anchoring and prospective evaluation (C3–C4)
Clinical trial design	Can twins augment or replace control arms?	Synthetic control arms [34]; pan-cancer twins [43]	Patient-level twins	Cellular mechanisms underlying patient-level forecasts

based digital twins of bone metastasis calibrated with *in vivo* data reproduced the anti-angiogenic effect of cabozantinib and the antiresorptive effect of zoledronic acid [63]. Across these examples the same gap remains: predictions are rarely evaluated prospectively in the same biological counterpart, and few models account for how surrounding tissue modifies the response.

4. Challenges and requirements for reliable virtual-cell models

Virtual cells, digital twins and agentic twins originate from different research traditions and therefore carry different conceptual expectations. A virtual cell should be distinguished from engineered or minimal cells in synthetic biology [64]. Digital twinning implies persistent correspondence with an identified biological counterpart rather than one-off personalization. Agentic twinning extends this relationship through context-dependent interaction, biologically constrained action and feedback, without implying cellular intentionality or prescribing a particular computational architecture. As these fields converge, terminology should therefore be defined by demonstrated capabilities rather than model architecture or branding.

4.1. Data and measurement limitations

Increasing data scale does not overcome the fundamental problem of biological observability. Most single-cell molecular assays are destructive, preventing repeated profiling of the same cell. Cellular twins must therefore rely on serial measurements from related cell populations or tissue regions, complemented by imaging, biomarkers and functional readouts. Sampling intervals may also exceed the timescale of the processes being modeled, leaving transient responses and early adaptation unobserved. Pseudotime and RNA velocity can provide temporal constraints, but cannot substitute for longitudinal measurements or establish causality [35,58,65].

No single modality captures the complete biological system. Bulk sequencing obscures cellular heterogeneity, dissociative single-cell assays remove spatial relationships, and spatial technologies trade molecular coverage against resolution [7]. Computational integration can align complementary measurements, but cannot recover information that was never observed [6]. Generative AI models can extrapolate beyond these observations, creating both an opportunity and a fundamental validation challenge. Their ability to generate plausible cellular states beyond the training data enables prediction in data-limited and previously unseen biological contexts, but can also produce biologically plausible yet incorrect states. Distinguishing faithful extrapolation from generative hallucination is therefore central to establishing predictive reliability.

Several factors limit transfer across datasets and biological systems. Donor heterogeneity is substantial: across 711 stem-cell lines from 301 individuals, 5–46% of phenotypic variation arose from differences between individuals [59], and single-cell analyses that ignore replicate variation can report hundreds of false-positive differentially expressed genes [66]. Tissue and cell-type context also matter: without fine-tuning, scGPT and Geneformer were outperformed by simpler methods on some tasks [67], and a benchmark of 27 methods across 29 datasets found that cellular-context embedding is needed for generalization [5].

Technical variation adds further limits. Batch effects span laboratories, protocols and conditions, and integration methods trade batch removal against biological signal [68]. Platform choice also affects evaluation: across 10 datasets and three platforms, systematic differences between perturbed and control cells inflated common performance metrics [69]. Models intended for transfer should therefore be evaluated on splits that hold out entire donors, batches, tissues and platforms to prevent data leakage, with performance reported per split.

The measurement challenge becomes greater as models incorporate interaction and feedback. Agentic twins require observations not only of molecular states, but also of cellular behaviors and their consequences for the surrounding environment. Transcriptional changes alone cannot establish proliferation, migration, secretion, differentiation or tissue function [3,35]. Errors in inferred states or interactions may also propagate through successive simulation steps, amplifying initially plausible but incorrect predictions. Validation systems should therefore increasingly combine perturbations with measurements of molecular state, cellular phenotype, spatial context and temporal dynamics.

4.2. Prospective validation and biological plausibility

Prospective validation is essential to distinguish prediction from reconstruction. Observations used to evaluate a twin

should therefore remain separate from those used for initialization or updating [36]. Future time points can be withheld to test whether model updates improve forward prediction rather than simply reconstructing the system's history. Counterpart-specific predictions should also be benchmarked against strong population-level models to determine whether personalization provides measurable predictive value.

No single metric is sufficient to establish that a digital twin is reliable, and we therefore recommend four complementary criteria (Supplementary Table S1). First, predictions should be evaluated on the change from control rather than on absolute expression, as metrics computed on absolute profiles reward reproduction of systematic differences between perturbed and control cells [69]. Each model should also be benchmarked against deliberately simple baselines, such as the mean response or an additive model, which current deep-learning models have not yet outperformed [70]. Second, predictive uncertainty should be calibrated against observed error, assessed through interval coverage and proper scoring rules that penalize both overconfidence and unnecessarily wide intervals [36,71]. Third, predictive error should be tracked across the forecast horizon, with recalibration judged by whether it reduces error at the next unseen timepoint rather than by improved fit to past observations [36]. Finally, for drug discovery the decisive test is often whether the model ranks interventions correctly, which can be evaluated against responses measured in organoids or perturbation screens [29,38].

Validation should become progressively more demanding as models extrapolate beyond observed biology. Performance should be evaluated across unseen donors, perturbations and biological contexts, particularly when interventions move the system beyond conditions represented during training [35]. Statistical similarity to observed states is not sufficient: a prediction may appear biologically plausible while violating regulatory constraints, bypassing obligatory intermediates or incorrectly representing interactions [35]. Predicted multicellular effects should therefore be challenged experimentally by perturbing relevant cells, signals or interaction channels and measuring both direct and downstream responses.

Ultimately, validation should establish not only whether a model produces plausible states, but whether it makes reliable predictions before the corresponding biology is observed. Predictive uncertainty should be calibrated against observed error, and translational claims evaluated prospectively against predefined experimental or clinical endpoints [36]. Biological plausibility constrains what could happen; prospective validation determines whether the model can predict what does happen.

4.3. Challenges specific to agentic twin cells

Agentic twin cells introduce an additional challenge: translating cellular state and environmental context into biologically meaningful actions. Some behaviors can be represented as discrete events, such as division, death or initiation of migration, whereas others, including secretion, motility and gene expression, vary continuously and are more difficult to quantify and validate. Practical models will therefore probably combine continuous state dynamics with probabilistic cellular events. Importantly, these actions should emerge from biological state and local context rather than from a predefined therapeutic objective. Directing simulated cells toward a 'healthy state,' for example, would encode the desired outcome rather than predict cellular behavior. Moreover, actions that benefit individual cells need not restore tissue homeostasis or improve organism-level outcomes.

As cellular actions become coupled, a second challenge is distinguishing genuine biological emergence from artifacts of the simulation. Different local rules can generate similar tissue-level patterns, creating substantial identifiability problems and motivating targeted perturbations, mechanistic constraints and controlled simulations with known ground truth [57]. Predicted interactions should respect receptor expression, ligand availability, diffusion, cellular contact and physical constraints [22,23,50]; otherwise, unconstrained communication can generate plausible but artificial collective behavior. Counterfactual-coupling tests can probe whether altering neighboring cells or signals changes predicted behavior, but such sensitivity alone does not establish biological validity. Experimental perturbation remains necessary to determine whether the same dependencies occur in the biological counterpart.

These challenges are amplified by the scale of multicellular simulation. Shared policies, hierarchical representations and event-driven updating may reduce computational demands, but increasing efficiency does not resolve biological uncertainty. The central challenge for agentic twin cells is therefore not simply to generate complex emergent behavior, but to demonstrate that the interactions, actions and feedback producing that behavior correspond to those operating in the biological system.

5. Conclusion

AI-powered virtual cells are beginning to transform cellular modeling from description toward prediction. Population-scale models can already learn transferable representations and generate plausible responses to perturbation, but longitudinal prediction, individualization and multicellular adaptation remain less established. Patient-derived organoids, *ex vivo* tissues and spatial perturbation systems provide tractable settings in which these capabilities can now be tested prospectively [29,52–54].

The immediate opportunity is to test three specific claims in biological systems that can be measured repeatedly. First, whether pretrained virtual cells adapted with limited counterpart data predict held-out perturbation responses more accurately than simple baselines [69,70]. Second, whether digital twins of patient-derived organoids, recalibrated between time points, reduce forecast error at the subsequent timepoint compared with a non-updated population model [11,29,36]. Third, whether perturbation effects propagate across cell types in spatially resolved screens as predicted by agentic twins [51–54].

The field can therefore progress from generating plausible cellular states toward building computational systems that remain predictively coupled to biology. Digital twin cells

provide this connection through counterpart-specific anchoring and longitudinal updating; agentic twin cells extend it through interaction, action and environmental feedback. Their value will ultimately depend not on the complexity of the model or plausibility of its simulations, but on whether it can predict biology that has not yet been observed.

6. Expert opinion

Cellular representation learning has advanced faster than our ability to predict how cells respond to intervention. Complex models often fail to outperform simple baselines when challenged with unseen cell types, perturbations or donors [5], precisely the settings in which predictive models could provide greatest value for drug discovery. This gap is partly evaluative: transcriptomic reconstruction remains a common endpoint, even when therapeutic success depends on cellular behavior, spatial interactions and responses that unfold over time.

We believe the greatest opportunity lies in treating AI-powered virtual cells as transferable foundation models rather than increasingly complex solutions to individual prediction tasks. Large-scale pretraining could encode biological knowledge that is subsequently adapted to specific tissues, perturbations and experimental systems, including settings where only limited data are available. This reverses the conventional modeling paradigm of increasing task-specific complexity until predictive performance saturates. The critical test is whether such models can extrapolate faithfully beyond their training distributions rather than merely generate biologically plausible states.

Realizing this potential will require datasets designed for prediction rather than reconstruction. Serial molecular, spatial, imaging and functional measurements should be linked to defined biological counterparts before and after controlled perturbations, supported by standardized pipelines for acquisition, integration and sharing. Predictions should be generated before subsequent observations become available and used to prioritize the next measurement or perturbation. Patient-derived organoids, *ex vivo* tissues and repeatedly sampled diseases provide particularly tractable settings for establishing this prospective prediction–experiment–update cycle [11,36,61].

Agentic twin cells could ultimately extend this cycle from individual cellular responses to adaptive multicellular systems. Their near-term value lies in generating ranked, uncertainty-aware hypotheses about perturbations, interactions and interventions that can be challenged experimentally. Added complexity should be justified by prospective improvements over simpler population models, direct experimental screening or established mechanistic approaches. Inflammatory skin disease provides an accessible setting in which epithelial–stromal–immune feedback can be repeatedly interrogated [45,46,61], whereas tumor ecosystems provide a more demanding test of whether models can anticipate clonal evolution and adaptive resistance [47,62].

The longer-term goal is not simply a more realistic simulation, but a computational model that remains predictively coupled to its biological counterpart: estimating its current state, forecasting its trajectory, comparing interventions, representing multicellular adaptation and recognizing when uncertainty precludes reliable prediction. Achieving this will require prospective benchmarks, mechanistic constraints, explicit uncertainty propagation and repeated experimental falsification. The defining advance will not be whether virtual cells can generate convincing biological futures, but whether they can reliably distinguish the futures that biology will actually realize.

Author contributions

CRediT: **Xixi Li:** Conceptualization, Project administration, Visualization, Writing – original draft, Writing – review & editing; **Nicolo Fabila:** Conceptualization, Visualization, Writing – original draft, Writing – review & editing; **Xuan Liu:** Conceptualization, Visualization, Writing – original draft, Writing – review & editing; **Christine R. Keenan:** Conceptualization, Writing – review & editing; **Robyn M. Brown:** Conceptualization, Writing – review & editing; **Phong B. H. Nguyen:** Conceptualization, Supervision, Writing – review & editing; **Michael P. Menden:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

This paper was not funded.

Declaration of interests

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

In compliance with the International Committee of Medical Journal Editors (ICMJE) recommendations, the authors acknowledge the utilization of AI services in this work: the authors of this article used ChatGPT (GPT-5.6 Sol) by OpenAI and Claude (Opus) by Anthropic exclusively for minor English corrections to improve clarity and flow in this work. All sentences revised by ChatGPT and Claude were reviewed and verified by the authors. No content was generated by ChatGPT, Claude or any other AI service.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Cell Science Program CZI, Abdulla S, Aevermann B, et al. CZ CELLxGENE discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. Nucleic Acids Res. 2025;53(D1):D886–D900.
2. Yu J, Moon J, Kim M, et al. HISSTA: a human in situ single-cell transcriptome atlas. Bioinformatics. 2025;41(4):btaf142. doi: 10.1093/bioinformatics/btaf142
3. Bunne C, Roohani Y, Rosen Y, et al. How to build the virtual cell with artificial intelligence: priorities and opportunities. Cell. 2024;187:7045–7063. doi: 10.1016/j.cell.2024.11.015

•• **Provides a roadmap for AI-powered virtual cells, identifying priorities and challenges central to predictive cellular modelling.**

4. Sun D, Gao W, Hu H, et al. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022;12(7):3049–3062. doi: 10.1016/j.apsb.2022.02.002
5. Wei Z, Wang Y, Gao Y, et al. Benchmarking algorithms for generalizable single-cell perturbation response prediction. Nat Methods. 2026;23(2):451–464. doi: 10.1038/s41592-025-02980-0
•• **Benchmarks single-cell perturbation prediction across diverse datasets, revealing generalisation limitations central to evaluating virtual-cell models.**
6. Vandereyken K, Sifrim A, Thienpont B, et al. Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet. 2023;24(8):494–515. doi: 10.1038/s41576-023-00580-2
7. Gulati G, D'silva J, Liu Y, et al. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. Nat Rev Mol Cell Biol. 2025;26(1):11–31. doi: 10.1038/s41580-024-00768-2
8. Lotfollahi M, Wolf F, Theis F. scGen predicts single-cell perturbation responses. Nat Methods. 2019;16(8):715–721. doi: 10.1038/s41592-019-0494-8
9. Lotfollahi M, Susmelj AK, De Donno C, et al. Predicting cellular responses to complex perturbations in high-throughput screens. Mol Syst Biol. 2023;19(6):e11517. doi: 10.15252/msb.202211517
10. Adduri AK, Gautam D, Bevilacqua B, et al. Predicting cellular responses to perturbation across diverse contexts with state. Cell. 2026;189(19):5914–5931.e20. doi: 10.1016/j.cell.2026.07.052
11. Bordukova M, Makarov N, Rodriguez-Esteban R, et al. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. Expert Opin Drug Discov. 2024;19(1):33–42. doi: 10.1080/17460441.2023.2273839
12. Bordukova M, Arneth AJ, Makarov N, et al. Generative AI and digital twins: shaping a paradigm shift from precision to truly personalized medicine. Expert Opin Drug Discov. 2025;20(7):821–826. doi: 10.1080/17460441.2025.2507376
13. De Domenico M, Allegri L, Caldarelli G, et al. Challenges and opportunities for digital twins in precision medicine from a complex systems perspective. Npj Digit Med. 2025;8(1):37. doi: 10.1038/s41746-024-01402-3
•• **Examines digital twins from a complex-systems perspective, informing the requirements for biological correspondence, dynamic updating and context-dependent prediction.**
14. Li B, Saini AK, Hernandez JG, et al. Agentic AI and the rise of in silico team science in biomedical research. Nat Biotechnol. 2026;44(5):711–725. doi: 10.1038/s41587-026-03035-1
15. Johnson J, Bergman D, Rocha H, et al. Human interpretable grammar encodes multicellular systems biology models to democratize virtual cell laboratories. Cell. 2025;188(17):4711–4733.e37. doi: 10.1016/j.cell.2025.06.048
• **Introduces a human-interpretable framework for multicellular modelling, providing a foundation for representing cellular interactions and behaviours in virtual-cell systems.**
16. Lopez R, Regier J, Cole MB, et al. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053–1058. doi: 10.1038/s41592-018-0229-2
17. Cui H, Wang C, Maan H, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods. 2024;21(8):1470–1480. doi: 10.1038/s41592-024-02201-0
• **Introduces scGPT, illustrating how large-scale pretraining can support transferable cellular representations and multiple downstream tasks.**
18. Bunne C, Stark SG, Gut G, et al. Learning single-cell perturbation responses using neural optimal transport. Nat Methods. 2023;20(11):1759–1768. doi: 10.1038/s41592-023-01969-x
19. Roohani Y, Huang K, Leskovec J. Predicting transcriptional outcomes of novel multigene perturbations with gears. Nat Biotechnol. 2024;42(6):927–935. doi: 10.1038/s41587-023-01905-6
20. Tejada-Lapuerta A, Schaar A, Gutgesell R, et al. Nicheformer: a foundation model for single-cell and spatial omics. Nat Methods. 2025;22(12):2525–2538. doi: 10.1038/s41592-025-02814-z
• **Integrates single-cell and spatial information in a foundation model, illustrating how cellular representations can incorporate tissue context.**
21. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088. doi: 10.1038/s41467-021-21246-9
22. Cang Z, Zhao Y, Almet AA, et al. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. Nat Methods. 2023;20(2):218–228. doi: 10.1038/s41592-022-01728-4
23. Liu J, Manabe H, Qian W, et al. CytoSignal detects locations and dynamics of ligand–receptor signaling at cellular resolution from spatial transcriptomic data. Nat Genet. 2026;58(6):1396–1408. doi: 10.1038/s41588-026-02624-9
24. Raghavan V, Zheng Y, Li Y, et al. Harnessing agent-based frameworks in CellAgentChat to unravel cell–cell interactions from single-cell and spatial transcriptomics. Genome Res. 2025;35(7):1646–1663. doi: 10.1101/gr.279771.124
25. Ghaffarizadeh A, Heiland R, Friedman SH, et al. PhysiCell: an open source physics-based cell simulator for 3-D multicellular systems. PLOS Comput Biol. 2018;14(2):e1005991. doi: 10.1371/journal.pcbi.1005991
• **Introduces PhysiCell, an agent-based framework coupling cellular behaviours with a dynamic microenvironment and providing a computational foundation for interactive multicellular simulation.**
26. Letort G, Montagud A, Stoll G, et al. PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics. 2019;35(7):1188–1196. doi: 10.1093/bioinformatics/bty766
27. Srivatsan SR, McFaline-Figueroa JL, Ramani V, et al. Massively multiplex chemical transcriptomics at single-cell resolution. Science. 2020;367(6473):45–51. doi: 10.1126/science.aax6234
28. Hetzel L, Böhm S, Kilbertus N, et al. Predicting cellular responses to novel drug perturbations at a single-cell resolution [preprint]. arXiv. 2022 [cited 2026 Nov 5]. Available from: https://arxiv.org/abs/2204.13545
29. Matthews JM, Schuster B, Kashaf SS, et al. OrganoID: a versatile deep learning platform for tracking and analysis of single-organoid dynamics. PLOS Comput Biol. 2022;18(11):e1010584. doi: 10.1371/journal.pcbi.1010584
• **Enables longitudinal tracking of individual organoids, providing a practical measurement approach for updating and evaluating cellular digital twins.**
30. Theodoris CV, Xiao L, Chopra A, et al. Transfer learning enables predictions in network biology. Nature. 2023;618(7965):616–624. doi: 10.1038/s41586-023-06139-9
• **Demonstrates transfer learning for network biology and candidate therapeutic target identification, illustrating the potential of pretrained cellular models in data-limited settings.**
31. Rosen Y, Brbić M, Roohani Y, et al. Toward universal cell embeddings: integrating single-cell RNA-seq datasets across species with SATURN. Nat Methods. 2024;21(8):1492–1500. doi: 10.1038/s41592-024-02191-z
32. Dixit A, Parnas O, Li B, et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell. 2016;167(7):1853–1866.e17. doi: 10.1016/j.cell.2016.11.038
33. Driessen A, Rajwade DA, Harsanyi B, et al. Conditional Monge gap enables generalizable single-cell perturbation modelling. Nat Mach Intell. 2026;8(6):984–996. doi: 10.1038/s42256-026-01242-8
34. Akbarialiabad H, Pasdar A, Murrell DF, et al. Enhancing randomized clinical trials with digital twins. NPJ Syst Biol Appl. 2025;11(1):110. doi: 10.1038/s41540-025-00592-0
35. Dimitrov D, Schrod S, Rohbeck M, et al. Interpretation, extrapolation and perturbation of single cells. Nat Rev Genet. 2026;27(5):349–370. doi: 10.1038/s41576-025-00920-4
36. Sel K, Hawkins-Daarud A, Chaudhuri A, et al. Survey and perspective on verification, validation, and uncertainty quantification of digital twins for precision medicine. Npj Digit Med. 2025;8(1):40. doi: 10.1038/s41746-025-01447-y
•• **Examines verification, validation and uncertainty quantification for digital twins, providing a foundation for assessing predictive reliability in precision medicine.**
37. Zhang J, Wu Y, Shen T, et al. Organoid research: new concepts and new technologies. Burns Trauma. 2026;14:tkag015. doi: 10.1093/burnst/tkag015

38. Vlachogiannis G, Hedayat S, Vatsiou A, et al. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. Science. 2018;359(6378):920–926. doi: 10.1126/science.aao2774
- **Links patient-derived organoid drug responses to clinical outcomes, supporting their use as experimentally tractable counterparts for treatment-response modelling.**
39. Ingber DE. Human organs-on-chips for disease modelling, drug development and personalized medicine. Nat Rev Genet. 2022;23(8):467–491. doi: 10.1038/s41576-022-00466-9
40. Dip SA, Mallick D, Acharjee Shuvo U, et al. Large language model agents for biological intelligence across genomics, proteomics, spatial biology, and biomedicine. Brief Bioinform. 2026;27(2):bbag110. doi: 10.1093/bib/bbag110
41. Mathis MW. Leveraging insights from neuroscience to build adaptive artificial intelligence. Nat Neurosci. 2026;29(1):13–24. doi: 10.1038/s41593-025-02169-w
42. Wong MYH, Ong JCL, Shah N, et al. Neuro-symbolic artificial intelligence in medicine. Nat Biomed Eng. 2026;10(8):1507–1512. doi: 10.1038/s41551-026-01728-1
43. Makarov N, Bordukova M, von Voithenberg LV, et al. TwinWeaver: an LLM-based foundation model framework for pan-cancer digital twins. Proc Mach Learn Res. 2026;306:85853–85893. doi: 10.48550/arXiv.2601.20906
44. Armingol E, Baghdassarian H, Lewis N. The diversification of methods for studying cell–cell interactions and communication. Nat Rev Genet. 2024;25(6):381–400. doi: 10.1038/s41576-023-00685-8
45. Reynolds G, Vegh P, Fletcher J, et al. Developmental cell programs are co-opted in inflammatory skin disease. Science. 2021;371(6527):eaba6500. doi: 10.1126/science.aba6500
46. Ma F, Plazyo O, Billi AC, et al. Single cell and spatial sequencing define processes by which keratinocytes and fibroblasts amplify inflammatory responses in psoriasis. Nat Commun. 2023;14(1):3455. doi: 10.1038/s41467-023-39020-4
47. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306–313. doi: 10.1038/nature10762
48. Janesick A, Shelansky R, Gottscho AD, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. Nat Commun. 2023;14(1):8353. doi: 10.1038/s41467-023-43458-x
49. Efremova M, Vento-Tormo M, Teichmann SA, et al. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc. 2020;15(4):1484–1506. doi: 10.1038/s41596-020-0292-x
50. Zohora FT, Paliwal D, Flores-Figueroa E, et al. CellNEST reveals cell–cell relay networks using attention mechanisms on spatial transcriptomics. Nat Methods. 2025;22(7):1505–1519. doi: 10.1038/s41592-025-02721-3
51. Sun E, Buendia A, Brunet A, et al. SpatialProp: tissue perturbation modeling with spatially resolved single-cell transcriptomics [preprint]. bioRxiv. 2025 [cited 2026 Nov 5]. Available from: https://pubmed.ncbi.nlm.nih.gov/41573962/
52. Baysoy A, Tian X, Renauer P, et al. Large-scale, spatially resolved panoramic CRISPR screening in native tissue environments using Perturb-DBiT. Nat Biotechnol. 2026. Available from: https://www.nature.com/articles/s41587-026-03127-y
53. Saunders RA, Allen WE, Pan X, et al. Perturb-multimodal: a platform for pooled genetic screens with imaging and sequencing in intact mammalian tissue. Cell. 2025;188(17):4790–4809.e22. doi: 10.1016/j.cell.2025.05.022
54. Binan L, Jiang A, Danquah SA, et al. Simultaneous CRISPR screening and spatial transcriptomics reveal intracellular, intercellular, and functional transcriptional circuits. Cell. 2025;188(8):2141–2158.e18. doi: 10.1016/j.cell.2025.02.012
- **Combines CRISPR screening with spatial transcriptomics to resolve intracellular and intercellular effects, supporting experimental evaluation of interaction-aware cellular models.**
55. Dimitrov D, Türei D, Garrido-Rodriguez M, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-seq data. Nat Commun. 2022;13(1):3224. doi: 10.1038/s41467-022-30755-0
56. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17(2):159–162. doi: 10.1038/s41592-019-0667-5
57. Li H, Zhang Z, Squires M, et al. scMultisim: simulation of single cell multi-omics and spatial data guided by gene regulatory networks and cell-cell interactions. Nat Methods. 2025;22(5):982–993. doi: 10.1038/s41592-025-02651-0
58. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381–386. doi: 10.1038/nbt.2859
59. Kilpinen H, Goncalves A, Leha A, et al. Common genetic variation drives molecular heterogeneity in human iPscs. Nature. 2017;546(7658):370–375. doi: 10.1038/nature22403
60. Collier JR, Monk NAM, Maini PK, et al. Pattern formation by lateral inhibition with feedback: a mathematical model of Delta-Notch intercellular signalling. J Theor Biol. 1996;183(4):429–446. doi: 10.1006/jtbi.1996.0233
61. Akbarialiabad H, Pasdar A, Murrell DF. Digital twins in dermatology, current status, and the road ahead. Npj Digit Med. 2024;7(1):228. doi: 10.1038/s41746-024-01220-7
62. Weerasinghe HN, Burrage PM, Nicolau DV Jr, et al. Agent-based modeling for the tumor microenvironment (TME). Math Biosci Eng. 2024;21(11):7621–7647.
63. Marsilio L, Barrios S, Maksimovic S, et al. In silico digital twins of bone metastasis enable investigation of tumor progression and therapy response. Cancer Res. 2025;85(21):4269–4284. doi: 10.1158/0008-5472.CAN-25-0088
64. Adamala KP, Dogterom M, Elani Y, et al. Present and future of synthetic cell development. Nat Rev Mol Cell Biol. 2024;25(3):162–167. doi: 10.1038/s41580-023-00686-9
65. La Manno G, Soldatov R, Zeisel A, et al. RNA velocity of single cells. Nature. 2018;560(7719):494–498. doi: 10.1038/s41586-018-0414-6
66. Squair JW, Gautier M, Kathe C, et al. Confronting false discoveries in single-cell differential expression. Nat Commun. 2021;12(1):5692. doi: 10.1038/s41467-021-25960-2
67. Kedzierska KZ, Crawford L, Amini AP, et al. Zero-shot evaluation reveals limitations of single-cell foundation models. Genome Biol. 2025;26(1):101. doi: 10.1186/s13059-025-03574-x
68. Luecken MD, Büttner M, Chaichoompu K, et al. Benchmarking atlas-level data integration in single-cell genomics. Nat Methods. 2022;19(1):41–50. doi: 10.1038/s41592-021-01336-8
69. Viñas Torné R, Wiatrak M, Piran Z, et al. Systema: a framework for evaluating genetic perturbation response prediction beyond systematic variation. Nat Biotechnol. 2026;44(6):1050–1059. doi: 10.1038/s41587-025-02777-8
- •• **Introduces an evaluation framework that accounts for systematic variation in perturbation datasets, helping distinguish predictive performance from confounding effects.**
70. Ahlmann-Eltze C, Huber W, Anders S. Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. Nat Methods. 2025;22(8):1657–1661. doi: 10.1038/s41592-025-02772-6
- **Shows that evaluated deep-learning models do not consistently outperform simple linear baselines in genetic perturbation prediction, underscoring the need for rigorous benchmarking.**
71. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc. 2007;102(477):359–378. doi: 10.1198/016214506000001437