From equations to agents: The artificial intelligence virtual cell reshaping precision oncology

Ruochen Wei,[1,7] Boquan Wang,[1,7] Bin Yan,[1] Yuxi Yang,[1] Wanyi Liu,[1] Jia Ding,[1] Dehua Hu,[1] Wei Zhang,[1] Shengjie Sun,[1] Yun Peng,[2] Peng Luo,[3] Yanru Cui,[4] Xue-Song Liu,[5] Jian-Guo Zhou,[6,*] and Shixiang Wang[1,*]

[1]Department of Biomedical Informatics, School of Life Sciences, Central South University, Changsha 410013, China
[2]Department of Urology, Peking University People's Hospital, Beijing 100044, China
[3]Department of Oncology, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, China
[4]Meinig School of Biomedical Engineering, Cornell University, Ithaca NY 14853, USA
[5]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
[6]Department of Oncology, The Second Affiliated Hospital of Zunyi Medical University, Zunyi 550002, China
[7]These authors contributed equally
*Correspondence: jianguo.zhou@zmu.edu.cn (J.-G.Z.); wangshx@csu.edu.cn (S.W.)
Received: December 19, 2025; Accepted: March 16, 2026; Published Online: March 27, 2026; https://doi.org/10.59717/j.xinn-oncol.2026.100002

Citation: Wei R., Wang B., Yan B., et al. (2026). From equations to agents: The artificial intelligence virtual cell reshaping precision oncology. *The Innovation Oncology* **1**:100002.

The Artificial Intelligence Virtual Cell (AIVC) is an advanced computational tool that combines detailed cellular data from single-cell and spatial technologies, including gene expression, epigenetic changes, and cell characteristics, while incorporating fundamental physical and biological rules. This approach creates accurate digital twins of cells that can simulate how tumors develop, evolve, and change over time across multiple scales, from molecules to tissues. Unlike traditional models based on fixed equations that struggle with cancer's extreme complexity and diversity, AIVC learns patterns directly from large biological datasets using modern AI methods. It models cells as intelligent agents operating in a mathematical space and refines predictions through ongoing feedback. This enables patient-specific computer simulations of drug effects, genetic alterations, and tumor environment interactions, moving precision oncology toward truly personalized and dynamic treatment strategies. Despite challenges like incomplete datasets, validation needs, and high computing demands, AIVC is maturing into a valuable research partner and advancing toward clinical use, positioned to transform cancer care.

INTRODUCTION

In the life sciences, the "virtual cell" concept emerged as a landmark effort to comprehend the essence of life through computational simulation.[1] At the turn of the century, the core ambition was to create a general-purpose platform for modeling cellular processes and testing hypotheses *in silico*. Early efforts relied on modular differential equation-based models of molecular reaction networks. However, when addressing highly complex diseases such as cancer, these traditional mathematical approaches harbor inherent limitations. Cancer involves multi-scale dysregulation—from genomic and epigenomic alterations to disrupted signaling pathways and tumor microenvironments. With millions of possible mutation combinations driving profound intratumoral heterogeneity, reductionist models dependent on fixed equations and predefined parameters struggle to link molecular genotypes to emergent clinical phenotypes.[2,3]

Today, single-cell and spatial omics technologies, together with high-throughput perturbation screening techniques exemplified by Perturb-seq, deliver unprecedented high-resolution maps of cellular states and their responses to interventions. Meanwhile, artificial intelligence (AI) provides scalable tools to extract meaningful patterns from these vast and complex datasets.[4] Their convergence has given rise to a transformative paradigm: the Artificial Intelligence Virtual Cell (AIVC). Far beyond conventional simulation, AIVC represents a disruptive platform that functions as a predictive digital twin of living cells. As Stephen Quake has envisioned, AI could invert the balance of cell biology "from being 90% experimental and 10% computational to the other way around." Powered by rapidly expanding, cost-effective single-cell atlases, AIVC enables rapid generation of actionable insights, such as tumor-cell drug response profiles, clonal evolution trajectories under therapy, or systemic effects of oncogenic perturbations that traditionally demand months of wet-lab experimentation.[5] In contrast to conventional AI methods that primarily perform pattern recognition, AIVC enables a mechanistic "hypothesis-simulation" loop. For instance, simulating *EGFR* inhibitor plus *MET* inhibitor perturbation to deduce potential bypass resistance pathways and generate novel candidate targets, moving beyond mere correlation predictions. In precision oncology, AIVC thus evolves from a descriptive tool into an agentive system capable of data-driven learning, multi-modal reasoning, and reliable prediction of unseen biological behaviors.

CORE ARCHITECTURE OF AIVC

Traditional equation-based modeling (Figure 1A) is limited by spatial discretization, solving complex partial differential equations, and requiring precise molecular parameters difficult to measure at scale in clinical tumor samples.[1] In contrast, the core innovation of AIVC lies in bypassing these barriers through multimodal data integration via encoders and large-scale pretraining (Figure 1B), mapping high-dimensional omics data to a compact latent space where cellular states form interpretable vectors for clustering similar cells and separating distinct ones.[4] This enables rapid patient-specific "virtual tumor" models from biopsy-derived single-cell data,[3] supporting *in silico* screening of personalized drug combinations and therapeutic response predictions.[2]

Far from opaque black boxes, modern AIVC thoughtfully incorporate prior biological knowledge. Classical principles, such as reaction-diffusion dynamics and biomechanical forces, are embedded as inductive biases in neural architectures to ensure outputs align with established physics and biology.[5] Simultaneously, AIVC embraces an agent-based paradigm: virtual cells act as autonomous agents navigating latent representations, with an AI module governing fate decisions (e.g., proliferation, apoptosis, migration) informed by data-learned molecular states and constrained by microenvironmental cues (e.g., cell crowding, extracellular matrix stiffness). Continuous refinement occurs via closed-loop learning,[6] enabling robust tracking of rare subclones and seamless cross-scale bridging, from molecular perturbations to altered cellular behaviors, population dynamics, and emergent tumor morphologies that recapitulate genotype-phenotype links.[5]

Generative capabilities further equip AIVC to probe uncharted biology, culminating in patient-specific digital twins for *in silico* "virtual clinical trials" (Figure 1C). These multiscale frameworks from single virtual cells to organoids facilitate comprehensive testing of unseen drugs, genetic edits, or combination therapies,[6] predicting efficacy and resistance mechanisms preemptively.[2,3] For example, the scDrugMap benchmark systematically evaluated multiple single-cell foundation models on over 495,000 cells from diverse perturbation datasets, demonstrating superior performance in predicting cancer cell drug responses to unseen compounds compared to traditional methods.[7] Similarly, models like CancerFoundation have shown enhanced accuracy in forecasting drug sensitivity profiles directly from single-cell RNA-seq data in tumor samples, enabling rapid *in silico* screening of personalized regimens.[7] This powerful exploratory capacity for probing hypothetical scenarios is increasingly attainable through recent methodological advances. For instance, the cell behavior hypothesis grammar—a human-interpretable framework recently introduced in *Cell*—enables researchers to construct multicellular models using near-natural-language rules (e.g., defining interactions between cancer cells and fibroblasts in pancreatic ductal adenocarcinoma to simulate hypoxia-driven migration, tumor invasion, and immunotherapy responses).[8] By bridging biological hypotheses with

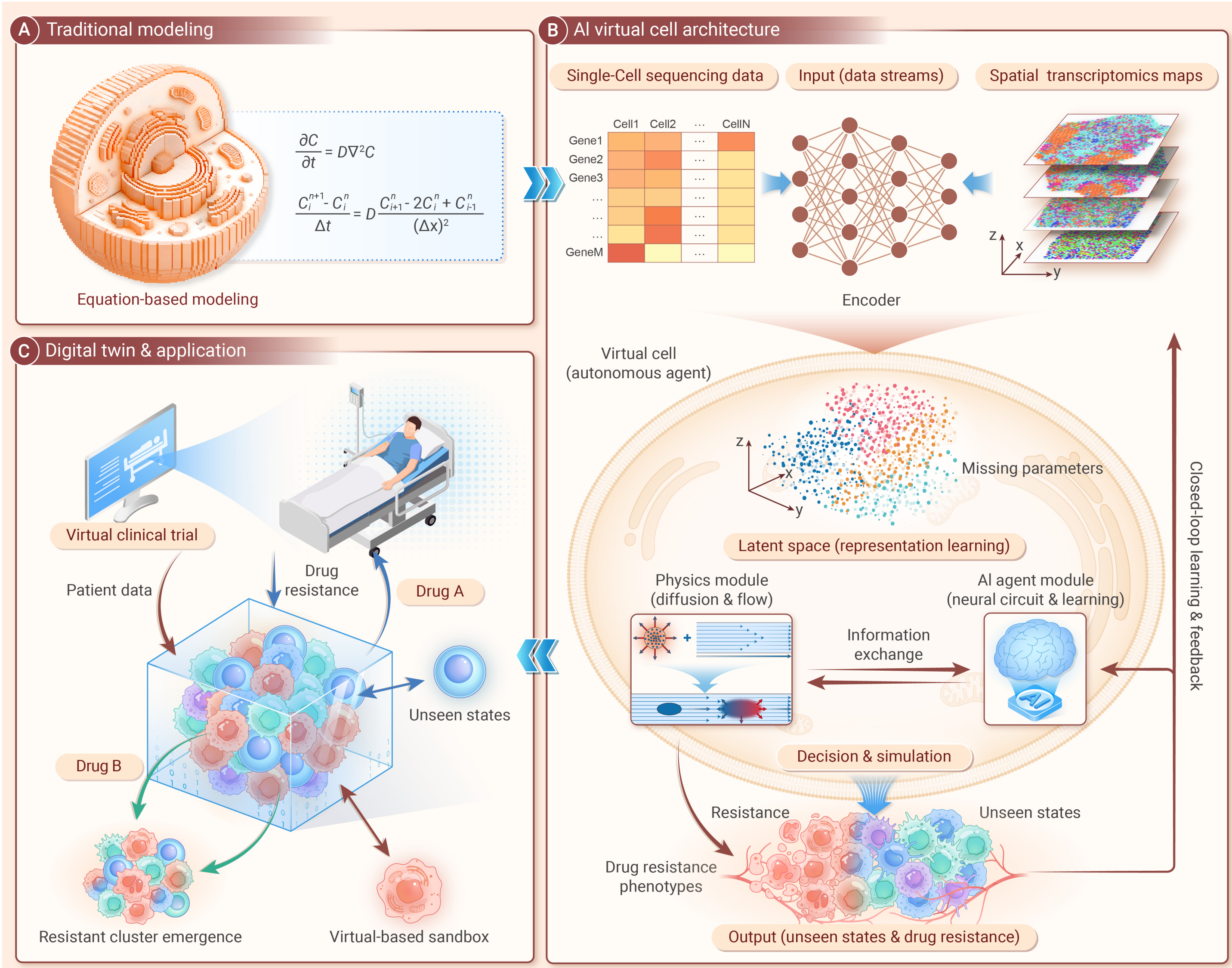

Figure 1. From equation-based models to AI-driven virtual vells: evolution toward digital twins in precision oncology (A) Traditional equation-based modeling. Uses partial differential equations and finite difference approximations to model spatiotemporal cellular dynamics. (B) Artificial intelligence virtual cell (AIVC) Architecture. Integrates single-cell and spatial omics data via encoders to project high-dimensional data into a latent space. Autonomous virtual agents here combine a physics module (for biophysical constraints) and an AI agent module (for neural circuit learning), enabling closed-loop prediction of drug resistance and unseen cellular states. (C) Patient-specific digital twins and clinical translation. Builds virtual tumors from patient data in a sandbox for *in silico* trials. Simulations compare therapeutic responses—such as resistance under Drug A versus tumor shrinkage and resistant cluster emergence under Drug B—to predict clinical outcomes and guide personalized treatment.

executable agent-based simulations in platforms,[9] this approach significantly lowers technical barriers to modeling complex tumor microenvironments and immune dynamics. As a result, high-fidelity virtual tumor models can now be built rapidly, facilitating *in silico* screening of personalized therapeutic regimens and strengthening the bridge from foundational discovery to clinical translation.

AIVC AS AN AUTONOMOUS RESEARCH PARTNER

The role of AI in virtual cell modeling is shifting from a passive tool to an active collaborator. In 2025, the AIVC landscape has witnessed explosive growth in both technological advances and resource availability: large-scale datasets such as scBaseCount, Tahoe-100M, BaseData, and X-Atlas/Orion have been successively released; models including STATE,[9] rBio, and Tahoe-x1 have been introduced; and foundational models for single-cell transcriptomics[4] and related modalities continue to emerge. Collectively, these developments are persistently driving the field toward a new level of maturity and capability. Frontier efforts now focus on integrating diverse foundation models into automated closed-loop platforms, coupling *in silico* predictions with wet-lab validation to create AI-powered scientific agents.[6] By assessing clonal fitness, efficacy, and off-target risks within tumor microenvironments, these agents expedite safer, potent immunotherapy design. Recent applications include AI-optimized CAR-T (Chimeric Antigen Receptor T-cell) constructs with enhanced specificity and potency via protein design and resistance prediction,[10] and hybrid frameworks combining foundation models with agent-based simulations to reveal tumor clonal evolution and emergent resistance trajectories under immunotherapy.

Interpretability remains paramount for life-critical oncology applications. A critical frontier involves developing structured formalisms, analogous to interpretability grammars. These formalisms translate AI decision processes into human-comprehensible biological terms, such as pathway activation scores, cell-cycle phases, or causal regulatory motifs.[3] Incorporating structured knowledge (e.g., via knowledge graphs or inductive biases) not only constrains models for biological fidelity but also aligns explanations with clinical reasoning, fostering the trust essential for translational adoption.[2]

Looking forward, these autonomous research assistants will optimize experimental trajectories and generate testable hypotheses, offloading exhaustive trial-and-error so scientists can prioritize creative innovation. In precision oncology, integrating single-cell multi-omics, mechanistic priors, and agent-based simulations into AIVC platforms promises patient-specific *in silico* evaluation of T-cell therapies (e.g., TCR/CAR [T-cell receptor/chimeric antigen receptor] constructs). Ultimately, this materializes a paradigm shift of AI from a mere tool to a collaborative partner.

CHALLENGES AND FUTURE DIRECTIONS

The transition of AIVC from a computational proof-of-concept to a clinical decision-support system faces distinct hurdles.[5,6] To clarify translational paths, distinguishing near-term technical challenges from long-term fundamental barriers is essential.

The most immediate bottleneck for clinical deployment is the input modality mismatch. While current AIVCs are trained on high-dimensional single-cell multi-omics data, routine clinical diagnostics rely heavily on hematoxylin and eosin (H&E) histology, immunohistochemistry, or bulk sequencing.[5] However, this is an addressable technical challenge. Emerging cross-modal AI frameworks are rapidly advancing in their ability to map accessible clinical data (e.g., pathology images) into the latent spaces of single-cell models. We anticipate that within the next few years, such "modality-bridging" encoders will enable AIVCs to construct high-fidelity virtual tumors directly from standard biopsy data, effectively mitigating the strict requirement for clinical single-cell sequencing. Similarly, while simulating billions of cells imposes high computational demands, ongoing breakthroughs in lightweight model architectures and specialized hardware accelerators suggest that the computational efficiency barrier will be overcome through engineering optimization rather than requiring fundamental scientific discovery.[5]

Specifically, high-fidelity agent-based simulation is implemented only in critical spatiotemporal regions where key biological processes, such as the tumor invasion front or the emergence of drug-resistant clones, occur. Simplified rules are employed in other stable regions. In this way, computational resources can be allocated dynamically to significantly reduce the overall burden. Concurrently, leveraging GPU/TPU (Graphics Processing Unit/Tensor Processing Unit) clusters for large-scale parallel simulation and adopting distributed computing frameworks to manage ultra-large-scale virtual cell populations represent feasible engineering pathways to realize such simulations.

In contrast, a more profound and persistent obstacle is the cross-scale extrapolation from cellular dynamics to patient-level outcomes. Successfully predicting cellular responses (e.g., drug-induced apoptosis) *in silico* does not automatically translate to predicting clinical endpoints like progression-free survival or systemic toxicity.[5] This discrepancy arises because patient outcomes are governed by systemic complexities, such as pharmacokinetics, whole-body immune surveillance, neuro-endocrine regulation, and vascular transport—that extend far beyond the local microenvironment currently modeled by AIVCs. Bridging this "cell-to-patient" gap represents a long-term fundamental challenge, requiring the evolution of AIVC from isolated tumor simulations to integrated systems that dynamically interact with comprehensive host physiological models, demanding deep fusion of systems biology and artificial intelligence beyond mere scaling of data or compute. An equally enduring barrier is the data silo effect: high-quality clinical and multi-omics datasets are fragmented across institutions due to privacy, competition, and infrastructural issues, restricting training data diversity, perpetuating biases, and necessitating global data-sharing solutions.

To surmount these barriers, rigorous and multidimensional validation is paramount. Current benchmarks, such as the "Virtual Cell Challenge", primarily focus on transcriptomic predictions.[9] However, clinical drug resistance often stems from post-translational modifications, protein relocalization, or metabolite shifts that RNA-seq misses. Future AIVCs must incorporate dynamic proteomics, phosphoproteomics, and other modalities to achieve mechanistic fidelity.[5] Furthermore, to build trust for clinical use, the field must advance toward closed-loop validation frameworks that integrate AIVC-driven hypothesis generation with rapid experimental feedback from patient-derived organoids, creating verifiable cycles essential for navigating human cancer's complexity.[6] For data silos specifically, emerging approaches like federated learning and policy-driven international data trusts offer promising avenues to enable collaborative model training without direct raw data transfer, gradually eroding these systemic barriers.

CONCLUSION

The emergence of the AIVC heralds a profound paradigm shift in oncology. This shift moves from traditional reductionist dissection of biological components to a generative, systems-level paradigm. The new paradigm integrates multiscale knowledge and dynamically grows holistic understandings of tumor ecosystems *in silico*. Despite ongoing challenges in data comprehensiveness, rigorous validation, computational scalability, and seamless translation to patient-level outcomes, AIVC is rapidly maturing into the core computational engine of precision oncology. As a transformative future direction, AIVC promises true patient-specific digital twins. These twins can simulate therapeutic responses, resistance trajectories, and microenvironmental dynamics prior to clinical intervention. This capability enables genuinely dynamic, personalized, and optimal decision-making. Realizing this vision, however, demands sustained efforts to bridge virtual experiments with real-world applications, fuse AIVC with emerging technologies such as patient-derived organoids, and drive theoretical innovations that redefine how we conceptualize and intervene in cancer as a complex, evolving system. Ultimately, AIVC not only opens powerful avenues for conquering cancer but also illuminates fundamental principles of life itself.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used Grok/Kimi in order to polish the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

1. Loew L.M. and Schaff J.C. (2001). The virtual cell: A software environment for computational cell biology. *Trends Biotechnol.* 19:401–406. DOI:10.1016/S0167-7799(01)01740-1
2. Kann B.H., Hosny A. and Aerts H.J.W.L. (2021). Artificial intelligence for clinical oncology. *Cancer Cell* 39:916–927. DOI:10.1016/j.ccell.2021.04.002
3. Yang T., Wang Y.-Y., Ma F., et al. (2025). Build the virtual cell with artificial intelligence: A perspective for cancer research. *Mil. Med. Res.* 12:4. DOI:10.1186/s40779-025-00591-6
4. Hao M., Gong J., Zeng X., et al. (2024). Large-scale foundation model on single-cell transcriptomics. *Nat. Methods* 21:1481–1491. DOI:10.1038/s41592-024-02305-7
5. Bunne C., Roohani Y., Rosen Y., et al. (2024). How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell* 187:7045–7063. DOI:10.1016/j.cell.2024.11.015
6. Qian L., Dong Z. and Guo T. (2025). Grow AI virtual cells: Three data pillars and closed-loop learning. *Cell Res.* 35:319–321. DOI:10.1038/s41422-025-01101-y
7. Wang Q., Pan Y., Zhou M., et al. (2025). scDrugMap: Benchmarking large foundation models for drug response prediction. *Nat. Commun.* 17:730. DOI:10.1038/s41467-025-67481-2
8. Johnson J.A.I., Bergman D.R., Rocha H.L., et al. (2025). Human interpretable grammar encodes multicellular systems biology models to democratize virtual cell laboratories. *Cell* 188:4711–4733. DOI:10.1016/j.cell.2025.06.048
9. Adduri A.K., Gautam D., Bevilacqua B., et al. (2025). Predicting cellular responses to perturbation across diverse contexts with State. Preprint at *bioRxiv*. DOI:10.1101/2025.06.26.661135
10. Luciani F., Safavi A., Guruprasad P., et al. (2025). Advancing CAR T-cell therapies with artificial intelligence: Opportunities and challenges. *Blood Cancer Discov.* 6:159–162. DOI:10.1158/2643-3230.BCD-23-0240

FUNDING AND ACKNOWLEDGMENTS

We thank Prof. Xingxu Huang (The Key Laboratory of Pancreatic Diseases of Zhejiang Province, The First Affiliated Hospital, Zhejiang University School of Medicine) for his insightful and critical comments, which substantially strengthened this manuscript. We also thank Herbert Fountain (Department of Biological and Biomedical Sciences, Cornell University) for proofreading and editing the English in this manuscript. This work was supported by the National Natural Science Foundation of China (Grant No. 82303953, 82060475, 82504050), the Hunan Provincial Natural Science Foundation of China (Grant No. 2025JJ40079), Noncommunicable Chronic Diseases-National Science and Technology Major Project (Grant No. 2023ZD0502105), Guizhou Province College Students Innovation and Entrepreneurship Training Program (Grant No. ZHCX2023005), the Central South University Startup Funding, MOE (Ministry of Education in China) Liberal arts and Social Sciences Foundation (Grant No. 24YJCZH462), Youth Science and Tech-nology Elite Talent Project of Guizhou Provincial Department of Education (Grant No.QJJ-2024-333), Zunyi City Science and Technology Plan Project (Grant No. Zunshi Kehe HZ (2023) 142, Zunshi Kehe HZ (2025) 256), Zunyi City Science and Technology Innovation Team (Grant No. Zun KCTD (2025) 63), Future Science and Technology Elite Talent Cul-tivation Project of Zunyi Medical University (ZYSE 2023-02), Guizhou Province High-Level Overseas-Educated Talents Innovation and Entrepreneurship Selective Funding Program (Grant No.3) and the Key Program of the Education Sciences Planning of Guizhou Province (Grant No.2024A007). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.