

Review Article

Leveraging AI for cell biology discovery

Adriana Simizo¹, Mauro de Moraes^{2,3}, Matheus Vesco¹ and Helder Nakaya^{1,2,3}

¹Hospital Israelita Albert Einstein, São Paulo, SP, Brazil; ²Department of Clinical and Toxicological Analyses, University of São Paulo, São Paulo, SP, Brazil; ³Institut Pasteur de São Paulo, São Paulo, SP, Brazil

Correspondence: Helder Nakaya (helder.nakaya@einstein.br)



Artificial intelligence (AI) has become a transformative tool in cell biology, driving discoveries through the analysis of complex biological data. This review explores the diverse applications of AI, including its impact on microscopy, imaging, drug discovery, and synthetic biology. AI methods have significantly advanced our ability to analyze cellular images at single-cell resolution, uncover complex patterns in biological data, and predict cellular responses to various stimuli. Deep learning approaches have improved cell segmentation and tracking, facilitated precise single-cell transcriptomics analysis, and enhanced our understanding of protein structures and interactions. The application of AI to high-throughput technologies has also enabled detailed modeling of cell behavior. Key challenges are addressed, such as data quality requirements, model interpretability, and the need to democratize AI tools for broader accessibility in biology. Finally, the review considers future directions, highlighting AI's potential to advance basic research and therapeutic applications.

Introduction

Cellular systems represent one of the most complex and dynamic entities in nature, comprising intricate networks of molecular interactions that operate across multiple spatial and temporal scales. At any given moment, a single cell orchestrates thousands of biochemical reactions, maintains precise protein levels, coordinates organelle functions, and responds to environmental cues through elaborate signaling cascades. The complexity emerges not only from the vast number of components – including proteins, nucleic acids, metabolites, and lipids – but also from their nonlinear interactions and the emergent properties that arise from these relationships.

Understanding cellular behavior presents formidable challenges due to the sheer volume of molecular players involved, the speed at which cellular processes occur, and the remarkable adaptability of cells to changing conditions. Traditional experimental approaches, while valuable, often struggle to capture the full scope of these dynamic interactions, particularly when attempting to observe multiple cellular processes simultaneously or track rapid molecular events in real time [1,2]. This inherent complexity has historically limited our ability to fully comprehend cellular systems and accurately predict their behavior under various conditions.

Artificial intelligence (AI) has emerged as a transformative force in cell biology research, advancing the analysis of complex biological data and driving innovation. This transformation is particularly evident in deep learning algorithms, which enable powerful analysis of cellular images at single-cell resolution, allowing researchers to extract detailed information about cell morphology and molecular processes [3,4]. Beyond imaging analysis, advanced AI approaches are rapidly transforming both data acquisition and analysis in biological research, with modern AI tools being particularly effective at uncovering mechanistic relationships across multiple layers of cellular biology [5–9]. These advanced modeling capabilities extend to digital twins of human cells, where AI can predict cellular responses to various stimuli, streamlining drug discovery and enabling the testing of billions of molecular combinations *in silico* [10,11]. Also, the emergence of contextual AI models has enhanced our ability to understand protein function across different cell types and tissues, generating hundreds of thousands of context-aware protein representations that reflect cellular and tissue organization [12]. In cancer research, AI models are solving key challenges such as data integration with high intra-individual variability, quantifying cellular plasticity, and identifying causal networks in oncogenesis [13–16].

This review focuses on the rapidly evolving landscape of AI applications in cell biology, ranging from microscopy to synthetic biology. The scope encompasses key areas of impact: imaging analysis, single-cell transcriptomics, protein structure prediction, and drug discovery. Figure 1 provides a visual summary of these key applications, highlighting how AI methodologies such as computer vision, graph neural networks, and generative AI are transforming research practices across these areas. In these domains, AI approaches can reveal biological interactions across molecular and cellular scales, from genomic regulation

Received: 24 January 2025
Revised: 22 October 2025
Accepted: 19 December 2025

Version of Record
Published: 8 January 2026

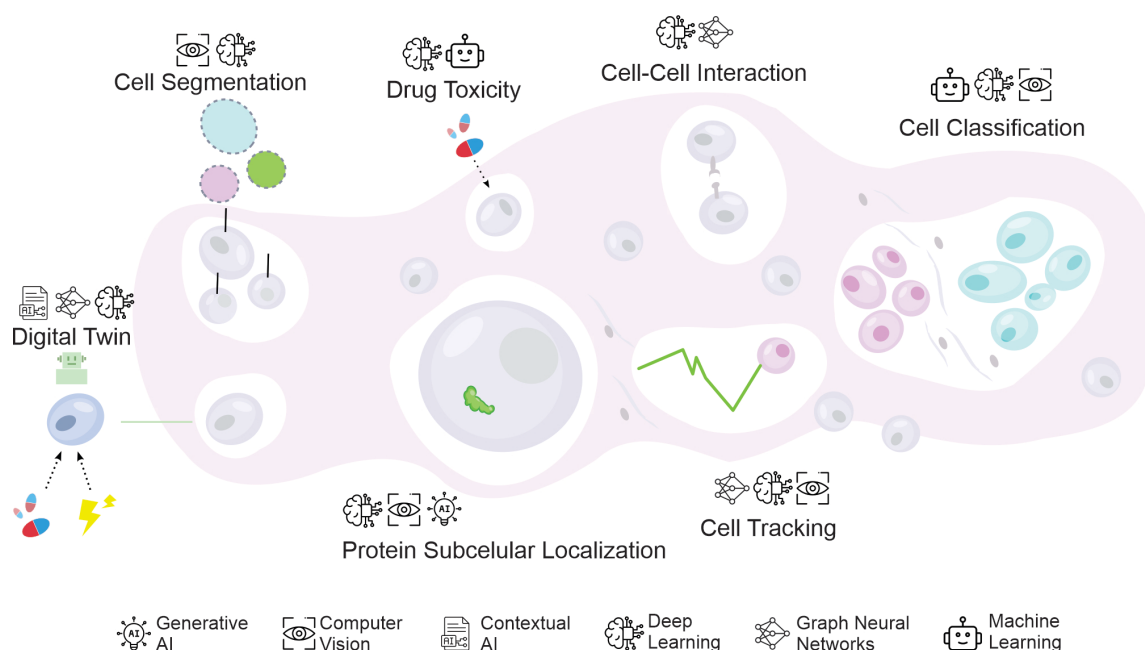


Figure 1: Overview of key artificial intelligence (AI) applications in cell biology.

AI has significantly advanced seven major areas in cell biology: (1) cell segmentation, using deep learning for precise identification of individual cells; (2) drug toxicity prediction, through computational modeling of molecular and cellular responses; (3) cell–cell interaction analysis, employing network-based approaches to study cellular communication; (4) digital twin development, enabling *in silico* modeling of cellular systems; (5) protein subcellular localization prediction, using AI to determine protein distribution within cells; (6) cell tracking, for monitoring cellular dynamics over time; and (7) cell classification, providing automated identification of cellular phenotypes. The icons at the bottom represent the foundational AI methodologies driving these advancements.

to protein function [17–19]. Recent advances include foundation models, improved imaging analysis tools, and novel computational platforms that enhance our understanding of cellular systems [20–23] (Table 1). These developments are essential for addressing the complexity of cellular systems, co-ordinating thousands of molecular interactions [43–45]. AI working alongside high-throughput technologies has advanced the modeling of cell behavior, driving progress in basic research and therapeutic applications [46]. While most reviews focus on specific AI applications, this work integrates advances across imaging, omics, protein structure prediction, and drug discovery, emphasizing key challenges and translational opportunities.

AI in microscopy and imaging

Deep learning approaches, including specialized architectures like U-Net [47] and Cellpose [28], have significantly improved cell segmentation and tracking. These AI models can effectively distinguish individual cells even in densely packed tissues and challenging imaging conditions, achieving accuracy rates above 90% [48]. Platforms like TrackMate [34] combine deep learning with segmentation algorithms, such as StarDist [33] and Cellpose [28], for robust tracking in transmitted light imaging. This combination has proven particularly valuable for long-term time-lapse imaging where phototoxicity must be minimized [49].

Beyond general cell and nucleus tracking, recent advances have enabled precise subcellular tracking and phenotyping. For example, deep learning-based frameworks have been applied to extract mitochondrial morphological and functional phenotypes from high-content images, providing insights into metabolic states and drug responses [50]. Similarly, AI approaches for spindle phenotype analysis allow detailed characterization of mitotic processes and structural variations at subcellular resolution [51]. These specialized trackers expand the scope of AI-powered imaging far beyond whole-cell analysis.

Automated phenotypic analysis, powered by machine learning models, classifies cell states and behaviors with high precision. Deep learning algorithms can identify complex morphological patterns and cellular states without requiring predefined features, allowing for unbiased detection of phenotypic variations [52–55]. These systems can process high-dimensional data from thousands of cells

Table 1.: Summary of artificial intelligence tools and their applications in cell biology.

Tool	Function	Performance metrics	URL
Drug discovery and synthetic biology			
AtomNet [24]	Prediction of molecule–protein binding	Hit rate 5–8%; validation ~90%; 318 targets with 16B compounds	https://www.atomwise.com/
DeepTox [25]	Pipeline for predicting chemical toxicity	AUC 0.86 (12 assays); Tox21 winner; 12.7k compounds	https://www.bioinf.jku.at/research/DeepTox/
MoleculeNet [26]	Molecular ML benchmark	>700k compounds; ROC-AUC >0.90 on Tox21/HIV	https://moleculenet.org/
Hypothesis generation			
FieldSHIFT [27]	Framework for generating biological hypotheses	1.4k domain pairs; GPT-4 success 58% vs 92% human; 74% validated gene overlap	https://huggingface.co/datasets/levinlab/neuroscience-to-dev-bio
Microscopy and imaging			
Cellpose [28]	Model for cell and nucleus segmentation	73% detection accuracy for cells; human-level segmentation with ~500 annotated cells	https://www.cellpose.org/
CellProfiler [29]	Feature extraction and segmentation	>12M cells processed; rand index = 0.919–0.976	https://cellprofiler.org/
Cytoself [30]	Model for protein localization features	Predicts subcellular localization for 88% of proteins; trained on 1.1M fluorescence images	https://github.com/royerlab/cytoself
DeepCell [31]	Nuclear and whole-cell segmentation	F1 = 0.82 (overall); trained on >1.3M cells	https://deepcell.org/
deepGPS [32]	Model for protein localization prediction	Accuracy 0.81; AUROC 0.83 (4-class); 62k images, 1.3k proteins	https://bits.fudan.edu.cn/opengps/home
StarDist [33]	Model for cell and nucleus segmentation	Average precision: 0.86 on real microscopy nuclei; robust in crowded images	https://stardist.net/
TrackMate [34]	Automated tracking	F1 > 0.96 for breast cancer cells and >0.99 for T cells; supports 2D/3D segmentation	https://imagej.net/plugins/trackmate/
U-Net [35]	Model for image segmentation	IoU up to 0.9 (near human-level accuracy) for 2D/3D cell segmentation	https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/
Predicting cellular dynamics and behavior			
NETISCE [36]	Cell fate reprogramming	85% accuracy across validated perturbations; remains >75% accurate under 50% simulated noise	https://github.com/VeraLiconaResearchGroup/Netisce
PERCEPTION [37]	Prediction of patient drug response	Trained on 488 cell lines; predictive for 33% FDA-approved drugs	https://github.com/ruppinlab/PERCEPTION
Transcriptomics and omics integration			
DeepTalk [38]	Model for cell–cell communication	AUC 0.84–0.95 for ligand–receptor pair prediction	https://github.com/JiangBioLab/DeepTalk
MIDAS [39]	Deep learning model for mosaic data integration	Outperforms 19 tools in multimodal integration; >185k cells	https://github.com/labomics/midas
MultiVI [40]	Model for multimodal data integration	Integrates RNA, ATAC, and protein data from >89k cells; Spearman $R = 0.8$ – 0.86 for imputation	https://scvi-tools.org/
scGNN [41]	Model for single-cell clustering	Outperforms 9 imputation and 4 clustering tools across 4 scRNA-seq benchmarks	https://github.com/juexinwang/scGNN
scGPT [6]	Generative AI for single-cell multi-omics	Trained and evaluated on 10.2M cells and 20–30% better temporal continuity than other existing tools	https://github.com/bowang-lab/scGPT
SpatialScope [42]	Spatial and single-cell transcriptomics integration	Integration across >1M spots; higher gene-level correlation ($r \approx 0.91$)	https://github.com/YangLabHKUST/SpatialScope

simultaneously, categorizing them based on morphology, behavior, and other characteristics [54–56]. Modern AI approaches are particularly effective at detecting subtle phenotypic changes that might be missed by traditional analysis methods, achieving up to 99% accuracy in phenotype prediction when combining both cellular and colonial morphological parameters [57].

Deep learning has transformed the mapping of protein subcellular localization through systems like cytoself, which can analyze the distribution of proteins across different cellular compartments without

requiring preexisting annotations [30]. AI models with high precision in identifying organelles and protein locations have enabled detailed maps of cellular organization [7,58–62]. Furthermore, advanced computational approaches such as deepGPS [32] now achieve protein subcellular localization prediction directly from sequences, producing both textual annotations and artificial fluorescence images that quantitatively represent cellular and tissue organization. The use of these methodologies has facilitated the development of comprehensive protein localization atlases, systematically capturing diverse patterns of cellular organization [30].

Quantitative image analysis tools measure cellular features with precision, leveraging advanced automation and computational pipelines. These systems extract parameters such as size, shape, texture, and fluorescence intensity [35,54,63–65]. Modern AI approaches can analyze both bright-field and fluorescence images, providing detailed measurements of cellular structures without the need for invasive staining techniques [66,67]. Deep learning combined with high-throughput imaging allows the extraction of up to 17 distinct morphological features per cell, supporting comprehensive quantitative profiling of cellular populations [68]. These measurements can be performed in real time, enabling dynamic analysis of cellular changes during experiments [69,70]. The same approach can also be applied to detect and count human parasites, such as *Trypanosoma cruzi* from micrographs of blood smear samples, achieving 90.5% sensitivity and 88.6% specificity even when using low-resolution photos taken by smartphone cameras attached to microscopes [71].

Transcriptomics and omics integration

Sophisticated deep learning approaches have transformed single-cell RNA sequencing (scRNA-seq) analysis, allowing precise identification of cell types and states. Advanced frameworks like scGNN [41] utilize graph neural networks to capture complex cell–cell relationships, achieving superior performance in gene imputation and cell clustering across diverse datasets. These models capture heterogeneous gene expression and predict cell trajectories by integrating regulatory signals [41]. Deep learning algorithms, such as scGPT, have improved the analysis of cellular dynamics, offering high performance in annotating cell types from scRNA-seq data [6,72].

The application of AI to spatial transcriptomics has transformed our understanding of tissue architecture and cellular interactions. SpatialScope [42], a unified deep generative model approach, enhances sequencing-based spatial transcriptomics data to achieve single-cell resolution while accurately inferring transcriptome-wide expression levels for image-based data. This technology facilitates the analysis of cellular communication and differentially expressed genes [42]. DeepTalk [38], an innovative method, uses graph attention networks to integrate scRNA-seq and spatial transcriptomics data, allowing accurate cell-type identification and insights into cell–cell communication.

Advanced computational methods have emerged as powerful tools for comprehensive cellular analysis. MIDAS [39], a deep probabilistic framework, achieves simultaneous dimensionality reduction, imputation, and batch correction of mosaic data through self-supervised modality alignment. The framework enables flexible integration of different omics combinations and has demonstrated superior performance compared with 19 other methods [39]. Similarly, MultiVI from the scVI suite [40] provides a probabilistic model for analyzing multi-omic data, synthesizing information from different laboratories and technologies while maintaining biological relevance. These advanced AI approaches enable researchers to construct comprehensive cellular atlases and derive meaningful insights from complex biological systems [73]. A summary of these tools and functionalities is presented in Figure 2.

AI tools for single-cell transcriptomics, spatial transcriptomics, and multi-omics data are organized by their input data type, AI tool, output, and biological insights. For single-cell transcriptomics, scGNN enables gene imputation, clustering, and differential expression gene (DEG) analysis, while scGPT identifies cell types and regulatory networks. Spatial transcriptomics tools, such as SpatialScope and DeepTalk, provide single-cell resolution, ligand–receptor interaction analysis, and spatially resolved effects. Multi-omics integration is supported by tools like MIDAS and MultiVI, which generate imputed data and integrated datasets, enabling the construction of comprehensive cell atlases and cross-modality cell state analysis. The foundational methodologies driving these tools are represented by icons at the bottom.

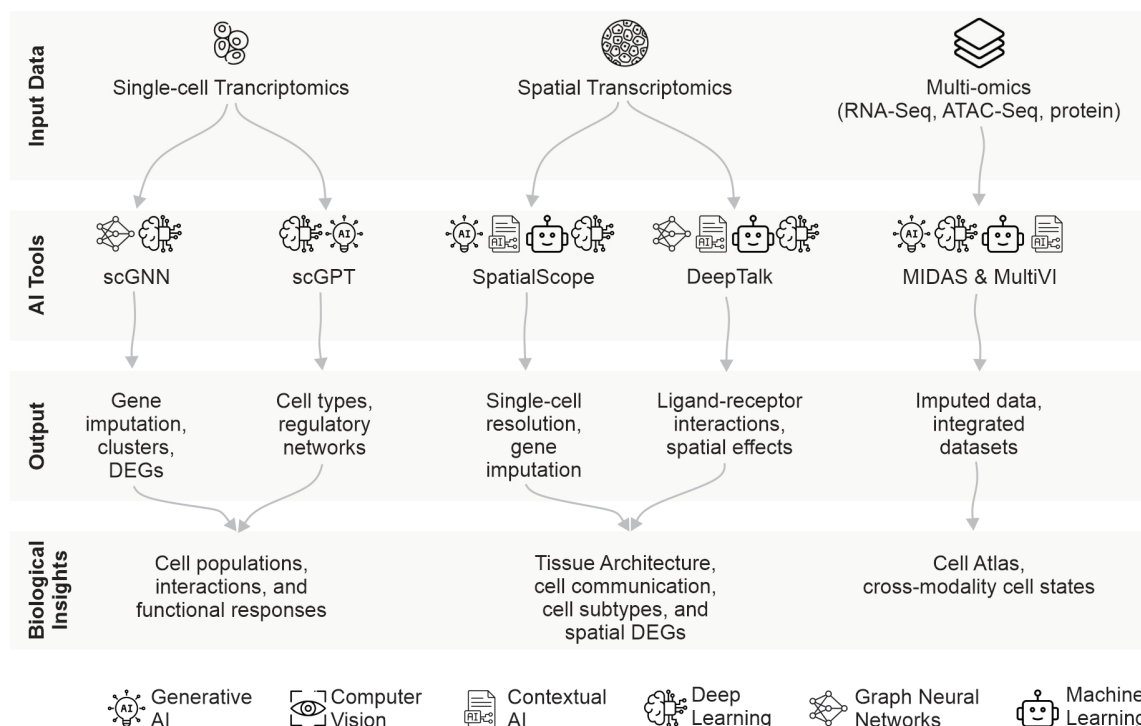


Figure 2: Artificial intelligence (AI) tools for transcriptomics and omics integration.

Predicting cellular dynamics and behavior

The cell cycle and signaling pathways represent intricate networks of molecular interactions operating across multiple scales. Cell signaling encompasses two essential processes: signal conduction for generation and intercellular transmission, and signal transduction for reception and processing of signals [74]. These pathways involve precise molecular events, including signal reception, amplification, distribution, and specific cellular responses that regulate cell behavior and function. Critical cellular determinations, such as cytoskeletal reorganization, cell cycle checkpoints, and programmed cell death, depend on stringent temporal regulation and specific spatial distribution of activated signal transducers [74].

The analysis of cellular signaling networks has significantly benefited from modern AI-based computational frameworks, which use non-parametric simulation techniques to provide accurate predictions without requiring detailed parameter knowledge [75]. Modern computational frameworks can simulate cell cycle progression through G1, S, G2, and M phases while incorporating checkpoint controls and regulatory mechanisms [76]. These systems employ various computational models, including cellular automata, Petri nets, Boolean models, rule-based systems, and artificial neural networks to analyze biological data [75]. During simulations, large volumes of data are generated, encompassing proteomics, genomics, interactomics, and metabolomics information, which serve as valuable input for predictive analytics [75].

Deep learning models can identify and predict cell fate decisions with remarkable accuracy. Using only bright-field images without artificial labeling, these models can identify differentiated cell types as early as 1 day of culture, achieving accuracy rates above 80% for various cell types [77]. AI systems have demonstrated success in predicting treatment responses in cancer, with models accurately forecasting how individual cells respond to both monotherapies and combination therapies [37,78]. The NETISCE computational tool [36] combines machine learning algorithms with signal flow analysis to estimate attractor landscapes and predict reprogramming targets, enabling the identification of cell fate transitions without requiring extensive experimental data.

AI-powered cellular dynamics modeling and fate prediction hold significant potential for advancing personalized medicine. These approaches enable precise matching of patients with effective drug treatments by analyzing scRNA-seq data [78]. AI models can also evaluate drug toxicity, identify specific molecular targets responsible for adverse effects, metabolic pathways, and potential clinical trial failures

[79]. For cancer treatment, AI systems can analyze high-dimensional datasets to identify new dynamic diagnostic and prognostic biomarkers, leading to more accurate cancer care strategies [80]. These tools have proven particularly valuable in understanding tumor heterogeneity and predicting treatment responses across different cancer types [81].

Applications in drug discovery and synthetic biology

Traditional drug discovery and toxicity assessment relied heavily on trial-and-error methods, where researchers tested compounds to identify therapeutic effects through extensive laboratory experiments. Animal studies served as the primary approach for toxicity evaluation, but these methods faced significant limitations due to cost, time, and ethical considerations. High-throughput screening techniques allowed researchers to screen large compound libraries, yet traditional methods often resulted in high failure rates, with 20% of drug candidates failing during clinical trials due to toxicity issues [82].

Various computational approaches have significantly improved drug toxicity prediction, marking a key advancement in pharmacological research. Machine learning algorithms can now evaluate drug toxicity using platforms like DeepTox [25] and MoleculeNet [26], which translate molecular structures into toxicity predictions. Advanced AI models combine chemical structure properties with protein target information to develop comprehensive toxicity scores, achieving high performance metrics (AUC = 0.83) and accuracy (ACC = 0.75) [82]. These systems can process thousands of compounds simultaneously, significantly reducing the time and cost of preclinical testing while maintaining high sensitivity (0.75) and specificity (0.74) [82].

Advancements in synthetic biology and computational tools have improved the design and optimization of cellular circuits for specific functions. These systems can predict the impact of bioengineering modifications on host cells and help create cells with enhanced properties such as increased productivity or stress resistance [83]. AI algorithms assist in dynamic pathway engineering by identifying architectures that optimize production and selecting optimal configurations for synthetic gene circuits [84,85]. These approaches enable the rapid construction and evaluation of engineered pathways, particularly in applications such as biofuel production and environmental remediation [84–86].

The pharmaceutical industry has witnessed significant growth in AI-driven companies. For example, Insilico Medicine has identified novel drug candidates, including a first-in-class fibrosis treatment using deep learning and generative adversarial networks [87]. Atomwise employs structure-based design via the AtomNet platform to predict small molecule binding, accelerating hit identification [24]. Exscientia has developed the first functional precision oncology platform using AI, with several AI-designed compounds currently in clinical development [88,89]. In personalized oncology, frameworks such as PERCEPTION have successfully predicted individual patient responses and resistance mechanisms to targeted therapies based on single-cell transcriptomic data [37]. AI is also being used to model protein–nanoparticle interactions. An example of this application is the prediction of the ‘protein corona’, which significantly affects nanoparticle behavior in biological environments. Accurate prediction of corona composition helps to optimize targeted delivery, minimize off-target effects, and enhance therapeutic efficacy [90].

Hypothesis generation and real-time experimentation

The scientific method traditionally relies on two main approaches: hypothesis-driven and data-driven research. In hypothesis-driven research, scientists make specific predictions based on prior knowledge and design targeted experiments to test these hypotheses. For example, in microbiome studies, researchers might predict the presence of specific bacteria and design PCR primers for detection. In contrast, data-driven approaches collect comprehensive datasets without predetermined hypotheses, such as using metagenomics sequencing to identify all microbes present in a sample [91,92]. Both approaches complement each other, with data-driven methods often generating hypotheses that can be tested through targeted experiments [93,94].

Sophisticated frameworks for literature and data analysis are transforming hypothesis generation. FieldSHIFT [27], an in-context learning framework, uses large language models to generate candidate scientific research directions from existing published studies. These AI tools can identify patterns and relationships across different scientific domains, leading to novel research directions. In oncology, AI-driven algorithms analyzing extensive omics datasets have uncovered intricate details of cancer etiology, revealing insights potentially overlooked by traditional methods [95].

Beyond hypothesis generation, AI systems are now being applied to experimental methodologies, including live-cell imaging. Advances in AI-powered image analysis allow real-time tracking of cellular events, linking theoretical predictions with experimental validation. These methods can detect and quantify cellular dynamics, monitor cell morphology, and analyze complex 3D cultures with high accuracy and speed [54,96]. Deep learning algorithms further address challenges such as image drift and phototoxicity, enhancing spatial and temporal resolution [70]. These systems can track cells in diverse biological contexts, such as moving worm brains and beating zebrafish hearts, with notable precision [97].

In omics research, AI methods have advanced the analysis of complex biological datasets. Deep learning models can effectively handle high-dimensional omics data, integrating multiple data types to identify biomarkers and regulatory mechanisms [98]. Recent studies have employed AI to analyze liver disease by incorporating clinical information with molecular profiles for better diagnosis and treatment strategies [98]. As these models become more integrated into clinical research, the need for transparent and interpretable approaches has grown rapidly. Explainable AI (XAI) techniques have become crucial in omics analysis, with over 400 publications between 2010 and 2023 dedicated to making AI-derived insights more understandable and biologically meaningful [99]. This growing emphasis on interpretability sets the stage for broader applications of XAI in other areas of biology.

Model interpretability remains a central challenge in biological AI applications, as many systems still operate as ‘black boxes’ with limited transparency in their decision-making. This hampers biological validation, bias detection, and clinical translation. To improve explainability, techniques such as SHAP [100] and LIME have been used to prioritize biomarkers and interpret complex gene expression or cytokine-based predictions [101], while Grad-CAM++ and Layer-wise Relevance Propagation (LRP) enhance interpretability in histopathology [102] and multi-cancer prognostic models [103], respectively. Recent omics frameworks embed XAI techniques for automated feature attribution and pathway-level interpretation [104,105], while attention-based and prototype-driven models further promote interpretability by design [106,107].

Promoting deep integration between computational and experimental expertise is equally critical: establishing joint funding initiatives, interdisciplinary workshops, and integrated training programs can foster a common language and mutual understanding between data scientists and biologists. Placing computational researchers within biology labs – and biologists within computational teams – can help ensure that AI-generated hypotheses are not only statistically sound but also biologically plausible and experimentally testable.

Challenges and future perspectives

High-quality datasets and standardized annotations remain critical challenges in computational cell biology [108]. Current research indicates that data quality has a significant impact on model performance, with issues such as batch effects and technical variations potentially leading to unreliable results [109], particularly in single-cell analysis where data heterogeneity can mask biological signals [110]. To address these issues, robust batch correction algorithms, such as scVI, have been developed and shown to improve cross-study integration and cell-type identification accuracy [111]. Additionally, community-driven efforts like the Human Cell Atlas are working toward developing standardized ontologies and annotation frameworks to improve consistency and data interoperability across laboratories [112].

Clinical translation of AI-driven discoveries remains a major hurdle. Regulatory barriers require model transparency and rigorous validation in diverse patient populations, often delaying adoption [46,94]. Beyond technical challenges, ethical and societal issues also deserve attention. Data privacy is critical when using patient-derived datasets, requiring compliance with regulations such as GDPR and HIPAA to maintain trust [113]. Moreover, AI models can perpetuate biases from training data, potentially reinforcing health disparities [114]. Additionally, the potential misuse of AI-generated biological insights underscores the need for strong governance and oversight [115,116].

The democratization of computational tools requires improved accessibility and training resources for researchers across different expertise levels. Platforms like the CZI AI Cell Models aim to make biologically relevant models and datasets easier to find, evaluate, and use [102]. Similarly, open-source tools such as DeepCell [31] and CellProfiler [29] offer accessible, cloud-based workflows and modular pipelines for image analysis. Commercial solutions like ZEISS arivis Cloud [117] further address computational resource challenges associated with analyzing large microscopy datasets.

Despite these advances, computational demands and scalability issues persist, particularly with large or multimodal datasets [118]. Current AI systems remain constrained by compute (high VRAM requirements, long training and inference times, and serving costs), scalability (batch effects, domain shift, and memory/IO bottlenecks across multi-institutional and multimodal data), and biological understanding (predictive features that lack mechanistic grounding, reliance on confounders, and limited ground truth for cell–cell interactions and perturbation responses). Overcoming these barriers will require more efficient architectures, streaming or distributed workflows, and closer integration with experimental validation to bridge prediction and biological insight. Addressing these challenges is also essential to ensure equitable global adoption of AI in biomedical research, as emphasized by recent Latin American initiatives that are building inclusive, AI-ready single-cell and spatial genomics datasets to counter regional and ancestral biases in model training [119,120].

Future directions in cell biology research point toward the development of comprehensive virtual cell models that integrate multiple data types and scales. Advances in spatial and single-cell multiomics and live-cell imaging enable reconstruction of dynamic cellular processes and tissue organization [42,69,121]. Recent developments in generative modeling, including diffusion models and flow matching, have provided state-of-the-art tools for simulating biological processes and temporal dynamics, enabling accurate prediction of cell or molecular state transitions over time [122]. Together, these technologies create a foundation for rapid hypothesis testing and for predicting cellular responses to diverse stimuli [122]. This integrative approach can lead to breakthroughs in biomedical research, personalized medicine, and drug discovery, particularly through the development of more sophisticated cell-scale models that can simulate complex biological processes.

Conclusion

Computational approaches have fundamentally advanced cell biology research by enabling sophisticated analysis of complex biological data. Deep learning has advanced our understanding of cellular processes through precise image analysis, protein modeling, and multi-omics integration, offering comprehensive views of cellular function. These technological advances have contributed to progress in fundamental cell biology, drug development, and disease modeling.

The synergy between computational methods and experimental biology represents a new paradigm in cell research. This approach facilitates hypothesis generation and testing at unprecedented scales, with machine learning models uncovering patterns in complex datasets that guide experimental design and interpretation. The democratization of these tools, coupled with improvements in interpretability and validation methods, promises to make advanced computational approaches accessible to a broader scientific community. As these technologies mature, their application in cell biology promises advances in personalized medicine, drug development, and our understanding of cellular systems, while preserving the essential partnership between computational prediction and experimental validation.

Perspectives

- The incorporation of Artificial Intelligence is transforming cell biology by facilitating sophisticated analyses of complex biological data across areas such as cellular imaging and drug development. This technological progress marks a pivotal change in how cellular systems are studied.
- The current AI landscape in cell biology is marked by significant advances in single-cell resolution image analysis, protein structure prediction, multi-omics data integration, and cell behavior modeling, although challenges such as model interpretability and data quality requirements persist.
- The field is moving towards the development of comprehensive virtual cell models that integrate multiple data types and biological scales, enabling rapid hypothesis testing and prediction of cellular responses, with potential breakthroughs in personalized medicine and drug discovery through more sophisticated cell-scale simulations.

Competing Interests

HIN previously served as the Chief Scientific Officer (CSO) of Synamics Therapeutics, a startup that utilizes artificial intelligence for drug discovery.

Funding

HIN is funded by FAPESP (grant numbers: 2018/14933-2, 2018/21934-5, and 2013/08216-2) and CNPq (313662/2017-7). We are particularly thankful to the following grants for fellows: AS (Hospital Israelita Albert Einstein), MCM (FAPESP Process 2020/12017-9), MFV (Hospital Israelita Albert Einstein).

Open Access

This article has been published open access under our Subscribe to Open programme, made possible through the support of our subscribing institutions, learn more here: https://portlandpress.com/pages/open_access_options_and_prices#conditional

CRediT Author Contribution

Writing - review and editing: AS; MCM; MFV; HIN. Design and creation of figures: AS; HIN.

Acknowledgments

We thank members of the Computational Systems Biology Laboratory for helpful discussions.

Abbreviations

ACC, accuracy; AI, artificial intelligence; AUC, area under the curve; AUROC, area under the receiver operating characteristic curve; CZI, Chan Zuckerberg Initiative; 2D, two-dimensional; 3D, three-dimensional; DEG, differentially expressed gene; DNA, deoxyribonucleic acid; GAN, generative adversarial network; GDPR, General Data Protection Regulation; GPU, graphics processing unit; HIPAA, Health Insurance Portability and Accountability Act; IO, input/output; IoU, intersection over union; LIME, Local Interpretable Model-agnostic Explanations; LRP, Layer-wise Relevance Propagation; ML, machine learning; PCR, polymerase chain reaction; PSC, pluripotent stem cell; RNA, ribonucleic acid; SHAP, Shapley Additive Explanations; SLA, service-level agreement; TEAD, TEA domain family member; VRAM, video random access memory; XAI, explainable artificial intelligence; scRNA-seq, single-cell RNA sequencing.

References

- Chen, Z., Seman, M., Fyodorova, Y., Farhat, A., Ames, A., Levashkevich, A. et al. (2024) Tracking live-cell single-molecule dynamics enables measurements of heterochromatin-associated protein-protein interactions. *Nucleic Acids Res.* **52**, 10731–10746 <https://doi.org/10.1093/nar/gkaf692> PMID: 39142658
- McSwiggen, D.T., Liu, H., Tan, R., Agramunt Puig, S., Akella, L.B., Berman, R. et al. (2025) A high-throughput platform for single-molecule tracking identifies drug interaction and cellular mechanisms. *Elife* **12**, RP93183 <https://doi.org/10.7554/eLife.93183> PMID: 39786807
- Wang, S., Rong, R., Zhou, Q., Yang, D.M., Zhang, X., Zhan, X. et al. (2023) Deep learning of cell spatial organizations identifies clinically relevant insights in tissue images. *Nat. Commun.* **14**, 7872 <https://doi.org/10.1038/s41467-023-43172-8>
- Wang, A., Zhang, Q., Han, Y., Megason, S., Hormoz, S., Mosaliganti, K.R. et al. (2022) A novel deep learning-based 3D cell segmentation framework for future image-based disease detection. *Sci. Rep.* **12**, 342 <https://doi.org/10.1038/s41598-021-04048-3> PMID: 35013443
- Bachman, J.A., Gyori, B.M. and Sorger, P.K. (2023) Automated assembly of molecular mechanisms at scale from text mining and curated databases. *Mol. Syst. Biol.* **19**, e11325 <https://doi.org/10.15252/msb.202211325> PMID: 36938926
- Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N. et al. (2024) scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat. Methods* **21**, 1470–1480 <https://doi.org/10.1038/s41592-024-02201-0> PMID: 38409223
- De Teresa-Trueba, I., Goetz, S.K., Mattausch, A., Stojanovska, F., Zimmerli, C.E. et al. (2023) Convolutional networks for supervised mining of molecular patterns within cellular context. *Nat. Methods* **20**, 284–294 <https://doi.org/10.1038/s41592-022-01746-2> PMID: 36690741
- Ren, F., Ding, X., Zheng, M., Korzinkin, M., Cai, X., Zhu, W. et al. (2023) AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. *Chem. Sci.* **14**, 1443–1452 <https://doi.org/10.1039/d2sc05709c> PMID: 36794205
- Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D. and DiMaio, F. (2024) Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* **21**, 117–121 <https://doi.org/10.1038/s41592-023-02086-5> PMID: 37996753
- Kaddi, C., Tao, M., Bergeler, S., George, K., Geerts, H., Van der Graaf, P.H. et al. (2025) Quantitative systems pharmacology-based digital twins approach supplements clinical trial data for enzyme replacement therapies in pompe disease. *Clin. Pharmacol. Ther.* **117**, 579–588 <https://doi.org/10.1002/cpt.3498> PMID: 39632463
- Joslyn, L.R., Huang, W., Miles, D., Hosseini, I. and Ramanujan, S. (2024) Digital twins elucidate critical role of T_{scm} in clinical persistence of TCR-engineered cell therapy. *NPJ Syst. Biol. Appl.* **10**, 11 <https://doi.org/10.1038/s41540-024-00335-7> PMID: 38278838
- Li, M.M., Huang, Y., Sumathipala, M., Liang, M.Q., Valdeolivas, A., Ananthakrishnan, A.N. et al. (2024) Contextual AI models for single-cell protein biology. *Nat. Methods* **21**, 1546–1557 <https://doi.org/10.1038/s41592-024-02341-3> PMID: 39039335

- 13 He, S., Jin, Y., Nazaret, A., Shi, L., Chen, X., Rampersaud, S. et al. (2025) Starfish integrates spatial transcriptomic and histologic data to reveal heterogeneous tumor-immune hubs. *Nat. Biotechnol.* **43**, 223–235 <https://doi.org/10.1038/s41587-024-02173-8> PMID: 38514799
- 14 Keyl, P., Bischoff, P., Dernbach, G., Bockmayr, M., Fritz, R., Horst, D. et al. (2023) Single-cell gene regulatory network prediction by explainable AI. *Nucleic Acids Res.* **51**, e20–e20, e20 <https://doi.org/10.1093/nar/gkac1212> PMID: 36629274
- 15 Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X. et al. (2024) Large-scale foundation model on single-cell transcriptomics. *Nat. Methods* **21**, 1481–1491 <https://doi.org/10.1038/s41592-024-02305-7> PMID: 38844628
- 16 Ribeiro-Dantas, M. da C., Li, H., Cabeli, V., Dupuis, L., Simon, F., Hettal, L. et al. (2024) Learning interpretable causal networks from very large datasets, application to 400,000 medical records of breast cancer patients. *iScience* **27**, 109736 <https://doi.org/10.1016/j.isci.2024.109736> PMID: 38711452
- 17 Yuan, Q. and Duren, Z. (2025) Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nat. Biotechnol.* **43**, 247–257 <https://doi.org/10.1038/s41587-024-02182-7> PMID: 38609714
- 18 Hwang, Y., Cornman, A.L., Kellogg, E.H., Ovchinnikov, S. and Girguis, P.R. (2024) Genomic language model predicts protein co-regulation and function. *Nat. Commun.* **15**, 2880 <https://doi.org/10.1038/s41467-024-46947-9> PMID: 38570504
- 19 Meng, L. and Wang, X. (2024) TAWFN: a deep learning framework for protein function prediction. *Bioinformatics* **40**, btae571 <https://doi.org/10.1093/bioinformatics/btae571> PMID: 39312678
- 20 Fu, X., Mo, S., Buendia, A., Laurent, A.P., Shao, A., Alvarez-Torres, M. del M. et al. (2025) A foundation model of transcription across human cell types. *Nature* **637**, 965–973 <https://doi.org/10.1038/s41586-024-08391-z>
- 21 Zeng, Y., Xie, J., Shangquan, N., Wei, Z., Li, W., Su, Y. et al. (2025) CellFM: a large-scale foundation model pre-trained on transcriptomics of 100 million human cells. *Nat. Commun.* **16**, 4679 <https://doi.org/10.1038/s41467-025-59926-5>
- 22 Moshkov, N., Bornholdt, M., Benoit, S., Smith, M., McQuin, C., Goodman, A. et al. (2024) Learning representations for image-based profiling of perturbations. *Nat. Commun.* **15**, 1594 <https://doi.org/10.1038/s41467-024-45999-1> PMID: 38383513
- 23 Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., Baldini, G. et al. (2024) CellViT: vision transformers for precise cell segmentation and classification. *Med. Image Anal.* **94**, 103143 <https://doi.org/10.1016/j.media.2024.103143> PMID: 38507894
- 24 Wallach, I., Bernard, D., Nguyen, K., Ho, G., Morrison, A., Stecula, A. et al. (2024) AI is a viable alternative to high throughput screening: a 318-target study. *Sci. Rep.* **14**, 7526 <https://doi.org/10.1038/s41598-024-54655-z>
- 25 Mayr, A., Klambauer, G., Unterthiner, T. and Hochreiter, S. (2016) DeepTox: Toxicity Prediction using Deep Learning. *Front. Environ. Sci.* **3** <https://doi.org/10.3389/fenvs.2015.00080>
- 26 Wu, Z., Ramsundar, B., Feinberg, E.N., Gomes, J., Geniesse, C., Pappu, A.S. et al. (2018) MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **9**, 513–530 <https://doi.org/10.1039/c7sc02664a> PMID: 29629118
- 27 O'Brien, T., Stremmel, J., Pio-Lopez, L., McMillen, P., Rasmussen-Ivey, C. and Levin, M. (2024) Machine learning for hypothesis generation in biology and medicine: exploring the latent space of neuroscience and developmental bioelectricity. *Digital Discovery* **3**, 249–263 <https://doi.org/10.1039/D3DD000185G>
- 28 Pachitariu, M. and Stringer, C. (2022) Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 <https://doi.org/10.1038/s41592-022-01663-4> PMID: 36344832
- 29 Stirling, D.R., Swain-Bowden, M.J., Lucas, A.M., Carpenter, A.E., Cimini, B.A. and Goodman, A. (2021) CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 <https://doi.org/10.1186/s12859-021-04344-9> PMID: 34507520
- 30 Kobayashi, H., Cheveralls, K.C., Leonetti, M.D. and Royer, L.A. (2022) Self-supervised deep learning encodes high-resolution features of protein subcellular localization. *Nat. Methods* **19**, 995–1003 <https://doi.org/10.1038/s41592-022-01541-z> PMID: 35879608
- 31 Greenwald, N.F., Miller, G., Moen, E., Kong, A., Kagel, A., Dougherty, T. et al. (2022) Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **40**, 555–565 <https://doi.org/10.1038/s41587-021-01094-0> PMID: 34795433
- 32 Yuan, G.-H., Li, J., Yang, Z., Chen, Y.-Q., Yuan, Z., Chen, T. et al. (2024) Deep generative model for protein subcellular localization prediction. *Bioinformatics*. <https://doi.org/10.1101/2024.10.29.620765>
- 33 Schmidt, U., Weigert, M., Broaddus, C. and Myers, G. (2018) Cell Detection with Star-Convex Polygons. In *editors Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds), pp. 265–273, Springer International Publishing, https://doi.org/10.1007/978-3-030-00934-2_30
- 34 Ershov, D., Phan, M.-S., Pylvänäinen, J.W., Rigaud, S.U., Le Blanc, L., Charles-Orszag, A. et al. (2022) TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat. Methods* **19**, 829–832 <https://doi.org/10.1038/s41592-022-01507-1> PMID: 35654950
- 35 Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J. and Maier-Hein, K.H. (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 <https://doi.org/10.1038/s41592-020-01008-z> PMID: 33288961
- 36 Marazzi, L., Shah, M., Balakrishnan, S., Patil, A. and Vera-Licona, P. (2022) NETISCE: a network-based tool for cell fate reprogramming. *NPJ Syst. Biol. Appl.* **8**, 21 <https://doi.org/10.1038/s41540-022-00231-y> PMID: 35725577
- 37 Papp, O., Jordán, V., Hetey, S., Balázs, R., Kaszás, V., Bartha, Á. et al. (2024) Network-driven cancer cell avatars for combination discovery and biomarker identification for DNA damage response inhibitors. *NPJ Syst. Biol. Appl.* **10**, 68 <https://doi.org/10.1038/s41540-024-00394-w> PMID: 38906870
- 38 Yang, W., Wang, P., Xu, S., Wang, T., Luo, M., Cai, Y. et al. (2024) Deciphering cell–cell communication at single-cell resolution for spatial transcriptomics with subgraph-based graph attention network. *Nat. Commun.* **15**, 7101 <https://doi.org/10.1038/s41467-024-51329-2>
- 39 He, Z., Hu, S., Chen, Y., An, S., Zhou, J., Liu, R. et al. (2024) Mosaic integration and knowledge transfer of single-cell multimodal data with MIDAS. *Nat. Biotechnol.* **42**, 1594–1605 <https://doi.org/10.1038/s41587-023-02040-y> PMID: 38263515
- 40 Ashuach, T., Gabitto, M.I., Koodli, R.V., Saldi, G.-A., Jordan, M.I. and Yosef, N. (2023) MultiVI: deep generative model for the integration of multimodal data. *Nat. Methods* **20**, 1222–1231 <https://doi.org/10.1038/s41592-023-01909-9> PMID: 37386189
- 41 Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R. et al. (2021) scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nat. Commun.* **12**, 1882 <https://doi.org/10.1038/s41467-021-22197-x>
- 42 Wan, X., Xiao, J., Tam, S.S.T., Cai, M., Sugimura, R., Wang, Y. et al. (2023) Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat. Commun.* **14**, 7848 <https://doi.org/10.1038/s41467-023-43629-w> PMID: 38030617

- 43 Zhang, D., Schroeder, A., Yan, H., Yang, H., Hu, J., Lee, M.Y.Y. et al. (2024) Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nat. Biotechnol.* **42**, 1372–1377 <https://doi.org/10.1038/s41587-023-02019-9> PMID: 38168986
- 44 Haviv, D., Remšik, J., Gatie, M., Snopkowski, C., Takizawa, M., Pereira, N. et al. (2025) The covariance environment defines cellular niches for spatial inference. *Nat. Biotechnol.* **43**, 269–280 <https://doi.org/10.1038/s41587-024-02193-4> PMID: 38565973
- 45 Schmauch, B., Cabeli, V., Domingues, O.D., Le Douget, J.-E., Hardy, A., Belbahri, R. et al. (2025) Deep learning uncovers histological patterns of YAP1/TEAD activity related to disease aggressiveness in cancer patients. *iScience* **28**, 111638 <https://doi.org/10.1016/j.isci.2024.111638> PMID: 39868035
- 46 Derraz, B., Breda, G., Kaempf, C., Baenke, F., Cotte, F., Reiche, K. et al. (2024) New regulatory thinking is needed for AI-based personalised drug and cell therapies in precision oncology. *NPJ Precis. Oncol.* **8**, 23 <https://doi.org/10.1038/s41698-024-00517-w> PMID: 38291217
- 47 Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y. et al. (2019) U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 <https://doi.org/10.1038/s41592-018-0261-2> PMID: 30559429
- 48 Stringer, C., Wang, T., Michaelos, M. and Pachitariu, M. (2021) Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 <https://doi.org/10.1038/s41592-020-01018-x> PMID: 33318659
- 49 Tinevez, J.-Y., Dragavon, J., Baba-Aissa, L., Roux, P., Perret, E., Canivet, A. et al. (2012) A quantitative method for measuring phototoxicity of a live cell imaging microscope. *Meth. Enzymol.* **506**, 291–309 <https://doi.org/10.1016/B978-0-12-391856-7.00039-1> PMID: 22341230
- 50 AI-recognized mitochondrial phenotype enables identification of drug targets. *Nat. Comput. Sci.* **4**, 563–564 <https://doi.org/10.1038/s43588-024-00682-9>
- 51 Dang, D., Efstathiou, C., Sun, D., Yue, H., Sastry, N.R. and Draviam, V.M. (2023) Deep learning techniques and mathematical modeling allow 3D analysis of mitotic spindle dynamics. *J. Cell Biol.* **222**, e202111094 <https://doi.org/10.1083/jcb.202111094> PMID: 36880744
- 52 Way, G.P., Kost-Alimova, M., Shibue, T., Harrington, W.F., Gill, S., Piccioni, F. et al. (2021) Predicting cell health phenotypes using image-based morphology profiling. *Mol. Biol. Cell* **32**, 995–1005 <https://doi.org/10.1091/mbc.E20-12-0784> PMID: 33534641
- 53 Zagajewski, A., Turner, P., Feehily, C., El Sayyed, H., Andersson, M., Barrett, L. et al. (2023) Deep learning and single-cell phenotyping for rapid antimicrobial susceptibility detection in Escherichia coli. *Commun. Biol.* **6**, 1164 <https://doi.org/10.1038/s42003-023-05524-4> PMID: 37964031
- 54 Salek, M., Li, N., Chou, H.-P., Saini, K., Jovic, A., Jacobs, K.B. et al. (2023) COSMOS: a platform for real-time morphology-based, label-free cell sorting using deep learning. *Commun. Biol.* **6**, 971 <https://doi.org/10.1038/s42003-023-05325-9> PMID: 37740030
- 55 He, B., Bergenstråhle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, Å. et al. (2020) Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat. Biomed. Eng.* **4**, 827–834 <https://doi.org/10.1038/s41551-020-0578-x> PMID: 32572199
- 56 Duran, I., Pombo, J., Sun, B., Gallage, S., Kudo, H., McHugh, D. et al. (2024) Detection of senescence using machine learning algorithms based on nuclear features. *Nat. Commun.* **15**, 1041 <https://doi.org/10.1038/s41467-024-45421-w> PMID: 38310113
- 57 Vedeneva, E., Gursky, V., Samsonova, M. and Neganova, I. (2023) Morphological Signal Processing for Phenotype Recognition of Human Pluripotent Stem Cells Using Machine Learning Methods. *Biomedicine* **11**, 3005 <https://doi.org/10.3390/biomedicine11113005> PMID: 38002005
- 58 Fischer, C.A., Besora-Casals, L., Rolland, S.G., Haeussler, S., Singh, K., Duchon, M. et al. (2020) MitoSegNet: easy-to-use deep learning segmentation for analyzing mitochondrial morphology. *iScience* **23**, 101601 <https://doi.org/10.1016/j.isci.2020.101601> PMID: 33083756
- 59 Moebel, E., Martinez-Sanchez, A., Lamm, L., Righetto, R.D., Wietrzynski, W., Albert, S. et al. (2021) Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nat. Methods* **18**, 1386–1394 <https://doi.org/10.1038/s41592-021-01275-4> PMID: 34675434
- 60 Thummuluri, V., Almagro Armenteros, J.J., Johansen, A.R., Nielsen, H. and Winther, O. (2022) DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* **50**, W228–W234 <https://doi.org/10.1093/nar/gkac278> PMID: 35489069
- 61 Gallusser, B., Maltese, G., Di Caprio, G., Vadakkan, T.J., Sanyal, A., Somerville, E. et al. (2023) Deep neural network automated segmentation of cellular structures in volume electron microscopy. *J. Cell Biol.* **222**, e202208005 <https://doi.org/10.1083/jcb.202208005> PMID: 36469001
- 62 Jiang, Y., Wang, D., Yao, Y., Eubel, H., Künzler, P., Möller, I.M. et al. (2021) MULocDeep: A deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation. *Comput. Struct. Biotechnol. J.* **19**, 4825–4839 <https://doi.org/10.1016/j.csbj.2021.08.027> PMID: 34522290
- 63 Landeros, C., Oh, J., Weissleder, R. and Lee, H. (2024) Deep learning pipeline for automated cell profiling from cyclic imaging. *Sci. Rep.* **14**, 23600 <https://doi.org/10.1038/s41598-024-74597-w> PMID: 39384907
- 64 Phillip, J.M., Han, K.-S., Chen, W.-C., Wirtz, D. and Wu, P.-H. (2021) A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. *Nat. Protoc.* **16**, 754–774 <https://doi.org/10.1038/s41596-020-00432-x> PMID: 33424024
- 65 Davies, S.P., Ronca, V., Wootton, G.E., Krajewska, N.M., Bozward, A.G., Fiancette, R. et al. (2024) Expression of E-cadherin by CD8⁺ T cells promotes their invasion into biliary epithelial cells. *Nat. Commun.* **15**, 853 <https://doi.org/10.1038/s41467-024-44910-2> PMID: 38286990
- 66 Lashen, A.G., Wahab, N., Toss, M., Miligy, I., Ghanaam, S., Makhlof, S. et al. (2024) Characterization of Breast Cancer Intra-Tumor Heterogeneity Using Artificial Intelligence. *Cancers (Basel)* **16**, 3849 <https://doi.org/10.3390/cancers16223849> PMID: 39594804
- 67 Helgadottir, S., Midtvedt, B., Pineda, J., Sabirsh, A., B. Adiels, C., Romeo, S. et al. (2021) Extracting quantitative biological information from bright-field cell images using deep learning. *Biophys. Rev.* **2**, 031401 <https://doi.org/10.1063/5.0044782>
- 68 Edlund, C., Jackson, T.R., Khalid, N., Bevan, N., Dale, T., Dengel, A. et al. (2021) LIVECell-A large-scale dataset for label-free live cell segmentation. *Nat. Methods* **18**, 1038–1045 <https://doi.org/10.1038/s41592-021-01249-6> PMID: 34462594
- 69 Yang, X., Chen, D., Sun, Q., Wang, Y., Xia, Y., Yang, J. et al. (2023) A live-cell image-based machine learning strategy for reducing variability in PSC differentiation systems. *Cell Discov.* **9**, 53 <https://doi.org/10.1038/s41421-023-00543-1> PMID: 37280224
- 70 Pylvänäinen, J.W., Gómez-de-Mariscal, E., Henriques, R. and Jacquemet, G. (2023) Live-cell imaging in the deep learning era. *Curr. Opin. Cell Biol.* **85**, 102271 <https://doi.org/10.1016/j.cceb.2023.102271> PMID: 37897927

- 71 Morais, M.C.C., Silva, D., Milagre, M.M., De Oliveira, M.T., Pereira, T, et al. (2022) Automatic detection of the parasite *Trypanosoma cruzi* in blood smears using a machine learning approach applied to mobile phone images. *PeerJ* **10**, e13470 <https://doi.org/10.7717/peerj.13470> PMID: 35651746
- 72 Zhang, S., Pyne, S., Pietrzak, S., Halberg, S., McCalla, S.G., Siahpirani, A.F. et al. (2023) Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. *Nat. Commun.* **14**, 3064 <https://doi.org/10.1038/s41467-023-38637-9>
- 73 Cao, Z.-J. and Gao, G. (2022) Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **40**, 1458–1466 <https://doi.org/10.1038/s41587-022-01284-4> PMID: 35501393
- 74 Su, J., Song, Y., Zhu, Z., Huang, X., Fan, J., Qiao, J. et al. (2024) Cell-cell communication: new insights and clinical implications. *Signal Transduct. Target. Ther.* **9**, 196 <https://doi.org/10.1038/s41392-024-01888-z> PMID: 39107318
- 75 Sánchez-Gutiérrez, M.E. and González-Pérez, P.P. (2022) Modeling and Simulation of Cell Signaling Networks for Subsequent Analytics Processes Using Big Data and Machine Learning. *Bioinform. Biol. Insights* **16** <https://doi.org/10.1177/11779322221091739> PMID: 35478994
- 76 Ma, C. and Gurkan-Cavusoglu, E. (2024) A comprehensive review of computational cell cycle models in guiding cancer treatment strategies. *NPJ Syst. Biol. Appl.* **10**, 71 <https://doi.org/10.1038/s41540-024-00397-7> PMID: 38969664
- 77 Zhu, Y., Huang, R., Wu, Z., Song, S., Cheng, L. and Zhu, R. (2021) Deep learning-based predictive identification of neural stem cell differentiation. *Nat. Commun.* **12**, 2614 <https://doi.org/10.1038/s41467-021-22758-0>
- 78 Sinha, S., Vegesna, R., Mukherjee, S., Kammula, A.V., Dhruva, S.R., Wu, W. et al. (2024) PERCEPTION predicts patient response and resistance to treatment using single-cell transcriptomics of their tumors. *Nat. Cancer* **5**, 938–952 <https://doi.org/10.1038/s43018-024-00756-7> PMID: 38637658
- 79 Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K. and Tekade, R.K. (2021) Artificial intelligence in drug discovery and development. *Drug Discov. Today* **26**, 80–93 <https://doi.org/10.1016/j.drudis.2020.10.010> PMID: 33099022
- 80 Liao, J., Li, X., Gan, Y., Han, S., Rong, P., Wang, W. et al. (2022) Artificial intelligence assists precision medicine in cancer treatment. *Front. Oncol.* **12**, 998222 <https://doi.org/10.3389/fonc.2022.998222> PMID: 36686757
- 81 Xu, Y., Su, G.-H., Ma, D., Xiao, Y., Shao, Z.-M. and Jiang, Y.-Z. (2021) Technological advances in cancer immunity: from immunogenomics to single-cell analysis and artificial intelligence. *Signal Transduct. Target. Ther.* **6**, 312 <https://doi.org/10.1038/s41392-021-00729-7> PMID: 34417437
- 82 Basile, A.O., Yahi, A. and Tatonetti, N.P. (2019) Artificial Intelligence for Drug Toxicity and Safety. *Trends Pharmacol. Sci.* **40**, 624–635 <https://doi.org/10.1016/j.tips.2019.07.005> PMID: 31383376
- 83 Goshisht, M.K. (2024) Machine learning and deep learning in synthetic biology: key architectures, applications, and challenges. *ACS Omega* **9**, 9921–9945 <https://doi.org/10.1021/acsomega.3c05913> PMID: 38463314
- 84 Liu, Z., Wang, J. and Nielsen, J. (2022) Yeast synthetic biology advances biofuel production. *Curr. Opin. Microbiol.* **65**, 33–39 <https://doi.org/10.1016/j.mib.2021.10.010> PMID: 34739924
- 85 Merzbacher, C. and Oyarzún, D.A. (2023) Applications of artificial intelligence and machine learning in dynamic pathway engineering. *Biochem. Soc. Trans.* **51**, 1871–1879 <https://doi.org/10.1042/BST20221542> PMID: 37656433
- 86 García-Granados, R., Lerma-Escalera, J.A. and Morones-Ramírez, J.R. (2019) Metabolic engineering and synthetic biology: Synergies, future, and challenges. *Front. Bioeng. Biotechnol.* **7**, 36 <https://doi.org/10.3389/fbioe.2019.00036> PMID: 30886847
- 87 Savage, N. (2023) Drug discovery companies are customizing ChatGPT: here's how. *Nat. Biotechnol.* **41**, 585–586 <https://doi.org/10.1038/s41587-023-01788-7> PMID: 37095351
- 88 Narganes-Carlón, D., Crowther, D.J. and Pearson, E.R. (2023) A publication-wide association study (PWAS), historical language models to prioritise novel therapeutic drug targets. *Sci. Rep.* **13**, 8366 <https://doi.org/10.1038/s41598-023-35597-4> PMID: 37225853
- 89 Kornauth, C., Pemovska, T., Vladimer, G.I., Bayer, G., Bergmann, M., Eder, S. et al. (2022) Functional precision medicine provides clinical benefit in advanced aggressive hematologic cancers and identifies exceptional responders. *Cancer Discov.* **12**, 372–387 <https://doi.org/10.1158/2159-8290.CD-21-0538> PMID: 34635570
- 90 Kopac, T. (2025) Leveraging Artificial Intelligence and Machine Learning for Characterizing Protein Corona, Nanobiological Interactions, and Advancing Drug Discovery. *Bioengineering (Basel)*. **12**, 312 <https://doi.org/10.3390/bioengineering12030312> PMID: 40150776
- 91 Cao, L. (2023) Trans-AI/DS: transformative, transdisciplinary and translational artificial intelligence and data science. *Int. J. Data Sci. Anal.* **15**, 119–132 <https://doi.org/10.1007/s41060-023-00383-y>
- 92 Sedrani, C. and Wilmes, P. (2022) Toward hypothesis-driven, personalized microbiome screening. *Cell Rep. Methods* **2**, 100139 <https://doi.org/10.1016/j.crmeth.2021.100139> PMID: 35474872
- 93 Algavi, Y.M. and Borenstein, E. (2023) A data-driven approach for predicting the impact of drugs on the human microbiome. *Nat. Commun.* **14**, 3614 <https://doi.org/10.1038/s41467-023-39264-0> PMID: 37330560
- 94 Liu, R., Rizzo, S., Whipple, S., Pal, N., Pineda, A.L., Lu, M. et al. (2021) Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* **592**, 629–633 <https://doi.org/10.1038/s41586-021-03430-5>
- 95 Xianyu, Z., Correia, C., Ung, C.Y., Zhu, S., Billadeau, D.D. and Li, H. (2024) The Rise of Hypothesis-Driven Artificial Intelligence in Oncology. *Cancers (Basel)* **16**, 822 <https://doi.org/10.3390/cancers16040822> PMID: 38398213
- 96 Zabolocki, M., McCormack, K., Van den Hurk, M., Milky, B., Shoubridge, A.P. et al. (2020) BrainPhys neuronal medium optimized for imaging and optogenetics in vitro. *Nat. Commun.* **11**, 5550 <https://doi.org/10.1038/s41467-020-19275-x> PMID: 33144563
- 97 Kim, D., Min, Y., Oh, J.M. and Cho, Y.-K. (2019) AI-powered transmitted light microscopy for functional analysis of live cells. *Sci. Rep.* **9**, 18428 <https://doi.org/10.1038/s41598-019-54961-x>
- 98 Ghosh, S., Zhao, X., Alim, M., Brudno, M. and Bhat, M. (2025) Artificial intelligence applied to 'omics data in liver disease: towards a personalised approach for diagnosis, prognosis and treatment. *Gut* **74**, 295–311 <https://doi.org/10.1136/gutjnl-2023-331740>
- 99 Toussaint, P.A., Leiser, F., Thiebes, S., Schlesner, M., Brors, B. and Sunyaev, A. (2023) Explainable artificial intelligence for omics data: a systematic mapping study. *Brief. Bioinformatics* **25**, bbad453 <https://doi.org/10.1093/bib/bbad453> PMID: 38113073
- 100 Aliouane, S.E., Chehili, H., Boulahrouf, K., Abdelaziz, A., Khelifa, N. and Hamidechi, M.A. (2025) Integrating deep learning and SHAP for breast cancer classification and biomarker discovery using gene expression data. *IEEE Access* **13**, 49693–49709 <https://doi.org/10.1109/ACCESS.2025.3552280>

- 101 Laatifi, M., Douzi, S., Ezzine, H., Asry, C.E., Naya, A., Bouklouze, A. et al. (2023) Explanatory predictive model for COVID-19 severity risk employing machine learning, shapley addition, and LIME. *Sci. Rep.* **13**, 5481 <https://doi.org/10.1038/s41598-023-31542-7> PMID: 37015978
- 102 Di Giammarco, M., Martinelli, F., Santone, A., Cesarelli, M. and Mercaldo, F. (2024) Colon cancer diagnosis by means of explainable deep learning. *Sci. Rep.* **14**, 15334 <https://doi.org/10.1038/s41598-024-63659-8> PMID: 38961080
- 103 Keyl, J., Keyl, P., Montavon, G., Hosch, R., Brehmer, A., Mochmann, L. et al. (2025) Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nat. Cancer* **6**, 307–322 <https://doi.org/10.1038/s43018-024-00891-1> PMID: 39885364
- 104 Zargari Marandi, R. (2024) Explainer: an R package to explain machine learning models. *Bioinform. Adv.* **4**, vbae049 <https://doi.org/10.1093/bioadv/vbae049> PMID: 38577543
- 105 Tang, X., Prodduturi, N., Thompson, K.J., Weinshilboum, R., O'Sullivan, C.C., Boughey, J.C. et al. (2024) OmicsFootPrint: a framework to integrate and interpret multi-omics data using circular images and deep neural networks. *Nucleic Acids Res.* **52**, e99–e99, e99 <https://doi.org/10.1093/nar/gkae915> PMID: 39445795
- 106 Gao, X., Zhang, F., Guo, X., Yao, M., Wang, X., Chen, D. et al. (2025) Attention-based deep learning for accurate cell image analysis. *Sci. Rep.* **15**, 1265 <https://doi.org/10.1038/s41598-025-85608-9>
- 107 Xiong, G., Bekiranov, S. and Zhang, A. (2023) ProtoCell4P: an explainable prototype-based neural network for patient classification using single-cell RNA-seq. *Bioinformatics* **39**, btad493 <https://doi.org/10.1093/bioinformatics/btad493> PMID: 37540223
- 108 Lubiana, T. and Nakaya, H.I. (2024) A reasonable request for true data sharing. *Lancet Reg. Health Am.* **35**, 100795 <https://doi.org/10.1016/j.lana.2024.100795> PMID: 38846807
- 109 Gao, Z. and Li, Y. (2023) Enhancing single-cell biology through advanced AI-powered microfluidics. *Biomicrofluidics* **17**, 051301 <https://doi.org/10.1063/5.0170050> PMID: 37799809
- 110 Chen, H., Lu, Y. and Rao, Y. (2024) A self-training interpretable cell type annotation framework using specific marker gene. *Bioinformatics* **40**, btac569 <https://doi.org/10.1093/bioinformatics/btac569> PMID: 39312689
- 111 Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J. et al. (2022) A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 <https://doi.org/10.1038/s41587-021-01206-w> PMID: 35132262
- 112 Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E. et al. (2017) The human cell atlas. *Elife* **6**, e27041 <https://doi.org/10.7554/eLife.27041>
- 113 Jonnagaddala, J. and Wong, Z.S.-Y. (2025) Privacy preserving strategies for electronic health records in the era of large language models. *NPJ Digit. Med.* **8**, 34 <https://doi.org/10.1038/s41746-025-01429-0> PMID: 39820020
- 114 Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 <https://doi.org/10.1126/science.aax2342> PMID: 31649194
- 115 Chen, Z., Chen, C., Yang, G., He, X., Chi, X., Zeng, Z. et al. (2024) Research integrity in the era of artificial intelligence: Challenges and responses. *Medicine (Abingdon)*. **103**, e38811 <https://doi.org/10.1097/MD.00000000000038811>
- 116 Bloomfield, D., Pannu, J., Zhu, A.W., Ng, M.Y., Lewis, A., Bendavid, E. et al. (2024) AI and biosecurity: The need for governance. *Science* **385**, 831–833 <https://doi.org/10.1126/science.adq1977> PMID: 39172825
- 117 Andrew Bergen Scalable and Automated AI Image Analysis for Volume Electron Microscopy. https://www.zeiss.com/microscopy/en/resources/insights-hub/life-sciences/scalable-and-automated-ai-image-analysis-for-volume-electron-microscopy.html?utm_source=chatgpt.com
- 118 Dessimoz, C. and Thomas, P.D. (2024) AI and the democratization of knowledge. *Sci. Data* **11**, 268 <https://doi.org/10.1038/s41597-024-03099-1> PMID: 38443367
- 119 Maracaja-Coutinho, V. and Nakaya, H.I. (2026) Single-cell research in Latin America and the Caribbean builds genomics datasets for equitable AI-powered precision medicine. *Nat. Rev. Mol. Cell Biol.* **27**, 1–3 <https://doi.org/10.1038/s41580-025-00913-5> PMID: 41068318
- 120 Possik, P.A., Adams, D.J., Aguiar, F.C., Alves, T.C., Alves-Hanna, F.S., Restrepo Arboleda, C.M. et al. (2025) Exploring Latin America one cell at a time. *Cell* **188**, 5790–5796 <https://doi.org/10.1016/j.cell.2025.09.013> PMID: 41106359
- 121 Yang, X., Chen, D., Sun, Q., Wang, Y., Xia, Y., Yang, J. et al. (2023) A live-cell image-based machine learning strategy for reducing variability in PSC differentiation systems. *Cell Discov* **9**, 53 <https://doi.org/10.1038/s41421-023-00543-1> PMID: 37280224
- 122 Bunne, C., Roohani, Y., Rosen, Y., Gupta, A., Zhang, X., Roed, M. et al. (2024) How to build the virtual cell with artificial intelligence: Priorities and opportunities. *Cell* **187**, 7045–7063 <https://doi.org/10.1016/j.cell.2024.11.015> PMID: 39672099