

Beyond Representation: AI in Cellular Discovery

Charlotte Bunne & Aviv Regev

A new paradigm is emerging at the intersection of artificial intelligence and experimental biology, where cells are no longer merely observed, but comprehensively modeled, queried, and predicted in silico. New measurement technologies and the ability to genetically manipulate cells precisely have opened the way to measure cells in extraordinary detail under tens of thousands of perturbations. Concomitantly, AI foundation models are learning to represent, simulate, and even anticipate cellular behavior. This essay traces the convergence of these revolutions, showing how they are giving rise to “virtual cells”: integrative models that unify diverse molecular and spatial data into coherent, functional representations that can generalize across biological contexts and conditions. Beyond representing and interpreting biological lab measurements, virtual cells aim to predict unseen outcomes, imagine new contexts, and guide discovery. In closing the loop between data generation and hypothesis testing, AI is transforming biology into a self-refining, interactive science, resulting in a profound shift: from observing life to actively modeling it, with implications for precision medicine, biotechnology, and the scientific method.

Understanding the immense complexity of biology requires characterizing life at its fundamental units – cells – in all their diversity of molecules, behaviors, and contexts. Humans are composed of an estimated thirty-seven trillion cells spanning many thousands of specialized types and dynamic states, each influenced by its genetics, current environment, and lived history. While cells within an individual all carry approximately the same set of instructions in their DNA, they operate by using (“expressing”) only specific components at a time (transcribing into RNA and translating the RNA into protein), as needed for their specific role and temporal context in the body, and thus vary substantially in their molecular characteristics and the functions they can execute across types, times, and conditions.

Deciphering this cellular kaleidoscope is essential for both basic science and medicine. Every process of human life, from brain activity to immune defense, emerges from the collective actions of diverse cells. Errors in these processes underlie disease: autoimmune conditions reflect immune cells turning against self-tissues, cancers emerge from a single cell with a mutated genome that ignores regula-

tory signals, and neurodegeneration occurs when vulnerable neuronal subtypes die because of the accumulation of toxic molecular aggregates and inappropriate immune cell activity. Understanding the molecular programs and interactions of cells is therefore the foundation for diagnosing disease, identifying therapeutic targets, and designing effective interventions.

Despite this cellular variation and its importance, molecular studies of cells have historically averaged across many of them at once, largely due to the limitations of measurement technologies. RNA levels, for example, were typically measured from cell cultures or tissues (and thus from many cells and not individual *single* cells) ground up into a molecular “smoothie.” While powerful, these bulk approaches hid the rich cell-to-cell heterogeneity that defines cells in both health and disease. Thus, a tumor sample’s average gene readout might miss a rare drug-resistant cell subpopulation; likewise, an immune organ’s bulk profile might overlook a few critical regulatory cells amid thousands of others.

Groundbreaking technological advances over the past decade – combining new lab measurements with powerful AI algorithms – have led to a historic shift from this aggregate view to a single-cell resolution of life. Innovations in single-cell sequencing and cytometry now allow researchers to profile millions of individual cells for their genomes, transcripts, proteins, and more – one cell at a time.¹ Early single-cell RNA sequencing (scRNA-seq) experiments revealed astonishing diversity even in what were thought to be uniform cell populations, overturning more categorical classifications of cell “types” and uncovering new cell states (for instance, previously unrecognized subtypes of neurons or complex, continuous immune cell states).² This laid the foundation for the Human Cell Atlas (HCA) initiative launched in 2016, which set out to create comprehensive reference maps of all human cells as a basis for both understanding human health and diagnosing, monitoring, and treating disease.³ Like a “Google Maps of the human body” that chart diverse landscapes at multiple levels of resolution and across different modalities (topography, street map, traffic patterns, and more), the Human Cell Atlas maps the multitude of cell types and states across tissues, organs, and systems in terms of different molecular and visual properties. By 2017, the HCA and similar efforts had demonstrated that it was possible to discover and catalog myriad cell types across the body using large-scale single-cell profiling; as of 2025, “version 1” atlases of eighteen organs and systems are near-complete.⁴ This massive data-gathering phase proved that a “census of cells” in the human body was within reach. In parallel, experimental strategies such as Perturb-seq began to extend the atlas concept beyond description: Perturb-seq couples pooled CRISPR-based gene perturbations with single-cell RNA-seq so that the effect of turning genes off or on can be read out for thousands of cells at once.⁵ In this way, Perturb-seq systematically links genetic changes to complex cellular responses, producing causal datasets that complement the observational maps of the atlas.⁶

Moving to single-cell genomic resolution fundamentally changed how we think about and interrogate biological systems and led to a plethora of basic scientific discoveries and medically relevant advances.⁷ We now focus on tissues as ecosystems of distinct cells interacting and changing over time within their native contexts. In cancer, for example, single-cell studies revealed a tumor and its microenvironment as an evolving mosaic of malignant and infiltrating immune cells, each with its own gene expression patterns. In neuroscience, single-neuron analyses have exposed hundreds of new subclasses of brain cells that were invisible in bulk measurements.⁸ In human development, the continuous processes by which the fertilized egg leads to thousands of different cell types could be recovered, showing far less discrete determinism than previously assumed. Across many organs, cell types as rare as 1 percent or 0.1 percent of the tissue cells have emerged as crucial nodes in their functioning, from the rare ionocytes in the lung and airways to enteric neurons in the gut.⁹ These insights also have direct medical implications: Single-cell atlases that link disease-associated genes to the specific cell types and contexts in which they act are clarifying mechanisms and guiding intervention. This enables more precise diagnostics, the discovery of novel drug targets, and the development of therapies across cancer and infectious, rare, and degenerative diseases.¹⁰

However, early single-cell methods typically captured only one layer of information at a time (most typically, RNA transcripts), dissociated cells from their native tissue context, and focused on the observed cells rather than the biological mechanisms that give rise to and control them. To fully understand cellular function, scientists have had to push further, combining multiple modalities per cell, deciphering the underlying causes of their actions, and putting them back into the spatial context from which they come.

Biology is inherently multidimensional, multimodal, and dynamic. A cell's identity is not defined by only one feature (say, a specific gene) or one category or modality of features (say, its mRNA levels) but by a constellation of many features (multidimensional) or many categories (multimodal) – DNA sequence and epigenetic marks, RNA transcripts, proteins, metabolites, physical shape at multiple length scales, location in a tissue, and interactions with neighbors – all of which change in time and space.¹¹

Recognizing this, the field has rapidly progressed through advances in technological measurements and computational algorithms from “snapshot” single-modality single-cell measurements to multimodal assays in cells, including dynamic ones. Today, it is feasible to profile multiple molecular layers from the same cell, for example, simultaneously capturing a cell's transcriptome (which genes are active) and its epigenome (how the DNA is marked and organized), or measuring both RNA and proteins.¹² Both experimental approaches and computational algorithms help recover dynamics, either by tracking live cells directly or by inferring dynamics from

static snapshots or molecular recorders and lineage tracers.¹³ These multimodal methods provide a more holistic view of each cell, linking gene regulation to gene expression to protein output. Instead of a one-dimensional profile, we get a richer “fingerprint” of each cell’s state, its location, and even its history and potential future.

In parallel, a revolution in spatial biology is recovering the two- and three-dimensional (2D/3D) context that traditional single-cell profiling lost. New spatial transcriptomics technologies allow us to map molecular information in 2D tissue sections and even 3D whole organs.¹⁴ Lab methods can now pinpoint individual RNA or protein molecules in cells at micron-scale resolution, whereas other techniques can recover the 2D location of a cell prior to profiling.¹⁵ The result is a tissue map showing what cell types are present and what they express, how they relate to histological images, and what tissue structures and other cells they touch. This information is crucial because cells’ behaviors and identity are profoundly influenced by their neighbors and proximal (micro) environment: a T cell nestled next to and precisely recognizing a tumor cell as “non-self” may act differently than the same T cell elsewhere; a neuron’s role in a circuit depends on its location and connections. Notably, profiling methods now span multiple scales, from subcellular imaging of organelles to cellular-resolution tissue maps to whole-organ scans. Thus, the rise of spatial profiling and advanced imaging enables us to fulfill the vision of the Human Cell Atlas not just through a cell census, but in a 3D, multiscale map of cells in the body.¹⁶

Moving from one-dimensional catalogs to multimodal, spatially resolved maps also shifts us from a cell biology centered on how cells are organized and function to a tissue biology focused on the holistic and integrated organization and function of tissues. We can begin to ask questions that were previously far less accessible: How do cells arrange into niches and communities? Which cell-to-cell interactions drive responses to injury or infection? How do signals propagate through the spatial architecture of a tissue, and how are they coordinated to perform the tissue’s and organ’s ultimate function? However, to address each of these questions requires generating, analyzing, and reasoning over enormous volumes of complex data. Indeed, it demands more than traditional analysis by calling for a partner that can navigate immense dimensionality, link disparate measurements, and surface hidden patterns and their underlying causes at scale.

This partner is artificial intelligence, which now serves as an essential component for making sense of the vast, multidimensional datasets produced by modern single-cell and spatial profiling. As experimental methods have scaled, from bulk assays to high-throughput and multi-omic measurements, machine-learning algorithms have matured. Indeed, this was not merely a fortunate coincidence but a mutually reinforcing dynamic: large volumes of rich, heterogeneous data fuel advances in AI architectures, while more capable AI tools in turn inspire and enable ever more ambitious experimental designs. Techniques such as representation learning com-

press tens of thousands of cellular features into biologically meaningful embeddings, enabling the discovery of hidden subpopulations and trajectories. Droplet-based single-cell RNA-seq, for instance, captures only 5 to 20 percent of each cell's transcripts, producing sparse and noisy profiles.¹⁷ Here, representation learning was essential to recover structure and make the method transformative.¹⁸ In parallel, integration frameworks aim to align disparate modalities – transcriptomics, proteomics, epigenomics, and imaging. These algorithms reconcile technical noise, resolve batch effects, and synthesize partial measurements so that each cell's composite profile emerges more coherently than any single assay could provide. They can also perform cross-modal inference: for example, predicting gene-expression patterns directly from histological images, thereby extending molecular insight to routine clinical specimens.¹⁹ Beyond these applications, AI contributes through predictive and generative modeling, allowing annotation of complex datasets and forecasting of cellular behavior under unseen conditions. Such models can anticipate how cells respond to perturbations or therapies or reconstruct dynamic processes from destructive measurements.²⁰

One of the most significant developments of the convergence of AI and biology is the emergence of foundation models for cell biology. In AI, “foundation models” are very large models trained on massive unlabeled datasets to perform a wide range of tasks that extend well beyond those used for pretraining.²¹ At the core of these foundation models lies the principle of self-supervised learning, a training approach that leverages inherent patterns in the data without relying on explicit labels provided by researchers. The first successful demonstration of foundation models was in natural language processing, where self-supervised methods train models by predicting masked or omitted components of text or by generating the next syllable in a sentence.²² The performance of the resulting large language models (LLMs) has been astonishing, transforming countless professional workflows – from writing, summarizing, and coding to research and customer service – and reshaping how individuals engage with and consume information in everyday life. Extending this paradigm to cell biology follows naturally, as large-scale experimental biology provides precisely the kind of data environment in which self-supervised AI thrives. High-throughput single-cell and spatial experiments routinely generate datasets of millions of cell profiles, each with thousands of molecular and spatial features, creating exactly the rich and diverse input that cell atlas foundation models need to learn robust, generalizable biological representations.²³ With this approach, scientists aim to mirror the success of LLMs, which derive powerful predictive capabilities by ingesting vast quantities of text data. Thus, where early single-cell genomics representation learning compressed noisy measurements into descriptive embeddings, foundation models are now trained on diverse, heterogeneous inputs across scales and modalities to

capture the underlying grammar of cellular systems: representations that should support prediction, querying, and generation well beyond the measured data.²⁴

However, biological datasets present distinct challenges. They are inherently multimodal (integrating transcriptomics, proteomics, epigenomics, and visual measurements) and multiscale (bridging molecular details to tissue-level organization and spanning nano, micro, and macro scales). Moreover, they represent dynamic biological systems captured through static snapshots, and they are governed by genetic and molecular causes. Addressing this complexity requires new computational architectures and training strategies specifically tailored to biological data types.

First, operationalizing large-scale foundation models within a cellular context requires reformatting raw experimental data into representations that algorithms can readily process. Experimental measurements are decomposed into standardized units, analogous to the “tokens” (basic units of text) in language models, that capture molecular features, spatial coordinates, and other cellular attributes. These tokenized profiles, derived from transcriptomic, proteomic, epigenomic, and spatial assays, are then processed by transformer-based architectures.²⁵ Such models learn statistical relationships – or a “grammar” – among these biological tokens across large cohorts of cells.

Next, the way in which these models are trained determines what kinds of patterns they capture. One family of self-supervised approaches uses masking, whereby a subset of features within a profile is hidden and the model must reconstruct them from the remaining context. This forces the model to capture covariation among genes and molecules and has proven highly effective in large single-cell and tissue foundation models.²⁶ A second strategy is integration-based alignment, whereby paired measurement types are co-embedded so that consistent structure across modalities reinforces a unified cellular representation. This allows, for instance, protein-localization patterns observed in imaging to be aligned with protein-interaction measurements.²⁷

A third novel strategy is cross-modal prediction, wherein the model is given one modality and asked to generate another. Tasks such as predicting protein abundance from RNA levels or inferring gene expression from histology images force the model to learn systematic correspondences across measurement spaces and enable translation between them.²⁸ In two other approaches, contrastive learning methods bring together representations of identical states and separate those of distinct states, structuring the embedding space so that distance reflects biological similarity rather than technical variation, while generative objectives require them to produce realistic new profiles that obey biological constraints.²⁹ Together, these objectives provide strong training signals, though it remains an open question whether they fully capture the biological rules at play, leaving room for the discovery of new learning objectives to train multiscale and multimodal foundation models for cellular biology.

Like other foundation models, once trained, these models can be applied to various downstream tasks distinct from those used for training. Using simple (linear) classification or regression models trained on the learned representations, they can distinguish subtle cellular states or predict quantitative responses, such as estimating how strongly a cell will activate under a given stimulus.³⁰ In retrieval tasks, their embeddings enable efficient matching, such as linking a rare or poorly characterized cell in a new dataset to its closest counterparts across large reference atlases, or identifying similar patients whose profiles can inform clinical decision-making.³¹

Beyond tokenization and learning objectives, the development of foundation models for cell biology has driven the design of novel architectures that move beyond the standard transformers used in language and vision. Whereas language models operate on sequential tokens and vision models on grid-structured pixels, biological data inhabit irregular, high-dimensional spaces: genes interact through complex regulatory networks, cells reside in continuous spatial and temporal contexts, and measurements are frequently sparse or missing. Meeting these demands has, for example, led to specialized attention modules, such as operators that extend context windows to capture long-range genomic interactions across millions of bases, and mechanisms that scale to the extreme dimensionality of spatial omics. In parallel, hierarchical transformer designs represent tissues across multiple levels of organization, linking molecules within cells, cells within microenvironments, and microenvironments within tissues. Another direction develops architectures that operate directly on sets of cells or continuous spatial and temporal coordinates, enabling models to capture local neighborhoods and smooth dynamics without imposing artificial sequence or grid structures. Strategies have also been introduced to ensure robustness across heterogeneous assays, such as technologies that capture thousands of RNA molecular species but only a limited number of protein species.

Advances in the design of foundation models for cell biology have already yielded practical scientific benefits. Models trained in a self-supervised manner on hundreds of millions of single-cell profiles – a scale of data integration across cell types, tissues, and organs not previously attempted in computational biology – outperform classical tools in recognizing cell types, identifying gene functions, and reconstructing gene interaction networks.³² Their embeddings also make it easier to pinpoint rare subtypes.³³ Another striking capability comes from models that integrate microscopy images with molecular profiles. Traditionally, pathologists could only qualitatively infer cellular activity from a tissue slide, but AI is changing that. Recently developed foundation models link histology to gene expression: by training on over 2.2 million paired images and RNA measurements from the same tissue “pixel,” the model learned to predict which genes are active in a section directly from an H&E (Hematoxylin and Eosin)-stained image.³⁴ In practice, this means we can take an ordinary histopathology slide of a patient’s biopsy and approximate the spatial pattern of gene activity – without any specialized measurement.

Foundation models are also beginning to address causality: For example, by identifying genes whose action controls the cellular phenotypes, like RNA levels. Lab experiments typically rely on genetic manipulations, changing genes one at a time (knockout or overexpression) and observing the effect. Single-cell genomics methods combined with CRISPR-based perturbations, as in Perturb-seq, allow thousands of genetic changes to be tested simultaneously, each cell carrying a different edit and being read out at single-cell resolution.³⁵ Even so, the combinatorial space of possible perturbations remains far larger than what can be experimentally tested. Foundation models offer a path forward by predicting perturbation outcomes, a key step toward true virtual experiments. In recent studies, models trained on large compendia of unperturbed cells were fine-tuned to anticipate how those cells' transcriptomes would shift after specific interventions (such as a drug or a CRISPR-based gene knockout).³⁶ These models generalized to new, unseen perturbations with partial success, effectively performing zero-shot predictions.³⁷ For example, without ever having observed a particular drug in training data, a model could predict with reasonable accuracy which genes increase or decrease because of it, and by how much, in a given cell type, extrapolating from patterns learned during broad pretraining. While still evolving, early results show that these models often match or exceed the accuracy of simpler predictive methods. Moreover, they do so in a mechanism-agnostic way: the model is not explicitly told about molecular pathways or networks; it infers likely cascades of change from data. This ability to successfully model causal effects in a cell or, essentially, to perform *in silico* perturbation experiments would represent a fundamental advance in computational biology.

Beyond foundation models, the convergence of AI and high-throughput cell biology is now entering a transformative new phase in which artificial intelligence moves past passive analysis to actively participate in, and even drive, the scientific discovery process. The next frontier involves developing AI agents: autonomous systems that can actively orchestrate entire research workflows. Unlike traditional AI models that passively analyze data or make predictions based on a user's prompt, an AI agent is a system that can perceive its environment, make decisions, and take actions to achieve specific goals. In the context of biological research, these agents can act as intelligent intermediaries between computational models and laboratory experiments, essentially becoming digital research assistants that can propose hypotheses, design experiments, and learn from results.³⁸

AI agents are beginning to transform the traditional scientific method through a lab-in-the-loop approach – a paradigm in which the agent integrates AI models, multiple analysis tools, and experimental platforms in a closed-loop cycle. Using this approach, it might propose a hypothesis or drug candidate based on computational predictions from a model, instruct robotic lab instruments to execute the experi-

ment, analyze the results, and then use those findings to refine its next hypothesis, all with minimal human intervention. This creates a highly iterative and adaptive workflow that has the potential to dramatically accelerate the pace of discovery. Rather than the traditional linear progression of design → experiment → analysis, the lab-in-the-loop paradigm enables massively parallel experimentation with continuous feedback loops, in which computational predictions and experimental results are seamlessly integrated at unprecedented scales.

Central to achieving this integration are sophisticated AI technologies that enable both virtual experimentation and intelligent decision-making. First, the AI agents need to be powered by effective generative models: machine-learning systems trained to learn the underlying patterns and distributions of complex biological data, enabling them to generate new, realistic biological scenarios that have never been observed.³⁹ Unlike traditional predictive models that merely classify or forecast outcomes, generative models can create entirely novel molecular structures, cellular behaviors, or experimental designs by sampling from learned biological distributions. In the lab-in-the-loop context, these models serve as powerful hypothesis-generating engines, simulating thousands of potential experiments *in silico* before any physical resources are consumed. For instance, an AI agent might use a generative model to explore how modifying specific nucleic acids in a regulatory DNA sequence could enhance or repress the expression of a gene, effectively running virtual experiments across a vast design space to identify the most promising candidates for laboratory validation.⁴⁰ Importantly, such closed-loop experimentation can be pursued with two distinct goals: either to address a concrete biological question, such as identifying a more stable protein variant, or to broaden and refine the generative model itself so that it captures biological principles more universally. Once the agent's suggested experiment is executed in the lab, the outcome data are immediately fed back into the system, enabling the generative model to refine its understanding of the biological landscape and improve future predictions.

Second, to fully realize the potential of autonomous scientific discovery, future lab-in-the-loop systems will need to incorporate reinforcement learning, a machine-learning paradigm in which agents learn optimal strategies through trial and error, receiving rewards or penalties based on the outcomes of their actions.⁴¹ Unlike supervised learning that requires labeled examples, reinforcement learning excels in sequential decision-making scenarios in which the best path forward must be discovered through exploration. In the envisioned framework, such systems would treat each experiment as an “action” and assign feedback based on the experimental outcome. An experiment yielding results that confirm a hypothesis or improve a target metric – such as enhanced enzyme activity or drug efficacy – would generate a positive reward, prompting the agent to update its strategy accordingly. This approach could enable AI agents to systematically navigate the vast experimental landscape (which cannot ever be fully tested in a lab), gradually learning optimal

experimentation policies that maximize scientific insight while minimizing time and resources. Maximizing the potential of reinforcement learning in this setting will require shifts in biological experimentation. Increased automation, miniaturization, and continuous monitoring and measurements can all greatly accelerate the success of this scientific process. Moreover, while large-scale experimental investments in biology were historically conducted in batch mode (one large experiment conducted to completeness prior to analysis), in a reinforcement learning setting, it is preferable to use the same experimental resources gradually across iterations. Finally, to reap the broader benefit for the model, scientists may need to accept that a given cycle may not provide the optimal experiment to address a specific question of interest.

This integration of AI agents with experimentation is particularly transformative when applied to the massive datasets generated by modern single-cell and multi-omics technologies. Although researchers can still design informative studies, systematically navigating the combinatorial space opened by datasets that profile millions (and more) of single cells across several molecular layers is beyond what unaided human reasoning can optimize efficiently. AI agents, on the other hand, thrive on such complexity. For instance, when analyzing a single-cell transcriptomic dataset from a tumor sample, an AI agent could identify rare subpopulations characterized by distinctive molecular signatures based on integrated pathway databases, prior biological literature, and predictive modeling of gene interactions. It might then use generative models to simulate the functional outcomes of perturbing hundreds of candidate regulatory genes *in silico*, selecting the most informative targets predicted to have the greatest impact on the cell phenotype or on the tissue niche. These selected targets can then be tested experimentally through high-throughput, multiplexed perturbation screens executed in cells in culture or in an animal model. Upon completion of these experiments, the AI agent could instantly integrate these new measurements, update its predictive models of the underlying regulatory networks, and refine subsequent experimental designs. By leveraging vast molecular knowledge, high-dimensional inference, and real-time feedback, such AI-driven workflows could help systematically identify critical biological insights far more efficiently than manual methods – either autonomously or collaboratively with a scientist user – maximizing the learnings achieved from a given volume of experiments and resources.

These capabilities are now converging toward an ambitious vision: virtual cells.⁴² Virtual cells are computational representations that integrate diverse molecular, spatial, and temporal data to capture, generalize, and simulate cellular behavior and dynamics. A virtual cell is a digital twin of a biological cell, and can be probed and perturbed in the computer just as a real cell might be in the lab. The prospects of this approach are intriguing. If successful, scientists could

predict a cell's response to virtually any change (a drug, a genetic mutation, a viral infection) without needing to perform every experiment in a dish. Much as LLMs learned to “decode” and “speak” natural language, these biological models aim to decode the experimental language of cells, enabling us to virtually “ask” a cellular system what would happen if we tweak a gene or expose it to a new stimulus.

The foundation models describe first steps toward but not yet realizations of the virtual cell approach. Their ability to capture cellular identities, predict perturbation outcomes, and bridge imaging with molecular data forms the scaffold on which more comprehensive virtual cells will be built.⁴³ At the same time, AI agents are emerging as an operational layer that can interrogate these models, generate hypotheses, design targeted experiments, and orchestrate model improvements, creating an interactive loop between *in silico* insight and *in vitro* validation.⁴⁴ Together, agents and foundation models hint at a future in which we can maintain computational mirrors of cells that behave with realistic biological fidelity and are continuously refined by experimental feedback. The concept of virtual cells thus encapsulates this momentum: with sufficient data, sophisticated learning, and agent-driven orchestration, we can create representations that capture a cell's essential behaviors and responses. Although such pioneering efforts are still in their early stages, they offer a glimpse of what is coming. If realized, virtual cells would make it possible to explore biological questions at a scale that no laboratory can match, rapidly screening interventions, uncovering mechanisms hidden from direct measurement, and tailoring therapies to patient-specific cellular states.

We stand at a moment akin to the advent of the microscope, but now empowered by silicon, molecules, and statistics. Just as seventeenth-century optics revealed an invisible world of cells and microbes, today's single-cell and spatial multi-omics analysis combined with AI is exposing hidden patterns, interactions, and potentials within and between cells. The relationship is synergistic and catalytic; high-throughput methods generate rich, multidimensional datasets, and AI transforms those data into coherent models, uncovering hidden patterns and pointing to the next questions to explore. Absent machine learning and artificial intelligence, single-cell genomics methods would have yielded uninterpretable data. Absent single-cell genomics, the data for learning AI models of molecular cell biology would not have been available at the needed scale. The notion of a virtual cell encapsulates this combination: a model trained on large collections of multiscale and multimodal measurements, so faithful and comprehensive that it can serve as an *in silico* surrogate for the real biological system.⁴⁵

The consequences of this shift will extend across science and medicine. In the clinic, diagnostics will move from organ-level descriptions toward pinpointing specific cellular subpopulations and the molecular interactions underlying disease. With models that translate and generate across modalities, even a simple blood

draw might be transformed into a comprehensive multi-omic profile, revealing spatial and functional states of cells that previously required invasive biopsies.⁴⁶ Therapeutic strategies, from small molecules to engineered cells, will increasingly rely on predictive models that simulate cellular responses and generate mechanistic explanations in partnership with AI agents.⁴⁷ Beyond medicine, similar approaches promise to illuminate how cellular ecosystems adapt in evolution and ecology. At a deeper level, these developments may alter our very conception of biology, shifting us from reductionist perspectives to viewing life as a complex, dynamic information system that we can systematically decode, rationally manipulate, and predict.

Many challenges remain, and humility is essential. Biology abounds with context-dependent phenomena, nonlinear feedback, multiple physical and temporal scales, many variables that are challenging to measure, and stochastic noise that confound even the most sophisticated models. Foundation models trained on current datasets may absorb technical artifacts or sampling biases, leading to predictions that fail to generalize across tissues, laboratories, or patient populations. Many molecules, structures, and interactions remain exceedingly difficult to measure, especially in three dimensions and over time; and in most demanding settings, models still cannot match laboratory experiments. Moreover, the “black box” character of deep architectures raises questions about interpretability and trust: without clear mechanistic explanations, it is hard to distinguish genuine insight from spurious correlation. To build confidence, models must undergo rigorous validation, reproducing results across cohorts, predicting outcomes of prospective perturbations, and withstanding real-world application.

Further, biological machine learning has so far focused primarily on prediction: identifying cell types, forecasting perturbation outcomes, or inferring spatial arrangements. What is needed next is the ability to query models in ways that surface concepts, much as one might prompt a colleague by asking not only what will happen but why. Beyond queryability lies reasoning, where models can weigh competing causal explanations, such as whether an immune response is driven by local cytokine gradients or by signals from distant tissues and what may be the functional reasons or consequences. Finally, true transformation will require long-term thinking: systems that can carry a biological problem over weeks or months, continuously refining hypotheses in the background, like a scientist who keeps returning to a puzzle until a unifying explanation emerges. Together, these capabilities would turn artificial intelligence from a predictive tool into a co-investigator, capable of uncovering mechanisms and reshaping how biology is practiced.

The path forward depends as much on advances in experimental biology as on algorithms. New biotechnologies are required to make the invisible measurable, such as molecules, states, and interactions beyond current reach, and to capture multiple modalities simultaneously within the same cell. Prediction alone is insufficient. True understanding demands deep knowledge of the biological system,

careful analysis, and often small-scale validation experiments that cannot necessarily be scaled away. Experimental biology therefore remains central. In this space, artificial intelligence will act less as a replacement than as a co-scientist, suggesting and refining experiments, exploring alternative hypotheses, and supporting interpretation, while experimental biology continues to provide the indispensable anchor for discovery.

The trajectory is unmistakable. In the coming years, we are likely to witness the first true AI-driven breakthroughs. In hindsight, these may seem obvious, but only because a new lens will have made them visible. The combination of artificial intelligence and multimodal cell biology shows that the secrets of life yield not only to those who look closely but also to those who can synthesize across angles, perspectives, levels of organization, and scales. By bringing computation into the very core of experimentation, we are breathing life into data, creating models that learn alongside our understanding. The result is a more nuanced grasp of how cells cooperate to create organisms and how we might wisely guide them, whether in treating disease or engineering biology. A new era of cellular insight is dawning and its defining feature is that artificial intelligence is not just studying biology, it is becoming part of it – an active participant in humanity’s quest to know itself.

ABOUT THE AUTHORS

Charlotte Bunne is Assistant Professor in Computer and Communication Sciences and Life Sciences at the École Polytechnique Fédérale de Lausanne (EPFL). She is also part of the Swiss Institute for Experimental Cancer Research (ISREC) and the EPFL AI Center. She has published in such journals as *Nature Methods*, *Cell*, and *Bioinformatics*.

Aviv Regev, a Member of the American Academy since 2022, is Head and Executive Vice President of Genentech Research and Early Development (gRED) and founding Co-Chair of the Human Cell Atlas. She has also served as Chair of the Faculty and Core Member at the Broad Institute of MIT and Harvard, Professor of Biology at MIT, and as a Howard Hughes Medical Institute Investigator. She is the author or coauthor of over four hundred peer-reviewed articles and has recently published in such journals as *Nature*, *Science*, and *Cell*.

ENDNOTES

- ¹ Evan Z. Macosko, Anindita Basu, Rahul Satija, et al., “Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets,” *Cell* 161 (5) (2015): 1202–1214; Allon M. Klein, Linas Mazutis, Ilke Akartuna, et al., “Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells,” *Cell* 161 (5) (2015): 1187–1201; Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, et al., “Massively Parallel Digital

- Transcriptional Profiling of Single Cells,” *Nature Communications* 8 (2017): 14049; Charlotte Giesen, Hao A. O. Wang, Denis Schapiro, et al., “Highly Multiplexed Imaging of Tumor Tissues with Subcellular Resolution by Mass Cytometry,” *Nature Methods* 11 (4) (2014): 417–422; and Marlon Stoeckius, Christoph Hafemeister, William Stephenson, et al., “Simultaneous Epitope and Transcriptome Measurement in Single Cells,” *Nature Methods* 14 (9) (2017): 865–868.
- ² Jellert T. Gaublomme, Nir Yosef, Youjin Lee, et al., “Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity,” *Cell* 163 (6) (2015): 1400–1412; and Dominic Grün, Anna Lyubimova, Lennart Kester, et al., “Single-Cell Messenger RNA Sequencing Reveals Rare Intestinal Cell Types,” *Nature* 525 (7568) (2015): 251–255.
- ³ Aviv Regev, Sarah A. Teichmann, Eric S. Lander, et al., “The Human Cell Atlas,” *eLife* 6 (2017), <https://doi.org/10.7554/eLife.27041>.
- ⁴ *Ibid.*
- ⁵ Atray Dixit, Oren Parnas, Biyu Liu, et al., “Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens,” *Cell* 167 (7) (2016): 1853–1866; Britt Adamson, Thomas M. Norman, Marco Jost, et al., “A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response,” *Cell* 167 (7) (2016): 1867–1882; and Joseph M. Replogle, Reuben A. Saunders, Angela N. Pogson, et al., “Mapping Information-Rich Genotype-Phenotype Landscapes with Genome-Scale Perturb-Seq,” *Cell* 185 (14) (2022): 2559–2575.
- ⁶ Thomas M. Norman, Max A. Horlbeck, Joseph M. Replogle, et al., “Exploring Genetic Interaction Manifolds Constructed from Rich Single-Cell Phenotypes,” *Science* 365 (6455) (2019): 786–793; and Jennifer E. Rood, Anna Hupalowska, and Aviv Regev, “Toward a Foundation Model of Causal Cell and Tissue Biology with a Perturbation Cell and Tissue Atlas,” *Cell* 187 (17) (2024): 4520–4545.
- ⁷ Jennifer E. Rood, Aidan Maartens, Anna Hupalowska, et al., “Impact of the Human Cell Atlas on Medicine,” *Nature Medicine* 28 (12) (2022): 2486–2496.
- ⁸ Kimberly Siletti, Rebecca Hodge, Alejandro Mossi Albiach, et al., “Transcriptomic Diversity of Cell Types Across the Adult Human Brain,” *Science* 382 (6667) (2023): 175.
- ⁹ Daniel T. Montoro, Adam L. Haber, Moshe Biton, et al., “A Revised Airway Epithelial Hierarchy Includes CFTR-Expressing Ionocytes,” *Nature* 560 (7718) (2018): 319–324; Lindsey W. Plasschaert, Rapolas Žilionis, Rayman Choo-Wing, et al., “A Single-Cell Atlas of the Airway Epithelium Reveals the CFTR-Rich Pulmonary Ionocyte,” *Nature* 560 (7718) (2018): 377–381; and Eugene Drokhlyansky, Christopher S. Smillie, Nicholas Van Wittenberghe, et al., “The Human and Mouse Enteric Nervous System at Single-Cell Resolution,” *Cell* 182 (6) (2020): 1606–1622.
- ¹⁰ Rood, Maartens, Hupalowska, et al., “Impact of the Human Cell Atlas on Medicine.”
- ¹¹ Sai Ma, Bing Zhang, Lindsay M. LaFave, et al., “Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin,” *Cell* 183 (4) (2020): 1103–1116; and Stoeckius, Hafemeister, Stephenson, et al., “Simultaneous Epitope and Transcriptome Measurement in Single Cells.”
- ¹² *Ibid.*
- ¹³ Hari Shroff, Ilaria Testa, Florian Jug, and Suliana Manley, “Live-Cell Imaging Powered by Computation,” *Nature Reviews Molecular Cell Biology* 25 (6) (2024): 443–463; Gioele La Manno, Ruslan Soldatov, Amit Zeisel, et al., “RNA Velocity of Single Cells,” *Nature* 560 (7719) (2018):

- 494–498; Laleh Haghverdi, Maren Büttner, F. Alexander Wolf, et al., “Diffusion Pseudotime Robustly Reconstructs Lineage Branching,” *Nature Methods* 13 (10) (2016): 845–848; Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, et al., “Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming,” *Cell* 176 (6) (2019): 928–943; and Dominik Klein, Giovanni Palla, Marius Lange, et al., “Mapping Cells Through Time and Space with Moscot,” *Nature* 638 (8052) (2025): 1065–1075.
- ¹⁴ Joshua A. Weinstein, Aviv Regev, and Feng Zhang, “DNA Microscopy: Optics-Free Spatio-Genetic Imaging by a Stand-Alone Chemical Reaction,” *Cell* 178 (1) (2019): 229–241; Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, et al., “Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics,” *Science* 353 (6294) (2016): 78–82; Samuel G. Rodrigues, Robert R. Stickels, Aleksandrina Goeva, et al., “Slide-Seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution,” *Science* 363 (6434) (2019): 1463–1467; and Nianchao Qian and Joshua A. Weinstein, “Spatial Transcriptomic Imaging of an Intact Organism Using Volumetric DNA Microscopy,” *Nature Biotechnology*, March 27, 2025, <https://doi.org/10.1038/s41587-025-02613-z>.
- ¹⁵ Kok Hao Chen, Alistair N. Boettiger, Jeffrey R. Moffitt, et al., “Spatially Resolved, Highly Multiplexed RNA Profiling in Single Cells,” *Science* 348 (6233) (2015): 1360–1363; Chee-Huat Linus Eng, Michael Lawson, Qian Zhu, et al., “Transcriptome-Scale Super-Resolved Imaging in Tissues by RNA seqFISH+,” *Nature* 568 (7751) (2019): 235–239; Giesen, Wang, Schapiro, et al., “Highly Multiplexed Imaging of Tumor Tissues with Subcellular Resolution by Mass Cytometry”; and Yury Goltsev, Nikolay Samusik, Julia Kennedy-Darling, et al., “Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging,” *Cell* 174 (4) (2018): 968–981.
- ¹⁶ Regev, Teichmann, Lander, et al., “Science Forum: The Human Cell Atlas”; Rood, Maartens, Hupalowska, et al., “Impact of the Human Cell Atlas on Medicine”; and Rood, Hupalowska, and Regev, “Toward a Foundation Model of Causal Cell and Tissue Biology with a Perturbation Cell and Tissue Atlas.”
- ¹⁷ Macosko, Basu, Satija, et al., “Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets.”
- ¹⁸ *Ibid*; and Romain Lopez, Jeffrey Regier, Michael B. Cole, et al., “Deep Generative Modeling for Single-Cell Transcriptomics,” *Nature Methods* 15 (12) (2018): 1053–1058.
- ¹⁹ Andrew Butler, Paul Hoffman, Peter Smibert, et al., “Integrating Single-Cell Transcriptomic Data Across Different Conditions, Technologies, and Species,” *Nature Biotechnology* 36 (5) (2018): 411–420; Tim Stuart, Andrew Butler, Paul Hoffman, et al., “Comprehensive Integration of Single-Cell Data,” *Cell* 177 (7) (2019): 1888–1902; Brian Hie, Bryan Bryson, and Bonnie Berger, “Efficient Integration of Heterogeneous Single-Cell Transcriptomes Using Scanorama,” *Nature Biotechnology* 37 (6) (2019): 685–691; Ilya Korsunsky, Nghia Milard, Jean Fan, et al., “Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony,” *Nature Methods* 16 (12) (2019): 1289–1296; Charles Comiter, Eeshit Dhaval Vaishnav, Metamia Ciampricotti, et al., “Inference of Single Cell Profiles from Histology Stains with the Single-Cell Omics from Histology Analysis Framework (SCHAF),” *bioRxiv* (2023), <https://www.biorxiv.org/content/10.1101/2023.03.21.533680v1>; and Weiqing Chen, Pengzhi Zhang, Tu N. Tran, et al., “A Visual-Omics Foundation Model to Bridge Histopathology with Spatial Transcriptomics,” *Nature Methods* 22 (7) (2025): 1568–1582.
- ²⁰ Charlotte Bunne, Stefan G. Stark, Gabriele Gut, et al., “Learning Single-Cell Perturbation Responses Using Neural Optimal Transport,” *Nature Methods* 20 (11) (2023): 1759–1768; and Yusuf Roohani, Kexin Huang, and Jure Leskovec, “Predicting Transcriptional

- Outcomes of Novel Multigene Perturbations with GEARS,” *Nature Biotechnology* 42 (6) (2024): 927–935.
- ²¹ Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, et al., “On the Opportunities and Risks of Foundation Models,” arXiv (2021), <https://arxiv.org/abs/2108.07258>.
- ²² Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Association for Computational Linguistics, 2019), 4171–4186, <https://aclanthology.org/events/naacl-2019/#n19-1>; and Alec Radford, Jeffrey Wu, Rewon Child, et al., “Better Language Models and Their Implications,” OpenAI News, February 14, 2019, <https://openai.com/index/better-language-models>.
- ²³ Rood, Maartens, Hupalowska, et al., “Impact of the Human Cell Atlas on Medicine”; Haotian Cui, Chloe Wang, Hassaan Maan, et al., “scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI,” *Nature Methods* 21 (8) (2024): 1470–1480; Minsheng Hao, Jing Gong, Xin Zeng, et al., “Large-Scale Foundation Model on Single-Cell Transcriptomics,” *Nature Methods* 21 (8) (2024): 1481–1491; Xi Fu, Shentong Mo, Alejandro Buendia, et al., “A Foundation Model of Transcription Across Human Cell Types,” *Nature* 637 (8047) (2025): 965–973; and Graham Heimberg, Tony Kuo, Daryle J. DePianto, et al., “A Cell Atlas Foundation Model for Scalable Search of Similar Human Cells,” *Nature* 638 (8052) (2025): 1085–1094.
- ²⁴ Cui, Wang, Maan, et al., “scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI”; Hao, Gong, Zeng, et al., “Large-Scale Foundation Model on Single-Cell Transcriptomics”; Fu, Mo, Buendia, et al., “A Foundation Model of Transcription Across Human Cell Types”; Heimberg, Kuo, DePianto, et al., “A Cell Atlas Foundation Model for Scalable Search of Similar Human Cells”; and Christina V. Theodoris, Ling Xiao, Anant Chopra, et al., “Transfer Learning Enables Predictions in Network Biology,” *Nature* 618 (7965) (2023): 616–624.
- ²⁵ Ashish Vaswani, Noam Shazeer, Niki Parmar, et al., “Attention Is All You Need,” in *Advances in Neural Information Processing Systems* 30 (NIPS 2017), ed. Isabelle Guyon, Ulrike Von Luxburg, Samy Bengio, et al. (Neural Information Processing Systems Foundation, 2017).
- ²⁶ Cui, Wang, Maan, et al., “scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI”; and Johann Wenckstern, Eeshaan Jain, Yexiang Cheng, et al., “AI-Powered Virtual Tissues from Spatial Proteomics for Clinical Diagnostics and Biomedical Discovery,” arXiv (2025), <https://doi.org/10.48550/arXiv.2501.06039>.
- ²⁷ Leah V. Schaffer, Mengzhou Hu, Gege Qian, et al., “Multimodal Cell Maps as a Foundation for Structural and Functional Genomics,” *Nature* 642 (8066) (2025): 222–231.
- ²⁸ Comiter, Vaishnav, Ciampricotti, et al., “Inference of Single Cell Profiles from Histology Stains with the Single-Cell Omics from Histology Analysis Framework (SCHAF)” ; and Chen, Zhang, Tran, et al., “A Visual-Omics Foundation Model to Bridge Histopathology with Spatial Transcriptomics.”
- ²⁹ Richard J. Chen, Tong Ding, Ming Y. Lu, et al., “Towards a General-Purpose Foundation Model for Computational Pathology,” *Nature Medicine* 30 (3) (2024): 850–862; Thomas Hayes, Roshan Rao, Halil Akin, et al., “Simulating 500 Million Years of Evolution with a Language Model,” *Science* 387 (6736) (2025): 850–858; and Eric Nguyen, Michael Poli, Matthew G. Durrant, et al., “Sequence Modeling and Design from Molecular to Genome Scale with Evo,” *Science* 386 (6723) (2024), <https://doi.org/10.1126/science.ad09336>.

- ³⁰ Cui, Wang, Maan, et al., “scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI”; and Theodoris, Xiao, Chopra, et al., “Transfer Learning Enables Predictions in Network Biology.”
- ³¹ Heimberg, Kuo, DePianto, et al., “A Cell Atlas Foundation Model for Scalable Search of Similar Human Cells”; and Wenckstern, Jain, Cheng, et al., “AI-Powered Virtual Tissues from Spatial Proteomics for Clinical Diagnostics and Biomedical Discovery.”
- ³² Cui, Wang, Maan, et al., “scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI”; Hao, Gong, Zeng, et al., “Large-Scale Foundation Model on Single-Cell Transcriptomics”; and Fu, Mo, Buendia, et al., “A Foundation Model of Transcription Across Human Cell Types.”
- ³³ Heimberg, Kuo, DePianto, et al., “A Cell Atlas Foundation Model for Scalable Search of Similar Human Cells.”
- ³⁴ Chen, Zhang, Tran, et al., “A Visual-Omics Foundation Model to Bridge Histopathology with Spatial Transcriptomics.”
- ³⁵ Dixit, Parnas, Liu, et al., “Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens”; Adamson, Norman, Jost, et al., “A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response”; and Replogle, Saunders, Pogson, et al., “Mapping Information-Rich Genotype-Phenotype Landscapes with Genome-Scale Perturb-Seq.”
- ³⁶ Theodoris, Xiao, Chopra, et al., “Transfer Learning Enables Predictions in Network Biology.”
- ³⁷ Bunne, Stark, Gut, et al., “Learning Single-Cell Perturbation Responses Using Neural Optimal Transport”; Mohammad Lotfollahi, F. Alexander Wolf, and Fabian J. Theis, “scGen Predicts Single-Cell Perturbation Responses,” *Nature Methods* 16 (8) (2019): 715–721; and Roohani, Huang, and Leskovec, “Predicting Transcriptional Outcomes of Novel Multi-gene Perturbations with GEARS.”
- ³⁸ Kexin Huang, Serena Zhang, Hanchen Wang, et al., “Biomni: A General-Purpose Biomedical AI Agent,” bioRxiv (2025), <https://www.biorxiv.org/content/10.1101/2025.05.30.656746v1>; and Eric Wang, Samuel Schmidgall, Paul F. Jaeger, et al., “TxGemma: Efficient and Agentic LLMs for Therapeutics,” arXiv (2025), <https://doi.org/10.48550/arXiv.2504.06196>.
- ³⁹ Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al., “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, ed. Zoubin Ghahramani, Max Welling, Corinna Cortes, et al. (Neural Information Processing Systems Foundation, 2014); Jonathan Ho, Ajay N. Jain, and Pieter Abbeel, “Denoising Diffusion Probabilistic Models,” in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, ed. Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, et al. (Neural Information Processing Systems Foundation, 2020); and Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, et al., “Flow Matching for Generative Modeling,” paper presented at the Eleventh International Conference on Learning Representations, Kigali, Rwanda, May 1–5, 2023, <https://openreview.net/pdf?id=PqvMRDCJT9t>.
- ⁴⁰ Eeshit Dhaval Vaishnav, Carl G. de Boer, Jennifer Molinet, et al., “The Evolution, Evolvability and Engineering of Gene Regulatory DNA,” *Nature* 603 (7901) (2022): 455–463.
- ⁴¹ Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction*, 1st ed. (MIT Press, 1998).

- ⁴² Charlotte Bunne, Yusuf Roohani, Yanay Rosen, et al., “How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities,” *Cell* 187 (25) (2024): 7045–7063.
- ⁴³ Cui, Wang, Maan, et al., “scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI”; Hao, Gong, Zeng, et al., “Large-Scale Foundation Model on Single-Cell Transcriptomics”; Fu, Mo, Buendia, et al., “A Foundation Model of Transcription Across Human Cell Types”; and Heimberg, Kuo, DePianto, et al., “A Cell Atlas Foundation Model for Scalable Search of Similar Human Cells.”
- ⁴⁴ Huang, Zhang, Wang, et al., “Biomni: A General-Purpose Biomedical AI Agent”; and Wang, Schmidgall, Jaeger, et al., “TxGemma: Efficient and Agentic LLMs for Therapeutics.”
- ⁴⁵ Bunne, Roohani, Rosen, et al., “How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities.”
- ⁴⁶ Chen, Zhang, Tran, et al., “A Visual-Omics Foundation Model to Bridge Histopathology with Spatial Transcriptomics”; and Ståhl, Salmén, Vickovic, et al., “Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics.”
- ⁴⁷ Bunne, Stark, Gut, et al., “Learning Single-Cell Perturbation Responses Using Neural Optimal Transport”; Roohani, Huang, and Leskovec, “Predicting Transcriptional Outcomes of Novel Multigene Perturbations with GEARS”; and Huang, Zhang, Wang, et al., “Biomni: A General-Purpose Biomedical AI Agent.”