

Building the Drug Discovery Engine of the Future with AI-Empowered Nodal Biology

Anna Greka

Inspired by the visionary predictions of Jules Verne, this essay proposes that integrating artificial intelligence with a paradigm called “nodal biology” will revolutionize the discovery of treatments and cures. Current drug development is slow, costly, and failure-prone, constrained by the challenge of drug target identification. Solving the “cell perturbation prediction problem” – predicting how human cells respond to any disease-causing perturbation – is key to accelerating successful drug target identification. Nodal biology, the discovery of shared druggable mechanisms (nodes) among seemingly disparate diseases, offers a scalable approach to generate the high-quality data needed to train cell prediction AI models. As an example, a cargo receptor node was identified, linking dozens of genetic diseases and leading to a new drug candidate. The synergistic combination of human scientific intuition and AI-empowered nodal biology is essential for building the biomedical innovation engine of the future, ultimately accelerating treatments for all human diseases.

Writing more than 150 years ago, Jules Verne was celebrated for his adventure novels, incredible stories that integrated science fiction with explorations of uncharted territories. Many of his predictions ultimately materialized. He imagined submarines diving to the depths of the oceans and rockets taking humans to space. Quite remarkably, in *The Begum's Fortune* (1879), Verne incorporated biology into his writing. In this lesser-known novel, Verne explores the concept of artificial life and the future of biology. His characters endeavor to create a living organism through the manipulation of biological processes, thereby anticipating ideas later explored in genetic engineering. Verne's intuition and imagination allowed him to make visionary predictions, well beyond the scientific knowledge of his time.

As we now stand at the precipice of a new era for humanity empowered by artificial intelligence, can we – like a modern-day Jules Verne – envision a future when we can cure all human diseases? Can we imagine fully understanding the functional architecture of all human cells, then organs, and ultimately the entire body in health and disease? And can this knowledge enable effective methods for the prediction, control, and optimization of their inner workings? Success in this realm will bring

a revolution to biomedicine: the ability to predict and control the response of any human cell to any disease-related perturbation will promote our diagnostic capabilities (detect disruptions in specific cells with precision) and, more important, our capabilities to treat diseases in an exquisitely targeted fashion (direct drugs to specific cell disruptions), ultimately extending healthy lifespan. Is this possible or is it science fiction? And if it is possible, what will it take to get there?

Recent advances in AI have propelled us into a golden era in biomedicine. This is exemplified by the groundbreaking work of Demis Hassabis, John Jumper, and David Baker to use AI-empowered models to predict the structure of proteins and to learn to create entirely new proteins, solving the “protein structure prediction problem,” for which they were awarded the 2024 Nobel Prize in Chemistry. Armed with these tools, the scientific community can now pursue the development of therapies with greater speed and efficiency. Furthermore, AI-driven automation systems and advanced imaging approaches that use computer vision and related deep-learning models combined with new capabilities in genetic engineering collectively empower scientists to pursue biological experiments at unprecedented scale. In this essay, I delve into how AI is transforming biomedicine, the inherent challenges, and the promising future trajectories. This perspective is shaped by my own research as a physician-scientist and cell biologist, and I consider how the combination of a new paradigm called *nodal biology* can be fully harnessed with AI-enabled tools to revolutionize the specificity, speed, and efficiency with which we develop much needed new treatments.

Building innovative therapies is difficult. To this day, the inherent challenges and risks associated with the development of novel, targeted therapies remain substantial. As of 2026, the probability of a new drug progressing from initial human studies to regulatory approval remains at a mere 13 percent, a rate that has not changed in decades. Of the drugs that are approved, the average cost of research and development (including clinical trials) is more than \$1 billion each. And the timelines to approval are very long: it often takes more than ten years for a treatment to be approved and become available to patients. Additional factors, such as shifts in regulatory policies and geopolitical considerations, exert profound effects on our ability to develop new therapies. Nevertheless, improving on the low success rate in research and development continues to be the core challenge to building a productive innovation engine.

For all practical purposes, the innovation challenge can be broken into three simple but critical components:

1. Target identification (ID) – finding the right drug targets with therapeutic significance for as many diseases as possible;

2. Optimal drug development – finding the right therapeutic modality for each target; and
3. Clinical deployment – giving the right therapy to a well-characterized patient population with clearly defined clinical readouts (that is, “endpoints”) to maximize therapeutic benefit.

AI is already proving massively catalytic for components two and three, with astonishing progress taking place in real time. Regarding component two – optimal drug development – AI models that have solved the protein structure prediction problem such as AlphaFold are turbocharging rational drug design.¹ This means that we can now monitor and control the process of designing the right drug to fit into the right “pocket” within a specific, well-defined molecular target, and we can do it by visualizing the drug bound to its target in 3D on our computer screens (*in silico*). The time it takes to build new drugs (such as small molecules or therapeutic antibodies) is thereby being markedly shortened and made more efficient every day. Regarding component three – clinical deployment – AI is accelerating patient clinical data aggregation, biomarker identification, and patient stratification, which should empower more-successful clinical trials in years to come. Despite ongoing challenges with data aggregation and harmonization, coupled with scientific, legal, and ethical barriers, AI is likely to facilitate more efficiently run clinical trials in the years ahead. Components two and three are already benefiting from the tools of AI. While critical improvements are still needed and will undoubtedly continue, these two components are not what currently constrains our ability to innovate.

The most challenging part of our innovation engine continues to be component one, target ID: the fundamental work required to identify molecular targets in human cells that can be perturbed for therapeutic benefit (a given molecular target can be inhibited or activated). I strongly believe that addressing the foundational target ID component of the innovation challenge will require solving the “cell perturbation prediction problem”: that is, understanding and predicting the repertoire of responses by human cells to any given disease-causing perturbation. Solving this challenge is likely to unleash an innovation engine ultimately capable of addressing human diseases at scale.

As a physician-scientist, I have training in both medicine and scientific research, and my core objective is to translate fundamental laboratory discoveries into therapeutic strategies that can benefit patients: that is, to treat all diseases for all patients, with utmost expediency. Given the complexities and difficulties previously outlined, how can we realistically aspire to reach this goal? I am optimistic that progress will come from creatively combining our finest current and future AI tools with human intelligence, or what I call “scientific intuition.” Marshaling our resources to solve the cell perturbation prediction problem will bring us closer to achieving the lofty goal of treating all human diseases.

The key bottleneck in solving this foundational problem is obtaining cell perturbation data of large enough magnitude and quality to train cell-prediction AI models. Deep-learning models require data of a scale that is unprecedented in cell biology. My current work is focused on generating such key datasets by starting with human genetic diseases, specifically monogenic disorders in which a mutation in the coding sequence of a given gene leads to a defined disease phenotype (for example, cystic fibrosis, a debilitating lung disease caused by specific genetic defects in a gene called *CFTR*). We have selected to begin with monogenic disorders for which the genetic cause is known because we can readily introduce these human pathogenic mutations into any cell of our choice to develop a faithful disease model using genetic engineering approaches like base editing.² Modeling thousands of human pathogenic mutations (introducing thousands of genetic perturbations) in diverse human cell types (neurons, cardiac myocytes, liver cells, and so on) and interrogating them with robust scalable readouts (single cell transcriptomics or proteomics, reading the repertoire of mRNA messages or proteins found in each of these cells) will generate a high-quality dataset with rigorous controls. In other words, this approach will generate data of the magnitude and quality needed to train AI models that can help us solve the cell perturbation prediction problem.

Addressing the cell perturbation prediction problem will be greatly facilitated by *nodal biology*. Over the last few years, my team and I have discovered that many clinically unrelated genetic disorders share common – most often uncharacterized – mechanisms, which we call *nodal pathways*. In other words, seemingly disparate diseases (for example, an inherited form of blindness, a familial form of Alzheimer’s disease, and a genetic kidney disease) ultimately converge on the same node, representing a shared druggable target (for example, TMED cargo receptors).³ We thus introduced the concept of *nodal biology*: the discovery of nonobvious, previously unseen connections between seemingly distinct diseases. This paradigm allows us to design large-scale interrogations of all monogenic diseases in a massively parallel manner to identify the full complement of biological nodes that connect subsets of these diseases in previously unseen nodal clusters.

How many monogenic diseases are there? With advances in DNA sequencing technologies, we can now readily detect mutations in more than eight thousand (out of a total of twenty thousand) human genes that drive these devastating chronic conditions; however, even as the pool of known mutations is growing, fewer than 5 percent of them have an available treatment. Genetic diseases collectively affect thirty million people across the United States, most arising during childhood, and more than four hundred million people globally.⁴ A key reason for the significant lack of therapies is that most research on genetic diseases has remained siloed

among small teams of scientists painstakingly working on one gene/one disease at a time. In other words, the fundamental challenges associated with the “one-by-one genetic defect” approach are: 1) scale – work happens in individual laboratories and few discoveries are able to reach critical mass in terms of validation and downstream clinical impact; 2) investability – the patient populations are fragmented into small subsets that do not attract sufficient biopharmaceutical interest; and 3) biological convergence – the one-by-one approach fails to leverage recent insights that different diseases converge on the same biological pathways. Nodal biology can address these challenges.

To be clear, the concept of convergent disease mechanisms is not new. In some cases, taking a known drug with a well-defined target from one disease into several others that share the same mechanism (what is sometimes called “indication extension”) is relatively obvious. For example, such an approach targeting the alternative complement pathway – an immune mechanism designed to destroy invading human pathogens – is an approved therapy that is helping patients with chronic and devastating kidney, blood, neuromuscular, and eye diseases.⁵ Where nodal biology truly shines is in uncovering nonobvious, previously unseen connections between seemingly different diseases. Our team discovered such a biological node, a point of convergence on a single druggable target at which mechanistic connections were identified between dozens of clinically disparate but connected genetic diseases affecting the eye, the kidney, the brain, and more.⁶ Given that potentially hundreds more nodes underlie approximately eight thousand human genetic diseases caused by up to one hundred thousand different mutations, bold innovation is crucial. To pave the path to much needed nodal therapies, we must revolutionize our ability to identify nodal pathways and druggable targets by deploying cell perturbation studies at scale. A bold vision for the future is to uncover as many druggable nodes as possible in a massively parallel fashion.

Gaining insights from nodal biology will require the use of scalable tools, including: genome editing techniques that can allow us to introduce disease-causing mutations in human cells at scale and thus to generate cellular models of thousands of different genetic diseases in parallel; high-throughput methods of reading the effects of thousands of genetic perturbations into a collection of human cells; and the ability to turn up or down any gene of interest in parallel in a collection of human cells.⁷ These tools are readily available today and can be scaled to serve the needs of any nodal biology project.

Nodal biology will also require computational methodologies (many still in development) to collect and analyze unprecedented amounts of human cell perturbation data to identify convergent nodes.⁸ We can subsequently use these data to train cell-perturbation AI models in an iterative fashion, bringing us closer to solving the cell prediction problem: that is, the ability to *predict* the response of human cells to a given perturbation that was not included in the training data – all *in silico*.

How would nodal biology solving the cell prediction problem – thereby securing component one of the innovation engine of the future – enhance our drug target discovery process? Imagine that scientists have uncovered multiple candidates as the potential drug target that could cure a devastating genetic disease. Without having to do expensive and time-consuming studies that could take years and millions of dollars, can they use our nodal biology–derived cell model to predict with confidence which of their candidates is the best drug target? Can they predict if turning the target up or down will be required for therapeutic benefit? And can they predict if drugging this target will be toxic to other cells or organs in the human body? Solving the cell prediction problem and building accessible models for the scientific community – much like AlphaFold for predicting protein structures – will vastly accelerate our innovation engine by empowering scientists to test multiple potential targets *in silico* with high confidence, ease, and efficiency.

To illustrate the arc of discovery in nodal biology, I will share an example from my own work. This discovery exemplifies how scientific intuition – derived from deep knowledge of cell biology and years of working in the laboratory – led us to a nodal target; and how combining this insight with AI tools like AlphaFold drove the discovery of a drug candidate for the treatment of several devastating diseases.

Several years ago, our team was tasked with solving a perplexing medical mystery. As is often the case with such inquiries, the stakes were considerable: the mystery involved a large family that has been suffering with kidney failure and untimely death for many generations. What, then, was the underlying cause of their affliction? Through concerted efforts, the etiology of this disease was finally revealed to be a mutation, specifically a single cytosine (C) misspelling in the DNA. This singular genetic error is sufficient to induce severe kidney damage in these patients.⁹

As a brief reminder, DNA is the genetic code, comprising three billion letters of information meticulously packaged within the nuclei of each of approximately thirty trillion cells in our bodies. Within cells, DNA is transcribed into RNA, and RNA, in turn, is translated into proteins, which are the functional units employed by our cells to execute their intricate functions.

Tragically, in the case of the family in question, the solitary DNA misspelling leads to the production of a malformed, dysfunctional, and toxic protein that progressively accumulates within their cells, ultimately resulting in cell death and kidney failure. In the absence of cell-prediction models (that is, in the absence of component-one solutions), we had no choice but to pursue answers by what I call “molecular sleuthing,” otherwise defined as the judicious integration of our most advanced experimental tools with scientific intuition. After a few years of such investigations, we finally found that this toxic protein accumulates within cells due to the involvement of another molecule, a “cargo receptor.”¹⁰ These cargo recep-

tors sequester the toxic protein and become ensnared in a cellular “traffic jam.” Over time, the toxic cargo accrues, ultimately leading to cell death. In essence, after more than five years of hard work, we found that the cargo receptors were the right (therapeutically relevant) drug target for this disease.

Based on these insights, our objective was to identify a drug capable of interfering with the cargo receptors and directing them to release their toxic cargo toward the cell’s waste disposal facility, known as the lysosome. We embarked on the search for a suitable drug candidate. We had some initial hints that a compound could be effective in removing toxic cargo, but that candidate proved relatively ineffective and raised concerns about potential toxicity to brain cells (otherwise known as “off-target” toxicity). Fortunately, by this time in the project, we were armed with AlphaFold, a powerful tool for drug discovery (and an excellent solution for component two of our innovation engine). Using AlphaFold, we were able to visualize the precise arrangement of each atom composing the cargo receptor target. This capability, in turn, facilitated the rational design of small molecules that can fit precisely into the exact right pocket in the cargo receptor target to achieve the desired specific effect: in this instance, to direct the cargo receptors and their toxic cargo toward the lysosome. Consequently, within an unusually short timeframe, we identified a new drug candidate that fits precisely into the designated pocket of this cargo receptor. When tested in human cells and mice, it performed exactly as intended: it steered the toxic cargo toward the lysosome, thereby clearing the cells of the toxic traffic jam. It is satisfying to know that after several years of work, this discovery is now advancing toward clinical trials in patients – and the family that inspired all this work is now our close partner in driving this program to the clinic.

Our investigations did not stop there. We subsequently discovered that the same drug target, the cargo receptors, are implicated in many different genetic diseases that involve the accumulation of different toxic proteins in different organs. These are severe disorders such as blindness, liver failure, and Alzheimer’s disease. In total, we estimate that more than fifty diseases, involving millions of patients, could ultimately be addressed by targeting the same cargo receptor system.

In essence, we discovered a nonobvious “node” that illuminated previously unknown connections between multiple unrelated genetic diseases, which in turn led to the realization that nodal biology could hold the key to accelerating the discovery of drug targets for the treatment of heretofore incurable diseases. In this new paradigm, we defined a node as a point of convergence on a single druggable target, enabled by the identification of previously unrecognized, shared biological pathways among seemingly disparate but now newly interconnected diseases. This work led us to ask the next fundamental question: Can we identify all the nodes that underpin all human genetic diseases? This forms the foundation for our current and future work on AI-empowered nodal biology.

Beyond monogenic diseases, solving the cell prediction problem requires us to perturb human cells with molecules and factors that drive devastating disorders such as obesity, cancer, autoimmune conditions, mental health conditions, and diseases of aging. We also need to scale our investigations to interrogating billions of human cells in parallel with readouts that capture disease-specific changes in cellular mRNA, proteins, and metabolites. Furthermore, progress in this grand challenge in biomedicine will require a shift to a new biomedical research model that incorporates AI-enabled robots and automation contributing alongside AI-empowered scientists. In the final analysis, the synergistic combination of scientific intuition and current and future AI tools can propel an unprecedented acceleration of our innovation engine toward new targeted therapies. I am confident that the next few years will prove catalytic, as we progress from identifying a few druggable nodes to progressively *predicting* which nodes to target until we can readily treat all human diseases with speed and precision.

As we pursue the merger of nodal biology with AI to build the drug discovery innovation engine of the future, it is worth noting that the intrinsic complexity of biological data poses significant challenges. Rigorous validation processes and iterative feedback loops are critical to ensure the reliability and accuracy of AI-driven insights. A key limitation lies in the ability to train deep-learning AI models with data of consistently high quality. Biological data often present with inherent variability due to experimental conditions, individual differences, and technological limitations, making it difficult to acquire sufficiently comprehensive datasets for robust model training. For example, a cellular perturbation dataset in Cambridge, Massachusetts, may not be identical to a dataset produced with the same protocol in Cambridge, England. The presence of noise, missing values, and inconsistencies within datasets further exacerbates the challenge, requiring advanced preprocessing techniques to clean and prepare the data for optimal model performance. Another potential consideration is the challenge of generating a massive dataset at a single lab/location with highly rigorous procedures before the data can be widely shared to train cell perturbation models. Without addressing these fundamental issues of biological data complexity and quality, the full potential of deploying cell-prediction AI models may remain constrained.

Nodal biology anchored on genetic diseases is limited in that it does not address a critical driver for many other human disorders: namely, the fact that disease processes also occur in the cross talk between different cell types, such as between immune cells interacting with brain cells or liver cells or heart cells. Developing the tools to measure perturbations in cell-cell interactions and more complex systems such as intact organs may ultimately be needed before we can train models that can give us predictive power over complex processes such as cancer metastasis or the interactions underlying schizophrenia. Our innovation engine will benefit from an exploration of systems of higher complexity once our nodal biology

paradigm for genetic diseases has given us the foundational tools to take on new challenges.

Advances in any complex scientific field, particularly one as multifaceted as the intersection of AI and biomedicine, require a deeply integrated, interdisciplinary approach. To achieve breakthroughs in building the innovation engine of the future, it is paramount to train a workforce that integrates expertise in AI/machine learning, biology, and medicine. Currently, a significant hurdle lies in the cross talk between these disciplines: each often operates with its own specialized lexicon and conceptual frameworks. Ongoing and future programs in our laboratories and universities should therefore focus on enhancing the opportunities for deep collaborative work between AI specialists and biomedical scientists.

AI models are only as good as the data they are trained on and the biological relevance of the questions they are designed to address. Biologists and physicians possess domain knowledge, understanding the intricacies of biological systems, experimental design, and the clinical implications of research findings. AI specialists offer the computational tools and analytical techniques to process vast datasets and build predictive models. The challenge lies in synthesizing these strengths, ensuring that AI algorithms are developed with a deep understanding of biology and human disease and that biological questions are framed in a way that can be effectively tackled by AI.

In Douglas Adams's *The Hitchhiker's Guide to the Galaxy*, the supercomputer Deep Thought, after eons of computation, reveals the answer to the "Ultimate Question of Life, the Universe, and Everything": it is "42." This answer was a huge disappointment to those awaiting a response – but they had not asked "the right question." The true challenge in AI-empowered biomedical innovation lies not in obtaining an answer, but in identifying and formulating the relevant question itself. This narrative serves as a poignant and cautionary tale for our engagement with AI in biomedicine. The intricate and highly dynamic nature of biological systems presents a humbling challenge. Unlike well-defined mathematical problems, biological questions are multilayered, interconnected, and subject to numerous variables. In this environment, AI, with its unparalleled capacity for data processing and pattern recognition, can only accelerate our innovation engines if we, the scientists, are exceptionally adept at posing the most relevant questions. Without precisely framed inquiries, AI might generate a deluge of statistically significant correlations that lack biological meaning or clinical utility – a "42" without an understanding of the question.

AI may never be able to fully replace human scientific intuition, nor the cautious validation work that has secured the discovery of clinically successful drugs in decades past. Nevertheless, rigorously trained AI models can serve as an indispensable scientific partner and a powerful instrument that amplifies and enriches our scientific capabilities, enabling us to analyze vast datasets, make good predictions,

and thus generate hypotheses at a scale and speed impossible for humans alone. This collaborative cross talk between human intelligence inventing new paradigms (such as nodal biology) and artificial intelligence will help us find the right targets or nodes connecting many human diseases (component one), build the best drug for each of these nodal targets (component two), and ultimately give each drug to the right patients for maximum benefit (component three). Of course, no drug can be developed and moved into human clinical studies without thorough prior validation using complex systems such as human organoids (human mini-organs grown in the laboratory) or animals.¹¹ However, the speed and cost of these validation studies can be markedly reduced if we have higher confidence in the drug targets that are moved forward to validation. If accurate predictions using AI can help us increase the success rate of target validation studies and the clinical trials that follow, such that overall drug development success rates increase from 13 percent today to above 30 percent in the decade ahead, the benefits to humanity will be incalculable, getting us closer than ever to our goal of curing as many diseases as possible in our lifetimes.

I began this essay with the ambition to conceptualize a framework with which to understand the inner workings of human cells in health and disease and enable effective methods to *predict* their responses to any disease-causing perturbation. I am convinced that solving the cell perturbation prediction problem will allow us to ultimately integrate it into models of entire organs and eventually the human body. Success in these efforts will revolutionize biomedicine: the ability to predict and control the behavior of human cells will radically enhance our ability to diagnose diseases (detect disruptions in cells) and, more important, treat them in an exquisitely targeted fashion (direct drugs to specific cell disruptions). This is how we will build the drug discovery innovation engine of the future, starting today. After all, as Jules Verne is quoted as saying, “all that is impossible remains to be accomplished.”

AUTHOR'S NOTE

This essay was researched in collaboration with Gemini 3 Pro. Every word is endorsed by the author.

ABOUT THE AUTHOR

Anna Greka is a physician-scientist, a Professor at Harvard Medical School and Mass General Brigham, and a Core Institute Member of the Broad Institute of MIT and Harvard, where she also serves as Founder and Chair of the Ladders to Cures Accelerator and a Member of the Institute's Executive Leadership Team. Her work has been published in *Science*, *Cell*, and *Nature*.

ENDNOTES

- ¹ John Jumper, Richard Evans, Alexander Pritzel, et al., "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature* 596 (7873) (2021): 583–589; and Josh Abramson, Jonas Adler, Jack Dunger, et al., "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3," *Nature* 630 (8016) (2024): 493–500.
- ² Nicole M. Gaudelli, Alexis C. Komor, Holly A. Rees, et al., "Programmable Base Editing of A·T to G·C in Genomic DNA without DNA Cleavage," *Nature* 551 (7681) (2017): 464–471.
- ³ Moran Dvela-Levitt, Maria Kost-Alimova, Maheswarareddy Emani, et al., "Small Molecule Targets TMED9 and Promotes Lysosomal Degradation to Reverse Proteinopathy," *Cell* 178 (8) (2019): 521–535; and Jillian L. Shaw, Juan Lorenzo Pablo, and Anna Greka, "Mechanisms of Protein Trafficking and Quality Control in the Kidney and Beyond," *Annual Review of Physiology* 85 (2023): 407–423.
- ⁴ Petra Kaufmann, Anne R. Pariser, and Christopher Austin, "From Scientific Discovery to Treatments for Rare Diseases—The View from the National Center for Advancing Translational Sciences—Office of Rare Diseases Research," *Orphanet Journal of Rare Diseases* 13 (2018): 196; *The Lancet Global Health*, "The Landscape for Rare Diseases in 2024," *The Lancet Global Health* 12 (3) (2024): e341; and Global Genes, "Numbers: RARE Disease Facts," <https://globalgenes.org/rare-disease-facts> (accessed December 22, 2025).
- ⁵ Vlado Perkovic, Jonathan Barratt, Brad Rovin, et al., "Alternative Complement Pathway Inhibition with Iptacopan in IgA Nephropathy," *The New England Journal of Medicine* 392 (6) (2025): 531–543; and Kate McKeage, "Ravulizumab: First Global Approval," *Drugs* 79 (3) (2019): 347–352.
- ⁶ Dvela-Levitt, Kost-Alimova, Emani, et al., "Small Molecule Targets TMED9 and Promotes Lysosomal Degradation to Reverse Proteinopathy"; and Shaw, Lorenzo Pablo, and Greka, "Mechanisms of Protein Trafficking and Quality Control in the Kidney and Beyond."
- ⁷ Gaudelli, Komor, Rees, et al., "Programmable Base Editing of A·T to G·C in Genomic DNA without DNA Cleavage"; Atray Dixit, Oren Parnas, Biyu Li, et al., "Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens," *Cell* 167 (7) (2016): 1853–1866; and Molly Gasperini, Andrew J. Hill, José L. McFaline-Figueroa, et al., "A Genome-Wide Framework for Mapping Gene Regulation via Cellular Genetic Screens," *Cell* 176 (1–2) (2019): 377–390.
- ⁸ Yang Xu, Stephen Fleming, Matthew Tegtmeyer, et al., "Modeling Interpretable Correspondence between Cell State and Perturbation Response with CellCap," bioRxiv (2024), <https://doi.org/10.1101/2024.03.14.585078>; and Mehrtash Babadi, Stephen Fleming, Yang Xu, et al., "Cellarium AI Platform," Cellarium, Broad Institute of MIT and Harvard, <https://cellarium.ai/tool/cellarium-ai-platform> (accessed December 22, 2025).

- ⁹ Andrew Kirby, Andreas Gnirke, David B. Jaffe, et al., “Mutations Causing Medullary Cystic Kidney Disease Type 1 Lie in a Large VNTR in MUC1 Missed by Massively Parallel Sequencing,” *Nature Genetics* 45 (3) (2013): 299–303.
- ¹⁰ Dvela-Levitt, Kost-Alimova, Emani, et al., “Small Molecule Targets TMED9 and Promotes Lysosomal Degradation to Reverse Proteinopathy”; and Le Xiao, Xiong Pi, Alissa C. Goss, et al., “Molecular Basis of TMED9 Oligomerization and Entrapment of Misfolded Protein Cargo in the Early Secretory Pathway,” *Science Advances* 10 (38) (2024): eadp2221.
- ¹¹ Ayshwarya Subramanian, Eriene-Heidi Sidhom, Maheswarareddy Emani, et al., “Single Cell Census of Human Kidney Organoids Shows Reproducibility and Diminished Off-Target Cells after Transplantation,” *Nature Communications* 10 (1) (2019): 5462.