Cell
Leading Edge

Voices

The impact of AI on research

Large and complex datasets have made artificial intelligence (AI) an invaluable tool for discovery across biological research. We asked experts how AI has impacted their work. Their experiences and perspectives offer thoughtful insights into potential offered by AI for their fields.

Aubrey Weigel
HHMI, Janelia Research Campus

Extending beyond what the eye can see

Over the last decade or more, microscopes have given researchers the ability to image larger samples, with higher speed and sensitivity. As the accessibility and functionality of these technologies continue to improve, the abundance of big data—oozing with feature-rich information—also increases. This is a bittersweet reality, unearthing many computational challenges and exciting new biological questions to be answered. Enter artificial intelligence (AI).

Frequently, the first steps in extracting information from these feature-rich datasets involve some form of classification, e.g., graphical annotations and pixel-based segmentation. These workflows often heavily rely on manual input and are therefore very time consuming, laborious, and often infeasible to complete. For example, dense segmentations of organelles in a cubic micron electron microscopy volume can take one person 2 weeks, extrapolating to an entire cell ~2,500 times the size, this task would take a century. AI makes this task possible in a fraction of the time, allowing researchers to subsequently analyze the data to answer their question(s) of interest.

The real excitement that AI brings to the field, however, is not just "speeding up" time-consuming human tasks but revealing questions we as researchers didn't even know to ask. Recent advances suggest that AI can use visual features inaccessible to humans. For instance, AI can categorize features in images to predict a functional property of the sample. Key features used for classification can even be visualized using "Explainable AI", an approach that can inform biologists on what the computer is "seeing" and allows hypothesis generation. We keep being surprised at what AI can extract from seemingly uninformative texture, hallmarking the potential of AI for biological discovery.

Carlos Caldas
Cancer Research UK Cambridge Institute

AI to deliver precision cancer medicine 2.0

AI through machine learning and deep learning enables computers to mimic the human ability to transform data into knowledge. In cancer biomedicine, there is an unprecedented deluge of data from bulk analyses of tumors at the genomic, proteomic, and metabolomic levels; single-cell analyses of both malignant cells and the normal cells of the tumor microenvironment; and spatially resolved analysis of tumors in 2D and 3D. These multi-dimensional data need to be interpreted and integrated to enable systems level views of tumor ecosystems and to deliver the next level of precision cancer medicine. For my lab and the field as a whole, AI is already an integral part of our research. We use AI to analyze linear sequencing data and predict cancer drivers and identify mutation signatures, to predict the consequences of sequence variants in 3D genome organzsation, to do *in silico* TCR/BCR characterization and predict their pairing with cognate neo-antigen, and to design new therapeutics. AI will also help convergence analysis of single-cell data and deconvolution of cellular composition from bulk data. Spatial analysis of molecularly annotated single cells *in situ* will require AI to objectively describe tissue neighborhoods, niches, and inter-cell communication. Transfer learning, the strengthening of machine learning models by applying what was learned from a different task, combined with deep learning will untap functional layers from routine tissue morphology captured in H&E tissue sections. AI is key to new systems level insights into tumor biology and to the next generation of diagnostic, prognostic, and predictive biomarkers.

Axel Meyer
University of Konstanz

Perspective on AI in evolutionary biology and genomics

Evolutionary biology and genomics have always been scientific disciplines that required computation. At the end of the 1980s, as graduate students and postdocs with Allan C. Wilson in the Biochemistry Department in Berkeley, we were the first to use PCR to produce DNA sequences to address important questions in molecular evolution. We used to read DNA sequences from autoradiographs of polyacrylamide gels, typing them into a keyboard by hand. Every day for many, many hours. We could see raw genetic variation in front of our eyes—how an individual differed from another. Every base pair meant something, and I could recognize genes instantly by the patterns of Gs, As, Ts, and Cs on the gel in front of my eyes. Now, I need reading glasses and, of course, much more sophisticated methods to analyze the avalanche of genetic data that can be obtained cheaply. Reading in DNA sequences by hand and analyzing them with, by today's standards, laughably simple algorithms seems quaint in hindsight, but it gave you a feeling for the veracity of the data. In the last three decades, all technological advances in DNA-sequencing technology and computational biology propelled our field forward. AI and machine learning tools are now prevalent everywhere in my research from population genomic patterns of historical population parameter estimation to predictions of future species' distribution due to global warming. But, we are now more removed from the raw data since today many layers of computational analysis are required to extract signals or patterns—that then still need to be interpreted by a human mind. In the end, I would argue that, despite the valuable new tools, biological knowledge, intuition, experience, and a skeptical eye for nonsense remain the most important skills a scientist must have for interpreting data.

Samantha A. Morris
Washington University School of Medicine

Dissecting cell identity with machine learning

As single-cell biology has gained sophistication across various modalities, the field is poised to leverage the immense amount of data generated by these approaches to gain new mechanistic insights into cell identity and behavior. Indeed, a new class of artificial intelligence-based methods is emerging to predict single-cell phenotypes in response to various perturbations, such as cell-cell signaling cues. My research program focuses on how gene regulatory networks (GRNs) control cell identity in development and reprogramming, and we are directly benefitting from these new computational strategies. We use machine learning to integrate single-cell gene expression and chromatin accessibility with prior biological knowledge of transcription factor (TF) binding motifs to construct GRN models. This construction of an interpretable network enables *in silico* TF perturbation, such as knockout or overexpression, simulating the subsequent changes in cell identity without requiring extensive training data. As a result, we can systematically screen through TFs to recover novel, context-dependent regulators of cell identity that we experimentally validate. We are applying this approach to understand how GRNs are "rewired" during reprogramming, identifying novel TFs to increase the yields of target cell types. We also use this method to better our understanding of gene regulation in developing organisms, from zebrafish to mouse, to human. Soon, I expect that such *in silico* perturbation methods will find broad application, revealing new insights into the establishment of cell identity in both established and emerging biological paradigms.