Deep learning on cell signaling networks establishes AI for single-cell biology
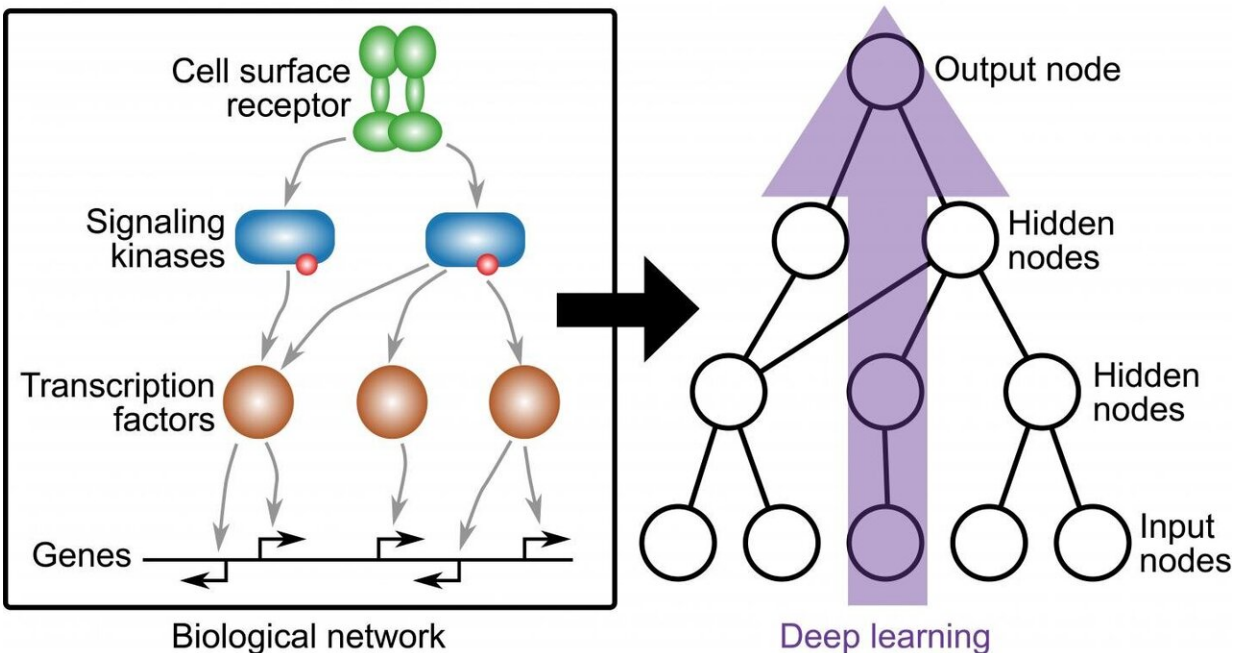
August 4 2020

Conceptual outline of knowledge-primed neural networks (KPNNs). KPNNs are artificial neural networks whose structure is based on biological knowledge, thereby enabling interpretable deep learning on biological networks. Credit: Nikolaus Fortelny/CeMM

Computer systems that emulate key aspects of human problem solving are commonly referred to as artificial intelligence (AI). This field has

seen massive progress over the last years. Most notably, deep learning enabled groundbreaking progress in areas such as self-driving cars, computers beating the best human players in strategy games (Go, chess), computer games, and in poker, and initial applications in diagnostic medicine. Deep learning is based on artificial neural networks—networks of mathematical functions that are iteratively reorganized until they accurately map the data describing a given problem to its solution.

In biology, deep learning has established itself as a powerful method to predict phenotypes (i.e., observable characteristics of cells or individuals) from genome data (for example gene expression profiles). Deep learning is usually a "black box" method: Neural networks are very powerful predictors when provided with enough training data. For example, they have been used to predict cell type from gene expression profiles, and protein structures from DNA sequence data. But standard neural networks cannot explain the learnt relationship of inputs to outputs in a human-understandable way. For this reason, deep learning has so far contributed little to advancing our mechanistic understanding of molecular functions within cells.

To address this lack of interpretability, CeMM Postdoctoral Fellow Nikolaus Fortelny and CeMM Principal Investigator Christoph Bock pursued the idea of performing deep learning directly on biological networks, instead of the generic, fully connected artificial neural networks used in conventional deep learning. They established "knowledge-primed neural networks" (KPNNs) that are based on signaling pathways and gene-regulatory networks. In KPNNs, each node corresponds to a protein or a gene, and each edge has a mechanistic biological interpretation (e.g., protein A regulates the expression of gene B).
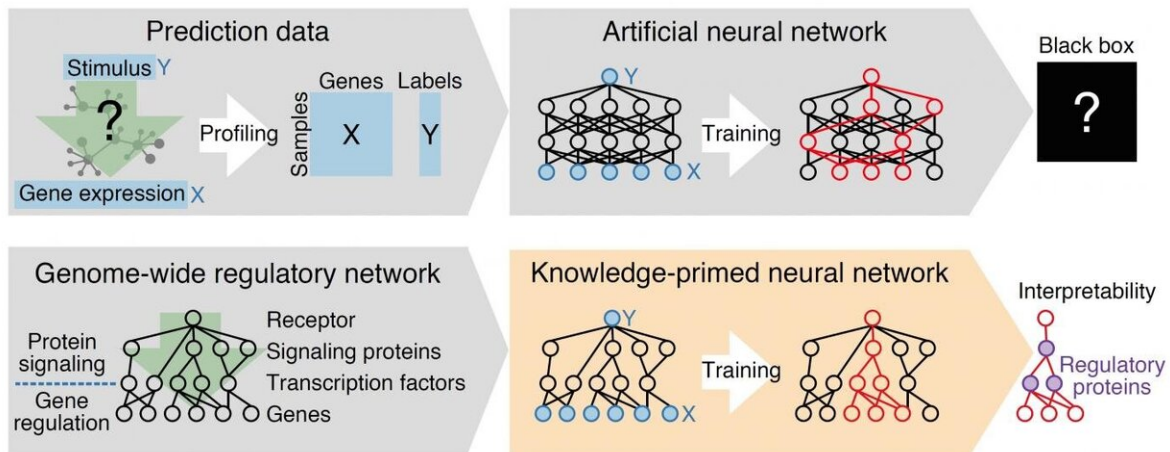
Detailed outline of the method: Artificial neural networks transform inputs (X) into outputs (Y), for example gene expression data (X) into phenotypes (Y) with high prediction accuracy. KPNNs incorporate knowledge of biological networks into deep learning, thus enabling interpretability in what is otherwise a black-box algorithm. Credit: Nikolaus Fortelny/CeMM

The CeMM researchers show in their new study published in *Genome Biology* that deep learning on biological networks is technically feasible and practically useful. By forcing the deep learning algorithm to stay close to gene-regulatory processes that are encoded in the biological network, KPNNs create a bridge between the power of deep learning and our rapidly growing knowledge and understanding of complex biological systems. As a result, the approach provides concrete insights into the investigated biological systems, while maintaining high prediction performance. This powerful new methodology uses an optimized approach for deep learning, which stabilizes node weights in the presence of redundancy, enhances the quantitative interpretability of node weights, and controls for the uneven connectivity inherent to biological networks.

CeMM researchers demonstrated their new KPNN method on large single-cell datasets, including a compendium of 483,084 single-cell transcriptomes for immune cells established by the Human Cell Atlas consortium. In this dataset, the scientists discovered unexpected diversity in the cell-type-defining regulatory networks between immune cells from bone marrow and cord blood.

The KPNN method combines the predictive power of deep learning and its ability to infer activity levels across multiple hidden layers with the functional interpretability of biological networks. KPNNs are particularly useful for the single-cell RNA-seq data, which are generated at massive scale using single-cell sequencing assays. Moreover, KPNNs are broadly applicable to other areas of biology and biomedicine where relevant prior knowledge can be represented as networks.

The predictions and biological insights obtained by KPNNs will be useful for dissecting cell signaling and gene regulation in health and disease, for identifying novel drug targets, and for deriving testable biological hypotheses from single-cell sequencing data. More generally, the study illustrates the future impact that artificial intelligence and deep learning, will have on mechanistic biology as the scientific community learns how to make AI results biologically interpretable.

More information: Nikolaus Fortelny et al, Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data, *Genome Biology* (2020). DOI: 10.1186/s13059-020-02100-5