

Deep learning on cell signaling networks establishes interpretable AI for single-cell biology

Researchers at CeMM, the Research Center for Molecular Medicine of the Austrian Academy of Sciences, have developed knowledge-primed neural networks (KPNNs), a new method that combines the power of deep learning with the interpretability of biological network models. KPNNs learn multiple layers of protein signaling and gene regulation from single-cell RNA-seq data, thereby providing a much-needed boost in our ability to convert massive single-cell atlas data into biological insights. These findings have now been published in the renowned scientific journal *Genome Biology*.

(Vienna, 4 August 2020) 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).

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.

Attached pictures:

1. The study authors Christoph Bock and Nikolaus Fortelny (© Klaus Pichler / CeMM)
2. 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. (© Nikolaus Fortelny / CeMM)
3. 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. (© Nikolaus Fortelny / CeMM)

The study “Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data” was published in *Genome Biology* on 3 August 2020. DOI: [10.1186/s13059-020-02100-5](https://doi.org/10.1186/s13059-020-02100-5)

Authors:

Nikolaus Fortelny and Christoph Bock

Funding:

This study was co-funded by an Austrian Science Fund (FWF) Special Research Programme grant (FWF SFB F 6102-B21), a New Frontiers Group award of the Austrian Academy of Sciences and the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant Agreement No 679146 awarded to Christoph Bock). Nikolaus Fortelny was supported by a fellowship from the European Molecular Biology Organization (EMBO ALTF 241-2017).

Christoph Bock joined CeMM as Principal Investigator in 2012. He pursues interdisciplinary research aimed at understanding the epigenetic and gene-regulatory basis of cancer, and advancing precision medicine with genomics technology. His research group combines experimental biology (high-throughput sequencing, epigenetics, CRISPR screening, synthetic biology) with computer science (bioinformatics, machine learning, artificial intelligence). He is also a guest professor at the Medical University of Vienna, scientific coordinator of the Biomedical Sequencing Facility (BSF) at CeMM, and key researcher at the Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases. He coordinates an EU Horizon 2020 project on the single-cell analysis of human organoids as a contribution to the Human Cell Atlas. Christoph Bock is an elected member of the Young Academy of the Austrian Academy of Sciences and has received major research awards, including the Max Planck Society’s Otto Hahn Medal (2009), an ERC Starting Grant (2016-2021), and the Overton Prize of the International Society of Computational Biology (2017).

The mission of **CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences** is to achieve maximum scientific innovation in molecular medicine to improve healthcare. At CeMM, an international and creative team of scientists and medical doctors pursues free-minded basic life science research in a large and vibrant hospital environment of outstanding medical tradition and practice. CeMM’s research is based on post-genomic technologies and focuses on societally important diseases, such as immune disorders and infections, cancer and metabolic disorders. CeMM operates in a unique mode of super-cooperation, connecting biology with medicine, experiments with computation, discovery with translation, and science with society and the arts. The goal of CeMM is to pioneer the science that nurtures the precise, personalized, predictive and preventive medicine of the future. CeMM trains a modern blend of biomedical

scientists and is located at the campus of the General Hospital and the Medical University of Vienna. www.cemm.oeaw.ac.at

For **further information** please contact

Laura Alvarez

Social Media and Communications Manager

CeMM

**Research Center for Molecular Medicine
of the Austrian Academy of Sciences**

Lazarettgasse 14, AKH BT 25.3

1090 Vienna, Austria

Phone +43-1/40160-70 057

Fax +43-1/40160-970 000

lalvarez@cemm.oeaw.ac.at

www.cemm.oeaw.ac.at