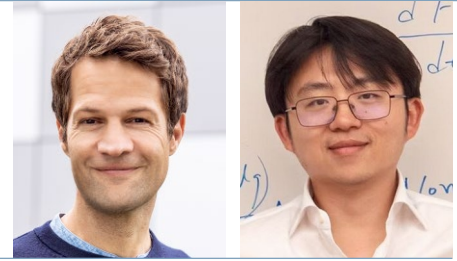


„LANGFRISTIG KÖNNTE EINE NEUE FORM VIRTUELLER BIOLOGIE ENTSTEHEN: HIN ZU EINER BIOLOGIE, DIE NICHT NUR BESCHREIBT, WAS IST, SONDERN ZUNEHMEND VORHERSAGEN KANN, WAS MÖGLICH WIRD.“



Fabian Theis

Weixu Wang

Von Einzelzellatlanten zu Zellsimulationen

DOI: 10.1007/s12268-026-2783-y

© The Author(s) 2026

Die moderne Einzelzellbiologie hat unser Verständnis biologischer Systeme grundlegend verändert. Technologien wie Einzelzell-RNA-Sequenzierung oder räumlich aufgelöste Omics-Verfahren ermöglichen heute die Vermessung biologischer Systeme mit bislang unerreichter Auflösung. Zelluläre Zustände lassen sich dadurch nicht mehr nur als Mittelwerte über Gewebe oder Populationen beschreiben, sondern als hochdimensionale Landschaften einzelner Zellen.

Erst Methoden der künstlichen Intelligenz (KI) ermöglichen es jedoch, in diesen riesigen Datenräumen biologische Struktur sichtbar zu machen: Zelltypen lassen sich identifizieren, Übergänge zwischen Zellzuständen rekonstruieren und Entwicklungslandschaften aus Millionen einzelner Messungen ableiten. Damit entstehen erstmals datengetriebene Karten biologischer Dynamik mit nahezu zellulärer Auflösung. Dennoch bleibt eine zentrale Einschränkung bestehen: Die meisten dieser Daten sind Momentaufnahmen. Sie zeigen, in welchem Zustand sich eine Zelle gerade befindet – nicht jedoch, wohin sie sich entwickelt oder welche regulatorischen Mechanismen diese Entwicklung steuern.

In den vergangenen Jahren wurden daher Methoden entwickelt, die Entwicklungsrichtungen aus Einzelzelldaten rekonstruieren können. Dabei werden Unterschiede zwischen verschiedenen RNA-Zuständen genutzt, um aus statischen Messungen wahrscheinliche zukünftige Zellzustände abzuleiten. Die meisten dieser Ansätze bleiben jedoch weitgehend statistisch: Sie rekonstruieren Übergänge zwischen Zellzuständen aus großen Datenmengen, modellieren die zugrunde liegenden regulatorischen Mechanismen aber nicht explizit. Dennoch markieren sie einen wichtigen Schritt hin zu einer dynamischen Zellbiologie.

Mit dem Modell RegVelo gehen wir einen Schritt weiter. Bisherige dynamische Modelle behandelten Gene häufig weitgehend unabhängig voneinander. RegVelo integriert dagegen regulatorisches Vorwissen direkt in das Modell und beschreibt Genexpression als gekoppelten

regulatorischen Prozess. Welche regulatorischen Faktoren treiben eine Zelldifferenzierung? Welche Eingriffe könnten krankhafte Entwicklungswege umlenken? Welche regulatorischen Programme bestimmen Therapieansprechen oder Resistenz? Es wird möglich, nicht nur Entwicklungswege zu rekonstruieren, sondern auch gezielt regulatorische Eingriffe *in silico* zu simulieren. Das Modell kann vorhersagen, wie sich Zellzustände verändern, wenn einzelne regulatorische Faktoren perturbiert werden – etwa durch genetische Veränderungen oder therapeutische Interventionen. Damit entsteht ein Übergang von rein beschreibenden hin zu prädiktiven Modellen biologischer Systeme.

Besonders relevant wird dies in Bereichen wie Entwicklungsbiologie, Immunologie oder der regenerativen Medizin und für ein mechanistisches Verständnis komplexer Krankheitsprozesse. Die aktuelle Entwicklung unterscheidet sich grundlegend von den „Whole Cell“-Modellen der Systembiologie. Statt biologische Systeme ausschließlich von unten nach oben mechanistisch zu modellieren, entstehen nun datengetriebene generative Modelle, die aus Millionen einzelner Zellen direkt dynamische Regeln biologischer Prozesse lernen.

Solche Modelle verdeutlichen einen breiteren Wandel in den Lebenswissenschaften. Künstliche Intelligenz wird zunehmend nicht nur zur Mustererkennung eingesetzt, sondern zur Simulation biologischer Dynamik. Es zeichnen sich derzeit zwei unterschiedliche Entwicklungsrichtungen ab. Einerseits entstehen immer größere generative KI-Modelle, darunter allgemeine biologische Foundation-Modelle ebenso wie spezialisierte Modelle zur Vorhersage von Perturbationseffekten, etwa unser CellFlow-Modell. Diese Modelle lernen biologische Zustände direkt aus sehr großen Datenmengen und können zunehmend als generative Simulationsmodelle biologischer Systeme verstanden werden. Andererseits verfolgen hybride Modelle wie RegVelo einen stärker biologisch informierten Ansatz, bei dem regulatorisches oder mechanistisches Vorwissen explizit integriert wird. Solche regulatorischen oder physikalischen Constraints könnten insbe-

sondere dann relevant werden, wenn Modelle robuste Vorhersagen außerhalb der Trainingsdaten treffen sollen. Welche dieser Strategien langfristig erfolgreicher sein wird – große datengetriebene Foundation-Modelle oder hybride, biologisch informierte Ansätze – ist derzeit eine der spannendsten offenen Fragen des Feldes.

Langfristig könnte daraus eine neue Form virtueller Biologie entstehen, hin zu einer Biologie, die nicht nur beschreibt, was ist, sondern zunehmend vorhersagen kann, was möglich wird – auf dem Weg zu einem „Virtual Cell“- oder sogar „Virtual Human“-Modell. ■

Prof. Dr. Dr. Fabian Theis, Leiter des Instituts für Computational Biology, Helmholtz Zentrum München

Weixu Wang, Doktorand am Computational Health Center, Helmholtz Zentrum München

Funding: Open Access funding enabled and organized by Projekt DEAL.
Open Access: This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Korrespondenzadressen:

Prof. Dr. Dr. Fabian Theis
Helmholtz Zentrum München
Computational Health Center und
Institut für Computational Biology – ICB

Weixu Wang
Institut für Computational Biology – ICB
Helmholtz Zentrum München
Ingolstädter Landstraße 1
D-85764 Neuherberg
fabian.theis@helmholtz-munich.de
weixu.wang@helmholtz-munich.de

Weiterführende Literatur:

- [1] Klein D, Fleck JS, Bobrovskiy D et al. (2025) CellFlow enables generative single-cell phenotype modeling with flow matching[J]. *bioRxiv*, <https://doi.org/10.1101/2025.04.11.648220>
- [2] Wang W, Hu Z, Weiler P et al. (2026) RegVelo: Gene-regulatory-informed dynamics of single cells. *Cell*, <https://doi.org/10.1016/j.cell.2026.04.02>