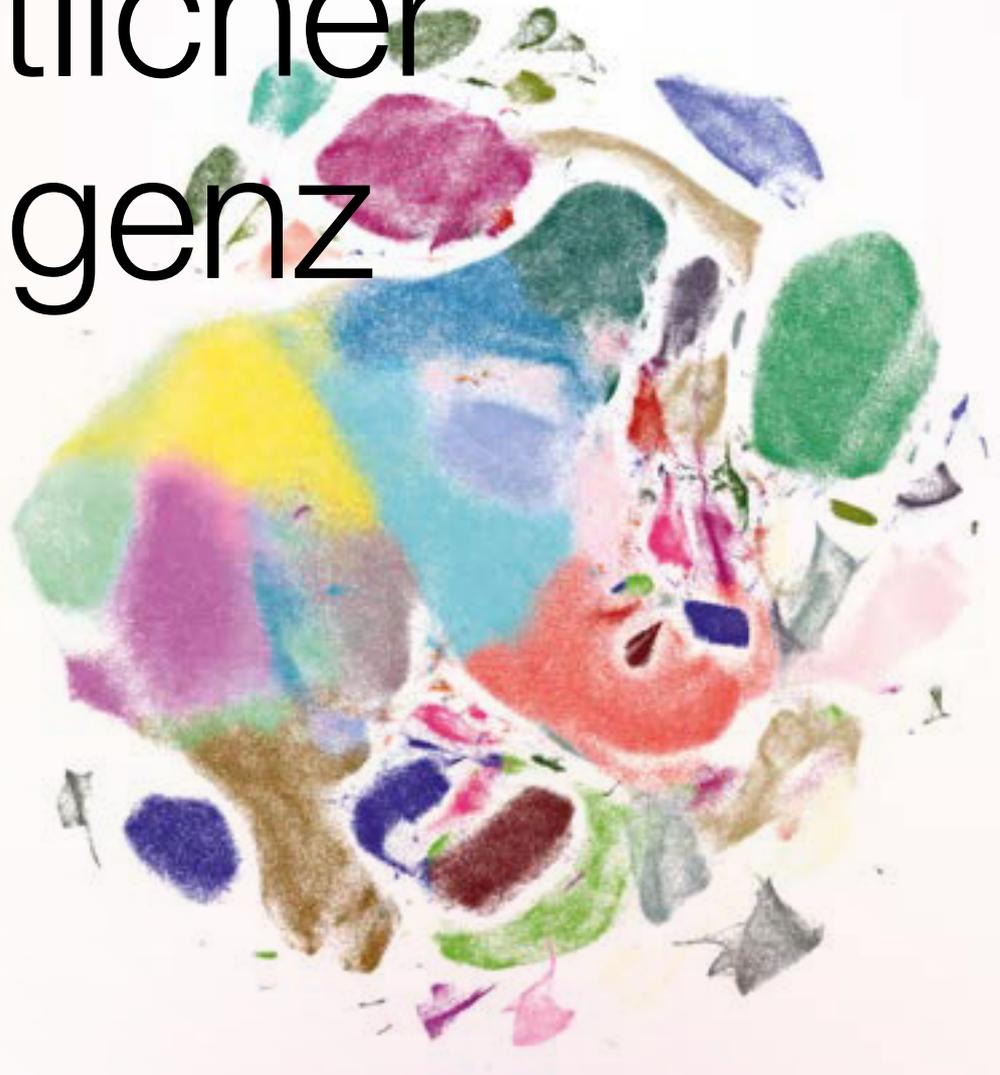


Krankheiten besiegen mit künstlicher Intelligenz



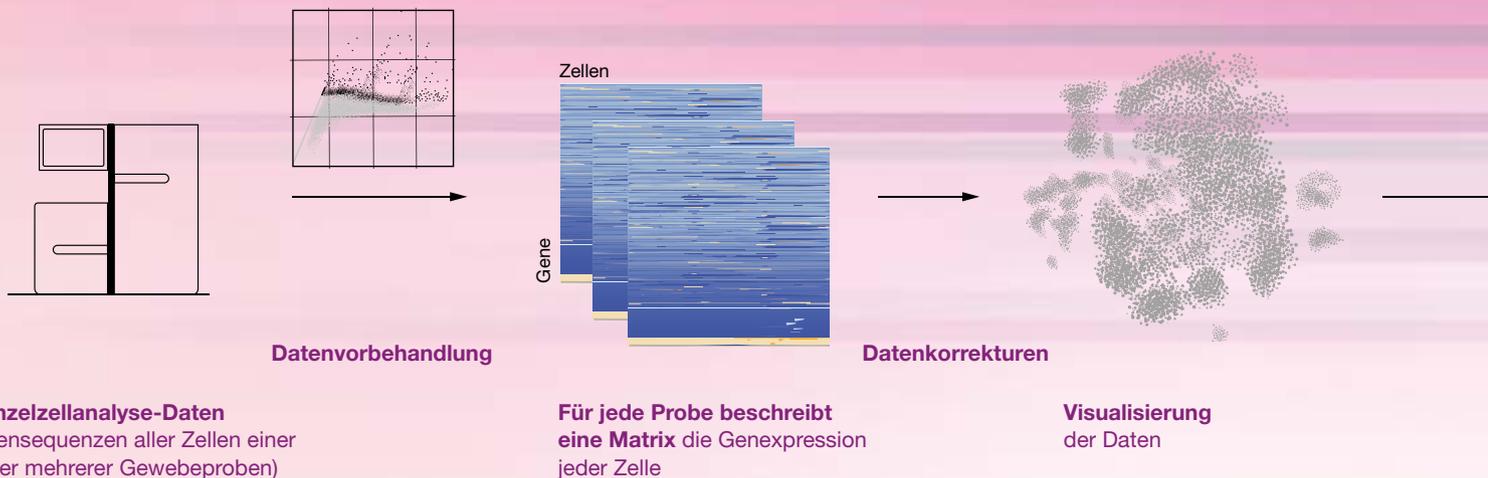
Die modernen Methoden der biomedizinischen Forschung erzeugen eine Unmenge an Daten. Darin Muster zu erkennen und daraus die richtigen Schlüsse zu ziehen, übersteigt die Fähigkeit von Menschen. Stattdessen kommen Computer zum Einsatz – und Mathematiker wie der diesjährige Leibnizpreisträger Prof. Fabian Theis, der in beiden Welten zu Hause ist: den Computer- und den Lebenswissenschaften.

Full Article (PDF, EN): www.tum.de/faszination-forschung-30

Defeating Diseases with Artificial Intelligence E

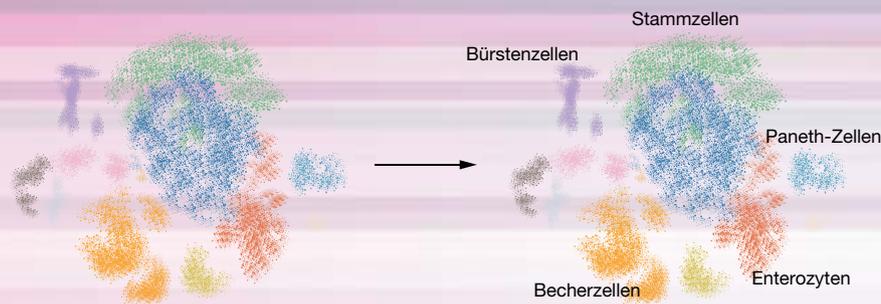
Modern methods used in biosciences produce a torrent of data that simply cannot be managed without the help of computers. Mathematician Fabian Theis is comfortable navigating the worlds of biosciences and computer sciences, harnessing artificial intelligence and machine learning to make large volumes of data usable for researchers. In his work, he strives above all to understand the behavior of cells. Theis uses single-cell genomics, an approach that enables him to determine which genes are active in a specific cell at a given time. This in turn makes it possible to examine the differences between cells in different tissues, in people with different lifestyles, or in healthy and sick individuals. In the Human Cell Atlas project, which aims to create a reference database of all human cell types, Theis has mapped the lungs. Another focus of his work is diabetes research. For example, Theis is developing tools with the ability to predict how insulin-producing pancreatic cells that lose their function due to diabetes will respond to different medications. □

Link
www.helmholtz-munich.de/en/icb/research-groups/theis-lab



In den Biowissenschaften hat längst die Mathematik Einzug gehalten. Dies liegt vor allem an den immensen Datenmengen – gerne plakativ als Big Data bezeichnet –, die inzwischen in vielen Forschungsdisziplinen anfallen und die ein menschliches Gehirn nicht mehr bewältigen kann. Dabei ist Big Data gar nicht so abstrakt, wie es scheint: „Im Prinzip“, so veranschaulicht es Fabian Theis, „fällt schon alles unter Big Data, was man nicht mehr mit einem Excel-Dokument auswerten kann.“ Der Professor für Mathematische Modellierung biologischer Systeme an der TUM und Leiter des Computational Health Center bei Helmholtz Munich muss es wissen, denn er hat sich auf die Auswertung großer Datenmengen spezialisiert. Dazu nutzt er künstliche Intelligenz – vereinfacht gesagt entwickelt er mit seinem Team ausgefeilte Software-Programme, die es ermöglichen, in den Daten nach Mustern zu suchen, daraus Schlüsse zu ziehen und so Forschungsergebnisse überhaupt erst nutzbar zu machen.

Dabei war Theis' Weg in die Biowissenschaften nicht vorgezeichnet. Erst nach dem Mathematik- und Physikstudium in Regensburg kristallisierte sich sein Interesse an Algorithmen heraus. Darunter versteht man eindeutige Handlungsvorschriften, die ein Programmierer einem Computer an die Hand gibt, um damit ein bestimmtes Problem zu lösen. Von da an stand das Problemlösen für Theis im Vordergrund. Herausforderungen fand er vor allem in den Biowissenschaften, weil sie auch viele medizinisch relevante Fragestellungen einschließen. „Ich finde es befriedigend, wenn man nicht nur Hypothesen aufstellt, sondern diese auch experimentell überprüfen kann“, erklärt der Mathematiker seine Vorliebe für die Biomedizin. „In der Biologie finden wir unglaublich komplexe Systeme wie Zellen, die noch gar nicht durch klare Modelle beschrieben werden können. Da gibt es für Mathematiker wie mich noch viel zu tun.“



Clusterbildung

Annotation: Jedem Cluster wird ein bestimmter Zelltyp zugeordnet

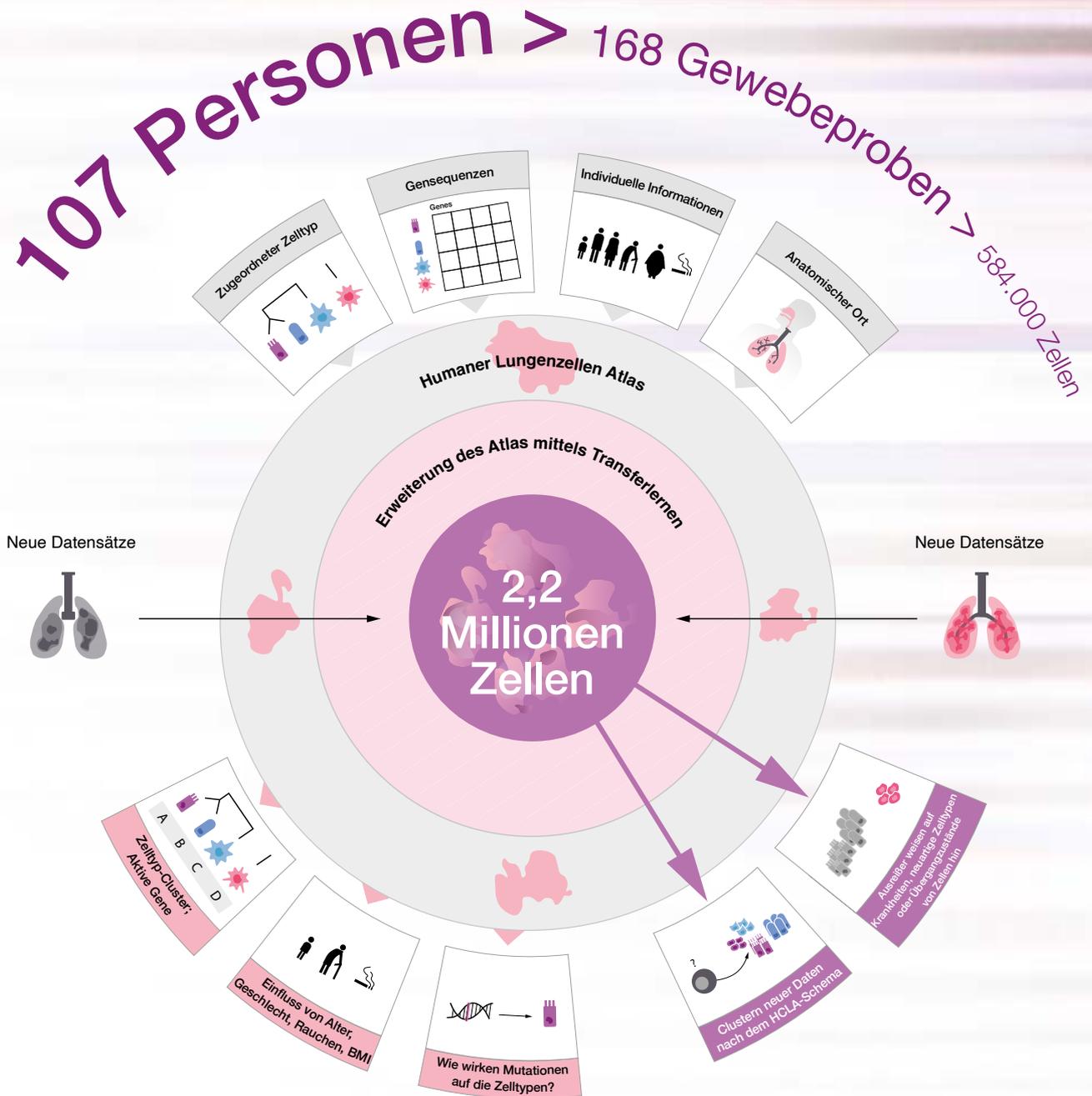
In beiden Welten zu Hause

Für seine Doktorarbeit wechselte Theis deshalb in die Biophysik; nur ein Jahr später folgte ein zweiter Dokortitel in Computerwissenschaften an der Universität Granada. Damit hatte der Wissenschaftler das Rüstzeug, um sich einem der komplexesten Systeme überhaupt zuzuwenden – der lebenden Zelle mit ihren tausenden von Genen, Proteinen und Stoffwechselwegen. Eine leistungsstarke Methode, mit der sich das Verhalten von Zellen verstehen lässt, ist die sogenannte Einzelzell-Genomik. Sie kann unterschiedliche Aspekte der Zelle in den Vordergrund rücken, wie das Erbgut selbst oder die Gesamtheit der Gene, die zu einem bestimmten Zeitpunkt aktiv sind. Letzteres interessiert Theis besonders, denn die aktiven Gene sind das, was die Identität einer Zelle ausmacht.

An dieser Stelle eine kurze Erinnerung: Jede Zelle eines Menschen besitzt das gleiche Erbgut und doch gibt es Leber-, Haut- Lungen- oder auch Immunzellen, die alle unterschiedliche Eigenschaften aufweisen. Der Grund dafür: In unterschiedlichen Zelltypen sind verschiedene Gene aktiv und nur diese tragen zu den Eigenschaften der Zelle bei. Forschende können die Aktivität eines Gens

nachweisen und auf diese Weise verschiedene Zelltypen miteinander vergleichen, aber genauso junge und alte Zellen eines Gewebes oder gesunde und kranke Zellen. Solche Untersuchungen werden schon länger gemacht, doch meistens werden dazu Gewebestücke aus einer Vielzahl an Zellen untersucht. Erst seit kurzem ist es möglich, mit sehr empfindlichen Methoden tatsächlich nur eine einzige, winzige Zelle zu untersuchen. Theis vergleicht die herkömmliche Methode gerne mit einem Obstsalat, denn sie liefert immer einen Mittelwert über viele Zellen – so wie das Geschmackserlebnis beim Essen des Obstsalats. „Heute können wir uns dagegen eine einzelne Frucht herauspicken und untersuchen“, freut sich der Forscher. Man kann sich also gut vorstellen, dass die Einzelzell-Genomik ein mächtiges Instrument ist, um der Funktion von Zellen auf den Grund zu kommen. Aber die Ergebnisse einer einzelnen Zelle können doch keine Big Data sein? „Doch“, sagt Theis. „Denn wir sehen uns natürlich in jeder Zelle sehr viele verschiedene Gene an und das oft bei Millionen von einzelnen Zellen. So kommen wir schnell auf ungeheure Datenmengen.“





Humane Zellatlanten sollten die Diversität der Menschen möglichst gut erfassen. Theis' Team entwickelte das sogenannte Transferlernen, um gezielt weitere Datensätze, vor allem von kranken Menschen, in den Zellatlas der Lunge (Human Lung Cell Atlas - HLCA) zu integrieren. Sie erweiterten den HLCA von 14 Datensätzen mit Daten von insgesamt 584.000 Zellen von 107 Menschen auf 44 Datensätze, die Daten 2,2 Millionen Zellen von 444 Menschen umfassen

Referenzdatenbank aller menschlichen Zelltypen

An dieser Stelle kommt das **Maschinelle Lernen**¹ zum Einsatz. Dabei werden Algorithmen mit großen Sätzen aus bekannten Daten trainiert, bis sie auf der Basis des Gelernten auch neue Datensätze interpretieren können. „Wenn es uns mit unseren Methoden gelingt, die in den Daten versteckten Botschaften zu verstehen, können wir anfangen, die richtigen Fragen zu stellen“, freut sich Theis. „Denn letztlich möchten wir nicht nur neue Methoden entwickeln, sondern auch neue Anwendungen in der biomedizinischen Forschung finden.“ Ein sehr schönes Beispiel dafür ist der Humane Zellatlas, ein wissenschaftliches Großprojekt, das von mehreren hundert Arbeitsgruppen auf der ganzen Welt gemeinsam durchgeführt wird. Ziel ist es, genaue Karten von allen Organen des menschlichen Körpers zu erstellen, um damit zu beantworten, wie diese Organe funktionieren und wie sie sich unterscheiden – bei Menschen unterschiedlichen Alters und Geschlechts, in Abhängigkeit vom Lebensstil und bei verschiedenen Krankheiten. Dafür haben die Arbeitsgruppen gemeinsam weltweit Millionen Zellen aus verschiedenen Organen und Geweben analysiert und katalogisiert. „Wir wollen mit unserem Atlas ein Modell der Wirklichkeit entstehen lassen“, schildert Theis, der Teil des Organisationskomitees ist, die Vision. „Für mich ist das ungeheuer spannend, denn ich wollte schon immer das große Bild sehen.“ Weil diese Aufgabe so komplex ist, wurde sie in Teilprojekte unterteilt: Für jedes Organ wurde ein eigener Atlas geschaffen. Theis kümmerte sich mit einem Team aus mehreren Arbeitsgruppen um die Lunge. Einige Gruppen arbeiteten dafür im Labor mit dem Lungengewebe und erzeugten so die Daten, die dann von anderen Gruppen wie der von Theis aufbereitet und ausgewertet wurden. Durch die Corona-Pandemie konnte der Münchner Forscher auch gleich die Leistungsfähigkeit der neuen Datenbank testen: „Ich frage mich immer, wo ich einen

1 Maschinelles Lernen

ist ein Zweig der Künstlichen Intelligenz, bei dem Algorithmen durch große Datensätze darauf trainiert werden, Muster und Korrelationen in großen Datensätzen zu finden.

sinnvollen Beitrag leisten kann“, so Theis. „Da COVID-19 in erster Linie die Lunge betrifft, wollten wir unseren Lungenatlas nutzen, um mehr über die Krankheit herauszufinden.“ Tatsächlich konnte Theis' Team erklären, warum insbesondere ältere Männer und Raucher anfällig sind für eine Infektion mit dem COVID-19-Erreger: Ihre Lungenzellen bilden mehr von dem Rezeptor, den das Virus als Einfallspforte in die Zellen nutzt. Natürlich ist so ein Atlas nie fertig, denn die Forschungsteams tragen immer neue Datensätze bei und verfeinern so die Ergebnisse. Aber die Hauptarbeit ist gemacht, wie Theis stolz berichtet: „Die 1.0-Version unseres Lungen-Atlas wurde jetzt veröffentlicht.“ ▶

Integration verschiedenster Datensätze in den Humanen Zellatlas

Ein Mammut-Projekt wie der Humane Zellatlas hat seine ganz eigenen Schwierigkeiten. Eine davon ist die Tatsache, dass die zugrunde liegenden Ergebnisse in einer Vielzahl von Laboren rund um die Welt erarbeitet wurden. Überall werden die Experimente zur Messung der Genaktivität in den Zellen, wenn auch nach den gleichen Protokollen, ein ganz kleines bisschen anders durchgeführt. Und das spiegelt sich auch in den Daten wider. Um die verschiedenen Datensätze trotzdem miteinander in Einklang zu bringen, kommt wieder Maschinelles Lernen ins Spiel, diesmal eine Variante, die als Transferlernen bezeichnet wird. Mit Hilfe einer von Theis' Team entwickelten Software kann der Computer das, was er an einem Datensatz gelernt hat, auf einen anderen, mit etwas andersartigen Daten, übertragen. So können jederzeit neue Datensätze aus anderen Laboren in den Atlas integriert werden, und andere Forscher können den Atlas nutzen, um ihre eigenen Datensätze zu verstehen.

„Ich will lernen mit Zellen zu sprechen.“

Fabian Theis

Die Besten nach München holen

Spitzenforschung ist heute nur noch mit Kooperationen möglich. So arbeitet auch Theis mit anderen Arbeitsgruppen zusammen, die im Nasslabor Experimente mit Zellkulturen machen und dann ihre Daten Theis zur Analyse übergeben. Die Zusammenarbeit ist dabei so eng, dass Doktoranden in der Regel zwei Arbeitsgruppen angehören und so schon während der Ausbildung beide Welten kennenlernen – die Welt der Biowissenschaften und die Computerwelt. „Auf diese Weise lernt der Nachwuchs, beide Sprachen zu verstehen“, ist Theis überzeugt. Um exzellente Studienabgänger für eine Doktorarbeit nach München zu bringen, hat er die Munich School of Data Science ins Leben gerufen, die aus rund 200 internationalen Bewerbungen pro Semester die besten 10 auswählt. Gefördert wird diese Einrichtung durch die beiden Münchner Hochschulen TUM und LMU, Helmholtz Munich und das Deutsche Luft- und Raumfahrtzentrum.

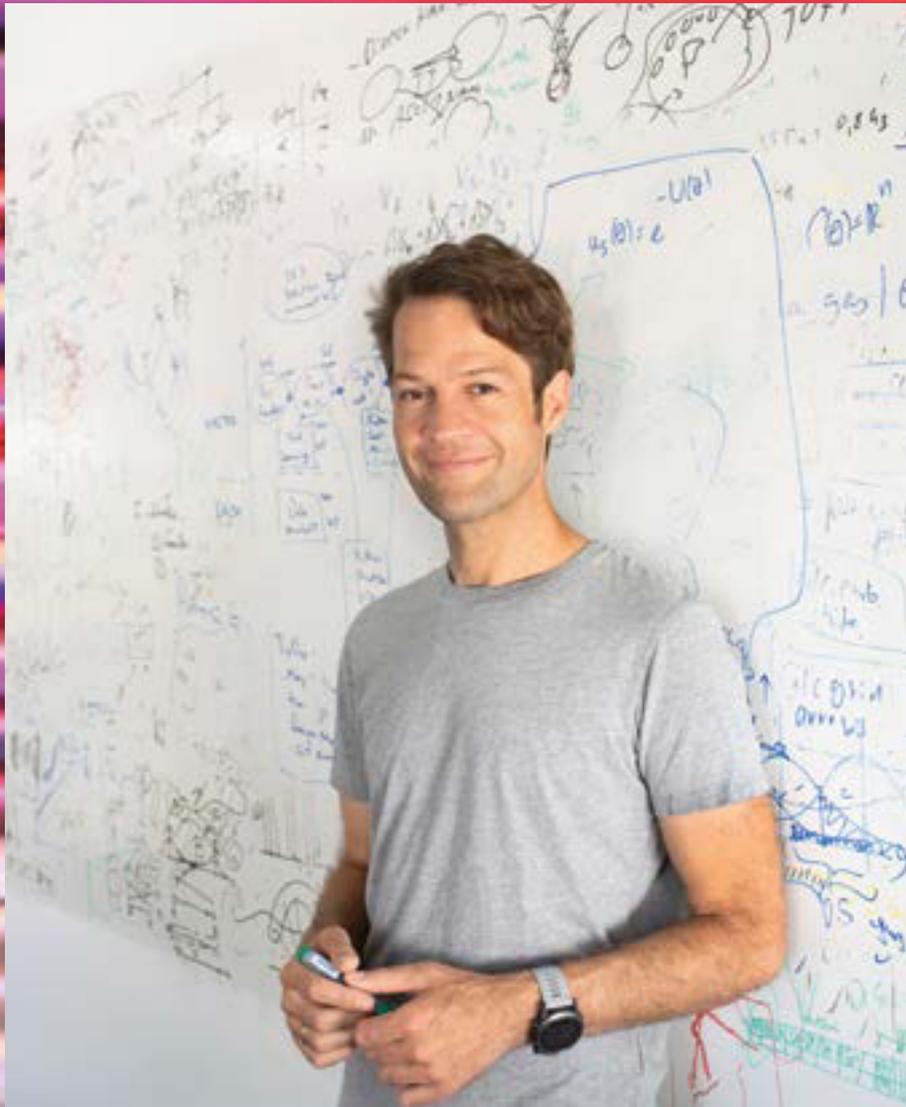
Erst kürzlich hat Theis mit dem Gottfried-Wilhelm-Leibniz-Preis der Deutschen Forschungsgemeinschaft den wichtigsten deutschen Forschungspreis erhalten. Dieser spült 2,5 Mio. Euro in die Kassen des Biomathematikers und beflügelt seine Forschung. „Ich will lernen mit Zellen zu sprechen“, sagt der Preisträger und meint das durchaus ernst. Indem die Forscher die Reaktion von Zellen auf Input von außen modellieren, können sie simulieren, wie zwei Zellen miteinander kommunizieren oder wie sie auf ein bestimmtes Medikament reagieren. „Wie die Zellen eines bestimmten Patienten auf ein Medikament reagieren werden“, präzisiert Theis. Herzlich willkommen – Sie sind in der Medizin der Zukunft angekommen! ■

Larissa Tetsch

Maschinelles Lernen in der Diabetes-Forschung

Vor dem Hintergrund, dass zu Helmholtz Munich ein auf Diabetes spezialisiertes Forschungszentrum gehört, überrascht es nicht, dass Theis verschiedene Fragestellungen rund um die tödliche Stoffwechselkrankheit bearbeitet. „Mit unserem Einzelzell-Genomik-Ansatz konnten wir eine Erklärung dafür finden, warum bestimmte Zellen der Bauchspeicheldrüse beim Typ-2-Diabetes – dem sogenannten Altersdiabetes – mit der Zeit die Fähigkeit verlieren, auf den Blutzucker zu reagieren“, bringt der Forscher ein Beispiel und führt aus, wie er gemeinsam mit seinem Kollegen Heiko Lickert – einem Spezialisten für die bei Diabetes veränderten Bauchspeicheldrüsenzellen – nach Wirkstoffen sucht, die diesen Prozess umkehren. Die Einzelzell-Genomik kann dabei den Erfolg der Medikamente überwachen und irgendwann vielleicht sogar ganz ohne Experimente vorhersagen, wie eine Zelle auf ein bestimmtes Medikament reagieren wird. Ein großer Schritt in Richtung „Personalisierte Medizin“. Das sieht wohl auch der renommierte Europäische Forschungsrat so, denn er fördert das Projekt finanziell großzügig.

In Zusammenarbeit mit der Augenklinik der LMU München verwendet Theis Maschinelles Lernen außerdem, um einen Algorithmus zur Erkennung der diabetischen Retinopathie zu trainieren. Diese Folgeerkrankung von Diabetes mellitus Typ 1 ist die häufigste Ursache für eine Erblindung im Erwachsenenalter. Zum Training wurden öffentliche Datensätze aus Bildern des Augenhintergrunds von Gesunden und Erkrankten verwendet. Nach dem Training erkannte der Computer Krankheitssymptome auch auf unbekanntem Bildern. „In München mit seiner großen Facharztdichte brauchen wir das nicht unbedingt“, schränkt Theis ein. „Aber in ländlichen Gebieten kann unser Algorithmus Hausärzte bei der Risikoabschätzung unterstützen, und eines Tages kann vielleicht sogar ein Arzt in Afrika die Augen seiner Patienten mit dem Handy untersuchen.“



Prof. Fabian Theis

hat in Regensburg Mathematik und Physik studiert und anschließend in Biophysik (Regensburg) und in Computerwissenschaften (Granada) promoviert. Nach Forschungsaufenthalten in Japan, den USA und am Max-Planck-Institut für Dynamik und Selbstorganisation in Göttingen sowie der Habilitation in Regensburg wechselte er als Associate Professor an die TUM. Dort hat er seit 2013 den Lehrstuhl für Mathematische Modellierung inne und leitet bei Helmholtz Munich das Computational Health Center sowie das Institute of Computational Biology. Theis ist Mitglied im Munich Data Science Institute der TUM, das sich als zentrale Schnittstelle und Innovationsplattform für Fragen und Lösungen aus Datenwissenschaften, Maschinellem Lernen und Künstlicher Intelligenz versteht. Zudem leitet er zusammen mit zwei Kollegen die Münchner Einheit des internationalen Forschungsnetzwerks European Laboratory for Learning and Intelligent Systems (ELLIS) mit derzeit rund 30 Standorten überall in Europa. Seit 2020 ist er Co-Vorsitzender des KI-Rats der Bayerischen Staatsregierung. Außerdem ist Theis wissenschaftlicher Direktor der Kooperationseinheit für Künstliche Intelligenz (Helmholtz AI) und assoziiertes Fakultätsmitglied am Wellcome Sanger Institute in Hinxton, UK. Seine wissenschaftliche Leistung wurde mehrfach ausgezeichnet, zuletzt mit einem ERC Advanced Grant und dem Leibniz-Preis der Deutschen Forschungsgemeinschaft.
