

# Defeating Diseases with Artificial Intelligence



**Modern biomedical research methods generate vast quantities of data. Identifying patterns within this data and drawing the right conclusions are tasks simply beyond human abilities. Instead, researchers turn to computers – and mathematicians like this year's Leibniz Award winner, Prof. Fabian Theis, who feels equally at home in both worlds: computer sciences and life sciences.**

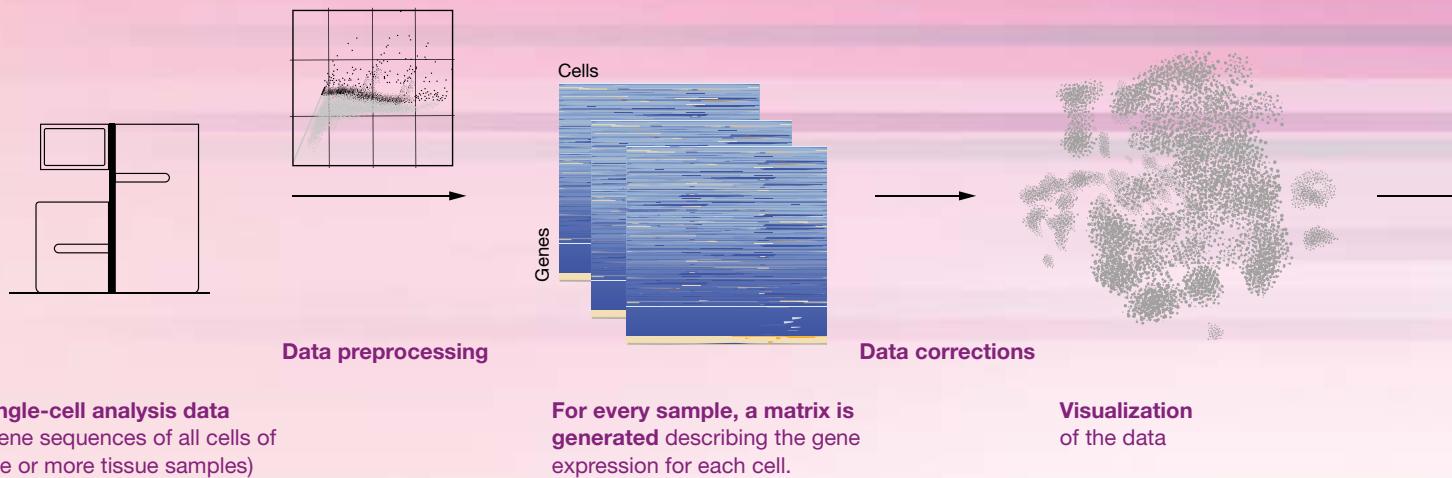
Gesamter Artikel (PDF, DE): [www.tum.de/faszination-forschung-30](http://www.tum.de/faszination-forschung-30)

## Krankheiten besiegen mit Künstlicher Intelligenz D

In den Biowissenschaften generieren moderne Methoden eine Flut von Daten, die ohne die Hilfe von Computern nicht mehr beherrschbar ist. Der Mathematiker Fabian Theis bewegt sich zwischen den Welten der Bio- und der Computerwissenschaften und nutzt Künstliche Intelligenz (KI) und Maschinelles Lernen, um große Datenmengen für Anwender nutzbar zu machen. Mit seiner Arbeit möchte er vor allem das Verhalten von Zellen verstehen. Dazu nutzt er die Einzelzell-Genomik, mit der er ermitteln kann, welche Gene in einer einzelnen Zelle zu einer bestimmten Zeit aktiv sind. Auf diese Weise lässt sich dann untersuchen, wie sich Zellen unterscheiden, die aus verschiedenen Geweben oder aus Menschen mit unterschiedlichem Lebensstil, aus gesunden oder aus kranken Menschen stammen. Im Rahmen des Humanen Zellatlas, der eine Referenzdatenbank aller menschlichen Zelltypen darstellen soll, hat Theis die Lunge kartiert. Ein weiterer Schwerpunkt seiner Arbeit ist die Diabetesforschung. So entwickelt Theis unter anderem Werkzeuge, um vorhersagen zu können, wie die insulinproduzierenden Zellen der Bauchspeicheldrüse, die bei Diabetes ihre Funktion verlieren, auf verschiedene Medikamente reagieren. □

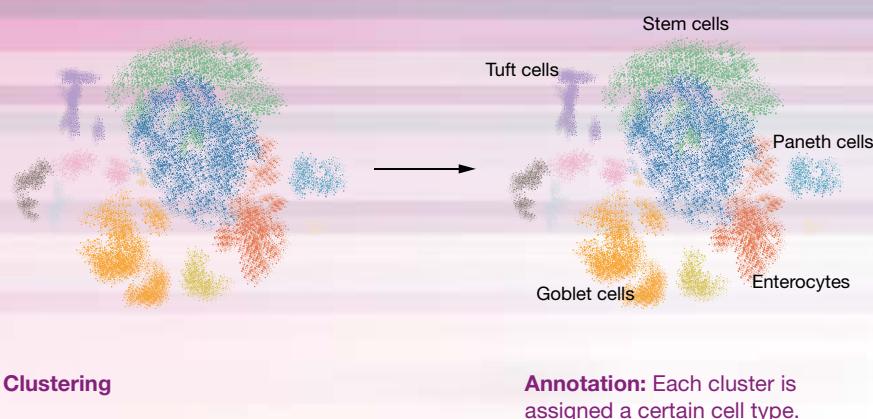
Link

[www.helmholtz-munich.de/en/icb/research-groups/theis-lab](http://www.helmholtz-munich.de/en/icb/research-groups/theis-lab)



**M**athematics has long since found its way into the biosciences. This is primarily due to the immense quantities of data, pithily referred to as big data, that are generated in many research disciplines, and which have become too vast for the human brain to handle. That said, big data is not quite as abstract as it might appear. “In principle,” explains Fabian Theis, “big data is anything that cannot be analyzed in an Excel spreadsheet.” A Professor of Mathematical Modeling of Biological Systems at TUM and Director of the Computational Health Centers at Helmholtz Munich, Theis is an authority in such matters, having specialized in the evaluation of large data volumes. He achieves this by harnessing artificial intelligence (AI). In simple terms, he and his team develop sophisticated software programs that make it possible to search for patterns in data, draw appropriate conclusions, and thereby make research results usable in the first place.

Fabian Theis’ path to the biosciences, however, has been far from direct. It was only after studying mathematics and physics in Regensburg that he began to develop an interest in algorithms – definite instructions that programmers input into computers to enable them to solve specific problems. From that point on, Theis was squarely focused on problem-solving. He saw particular challenges in the biosciences given their links to medically significant issues. “I find it gratifying that, in addition to proposing hypotheses, I can also test them in experiments,” enthuses the mathematician, explaining his love of biomedicine. “In biology, we encounter unbelievably complex systems, such as cells, which we are still unable to describe in clear models. That means mathematicians like me still have a lot of work to do.”

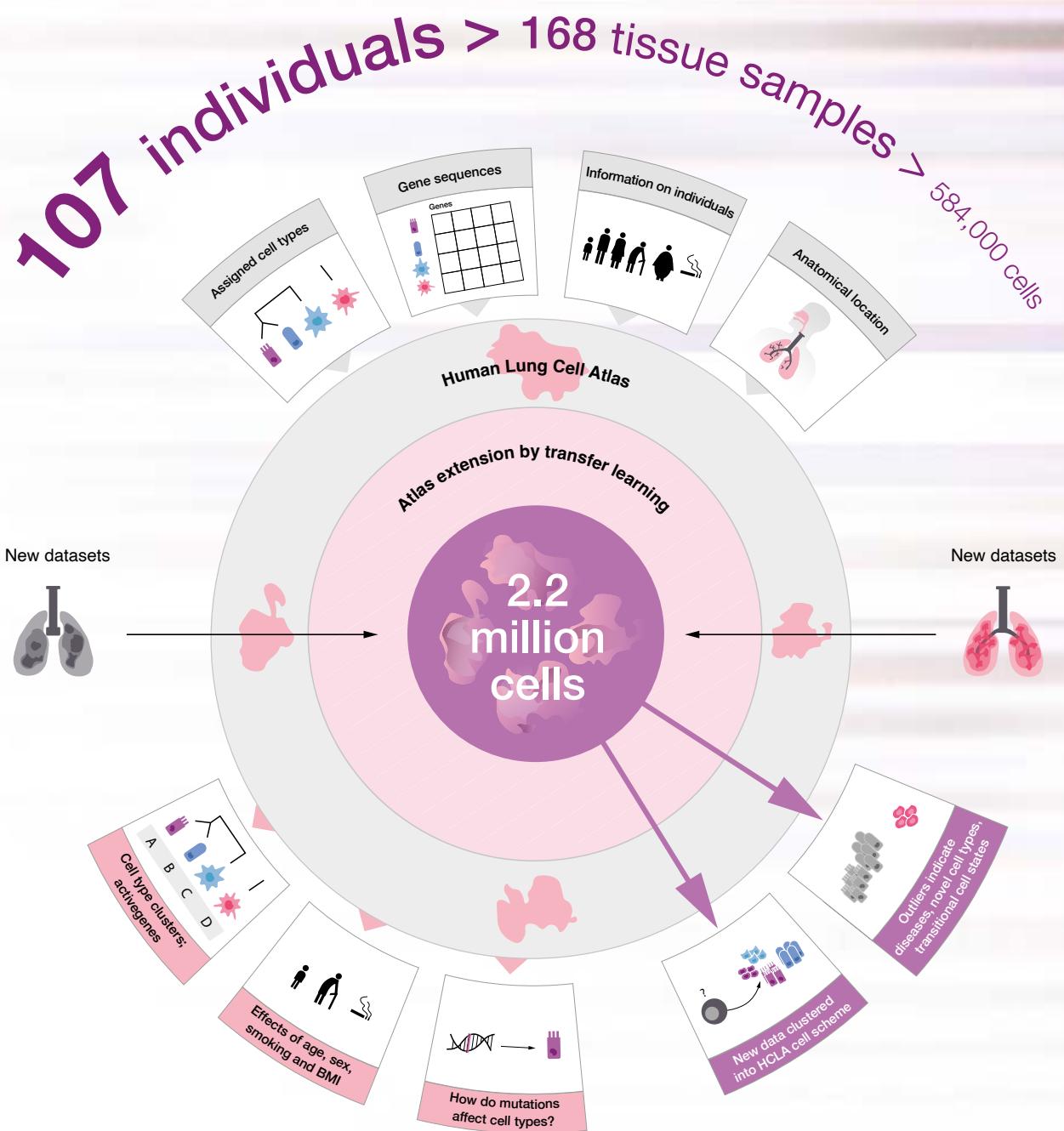


### A man at home in both worlds

Theis decided to focus on biophysics for his doctoral thesis. A second doctorate followed a year later, this time in computer science from the University of Granada. These studies equipped Theis with the tools required to tackle one of the most complex systems in existence: the living cell, with its thousands of genes, proteins and metabolic pathways. One promising method for understanding cell behavior is single-cell genomics. It can shine a spotlight on various aspects of cells, such as the genome itself or the totality of all genes active at a given moment in time. Theis is particularly interested in the latter aspect, as active genes determine the identity of a cell.

Before delving any deeper, a quick refresher may be helpful. Each human cell contains the same genetic material and yet we have liver, skin and lung cells or also immune cells, for example, all with different characteristics. This is because different genes are active in different cell types – and only these active genes influence a cell's characteristics. Researchers can detect gene activity, which in turn allows them to compare different types of cells. They can also compare young and old cells in a given type of tissue, as well as healthy and diseased cells.

Although these kinds of investigations have been conducted for some time, they usually examine a piece of tissue made up of numerous cells. It is only very recently that highly sensitive methods have made it possible to examine a single, tiny cell. Theis likes to compare the conventional method to a fruit salad, as it always provides an average value across numerous cells – just like the taste experience of eating a fruit salad. “By contrast, we are now able to pick out individual fruits and examine them separately,” effuses the researcher. This clearly illustrates the potential of single-cell genomics as a powerful tool for getting to the bottom of how cells function. But surely analyzing a single cell can’t produce big data? “On the contrary,” says Theis. “Every single cell obviously contains vast numbers of different genes and there are often millions of individual cells to look at. This means we rapidly generate vast amounts of data.” ▶



**Human cell atlases** should capture the diversity of people as thoroughly as possible. Theis' team developed so-called transfer learning in order to integrate additional datasets, especially from sick people, into the Human Lung Cell Atlas (HLCA). They expanded the HLCA from 14 datasets with data from a total of 584,000 cells from 107 people to 44 datasets comprising data from 2.2 million cells from 444 people.

## Reference database of all human cell types

This is where **machine learning**<sup>1</sup> comes in. Algorithms are trained with huge sets of known data until they reach a point where what they have learned enables them to interpret new datasets. "If our methods allow us to understand the messages hidden in this data, we can start to ask the right questions," says a delighted Theis. "Ultimately, we not only want to develop new methods, we also hope to find new applications in biomedical research." A prime example is the Human Cell Atlas (HCA), a major scientific undertaking involving hundreds of working groups all around the world. The project's goal is to produce precise maps of every organ in the human body and thereby explain how these organs function and how they differ – in people of different ages and sexes, people with different lifestyles, and with various diseases. Working groups around the world have come together to analyze and catalog millions of cells from different organs and tissues. "We hope that our atlas will give rise to a model of reality," says Theis, a member of the HCA Organizing Committee, as he outlines the project's vision. "I find it incredibly exciting because I've always wanted to see the big picture." In light of the project's complexity, it has been organized into sub-projects, with a separate atlas created for each organ. Theis and a team comprising several working groups focused on the lungs. This saw some groups working with lung tissue in laboratories and generating data for other groups – like the group led by Theis – to process and analyze. The coronavirus pandemic

### 1 Machine learning

is a subset of artificial intelligence involving the use of extensive datasets to train algorithms to detect patterns and correlations in big data.

gave the Munich-based researcher the opportunity to test the capabilities of the new database. "I'm always asking myself where I can make a meaningful contribution," says Theis. "Given that COVID-19 affects the lungs above all, we wanted to use our lung atlas to find out more about the disease." In fact, Theis' team were able to identify why elderly people and smokers were particularly susceptible to a COVID-19 infection: their lung cells form more of the receptor that the virus uses to gain access to the cells. Of course, an atlas like this will never be complete, as the research teams are constantly adding new datasets and refining their findings. The main bulk of the work has been done, however, as Theis proudly reports: "Version 1.0 of our lung atlas has now been published." ▶

### Integrating countless datasets in the Human Cell Atlas

A mammoth undertaking like the Human Cell Atlas faces very specific challenges. One is the fact that the results that underpin the entire project are being developed and produced at numerous laboratories around the world. Even when the same protocols are followed, experiments to measure gene activity in cells differ ever so slightly across the various locations. This is also reflected in the data. In order to harmonize the numerous datasets despite these minor discrepancies, the researchers once again turn to machine learning – this time using a variant called transfer learning. Software developed by Theis' team enables the computer to translate what it has learned from one dataset to another dataset with slightly different data. This means that new datasets from other labs can be integrated into the atlas at any time, while other researchers can use the atlas to better understand their own datasets.

# *“I want to learn to speak with cells.”*

Fabian Theis

## **Bringing the best minds to Munich**

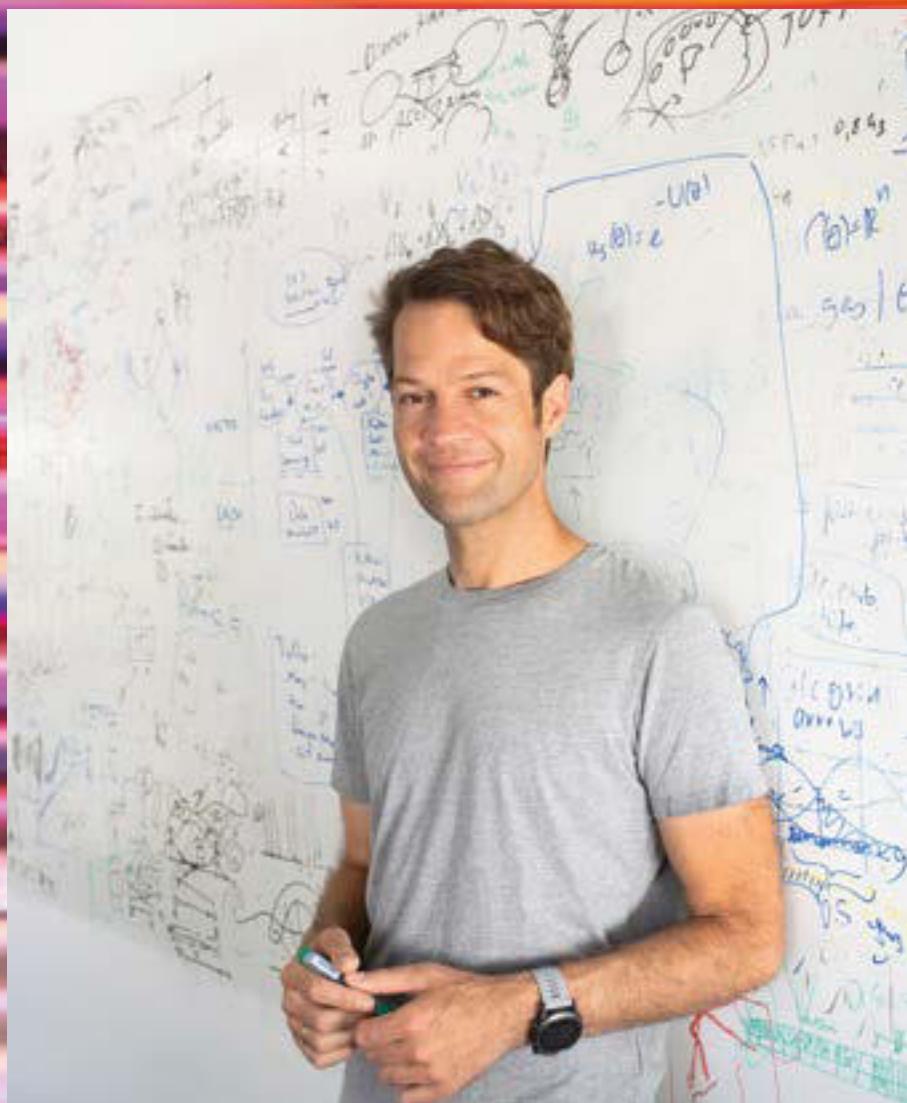
Nowadays, cooperation is the only route to cutting-edge research. With this in mind, Theis collaborates with other working groups who conduct experiments with cell cultures in the wet lab before sending their data to Theis for analyses. The cooperation is so close that doctoral students usually belong to two working groups, which enables them to explore both worlds during their studies – the world of biosciences and the world of computing. “This way, junior researchers learn to speak both languages,” states a confident Theis. In an effort to attract outstanding graduates to Munich for doctoral study, Theis has founded the Munich School of Data Science. It receives around 200 international applications per semester and selects only the best ten. This new institution is supported by Munich’s two public universities, TUM and LMU, as well as Helmholtz Munich and the German Aerospace Center.

Recently, Theis received one of the most prestigious research awards in Germany, the Gottfried Wilhelm Leibniz Prize, from the German Research Foundation. The prize money of €2.5 million has swelled the biomathematician’s coffers and spurred on his research. “I want to learn to speak with cells,” the award winner claims, with complete sincerity. By modeling cells’ responses to external inputs, researchers can simulate how two cells will communicate with each other or how cells will react to a given medication. “How a specific patient’s cells will respond to a given medication,” specifies Theis. You have arrived in the medical world of the future. Welcome! ■ *Larissa Tetsch*

## **Machine learning in diabetes research**

Given that Helmholtz Munich has a specialist research center dedicated to diabetes, it is no surprise that Theis is examining various issues related to the deadly metabolic disorder. “Based on our single-cell genomics approach, we were able to find an explanation for why, in the case of type 2 diabetes – also known as adult-onset diabetes – certain cells in the pancreas lose the ability to respond to blood sugar over time,” explains Theis. He describes how he and his colleague Heiko Lickert, a specialist in pancreatic cells altered by diabetes, are searching for substances to reverse this process. In this context, single-cell genomics can monitor a medication’s success and potentially, at some point, even predict how a cell will respond to a given medication before any experiments are conducted. It represents a major step towards personalized medicine. The prestigious European Research Council certainly agrees, having provided generous financial support for the project.

Working in cooperation with the Department of Ophthalmology at the LMU Klinikum, Theis has also used machine learning to train an algorithm to identify diabetic retinopathy. This secondary disease of type 1 diabetes mellitus is the most common cause of blindness in adults. The training used public datasets comprising images of the fundus of healthy and diseased eyes. Once the training was complete, the computer was also able to identify symptoms of disease on unknown images. “In Munich, the high numbers of specialist doctors means we don’t necessarily need this tool,” qualifies Theis. “But, in rural areas, our algorithm can help family doctors to assess risks and, one day, a doctor in Africa might be able to examine their patient’s eyes with their phone.”



---

**Prof. Fabian Theis**

---

studied mathematics and physics in Regensburg before receiving doctorates in biophysics (Regensburg) and computer science (Granada). Following research stays in Japan, the USA and at the Max Planck Institute for Dynamics and Self-Organization in Göttingen, he completed his post-doctoral lecturing qualification (Habilitation) in Regensburg before accepting a position as Associate Professor at TUM. Theis has held the Chair of Mathematical Modeling at TUM since 2013 and is Director of the Computational Health Center and the Institute of Computational Biology at Helmholtz Munich. He is a member of the Munich Data Science Institute at TUM, a central interface and innovation platform for issues and solutions arising from data sciences, machine learning and artificial intelligence. Additionally, he and two colleagues lead the Munich branch of the European Laboratory for Learning and Intelligent Systems (ELLIS), an international research network with around 30 locations throughout Europe at present. In 2020, he became Co-Chair of the Bavarian state government's AI Council. Theis is also Scientific Director of the Helmholtz Artificial Intelligence Cooperation Unit (Helmholtz AI) and an Associate Faculty Member at the Wellcome Sanger Institute in Hinxton, UK. His scientific achievements have been recognized on multiple occasions, most recently with the ERC Advanced Grant and the German Research Foundation's Leibniz Prize.

---