Research Interests – Prof. Fabian Theis

How do cells make decisions? What steps turn a single fertilised egg into a whole organism? And how can we recognise early warning signs of disease and restore the tissue to a healthy state with optimal medications?

Answering these types of questions used to fall exclusively in the realm of biology or medicine. But with the flood of data seen in many areas, these have become questions to be addressed with data science. Especially in the field of biology, where central governing equations are lacking, statistical descriptions of complex biological systems are necessary. These motivations lie at the heart of Prof. Dr. Dr. Fabian Theis' research, aiming to advance machine learning and artificial intelligence (AI) in the context of computational biology. The scientist develops machine learning methods for the analysis of large-scale biological data, such as single-cell genomics data. His goal: to understand the underlying mechanisms of cellular processes. Methodologically, Theis and his research group focus on deep representation learning, which they apply to study trajectories not only of cells under development or disease but also of patients in clinical cohorts.

Fabian Theis and his team use single-cell sequencing to model cells and their activities. This method is based on the overwhelming success of genomics, which has become increasingly important with the Human Genome Project as a core driver of molecular biology data. Over the past decades, whole genome sequencing costs have been reduced from an initial 100 M$ to a few 100 $ over the past two decades. Combining this leap in data production with the recent advances in microfluidics, it has become possible to now measure cellular state at the single-cell level with unprecedented throughput and resolution, thus providing one of the most dynamic sources of big data in molecular biology. The triple award "Nature Method of the Year" (2013, 2019, and 2020 in various extensions) shows the enormous relevance of these technologies to biomedical research. Single-cell genomics serves as a basis for highly ambitious initiatives such as the Human Cell Atlas project, which aims to quantify cellular heterogeneity in all human tissues and is poised, as a spiritual successor to the game-changing Human Genome Project, to become a key resource for many future biomedical studies.

Analysis and interpretation of such big data are of key importance and a bottleneck in exploiting to a maximum the insights hidden in the datasets. AI methods are therefore indispensable to

Forschungsschwerpunkte – Gottfried Wilhelm Leibniz-Preis 2023
Prof. Fabian Theis
Januar 2023

DFG

initiatives like the Human Cell Atlas to make them the foundations for modern medical research. Fabian Theis addresses these challenges by developing methods for analysing, visualising, and modelling cell heterogeneities and building cell atlases. One major goal is to make the lab's methods accessible in analysis software packages with large user bases, which in parts have become de facto standards in their application areas. The analysis framework "Scanpy", for instance, has found broad applications in various communities and has become one of the most popular frameworks in the field.

Early on in his research career, Theis has been focusing on contributing to the systematic characterisation of cellular processes at a molecular level. Network description, as a key element of efficient data integration, played a prominent role. Parallel to the network approach, the researcher started to work on the machine learning-based extension of quantitative models by latent variables in an ERC Starting Grant in 2010. The core area of application was stem cell biology and therapy, in particular, the question of how a cell makes developmental decisions. For this purpose, Theis developed computer-based analytical methods for microscopy images, in which the activity of regulatory molecules in individual cells was quantified by fluorescence. Using neural networks and deep learning, it was possible, for instance, to predict at what point a blood cell in a whole genealogy of cells would choose to become an erythrocyte or a lymphocyte. In an interdisciplinary joint endeavour with experimental colleagues, Theis and his team developed tools to process and analyse single-cell data from both time-resolved microscopy and single-cell genomics. This work detailed protein dynamics quantitatively over many days and also cellular generations, which toppled a standard model of lineage choice and unravelled the role of important genes in early blood development.

Beyond these early lineage models, Fabian Theis' team has extended their approaches towards general representation learning of cellular space and developed methods for estimating and visualising lineages and branching trajectories, contributing to the popularity of machine learning methods in single-cell genomics. More recently, the team leveraged unsupervised deep learning to describe latent space structures in cell manifolds. The so-called Deep Count Autoencoder, for example, compresses an enormous amount of gene expression profiles from cells: This reduction of information removes unnecessary data and thus helps to de-noise single-cell data sets. Surprisingly, the method also learned to describe diverse cellular variations and helped to interpolate processes in stem cell differentiation or disease development. Fabian Theis and colleagues tested how the method learned and predicted perturbations, such as drug combinations, laying the groundwork for the ERC Advanced Grant that Theis acquired in

2022. In this project, the researchers aim to model the effects of drugs on cellular and organoid screens to support pharmacology in screening for drug effects with in-silico models.

The methods developed by Fabian Theis and his research group have led to the discovery of new biological insights, such as the identification of new cell types and the characterisation of cell-to-cell variability in gene expression. These approaches were applied to build cell atlases within the Human Cell Atlas consortium, for example, in the lung. Corresponding results can be used to investigate, for example, how the expression of the SARS-CoV-2 receptor and associated proteases are associated with age, sex, and smoking status, assessing the molecular basis for increased risk of infection and disease severity in different populations.

Beyond single-cell genomics, Fabian Theis has leveraged AI in other areas ranging from biomedical research via healthcare to precision medicine and has successfully demonstrated the value of AI-based technologies in many applications, from prevention and diagnostics to therapy. The researcher applied AI to medical diagnostics and risk identification, for example, for the diagnosis of diabetic retinopathy and differential diagnosis of skin diseases such as psoriasis and atopic eczema. In a long-term collaboration with diabetes researchers, he further developed a classification for disease risk for type 1 diabetes based on genetic data and contributed towards novel non-invasive therapies for the prevention, treatment, and eventual reversion of type 2 diabetes.

Overall, the pioneering research of Fabian Theis in the field of data science led to the development of ground-breaking methods and tools for the analysis of single-cell data and the discovery of new biological insights, with the potential to inform the development of new therapies for various diseases.