Abstract

Multiple sclerosis (MS) is an inflammatory neurodegenerative disease. Despite improvements in treatment and extensive research efforts, the underlying mechanism causing and driving the disease is barely understood. White matter brain lesions are an apparent characteristic of MS, and their evolution throughout disease progression has not been holistically examined. Consequently, we aimed to combine expertise in next-generation sequencing with bioinformatics approaches to analyze the whole lesion transcriptome.

In the first manuscript, we were able to extract the transcriptome of 98 samples from 10 MS and five non-neurological control individuals. We extracted differentially expressed genes (DEGs) based on generalized linear models and corrected for confounding factors such as age and sex. With *de novo* network enrichment, we were able to extract lesion-specific PPI networks. We found that chronic active lesions differed the most from all other lesion types. Furthermore, we found TGFβ-R2 to be a central hub throughout all lesions and could validate those findings with immunohistochemistry. And found TGFβ-R2 to be expressed in astrocytes.
The second manuscript introduces the MS Atlas, a user-friendly web service providing easy access to our results. The MS Atlas is a tool that allows testing for research hypotheses and potential drug targets. It fosters *de novo* network enrichment able to extract mechanistic markers allowing pathway-based lesion type comparison. We provide an application scenario based on Natalizumab, a drug targeting VLA-4 that appears to be ineffective in progressive MS.
In conclusion, we present the first unbiased, holistic transcriptome profile of all white matter lesions. We identified lesion-specific as well as disease-specific de-regulated networks. Our results demonstrate differences and similarities between lesion types. Furthermore, our web-service represents the first comprehensive database allowing easy access to the full transcriptome.

The developments in next-generation sequencing techniques and bioinformatics have changed medical research significantly. Instead of investigating the relation of single genes and molecules with respect to a disease, omics approaches can measure thousands of features. This led to an increased demand in the number of samples to ensure statistical reliability. At the same time, patients are more aware of the potential privacy issues introduced by genomic data. Federated learning is a privacy-aware, decentralized approach of machine learning and trains algorithms on distributed data by

exchange model parameters instead of raw data.
In our third manuscript, we developed a user-friendly federated tool named sPLINK that can perform large scale GWAS analysis on distributed datasets. We demonstrate its accuracy compared to traditional aggregated analysis and its superiority over previously applied meta-analysis in case of unevenly distributed phenotypes. We conclude that federated learning, combined with further privacy-ensuring mechanisms, can access previously unattainable patient data at hospitals. This will allow researchers global yet anonymous access to primary medical data to study and understand disease mechanisms.