## Abstract

For a long time, doctors have relied on the olfactory sensations of patients' breath and body odors due to their diagnostic powers. Since then, more robust and modern non-invasive methods have been developed and promise to improve clinical practice. With these revolutionary developments in high-throughput omics technologies, it has become possible to characterize molecular processes more accurately than ever before. Investigators are presented with data in increasing amounts and complexity covering a variety of processes and molecules in the human body. Therefore, robust methods and standardized software, both available to computational experts and layman users, are required to disentangle and discover patterns in biological data sources.
This thesis is a collection of manuscripts in which we developed computational tools or methods to extract biomarkers from breath and genetic data.

In the first manuscript, BreathPy is introduced, a python library for the analysis of breath gas data. It implements state-of-the-art methods for preprocessing, automatic analysis-pipelines featuring cross-validation, and visualization capabilities for breath data. BreathPy is the first open-source library to support the processing of MCC-IMS data, is compatible with other analysis platforms and available in the python library ecosystem.

The next manuscript introduces BALSAM, our web-platform for the fully automated analysis of breath data. Aimed at clinical researchers, it offers an intuitive interface and enforces machine learning best practices for the reproducible discovery of biomarker patterns. We present two example studies that highlight the platform's capabilities and compare the detected markers with previous studies.

In manuscript three, we investigate patterns in the nATF6 mediated development of colonic tumors. For this purpose, we analyzed gene-expression data of transgenic mice and developed a novel feature selection approach. As a result, we present genes showing a strong allele dosage effect, enriched pathways, and a set of genes predictive of tumorigenesis.

The last manuscript highlights improvements to the multi-omics integration software netDx. In addition to improved performance and usability, it introduces four use-case examples that demonstrate the integration of several omics data types using patient similarity networks.

netDx is available on GitHub and has been added to the bioconductor gateway, enabling a streamlined installation process.

While complexities in metabolite identification and feature selection remain, the integration of high-throughput omics and advancements in non-invasive technologies promise improvements in the characterization of molecular processes. We developed tools for the discovery of such biomarkers and made them accessible on widely used community platforms and the web.