

## BEYOND ASSOCIATION AND CORRELATION: DATA MINING BASED ON REAL WORLD MICROBIOME PROFILES

Keywords: Data mining, Real-world dataset, Nutrition, Microbiome

Isabel Dorst<sup>1</sup>, Plamena Dikarlo<sup>1</sup>, Dr. Kai Stieger<sup>1</sup>

<sup>1</sup> BIOMES NGS GmbH; Wildau – Germany

**Objective:** Microbiome studies have identified a variety of intrinsic and extrinsic factors associated with specific patterns of the gut microbiome. However, most large microbiome datasets display a lack of methodological consistency and usually examine only one microbial snapshot, thereby neither representing intraindividual dynamics over time nor allowing causal inferences. Moreover, there is no real-world evidence that accurately reflects circumstances of daily life and natural consumer behavior. This study provides insights into typical application of BIOMES' complex, heterogeneous and population-based microbiome data.

**Methods:** Up to 80 variables of BIOMES' customer profiles on environment, diet, lifestyle, anthropometrics and medical history as well as simulated 16S rDNA amplicon sequence data are showcased in this illustrative data use case. A multivariate and compositional analysis of grouped variables is conducted to rank the impact of various host traits on the gut microbiota. The most influential variable group is further analyzed by correlating aggregated indices with genus abundances. Lastly, a stratified longitudinal analysis is pursued for one highly correlated variable to determine a potential causal link to the identified microbial change.

**Results:** Overall, host diet is identified as the factor most strongly associated with the gut microbiome. This association is further explored and several interesting links between dietary patterns and the gut microbiome are revealed. The strongest link is found between a particular genus and a nutritional index that reflects the consumption of high-fibre foods, suggesting a potential fibre-degrading property of the microbe. This hypothesis is further supported by the finding that said genus increases significantly over time only when fibre consumption is increased between microbial testing, whereas abundance remains unchanged in case of unaltered fibre consumption. This evidence strongly hints towards a putative causal relationship between dietary fiber consumption and the investigated genus.

**Conclusion:** This illustrates how BIOMES' dataset of over 40,000 microbiome profiles are used for exploratory purposes, hypothesis generation, and hypothesis validation. Cross-sectional and longitudinal real-world data can be used to establish correlations and possible causality to address the current gap in microbiome research.