Mass spectrometry-based metatranscriptomics, which characterizes the expressed microbial proteins within a complex ecosystem, has emerged as an insightful method to unravel the mechanisms underlying microbiome dynamics. Metatranscriptomics data analysis presents challenges, including large protein sequence database searches which can lead to low sensitivity and/or detection of a high proportion of false positives. The emergence of deep learning-based predicted spectral library searching methods offer an opportunity to improve the detection sensitivity. In this study, we use a ground-truth dataset to test the capabilities of deep learning-based spectral library searching.

THE DATABASE

A ground-truth dataset of digested mixture of 32 microbial species and strains of Archaea, Bacteria, Eukaryotes, and Bacteriophages with known species abundances was used (Kleiner M et al (2017)). Some of the bacterial strains were very closely related, but still distinguishable at the protein and nucleotide sequence level. The uneven mock community was designed to cover a large range of species abundances both at the level of cell number and proteincnc biosmasses to test for the dynamic range and detection limits of the quantification methods.

Results

Organism abundance detection decreases as database size increases for all levels of organism abundance. Scribe shows consistent decrease for all organism abundance levels.

The ground-truth dataset was searched against 1X (original database), 2X (original database + 1X IGC database), 4X (original database + 3X IGC database) and 10X (original database + 3X IGC database) databases using MaxQuant, FragPipe and Scribe search algorithms.

Organism abundance was calculated for high-abundance organisms, intermediate abundance and low-abundance based on protein abundance levels for the organisms.

Quantitative measurements from the three search algorithms was compared with known values (Reference%) for the four database searches.

REFERENCES


5: Scribe detected low-abundance phage proteins such as M13 phage, ES18 phage, F0 phage and F2 phage.

Only Scribe detected low-abundance phage proteins such as M13 phage, ES18 phage, F0 phage and F2 phage.

Quantitative analysis of the DDA-M5s data using peptide intensity values showed that none of the algorithms performed well. We plan to use this database to determine an appropriate method of quantitative estimation.