[BIOINFO 2023]

Microbial Risk Score for disease prediction using amplicon or shotgun metagenomic sequences

Boram Kim[1], Taesung Park[2,*]

[1] *Interdisciplinary Program in Bioinformatics, Seoul National University*
[2] *Department of Statistics, Seoul National University*
**Corresponding author: tspark@stats.snu.ac.kr*

Numerous human microbiome studies have discovered that the microbiome plays a crucial role in various diseases. These studies provide great opportunities to investigate the microbiome's potential use in evaluating and predicting disease status. In genetic studies, a polygenic risk score has been widely used to measure genetic risk for disease. Similarly, a microbial risk score based on alpha diversity has been developed. However, alpha diversity generally has the low predictive performance. Motivated by the polygenic risk score, we propose the weighted sum-based microbial risk score (wsMRS) by aggregating microbial profiles into a simplified score. Specifically, wsMRS is calculated by summing the weighted abundances of disease associated microbes. We demonstrate the utility of wsMRS for disease prediction through four real data sets. We compare the performance of wsMRS with existing alpha diversity-based MRS. We show that wsMRS outperforms alpha diversity-based MRS. Furthermore, combining wsMRS and alpha diversity-base MRS improves prediction performance. For example, wsMRS for hepatocellular carcinoma data achieves an AUC of 0.96 on the testing dataset, which is higher than the test AUC of 0.8 for alpha diversity-based MRS. The combined MRS shows a test AUC of 0.97.