Early Detection of Cancer Using Microbial Taxonomic and Functional Information

Ding, Yue (School: Nikola Tesla STEM High School)

Microbial DNA is a novel analyte for early cancer detection. Although recent studies have found distinct communities of bacteria and fungi between tumor types, they have not investigated microbial functions associated with these taxa as potential biomarkers. Such functional information may improve the accuracy of predicting cancer presence and type, including when combined with taxonomic data. Therefore, this project explores the diagnostic utility of tumor and blood-derived microbial functional information alongside microbial taxonomic information in more than 2000 cancer tissue and blood samples. Specifically, 75 machine learning models were trained and evaluated with microbial information to discriminate between (i) tumors of 10 cancer types and (ii) blood samples from patients with 7 types of cancer versus healthy controls. In general, functional data provides weaker predictive performance compared to taxonomy, perhaps due to greater conservation of tumor microbiome function than taxonomic community composition. Synergism between taxonomic and functional information was observed most among cancer tissues, with an 18.04% relative increase in mean balanced accuracy over equivalent models built with taxonomy or functions alone, achieving an accuracy of 82.10% and an AUROC of 0.9249. This synergy was weaker in blood/plasma, potentially due to the fragmented nature of cell-free DNA, making it difficult to obtain full function representation. Differential abundance analysis identified 190 taxonomic and 118 functional biomarkers, all statistically significant with a p-value < 0.05. Overall, combining microbial functions with taxonomy may provide more powerful diagnostic tests, enhancing early detection of multiple types of cancer and improving patient survival.

Awards Won:
Third Award of $1,000