Functional Modeling of the Gut Microbiome Reveals Systemic Disease-Associated Signatures Related to Extracellular Protease Activity

Chakravarthy, Arhan (School: Homestead High School)

Systemic inflammatory diseases like cirrhosis, ankylosing spondylitis (AS), and hypertension affect over 1 billion people and are often detected after significant organ damage. Growing evidence links gut microbial composition to these diseases, yet most studies overlook bacterial function, such as extracellular protease (ECP) production. Because ECP production varies among bacteria and microbiome composition varies between individuals, taxonomic signatures are inconsistent disease predictors. Although ECP activity is linked to intestinal barrier disruption and inflammation, its non-invasive biomarker potential remains unexplored. Here I introduce a framework for evaluating the gut microbiome based on its ECP capacity. I mapped 116 bacterial species to their ECP production by screening genomic data for catalytic domains and secretion signals. I then engineered a metric, Community Proteolytic Potential (CPP), weighting normalized bacterial abundances with these ECP profiles to quantify a microbial community's functional output. I applied CPP to stool metagenomic data from two controls (healthy individuals; Type 1 Diabetes) and three patient cohorts: cirrhosis, AS, and hypertension. My analysis revealed that specific ECP classes drive consistent CPP score elevations across all patient cohorts ($p < 0.05$), suggesting these distinct diseases share a common association with high proteolytic activity. CPP scores in T1D negative controls where microbiome disruption is not protease driven showed no shift. Classifiers trained on CPP features outperformed taxonomic models across all three diseases, achieving ~10% better discrimination between diseased and healthy samples (AUCROC). My findings suggest ECP signatures extend beyond the intestine and may inform stool-based diagnostics.

Awards Won: