A METHOD TO IDENTIFY MICROBIAL METABOLIC IMMUNOREGULATORS OF COLORECTAL CANCER: AN AI DISCOVERY ENGINE

Eric Luellen, MSc, MPH, Chief Data Scientist, Bioinformatix.io

Introduction
The microbiota-generated metabolome is recently understood to significantly impact initiating, progressing, and proliferating cancers and predict immunotherapy efficacy. Two challenges have prevented breakthrough results: identifying immunoregulators is glacial in wet-lab trial-and-error discovery because of complexity, ~ 1.69×10^{14} permutations, and historic computational analysis errs using linear, one-to-one biostatistical analysis instead of conditional, complex analysis. Thus, new tools are urgently needed. This study proposes and shows proof of a new method using machine learning to discover the defining microbial-metabolic immunoregulators of colorectal (CRC) cancer at scale rapidly.

Objectives
The objective of this study was to determine if machine learning could quickly identify the microbial-metabolic immunoregulators of CRC into accurate and replicable predictive models.

Methods
This study analyzed annotated secondary integrated microbiome-metabolome datasets from fecal samples for CRC versus healthy controls with proprietary contests of machine learning algorithms. It then juxtaposed the results against 24 papers published in peer-reviewed journals.

Results
Predictive models were trained and tested using a data frame of 347 observations across 9112 gut bacteria and metabolites, yielding an average accuracy of 91.7% (confusion matrix) and 98% (area under ROC curve). While two dozen previously published CRC studies identified 13 CRC immunoregulators (eight bacteria and five metabolites), this study identified 15 CRC immunoregulators (eight metabolites and seven bacteria) 77% of which were different, and most of which appear novel in the literature. Moreover, the 23% overlap were less important in the AI models, and two critical bacteria appear also connected to periodontal health. Finally, while an integrated microbiome-metabolome model was superior, a metabolite-only model was only .15% less accurate.

Conclusions
Robust machine-learning tools appear capable of sorting through large orders of microbiome-metabolome permutations to discover candidate biomarkers, diagnostic,

and therapeutics for CRC, yielding new insights into mechanisms of action and accelerating discovery by orders of magnitude. This method goes beyond linear associations to mathematically model conditional and complex relationships suitable for causal inference. It is a Rosetta stone at best and a hypothesis generator at worst. Moreover, this method may be generalizable to disease states and phenotypes beyond CRC.

Keywords: Microbial-metabolic immunoregulators, colorectal cancer, AI