SMbiot: A Shared Latent Model for Microbiomes and their Hosts

Madan Krishnamurthy[1], Lukas Herron[2], German Plata[3], and Purushottam Dixit[4,5]

***Short Abstract* — Animal hosts are colonized by microbial species across several kingdoms and host sites that form a tightly knit metacommunity. Yet, we do not have models that integrate these data and predictively capture the context-specific variation in microbial taxa abundances in different communities and the hosts' phenotypic states. To that end, we present SMbiot: Shared Latent Model for MicroBIOmes and their HosTs, a dimensionality reduction framework that integrates measurements of several microbiomes and host phenotypes for predictive modeling. Using two important animal host systems, we show that SMbiot predicts microbial abundances across various organ sites and kingdoms of life as well as host phenotypic variables. We then use SMbiot to identify microbial taxa that are predictive of specific host phenotypes, in a context-specific manner, leading to novel biological insight. We experimentally validate SMbiot using 16S sequencing data to predict the gut metabolomic response to antibiotics in poultry. SMbiot will be a significant tool to investigate mechanistic relationships between multidimensional and multimodal data.**

***Keywords* — Microbiome, Host-microbiome interactions,**

I. Overview

Animal associated microbial communities colonize multiple organs, comprise microbial taxa across kingdoms of life, and are now recognized to be important determinants in host health as well as environmental footprint. Notably, the associations between microbial taxa and host phenotypes are context-dependent, i.e. effect of microorganisms on host phenotypes depends on the presence/absence of other organisms and vice versa.

New developments in high and low throughput multi-omics technologies provide a multimodal picture of animal hosts and their associated microbial communities, including abundances of several microbes at several host sites and across kingdoms of life, and phenotypic details of the host. Correlation-based analyses can identify specific microbe-phenotype associations. However, the correlation between host variables and specific species are typically weak, and as a result, the high dimensionality of *omics* datasets often limits the power to identify statistically significant associations due to the very large number of tested hypotheses. Regression methods identify models to predict phenotypes/bacterial abundances across all data and therefore cannot find context-dependent interactions. Modern approaches such as neural networks are in principle able to model context-dependence but typically require large training data sets.

Shared latent space approaches have been used to integrate coupled variation in multi-omics information about the same samples. We present SMbiot: A **S**hared Latent **M**odel for Micro**BIO**mes and their Hos**T**s, to predictively model the covariation between host-associated microbiomes and host phenotypes. Briefly, in SMbiot, we use a shared latent space representation that simultaneously models available microbiome data, for example, microbiota compositions at different host sites or from different kingdoms of life, as well as multidimensional data on the hosts' phenotypic states. SMbiot can obtain an approximate latent space representation when only microbiome or host phenotypic data are available, thereby allowing us to predict unobserved microbiome or phenotypic data. SMbiot can systematically identify context-dependent interactions between microbial taxa and host phenotypes, thereby generating novel biological hypotheses.

Using multiple datasets from livestock and humans we demonstrate the accuracy of SMbiot in quantitatively describing and enabling predictions of host phenotypes and the microbiome. SMbiot further allows us to identify a handful of host variables that can accurately predict the microbiome composition or subsets of the microbiome that best predict individual phenotypic variables. SMbiot can be used to jointly model and predict microbial communities at different host body sites or of microorganisms in different kingdoms of life. Finally, we apply SMbiot to predict changes in the gut metabolome of chickens upon antibiotic treatment based on 16S rRNA sequencing data and show that model predictions agree with experimental data. We believe that SMbiot will be an invaluable predictive model of animal hosts and their associated microbiomes.

Acknowledgements: This work was funded by NIH grant 1R35GM142547-01.

[1]Elanco Animal Health Inc. Greenfield, Indiana
[2]Institute of Physical Sciences and Technology, University of Maryland
[3]MBRD Service Company, Indianapolis, Indiana
[4]Department of Physics, University of Florida, E-mail: pdixit@ufl.edu
[5]Department of Chemical Engineering, Cancer Center, and Genetics Institute, University of Florida