2017 Innovation Lab: Quantitative Approaches to Biomedical Data Science Challenges in our Understanding of the Microbiome

Goal

The goal of the 2017 Innovation Lab is to foster the formation of *new* interdisciplinary collaborations that will generate creative strategies for addressing challenges associated with the visualization, modeling, and analysis of microbiome big data. Such challenges arise from multifaceted data structures like networks and images, missingness or sparsity of data, streaming of non-stationary time series data, the need for integration from multiple sources of data, interaction effects, etc. This Innovation Lab is intended to bring together expertise from the mathematical, statistical, and biomedical fields, to address interdisciplinary topics in biomedical data science critical to the study of the microbiome (In this document, the term "biomedical" will be used in the broadest sense to include biological, biomedical, behavioral, ecological, social, environmental, and clinical studies that relate to understanding health and disease). Microbiome is defined for the purposes of this workshop as the relevant data associated with the microbiota with a health or biomedical research objective (e.g. the underlying metagenomic diversity, metabolomic profiles or other high-dimensional systems biology data coming from such organisms and the environments they inhabit). It is anticipated that collaborations formed during the Innovations Lab will result in NSF or NIH grant proposals to further develop, refine, and test hypotheses.

Description

The Innovation Lab will promote collaboration between mathematicians, statisticians, and biomedical researchers towards the development of novel or significantly adapted models, methods, and approaches for overcoming difficult data science challenges arising from the collection and analysis of microbiome big data. The study of the microbiome is a rapidly developing area with the aim of identifying, treating, and preventing disease as well as promoting health by understanding the interaction of microbes with humans in a variety of ecological niches both within the human body and the external environment. Achieving the potential of microbiome big data to transform our knowledge will necessitate the integration of data formats, including structured and unstructured data, real-time and batch data, and observational and experimental data, from diverse populations interacting with distinct environments. Integrated data will require the development of computational and analytical methods to enable high-confidence for researchers or doctors with a personalized user interface and experience (UI/UX).

Network analysis, causal analysis, real time data analysis, UI/UX, and machine learning are among the important tools for modeling, simulation, and visualization of large-scale, multi-dimensional, integrated data coming from microbiome data sources. The data on which these methods will be used are complex due to what is not present (missing data), and also what is implicit, such as an underlying network structure of dependencies or the trends, cycles, or other patterns arising from high-frequency, non-stationary, time series data. Collaborations between biomedical and quantitative scientists (e.g. mathematicians, statisticians, computer scientists, physicists, and engineers) with relevant expertise could lead to better approaches to implementation and interpretation of such information.

2017 Innovation Lab: Quantitative Approaches to Biomedical Data Science Challenges in our Understanding of the Microbiome

At the Innovation Lab, interdisciplinary teams will work together to ideate and develop pilot projects for tackling selected problems. Potential biomedical topics for the Innovation Lab range across scales, from real-time molecular biomarker monitoring from complex subsets of microbial populations to observational data on behavior in humans that are affected by the microbiome. Topics may lead to new insights and lay the groundwork for future advances in the study of the microbiome. Examples could include inferring behavioral change or mental state from interactions with microbiota, predicting the health outcomes in subpopulations of humans by integrating biomarker data from high-dimensional data from microbiota with more traditional clinical or genomic data sources, analyzing the effects of microbial interventions on health or disease outcomes.

Publicly available multi-scale, multi-modality data that may be used include, but are not limited to, the Human Microbiome Project, Earth Microbiome Project, Human Oral Microbiome Database, LINCS data, dbGaP data.