**Gut Microbiota, Host Plasma Metabolomics, Diet & Cardiometabolic Risk in Urban vs Rural Chinese**

**Contact: Penny Gordon-Larsen PI pglarsen@unc.edu**

We have an R01 to examine microbiome and metabolome changes in association with environmental and diet change and how those changes together may modulate risk of inflammation mediated diseases. China is a perfect setting in which to investigate this question as it is rapidly modernizing country experiencing substantial transition from traditional to Westernized diet at different points in time, and across diverse geographic regions, along with increasing burden of obesity, diabetes and inflammation, providing a unique model for examining development of risk. While the larger R01 is just underway we have pilot data ready for analysis.

The data in this project come from the China Health and Nutrition Survey (CHNS), an NIH-funded study of more than 15,000 individuals followed over 25 years, provides high quality longitudinal data and captures the transition from traditional to Western diet in parallel with urbanization and emergence of obesity, hypertension, insulin resistance, type 2 diabetes and cardiovascular disease during the past two decades. We will use these data to generate insights that would not be obtainable in studies with subjects who consume only a Westernized diet. Using sophisticated statistical models we propose to examine whether information on gut microbiota composition along with data on the plasma metabolome can predict specific health outcomes, allowing us to implicate microbiota in disease pathways. We have collected fecal samples from 10,000 adult CHNS participants aged 30-65 at the 2015 exam. In a subsample from two neighboring Southern provinces (Guizhou, low income, actively urbanizing, n=600; and Hunan, high income, already urbanized, n=600) varying in current (and long-term change) in Westernized diet and sequence the fecal samples. We will select a subsample with 16S data who have consumed a traditional diet over the 25 years (n=400) and conduct plasma metabolomics of the host. In the CHNS subsample with microbiota plus metabolome (n=400) data, we will collect replicate blood and fecal samples in CHNS2017 to derive 16S and plasma metabolomics data to assess two-year changes (2015-2017) in diet, gut microbiota, plasma metabolites and in markers of cardiometabolic disease (body mass index, central adiposity, diabetes and inflammation). We will examine whether gut microbiota and plasma metabolites differ depending upon when, within the 25-year period, diet changes occur, and if they are associated with health outcomes. In our longitudinal subsample we will examine changes in markers of Western diet in relation to concurrent changes in microbial diversity and community composition, and in metabolites. We first use a series of standard regression models and then high dimensional regression analysis with variable selection and validation to build predictive models. We capitalize upon an established and well-characterized, large cohort with far greater variability in diet and microbial communities than studies in cohorts on only Western diets. The proposed project will substantially transform current understanding of the intersection of diet, gut microbiota, host metabolism and cardiometabolic disease.

Our preliminary studies suggest metabolite microbial changes linked to urbanization consistent with a model in which selection pressure rapidly impacts the relative abundance of a metagenomic pool of microbes through niche specialization, driving selection at more derived levels of the phylogenetic tree. While a large majority of the taxa in our study could be mapped to microbes sampled from an American cohort from the Human Microbiome Project (HMP) with high identity, microbes with higher relative abundance in Chinese urban samples were substantially more prevalent in the HMP. Most of the preliminary work has been with the 16S data and we have begun to work with the Whole Genome Sequence data.

We have two projects: (1) we have pilot data for immediate analysis including 16S and Whole Genome Sequencing as well as host plasma metabolomics and a vast array of exposure and outcome metadata in Hunan rural (n=20) and urban (n=20) Chinese adults; and (2) forthcoming 16S and Whole Genome Sequencing as well as host plasma metabolomics and a vast array of exposure and outcome metadata in 1200 Chinese adults. We are eager to bring students on to work with the existing 16S, whole genome, and metabolite data.

We have a vibrant and cohesive research group for this project that represents a decade’s work with the China Health and Nutrition Survey data that forms the basis of the project. The project is led by PI: Gordon-Larsen (NUTRITION) and includes collaborators in Nutrition (Popkin), Biostatistics (Herring, Howard), Bioinformatics (Fodor & Wei, UNC-Charlotte; Wu & Li, U Penn). We currently have a team of students and postdocs with expertise in nutrition, biostatistics, and microbiology. We are eager to add students in bioinformatics. The primary areas of interest we are looking for in a BD2K student are: bioinformatics, machine learning, innovative tools to reach across the microbiome, metabolome and nutrient data, and novel tools for dealing with repeated testing/false discovery.