The Children's Hospital of Philadelphia

Annual Progress Report: 2014 Formula Grant

Reporting Period

January 1, 2015 – June 30, 2015

Formula Grant Overview

The Children's Hospital of Philadelphia received $3,695,893 in formula funds for the grant award period January 1, 2015 through December 31, 2018. Accomplishments for the reporting period are described below.

Research Project 1: Project Title and Purpose

The Human Microbiome in Pediatric Health and Disease – Our purpose is to study the pediatric microbiome with the goal of improving diagnosis and treatment of disease. Humans live in association with myriad microbes that strongly influence health and disease. Proper microbial populations are a key component of health. Dysbiotic microbes are associated with diseases such as pediatric inflammatory bowel disease (IBD). This project will support research efforts to better understand the microbiome in healthy pediatric patients and pediatric patients with IBD. It will further investigate the effects of antibiotic exposure, *Clostridium difficile* (*C. difficile)* infection, the metabolome, and human genetic variations on microbial populations in these children.

Anticipated Duration of Project

1/1/2015 – 12/31/2018

Project Overview

Humans live in association with myriad microbes that strongly influence health and disease. Our guts contain approximately one hundred billion microbes per gram of intestinal contents. Distinctive microbial communities also inhabit our mouth, skin, lungs, and other body sites. These communities are diverse and comprised of bacteria, fungi, archaea, and viruses. Studies of the human microbiome have exploded in recent years, in part as the result of the availability of ultradeep sequencing methods and bioinformatic tools for the analysis of the resulting large data sets. The overall goal of "The Human Microbiome in Pediatric Health and Disease" is to characterize the microbiome in healthy children and children suffering from inflammatory bowel disease (IBD). Pediatric IBD is unfortunately increasing in incidence and prevalence. Microbiome samples including stool, oral and rectal swabs will be collected from 100 control, 100 IBD and 25 very early onset IBD pediatric subjects. Data collected associated with the

Note: This project is funded, in part, under a grant with the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations or conclusions.

samples includes information pertaining to diagnosis, disease location, demographics, procedures environmental and family history, and medications. Microbial DNA will be extracted and analyzed by various sequencing techniques to determine the microbial populations. We believe that this comprehensive multidisciplinary project of the pediatric microbiome will set the stage for improving diagnosis and treatment of children suffering from IBD and other disorders. We propose to accomplish the following 4 aims:
Aim 1. Survey the breadth of microbial populations in healthy children at multiple body sites.
Aim 2. Investigate the associations of microbial population structure and inflammatory bowel disease in children, including the burgeoning variant very early onset IBD (VEO-IBD--defined as IBD initially diagnosed in children <5 years of age).
Aim 3. Assess the effects of antibiotic exposure on microbial populations resident in healthy and ill children, and the association with *C. difficile* infection.
Aim 4. Assess interactions among human genetic variation, microbial population structure, antibiotic use, metabolomics and outcome in pediatric IBD.

Principal Investigator

Robert N. Baldassano, MD
Professor of Pediatrics
The Children's Hospital of Philadelphia
34th Street and Civic Center Boulevard, GI 7NW
Philadelphia, PA 19104

Other Participating Researchers

Judith R. Kelsen, MD – employed by The Children's Hospital of Philadelphia
Frederic D. Bushman, PhD - employed by University of Pennsylvania
Gary D. Wu, MD – employed by University of Pennsylvania
Kyle Bittinger, PhD – employed by The Children's Hospital of Philadelphia
Jessi Erlichman, MPH – employed by The Children's Hospital of Philadelphia
Maren Wennberg, BS – employed by The Children's Hospital of Philadelphia
Alejandro Martinez, MS – employed by The Children's Hospital of Philadelphia

Expected Research Outcomes and Benefits

The project will generate a deep sequencing data set of microbiome samples from 100 healthy children, 100 children with IBD, and 25 children with VEO-IBD. This will make possible an exploration of the role of dysbiosis in pediatric IBD, allowing us to assess the design of improved techniques for diagnosis and treatment of pediatric IBD. These data sets will be uploaded to public archives and will provide a reference set for many workers to investigate in further detail. We will also provide detailed data on the effects of antibiotic use which will help guide the design of optimally effective antibiotic therapies. In addition, we will investigate the genetic structure of the pathogen *C. difficile* in pediatric samples, with the goals of tracing the origin of strains and developing effective targeting treatments. Lastly, we will characterize the relationship of microbiome composition, human genetics, and circulating serum metabolites.

This will advance our understanding of pediatric IBD and potentially provide biomarkers for improved diagnosis and improved forms of therapy.

Summary of Research Completed

During the research reporting period, the researchers accomplished the milestones established for the reporting period. This includes the following accomplishments: completed preliminary hiring, completed IRB approvals, performed method testing for swab storage, and tests of human DNA proportions in swabs and stool used in the study. Detailed descriptions of these research efforts is described below:

Begin hiring staff

Identification and hiring of qualified personnel to support the grant research project was initiated. In the reporting period, we hired one research coordinator, one technician, a data integration analyst, and a project manager.

Complete acquisition of IRB approvals

All protocols being used within this study have received Institutional Review Board (IRB) approvals. This includes amending the previously cited IRB protocols to encompass work specific to this study and increase sample size where warranted. The IRB protocol titled, "The Human Microbiome in Pediatric Health and Disease" was submitted and approved by the Children's Hospital of Philadelphia IRB to allow use of data and specimens collected in one of the existing IBD studies. Prospective subjects can be enrolled directly into "The Human Microbiome in Pediatric Health and Disease"

Tests of human DNA proportions in swabs and stool

To evaluate the accuracy of human-derived sequence attribution, we simulated sequence read pairs from a variety of genomic sources. Synthetic datasets of 10,000 read pairs were generated with 50% derived from human, 40% from bacterial, 9% from viral, and 1% from phage ΦX174 source genomes. Datasets were generated at read lengths of 50-250bp and simulated sequencing error rates of 0-5%. The synthetic datasets were used to analyze 6 popular human filtering and alignment tools: BLAST-Like Alignment Tool (Blat), Best Match Tagger (BMTagger), Best Match Filter (Bmfilter), Bowtie2, Burrows-Wheeler Aligner (BWA), and Scalable Nucleotide Alignment Program (SNAP).

BMTagger and the related program Bmfilter misclassified large percentage of human reads as non-human (false negatives, Figure 1). SNAP performed well with 2% sequencing error rate but had an increased false negative rate at 5% sequencing error. BWA and Bowtie2 had the lowest overall rate of false negatives, but both tools misclassified a small percentage of non-human reads as human (false positive assignments). In the case of BWA, we were able to remedy the false positive errors by introducing thresholds for percent identity and alignment length.

BWA was selected as default aligner for human filtering and was applied to ten well-understood fecal samples obtained from the Sequence Read Archive (SRA) at National Center for Biotechnology Information (NCBI). The percentage of human reads observed in these samples ranged from under 10% to about 98% in subjects with IBD (Figure 2).

Swab storage study

As a step in preparing for large-scale sampling, a study of the storage conditions suitable for oral swabs was completed. Swabs were stored under 10 different conditions for up to one week. Conditions involved various combinations of time intervals and storage at three temperatures (room temperature, 4^{o}C, and -80^{o}C).

After storage, swabs were cut into tubes, DNA was extracted, bacterial 16S ribosomal ribonucleic acid (rRNA) gene segments from the first and second variable region (V1V2 region) were polymerase chain reaction (PCR) amplified with bar-coded primers, and amplicons were sequenced using the 454/Roche Pyrosequencing method. Sequence reads were parsed into samples by bar code, then sequence reads formed into operational taxonomic units (clusters) based on 97% identity. Each sample was expressed as a vector of counts for each operational taxonomic unit (OTU). Taxonomy was assigned using the Ribosomal Database Project (RDP) database, which showed that Streptococcus predominated in all of the swab samples, and that many other lineages were present in each sample as well.

To assess the effects of storage time and temperature, distances between all pairs of samples were calculated, then sample similarities were assessed using principal coordinate analysis (Figure 3). In this display, samples more similar to each other are more nearby, those more dissimilar are farther apart.

Samples from the three individuals are colored yellow, blue, and red. As can be seen in the plot, individuals are notably different from each other, and this dominates over effects of swab storage. A more detailed statistical analysis showed no major effects of storage conditions, even when samples were stored at room temperature for several days.

We conclude that results for oral samples will not usually be affected by details of the storage methods. Going forward, as a conservative approach, our standard operating procedures will involve storage on ice as early as possible after sampling and transfer to -80 ^{o}C as early as possible, but we do not expect storage time to be a major source of variation in our experiments.

Figure 1: Removal of human sequence from whole-genome shotgun metagenomics data. Synthetically-generated datasets of 10,000 read pairs were used to analyze human sequence filtering results from Blat, Bmfilter, BMTagger, Bowtie2, BWA, and SNAP. The false positive and false negative rates for human-derived reads are shown for a range of read lengths and sequencing error rates. BMTagger misclassified approximately 25% of human-derived reads as non-human. BWA and Bowtie2 had the lowest rate of false negative human classifications, and a low rate of false positive attributions.

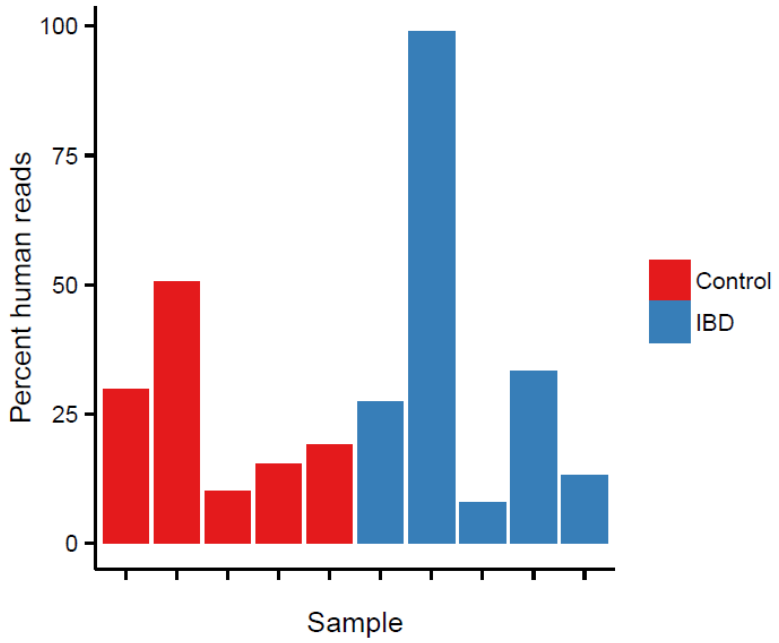

Figure 2: Removal of human sequence from whole-genome shotgun metagenomics data. Ten well-characterized fecal samples were filtered for Human reads using the best parameters determined by synthetic datasets. As much as 98% of the reads in one sample, from a subject with Inflammatory Bowel Disease (IBD), were attributable to human genomic sources.

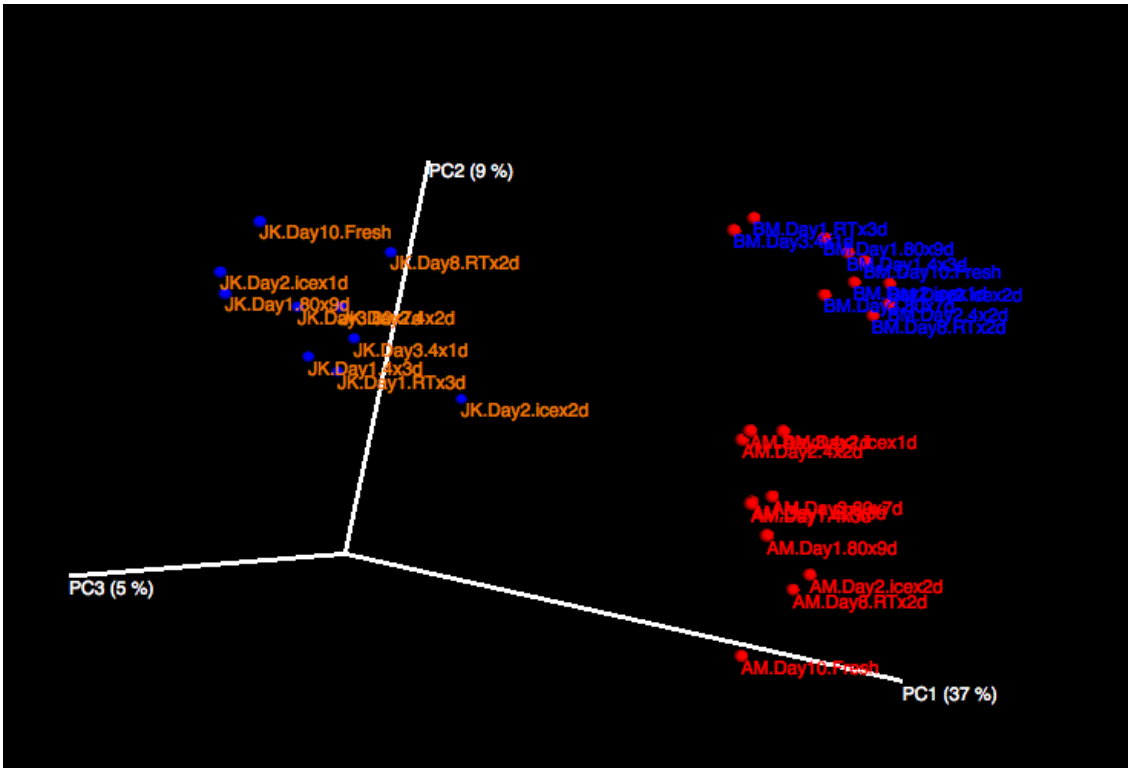

Figure 3: Analysis of temperature and storage time on sample pairs