# PROJECT SUMMARY

## The Human Virome in Children and Its Relationship to Febrile Illness.

Gregory Storch MD, Washington University

## I. PROJECT ID NUMBER, PUBLICATION MORATORIUM INFORMATION, PROJECT DESCRIPTION:

This manuscript is part of a pilot effort on the part of NIH staff and the Nature publishing group to provide a more convenient archive for "marker papers" to be published. These "marker papers" are designed to provide the users of community resource data sets with information regarding the status and scope of individual community resource projects. For further information see editorial in September 2010 edition of Nature Genetics (*Nature Genetics*, **42**, 729 (2010)), and the Nature Precedings HMP summary page.

Project # 46335. No publication moratorium.

Definition of the human microbiome is an important scientific priority. This study will expand the scope of the investigation to include viruses, which account for a substantial proportion of infectious disease morbidity and mortality, especially in children. The long-term goal of this project is to describe the human virome in children and to investigate its relevance to febrile illnesses in children. The project will also seek to understand the relationship of the immune system to the composition of the virome. Thus, the project's specific aims are 1) To elucidate the spectrum of viruses that can be detected using non-biased, high throughput sequencing on samples of blood, respiratory, and gastrointestinal secretions from healthy children and to use this information as a basis for understanding the role of viruses in acute febrile illnesses without an obvious source, and 2) To investigate the effect of various forms of immunosuppression on the spectrum of viruses detected in children, and to use this information as a basis for understanding the role of viruses in acute febrile illnesses occurring in these children. Our preliminary studies show that diverse viruses can be detected in children having undiagnosed fever. To carry out the specific aims, well children will be enrolled prior to having elective surgery, and febrile but otherwise well children will be enrolled from the Emergency Department at St. Louis Children's Hospital. Immunocompromised children will be recruited from hematopoietic stem cell and solid organ transplant clinics, the HIV/AIDS clinic, and the rheumatology/immunology clinic from the same hospital.

Children with fever will have samples obtained at the time of the febrile illness and at 1 and 6-month follow-up visits. Selected samples from each study group will be analyzed at the Genome Center at Washington University (GCWU) using next generation 454 high throughput

sequencing to detect and sequence all viral sequences present. We anticipate detecting and sequencing a broad range of viruses, including previously unrecognized agents. A variety of techniques will be used to investigate the significance of viruses detected.

Virus-specific PCR assays will be used to determine the frequency and extent of viruses detected by sequencing, using the full range of samples collected. Host response to the detected viruses will be investigated using serologic analysis, cytokine profiling, and microarrays to characterize host gene expression.

These studies will take advantage of follow-up samples to compare the acute response with the response in the convalescent period. This study will draw upon the expertise and technological assets of one of the world's most powerful sequencing centers to provide the research community with a comprehensive sequence data base of the viruses that are present in children, which can be used to improve our understanding of the causes of febrile illnesses in young children, many of which are currently undiagnosed.

## II. DATA QUALITY:

The quality of capillary sequencing data (Sanger sequencing on the AB3730 instrument) at the GCWU is measured by assessing the failure rate of each individual set of 96 lanes within one full run. Successful runs are those with fewer than 20% failures, although this number is often set more stringently. In addition, the expected read length at this time is 700 bases of a quality score greater than phred Q20.

The Illumina production pipeline is evaluated by the number of passing reads that contain high quality data. Successful runs are those producing an expected full set of reads with a low error rate. The full set of reads depends on the sequencing conditions while error rates are typically < 1%, analogous to phred Q20.

Sequencing on the Roche-454 platform is similarly evaluated by the number of reads and bases produced per run, the read length distribution, as well as an error rate in base calling.

## III. DATA ANALYSIS AND PUBLICATION PLANS:

Initial data analysis will focus on the results of panels of virus-specific PCRs performed on each sample type. We will compare the findings in infebrile patients and in afebrile control subjects. We will also compare viruses detected in different samples from the same subject. We will also compare the viruses detected in normal and immunocompromised children. We will then proceed to an analysis of high throughput sequencing results. We will use the PCR results as a basis for assessing the ability to detect viruses known to be present in the samples. We will relate detected sequences to known reference viral sequences, assess polymorphism of these known viruses, assemble more complete genomes when possible, analyze sequences that do not match known viruses or other microbes as potentially representing novel viruses. Demographic and clinical findings related to potential new viruses will be analyzed. We plan to submit papers reporting on virus-specific PCR data and shotgun sequencing and virus detection by the end of 2010.

## IV. DATA RELEASE PLAN:

The GCWU plans to release the sequence reads to the NCBI Trace Archive without delay and has a data submission pipeline in place that has been doing this for years. Deposition in dbGaP will be used for the open access Short Read Archive for nextgen data. We will also release genome assemblies without delay, although the NIAID policy allows for a brief delay. Again, we have a pipeline in place that has routinely submitted assemblies and annotations to GenBank and worked with them to resolve discrepancies.

## V. CONTACT PERSON:

Dr. Gregory Storch, Washington University, storch@kids.wustl.edu