PROJECT SUMMARY

Skin Microbiome in Disease States: Atopic Dermatitis and Immunodeficiency

Julia A. Segre, National Human Genome Research Institute, NIH
Heidi H Kong, Center for Cancer Research, National Cancer Institute, NIH

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 ID: 46333. No publication moratorium.

The Human Microbiome Project seeks to explore the diversity of microbiota that resides in and on the human body, including the skin, in health and disease. The NIH Intramural Skin Microbiome Consortium (NISMC) is a trans-disciplinary group of experts engaged in the practice of genomics, bioinformatics, large-scale DNA sequencing, dermatology, immunology, allergy, infectious disease, and clinical microbiology. Atopic dermatitis (AD, “eczema”) is a chronic relapsing skin disorder that affects ~15% of U.S. children and is associated with $1 billion of medical costs annually. AD is characterized by dry, itchy skin, infiltrated with immune cells. Colonization by *Staphylococcus aureus* (*S. aureus*) is ten-fold more common in AD patients and is associated with disease flares. We hypothesize that, in addition to *S. aureus*, AD may also be associated with additional microbes and/or selective shifts of commensal microbes that are relevant to disease progression. The NISMC seeks to define the microbiota that resides in and on the skin and nares of three patient groups, all of whom have eczematous lesions and are currently seen at the NIH Clinical Center: (1) AD patients; (2) Wiskott-Aldrich syndrome (WAS) patients; and (3) Hyper IgE syndrome (HIES) syndrome patients. Examination of the microbiome of patients with WAS or HIES syndromes, both rare immunodeficiencies, will advance our understanding of how an individual’s immune system shapes their cutaneous microbial community. We are performing a prospective longitudinal study that follows these groups of patient thorough the cycles of eczema flares, ascertaining clinical data and samples at each stage. To analyze the longitudinal samples obtained from both affected and unaffected skin sites, we will use an integrated approach including: (1) extensive characterization of bacterial diversity; (2) analysis of fungal diversity; (3) generation of whole-genome sequences of novel skin microbial isolates; and (4) metagenomic sequencing to assess microbial abundance and deduce metabolic activities. Examination of the microbial communities on affected and

unaffected skin of patients with eczema will yield insights into the gene-environment interactions relevant to these diseases.

II. DATA QUALITY:

For primary Sanger-based reads, the values for near-full length sequences include a 94.1% success rate, an average of 676 Q20 bases per sequence read, and a 92% read-pair success rate. For Roche/454 sequence reads, V1-V3, V3-5 and V6-V9 amplicons are all quality controlled to exclude reads that do not exactly match barcodes/molecular IDs, which are embedded in the primer sequences. All sequences are filtered for minimum read length and to remove contaminating human DNA by the NHGRI sequencing center.

III. DATA ANALYSIS AND PUBLICATION PLANS:

Data is analyzed with tools available within MOTHUR (DOTUR, SONS) and UniFrac. Chimeras are removed with ChimeraSlayer. Samples are compared intra-personally within longitudinal survey and inter-personally between affected subjects and unaffected (healthy controls). Sequencing results are correlated with patients' disease severity and specific affected status, such as performed in previous publications from this group. For full-length 16S rRNA sequencing, we resolve the bacterial phylogeny to the species level when possible to further power the studies. We anticipating submiting for publication the initial results of these 16S rRNA sequencing results within the next 12 months.

IV. DATA RELEASE PLAN:

We recognize the importance of sharing clinical and sequencing data in a timely and unrestricted manner, and have written a data sharing policy that addresses the requirements set forth by the NIH and NHGRI. Our IRB-approved consents and microbiome sampling protocol include release of microbial sequences and aggregate human sequences to public databases and coded human sequence data to access-controlled databases. We have deposited our 16S rRNA sequences in Genbank, linked to our NCBI genome project page (#46333). We have no data embargo for the use of these sequences by other investigators. We have deposited into dbGaP (Study Accession: phs000266) subject and clinical data, linked to microbial Genbank ID numbers. We are putting in place bioinformatic tools to ensure that only high quality data and metadata are produced, and are working to provide that data to the community as rapidly as possible through the HMP data acquisition and coordination center (DACC) and public sequence databases via dbGaP, SRA and Trace Archives.

Release of clinical data

Clinical data are managed using the LabMatrix software package (v 3.6, BioFortis Inc.). LabMatrix collects, organizes, and maintains clinical and molecular research data. Labmatrix is used to maintain detailed subject records that include patient history, phenotypic data, and the results of laboratory tests. These records are linked to biorepositories that track specimen collection, usage, and analysis. Access to data in the Labmatrix database is regulated in a

HIPAA- and institutionally-compliant manner. Detailed demographic and clinical data are collected for each subject upon entry into the study. Each subject is assigned a unique identifier and each body site is assigned a unique site code and these identifiers, along with the collection date, uniquely identify a clinical sample.

Consent is obtained to share coded information (not containing any unique identifying patient information) with researchers at NIH and other institutions. Consent is also obtained to provide this coded information in an access-controlled database. The clinical data relating to these studies will be submitted semi-annually to the database of Genotype and Phenotype (dbGaP) at NCBI. This database provides both open- and controlled-access to clinical data. Open access is typically granted to non-sensitive summary data. For instance, the prevalence of MRSA in different study populations could be considered non-sensitive since it does not reveal personally identifiable data. Access to genotypic or phenotypic data for a single study participant, however, would be access controlled.

Release of metadata associated with sequence traces or other types of data

Metadata are critical to the interpretation of the sequence data generated as part of the HMP. The Genomic Standards Consortium published standards regarding the minimum information for a genome sequence (Field et al. 2008), and work is ongoing to extend this to metagenomic data sets (Kottmann et al. 2008). Within the NISMC, we are implementing a standard set of metadata that will accompany reads and assemblies. Where possible, we are referencing standard ontologies and enforcing those ontologies using checks within the database.

Intellectual Property Management Plan

We adhere to the NIH's Best Practices for the Licensing of Genomic Inventions. Software, biological resources, technologies, and protocols developed with funds from this award will be made readily available for research purposes to qualified individuals within the scientific community after publication.

V. CONTACT PERSONS:

Clinical questions: Research Nurse, 888-NIH-DERM

Sequencing and analysis questions: Julia Segre, National Institute of Human Genome Research, NIH 301-402-2314, jsegre@nhgri.nih.gov