Human Microbiome Science
Vision for the Future

JULY 24 - 26, 2013

Owen White
Institute for Genome Sciences
University of Maryland
Welcome to the Data Analysis and Coordination Center (DACC) for the National Institutes of Health (NIH) Common Fund supported Human Microbiome Project (HMP). This site is the central repository for all HMP data. The aim of the HMP is to characterize microbial communities found at multiple human body sites and to look for correlations between changes in the microbiome and human health. More information can be found in the menus above and on the NIH Common Fund site.

Areas of Interest

hmpdacc.org

Human Microbial Sampling

16S RNA and whole metagenome sequencing of samples collected from 300 healthy human participants, to characterize complexity of microbial communities at individual body sites and to provide insights into functions performed by the human microbiome.

DACC Member Organizations

Related Sites

NIH Common Fund
NCBI HMP Data Repository
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1001377
Figure 1. Timeline of microbial community studies using high-throughput sequencing.

http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1001377
# Current and Previous RFAs

<table>
<thead>
<tr>
<th>Title</th>
<th>RFA</th>
<th>Expired</th>
<th>NIH Institute/Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Role of Microbial Metabolites in Cancer Prevention and Etiology (U01)</td>
<td>PAR-11-152</td>
<td>Yes</td>
<td>NCCAM; NCI</td>
</tr>
<tr>
<td>Prevention of HIV Transmission/Acquisition through a better understanding of Reproductive Health (R01)</td>
<td>RFA-HD-13-008</td>
<td>Yes</td>
<td>NIAID; NICHD</td>
</tr>
<tr>
<td>Prevention of HIV Transmission/Acquisition through a better understanding of Reproductive Health (R03)</td>
<td>RFA-HD-13-009</td>
<td>Yes</td>
<td>NIAID; NICHD</td>
</tr>
<tr>
<td>Mircobiome and Viral Metagenomics Lab for TEDDY Study ITN#12-17</td>
<td>N/A</td>
<td>Yes</td>
<td>NIDDK</td>
</tr>
<tr>
<td>Genomes to Natural Products (U01)</td>
<td>N/A</td>
<td>Yes</td>
<td>NCCAM; NIGMS</td>
</tr>
<tr>
<td>The Influence of the Microbiome on Preterm Labor and Delivery (R01)</td>
<td>RFA-NR-13-009</td>
<td>Yes</td>
<td>NINR</td>
</tr>
<tr>
<td>Environmental Influences on the Microbiome (R01)</td>
<td>RFA-ES-12-009</td>
<td>Yes</td>
<td>NIEHS</td>
</tr>
<tr>
<td>Gut-Microbiome-Brain Interactions and Mental Health (R21/R33)</td>
<td>N/A</td>
<td>Yes</td>
<td>NIMH</td>
</tr>
<tr>
<td>Evaluation of Multi-'omic Data in Understanding the Human Microbiome (U54)</td>
<td>N/A</td>
<td>Yes</td>
<td>NIDDK; ODS</td>
</tr>
<tr>
<td>Mechanistic Research on CAM Natural Products (R01)</td>
<td>N/A</td>
<td>Yes</td>
<td>NCCAM; NCI; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Diet Composition and Energy Balance (R01)</td>
<td>N/A</td>
<td>Yes</td>
<td>NCCAM; NCI; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Genomic Centers for Infectious Diseases (U19)</td>
<td>N/A</td>
<td>Yes</td>
<td>NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Preliminary Clinical Studies in Preparation for Large Interventional Trials of Complementary and Alternative Medicine (CAM) Therapies (R34)</td>
<td>N/A</td>
<td>Yes</td>
<td>NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Role of the Microflora in the Etiology of Gastro-Intestinal Cancer (R01)</td>
<td>RFA-NR-13-002</td>
<td>Yes</td>
<td>NINR</td>
</tr>
<tr>
<td>Biomarkers of Infection-Associated Cancers (R01)</td>
<td>N/A</td>
<td>Yes</td>
<td>NIAID; NIDA; NIDCR</td>
</tr>
<tr>
<td>Enhancing Tumorcidal Activity of Natural Killer Cells for Cancer Prevention (R01)</td>
<td>N/A</td>
<td>Yes</td>
<td>NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Unconventional Roles of Ethanol Metabolizing Enzymes, Metabolites, and Cofactors in Health and Disease</td>
<td>N/A</td>
<td>Yes</td>
<td>NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Multidisciplinary Studies of HIV/AIDS and Aging (R21)</td>
<td>PAR-12-171</td>
<td>No</td>
<td>NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Multidisciplinary Studies of HIV/AIDS and Aging (R01)</td>
<td>PAR-12-175</td>
<td>No</td>
<td>NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Multidisciplinary Studies of HIV/AIDS and Aging (R03)</td>
<td>PAR-12-176</td>
<td>No</td>
<td>NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NIDCR; NIDDK; NIEHS; NIGMS; NIMH; NINDS; NINR</td>
</tr>
<tr>
<td>Exploratory/Developmental Clinical Research Grants in Obesity (R21)</td>
<td>PA-12-179</td>
<td>No</td>
<td>NCCAM; NCI; NHLBI; NIDCR; NIDDK; ODS</td>
</tr>
<tr>
<td>The Role of Microbial Metabolites in Cancer Prevention and Etiology (U01)</td>
<td>PA-13-159</td>
<td>No</td>
<td>NCI</td>
</tr>
</tbody>
</table>
# Changes in the Gut Microbiota Associated with Health and Disease

<table>
<thead>
<tr>
<th>Disease/Health component</th>
<th>Microbiotic association</th>
<th>PubMedID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism</td>
<td>Modulating gut bacteria leads to behavioral improvement.</td>
<td>22114588</td>
</tr>
<tr>
<td>Allergies</td>
<td>Early colonization with Lactobacillus associated w/decreased allergies.</td>
<td>21512004</td>
</tr>
<tr>
<td>Anorexia</td>
<td>Bacteroidetes, Firmicutes, and Lactobacillus similar to lean patients though M. smithii significantly increased.</td>
<td>19774074</td>
</tr>
<tr>
<td>Bacterial vaginosis</td>
<td>Correlation of community profiles and Nugent criteria.</td>
<td>20534435</td>
</tr>
<tr>
<td>Celiac's disease</td>
<td>Higher diversity (Shannon-Wiener index) in Celiac's disease patients versus controls.</td>
<td>21565393</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>Gut microbiota-dependent metabolism of phosphatidylcholine .</td>
<td>21475195</td>
</tr>
<tr>
<td>Colorectal carcinoma</td>
<td>Fusobacterium nucleatum in colon cancer tissue.</td>
<td>22009990</td>
</tr>
<tr>
<td>Gastric Cancer</td>
<td>Carcinogenic pathway for developing gastric adenocarcinomas.</td>
<td>21937990</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>Neisseria elongata and Streptococcus mitis levels increase in saliva</td>
<td>21994333</td>
</tr>
<tr>
<td>IBD - Crohn's Disease</td>
<td>Less diversity in patients with Crohn's disease compared to healthy patients.</td>
<td>18401439</td>
</tr>
<tr>
<td>IBD (General)</td>
<td>IBD associated with overall community dysbiosis rather than single causal bacterial species</td>
<td>23013615</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>Vitamin D metabolism</td>
<td>20333556</td>
</tr>
<tr>
<td>Obesity</td>
<td>Significant changes in gut microbiota are associated with increasing obesity.</td>
<td>20368178</td>
</tr>
<tr>
<td>Reflux esophagitis</td>
<td>Prevalence of gram-negative anaerobes.</td>
<td>11902583</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Altered epithelial and mucosal permeability, loss of immune tolerance</td>
<td>23378145</td>
</tr>
<tr>
<td>Sexual transmission</td>
<td>Vaginal bacterial communities and sexually transmitted infections.</td>
<td>22133886</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>Betaproteobacteria enriched in diabeticc persons.</td>
<td>20140211</td>
</tr>
<tr>
<td>Type 1 Diabetes</td>
<td>Lactate-producing and butyrate-producing species v β-cell autoimmunity</td>
<td>23274889</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>Increased ratio of Firmicutes to Actinobacteria.</td>
<td>18648509</td>
</tr>
</tbody>
</table>
Host Immune System/Microbiota Interactions
Microbiome and Disease Associations
Functional Interactions between Host & Microbiome
Diet and the Microbiome
Translational Research and the Microbiome
Probiotics, MB Vaccines, Fecal Transplants
Summarizing, we are...

- Diverse in research areas and expertise
- Supported across several NIH institutes
- Hypothesis driven
- Many diseases, many systems
- Multi-disciplinary
- “Big” data generators and diverse data generators.
Features:

• common sample provenance
• common data provenance
• common protocol design
• centralized infrastructure
• centralized analysis infrastructure
The Human Microbiome Community

¿ Can we operate as both?

• Coordinated consortium
• Independent hypothesis-driven researchers

• Benefits:
  – Increased certainty about the data
  – Increased usability of the data
  – Streamlining the IRB process
  – Assist w/computational needs
  – Training
human microbiome stool

Results: 1 to 20 of 746

1. Human Feces
   1. ILLUMINA (Illumina HiSeq 2000) run: 58.8M spots, 5.9G bases, 4Gb downloads
      Accession: SRX323015

2. Human Feces
   2. ILLUMINA (Illumina HiSeq 2000) run: 44.3M spots, 4.5G bases, 3Gb downloads
      Accession: SRX323014

3. Human Feces
   3. ILLUMINA (Illumina HiSeq 2000) run: 52.1M spots, 5.3G bases, 3.5Gb downloads
      Accession: SRX323013

4. Human Feces
   4. ILLUMINA (Illumina HiSeq 2000) run: 45.6M spots, 4.6G bases, 3.1Gb downloads
      Accession: SRX323011

5. Human Feces
   5. ILLUMINA (Illumina HiSeq 2000) run: 31.8M spots, 3.2G bases, 2.2Gb downloads
      Accession: SRX323009

6. Human Feces
   6. ILLUMINA (Illumina HiSeq 2000) run: 34.6M spots, 3.5G bases, 2.4Gb downloads
      Accession: SRX323008

7. Human Feces
   7. ILLUMINA (Illumina HiSeq 2000) run: 27.2M spots, 2.7G bases, 1.9Gb downloads
      Accession: SRX322985

Top Organisms
- Human metagenome (681)
- Human gut metagenome (32)
- Unidentified (32)
- Human lung metagenome (1)

Search in related databases
- BioSample
  - Access: 50, 1.337, 1.387
- BioProject
  - Access: 10, 10
- dbGaP
  - Access: 9, 9
- GEO Datasets

Find related data
- Database: Select
- Find items

Search details
- "human metagenome"
- [Organism] OR human microbiome[All Fields] AND stool[All Fields]
Data uncertainty

• Origin, library prep, nucleic acid prep
• Biological sample still available?
• Subject/volunteer still available?
• Publication, downstream citations
• Quality in comparison to all others
• Patient phenotype
• Associated with disease?
Data uncertainty
....with increased coordination

• Improved submission standards, with provenance
• Create a investigator registry, track biosamples
• ....and volunteer availability
• ....and their publications
• Large scale QC operations
• Improve dbGaP submissions
• ....including things like disease phenotype
Models for sequence submission

• Centralized submission broker?
• Improved submission tools?
  – Sequence/metadata standards/provenance
• Enroll help from journals, requiring
  – Common submission formats
  – Use of metadata standards
  – Descriptions of sample/volunteer availability
  – Protocols
  – All of above as supplemental data hosted @ journal?
Models for an investigator registry

- Track PIs based on literature and grant funding
- Track volunteers using a registry
- Improve coordination at IRB/approval level
Clinical research subject recruitment: the Volunteer for Vanderbilt Research Program
www.volunteer.mc.vanderbilt.edu.

Harris PA, Lane L, Biaggioni I.
General Clinical Research Center, Vanderbilt University, Nashville, TN 37212, USA. paul.harris@vanderbilt.edu

Abstract
This article provides information concerning a novel research subject recruitment registry developed at Vanderbilt University. Project goals were (1) to provide a mechanism for lay individuals to self-enter information conveying interest in volunteering for clinical research and (2) provide tools for researchers to select and contact potential volunteers based on study-specific inclusion criteria. The registry was built and offered as an institutional resource to all university scientists conducting institutional review board-approved research. The authors present (1) a model for redesigning workflow associated with subject registration, volunteer retrieval, and subject contact; (2) details of a Web-based software application used as a focal point in designing workflow for our system; (3) descriptive statistics for volunteer and researcher use of the system during the first 32 months of operation; (4) cost estimates for the project; and (5) a set of recommendations for other medical centers wishing to adopt similar methodology.

ResearchMatch: a national registry to recruit volunteers for clinical research.

Harris PA, Scott KW, Lebo L, Hassan N, Lightner C, Pulley J.
Office of Research Informatics, Vanderbilt University, Nashville, Tennessee 37203, USA. paul.harris@vanderbilt.edu

Abstract
The authors designed ResearchMatch, a disease-neutral, Web-based recruitment registry to help match individuals who wish to participate in clinical research studies with researchers actively searching for volunteers throughout the United States. In this article, they describe ResearchMatch’s stakeholders, workflow model, technical infrastructure, and, for the registry’s first 19 months of operation, utilization metrics. Having launched volunteer registration tools in November 2009 and researcher registration tools in March 2010, ResearchMatch had, as of June 2011, registered 15,871 volunteer participants from all 50 states. The registry was created as a collaborative project for institutions in the Clinical and Translational Science Awards (CTSA) consortium. Also as of June 2011, a total of 751 researchers from 61 participating CTSA institutions had registered to use the tool to recruit participants into 540 active studies and trials. ResearchMatch has proven successful in connecting volunteers with researchers, and the authors are currently evaluating regulatory and workflow options to open access to researchers at non-CTSA institutions.
Shared IRB review for multi-site studies

- Common institutions submit review documents
- Divide/share review process
- Promoting consistency and compliance
- Easing IRB approval through cooperation
Metadata matters
<table>
<thead>
<tr>
<th>Metadata assessment across all demonstration projects</th>
<th>Thanks: Steve Sherry et al at dbGaP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IHMC Variable</strong></td>
<td><strong>SUBJID</strong></td>
</tr>
<tr>
<td><strong>SUBJID</strong></td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>0.88</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td>0.81</td>
</tr>
<tr>
<td><strong>Other Race</strong></td>
<td>0.56</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td>0.38</td>
</tr>
<tr>
<td><strong>Smoking_duration</strong></td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Drugs</strong></td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Weight_kg</strong></td>
<td>0.25</td>
</tr>
<tr>
<td><strong>BP</strong></td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Weight_lbs</strong></td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Height</strong></td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Disease</strong></td>
<td>0.19</td>
</tr>
<tr>
<td><strong>Institution</strong></td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Dose</strong></td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Start_date</strong></td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Finish_date</strong></td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Other Country</strong></td>
<td></td>
</tr>
</tbody>
</table>
Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MlxS) specifications.


Microbial Genomics and Bioinformatics Group, Max Planck Institute for Marine Microbiology, Bremen, Germany.

Abstract

Here we present a standard developed by the Genomic Standards Consortium (GSC) for reporting marker gene sequences—the minimum information about a marker gene sequence (MIMARKS). We also introduce a system for describing the environment from which a biological sample originates. The 'environmental packages' apply to any genome sequence of known origin and can be used in combination with MIMARKS and other GSC checklists. Finally, to establish a unified standard for describing sequence data and to provide a single point of entry for the scientific community to access and learn about GSC checklists, we present the minimum information about any (x) sequence (MlxS). Adoption of MlxS will enhance our ability to analyze natural genetic diversity documented by massive DNA sequencing efforts from myriad ecosystems in our ever-changing biosphere.
PhenX measures to identify opportunities for cross-study analysis.

- **SC defines scope of the domain**
- **WG identifies broad list of measures**
- **WG selects preliminary measures**
- **WG seeks input from research community**
- **WG reviews data from outreach**
- **WG selects final 15 measures**
Welcome to the PhenX Toolkit

The Toolkit provides standard measures related to complex diseases, phenotypic traits and environmental exposures. Use of PhenX measures facilitates combining data from a variety of studies, and makes it easy for investigators to expand a study design beyond the primary research focus. All Toolkit content is available to the public at no cost.

Information about the project is available at www.phenx.org

More »

Please Read Toolkit Guidance

How to cite use of PhenX measures:
Measures incorporated in this study were selected from the PhenX Toolkit version April 29 2013, Ver 5.4. More »

How to cite the PhenX Toolkit:

Funding for PhenX and the PhenX Toolkit was provided by NHGRI 5U01HG004597 and 3U01HG004597-03S3.
Scenario

• You’re a member of the HM-PI registry
• You submit
  – IRB forms with specific description of patient phenotypes in your study
  – Biosamples or volunteers available to consortium
• ID other volunteers relevant to your study
• We monitor publications
• Assist journals with metadata standards
• Link all above data to SRA data
One possible protocol harmonization cross-validation approach.

• For a particular body site
• Gather protocols (~5)
• Test in the hands of a single technician
• Or randomize with multiple techs
• Review results, ID similar systems
• Distribute to larger testing network
2012 performed an email poll for analysis needs

- Most popular response: metagenomic assembly.

In May held a two-hour webinar on assembly topics with:

- C. Titus Brown (Michigan State University)
- Michael Schatz (Cold Spring Harbor Lab).

**Presenters:**

**Dr. C. Titus Brown**

*Assistant Professor*
Depts. of Computer Science and Engineering,
Microbiology and Molecular Genetics
Michigan State University

**Michael C. Schatz, Ph.D.**

*Assistant Professor*
Quantitative Biology
Watson School of Biological Sciences
Cold Spring Harbor

*Laboratory Adjunct Assistant Professor*
Computer Science
Stony Brook University
Gaps

- Training
- PI registry
- Metagenomic QC system
- Centralized processing/value addition to SRA
- Harmonized protocols
- Standards, submission tools
  - SRA / dbGaP
  - Journals