Human Microbiome Science
Vision for the Future

JULY 24 - 26, 2013

Owen White
Institute for Genome Sciences
University of Maryland

A. Initiatives and Organization of the HMP
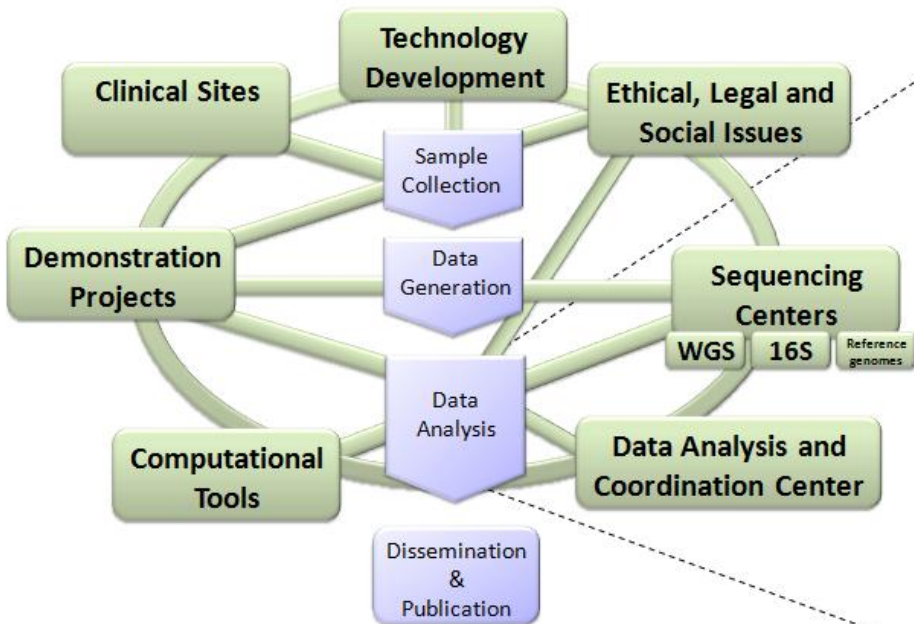
B. 300 Subjects Enrolled — Sampling 15/18 body sites, DNA extraction, High throughput sequencing
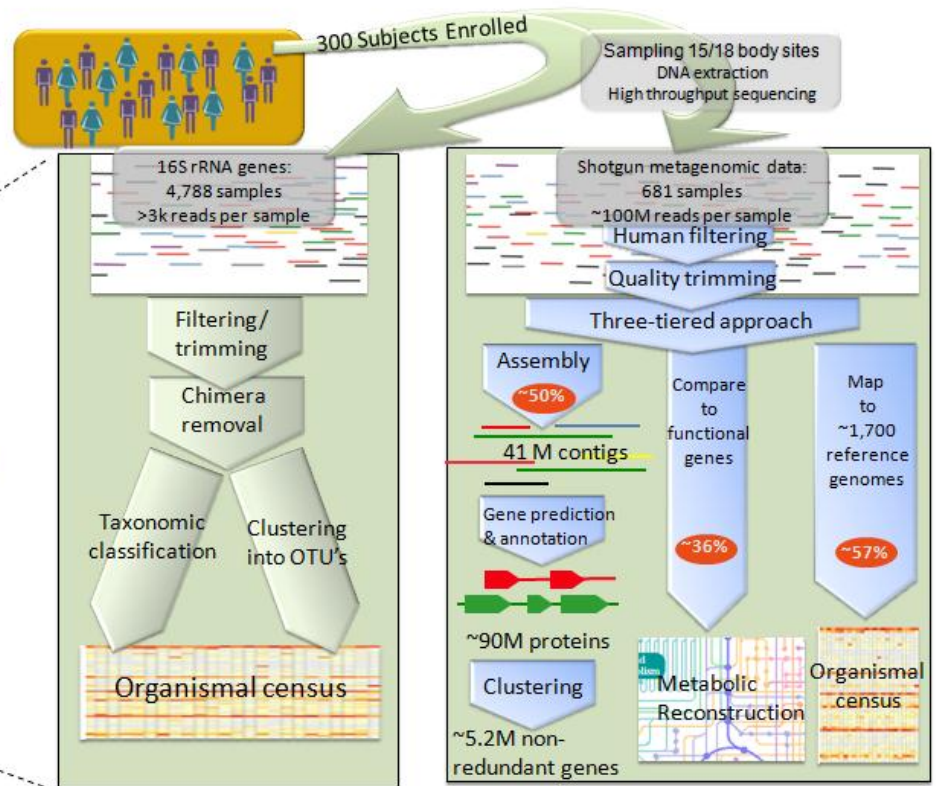
Gevers D, Knight R, Petrosino JF, Huang K, et al. (2012) The Human Microbiome Project: A Community Resource for the Healthy Human Microbiome. PLoS Biol 10(8): e1001377. doi:10.1371/journal.pbio.1001377
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1001377

PLOS | BIOLOGY

TENTH ANNIVERSARY

Figure 1. Timeline of microbial community studies using high-throughput sequencing.
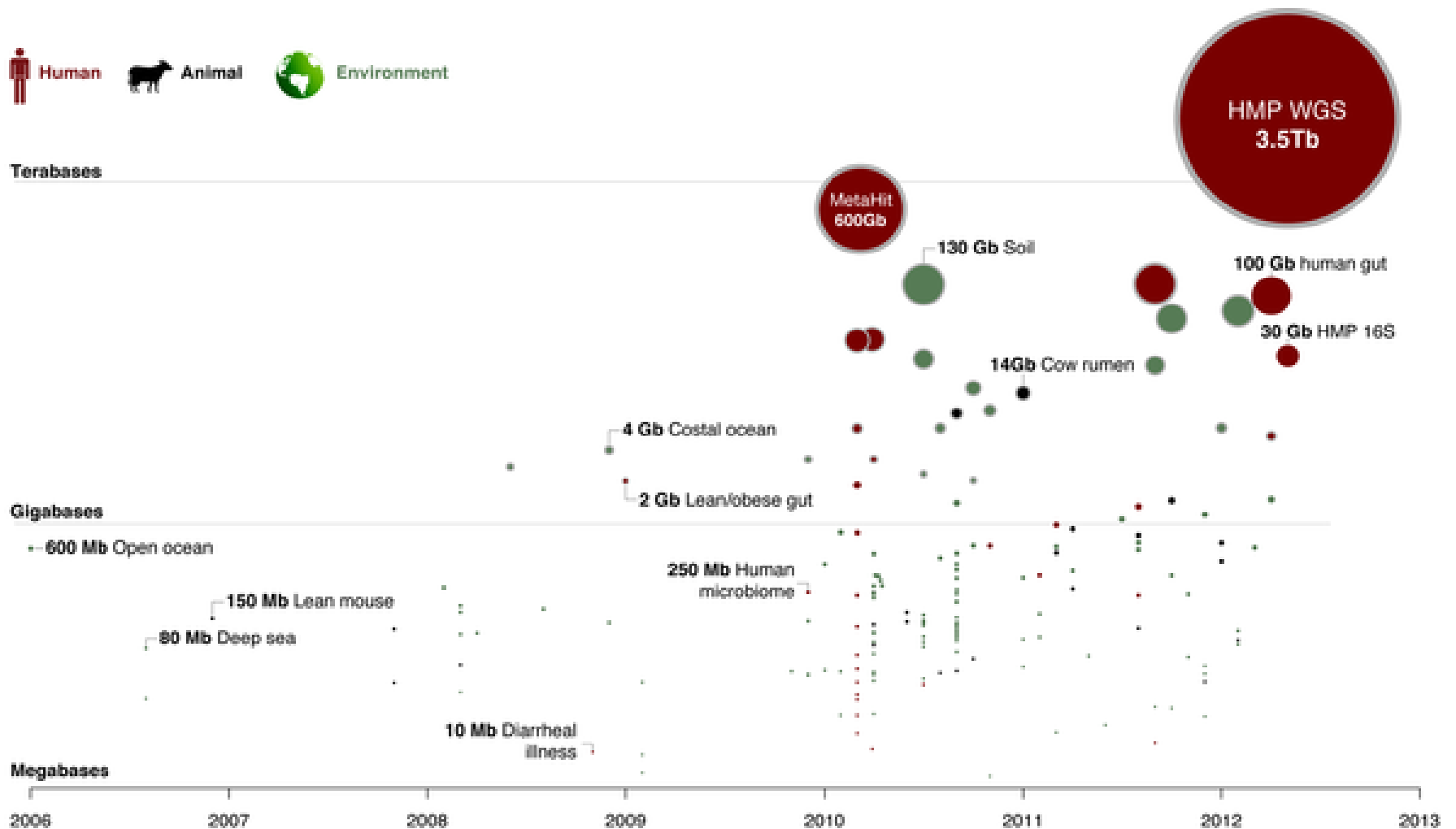
Gevers D, Knight R, Petrosino JF, Huang K, et al. (2012) The Human Microbiome Project: A Community Resource for the Healthy Human Microbiome. PLoS Biol 10(8): e1001377. doi:10.1371/journal.pbio.1001377
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1001377

nature
The **mouse** genome

Science
24 April 2009 | $15
Livestock Decoded
AAAS

nature
The rat genome
Insights into mammalian evolution

The Human Genome Project

National Human Genome Research Institute, NIH

ENCODE

HMP

International HapMap Project

eMERGE Network
electronic medical records & genomics
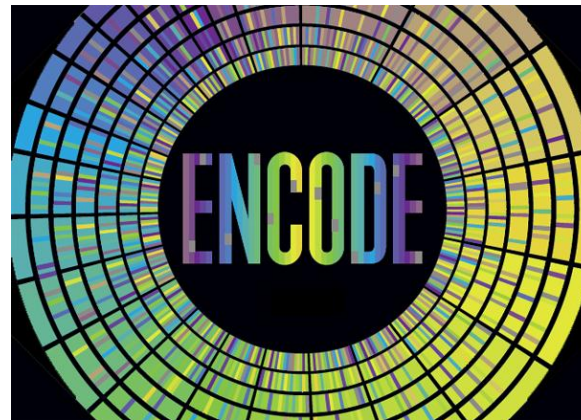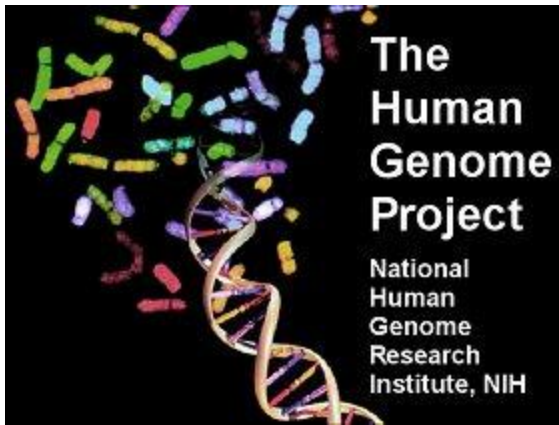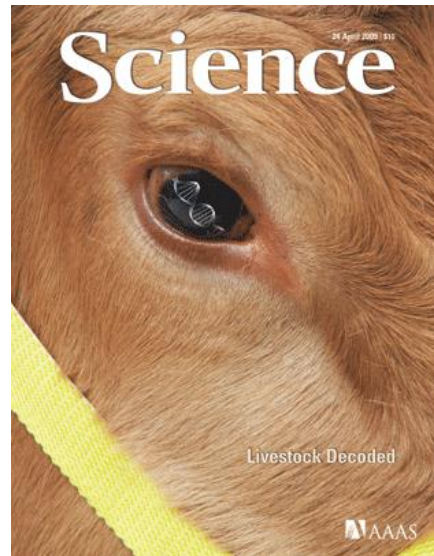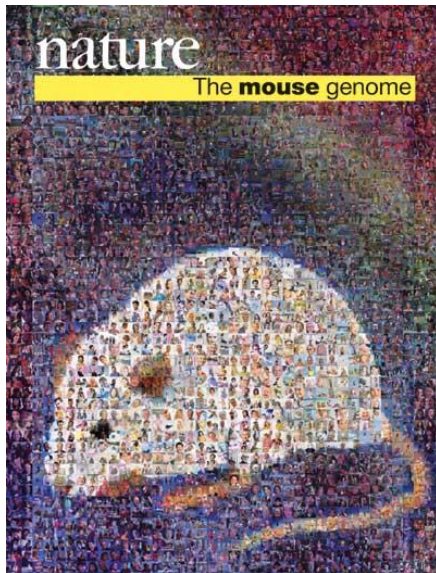
Current and Previous RFAs

Title	RFA	Expired	NIH Institute/Center
The Role of Microbial Metabolites in Cancer Prevention and Etiology (U01)	PAR-11-152	Yes	NCCAM; NCI
Prevention of HIV Transmission/Acquisition through a better understanding of Reproductive Health (R01)	RFA-HD-13-008	Yes	NIAID; NICHD
Prevention of HIV Transmission/Acquisition through a better understanding of Reproductive Health (R03)	RFA-HD-13-009	Yes	NIAID; NICHD
Mircobiome and Viral Metagenomics Lab for TEDDY St	ITN#12-17		
Genomes to Natural Products (U01)			
The Influence of the Microbiome on Preterm Lal			
Environmental Influences on the Microbiome (R			
Gut-Microbiome-Brain Interactions and Mental			
Dynamics of Host-Associated Microbial Commu			
Evaluation of Multi-'omic Data in Understanding Disease (U54)			
Mechanistic Research on CAM Natural Products			
Diet Composition and Energy Balance (R01)			DK; ODS
Genomic Centers for Infectious Diseases (U19)			
Preliminary Clinical Studies in Preparation for La and Alternative Medicine (CAM) Therapies (R34			
Role of the Microflora in the Etiology of Gastro-I			
Biomarkers of Infection-Associated Cancers (R01			
Biomarkers of Infection-Associated Cancers (R21			
Enhancing Tumoricidal Activity of Natural Killer Prevention (R01)			
Unconventional Roles of Ethanol Metabolizing E and Disease			
Unconventional Roles of Ethanol Metabolizing Enzyn and Disease			
Multidisciplinary Studies of HIV/AIDS and Aging (R21)	PAR-12-174		NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NICHD; NIDA; NIMH; NINDS
Multidisciplinary Studies of HIV/AIDS and Aging (R01)	PAR-12-175		NCCAM; NCI; NHLBI; NIA; NIAAA; NIAID; NIAMS; NICHD; NIDA; NIDDK; NIMH; NINDS; NINR
Multidisciplinary Studies of HIV/AIDS and Aging (R03)	PAR-12-176		NCI; NIA; NIAAA; NIAID; NICHD; NIDA; NIMH
Exploratory/Developmental Clinical Research Grants in Obesity (R21)	PA-12-179		NCCAM; NCI; NHLBI; NICHD; NIDDK; ODS
The Role of Microbial Metabolites in Cancer Prevention and Etiology (U01)	PA-13-159		NCI

Common Fund
NCCAM, NCI, NHLBI, NIA, NIAAA, NIAID, NIAMS, NICHD, NIDA, NIDCR, NIDDK, NIEHS, NIGMS, NIMH, NINDS, NINR

## Changes in the Gut Microbiota Associated with Health and Disease

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

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.
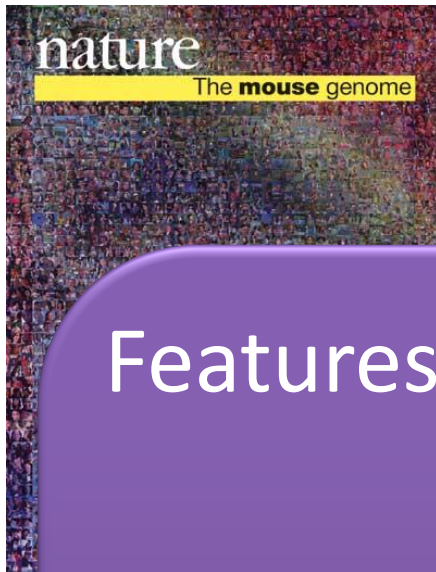
Features:

- common sample provenance
- common data provenance
- common protocol design
- centralized infrastructure
- centralized analysis infrastructure
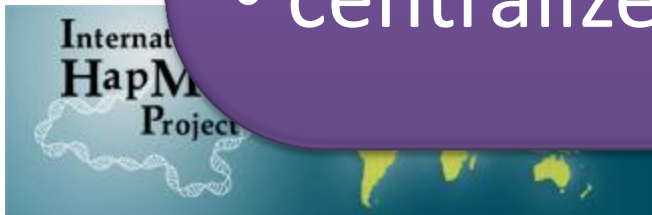
electronic medical records & genomics

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

NCBI   Resources ⊙ How To ⊙   Sign in to NCBI

SRA | SRA   ▾ | human microbiome stool   ⊗ | Search
                              Save search   Advanced   Help

Show additional filters   Display Settings: ⊙ Summary, 20 per page   Send to: ⊙   Filters: Manage Filters

Access   Results: 1 to 20 of 746   ▼ Top Organisms [Tree]
Controlled (641)   << First < Prev Page 1 of 38 Next > Last >>   human metagenome (681)
Public (746)   human gut metagenome (32)
                             ☐ Human Feces   unidentified (32)
Source   1.   human lung metagenome (1)
DNA (675)   1 ILLUMINA (Illumina HiSeq 2000) run: 58.8M spots, 5.9G bases, 4Gb downloads
RNA (26)   Accession: SRX323015
metagenomic (55)
                             ☐ Human Feces   Search in related databases   ▲
Type   2.
genome (424)   1 ILLUMINA (Illumina HiSeq 2000) run: 44.3M spots, 4.5G bases, 3Gb downloads
                                Accession: SRX323014

                             ☐ Human Feces
Clear all   3.
                                1 ILLUMINA (Illumina HiSeq 2000) run: 52.1M spots, 5.3G bases, 3.5Gb downloads
Show additional filters   Accession: SRX323013

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

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

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

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

Database	Access		all
	public	controlled	
BioSample	50	1,337	1,387
BioProject	10		10
dbGaP		9	9
GEO Datasets			

Find related data   ▲
Database: | Select   ▾ |
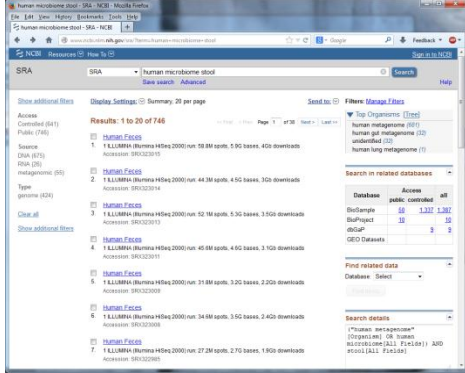   Find items

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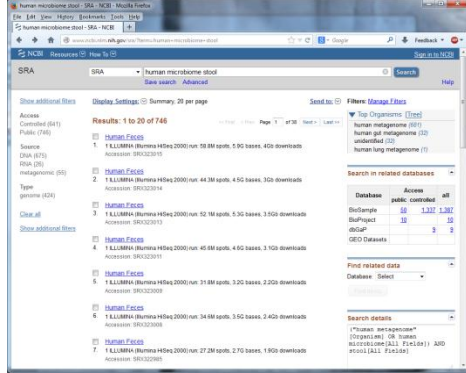
Data uncertainty

- 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

Research Portfolio Online Reporting Tools (RePORT)

Search [   ] 🔍

HOME | ABOUT RePORT | FAQs | GLOSSARY | CONTACT US

My RePORTER   Login | Register   System Health: YELLOW

NIH RePORTER
Version: 5.7.0

About RePORTER DATA   FAQ   ExPORTER   RePORTER Manual   RSS of Newly Added Projects 📶 ❓

QUERY ᴺᴱᵂ   BROWSE NIH ᴮᴱᵀᴬ

SUBMIT QUERY   CLEAR QUERY

Fiscal Year (FY): ❓   [Active Projects]   SELECT
Current FY is 2013

RESEARCHER AND ORGANIZATION

Principal Investigator (PI) / ❓   [   ] , [   ]
Project Leader:
(Last Name, First Name)   Use '%' for wildcard
Enter several PI/Project Leader names

Organization: ❓   [   ]   LOOKUP
Please enter at least 3 characters to use Lookup.
◉ Contains ○ Begins with ○ Exact

Department: ❓   [   ]   SELECT

Organization Type: ❓   [   ]   SELECT

City: ❓   [   ]
Use '%' for wildcard

State: ❓   [   ]   SELECT

Country: ❓   [   ]   SELECT

Congressional District: ❓   [   ]   SELECT

DUNS Number: ❓   [   ]

TEXT SEARCH

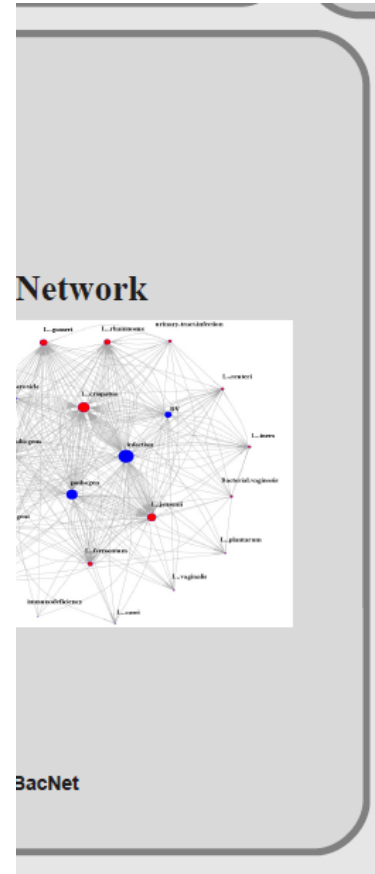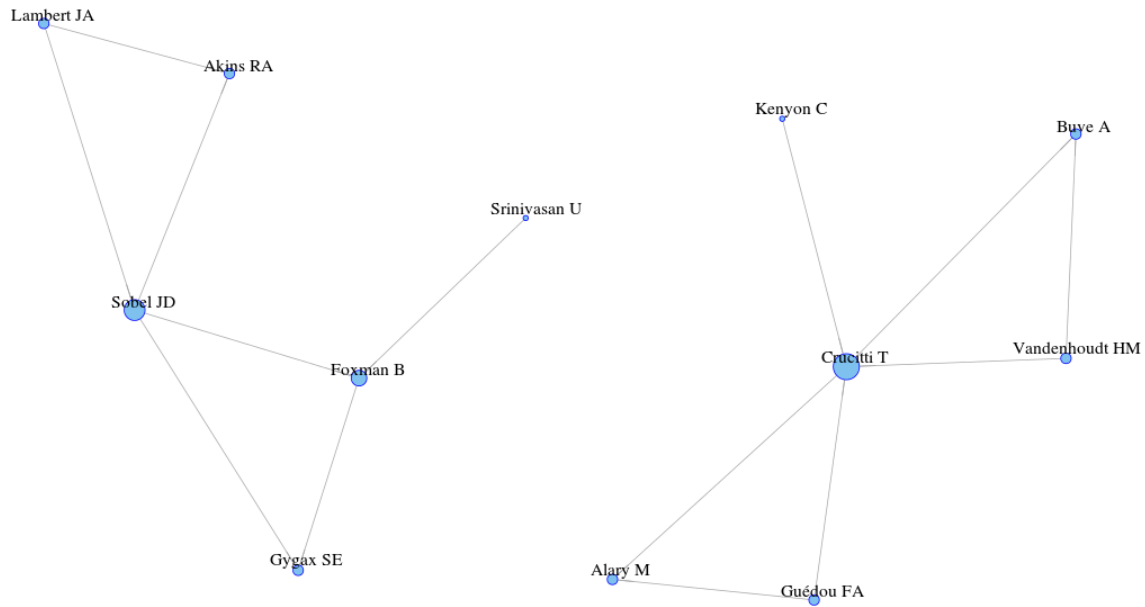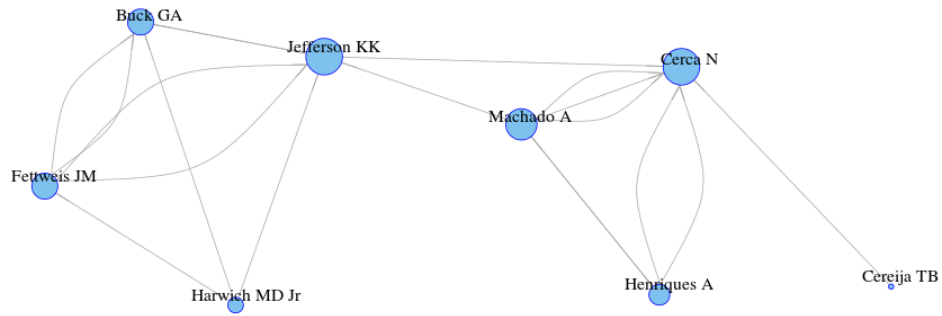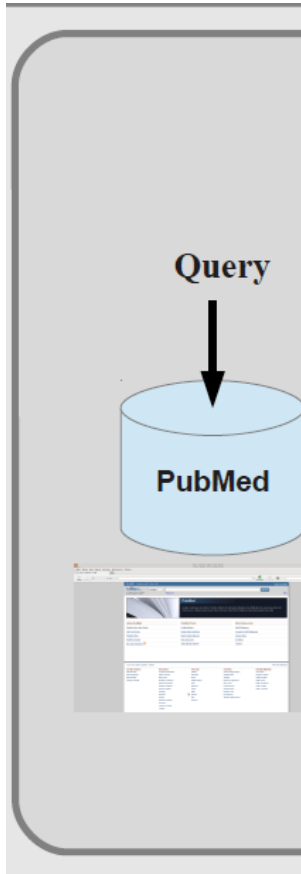Text Search (Logic): ❓   [   ]

◉ And
○ Or

Search in ❓
◉ Projects
○ Publications
○ Projects & Publications

Limit Project search to
☐ Project Title
☐ Project Terms
☐ Project Abstracts

Limit Publication search to
Start Year [2012 ▾]
End Year   [2013 ▾]

Query

PubMed

BacNet

nikov,
Idaho

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.

IRBshare

= Full Board Review Sites

= Shared Review Sites

What is IRBs
The IRBshare Sys
shared IRB review
multi-site studies
Participating Insti
Shared Review D
Shared Review P
supported by a c
secure web porta
IRBshare Master
("IMA"). IRBshare
human research
communication,
nationwide.

Shared IRB review for <u>multi-site studies</u>

- Common institutions submit review documents
- Divide/share review process
- Promoting consistency and compliance
- Easing IRB approval through cooperation

Metadata matters

Metadata assessment across all demonstration projects : Thanks: Steve Sherry et al at dbGaP

IHMC Variable	Total Fraction Mappable (Identical+Mappable)	Fraction Identical	Fraction Mappable	Fraction Not mappable	Fraction Not present	Foregut Microbiome in Development of Esophageal Adenocarcinoma	Urethral microbiome in adolescent males	The Thrifty Microbiome: The Role of the Gut Microbiota in Obesity in the Amish	Metagenomic Analysis of the Structure and Function of the Human Gut Microbiota in Crohn's Disease	Effect of Crohn's Disease Risk Alleles on Enteric Microbiota	Metagenomic study of the human skin microbiome associated with acne	The Microbial Ecology of Bacterial Vaginosis: A Fine Scale Resolution Metagenomic Analysis	NIH Intramural Sign Microbiome... Atopic Dermatitis
SUBJID	1.00	0.88	0.13			SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID
Gender	0.94	0.94			0.06	Gender		SEX	Sex	SEX	sex		
Age	0.88	0.81	0.06	0.06	0.06	Age_at_first_visit	AgeAtEnrollment	AGE	DOB	AGE	age	age	Age_status
Race	0.81	0.44	0.38		0.19	Race	Race_Other_Text			RACE	ethnicity	ethnicity	Race
Other Rrace	0.56	0.31	0.25		0.44	Other Race	Race_Other				ethnicity	ethnicity	
Smoking	0.38	0.31	0.06		0.63	Smoking_status				SMOKING_STATUS		smoke_ever	
Lab	0.31	0.19	0.13	0.06	0.63	Diagnosis	TID			CLOSTRIDIA_DIFFICILE		nugent_score	
Smoking_duration	0.31	0.19	0.13		0.69	Smoking_status						smoke_ever	
Drugs	0.31	0.19	0.13		0.69	Antacids, Steroids, Antibiotics			Antibiotics		Currently Treated?	treatment1, treatment2	
Weight_kg	0.25	0.25		0.06	0.69								Weight
BP	0.19	0.19			0.81			SBP, DBP					
Weight_lbs	0.19	0.19			0.81								Weight
Height	0.19	0.19			0.81								
Disease	0.19	0.06	0.13		0.81				Status		face, nose		
Institution	0.13	0.00	0.13		0.88							highest_grade	
Dose	0.13	0.06	0.06		0.88								
Duration	0.13	0.06	0.06		0.88								
Start_date	0.13	0.13			0.88		TID		Disease duration				
Finish_date	0.13	0.13			0.88		TID		Disease duration				
Location	0.13	0.13		0.06	0.81	Other Country				DURATION			

Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications.

Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PS, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, San Gil I, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E, Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spo A, Stombaugh J, Tiedje JM, Ward DV, Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO.

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 (MIxS). Adoption of MIxS 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.
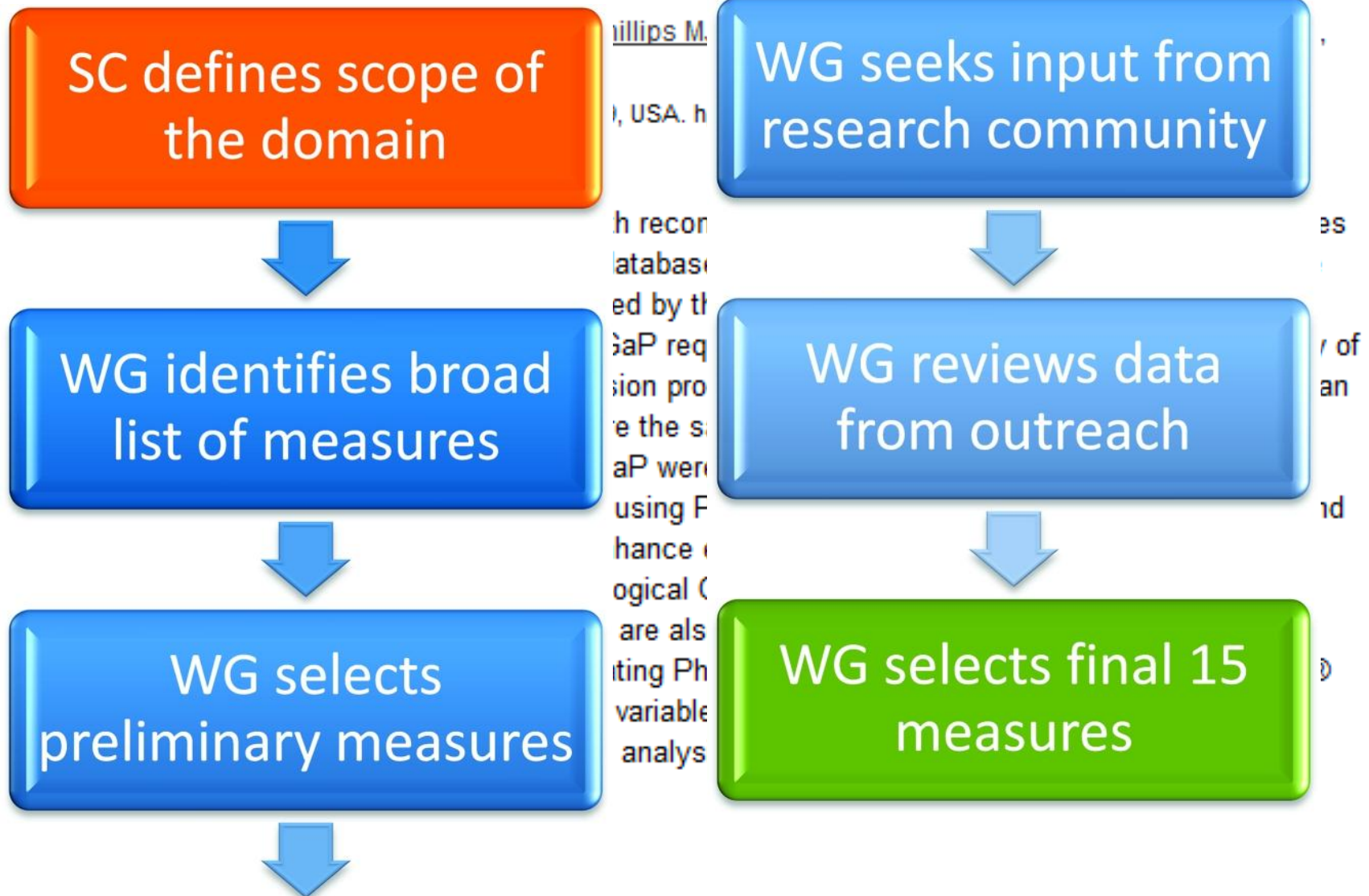
PhenX Toolkit

New NIH Common Data Element (CDE) Portal, more here

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:

Hamilton, et al. (2011) The PhenX Toolkit: Get the Most From Your Measures. *American Journal of Epidemiology*, **174**(3), 253-60.

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

QuickStart

Tutorial

Substance Abuse and Addiction

Browse

Search

My Toolkit

Registration

Registration

You may browse the PhenX Toolkit, but to save your work, you need to Register. See additional benefits for registered users.

Existing users may login:

User ID:

Password:

Log In

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
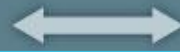
One possible protocol harmonization cross-validation approach.

- 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

201[...] [...]needs
• M[...]

•In ma[...]embly topics[...]

• C[...]ersity)
• M[...]Lab).
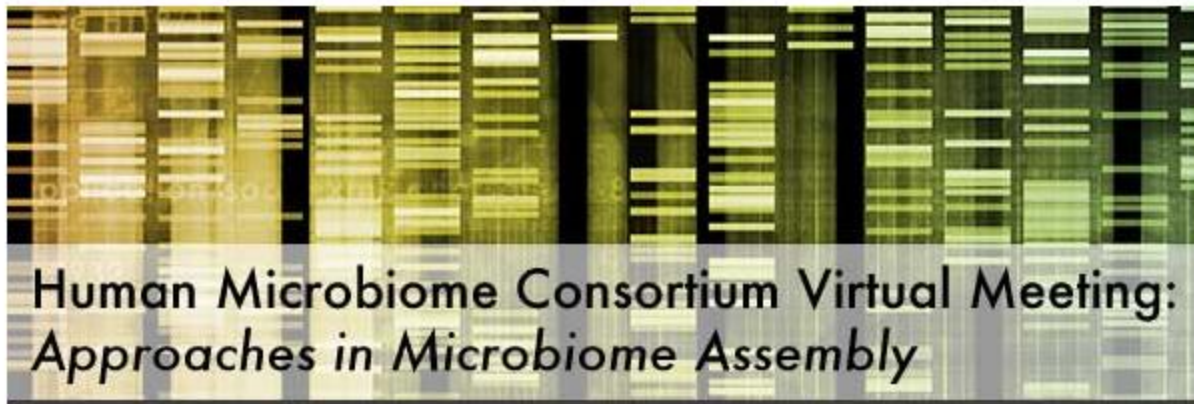
Human Microbiome Consortium Virtual Meeting:
Approaches in Microbiome Assembly

Presenters:

Dr. C. Titus Brown

Assistant Professor
Departments of Computer Science and Engineering
Microbiology and Molecular Genetics
Michigan State University

Michael C. Schatz, Ph.D.

Assistant Professor
Quantitive 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