GABRIELLA MILLER KIDS FIRST PEDIATRIC RESEARCH PROGRAM

Lorette C. Javois, Ph.D.
Working Group Coordinator

April 26, 2016

National Institutes of Health
Turning Discovery Into Health

Form **1040** Department of the Treasury—Internal Revenue Service (99)

U.S. Individual Income Tax Return **2014** OMB No. 1545-0074 IRS Use Only—Do not write or staple in this space.

For the year Jan. 1–Dec. 31, 2014, or other tax year beginning , 2014, ending , 20 See separate instructions.

Your first name and initial	Last name	Your social security number
If a joint return, spouse's first name and initial	Last name	Spouse's social security number

Home address (number and street). If you have a P.O. box, see instructions. Apt. no.

▲ Make sure the SSN(s) above and on line 6c are correct.

City, town or post office, state, and ZIP code. If you have a foreign address, also complete spaces below (see instructions).

Foreign country name	Foreign province/state/county	Foreign postal code

Presidential Election Campaign

Check here if you, or your spouse if filing jointly, want $3 to go to this fund. Checking a box below will not change your tax or refund. ☐ You ☐ Spouse

Filing Status

Check only one box.

1 ☐ Single

2 ☐ Married filing jointly (even if only one had income)

3 ☐ Married filing separately. Enter spouse's SSN above and full name here. ▶

4 ☐ Head of household (with qualifying person). (See instructions.) If the qualifying person is a child but not your dependent, enter this child's name here. ▶

5 ☐ Qualifying widow(er) with dependent child

Exemptions

6a ☐ **Yourself.** If someone can claim you as a dependent, **do not** check box 6a

b ☐ **Spouse**

Boxes checked on 6a and 6b

c **Dependents:**

(1) First name Last name	(2) Dependent's social security number	(3) Dependent's relationship to you	(4) ✓ if child under age 17 qualifying for child tax credit (see instructions)
			☐
			☐
			☐
			☐

If more than four dependents, see instructions and check here ▶ ☐

No. of children on 6c who:
- **lived with you**
- **did not live with you due to divorce or separation (see instructions)**

Dependents on 6c not entered above

d Total number of exemptions claimed

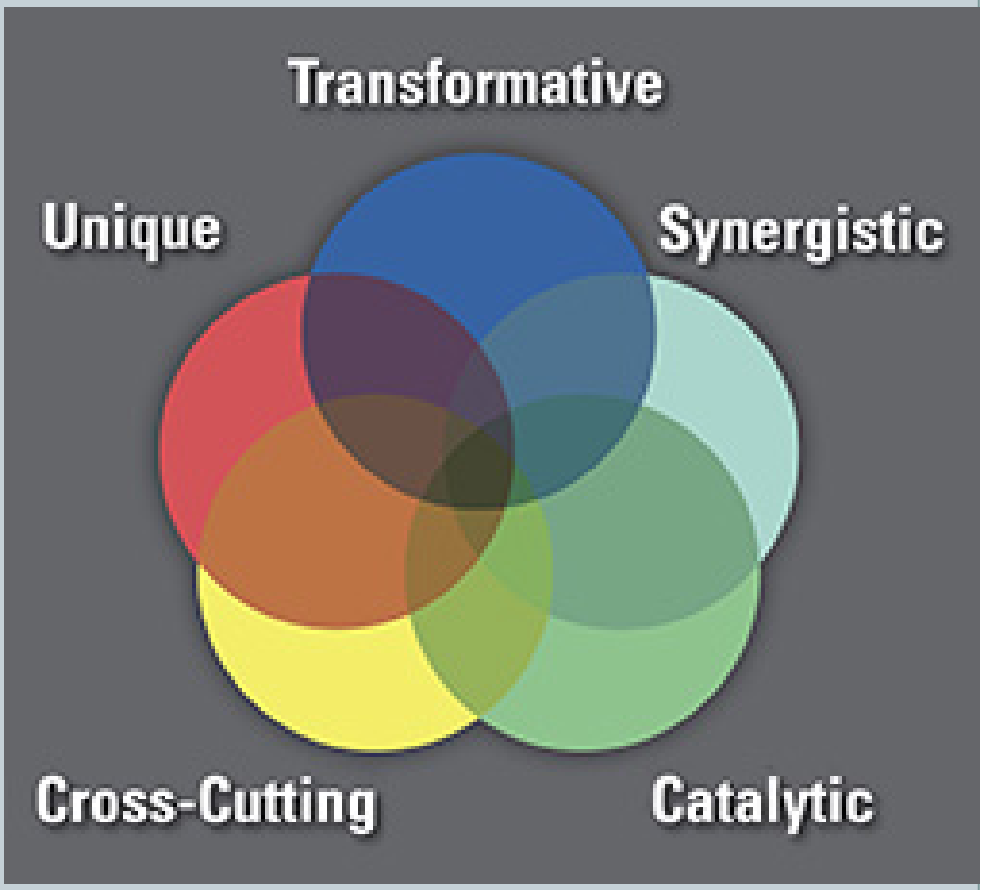
Add numbers on lines above ▶

Gabriella Miller Kids First Research Act

- Signed into law on April 3, 2014
- Ended taxpayer contribution to presidential nominating conventions
- Transferred remaining $126 million into the Pediatric Research Initiative Fund
- Authorized appropriation of $12.6 million per year for 10 years to the NIH Common Fund for pediatric research; first appropriation was for FY2015

NIH Common Fund

- Run by the Office of Strategic Coordination, Office of the NIH Director
- Trans-NIH programs involving multiple NIH Institutes



Gabriella Miller Kids First Planning

- January 2015 – A Trans-NIH Working Group consisting of Dr. Collins, other IC Directors, and NIH staff considered the challenges and opportunities for transformative pediatric research
- Ideas coalesced around the need for a pediatric data resource consisting of well-curated clinical and genetic sequence data
- Focus on pediatric cancers and structural birth defects
- Leadership by NICHD, NHLBI, NHGRI and NCI

Kids First Common Fund Working Group

- Trans-NIH Structural Birth Defects Working Group: NICHD, NHLBI, NHGRI, NIAAA, NIAMS, NIDCR, NIDDK, NIDA, NIEHS, NINDS, OD/ORIP, CDC
- NCI's Pediatric Oncology Preclinical & Clinical Programs, NEI, NIAID
- Bioinformatics Specialists

- Extramural research community stakeholders
- Advocacy Communities

Overall Goal

To develop a data resource for the pediatric research community of well-curated phenotype and sequence data that will help determine the biological basis of childhood cancers and structural birth defects

- Cohort identification and enrichment
- Data Resource development integrating genomic and other data with community portal
- Pilot projects using the data resource to mine, aggregate, link, and analyze data

Limited funds mandate focused effort

Progress to Date

- PAR-15-259 Discovery of the Genetic Basis of Structural Birth Defects and of Childhood Cancers: Gabriella Miller Kids First Pediatric Research Program (X01) → 7 cohorts in the pipeline
- PAR-16-150 X01 applications due June 17, 2016
- RFA-RM-16-001 → Kids First Sequencing Center applications to be reviewed June 2016
- Data Resource Center RFA → coming soon

GETTING STRUCTURAL: BIRTH DEFECTS AND THE KIDS FIRST DATA RESOURCE

Bruce D. Gelb, M.D.
Mindich Child Health and Development Institute
Departments of Pediatrics and Genetics & Genomic Sciences

Icahn School of Medicine at Mount Sinai

DISCLOSURES

- None for this presentation

How Might Investigators Use the Kids First Data Resource?

Bench to Bassinet

Funded by NHLBI

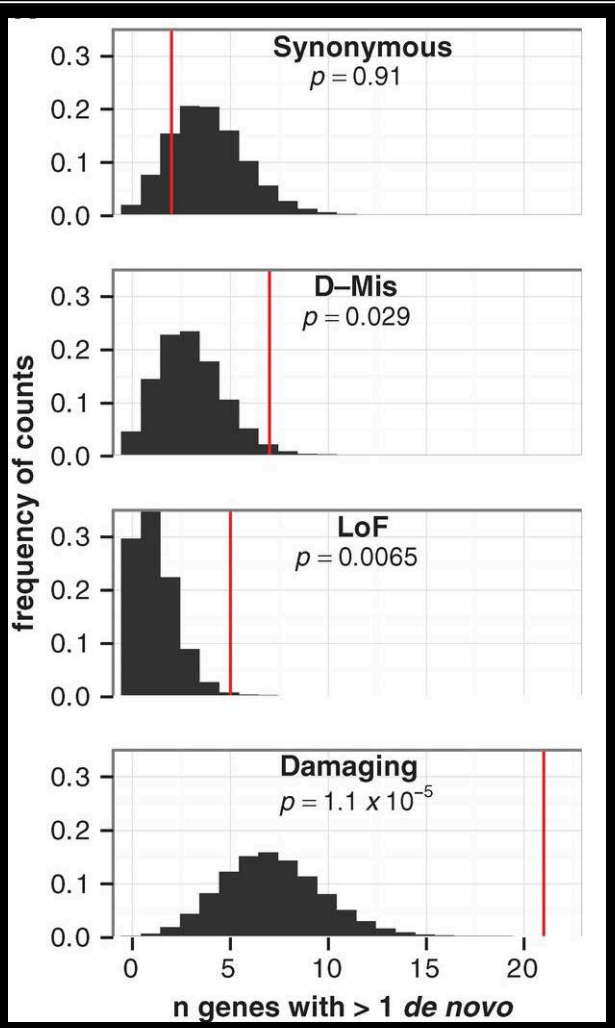
Pediatric Cardiac Genomics Consortium

Boston Children's/Brigham and Women's -CHOP-Columbia-Mount Sinai-Yale

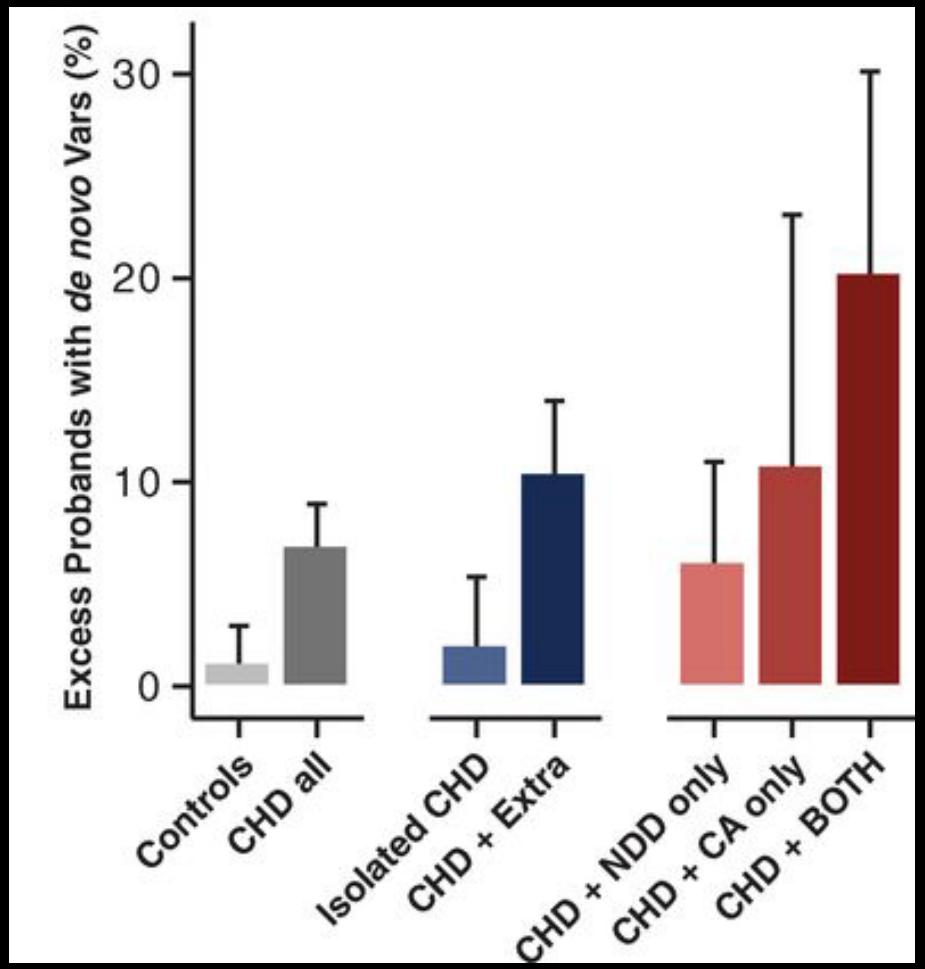
National Heart Lung and Blood Institute

People Science Health

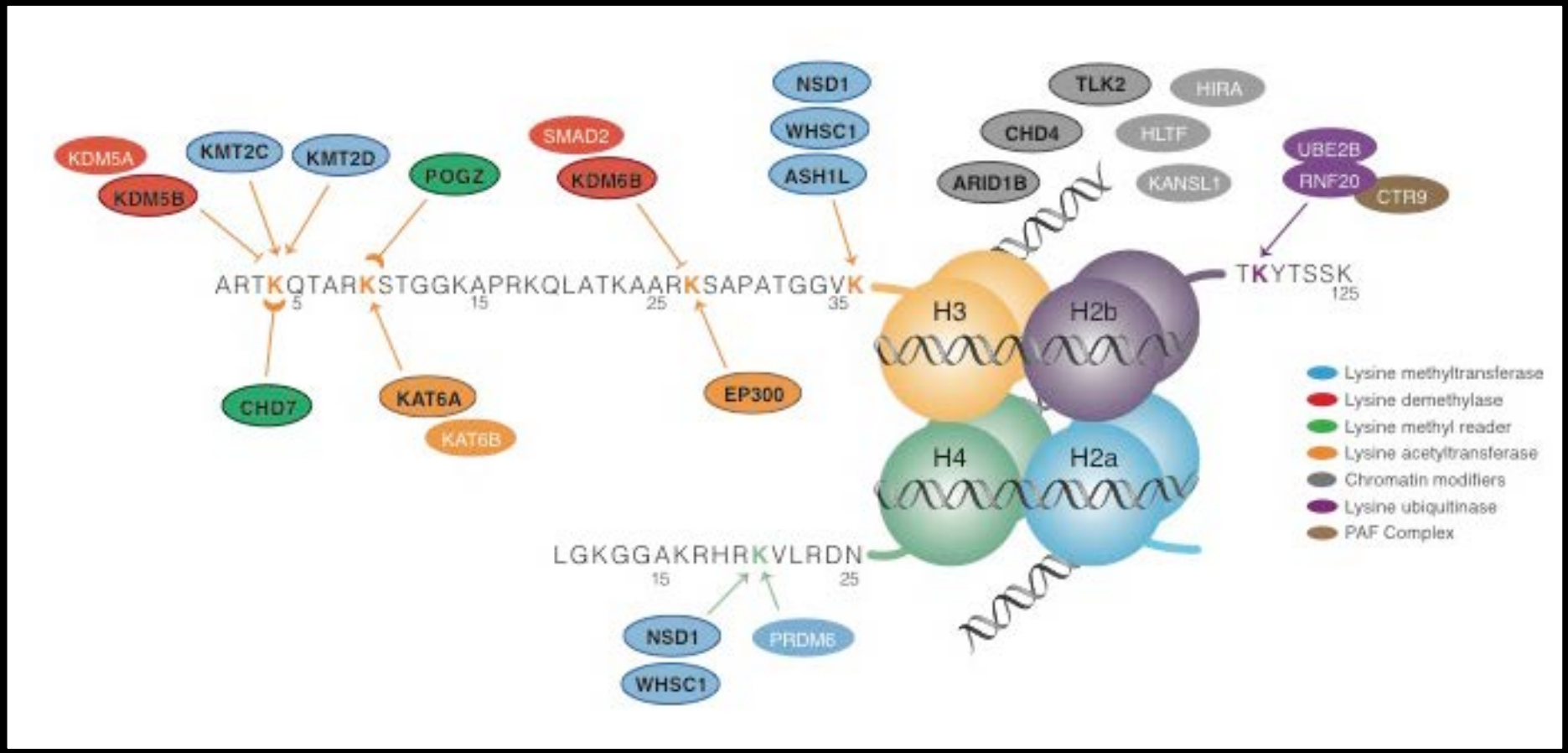
WES STUDY #2, *Science* 2015
Gene with Multiple Mutations



WES STUDY #2, *Science* 2015
CHD +/- Extracardiac



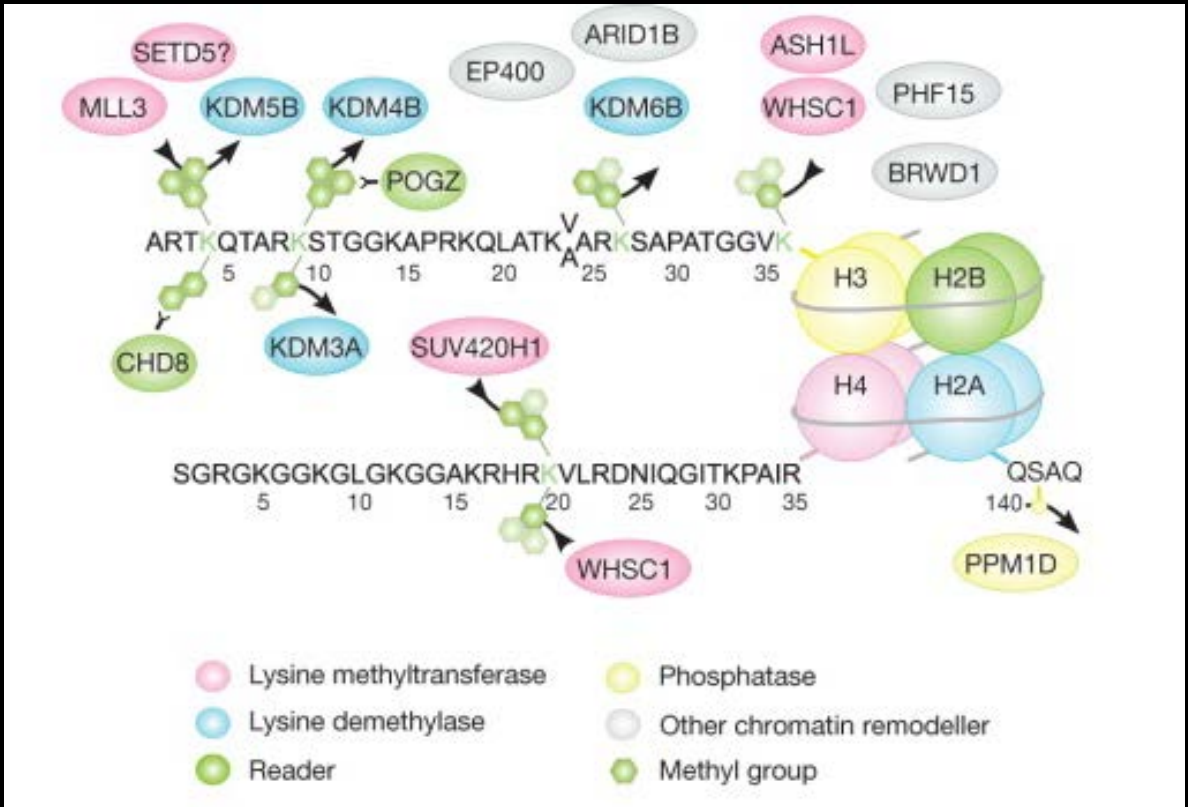
WES STUDY #2, *Science* 2015
Findings



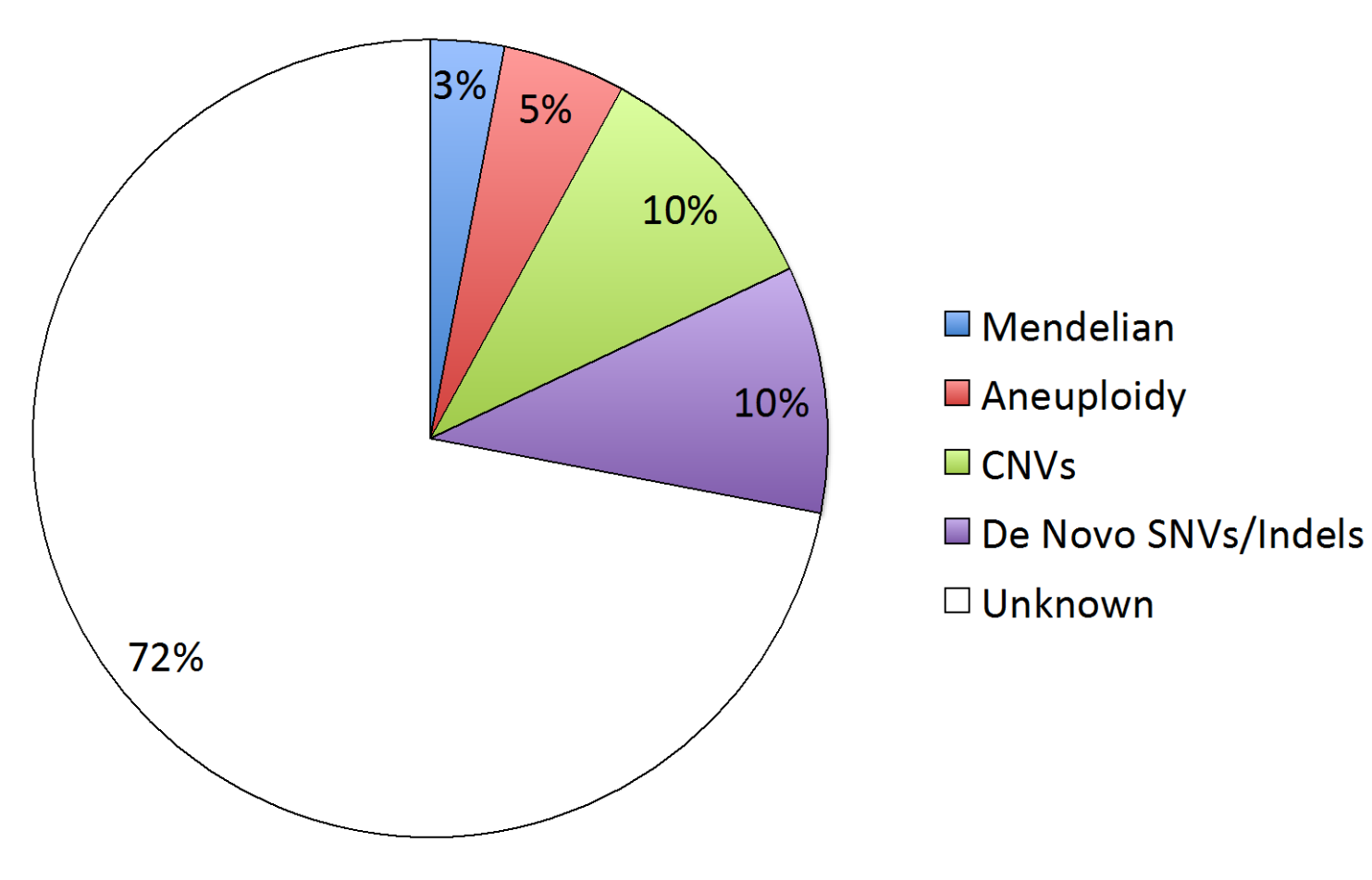
nature International weekly journal of science

Synaptic, transcriptional and chromatin genes disrupted in autism

De Rubeis *et al.*, November 2014



CHD GENETICS
Architecture Understood

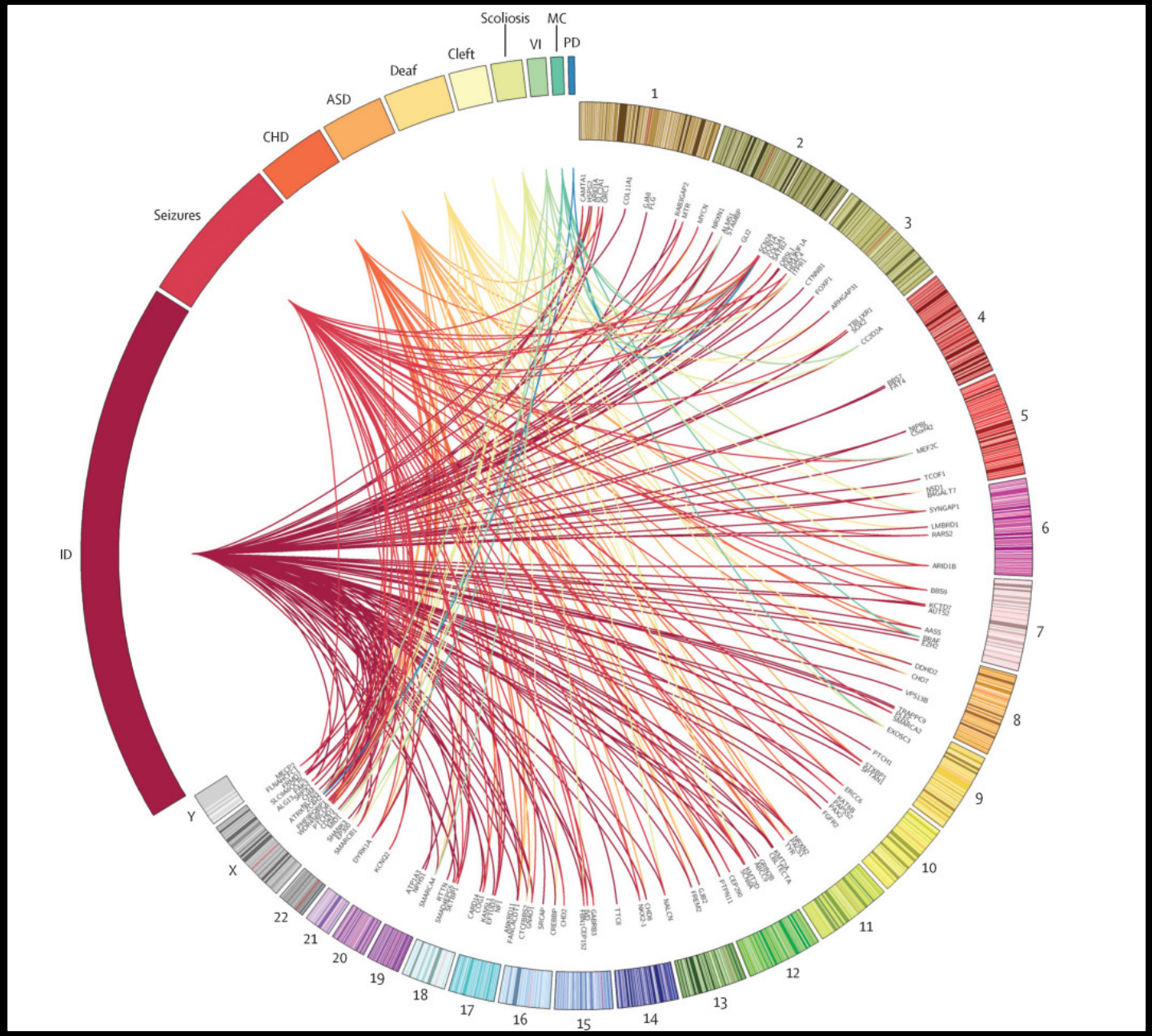


CHD GENETICS
What's Missing?

- Exomic Mutations
  - Saturation
  - Other Modes (Recessive, Somatic)
- Non-Exomic
  - Isolated CHD = Cardiac Enhancer/Promoter Mutations?
- Epigenetics

DECIPHERING DEVELOPMENTAL DISORDERS STUDY

- United Kingdom
- Recruitment
  - April 2011 – April 2015
  - 14,000 Patients with DDs
- Genomics
  - Arrays
  - Exome Sequencing



Wright *et al.*
Lancet 2015

DDD STUDY
Meta-Analysis

Table 2 | Novel genes with compelling evidence for a role in developmental disorder

Evidence	Gene	*De novo* DDD (missense, LOF)	*De novo* meta (missense, LOF)	*P* value	Test	Mutation clustering	Predicted haploinsufficiency (%)
De novo enrichment	*COL4A3BP*	3 (3,0)	5 (5,0)	4.10×10^{-12}	Meta	Yes	14.7
	PPP2R5D	4 (4,0)	5 (5,0)	6.01×10^{-12}	DDD	Yes	19.7
	ADNP	4 (0,4)	5 (0,5)	4.59×10^{-11}	Meta	No	9.8
	POGZ	2 (0,2)	5 (0,5)	4.31×10^{-10}	Meta	No	30.0
	PPP2R1A	3 (3,0)	3 (3,0)	2.03×10^{-8}	DDD	Yes	23.5
	DDX3X	4 (3,1)	5 (3,2)	2.26×10^{-7}	DDD	No	12.7
	CHAMP1	2 (0,2)	3 (0,3)	4.58×10^{-7}	Meta	No	52.9
	BCL11A	3 (3,0)	4 (3,1)	1.03×10^{-6}	DDD	Yes	0.6
	PURA	3 (1,2)	3 (1,2)	1.14×10^{-6}	DDD	No	9.4
De novo enrichment plus additional evidence	*DNM1*	3 (3,0)	5 (5,0)	1.43×10^{-6}	Meta	No	13.5
	TRIO	2 (2,0)	7 (7,0)	5.16×10^{-6}	Meta	Yes	25.7
	PCGF2	2 (2,0)	2 (2,0)	1.08×10^{-5}	DDD	Yes	37.7

The Deciphering Developmental Disorders Study, *Nature* 2015

Kids First Pediatric Research Program
Structural Birth Defects

- Congenital Diaphragmatic Hernias
- Disorders/Differences of Sex Development
- Orofacial Cleft Birth Defects
- Structural Heart and Other Birth Defects
- Syndromic Cranial Dysinnervation Disorders

WHOLE GENOME SEQUENCING
What To Expect

- *De Novo* Mutations
  - **~ 50-75 SNVs and Indels/Proband**
  - **Improved Exome Coverage; ~1.5 SNVs/Proband**
- Structural Variation
  - Far Wider Range in Sizes and Types
- Enormous Numbers of Inherited Variation

KIDS FIRST DATA RESOURCE
Case Uses for Birth Defects WGS

- Genomics-Capable User
  - Primary Need: Access to VCF or BAM Files
  - Functionalities
    - Stable Large Pipe for Downloading
    - Phenotypic Data
- Genomics-Naïve User
  - Primary Need: Look-Up Functionality
  - Functionalities
    - Gene- and Variant-Centric Look-Up
    - Phenotype-Centric Look-Up

KIDS FIRST DATA RESOURCE
Interface

- Graphical
- Range of Variation
  - SNVs
  - Indels
  - Structural Variation
- Phenotype Data
  - Primary Need: Look-Up Functionality
- Functional Genomics
  - ChIP-Seq
  - Epigenome Roadmap

Clinical, Genetic, and Epidemiologic Approaches to Understanding Cancer Etiology

Payal P. Khincha, MD, MSHS
Clinical Fellow, Clinical Genetics Branch
Division of Cancer Epidemiology and Genetics

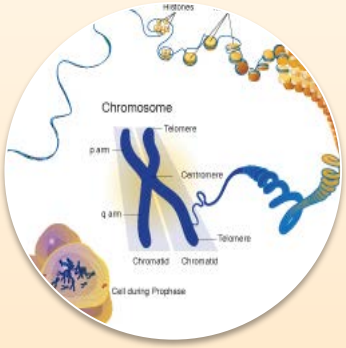
NIH NATIONAL CANCER INSTITUTE

April 26, 2016

A Multidisciplinary Approach

Clinical Studies

- Phenotype
- Family Studies
- Psychosocial Studies
- Prevention
- Treatment



Basic Science

- Genetics
- Molecular Biology
- Biochemistry

Epidemiology

- Populations
- Biomarkers
- Environment
- Genetics

Rare ⟷ **Common**

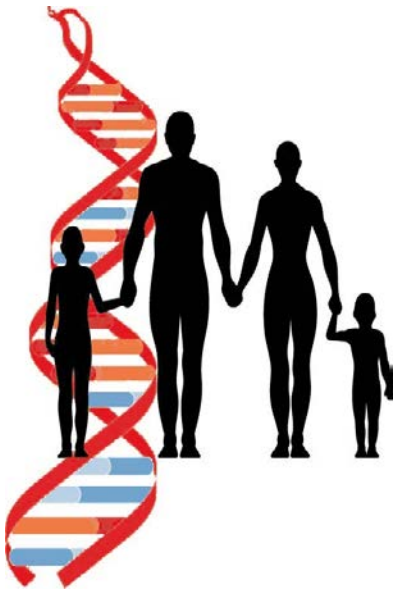
Early Family Studies			Current Research
Li-Fraumeni Syndrome	Families with multiple cancers in children, young adults	*TP53* tumor suppressor gene	International consortium; medical, genetic, & psychosocial studies
Wilms' Tumor, Aniridia, Congenital Malformations	Multicenter study of childhood cancer WAGR syndrome		New avenues of research in molecular and cellular biology
Familial Chronic Lymphocytic Leukemia	Family and population studies	Precursor monoclonal B-cell lymphocystosis	International CLL consortium, genomics
Myotonic Dystrophy (*Dystrophica Myotonica, DM*)	Leukemia developed in patient with DM	Is DM a cancer susceptibility disorder?	DM patients have 2-fold increased risk

Advancing understanding of cancer etiology

Clinical Genetics Branch

"Saving lives & improving the quality of life for individuals predisposed to cancer"

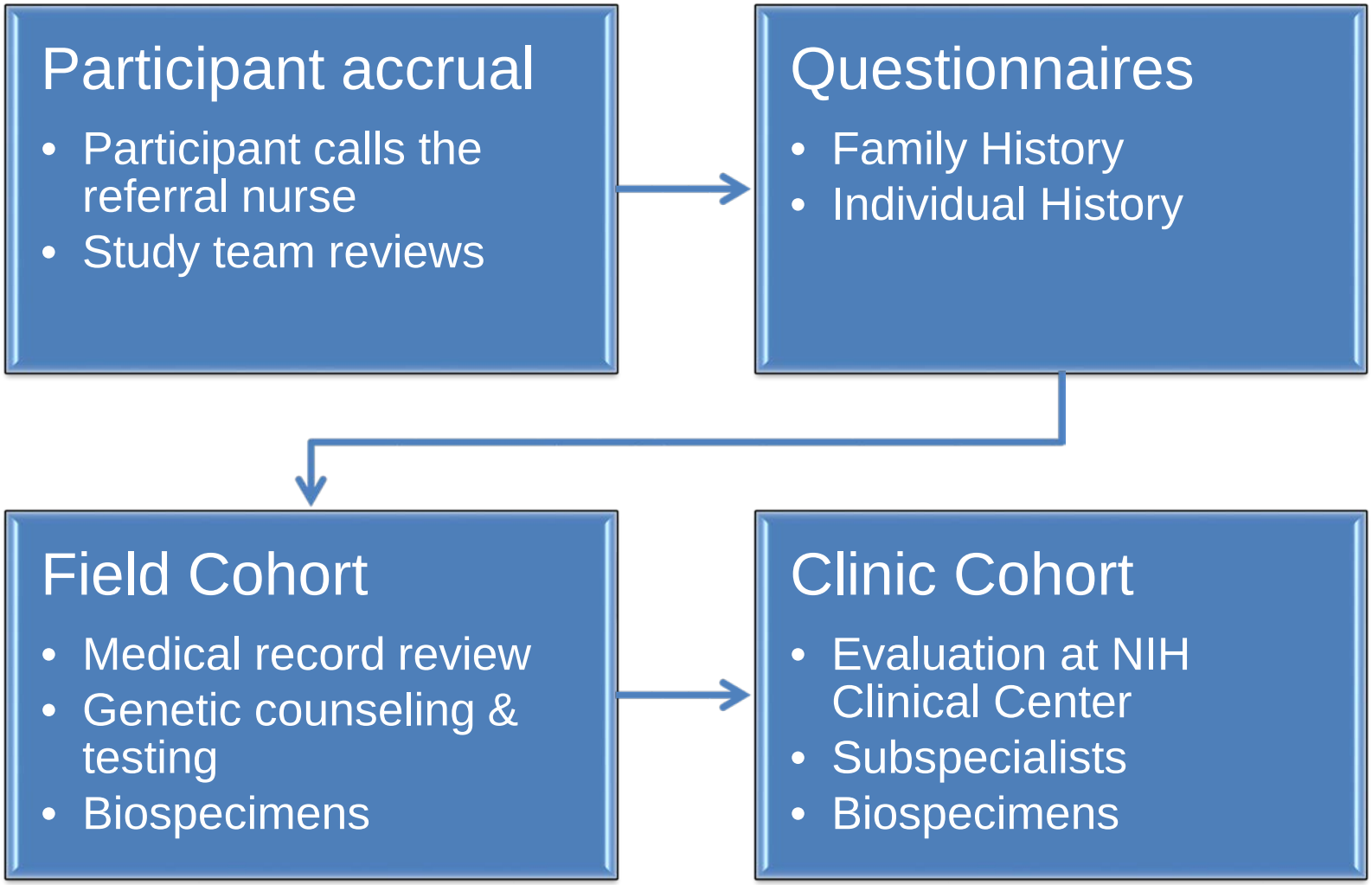
- **Family Studies**
  - Inherited Bone Marrow Failure Syndromes*
  - *DICER1* Pleuropulmonary Blastoma Cancer Predisposition Syndrome*
  - Li-Fraumeni Syndrome*
  - Familial Testicular Cancer
  - Neurofibromatosis Type 1
  - Breast/Ovarian Cancer
  - Behavioral, Counseling, Psychological, and Social Implications of Increased Cancer Risk
- **Clinical Epidemiology Unit**
  - HPV Epidemiology, Screening, and Prevention
  - Gynecologic Malignancies
  - Myotonic Dystrophy and Cancer Susceptibility
  - Telomere Molecular Epidemiology



*Currently recruiting participants

http://dceg.cancer.gov

Family Studies- common schema



National Cancer Institute at the National Institutes of Health | www.cancer.gov

Inherited Bone Marrow Failure Syndromes (IBMFS)

- Home
- About this Study
- Participation
- IBMFS Disorders
- What are Gene Mutations?
- Study Team
- Publications
- IBMFS Resources

Inherited bone marrow failure syndromes (IBMFS) are rare disorders; usually these patients have some form of aplastic anemia (failure of the bone marrow to produce blood), and may have a family history of the disorder. There are several well-described syndromes that can be recognized by healthcare experts either by physical characteristics in the patients or from laboratory findings. There are also patients who are harder to classify.

The NCI IBMFS Cohort Study brochure is available for patients and families.

Read our Brochure

Patients with these syndromes are of interest to the NCI because they have have a very high risk of developing cancer (either leukemia or certain solid tumors). At the moment, we cannot predict which specific patient with an IBMFS is going to develop cancer, and we want to study all patients with an IBMFS to learn more about those without and those who may develop cancer.

The NCI IBMFS Cohort Study enrolls families from North America that have at least one member with an IBMFS. The study includes individuals known to have an IBMFS as well as their first degree relatives (brothers, sisters, parents, and children) as well as other relatives where appropriate.

Our overall goal is to reach a better understanding of how cancers develop in persons with IBMFS, so that we may improve the health care that can be offered to persons with these disorders.

Join Us

How can I join?

Individuals with one of the inherited bone marrow failure syndromes and their family members are encouraged to participate.

Phone: **1-800-518-8474** to speak with the referral nurse

Email: **NCI.IBMFS@westat.com**

https://www.marrowfailure.cancer.gov/

NCI IBMFS Study

- **Disorders studied**
  - Fanconi Anemia
  - Dyskeratosis Congenita
  - Diamond-Blackfan Anemia
  - Shwachman-Diamond Syndrome
  - Congenital Neutropenia
  - Thrombocytopenia Absent Radii Syndrome
  - Others

- **Evaluations**
  - Questionnaires
  - Medical Record Review
  - IBMFS Team
  - Genetic Counseling
  - Subspecialists
  - Biospecimens

Cancer susceptibility
Dysmorphology

https://www.marrowfailure.cancer.gov/

NCI IBMFS Cohort, 2002-2015

	Screened Families	Consented Families*	Affected Individuals	Relatives	Total
FA	211+3	127+3	147+3	380+12	527+15
DBA	122	87	106	301+45	407+63
DC	120+34	103+29	152+38	329+77	481+115
SDS	55+17	31+17	27+18	89	116
AMEGA	5	1	1	1	2
TAR	10	7	8	21	29
OTHERS	144	53	57	120	177
TOTAL	**721**	**458**	**557**	**1375**	**1932**

*Consented less than screened: some not eligible, some still in progress.
"+" means resembling the syndrome but unable to prove, called "XX-like".

Pleuropulmonary Blastoma *DICER1* Syndrome Study

Home

What is PPB?

What is the *DICER1* gene?

What is the study about?

What is involved?

Who is on the study team?

What is a clinical study?

Pleuropulmonary blastoma (PPB) is a rare tumor of the lung. Research has shown that PPB may be part of an inherited cancer predisposition syndrome caused by changes in a gene known as *DICER1*.

The PPB *DICER1* Syndrome Study is an observational study of individuals with PPB and/or other *DICER1*-related tumors and their families. The study will help researchers define the tumor types and risks associated with the *DICER1* syndrome. We also would like to understand why some individuals with a change in *DICER1* develop cancer while others remain healthy. Read the study summary on ClinicalTrials.gov.

To carry out this study, the National Cancer Institute's (NCI) Clinical Genetics Branch (CGB) is partnering with leading research groups, including:

- International Pleuropulmonary Blastoma Registry (also see their Facebook page)
- International Ovarian and Testicular Stromal Tumor Registry
- Department of Pathology at Children's National Medical Center in Washington, D.C.
- St. Louis Children's Hospital, part of Washington University in St. Louis (read an article about their research).

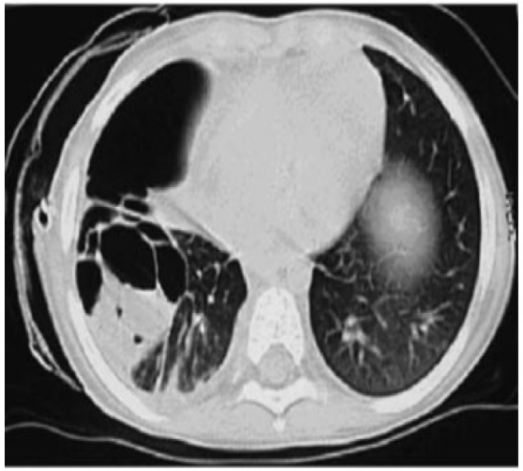
How can I join?

All individuals with PPB or related tumors and conditions and their family members are encouraged to participate.

Phone: **1-800-518-8474** to speak with the referral nurse
Email: **NCI.PPB@westat.com**

Pleuropulmonary Blastoma in *DICER1* Syndrome

Type I PPB
Cystic tumor
5-year OS: 91%
Median age: 8 mo



Type II PPB
Mixed cystic/solid tumor
Mixed-pattern primitive sarcoma
5-year OS: 71%
Median age: 35 mo

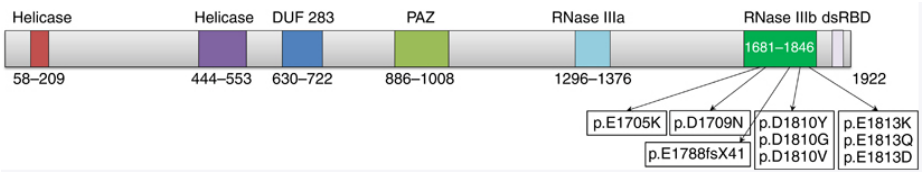
Type III PPB
Solid tumor
High-grade sarcoma
5-year OS: 53%
Median age: 41 mo

https://ppb.cancer.gov/

Data: Messinger, Stewart *et al., Cancer* 2015
Images: Priest *et al., Pediatric Pulmonology* 2009

DICER1 Syndrome

DICER1: an endoribonuclease critical to the generation of small noncoding regulatory RNAs contains a “hotspot” for mutations



Witkowski et al, *British Journal of Cancer* (2013) 109; 2744–2750

***DICER1* Mutations in Familial Pleuropulmonary Blastoma**

D. Ashley Hill,[1,2]* Jennifer Ivanovich,[1] John R. Priest,[2] Christina A. Gurnett,[1] Louis P. Dehner,[1] David Desruisseau,[1] Jason A. Jarzembowski,[3] Kathryn A. Wikenheiser-Brokamp,[4] Brian K. Suarez,[1] Alison J. Whelan,[1] Gretchen Williams,[2,5] Dawn Bracamontes,[1,2] Yoav Messinger,[2,5] Paul J. Goodfellow[1]

Science (2009) 325; 965

- Multi-nodular goiter
- Sex-cord stromal tumors
- Cystic nephroma + Wilms
- Embryonal rhabdomyosarcoma of cervix
- Nasal chondromesenchymal hamartoma
- Ocular medulloepithelioma

https://ppb.cancer.gov/

NCI *DICER1* Syndrome Cohort

Accrual	Clinical Center Cohort	Total Enrolled
Families	37	82
DICER1 mutation carriers	99	173
DICER1-negative controls	69	220
Not yet tested	0	67
Total individuals	**168**	**460**

NATIONAL CANCER INSTITUTE
Division of Cancer Epidemiology & Genetics

Li-Fraumeni Syndrome Study

- Home
- About this Study
- Our Goals
- Research Team
- Participation
- Publications
- Resources

Li-Fraumeni Syndrome (LFS) is a rare, inherited disorder which leads to a higher risk of developing certain cancers. These cancers tend to occur at younger ages in patients with LFS than in the general population. The types of tumors most frequently seen in LFS include bone and soft tissue cancers (called "sarcomas",), breast cancer, brain tumors, and cancer of the adrenal gland. The diagnosis of LFS is based on an individual's personal and family history of cancers. Heritable disease-causing changes in a gene called TP53 is currently the only known cause of LFS and is identified in about 70% of families with a clinical diagnosis of LFS.

Our **2014 Study Newsletter** is available. Learn more about our newly added research objectives.

Read our Newsletter

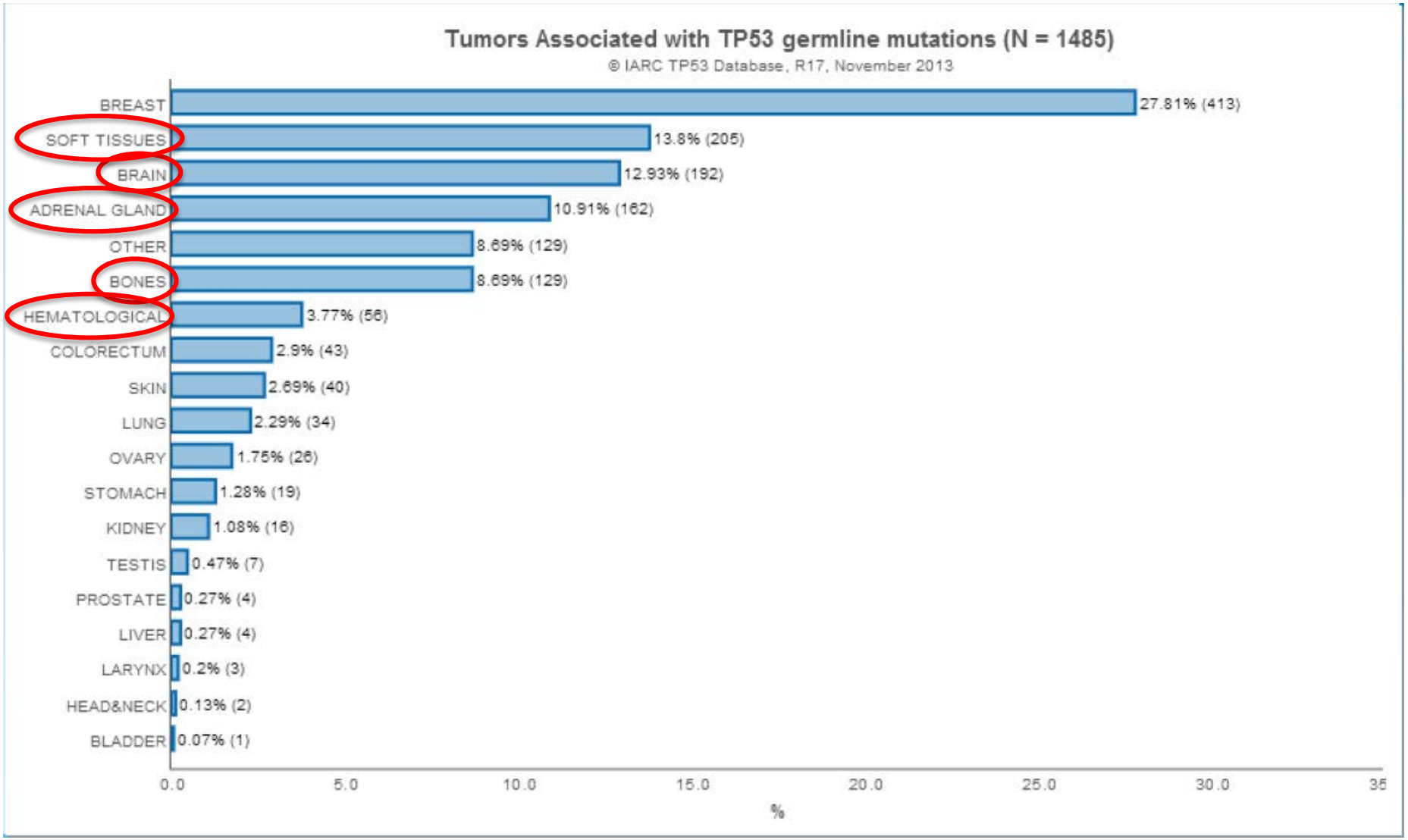
In order to more efficiently study the medical, genetic, psychological, and social functioning issues associated with LFS, the Clinical Genetics Branch (CGB) of the National Cancer Institute (NCI) joined with researchers from around the world to form the Li-Fraumeni Exploration (LiFE) Research Consortium. The most pressing research questions related to LFS require assembling information from multiple research institutions, in order to have the large number of affected and unaffected family members required for research of this kind. In support of both the LiFE consortium and its own research program, CGB is also conducting a detailed study (based at the National Institutes of Health/National Cancer Institute in Bethesda, MD) of individuals with LFS and their family members.

Research Highlights

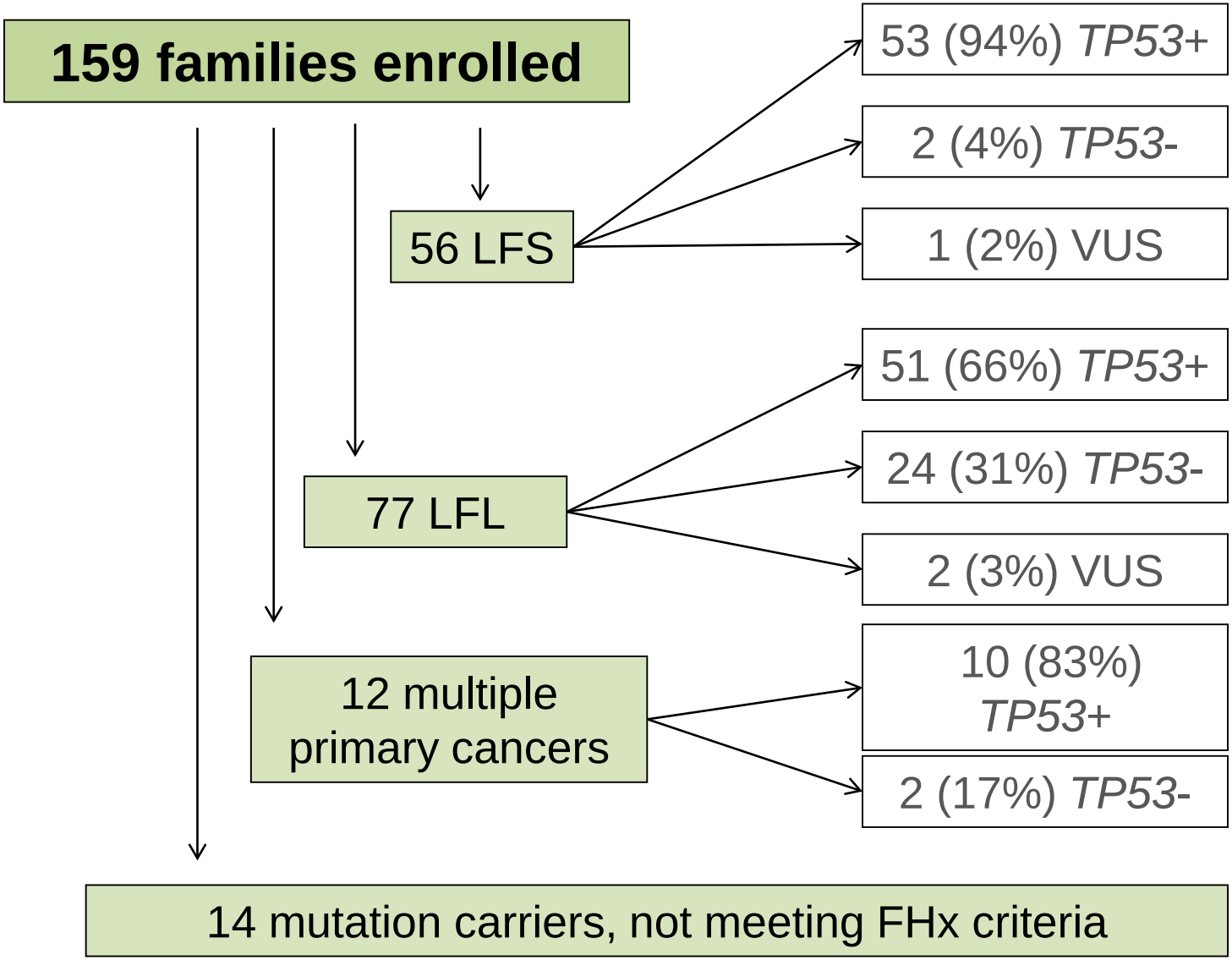
CGB has just launched a pilot study targeting Physical Activity and Diet (PAD) in LFS families. This is a first step towards determining whether it is feasible, practical, to collect data of this kind from our study participants. The long-range goal of this effort is to determine whether there might be opportunities to attempt modification of these important lifestyle variables, with an eye towards reducing cancer risk and improving overall health and well-being in LFS families. A random sample of LFS study participants will be invited to participate, by providing information about their daily diet and physical activity using a web-based data collection tool. If you have questions about this effort, please do not hesitate to contact us.

http://lfs.cancer.gov

Li-Fraumeni Syndrome



Li-Fraumeni Syndrome at NCI: Family Accrual

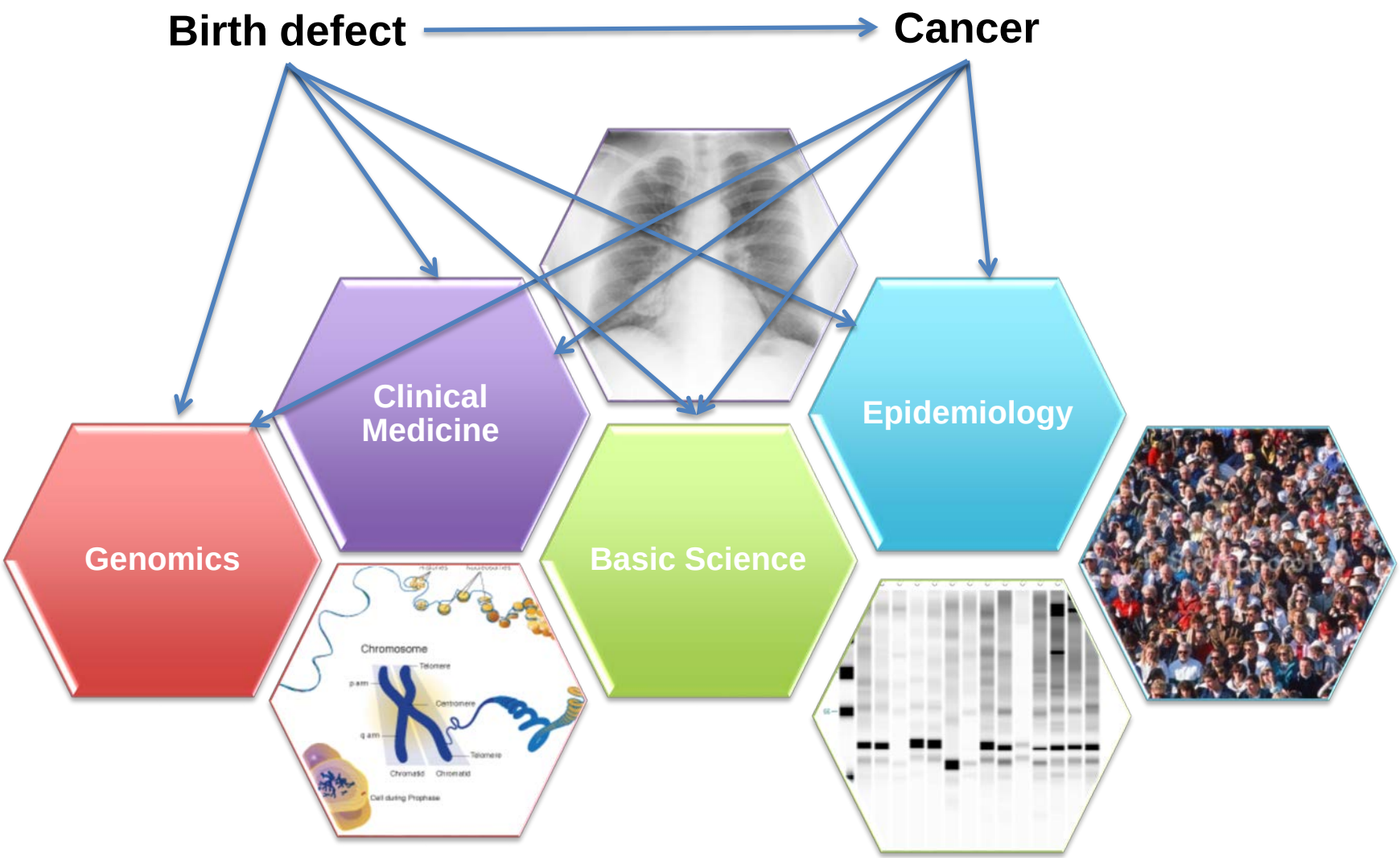


Li-Fraumeni Syndrome at NCI

- More precisely define the clinical spectrum
- Identify potential risk modifying factors
- Establish an effective cancer screening program
- Identify other potential novel causative gene(s)
- Evaluate the psychological and social functioning effects in affected families
- Genetic counselling and testing of appropriate family members

http://lfs.cancer.gov

Integrated Approach to Understanding Cancer Susceptibility



Acknowledgements

- **Clinical Genetics Branch**
  - **Sharon Savage**
  - Blanche Alter
  - Neelam Giri
  - Phuong Mai
  - Douglas Stewart
  - Mark Greene
  - June Peters
  - Rosamma DeCastro
  - Jennifer Loud
  - Renee Bremer
  - Jennifer Young
- **Cancer Genomics Research Laboratory**
- **Research support**
  - Lisa Leathwood
  - Maureen Risch
  - Janet Bracci
  - Katherine Beebe
  - Kathryn Nichols
  - Laura Harney
  - Ann Carr
- **Patients and families**

The Human Microbiome Data Coordination Center

Owen White
University of Maryland School of Medicine
Gabriella Miller Kids First Pediatric Research Program NIH Data Workshop
April 2016.

NIH HUMAN MICROBIOME PROJECT

REFERENCE GENOMES | MICROBIOME ANALYSIS | IMPACTS ON HEALTH | TOOLS & TECHNOLOGY | ETHICAL IMPLICATIONS | OUTREACH | HMPDACC DATA BROWSER

Feedback

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.

GET DATA

GET TOOLS

Current News

July 2013
Human Microbiome Science: Vision for the Future conference to be held in Bethesda, MD July 24-26

May 2013
Human Microbiome Consortium Virtual Meeting: Approaches in Microbiome Assembly

May 2013
Booth at ASM 2013 (#639)

More News Items

Publications

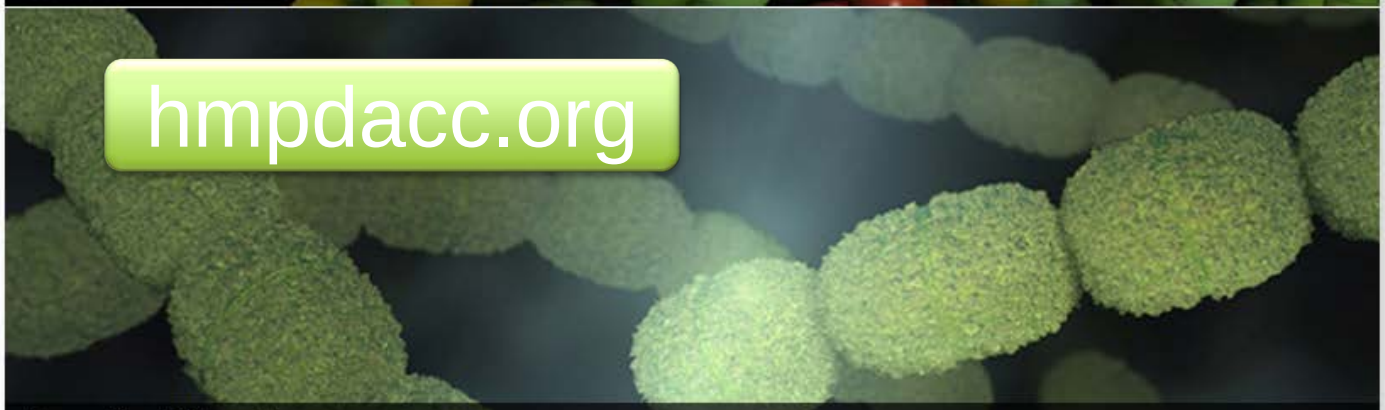
- Colitis-induced Bone Loss is Gender Dependent and Associated with Incr...
- Topographic diversity of fungal and bacterial communities in human ski...
- Comparative metagenomic and rRNA microbial diversity characterization ...

More Publications

Partner Resources

- NIH Common Fund
- NCBI HMP Data Repository

Areas of Interest



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...

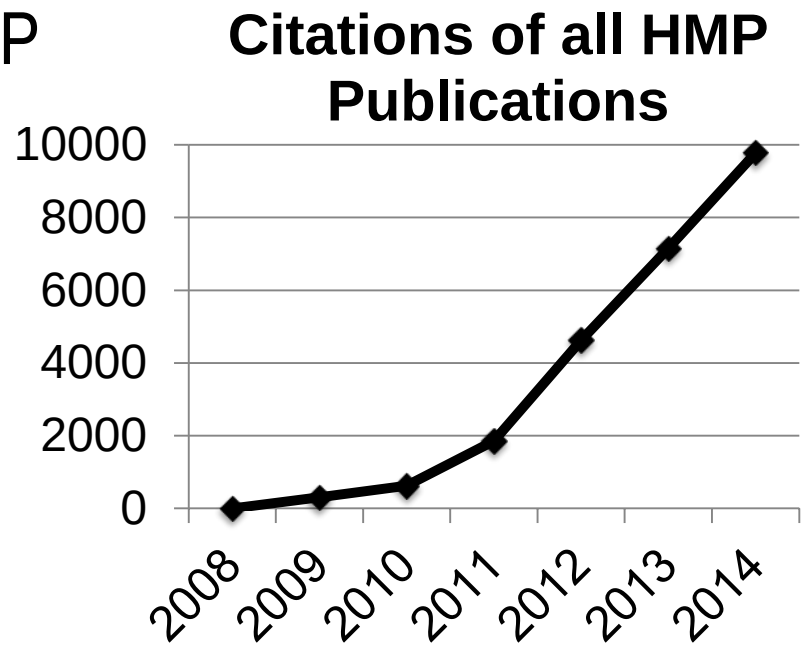
DACC Member Organizations

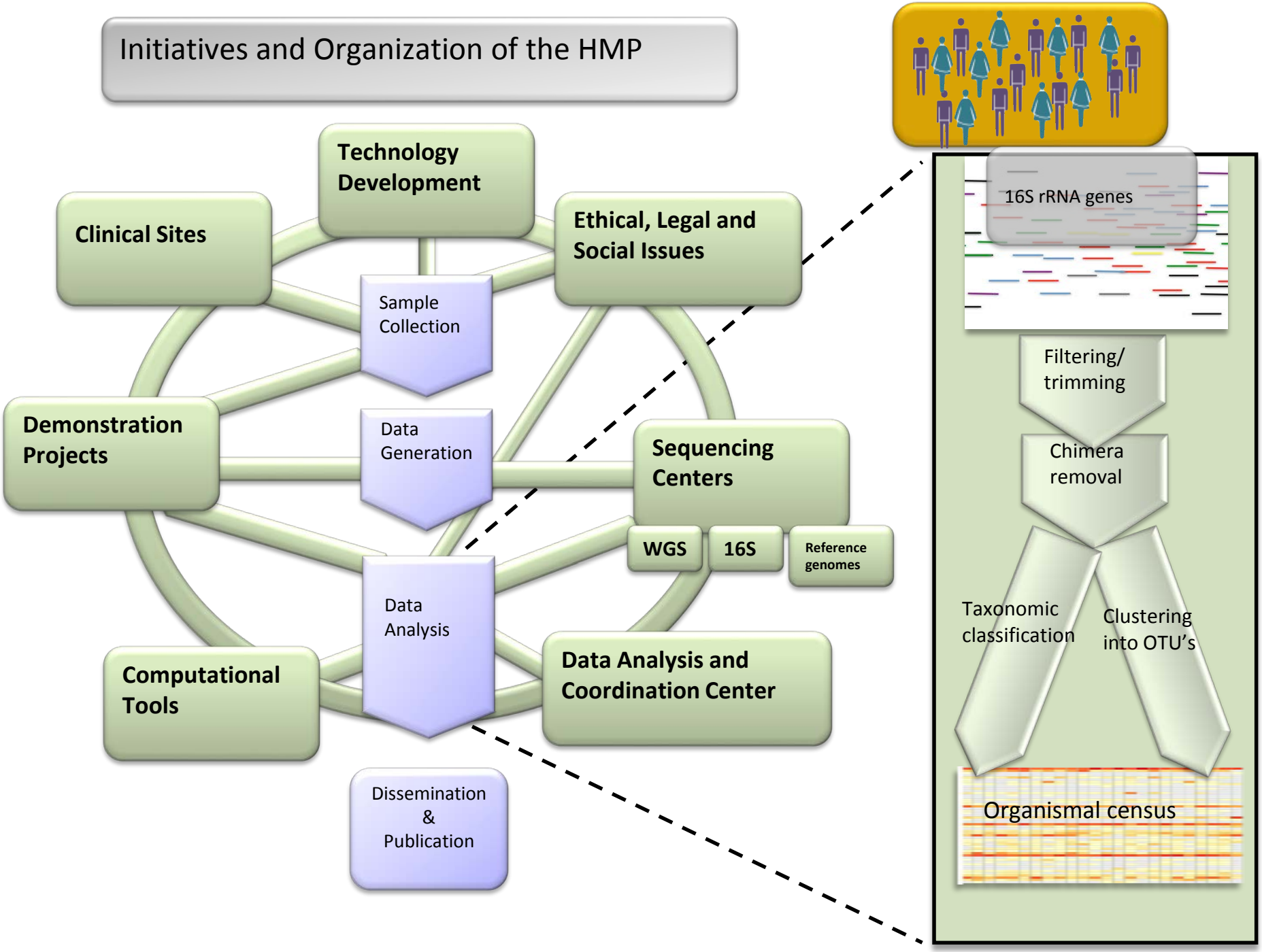
Related Sites

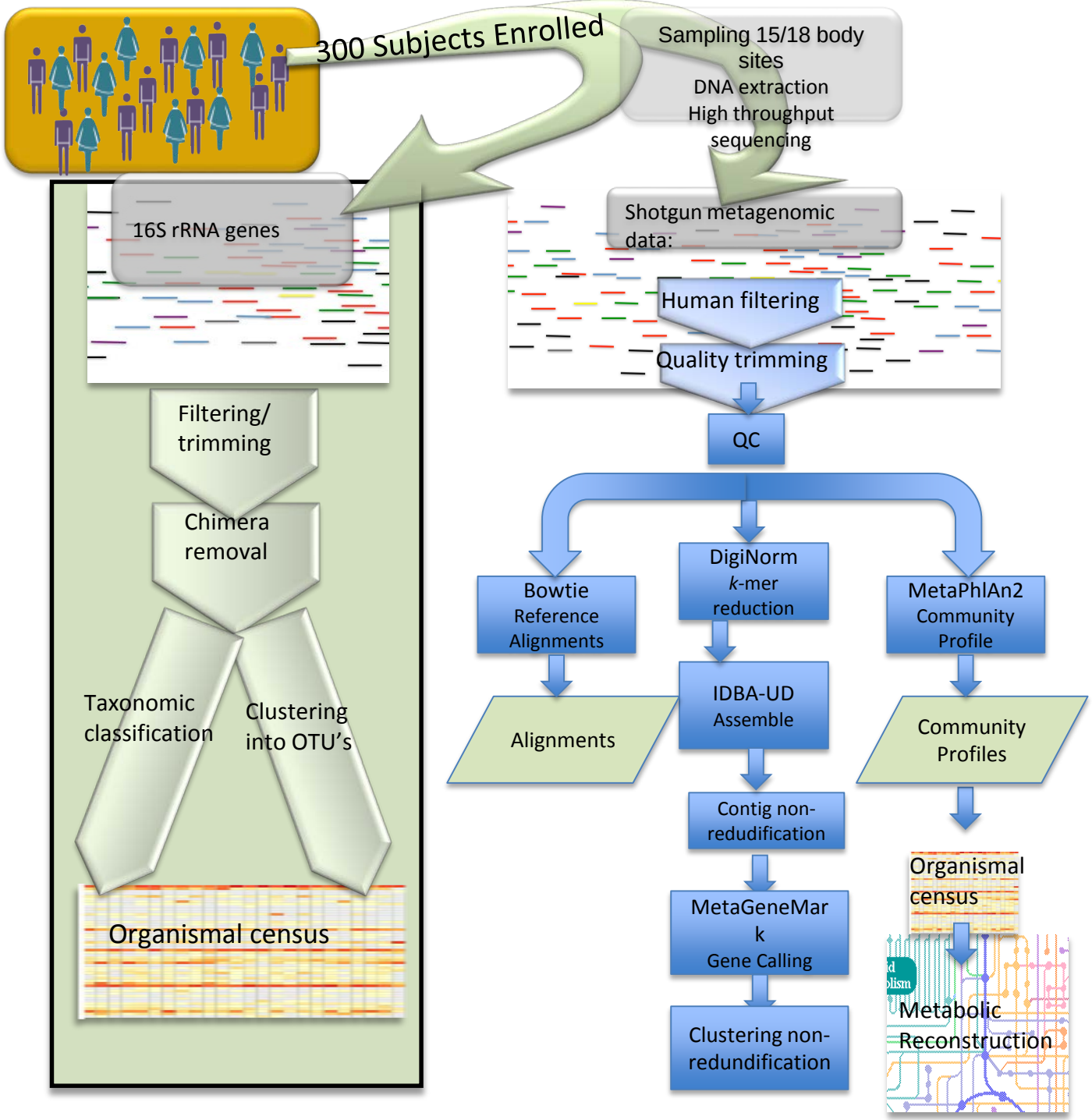
img/hmp m

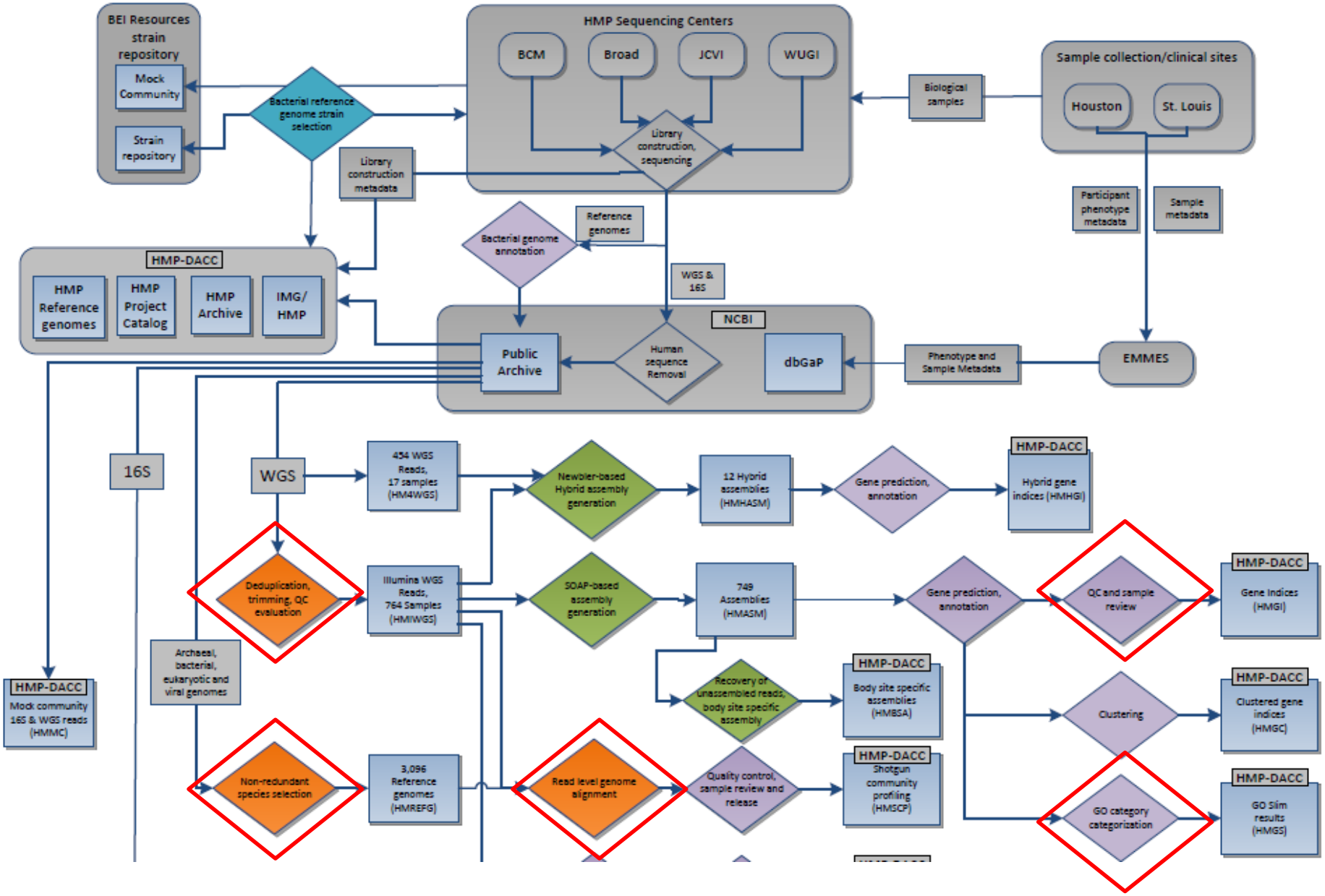
CloVR

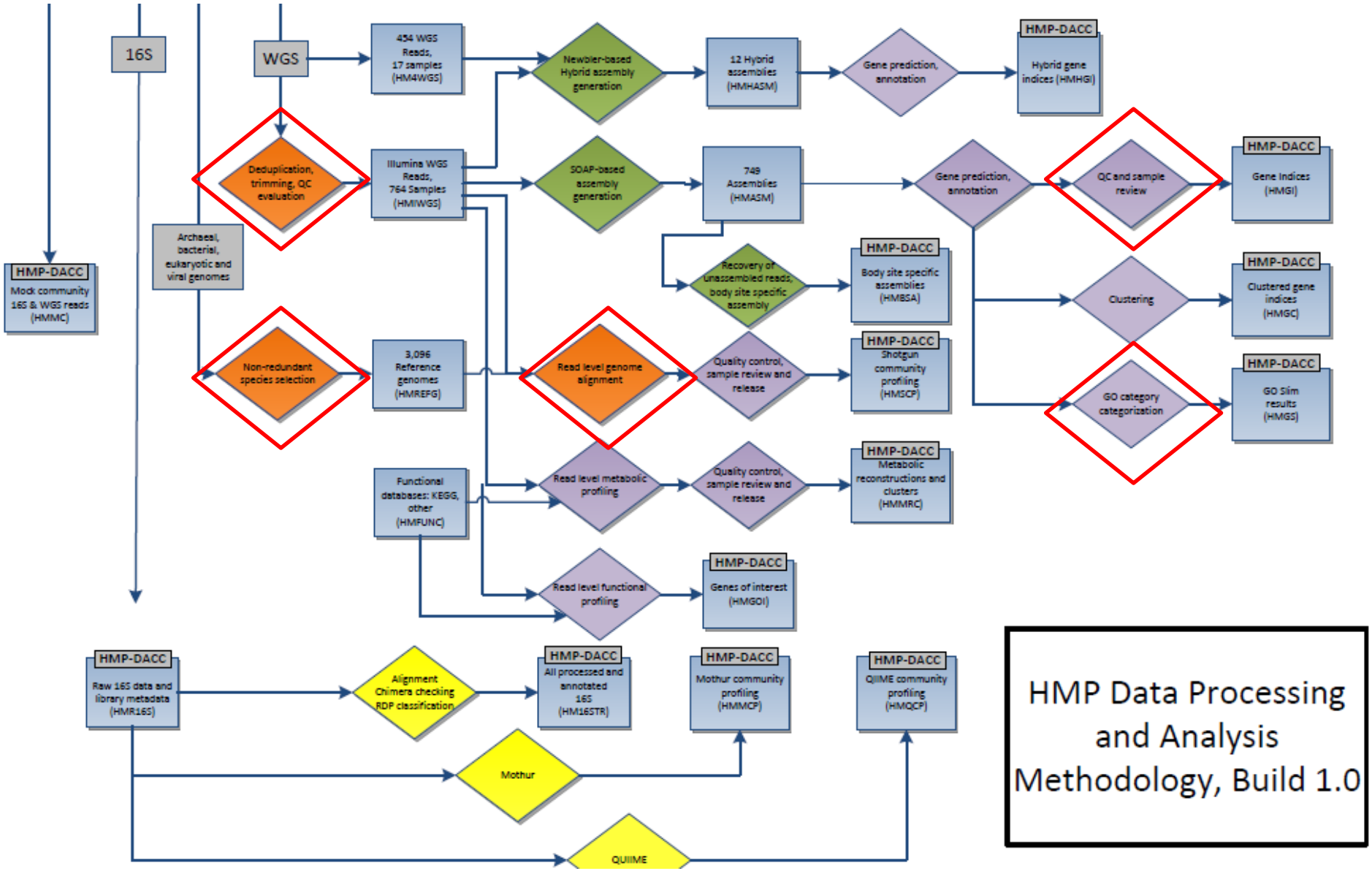
- Citations of original HMP publications : 470
- Total number of DACC website users: 157,855 (45% returning visits)
- Number of users in 2014: 50,733




Initiatives and Organization of the HMP
Technology Development
Clinical Sites
Ethical, Legal and Social Issues
Sample Collection
Demonstration Projects
Data Generation
Sequencing Centers
WGS
16S
Reference genomes
Data Analysis
Computational Tools
Data Analysis and Coordination Center
Dissemination & Publication
16S rRNA genes
Filtering/ trimming
Chimera removal
Taxonomic classification
Clustering into OTU's
Organismal census


300 Subjects Enrolled
Sampling 15/18 body sites
DNA extraction
High throughput sequencing
16S rRNA genes
Filtering/ trimming
Chimera removal
Taxonomic classification
Clustering into OTU's
Organismal census
Shotgun metagenomic data:
Human filtering
Quality trimming
QC
Bowtie Reference Alignments
DigiNorm k-mer reduction
MetaPhlAn2 Community Profile
Alignments
IDBA-UD Assemble
Community Profiles
Contig non-redudification
MetaGeneMark Gene Calling
Clustering non-redundification
Organismal census
Metabolic Reconstruction


BEI Resources strain repository
Mock Community
Strain repository
Bacterial reference genome strain selection
HMP Sequencing Centers
BCM
Broad
JCVI
WUGI
Library construction, sequencing
Biological samples
Sample collection/clinical sites
Houston
St. Louis
Library construction metadata
Participant phenotype metadata
Sample metadata
Bacterial genome annotation
Reference genomes
WGS & 16S
HMP-DACC
HMP Reference genomes
HMP Project Catalog
HMP Archive
IMG/ HMP
NCBI
Public Archive
Human sequence Removal
dbGaP
Phenotype and Sample Metadata
EMMES
16S
WGS
454 WGS Reads, 17 samples (HM4WGS)
Newbler-based Hybrid assembly generation
12 Hybrid assemblies (HMHASM)
Gene prediction, annotation
HMP-DACC
Hybrid gene indices (HMHGI)
Deduplication, trimming, QC evaluation
Illumina WGS Reads, 764 Samples (HMIWGS)
SOAP-based assembly generation
749 Assemblies (HMASM)
Gene prediction, annotation
QC and sample review
HMP-DACC
Gene Indices (HMGI)
Archaeal, bacterial, eukaryotic and viral genomes
HMP-DACC
Mock community 16S & WGS reads (HMMC)
Recovery of unassembled reads, body site specific assembly
HMP-DACC
Body site specific assemblies (HMBSA)
Clustering
HMP-DACC
Clustered gene indices (HMGC)
Non-redundant species selection
3,096 Reference genomes (HMREFG)
Read level genome alignment
Quality control, sample review and release
HMP-DACC
Shotgun community profiling (HMSCP)
GO category categorization
HMP-DACC
GO Slim results (HMGS)



HMP Data Processing and Analysis Methodology, Build 1.0

rotocols

Author: Brandi Cantarel
Version:
Effective Date:

1 Abstract

2 Introduction

This SOP describes creation of a non-redundant catalog of bacterial genes by body site. This was done by clustering the gene predictions coming out of the Illumina metagenomic wgs sequence, using the same parameters used by the Metahit project[1] to cluster their human intestinal tract data.

3 Requirements

3.1 Software requirements

USEARCH (http://www.drive5.com/usearch/)

4 Procedure

4.1 Clustering

Clustering was performed as described in Qin et al[1], with the exception that clustering was done here using USEARCH, rather than BLAT.

MetaGeneMark predicted ORFs (available for download at http://hmpdacc.org/HMGI/) were aligned to one another using Usearch[2] (http://www.drive5.com/usearch/), version 5.0.144. The following options were used:

-id 0.95
-targetfract 0.90

5 Implementation

6 Discussion

7 Related Documents & References

[1] Qin, J. et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature;464(7285):59-65.

[2] Edgar, RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics; 26(19):2460-1.

Documented Protocols

Reference genomes - 14

Annotation, gene naming, core gene evaluation, genome assembly metrics

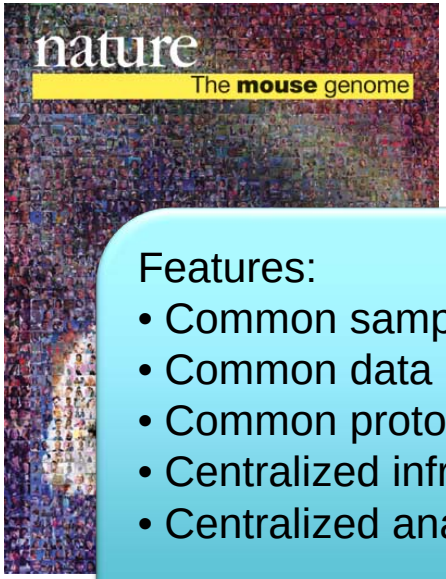
16S - 8

Human sequence removal, community profiling, file generation

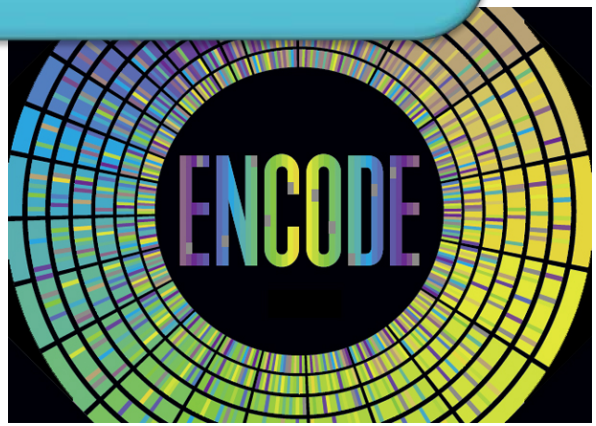
WGS – 11

Human screening, read processing, assembly, annotation, metabolic reconstruction

[2] Edgar, RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics; 26(19):2460-1.



Features:
- Common sample provenance
- Common data provenance
- Common protocol design
- Centralized infrastructure
- Centralized analysis infrastructure



human microbiome stool - SRA - NCBI - Mozilla Firefox

File Edit View History Bookmarks Tools Help

human microbiome stool - SRA - NCBI

www.ncbi.nlm.nih.gov/sra/?term=human+microbiome+stool

Google Feedback

NCBI Resources How To Sign in to NCBI

SRA SRA human microbiome stool Search

Save search Advanced Help

Show additional filters

Access
Controlled (641)
Public (746)

Source
DNA (675)
RNA (26)
metagenomic (55)

Type
genome (424)

Clear all

Show additional filters

Display Settings: Summary, 20 per page

Send to:

Results: 1 to 20 of 746

<< First < Prev Page 1 of 38 Next > Last >>

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

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

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

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

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

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

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

Filters: Manage Filters

Top Organisms [Tree]
human metagenome *(681)*
human gut metagenome *(32)*
unidentified *(32)*
human lung metagenome *(1)*

Search in related databases

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?

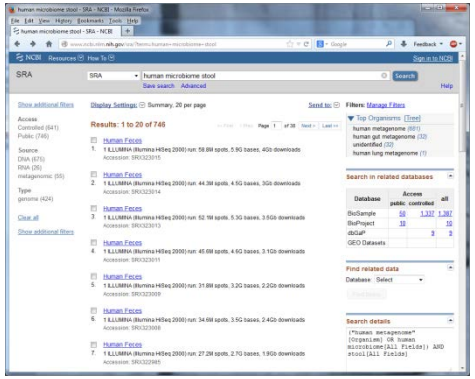
Subject/volunteer still available?

Publication, downstream citations

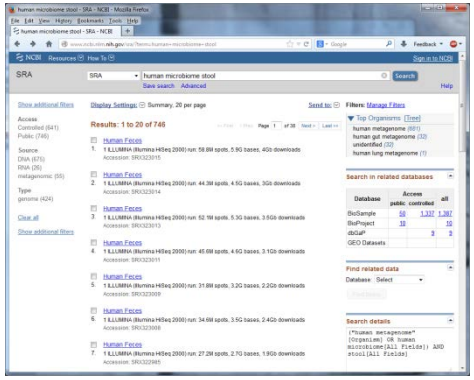
Quality in comparison to all others

Patient phenotype

Associated with disease?



Protocol harmonization



Phenotyping

Exome sequencing

Functional genomics/`omics preparation

Informatics:

- Variant types: (CNV, de novo mutations)
- Leveling across all known datasets

Insert IHMC code stuff.

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 of 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 Skin Microbiome Consortium Study of Atopic Dermatitis and Primary Immunodeficiencies	The Human Virome in Children and Its Relationship to Febrile Illness	The Neonatal Microbiome and Necrotizing Enterocolitis	The Human Gut Microbiome and Recurrent Abdominal Pain in Children	The Vaginal Microbiome: Disease, Genetics and the Environment	Ulcerative Colitis Human Microbiome Project	Human Microbiome Demo Evaluation of the C
SUBJID	1.00	0.88	0.13			SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	SUBJID	Subject_ID	SUBJID	Patient ID	Subject ID
Gender	0.94	0.94			0.06	Gender		SEX	Sex	SEX	sex			Gender	GENDER	Gender		Gender	Gender
Age	0.88	0.81	0.06	0.06	0.06	Age_at_first_visit	AgeAtEnrollment	AGE	DOB	AGE	age	age	Age_status	Age_years, Age_mth	AGE	Age	age		Age
Race	0.81	0.44	0.38		0.19	Race	Race_Other_Text			RACE	ethnicity	ethnicity	Race	Race	RACE	Ethnicity	racial_background		Race
Other Rrace	0.56	0.31	0.25		0.44	Other Race	Race_Other				ethnicity	ethnicity		Race			racial_background		Race
Smoking	0.38	0.31	0.06		0.63	Smoking_status				SMOKING_STATUS		smoke_ever					smoking_status	Tob	
Lab	0.31	0.19	0.13	0.06	0.63	Diagnosis	TID			CLOSTRIDIA_DIFFICILE		nugent_score			VITAL_STATUS				PASI
Smoking_duration	0.31	0.19	0.13		0.69	Smoking_status						smoke_ever					age_startsmoking	Tob Year Start_Tob	
Drugs	0.31	0.19	0.13		0.69	Antacids, Steroids, Antibiotics			Antibiotics		Currently Treated?	treatment1, treatment2							ABX
Weight_kg	0.25	0.25		0.06	0.69								Weight	Weight_kg	BIRTHWT				
BP	0.19	0.19			0.81			SBP, DBP						S_BP, D_BP					
Weight_lbs	0.19	0.19			0.81								Weight						
Height	0.19	0.19			0.81										FEET, INCHES				
Disease	0.19	0.06	0.13		0.81				Status		face, nose								Family Hist
Institution	0.13	0.00	0.13		0.88							highest_grade					education		
Dose	0.13	0.06	0.06		0.88										FEEDING_START				
Duration	0.13	0.06	0.06		0.88										FEEDING_ACHIEVED		days_since_most_recent_use_doctor		
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													Body site
Drug_name	0.06		0.06	0.06	0.88					DURATION									Duration

Insert IHMC code stuff.

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 of 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 Scale Resolution
SUBJID	1.00	0.88	0.13			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
Race	0.81	0.44	0.38		0.19	Race	Race_Other_Text			RACE	ethnicity	ethnicity
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							
BP	0.19	0.19			0.81			SBP, DBP				
Weight_lbs	0.19	0.19			0.81							
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						
Drug_name	0.06		0.06	0.06	0.88					DURATION		

Insert IHMC code stuff.

dbGaP

The Microbial Ecology of Bacterial Vaginosis: A Fine Scale Resolution Metagenomic Analysis	NIH Intramural Skin Microbiome Consortium Study of Atopic Dermatitis and Primary Immunodeficiencies	The Human Virome in Children and Its Relationship to Febrile Illness	The Neonatal Microbiome and Necrotizing Enterocolitis	The Human Gut Microbiome and Recurrent Abdominal Pain in Children	The Vaginal Microbiome: Disease, Genetics and the Environment	Ulcerative Colitis Human Microbiome Project	Human Microbiome Demo Evaluation of the C
SUBJID	SUBJID	SUBJID	SUBJID	Subject_ID	SUBJID	Patient ID	Subject ID
		Gender	GENDER	Gender		Gender	Gender
age	Age_status	Age_years, Age_mth	AGE	Age	age		Age
ethnicity	Race	Race	RACE	Ethnicity	racial_background		Race
ethnicity		Race			racial_background		Race
smoke_ever					smoking_status	Tob	
nugent_score			VITAL_STATUS				PASI
smoke_ever					age_startsmoking	Tob Year Start_Tob	
treatment1, treatment2							ABX
	Weight	Weight_kg	BIRTHWT				
		S_BP, D_BP					
	Weight						
			FEET, INCHES				
							Family Hist
highest_grade					education		
			FEEDING_START				
			FEEDING_ACHIEVED		days_since_most_recent_use_doctor		
							Body site
							Duration

Things are likely to be worse here

Many different submitters

Much less motivation to be consistent

It is imperative to compare to external datasets

Sharing data is much larger in scope

Loss in power

Nat Biotechnol. 2011 May;29(5):415-20. doi: 10.1038/nbt.1823.

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.

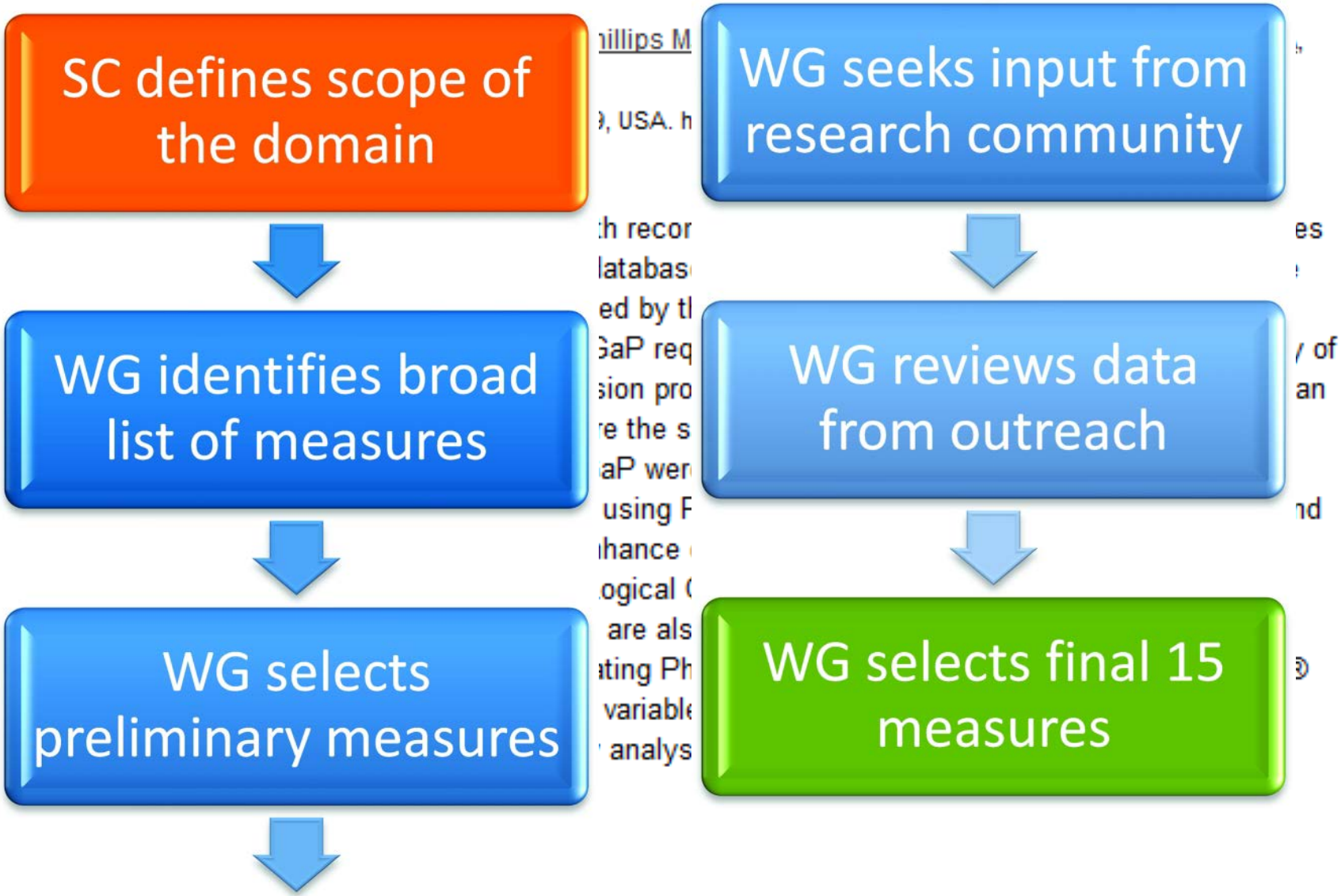
Pan H, Tryka KA, Vreeman DJ, Huggins W, Phillips MJ, Mehta JP, Phillips JH, McDonald CJ, Junkins HA, Ramos EM, Hamilton CM.

RTI International, Research Triangle Park, NC 27709, USA. hpan@rti.org

Abstract

The PhenX Toolkit provides researchers with recommended, well-established, low-burden measures suitable for human subject research. The database of Genotypes and Phenotypes (dbGaP) is the data repository for a variety of studies funded by the National Institutes of Health, including genome-wide association studies. The dbGaP requires that investigators provide a data dictionary of study variables as part of the data submission process. Thus, dbGaP is a unique resource that can help investigators identify studies that share the same or similar variables. As a proof of concept, variables from 16 studies deposited in dbGaP were mapped to PhenX measures. Soon, investigators will be able to search dbGaP using PhenX variable identifiers and find comparable and related variables in these 16 studies. To enhance effective data exchange, PhenX measures, protocols, and variables were modeled in Logical Observation Identifiers Names and Codes (LOINC®). PhenX domains and measures are also represented in the Cancer Data Standards Registry and Repository (caDSR). Associating PhenX measures with existing standards (LOINC® and caDSR) and mapping to dbGaP study variables extends the utility of these measures by revealing new opportunities for cross-study analysis.

PhenX measures to identify opportunities for cross-study analysis.



PhenX Toolkit

My Toolkit

Register | Log In

Home | Browse | Search | Registration | Resources | News | Help | About

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.

Registration

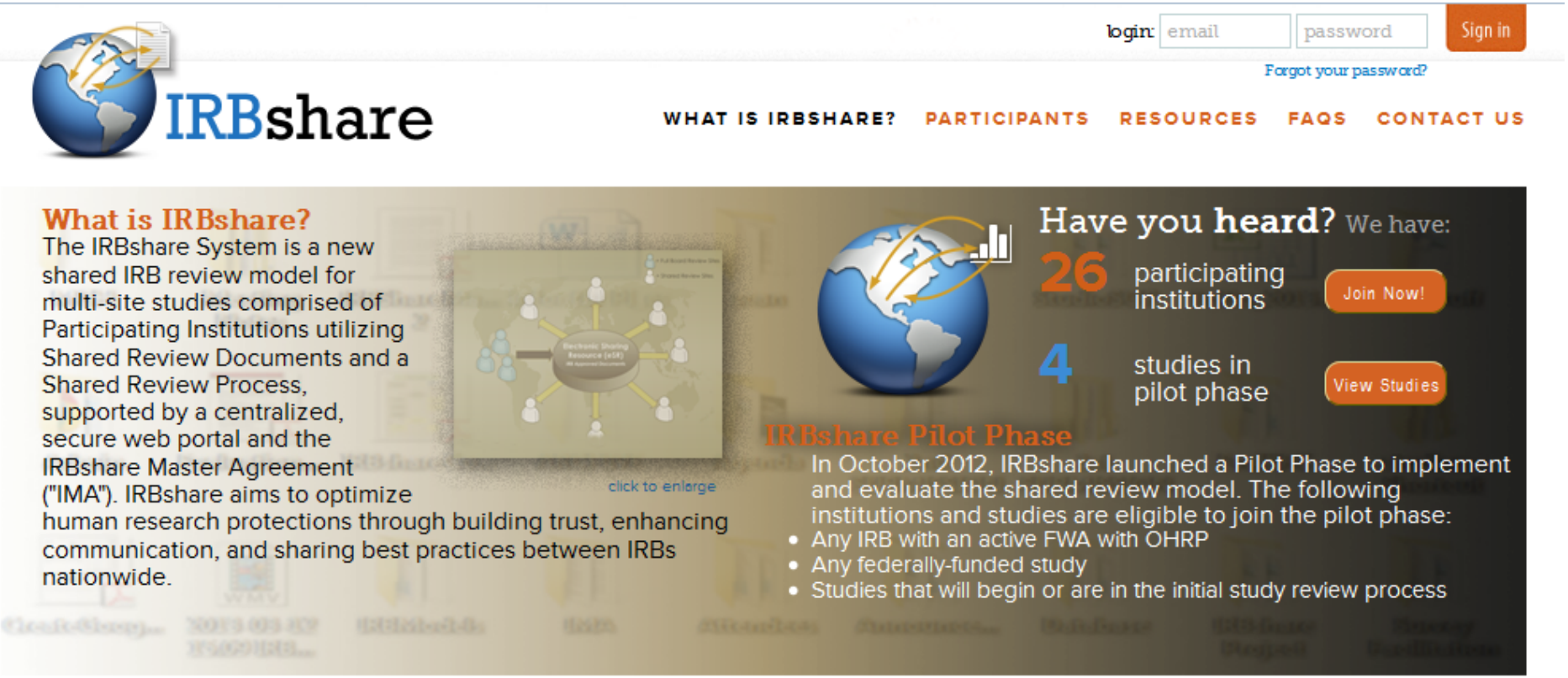
You may browse the PhenX Toolkit, but to save your work, you need to Register. See additional benefits for registered users.

Existing users may login:

User ID:

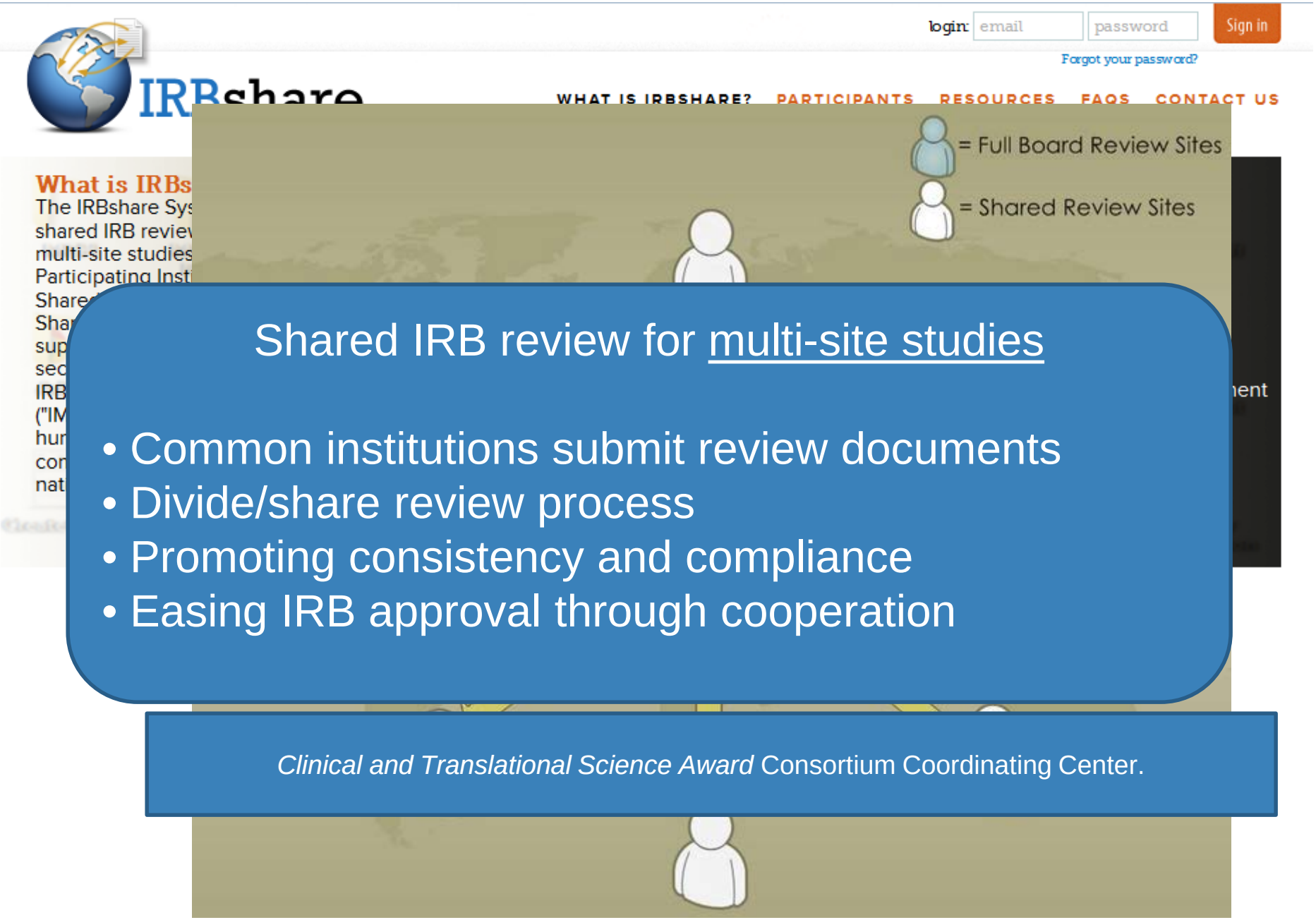
Password:

Log In



Clinical and Translational Science Award Consortium Coordinating Center.

login: email
password
Sign in
Forgot your password?
IRBshare
WHAT IS IRBSHARE?
PARTICIPANTS
RESOURCES
FAQS
CONTACT US
What is IRBs
The IRBshare Sys
shared IRB revie
multi-site studies
Participating Inst
Shared Review D
Shared Review P
supported by a c
secure web porta
IRBshare Master
("IMA"). IRBshare
human research
communication,
nationwide.
= Full Board Review Sites
= Shared Review Sites
IRBshare System
IRB Approved Documents



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

Clinical and Translational Science Award Consortium Coordinating Center.

PI Registry

Contact PIs, recuit

Capture:

- IRB forms with specific description of patient phenotypes in your study
- Biosamples or volunteers available to consortium

ID:

- Other volunteers relevant to your study
- Samples relevant to your
- Reference samples/resources
- Methods
- Potential role in protocol in harmonization

Can the cloud solve our problems?

Maybe, but not a panacea

Not necessarily cheaper

Not necessarily easier

<u>Absolutely will not improve standards</u>

<u>Absolutely will not solve sharing</u>

Thank you

IGS

Michelle Giglio
Jonathan Crabtree
Heather Creasy
Victor Felix
Joshua Orvis
Anup Mahurkar

Harvard

Curtis Huttenhower

UCSD

Rob Knight

NIH – NHGRI
Common Fund

iHMP Centers

Baylor College of Medicine
Texas Medical Center, Houston, Texas

BROAD INSTITUTE

J. Craig Venter INSTITUTE

Washington University in St. Louis
School of Medicine

Bench to Bassinet

Data Sharing: The PCGC Experience

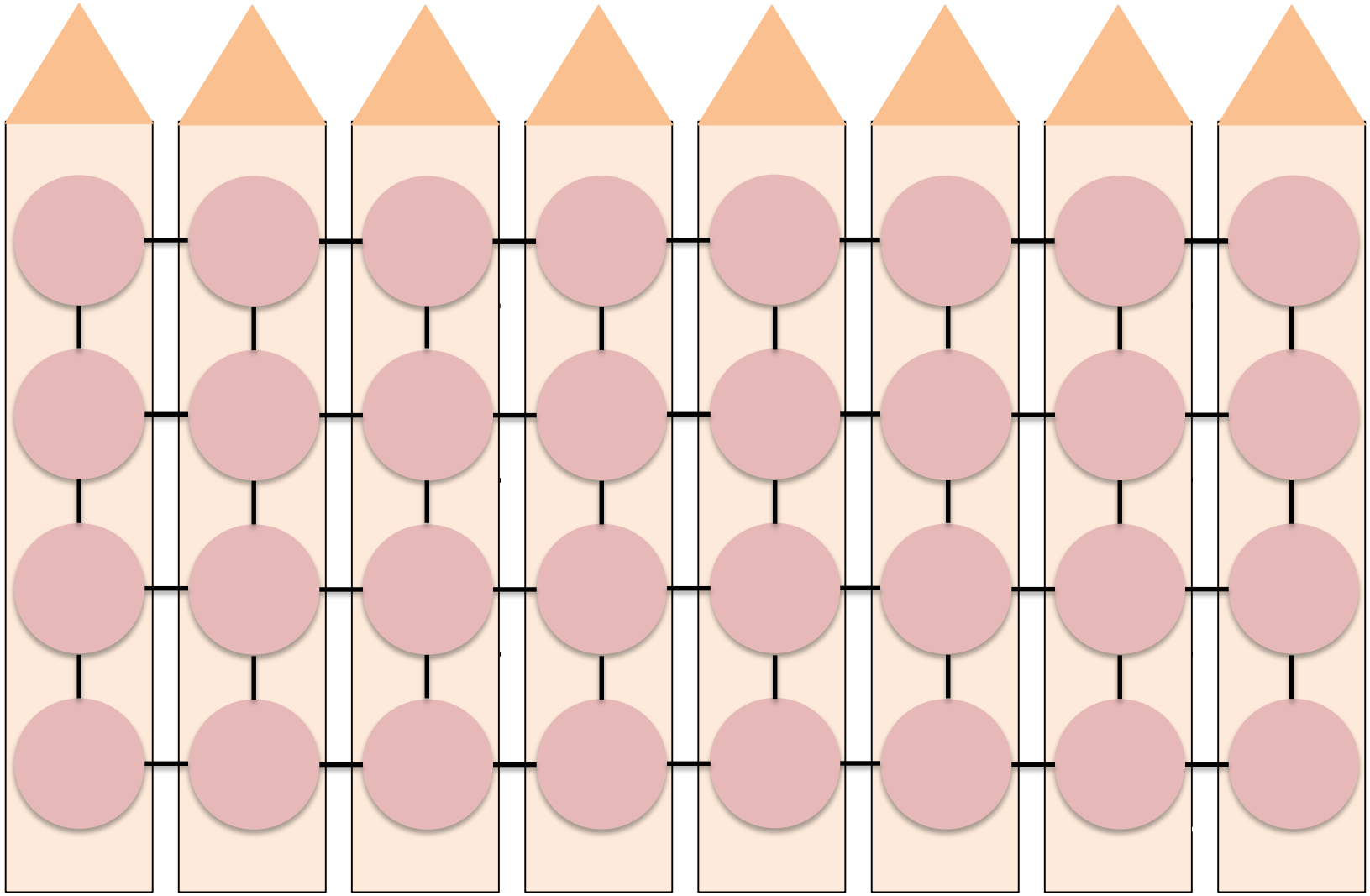
Gabriella Miller Kids First Data Resource Workshop
National Institutes of Health

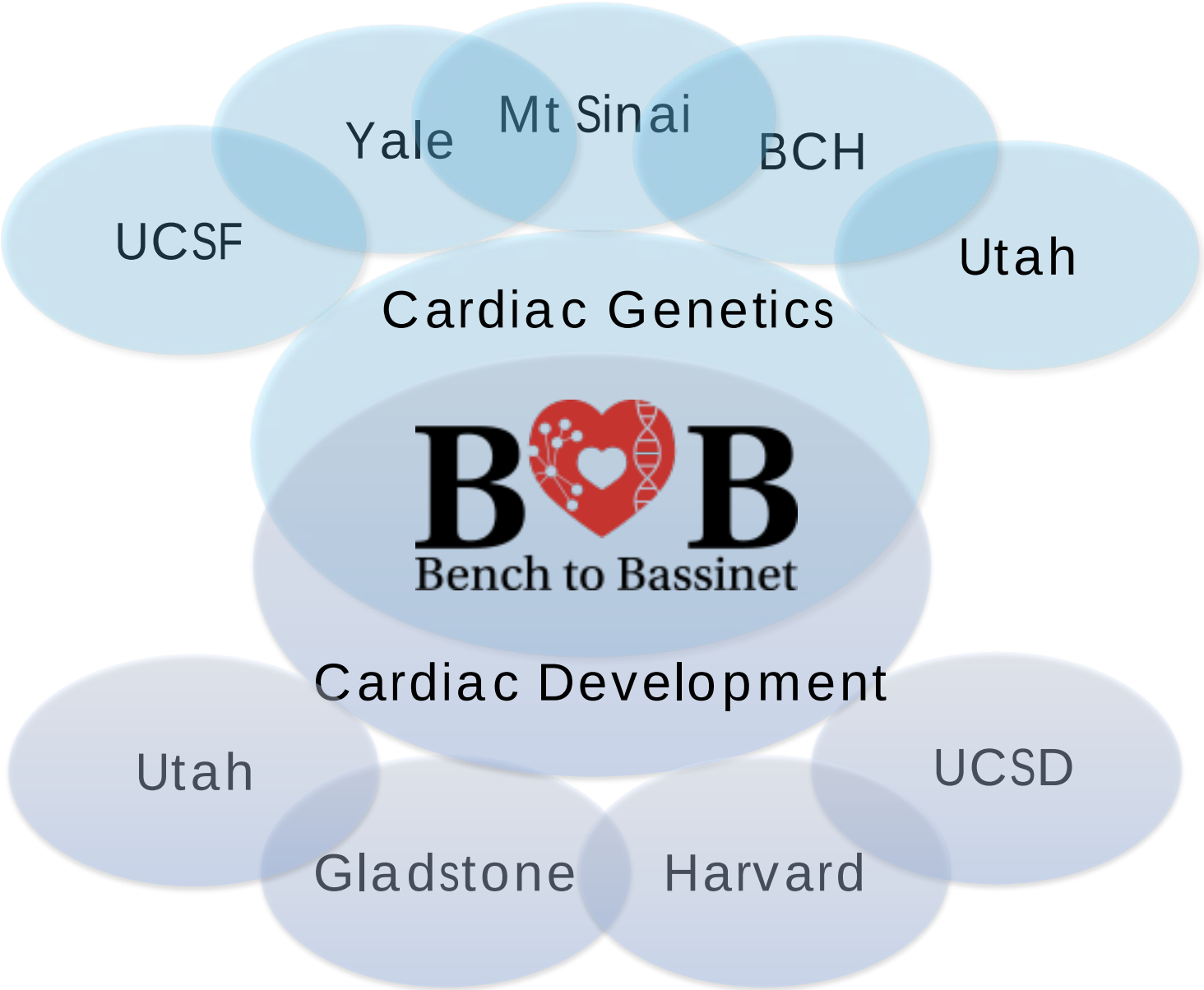
April 26, 2016

Opportunity

Bench to Bassinet

Themes




Yale
Mt Sinai
BCH
UCSF
Utah
Cardiac Genetics
B B
Bench to Bassinet
Cardiac Development
Utah
UCSD
Gladstone
Harvard

Bench to Bassinet

24770 subjects

9702 genomes

140 TB of data

676 elements

Data Resource

Bench to Bassinet

Awareness creates opportunity

PCGC Data Hub

- Query, reporting, discovery
- Genotype-phenotype correlations
- Virtual cohorts

Principles

- Access immediacy
- Interoperability, reproducibility, provenance
- Secure environment
- Distribution node (dbGaP)

Search HeartsMart

Login to HeartsMart to use the interactive query builder to:

- **Identify** cohorts of subjects.
- **Locate** raw and analyzed molecular data sets.
- **Export** phenotype data for further investigation.
- **Download** data files matching your criteria.

News

Dec. 14, 2015
Exomes published
Dec. 3, 2015
Exomes published
Nov. 30, 2015
Data refresh (NERI November 10, 2015 update)

More news

Support

Have a question or need some help? Leave the PCGC Support Team a message using our support form.

Upload

Genotyping, **Candidate Gene Evaluation**, and **Whole Exome** cores can submit data by following these procedures.

Clinical Phenotypes for **existing samples** can be provided to the core by using our REDCap database.

Confirmation Core CNV and variant call confirmations can also be provided to the core by using our REDCap database.

Download

Molecular data including whole **whole exome**, **target sequencing**, and **SNP genotyping** results are available via our secure, high-speed Data Expedition server.

Download the MTPexpedat (GUI) or movedat (CLI) client applications on our Data Expedition page.

SFTP is no longer supported for PCGC file transfers.

Harvest

Demos | Docs | Download | Articles | Webinars | About | Resources | Contact

Developed by CBMi at The Children's Hospital of Philadelphia Research Institute

Discover Harvest

Explore your data, *not* your database.

Designed by and for biomedical researchers, The Harvest Stack is an **open source BSD-licensed** toolkit for building web applications for integrating, discovering, and reporting data.

Try A Demo | Watch | Read More

Open Source & Available on GitHub

We believe in open source software and open-source our work. Harvest is licensed under BSD 2-clause License.

Web-Based Technologies

Harvest comes with an HTML5 web client backed by a set of discoverable REST APIs.

Designed for Biomedical Data First

Created by the Center for Biomedical Informatics at The Children's Hospital of Philadelphia Research Institute, Harvest addresses the unique needs of biomedical researchers.

Nationally Funded

National Human Genome Research Institute (NHGRI), National Institute on Deafness and Other Communication Disorders (NIDCD), National Heart, Lung and Blood Institute (NHLBI), and others have funded Harvest applications.

Pennington et al, JAMIA, 2014

https://rescows02.research.chop.edu/pcgc/query/

PCGC DataHub | Workspace | Query | Results | Genomic Data | Download Queue | Admin | Michael Italia

Search by name, description, or data.

- Subject +
- Sample +
- Variants (Beta) +
- Demographics +
- Diagnoses −

B2B (FYLER) DIAGNOSES

- Conotruncal Lesion
- Heterotaxy
- Left Ventricular Outflow Tract Obstruction
- Tetralogy Of Fallot
- Hypoplastic Left Heart Syndrome
- Atrial Septal Defect
- Fyler Diagnosis
- Age Reported

CARDIAC ABNORMALITIES

- Abdominal Situs
- Cardiac Situs
- Systemic Vein
- Hepatic Vein
- Pulmonary Vein
- Right Atrium

Diagnoses > B2B (Fyler) Diagnoses

Heterotaxy

Heterotaxy per B2B Comm definition, based on Fyler codes and extracardiac findings

Unique values 2

Exclude selected values

false	7067 (91%)
true	677 (8%)

Update Filter

The PCGC Data Hub 3.0b6, powered by Cilantro 2.2.34-beta · Funded by the NHLBI's B2B Program
Developed, hosted, and maintained by The Children's Hospital of Philadelphia
Terms of Use · Privacy Policy

70 individuals

- Gene Symbol is either KCNA1, CFC1, FOXH1 or GDF1
- Effect Region is Exon
- Heterotaxy is True

View Results

https://rescows02.research.chop.edu/pcgc/query/

PCGC DataHub | Workspace | Query | Results | Genomic Data | Download Queue | Admin | Michael Italia

Search by name, description, or data.

- Subject +
- Sample +
- Variants (Beta) −
  - Gene Symbol
  - RefSeq ID
  - Variant Type
  - Genotype
  - Coverage/Quality
  - Genomic Position
  - Deleterious Predictions
  - **Predicted Variant Effects**
  - dbSNP ID
- Demographics +
- Diagnoses +
- Genomics +
- Extracardiac +
- Procedures +
- Genetic Testing +

Unique values 11

‹ Prev | Next › | page 1 / 2 | ×Clear

Search Effect Regions...

- Intergenic +
- Upstream +
- UTR-5 Prime +
- Slice-Site Acceptor +
- Slice-Site Donor +
- Exon −
- Intron +
- UTR-3 Prime +
- Downstream +
- Regulation +

Exon

Update Filter

Functional Class

Unique values 3

‹ Prev | Next › | page 1 / 1 | ×Clear

Enter or paste a list of values (one per line) or search

1 individual

- ✓ Gene Symbol is either KCNA1, CFC1, FOXH1 or GDF1 ×
- ✓ Effect Region is Exon ×
- ✓ Heterotaxy is True ×
- ✓ Structural Heart Disease is No ×
- ✓ Structural Heart Disease is No ×
- ✓ Variant Type is INDEL ×

View Results

https://rescows02.research.chop.edu/pcgc/query/

PCGC DataHub | Workspace | Query | Results | Genomic Data | Download Queue | Admin | Michael Italia

Search by name, description, or data.

Translator Used

MATERNAL DEMOGRAPHICS

- Maternal Structural Heart Disease
- Maternal Age at Proband Birth
- Maternal Age
- Maternal Race
- Maternal Race (Multiple Races)
- Maternal Race (Other)
- Hispanic or Latino Ethnicity (Maternal)
- Maternal Birth Country
- Maternal Education Level
- Maternal Education Level (UK)
- Maternal Education (Other)

PATERNAL DEMOGRAPHICS

- Paternal Structural Heart Disease
- Paternal Age at Proband Birth
- Paternal Age
- Paternal Race
- Paternal Race (Multiple Races)
- Paternal Race (Other)
- Hispanic or Latino Ethnicity (Paternal)
- Paternal Birth Country

Demographics > Paternal Demographics

Paternal Structural Heart Disease

Unique values 4

Exclude selected values

No	7047 (94%)
Form not entered	206 (2%)
Yes	172 (2%)
Not applicable	47 (<1%)

Update Filter

The PCGC Data Hub 3.0b6, powered by Cilantro 2.2.34-beta · Funded by the NHLBI's B2B Program
Developed, hosted, and maintained by The Children's Hospital of Philadelphia
Terms of Use · Privacy Policy

60 individuals

- Gene Symbol is either KCNA1, CFC1, FOXH1 or GDF1
- Effect Region is Exon
- Heterotaxy is True
- Structural Heart Disease is No
- Structural Heart Disease is No

View Results

Query:

Subjects with heterotaxy

No evidence of structural heart disease in either parent

Pathogenic variants in genes *KCNA1, CFC1, FOXH1, GDF1*

Bench to Bassinet

PCGC Data Hub successes:

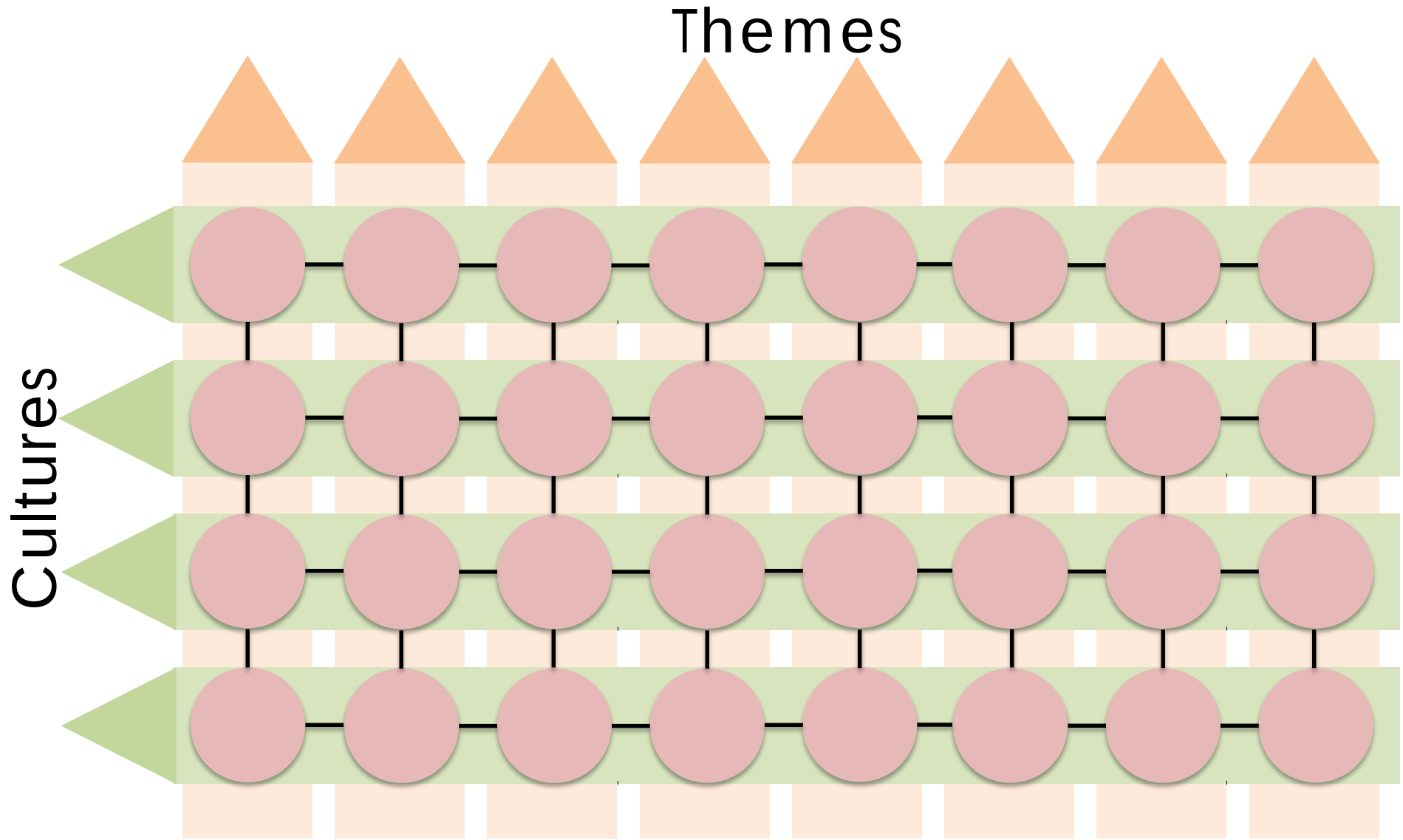
- Data management
- Access immediacy
- Traditional dissemination (inc GMKF)

Challenges:

- Volume
- Cost
- Distributed analysis
- Community empowerment

Challenge

Bench to Bassinet



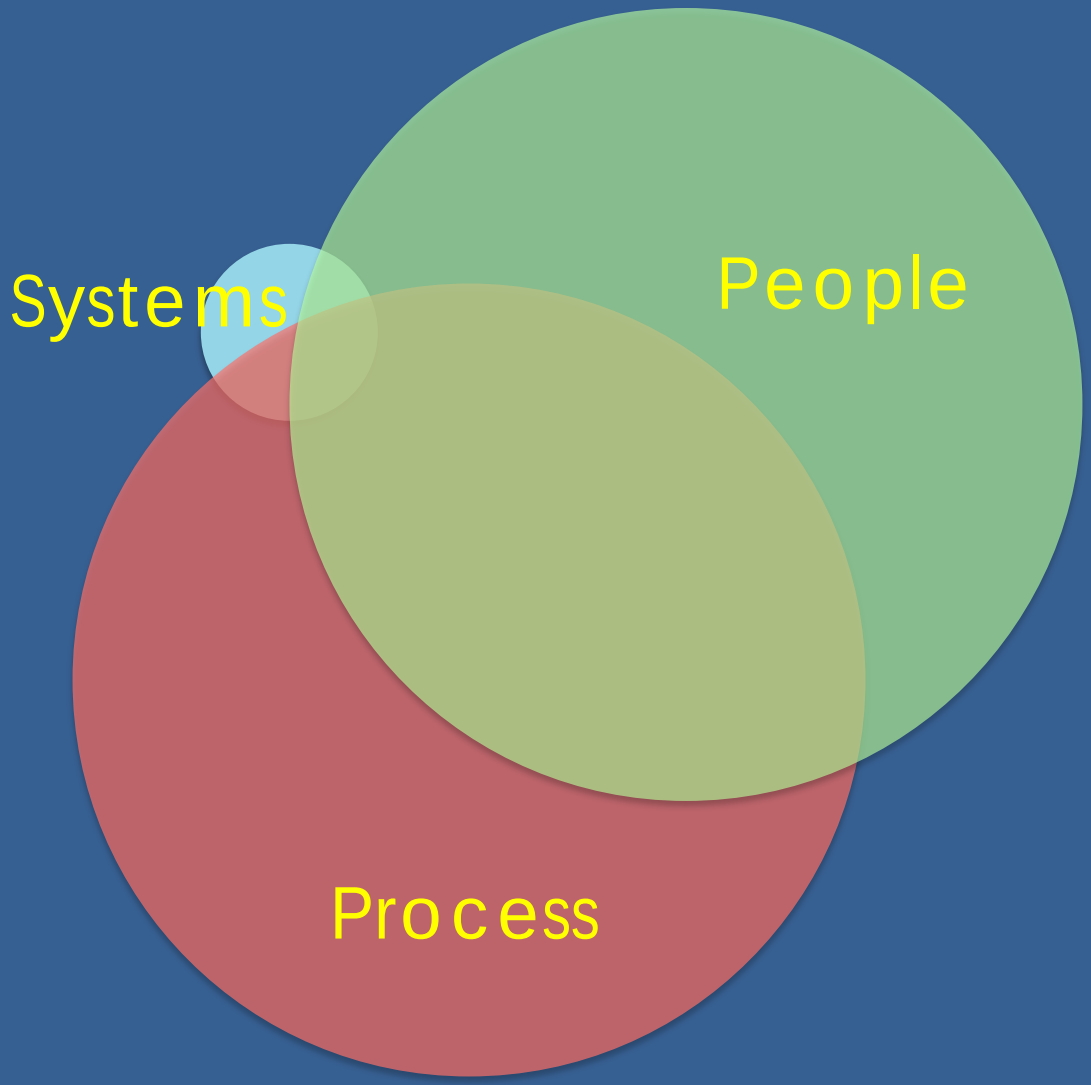
Landscape

Bench to Bassinet

Questions:

- Can I share?
- What should I share?
- How can I share?
- Why would I share?

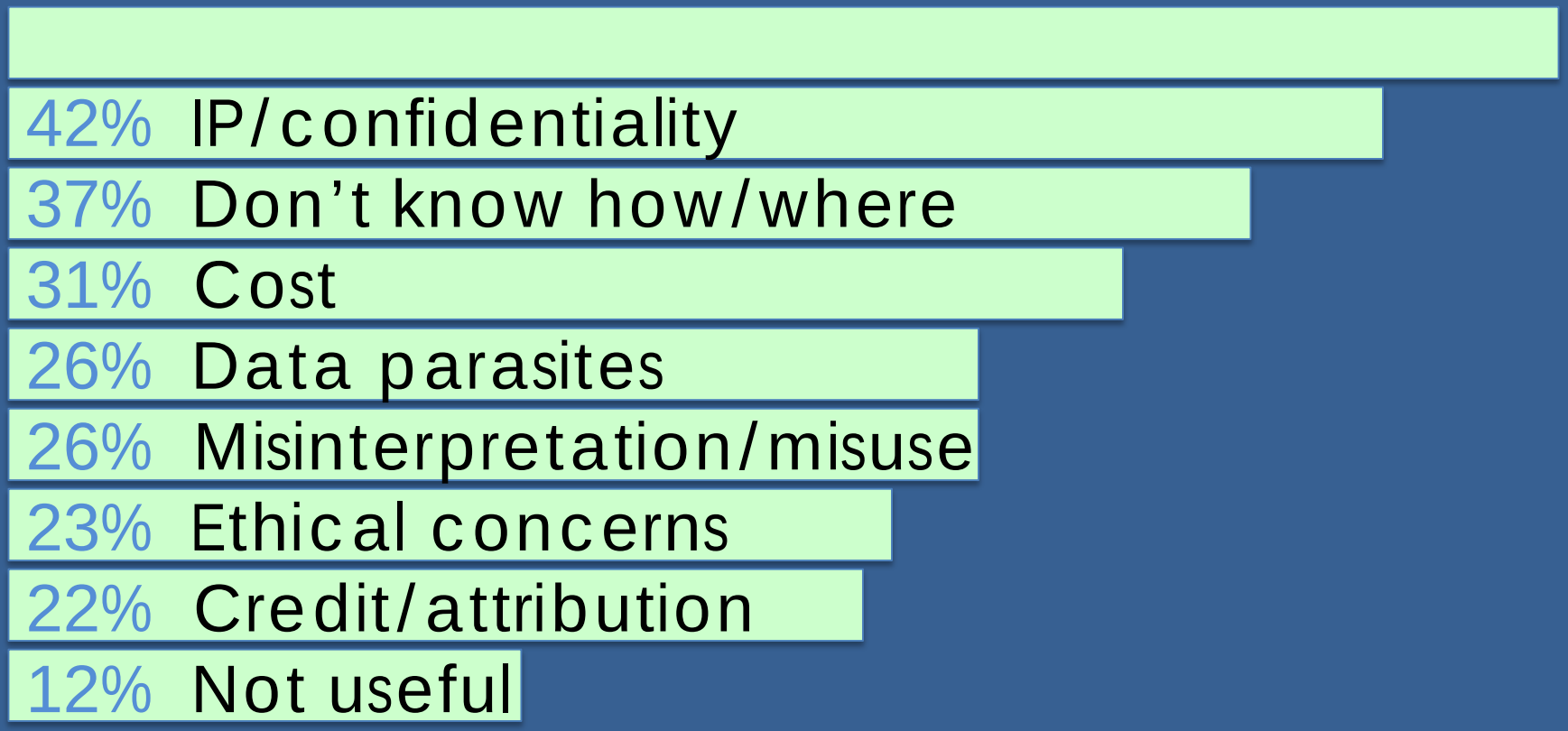
Data socialization



Uncertainty

Bench to Bassinet

Reasons given for not sharing data



Derived from: Ferguson L, Exchanges Blog, Wiley Publishing, 11-3-14

Behavioral economics

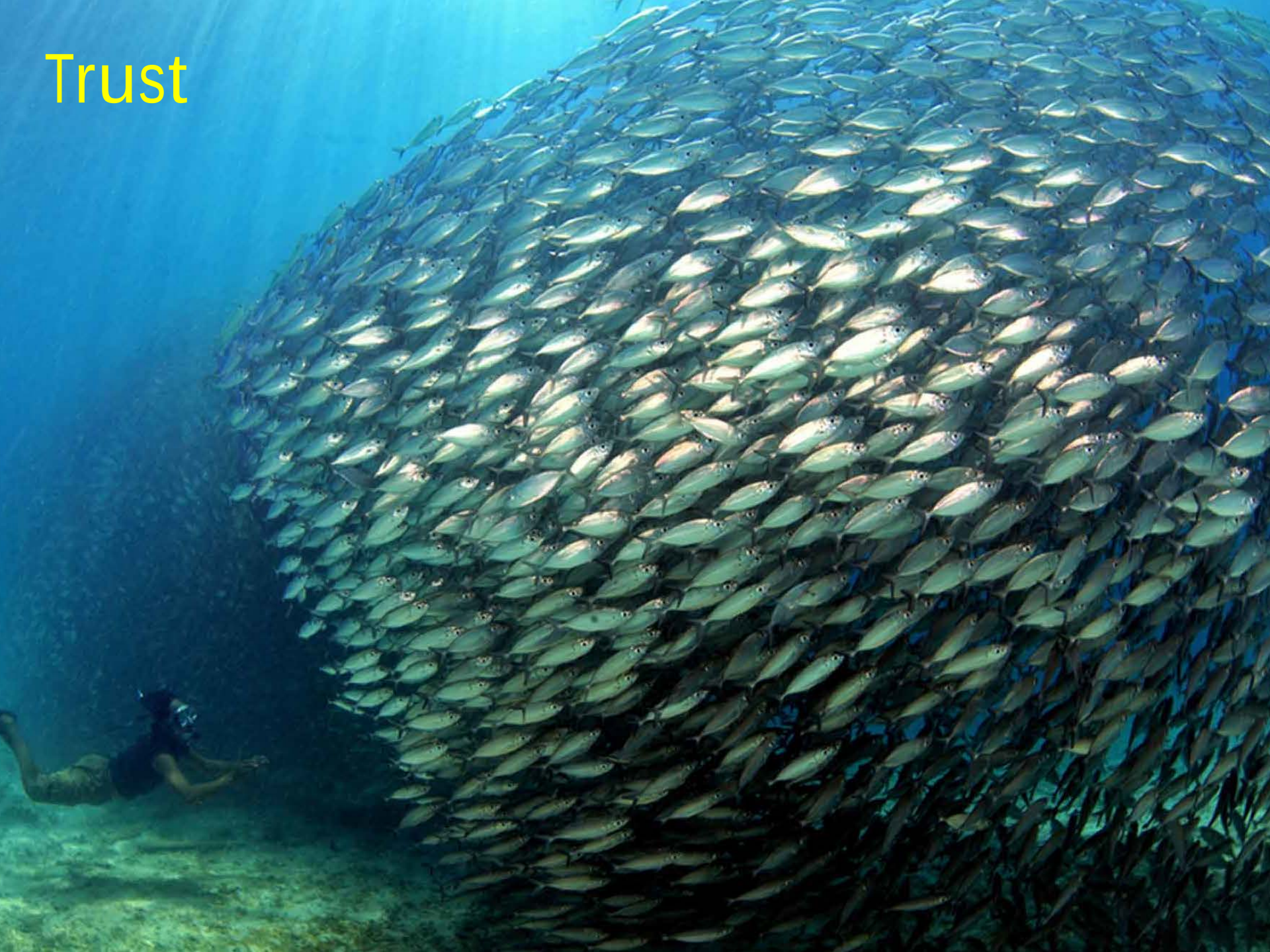
Bench to Bassinet

Vibrant data communities

- PubMed Central
- ExAC server
- TCGA
- Facebook
- Linux

- Well-defined roles/rules
- Simple roles/rules
- Actor-oriented
- Incentivized participation
- Product lead vs. operational excellence

Trust



B2B community

Bench to Bassinet

CCHMC ACC

- Rachel Akers
- Jim Cnota
- Stacy France
- Nathan Hawk
- Stephen Hope
- Bridget Kellner
- Eileen King (co-PI)
- Mojtaba Kohram
- Michal Kouril
- Michael Kuhlmann
- Guillaume Labilloy
- Chunyan Liu
- Lauren Lykowski
- Lisa Martin
- Diana McClendon
- Nick Ollberding
- Andy Rupert
- Ting Sa
- Nathan Salomonis
- Michael Wagner
- Alison Wheeler
- Pete White (co-PI)

Site PIs/co-PIs

- Daniel Bernstein
- Martina Brueckner
- Benoit Bruneau
- Neil Chi
- Sylvia Evans
- Bruce Gelb
- Richard Lifton
- Jane Newberger
- Len Pennachio
- Katie Pollard
- William Pu
- Amy Roberts
- Christine Seidman
- Jon Seidman
- Deepak Srivastava
- Mariin Tristani-Firouzi
- Axel Visel
- Joe Yost

NIH

- Jon Kaltman
- Melissa Parisi
- Gail Pearson
- Charlene Schramm

ClinGen, ClinVar, The Seqr Platform and the Matchmaker Exchange

Heidi L. Rehm, PhD, FACMG
Director, Partners HealthCare Laboratory for Molecular Medicine
Medical Director, Broad Institute Clinical Research Sequencing Platform
Associate Professor of Pathology, Brigham and Women's Hospital and
Harvard Medical School

PARTNERS HEALTHCARE | PERSONALIZED MEDICINE | BWH | HARVARD MEDICAL SCHOOL | BROAD INSTITUTE