GABRIELLA MILLER KIDS FIRST
PEDIATRIC RESEARCH PROGRAM

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

April 26, 2016
### 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

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Single</td>
<td>4</td>
<td>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.</td>
</tr>
<tr>
<td>2</td>
<td>Married filing jointly (even if only one had income)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Married filing separately. Enter spouse's SSN above and full name here.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Qualifying widow(er) with dependent child</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

### Exemptions

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6a</td>
<td>Yourself. If someone can claim you as a dependent, do not check box 6a.</td>
<td></td>
<td></td>
<td>Boxes checked on 6a and 6b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>Spouse</td>
<td></td>
<td></td>
<td>No. of children on 6c who:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>Dependents:</td>
<td></td>
<td></td>
<td>- lived with you</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) First name</td>
<td>(2) Dependent's social security number</td>
<td>(3) Dependent's relationship to you</td>
<td>(4) ✓ if child under age 17 qualifying for child tax credit (see instructions)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Last name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If more than four dependents, see instructions and check here ▶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>Total number of exemptions claimed</td>
<td></td>
<td></td>
<td>Add numbers on lines above ▶</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
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
DISCLOSURES

• None for this presentation
How Might Investigators Use the Kids First Data Resource?
Pediatric Cardiac Genomics Consortium

Boston Children’s/Brigham and Women’s - CHOP-Columbia-Mount Sinai-Yale
Gene with Multiple Mutations

- Synonymous: $p = 0.91$
- D–Mis: $p = 0.029$
- LoF: $p = 0.0065$
- Damaging: $p = 1.1 \times 10^{-5}$
WES STUDY #2, Science 2015 Findings
Synaptic, transcriptional and chromatin genes disrupted in autism

De Rubeis et al., November 2014
CHD GENETICS
What’s Missing?

• Exomic Mutations
  – Saturation
  – Other Modes (Recessive, Somatic)

• Non-Exomic
  – Isolated CHD = Cardiac Enhancer/Promoter Mutations?

• Epigenetics
• United Kingdom

• Recruitment
  – April 2011 – April 2015
  – 14,000 Patients with DDs

• Genomics
  – Arrays
  – Exome Sequencing
## DDD STUDY

Meta-Analysis

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

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Gene</th>
<th>De novo DDD (missense, LOF)</th>
<th>De novo meta (missense, LOF)</th>
<th>p value</th>
<th>Test</th>
<th>Mutation clustering</th>
<th>Predicted haploinsufficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>De novo enrichment</td>
<td>COL4A3BP</td>
<td>3 (3.0)</td>
<td>5 (5.0)</td>
<td>4.10 × 10^{-12}</td>
<td>Meta</td>
<td>Yes</td>
<td>14.7</td>
</tr>
<tr>
<td></td>
<td>PPP2R5D</td>
<td>4 (4.0)</td>
<td>5 (5.0)</td>
<td>6.01 × 10^{-12}</td>
<td>DDD</td>
<td>Yes</td>
<td>19.7</td>
</tr>
<tr>
<td></td>
<td>ADNP</td>
<td>4 (0.4)</td>
<td>5 (0.5)</td>
<td>4.59 × 10^{-11}</td>
<td>Meta</td>
<td>No</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td>POGZ</td>
<td>2 (0.2)</td>
<td>5 (0.5)</td>
<td>4.31 × 10^{-10}</td>
<td>Meta</td>
<td>No</td>
<td>30.0</td>
</tr>
<tr>
<td></td>
<td>PPP2R1A</td>
<td>3 (3.0)</td>
<td>3 (3.0)</td>
<td>2.03 × 10^{-8}</td>
<td>DDD</td>
<td>Yes</td>
<td>23.5</td>
</tr>
<tr>
<td></td>
<td>DDX3X</td>
<td>4 (3.1)</td>
<td>5 (3.2)</td>
<td>2.26 × 10^{-7}</td>
<td>DDD</td>
<td>No</td>
<td>12.7</td>
</tr>
<tr>
<td></td>
<td>CHAMP1</td>
<td>2 (0.2)</td>
<td>3 (0.3)</td>
<td>4.58 × 10^{-7}</td>
<td>Meta</td>
<td>No</td>
<td>52.9</td>
</tr>
<tr>
<td></td>
<td>BCL11A</td>
<td>3 (3.0)</td>
<td>4 (3.1)</td>
<td>1.03 × 10^{-6}</td>
<td>DDD</td>
<td>Yes</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>PURA</td>
<td>3 (1.2)</td>
<td>3 (1.2)</td>
<td>1.14 × 10^{-6}</td>
<td>DDD</td>
<td>No</td>
<td>9.4</td>
</tr>
<tr>
<td>De novo enrichment plus additional evidence</td>
<td>DNM1</td>
<td>3 (3.0)</td>
<td>5 (5.0)</td>
<td>1.43 × 10^{-6}</td>
<td>Meta</td>
<td>No</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td>TRIO</td>
<td>2 (2.0)</td>
<td>7 (7.0)</td>
<td>5.16 × 10^{-6}</td>
<td>Meta</td>
<td>Yes</td>
<td>25.7</td>
</tr>
<tr>
<td></td>
<td>PCGF2</td>
<td>2 (2.0)</td>
<td>2 (2.0)</td>
<td>1.08 × 10^{-5}</td>
<td>DDD</td>
<td>Yes</td>
<td>37.7</td>
</tr>
</tbody>
</table>
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
• 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
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
<table>
<thead>
<tr>
<th>Early Family Studies</th>
<th>Current Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Li-Fraumeni Syndrome</strong></td>
<td>Families with multiple cancers in children, young adults</td>
</tr>
<tr>
<td><strong>Wilms’ Tumor, Aniridia, Congenital Malformations</strong></td>
<td>Multicenter study of childhood cancer WAGR syndrome</td>
</tr>
<tr>
<td><strong>Familial Chronic Lymphocytic Leukemia</strong></td>
<td>Family and population studies</td>
</tr>
<tr>
<td><strong>Myotonic Dystrophy (Dystrophica Myotonica, DM)</strong></td>
<td>Leukemia developed in patient with DM</td>
</tr>
</tbody>
</table>
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

Participant accrual
- Participant calls the referral nurse
- Study team reviews

Questionnaires
- Family History
- Individual History

Field Cohort
- Medical record review
- Genetic counseling & testing
- Biospecimens

Clinic Cohort
- Evaluation at NIH Clinical Center
- Subspecialists
- Biospecimens
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.

Patients with these syndromes are of interest to the NCI because they 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.

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

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

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

https://www.marrowfailure.cancer.gov/
Pleuropulmonary Blastoma *DICER1* Syndrome 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

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

DICER1 Mutations in Familial Pleuropulmonary Blastoma

D. Ashley Hill,1,2 Jennifer Ivanovich,1 John R. Priest,2 Christina A. Garnett,1 Louis P. Dehner,1 David Desruisseau,1 Jason A. Jarzembski,3 Kathryn A. Wikenheiser-Brodkamp,4 Brian K. Suarez,1 Alison J. Whelan,1 Gretchen Williams,2,3 Dawn Bracamontes,1,2 Yoav Messinger,2,5 Paul J. Goodfellow1

Science (2009) 325; 965

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

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

https://ppb.cancer.gov/
## NCI *DICER1* Syndrome Cohort

<table>
<thead>
<tr>
<th>Accrual</th>
<th>Clinical Center Cohort</th>
<th>Total Enrolled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Families</td>
<td>37</td>
<td>82</td>
</tr>
<tr>
<td><em>DICER1</em> mutation carriers</td>
<td>99</td>
<td>173</td>
</tr>
<tr>
<td><em>DICER1</em>-negative controls</td>
<td>69</td>
<td>220</td>
</tr>
<tr>
<td>Not yet tested</td>
<td>0</td>
<td>67</td>
</tr>
<tr>
<td><strong>Total individuals</strong></td>
<td><strong>168</strong></td>
<td><strong>460</strong></td>
</tr>
</tbody>
</table>

[https://ppb.cancer.gov/](https://ppb.cancer.gov/)
Li-Fraumeni Syndrome Study

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.

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.
Li-Fraumeni Syndrome

- Autosomal dominant inheritance
- Multiple different cancers
- TP53 accounts for about 70%
- Early onset cancers
- Child onset cancer
- By age 30 ~50% of women and ~30% of men
- By age 60 nearly everyone
- More than half develop multiple cancers
159 families enrolled

- 56 LFS
  - 53 (94%) TP53+
  - 2 (4%) TP53-
  - 1 (2%) VUS

- 77 LFL
  - 51 (66%) TP53+
  - 24 (31%) TP53-
  - 2 (3%) VUS
  - 10 (83%) TP53+
  - 2 (17%) TP53-

- 12 multiple primary cancers

- 14 mutation carriers, not meeting FHx criteria

http://lfs.cancer.gov
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
  - 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.
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.
• 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

- Clinical Sites
- Demonstration Projects
- Technology Development
  - Sample Collection
  - Data Generation
  - Data Analysis
  - WGS
  - 16S
  - Reference genomes
- Ethical, Legal and Social Issues
- Sequencing Centers
  - Filter trimming
  - Chimera removal
  - Taxonomic classification
  - Clustering into OTU’s
- Data Analysis and Coordination Center
- Computational Tools
  - Dissemination & Publication
- Data Generation & Analysis
  - 16S rRNA genes
300 Subjects Enrolled

Sampling 15/18 body sites
DNA extraction
High throughput sequencing

Shotgun metagenomic data:

Human filtering
Quality trimming

QC

DigiNorm
k-mer reduction

IDBA-UD
Assemble

Bowtie
Reference Alignments

MetaPhlAn2
Community Profile

Contig non-redundification

MetaGeneMar
k Gene Calling

Clustering non-redundification

Community Profiles

Organismal census

16S rRNA genes

Filtering/trimming
Chimera removal
Taxonomic classification
Clustering into OTU’s

Organismal census

Metabolic Reconstruction
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:

- \texttt{-id 0.95}
- \texttt{-targetfract 0.90}

5 Implementation

6 Discussion

7 Related Documents & References


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

Features:
- Common sample provenance
- Common data provenance
- Common protocol design
- Centralized infrastructure
- Centralized analysis infrastructure
Results: 1 to 20 of 746

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

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

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

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

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

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

7. Human Feces
   7. ILLUMINA (Illumina HiSeq 2000) run: 27.2M spots, 2.7G bases, 1.9G downloads
      Accession: SRX322985
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
## Metadata assessment across all demonstration projects: Thanks: Steve Sherry et al at dbGaP

**IHMC Variable**

<table>
<thead>
<tr>
<th>IHMC Variable</th>
<th>SUBJID</th>
<th>SUBJID</th>
<th>SUBJID</th>
<th>SUBJID</th>
<th>SUBJID</th>
<th>SUBJID</th>
<th>SUBJID</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight_kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight_Lbs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start_date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finish_date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug_name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHMC Code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Metadata assessment

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Assessment</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Other Race</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>Smoking duration</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Drugs</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Weight_kg</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Weight_Lbs</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Dose</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Start_date</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Finish_date</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Drug_name</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

**Insert IHMC code stuff.**
Insert IHMC code stuff.

<table>
<thead>
<tr>
<th>DBGaP</th>
<th>SUB JID</th>
<th>SUB JID</th>
<th>SUB JID</th>
<th>Subject ID</th>
<th>SUB JID</th>
<th>Patient ID</th>
<th>Subject ID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gender</td>
<td>gender</td>
<td>gender</td>
<td>gender</td>
<td>gender</td>
<td>gender</td>
<td>gender</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>age</td>
<td>age</td>
<td>age</td>
<td>age</td>
<td>age</td>
<td>age</td>
</tr>
<tr>
<td></td>
<td>ethnicity</td>
<td>ethnicity</td>
<td>ethnicity</td>
<td>ethnicity</td>
<td>ethnicity</td>
<td>ethnicity</td>
<td>ethnicity</td>
</tr>
<tr>
<td></td>
<td>race</td>
<td>race</td>
<td>race</td>
<td>race</td>
<td>race</td>
<td>race</td>
<td>race</td>
</tr>
<tr>
<td></td>
<td>smoking ever</td>
<td>smoking ever</td>
<td>smoking ever</td>
<td>smoking ever</td>
<td>smoking ever</td>
<td>smoking ever</td>
<td>smoking ever</td>
</tr>
<tr>
<td></td>
<td>nugent score</td>
<td>nugent score</td>
<td>nugent score</td>
<td>nugent score</td>
<td>nugent score</td>
<td>nugent score</td>
<td>nugent score</td>
</tr>
<tr>
<td></td>
<td>smoke ever</td>
<td>smoke ever</td>
<td>smoke ever</td>
<td>smoke ever</td>
<td>smoke ever</td>
<td>smoke ever</td>
<td>smoke ever</td>
</tr>
<tr>
<td></td>
<td>treatment</td>
<td>treatment</td>
<td>treatment</td>
<td>treatment</td>
<td>treatment</td>
<td>treatment</td>
<td>treatment</td>
</tr>
<tr>
<td></td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
</tr>
<tr>
<td></td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
<td>weight</td>
</tr>
<tr>
<td></td>
<td>v vital status</td>
<td>v vital status</td>
<td>v vital status</td>
<td>v vital status</td>
<td>v vital status</td>
<td>v vital status</td>
<td>v vital status</td>
</tr>
<tr>
<td></td>
<td>age starts smoking</td>
<td>age starts smoking</td>
<td>age starts smoking</td>
<td>age starts smoking</td>
<td>age starts smoking</td>
<td>age starts smoking</td>
<td>age starts smoking</td>
</tr>
<tr>
<td></td>
<td>Tob year</td>
<td>Tob year</td>
<td>Tob year</td>
<td>Tob year</td>
<td>Tob year</td>
<td>Tob year</td>
<td>Tob year</td>
</tr>
<tr>
<td></td>
<td>PASI</td>
<td>PASI</td>
<td>PASI</td>
<td>PASI</td>
<td>PASI</td>
<td>PASI</td>
<td>PASI</td>
</tr>
<tr>
<td></td>
<td>ABX</td>
<td>ABX</td>
<td>ABX</td>
<td>ABX</td>
<td>ABX</td>
<td>ABX</td>
<td>ABX</td>
</tr>
<tr>
<td></td>
<td>FEET, INCHES</td>
<td>FEET, INCHES</td>
<td>FEET, INCHES</td>
<td>FEET, INCHES</td>
<td>FEET, INCHES</td>
<td>FEET, INCHES</td>
<td>FEET, INCHES</td>
</tr>
<tr>
<td></td>
<td>education</td>
<td>education</td>
<td>education</td>
<td>education</td>
<td>education</td>
<td>education</td>
<td>education</td>
</tr>
<tr>
<td></td>
<td>family hist</td>
<td>family hist</td>
<td>family hist</td>
<td>family hist</td>
<td>family hist</td>
<td>family hist</td>
<td>family hist</td>
</tr>
<tr>
<td></td>
<td>body site</td>
<td>body site</td>
<td>body site</td>
<td>body site</td>
<td>body site</td>
<td>body site</td>
<td>body site</td>
</tr>
<tr>
<td></td>
<td>duration</td>
<td>duration</td>
<td>duration</td>
<td>duration</td>
<td>duration</td>
<td>duration</td>
<td>duration</td>
</tr>
</tbody>
</table>
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
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.

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.

- 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

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.
What is IRBshare?
The IRBshare System is a new shared IRB review model for multi-site studies comprised of Participating Institutions utilizing Shared Review Documents and a Shared Review Process, supported by a centralized, secure web portal and the IRBshare Master Agreement ("IMA"). IRBshare aims to optimize human research protections through building trust, enhancing communication, and sharing best practices between IRBs nationwide.

Have you heard? We have:

- 26 participating institutions
- 4 studies in pilot phase

IRBshare Pilot Phase
In October 2012, IRBshare launched a Pilot Phase to implement and evaluate the shared review model. The following institutions and studies are eligible to join the pilot phase:

- Any IRB with an active FWA with OHRP
- Any federally-funded study
- Studies that will begin or are in the initial study review process

Clinical and Translational Science Award Consortium Coordinating Center.
What is IRBshare?

The IRBshare System is a multi-site study coordinating center for IRB review. Participating Institutions (PIs) share a common, secure web portal to communicate, approve, and manage IRB documents and review sites. The IRBshare System is an initiative of the Clinical and Translational Science Award (CTSA) Program of the National Institutes of Health (NIH). It facilitates the review of clinical trials and other research involving human subjects across multiple institutions, making it easier to conduct research efficiently and ethically.

- **Full Board Review Sites**: Sites that conduct full board-level reviews of research protocols.
- **Shared Review Sites**: Sites that participate in the shared review process, allowing for expedited review and approval.

The IRBshare System promotes collaboration among institutions, enabling streamlined communication and efficient review of research protocols.
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, recruit

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

Absolutely will not improve standards

Absolutely will not solve sharing
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
Data Sharing: The PCGC Experience

Gabriella Miller Kids First Data Resource Workshop
National Institutes of Health
April 26, 2016
Cardiac Genetics

Bench to Bassinet

Cardiac Development

UCSF
Yale
Mt Sinai
BCH
Utah
Harvard
Utah
UCSD
Gladstone
24770 subjects
9702 genomes
140 TB of data
6766 elements
Data Resource

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.

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.

Support

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

Open Source & Available on GitHub
We believe in open source software and open-source our work. Harvest is licensed under BSD 2-clause License.

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.

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

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
Diagnoses > B2B (Fyler) Diagnoses

Heterotaxy

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

Unique values: 2

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

Update Filter
### Functional Class

| Unique values | 3 |

### Predicted Variant Effects

- Intergenic
- Upstream
- UTR-5 Prime
- Slice-Site Acceptor
- Slice-Site Donor
- Exon
- Introns
- UTR-3 Prime
- Downstream
- Regulation

- Gene Symbol is either KCNA1, CFC1, FOXH1 or GDF1
- Effect Region is Exon
- Heterolaxy is True
- Structural Heart Disease is No
- Structural Heart Disease is No
- Variant Type is INDEL
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
Questions:

• Can I share?
• What should I share?
• How can I share?
• Why would I share?
Data socialization

Systems

People

Process
Uncertainty

Reasons given for not sharing data

42% IP/confidentiality
37% Don’t know how/where
31% Cost
26% Data parasites
26% Misinterpretation/misuse
23% Ethical concerns
22% Credit/attribution
12% Not useful

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

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
<table>
<thead>
<tr>
<th>B2B community</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CCHMC ACC</strong></td>
</tr>
<tr>
<td>- Rachel Akers</td>
</tr>
<tr>
<td>- Jim Cnota</td>
</tr>
<tr>
<td>- Stacy France</td>
</tr>
<tr>
<td>- Nathan Hawk</td>
</tr>
<tr>
<td>- Stephen Hope</td>
</tr>
<tr>
<td>- Bridget Kellner</td>
</tr>
<tr>
<td>- <strong>Eileen King (co-PI)</strong></td>
</tr>
<tr>
<td>- Mojtaba Kohram</td>
</tr>
<tr>
<td>- Michal Kouril</td>
</tr>
<tr>
<td>- Michael Kuhlmann</td>
</tr>
<tr>
<td>- Guillaume Labilloy</td>
</tr>
<tr>
<td>- Chunyan Liu</td>
</tr>
<tr>
<td>- Lauren Lykowski</td>
</tr>
<tr>
<td>- Lisa Martin</td>
</tr>
<tr>
<td>- Diana McClendon</td>
</tr>
<tr>
<td>- Nick Ollberding</td>
</tr>
<tr>
<td>- Andy Rupert</td>
</tr>
<tr>
<td>- Ting Sa</td>
</tr>
<tr>
<td>- Nathan Salomonis</td>
</tr>
<tr>
<td>- <strong>Michael Wagner</strong></td>
</tr>
<tr>
<td>- Alison Wheeler</td>
</tr>
<tr>
<td>- Pete White (co-PI)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Site PIs/co-PIs</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Daniel Bemstein</td>
</tr>
<tr>
<td>- Martina Brueckner</td>
</tr>
<tr>
<td>- Benoit Bruneau</td>
</tr>
<tr>
<td>- Neil Chi</td>
</tr>
<tr>
<td>- Sylvia Evans</td>
</tr>
<tr>
<td>- Bruce Gelb</td>
</tr>
<tr>
<td>- Richard Lifton</td>
</tr>
<tr>
<td>- Jane Newberger</td>
</tr>
<tr>
<td>- Len Pennachio</td>
</tr>
<tr>
<td>- Katie Pollard</td>
</tr>
<tr>
<td>- William Pu</td>
</tr>
<tr>
<td>- Amy Roberts</td>
</tr>
<tr>
<td>- Christine Seidman</td>
</tr>
<tr>
<td>- Jon Seidman</td>
</tr>
<tr>
<td>- Deepak Srivastava</td>
</tr>
<tr>
<td>- Marin Tristani-Firouzi</td>
</tr>
<tr>
<td>- Axel Visel</td>
</tr>
<tr>
<td>- Joe Yost</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>NIH</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Jon Kaltman</td>
</tr>
<tr>
<td>- Melissa Parisi</td>
</tr>
<tr>
<td>- Gail Pearson</td>
</tr>
<tr>
<td>- Charlene Schramm</td>
</tr>
</tbody>
</table>