

Tracking Utility of Common Fund Data Sets

Stephanie Courchesne-Schlink, PhD
Team Leader for Policy, Planning,
Evaluation, and Communication
Office of Strategic Coordination



National Institutes of Health
Office of Strategic Coordination - The Common Fund

Tracking Use of Data Sets



Sustainability Plans

Where does the data go when Common Fund support ends?

Data sets are proving sufficiently useful that other entities are taking them up

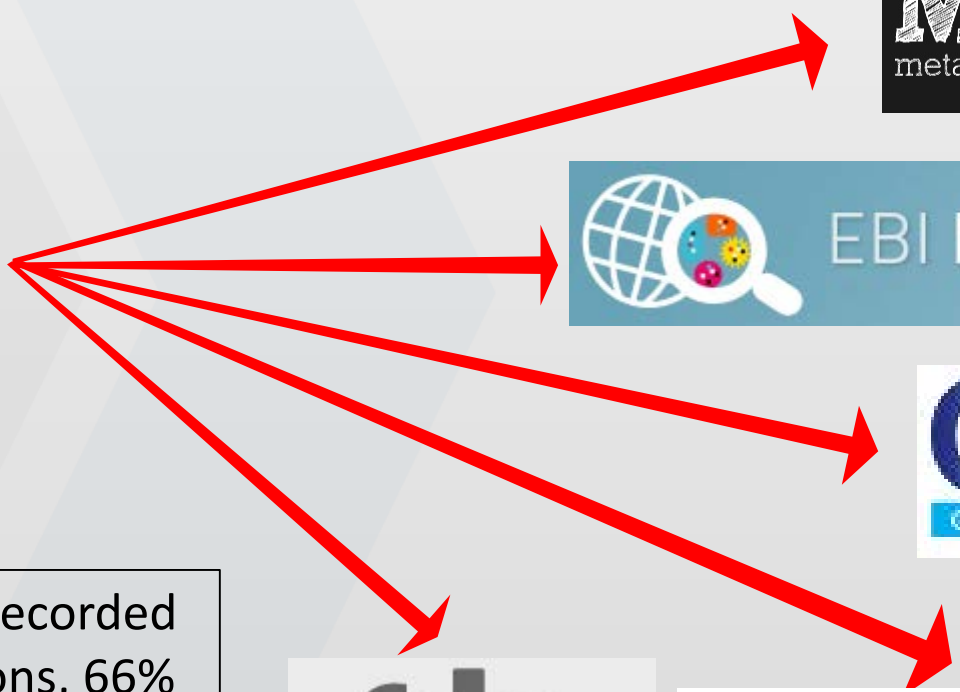
- Housed at NCBI (NIH)
- NIH ICs or other entities take over database
- Data uploaded to cloud server (NIH Data Commons)



HMP data sets: Centralized at the DACC and incorporated into other databases



HMP-DACC



Ribosomal Database Project



In FY16, the DACC recorded 121,000 user sessions, 66% of whom were new users. (20% of these sessions had dwell times of 10-30 mins)

Tracking Research Outputs from HMP Data Sets

Create a library of papers that mention HMP or that cite the main HMP consortium publications

Remove papers that acknowledge funding from HMP grants

Manually curate remaining papers to determine which used the HMP datasets in their study

(107 papers as of Oct 2016)

ARTICLE

PUBLISHED ONLINE: 17 NOVEMBER 2013 | DOI: 10.1038/NCHEMBIO.1386

nature
chemical biology

A widespread self-cleaving ribozyme class is revealed by bioinformatics

Adam Roth^{1,2,5}, Zasha Weinberg^{1,2,5}, Andy G Y Chen¹, Pete Ronald R Breaker^{1-3*}

Clinical Infectious Diseases

MAJOR ARTICLE



The Journal of Urology
Volume 195, Issue 2, February 2016, Pages

Investigative Urology

The Presence of *Oxalobacter formigenes* in the Microbiome of Healthy Young Adults

Clea Barnett*, Lama Nazzari*, David S. Goldfarb, Martin J. Blaser

Show more



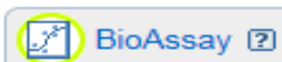
Fecal Microbial Transplants Reduce Antibiotic-resistant Genes in Patients With Recurrent *Clostridium difficile* Infection

Braden Millan,^{1,*} Heekuk Park,^{1,*} Naomi Hotte,¹ Olivier Mathieu,² Pierre Burguiere,² Thomas A. Tompkins,² Dina Kao,^{1,b} and Karen L. Madsen^{1,b}

¹Department of Medicine, University of Alberta, Edmonton, and ²Lallemand Health Solutions, Montreal, Quebec, Canada

PubChem

A component of the Molecular Libraries program



Today's Statistics >

Compounds:	92,877,506
Substances:	225,657,975
BioAssays:	1,218,736
Tested Compounds:	2,283,551
Tested Substances:	3,597,363
RNAi BioAssays:	95
BioActivities:	231,328,738
Protein Targets:	10,182
Gene Targets:	20,208

Dec. 19, 2016

- Averages between 2 and 10 million page views per day
- 1.5 million unique interactive users per month
- Integrated into key scientific web-based resource sites (PubMed, MeSH, ChemIDplus, FDA, NIST, USTPO, DrugBank, GSK, IBM, Nature, Sigma-Aldrich, etc.)



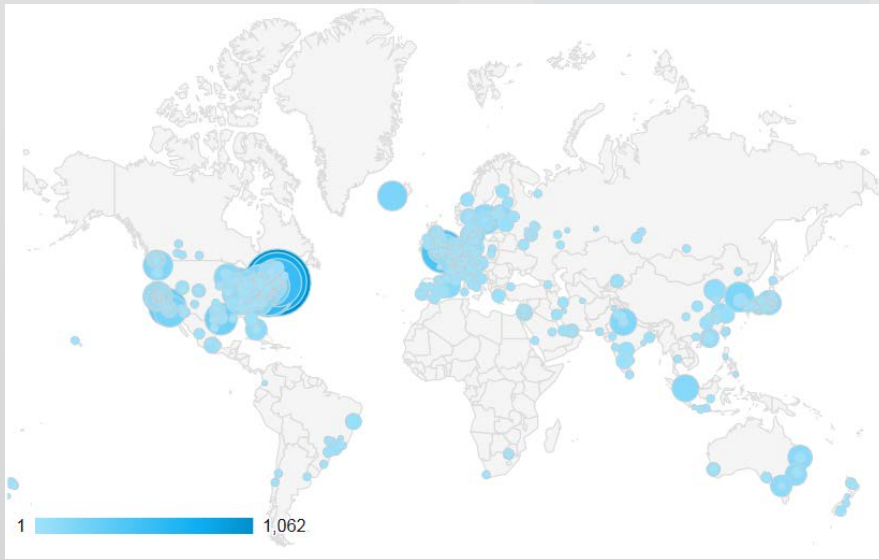
Genotype-Tissue Expression

Data sets to study relationship between genetic variation and gene expression

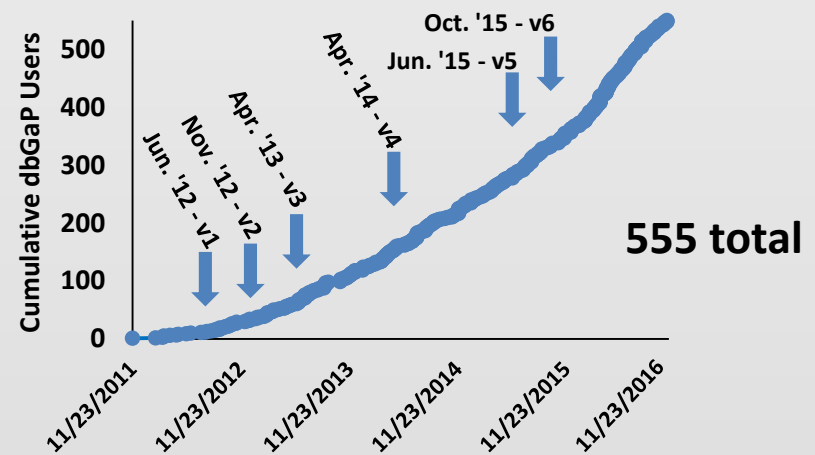
GTEx Portal Users

Registered users	Unique users	# different countries
9,210	48,454	>90

~8000 unique visitors/month; ~53% academic & ~25% Biotech/Pharma

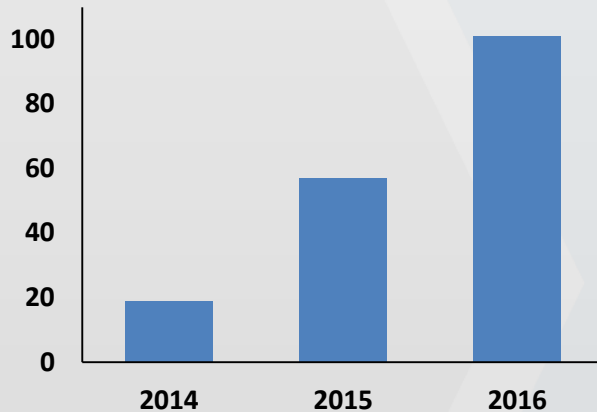


dbGaP users



Use of GTEx Data

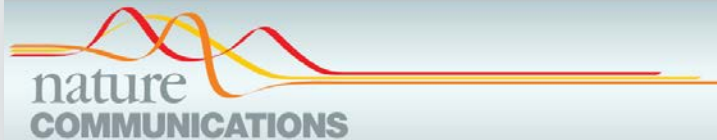
Publications using GTEx Data



177 total

Cell

Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity



Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility

Blekhman *et al. Genome Biology* (2015) 16:191
DOI 10.1186/s13059-015-0759-1



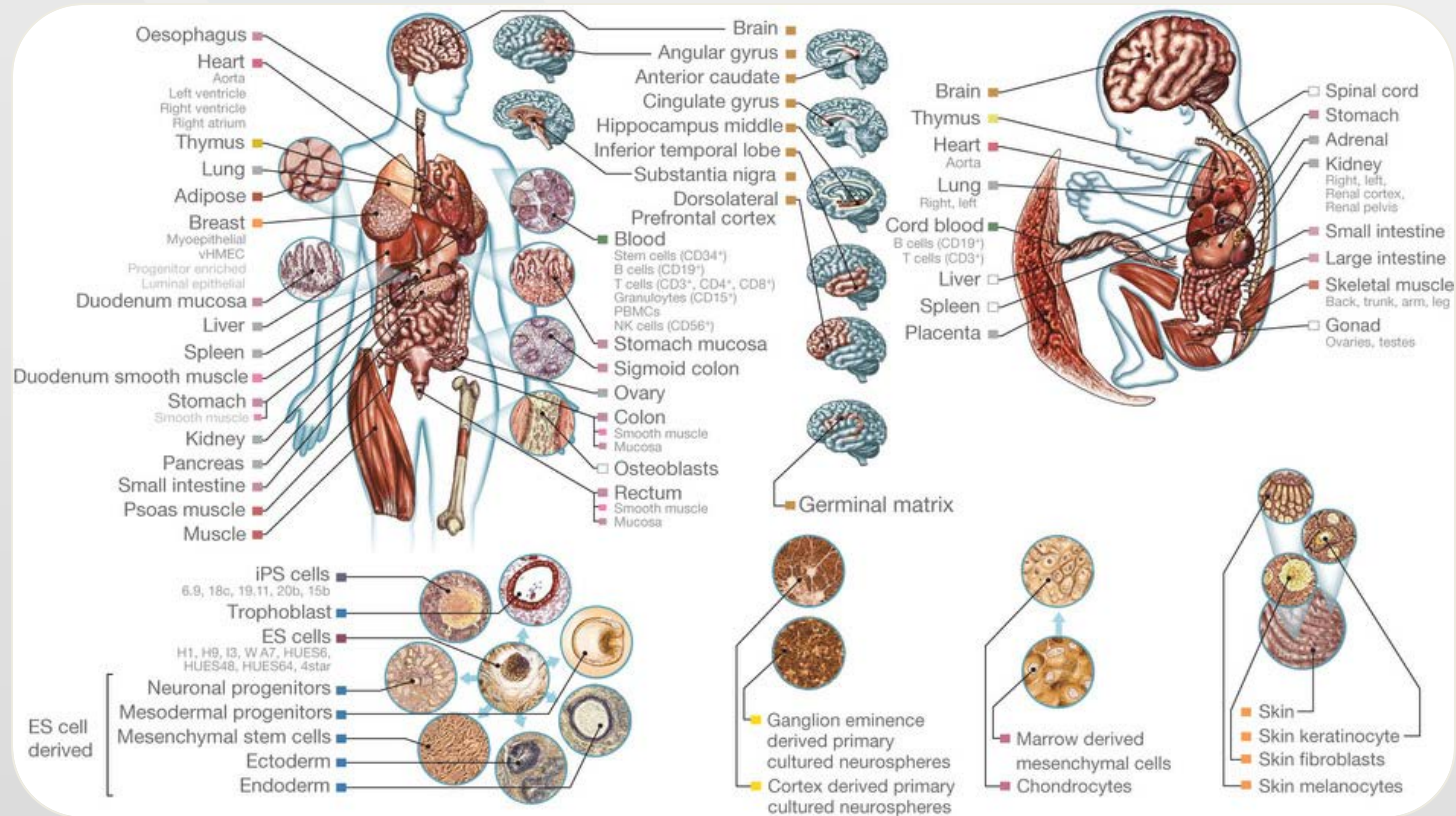
RESEARCH

Open Access

Host genetic variation impacts microbiome composition across human body sites



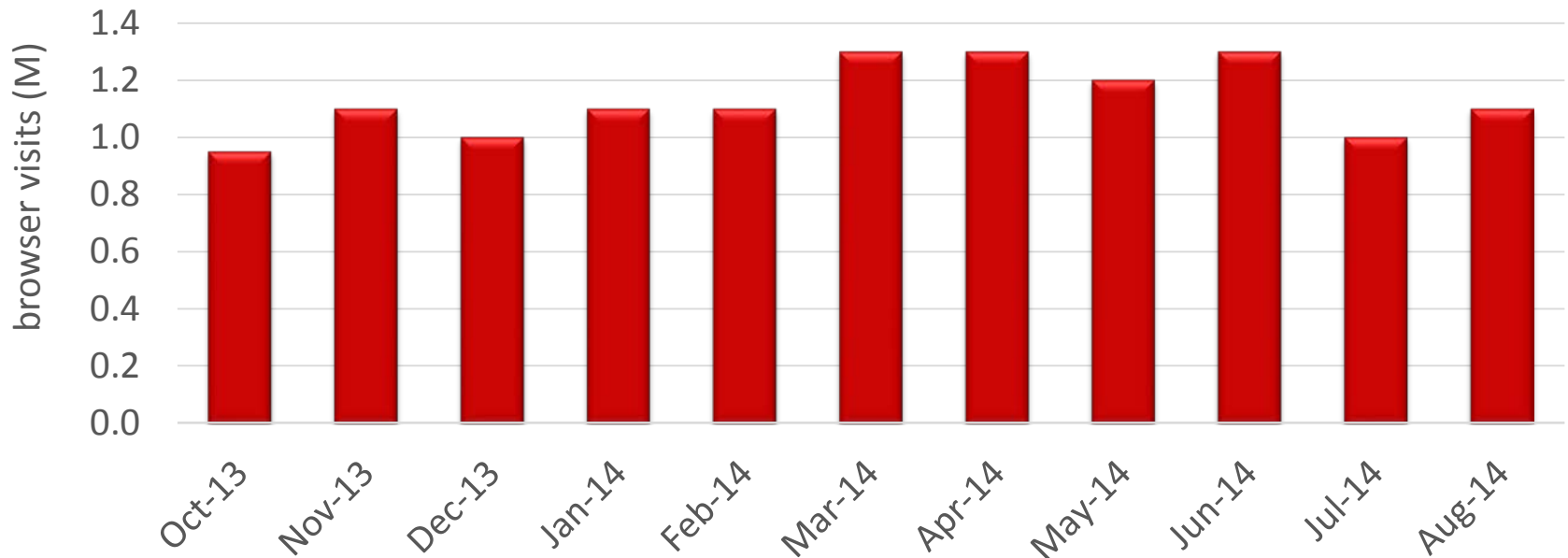
Roadmap Epigenomics Program



Developed 111 reference epigenomes from a variety of human cells and tissues

Reference Epigenomes Data Usage

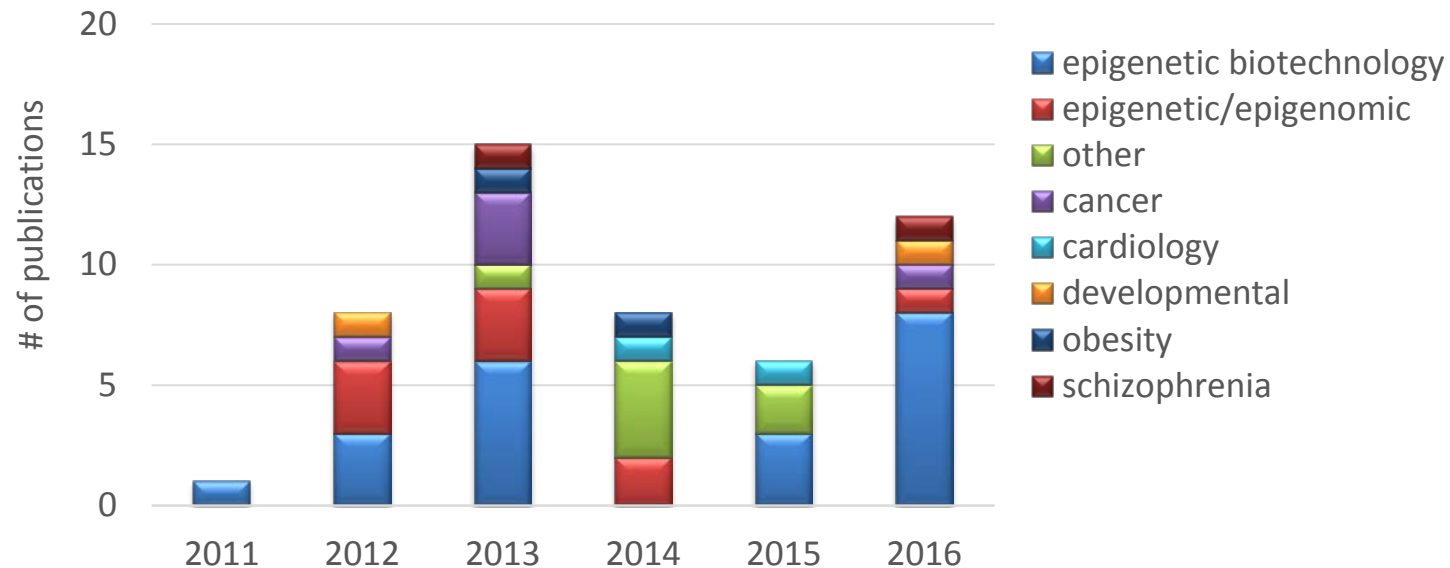
Browser Visits



- Approximately 12M visitors in 1 year alone; data combined across multiple CF-supported browsers
- Data now accessed mostly through ENCODE and IHEC

Using Reference Epigenome Maps

Research Areas of Publications
Using Roadmap Epigenomics Data



RESEARCH



Genome-wide signatures of differential methylation in pediatric acute lymphoblastic leukemia

RESEARCH ARTICLE

Disproportionate Contributions of Select Genomic Compartments and Cell Types to Genetic Risk for Coronary Artery Disease

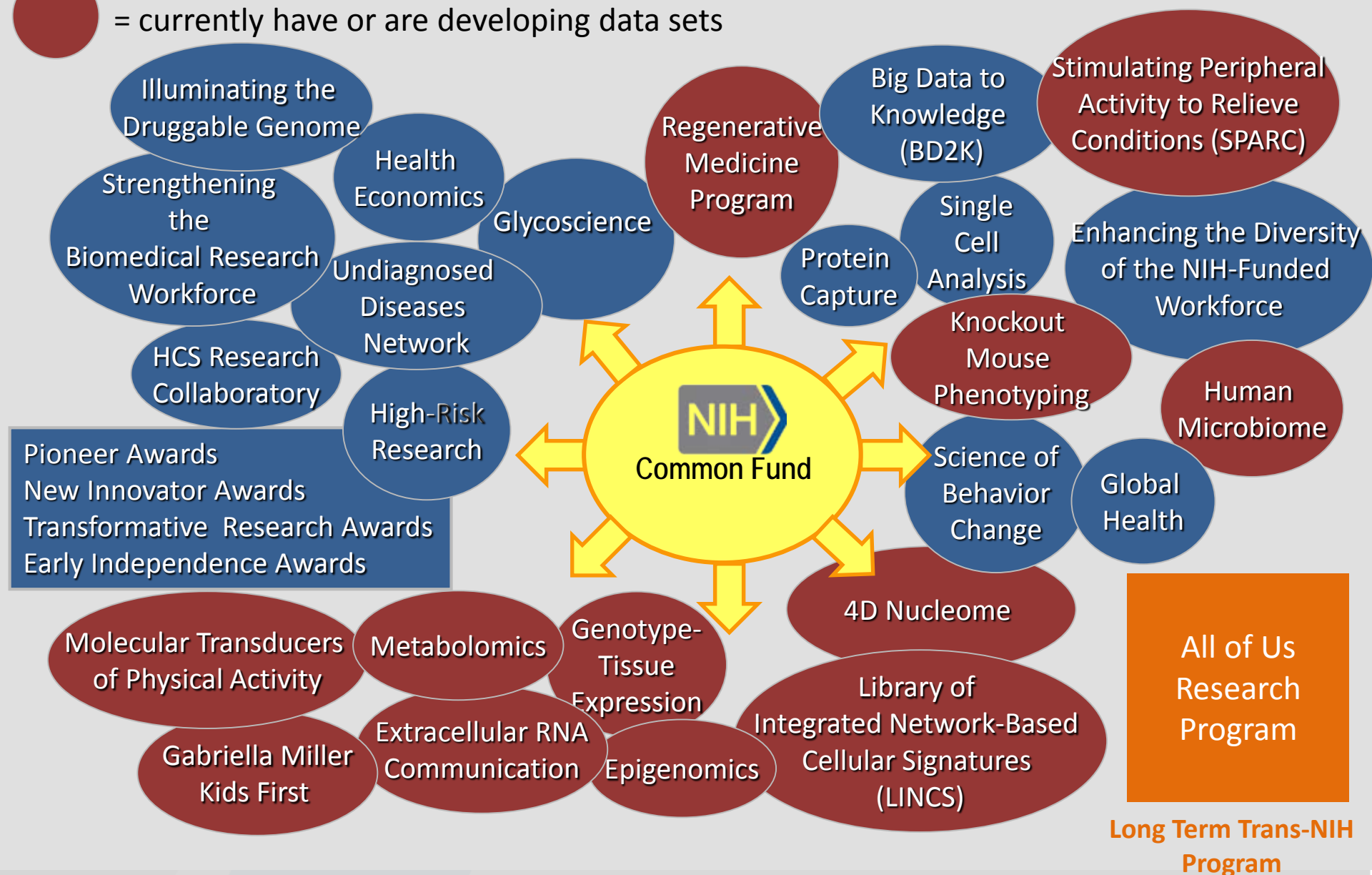


Thank you!

Questions?

Current Common Fund Programs (FY17)

 = currently have or are developing data sets



Long Term Trans-NIH Program