Open science, public-private partnerships, and the NIH

David Wholley,
Senior Vice-President, Research Partnerships
Foundation for the National Institutes of Health (FNIH)

BIO International Convention

25 YEARS OF Bio INNOVATION

NIH rules support open science

NIH requires broad and responsible data sharing:

- **Genomic Data Sharing Policy**: data to be shared via NIH repositories "no later than 6 months after initial data sharing begins, or at time of acceptance of first publication, whichever comes first"

- The **21st Century Cures Act** adds authority to require data sharing on NIH grants
  - NIH Data Commons Pilot is testing usefulness of datasets, patient protections, interoperability, user tools needed (12 grants funded in November 2017)

- NCI grants awarded under the **Cancer Moonshot** require process for making resulting publications, and to the extent possible, underlying primary data "immediately and broadly available to the public"

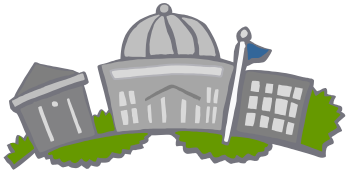
- **Clinical Trials Data Policy**: results are to be submitted to clinicaltrials.gov within 12 months of trial completion

- **Publications Policy**: all final versions of publications resulting from NIH-funded research to be made publicly available through PubMed Central within 12 months of initial publication

A snapshot of the Foundation for the NIH

The FNIH was established by Congress in 1990 as a not-for-profit charitable organization

The Foundation began its work in 1996 to facilitate groundbreaking research at the NIH and worldwide



By creating effective alliances to advance biomedical research

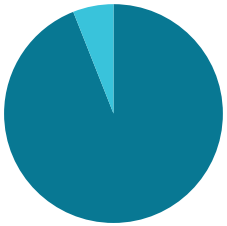
FNIH
Foundation for the National Institutes of Health

501(c)(3)
Non-governmental not-for-profit & independent Board of Directors

Over 550
projects supported

120+
Active research partnerships, scientific education/training, conferences/events and capital programs

Over $1 billion
raised by the FNIH since 1996



93%
of funds directly support programs

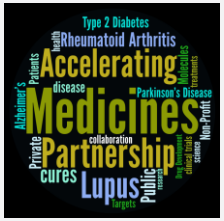
14 years
of outstanding Charity Navigator ratings

Research Partnerships at FNIH

LUNG-MAP

PredictTB

Development of a Consensus Pathway For Field Testing Gene Drive-Modified Mosquitoes



ADNI

SHORTEN-TB (Developing Lead Compounds to Shorten the duration of Tuberculosis Chemotherapy)

THE biomarkers CONSORTIUM

Development of new technologies for controlling transmission of Mosquito-borne diseases

CCVIMC
Comprehensive Cellular Vaccine Immune Monitoring Consortium

Example: AMP T2D Knowledge Portal

Provides open access to data to the entire research community via sophisticated query engine and analytical tools (7000+ current users worldwide)

Software platform and analytical tools are all available via open source repository (GitHub)

Uses federated and mirrored pipelines to share data without moving them from their source

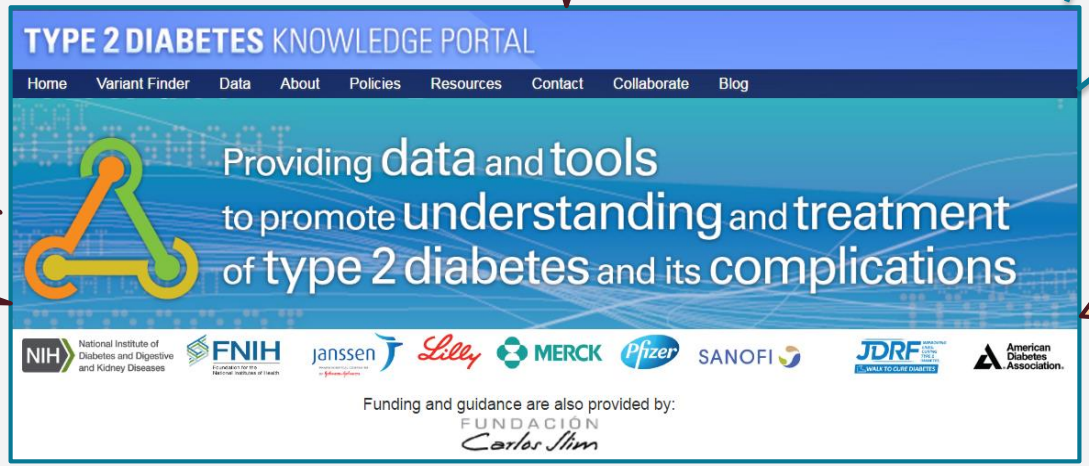
Epigenomics data

Diabetes Epigenome Atlas

European Data Hub

EMBL-EBI

Currently includes data from >1M GWAS, 228K exome chip, 19K whole exomes, 33 datasets, 62 traits, 5 ancestries



Connects genomic data and disease insights

Same software platform in use at Broad to develop portals for AMP AD and PD, as well as in CVD, cerebro-vascular, IBD, epilepsy, etc.

BROAD INSTITUTE

Rapid Data Release

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0084-1

Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes

Anubha Mahajan*

We aggregated coding variant data for 81,412 type 2 diabetes cases and 370,832 controls of diverse ancestry, identifying 40 coding variant association signals ($P < 2.2 \times 10^{-7}$); of these, 16 map outside known risk-associated loci. We make two important observations. First, only five of these signals are driven by low-frequency variants: even for these, effect sizes are modest (odds ratio ≤1.29). Second, when we used large-scale genome-wide association data to fine-map the associated variants in their regional context, accounting for the global enrichment of complex trait associations in coding sequence, compelling evidence for coding variant causality was obtained for only 16 signals. At 13 others, the associated coding variants clearly represent 'false leads' with potential to generate erroneous mechanistic inference. Coding variant associations offer a direct route to biological insight for complex diseases and identification of validated therapeutic targets; however, appropriate mechanistic inference requires careful specification of their causal contribution to disease predisposition.

Results:

- Coding variant data for 81,412 type 2 diabetes cases and 370,832 controls of diverse ancestry (many cohorts analyzed)
- Identification of 40 coding variant associated traits
- 16 variants map to new loci **= potential new targets**

Paper Published	April 9, 2018
Data Uploaded to the KP	April 9, 2018
Data available for querying alongside other data in KP	April 9, 2018