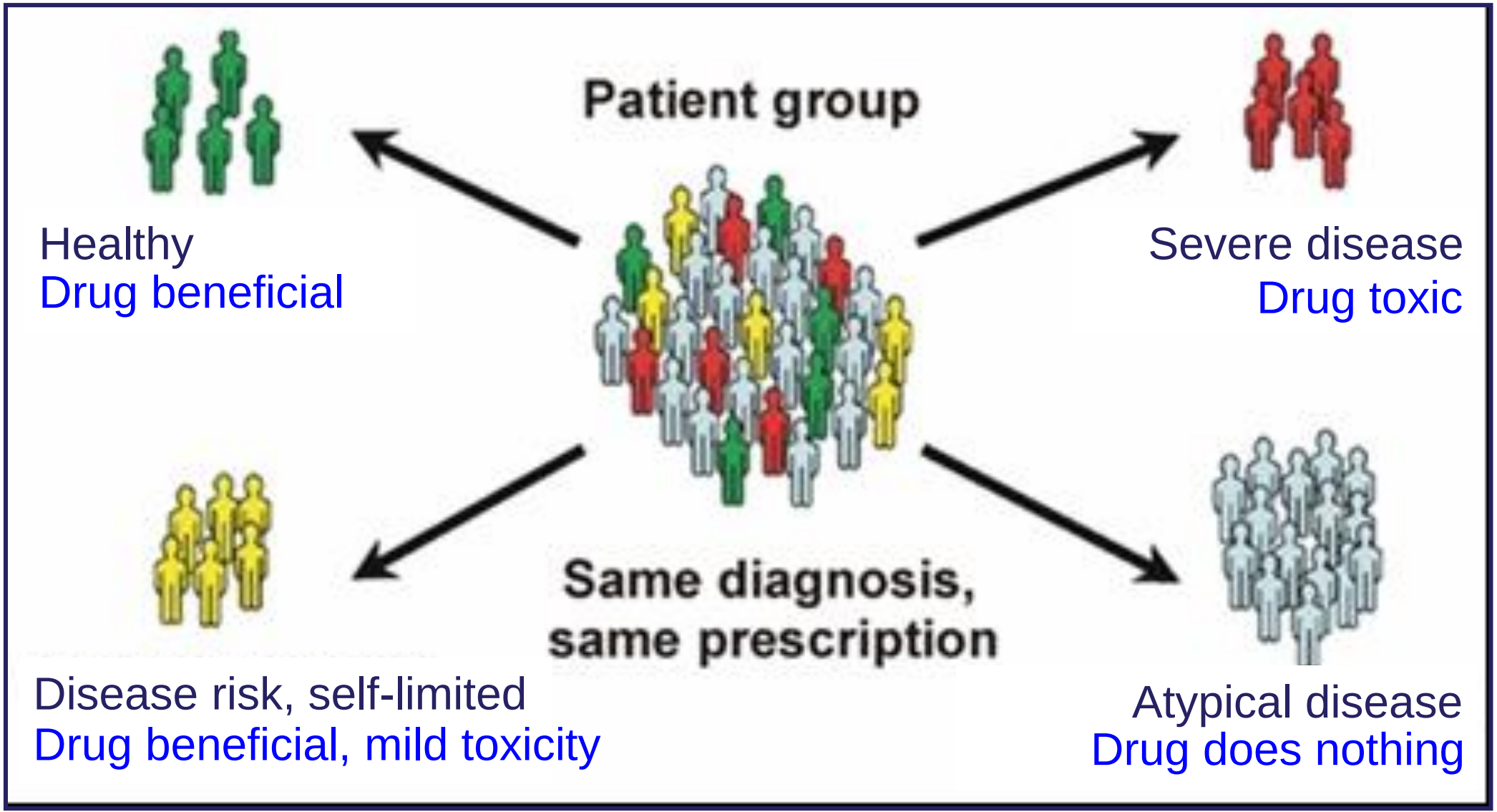
The *All of Us* Research Program – Program Overview

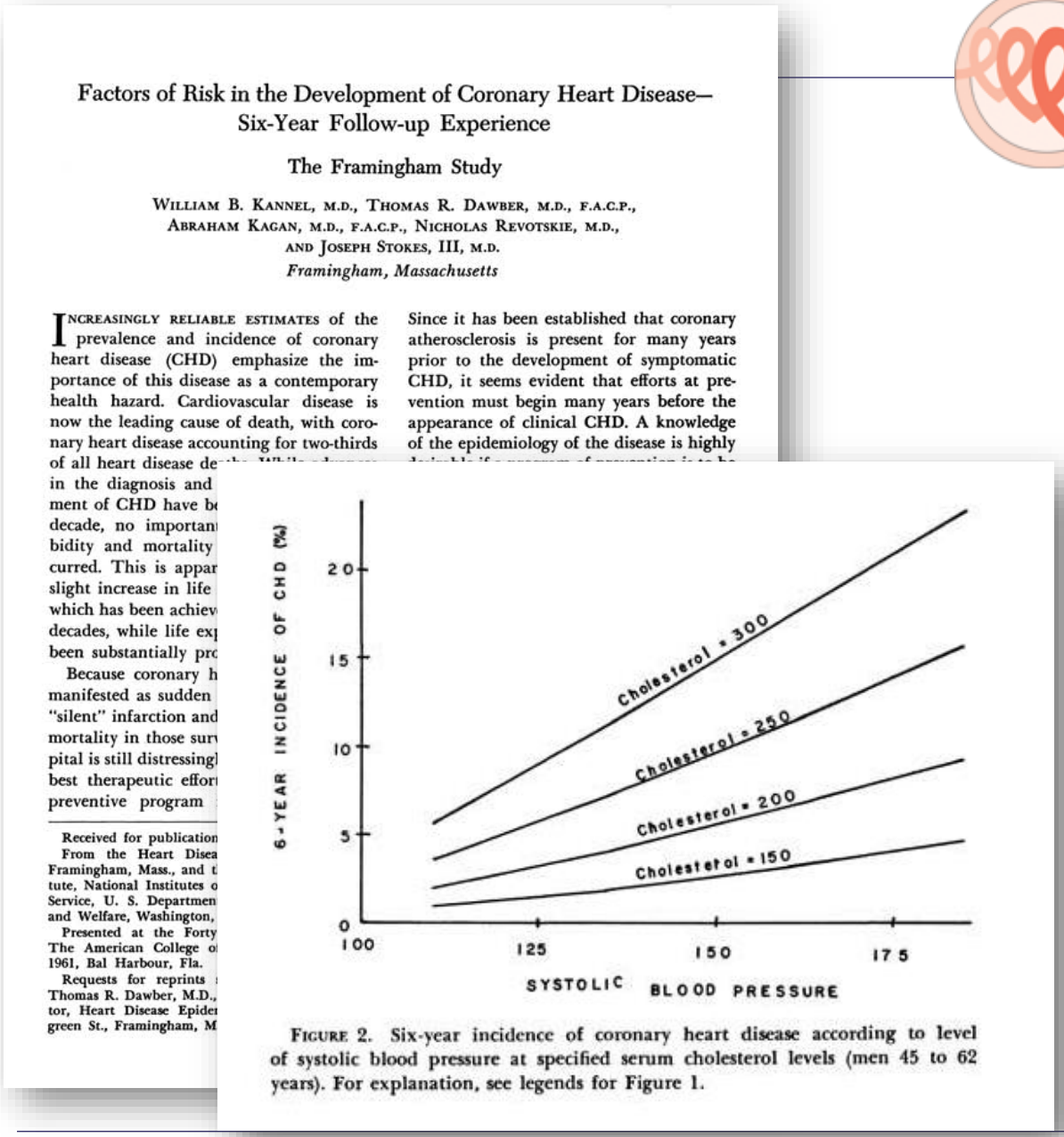
NIH Collaboratory Grand Rounds
Feb 2, 2018

Josh Denny, MD, MS
Professor of Biomedical Informatics and Medicine
Vanderbilt University Medical Center

People have different disease risk and variable drug response



Framingham Heart Study



Factors of Risk in the Development of Coronary Heart Disease—
Six-Year Follow-up Experience

The Framingham Study

WILLIAM B. KANNEL, M.D., THOMAS R. DAWBER, M.D., F.A.C.P.,
ABRAHAM KAGAN, M.D., F.A.C.P., NICHOLAS REVOTSKIE, M.D.,
AND JOSEPH STOKES, III, M.D.
Framingham, Massachusetts

INCREASINGLY RELIABLE ESTIMATES of the prevalence and incidence of coronary heart disease (CHD) emphasize the importance of this disease as a contemporary health hazard. Cardiovascular disease is now the leading cause of death, with coronary heart disease accounting for two-thirds of all heart disease de

in the diagnosis and
ment of CHD have b
decade, no importan
bidity and mortality
curred. This is appar
slight increase in life
which has been achiev
decades, while life ex
been substantially pr

Because coronary h
manifested as sudden
"silent" infarction and
mortality in those sur
pital is still distressingl
best therapeutic effor
preventive program

Since it has been established that coronary atherosclerosis is present for many years prior to the development of symptomatic CHD, it seems evident that efforts at prevention must begin many years before the appearance of clinical CHD. A knowledge of the epidemiology of the disease is highly

Received for publication
From the Heart Disea
Framingham, Mass., and t
tute, National Institutes o
Service, U. S. Departmen
and Welfare, Washington,
Presented at the Forty
The American College o
1961, Bal Harbour, Fla.
Requests for reprints
Thomas R. Dawber, M.D.,
tor, Heart Disease Epide
green St., Framingham, M

FIGURE 2. Six-year incidence of coronary heart disease according to level of systolic blood pressure at specified serum cholesterol levels (men 45 to 62 years). For explanation, see legends for Figure 1.

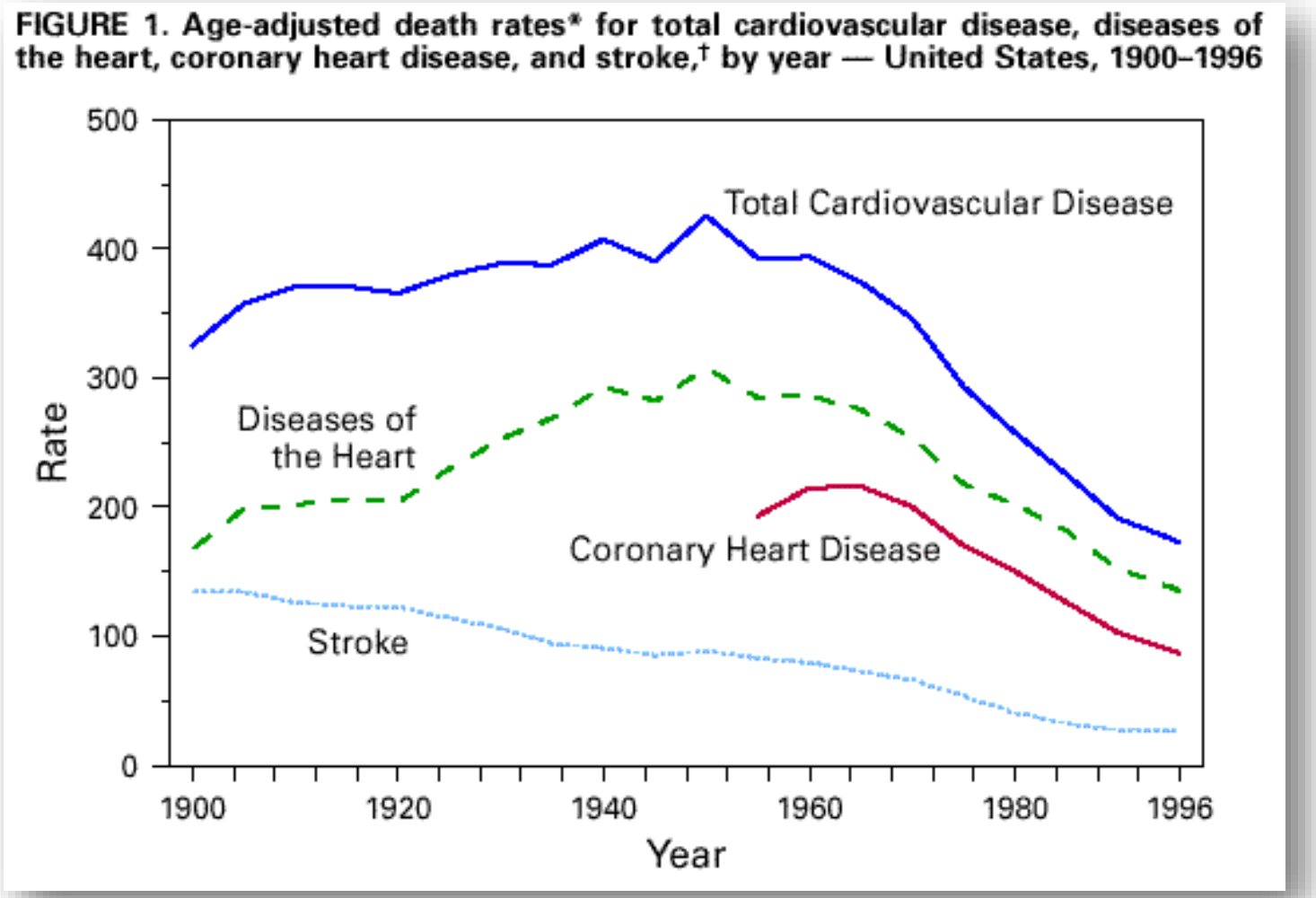
Enrolled 5209 men and women in 1948

Some Framingham early discoveries:

- 1960 – Cigarettes increase heart disease
- 1961 – cholesterol, blood pressure increase heart disease
- 1967 – exercise decreases risk of heart disease; obesity increases it
- 1970 – high blood pressure and atrial fibrillation cause stroke

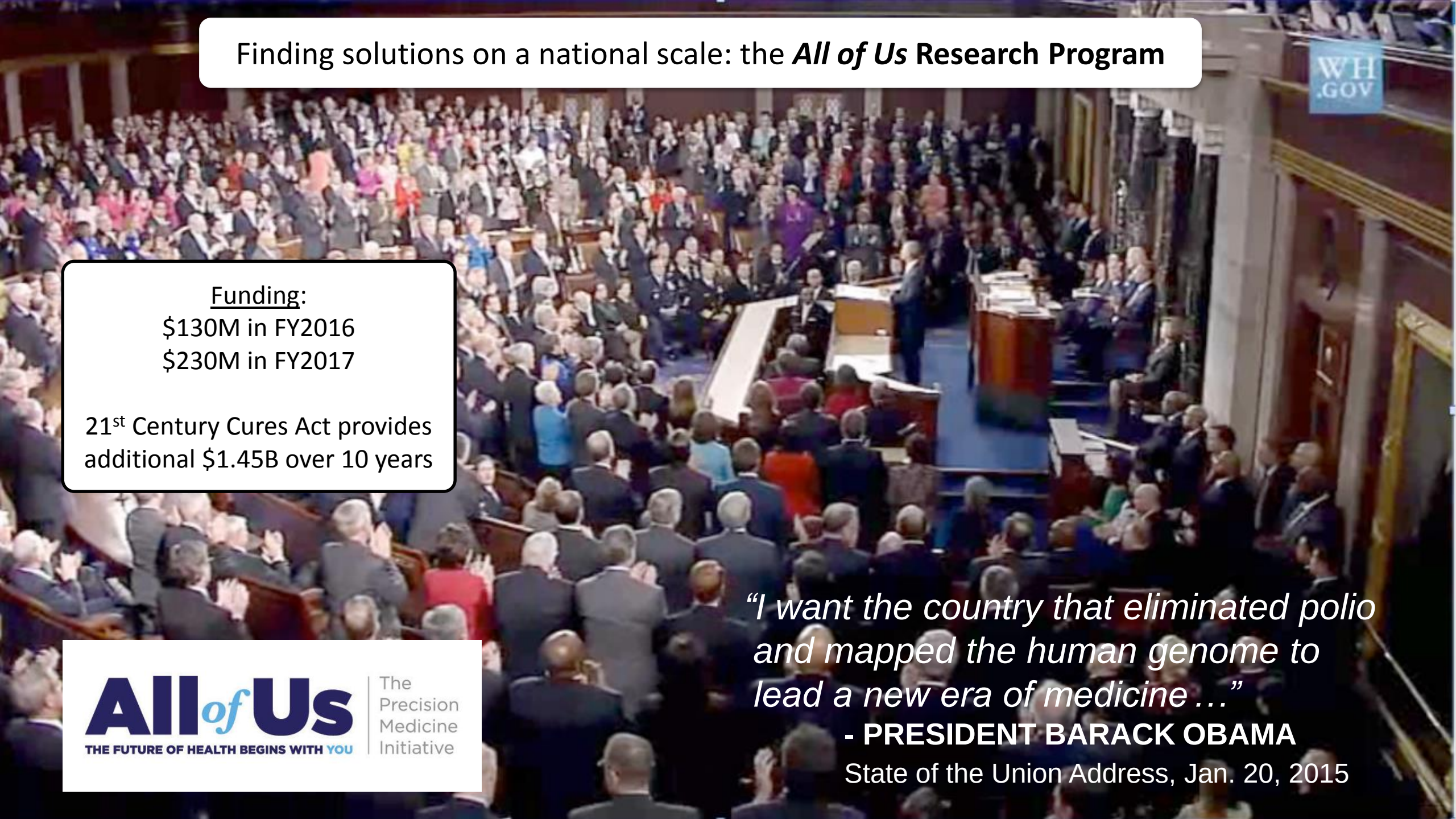
Kannel et al. Annals of Intern Med, 1961

The impact of Framingham (and similar cohorts) has been dramatic



https://www.cdc.gov/Mmwr/preview/mmwrhtml/mm4830a1.htm

Finding solutions on a national scale: the *All of Us* Research Program



Funding:
$130M in FY2016
$230M in FY2017

21st Century Cures Act provides additional $1.45B over 10 years

All of Us
THE FUTURE OF HEALTH BEGINS WITH YOU
The Precision Medicine Initiative

"I want the country that eliminated polio and mapped the human genome to lead a new era of medicine …"

- PRESIDENT BARACK OBAMA

State of the Union Address, Jan. 20, 2015

Core goals for the *All of Us* Research Program

The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine

Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director, NIH

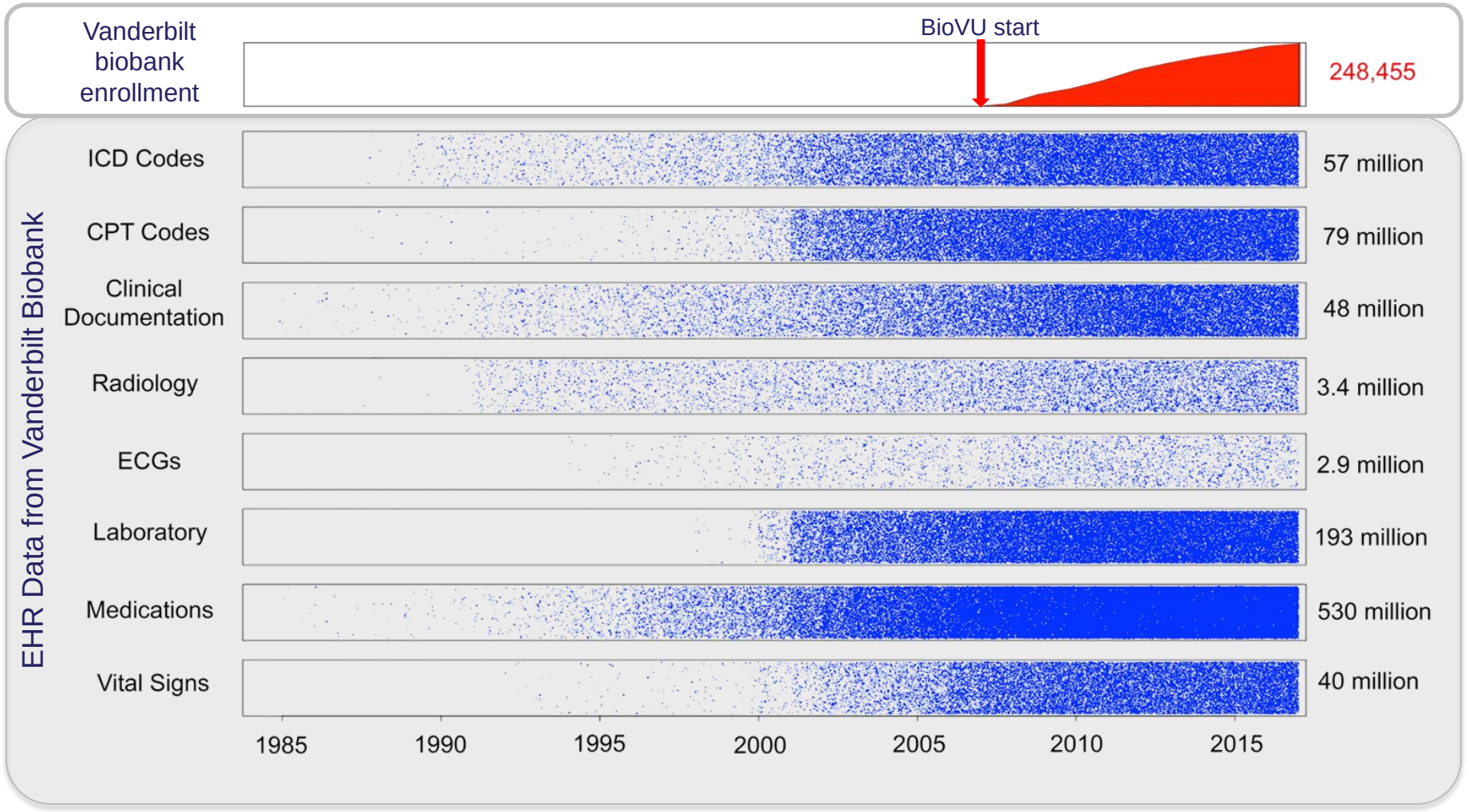
September 17, 2015

- **1 million or more**
- Longitudinal, **recontactable**
- **EHR data**, biospecimens, baseline evaluations, and health surveys
- Focus on **engagement**
- Focus on **diversity**

- **Network formed – July 2016**

https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf
https://allofus.nih.gov/news-events-and-media/announcements/all-us-research-program-initial-protocol

EHR data are dense and efficient for discovery: Vanderbilt's experience (BioVU)



All of Us Research Program - Mission and Objectives

1. Nurture relationships
with **one million** or more participant partners, from all walks of life, for decades, reflecting the **broad diversity** of the U.S, especially those **underrepresented in biomedical research**

Our mission

To accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care for all of us

2. Deliver the largest, richest, and most accessible biomedical dataset
providing the **tools & capabilities** that make it easy for researchers, participants, and citizen scientists to make discoveries

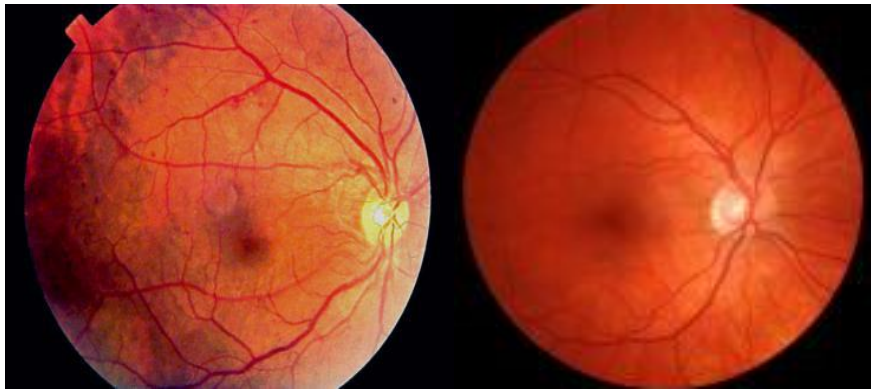
3. Catalyze a robust ecosystem
of researchers and funders to use and support the **rich, longitudinal resource** of deep clinical, **environmental, lifestyle, & genetic** data.

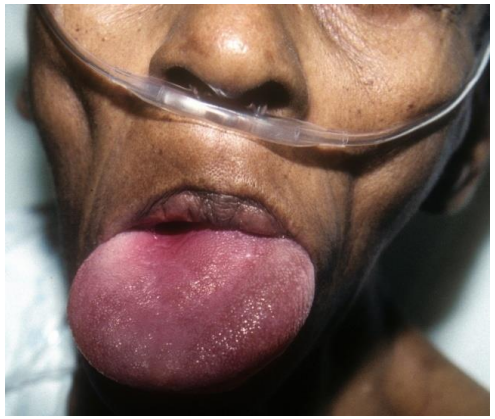
Some of the *All of Us* research goals

Studying exposures and habits

New biomarkers (and their interactions)

Machine learning on Big Data

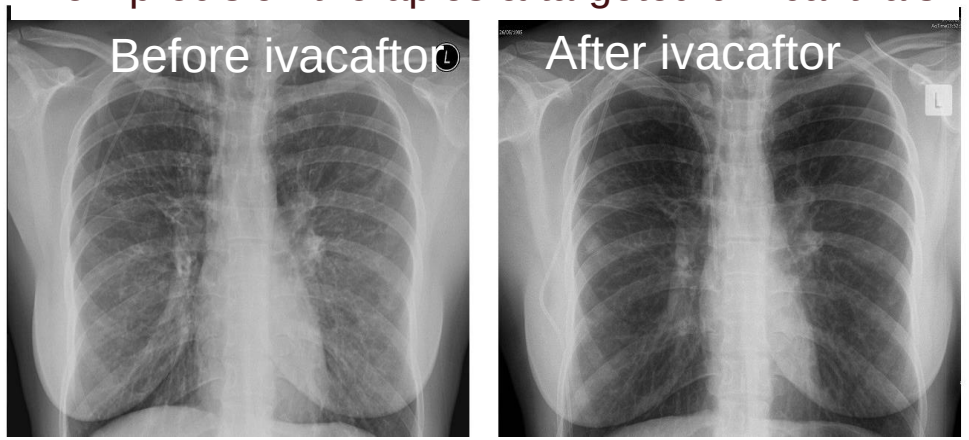




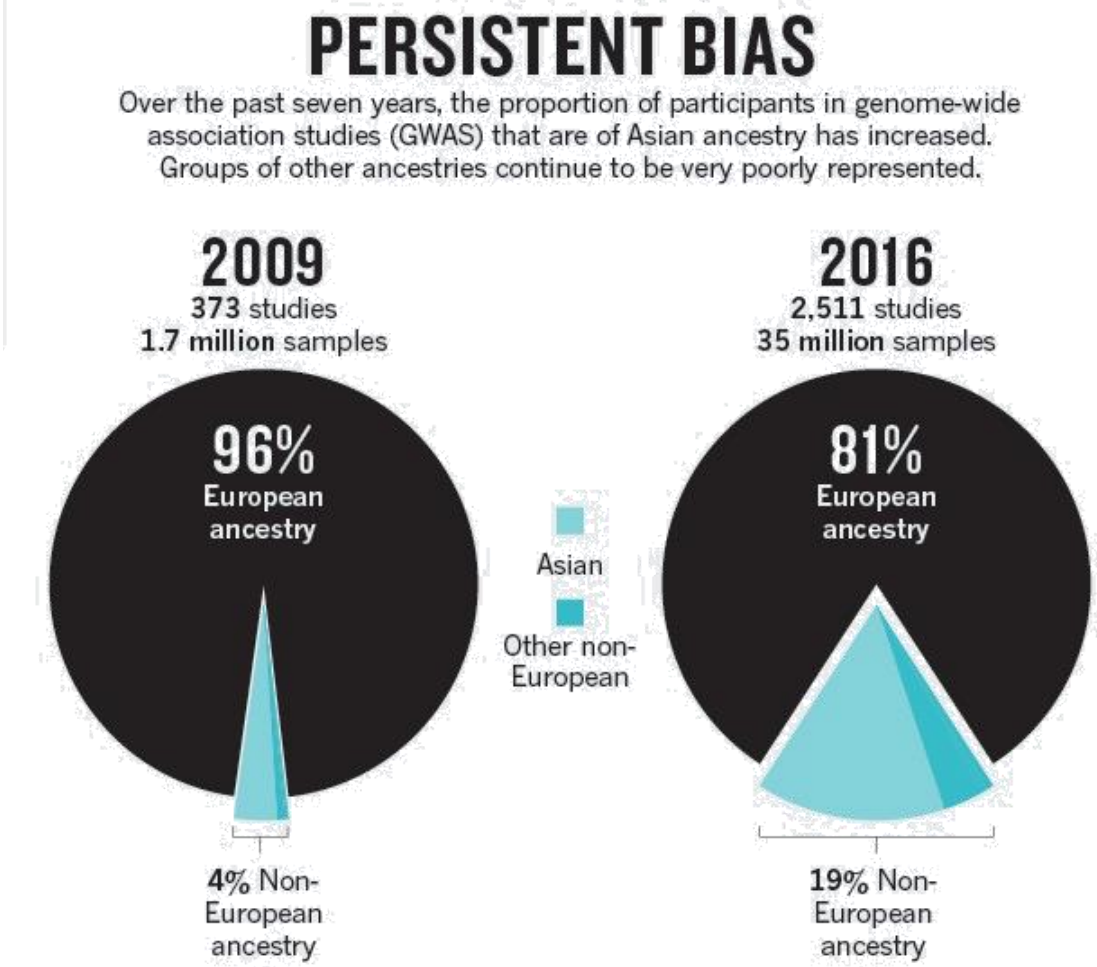
Improving drug action

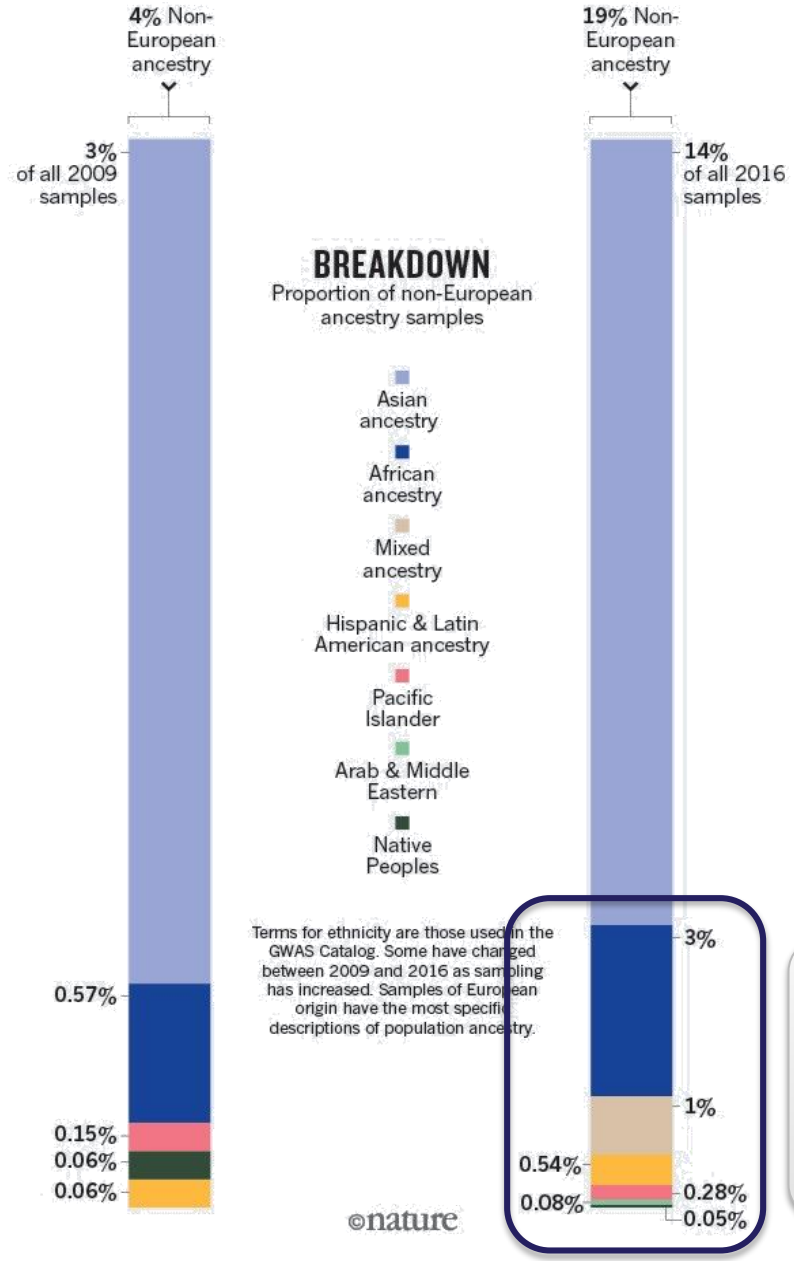
Engaging diverse participants

New precision therapies & targeted clinical trials



Why Diversity?





4% GWAS represents ~1/3 of US population

Popejoy & Fullerton, *Nature* 2016

Major Building Blocks of the Research Program

DATA AND RESEARCH CENTER (DRC)

Big data capture, cleaning, curation, & sharing in secure environment

Vanderbilt, Verily, Broad Institute

BIOBANK

Repository for processing, storing, & sharing biosamples

Mayo Clinic

PARTICIPANT CENTER

Direct volunteer participant enrollment, digital engagement innovation, & consumer health technologies

Scripps Research Institute (with multiple partners)

PARTICIPANT TECHNOLOGY SYSTEMS CENTER

Web & phone-based platforms for participants

Vibrent Health

HEALTH CARE PROVIDER ORGS (HPOs)

Clinical & scientific expertise network, enrollment & retention of participants

20+ regional med centers, FQHCs, VA, future awards to grow network

COMMUNICATIONS & ENGAGEMENT

Comms, marketing, & design expertise; Engagement coordination & community partners network

Wondros, HCM, future awards to grow network of community partners

National Network of Inaugural Partners

National Partners

Regional Medical Centers

FQHCs

Community Engagement Studios

- **77** Studios; **654** community members; Avg **8-9** community members/studio; **46%** self-identified as a racial/ethnic minority
- Studios in **17 cities** including Nashville, Los Angeles, Rochester, NY, **Chicago**, Sioux Falls, SD, Miami, New Orleans, San Diego, Appalachian Mountains
- 15 engagement studios on Return of Value

Community Engagement Partners – led by Dara Richardson-Heron

- **American Academy of Family Physicians**, Leawood, Kansas
- **American Academy of HIV Medicine,** Washington, D.C.
- **American Association of Colleges of Nursing**, Washington, D.C.
- **American Medical Association**, Chicago, Illinois
- **Arab Community Center for Economic and Social Services**, Dearborn, Michigan
- **Asian & Pacific Islander American Health Forum**, Oakland, California
- **Association of Nurses in AIDS Care**, Uniontown, Ohio
- **Black Women's Health Imperative**, Washington, D.C.
- **Cobb Institute (W. Montague Cobb/National Medical Association Health Institute)**, Washington, D.C.
- **Delta Research and Educational Foundation**, Washington, D.C.
- **FiftyForward**, Nashville, Tennessee
- **League of United Latin American Citizens**, Washington, D.C.
- **National Alliance for Hispanic Health**, Washington, D.C.
- **National Baptist Convention,** Nashville, Tennessee
- **National Hispanic Medical Association**, Washington, D.C.
- **National Minority Quality Forum**, Washington, D.C.
- **National Network of Libraries of Medicine**, Bethesda, Maryland
- **San Francisco General Hospital Foundation**, San Francisco, California
- **UnidosUS**, Washington, D.C.

Summary of the approach and protocol

Direct Volunteers

Health Care Provider Organizations

Enroll, Consent, EHR sharing

Health Surveys

Baseline measurements

Bio-specimens

Smartphones & Wearables

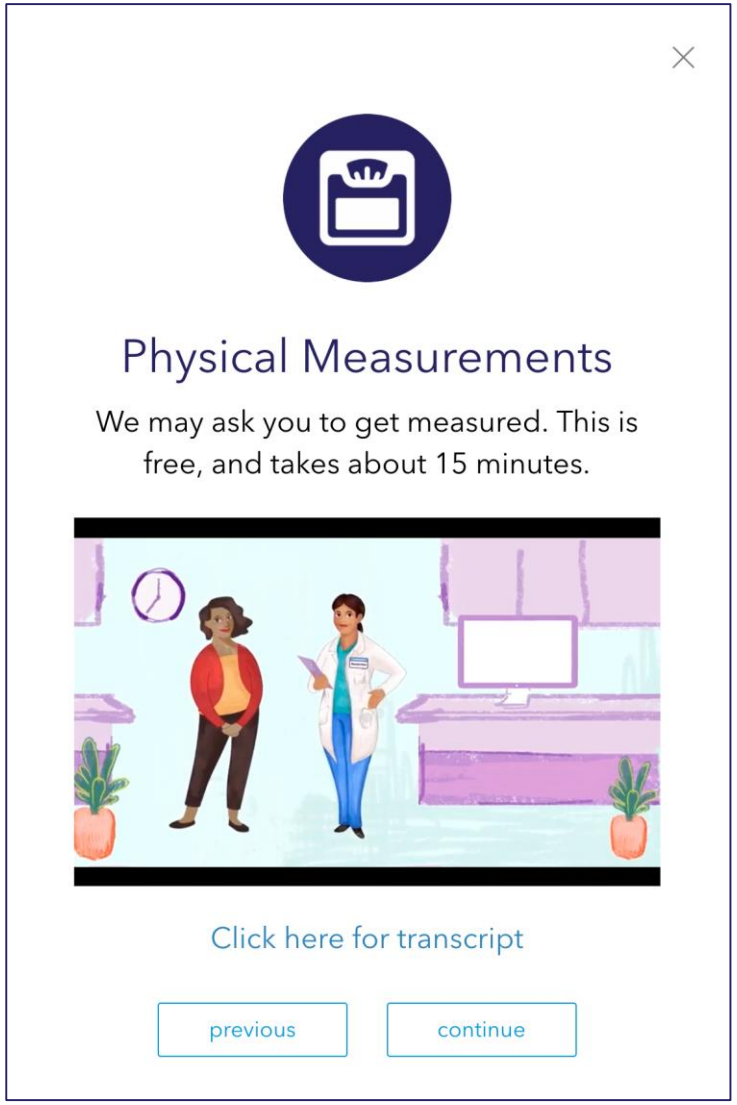
Build direct volunteer capacity

- **Vision**: Make it possible for anyone, anywhere in the country to participate in biomedical research.

- **Where we are**:
  - Built a network of partners
  - 3 DV locations online now, with more coming online regularly
  - Reach of 37,000 facilities or providers covering 97% of US …but we cannot fully staff all at once.

Consent / e-Consent

- Recruit 18+ years old initially; working on pediatrics plan
- eConsent (paper long-form being developed)
- 6th grade reading level; English & Spanish
- eConsent process includes modules on:
  - Participant Provided Info (PPI) + Linkage + Re-contact
  - Physical Measurements (PM) + Biospecimen
  - Sensors or wearable devices
  - EHR
  - Genetic information
- Videos expand on key concepts
- Separate opt-in & signature for some modules, including EHR and genetics (state laws)



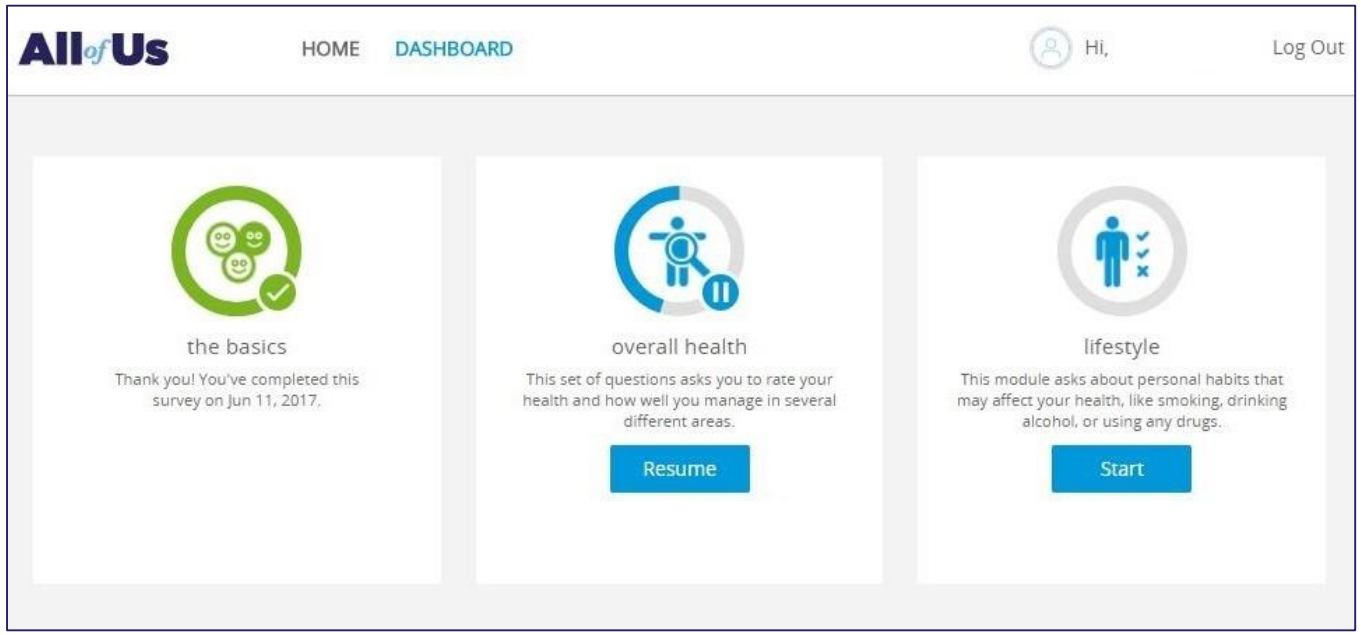
PPI/Survey Modules for Launch

PPI Enrollment Surveys

1. The Basics
2. Overall Health
3. Lifestyle

In Development

4. Personal Health History
5. Medications
6. Family History
7. Health Care Access and Utilization
8. Sleep
9. Environment/exposures



Physical Measurements

Physical Measurements

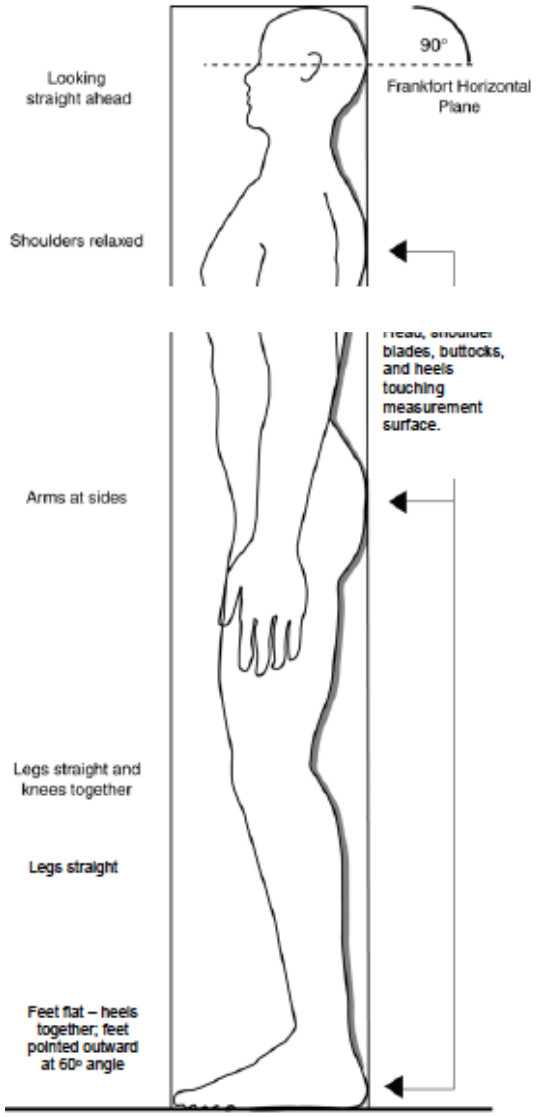
- Blood pressure
- BMI
- Heart rate
- Height
- Hip circumference
- Waist circumference
- Weight

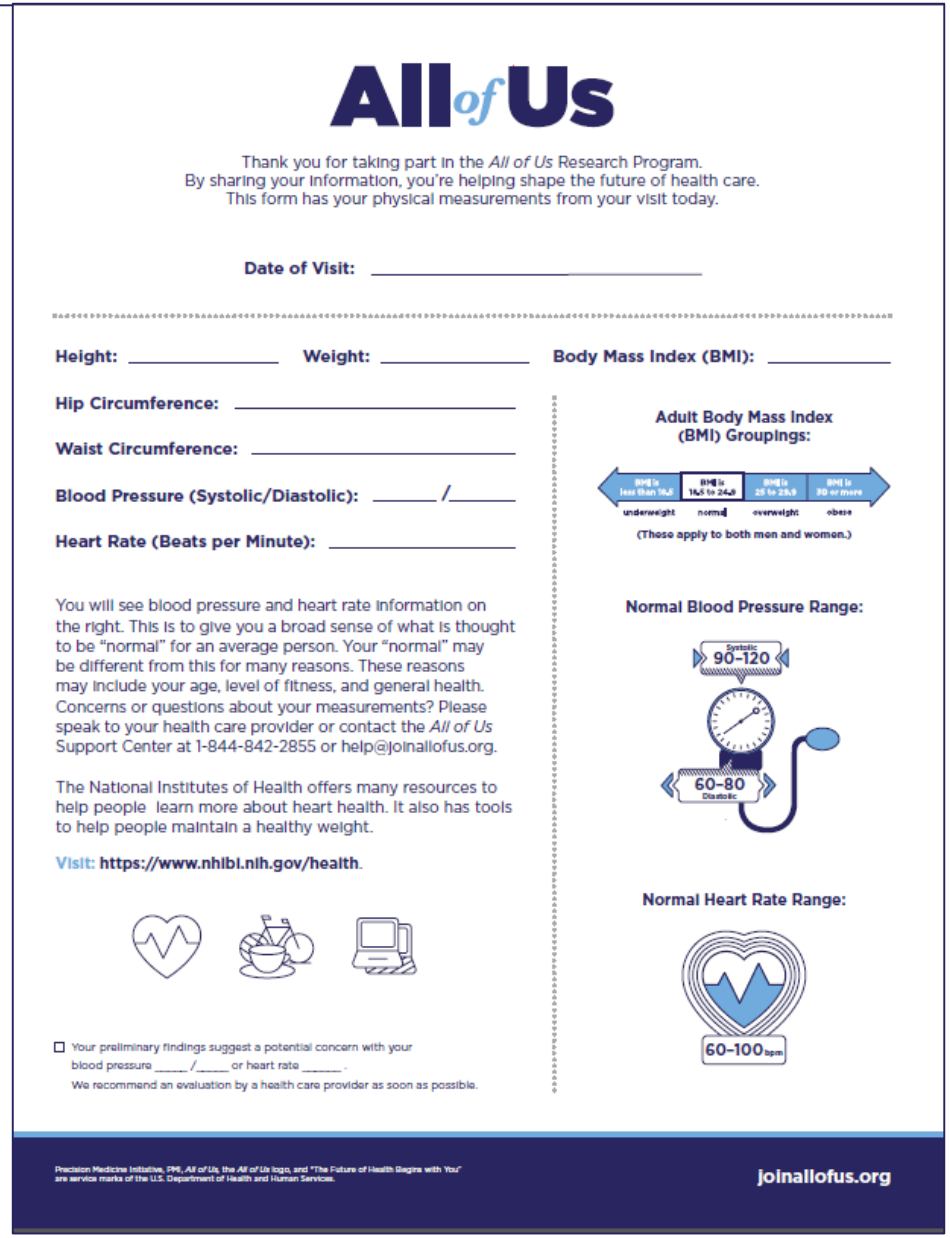
Biospecimen Collection

- Blood (or saliva)
- Urine

Participants will have access to their physical measurements through:

- The Participant Portal
- In Writing





All of Us

Thank you for taking part in the *All of Us* Research Program.
By sharing your information, you're helping shape the future of health care.
This form has your physical measurements from your visit today.

Date of Visit: _____

Height: _____ **Weight:** _____ **Body Mass Index (BMI):** _____

Hip Circumference: _____

Waist Circumference: _____

Blood Pressure (Systolic/Diastolic): _____ /_____

Heart Rate (Beats per Minute): _____

You will see blood pressure and heart rate information on the right. This is to give you a broad sense of what is thought to be "normal" for an average person. Your "normal" may be different from this for many reasons. These reasons may include your age, level of fitness, and general health. Concerns or questions about your measurements? Please speak to your health care provider or contact the *All of Us* Support Center at 1-844-842-2855 or help@joinallofus.org.

The National Institutes of Health offers many resources to help people learn more about heart health. It also has tools to help people maintain a healthy weight.

Visit: **https://www.nhlbi.nih.gov/health**.

☐ Your preliminary findings suggest a potential concern with your blood pressure _____ /_____ or heart rate _____.
We recommend an evaluation by a health care provider as soon as possible.

Adult Body Mass Index (BMI) Groupings:

BMI is less than 18.5 underweight | BMI is 18.5 to 24.9 normal | BMI is 25 to 29.9 overweight | BMI is 30 or more obese

(These apply to both men and women.)

Normal Blood Pressure Range:

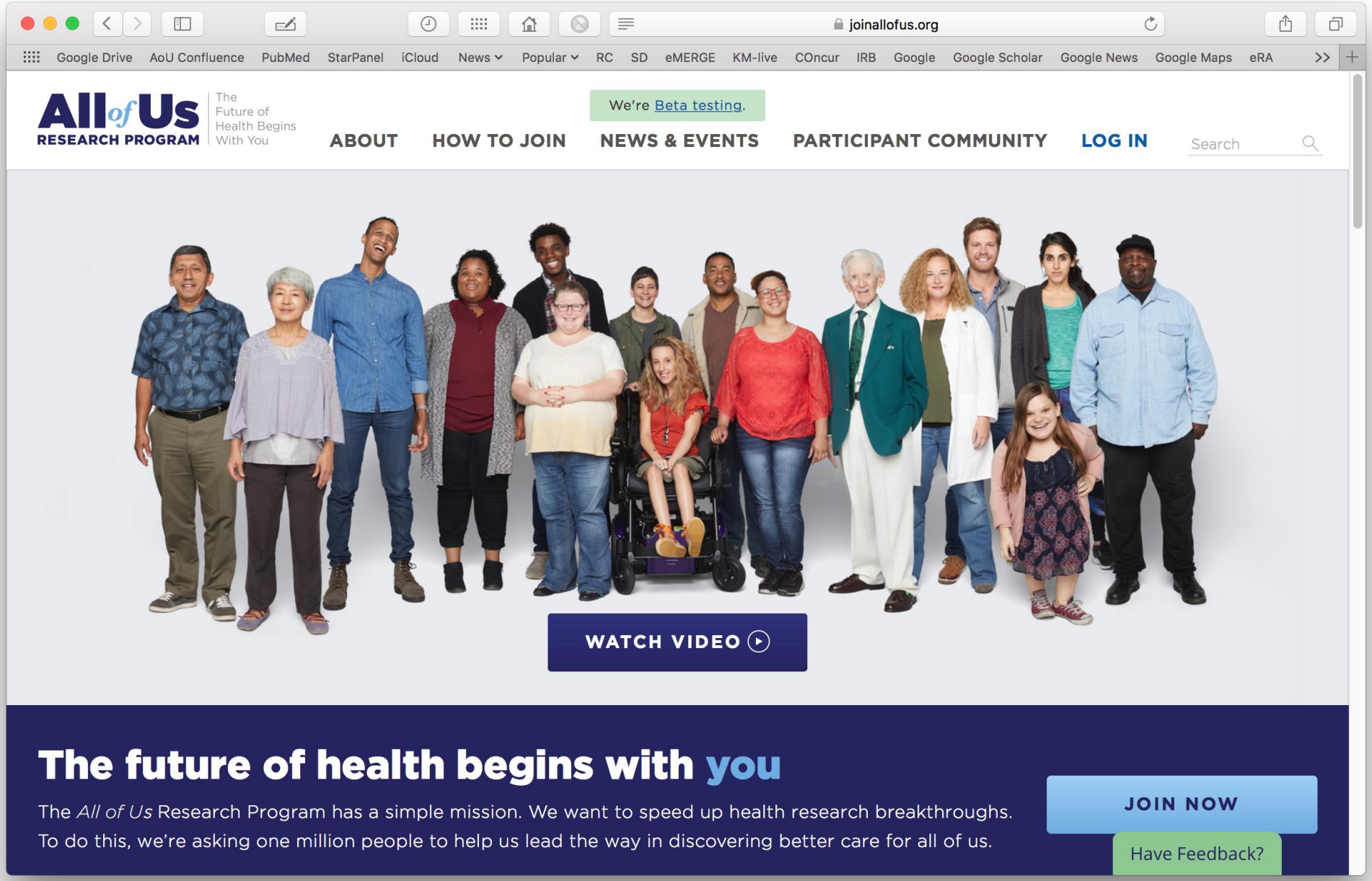
Systolic 90–120
60–80 Diastolic

Normal Heart Rate Range:

60–100 bpm

joinallofus.org

In beta testing now…

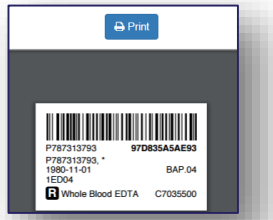


May 31, 2017: Launched Beta phase

- Version 1 protocol tested & IRB approved
- Completed security plan/tests
- Completed end-to-end “dress rehearsals” nationally
- Enrollment website & participant portal up & running
- Call center & command center up & running
- HPO network & Direct Volunteer capability established
- New sites launched every 1-3 weeks
- Kicked off mobile exhibit, the *All of Us* Journey
- Announced Fitbit pilot with 10,000 individuals

>13,000 participants in beta phase (slow ramp up)

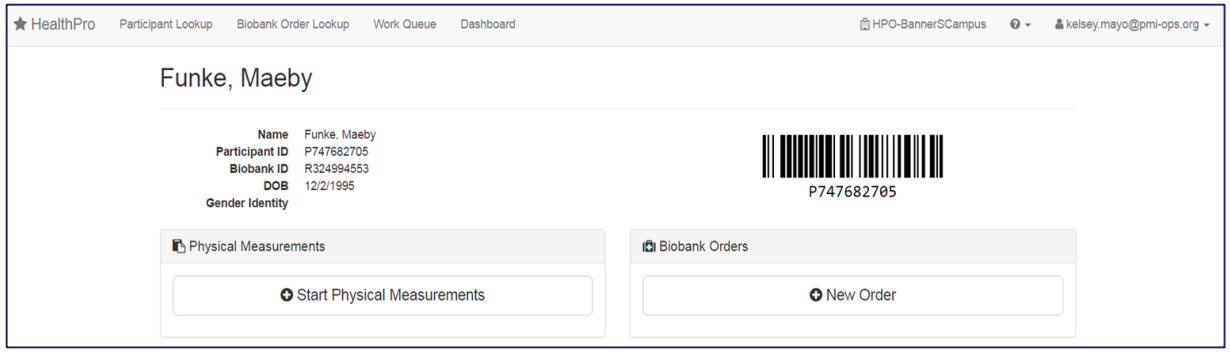
Goal: 1 million in ~4-5 years

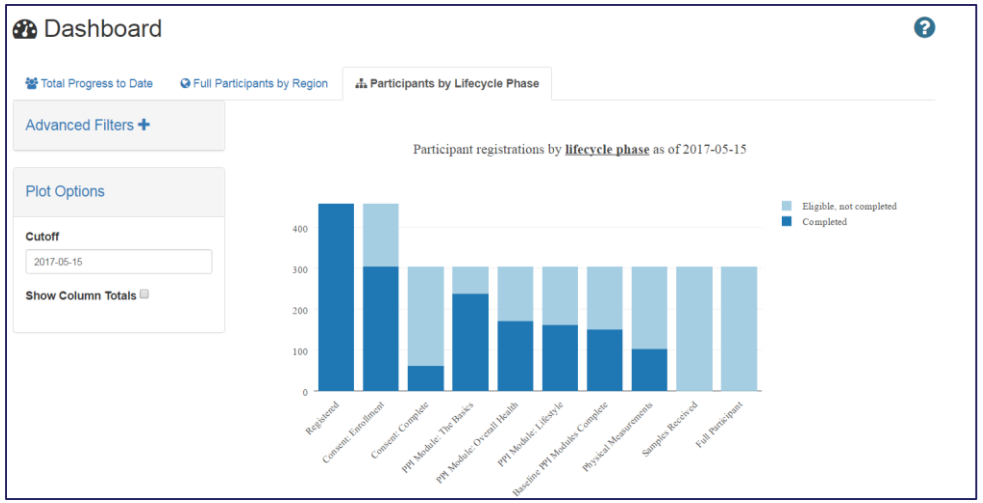


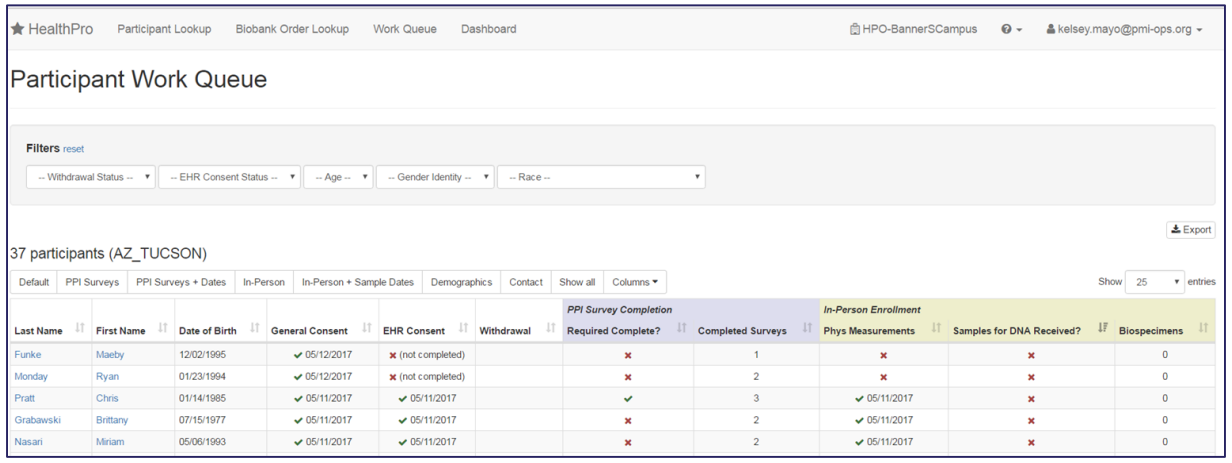
Data and Research Center (DRC): what we do

Our charter

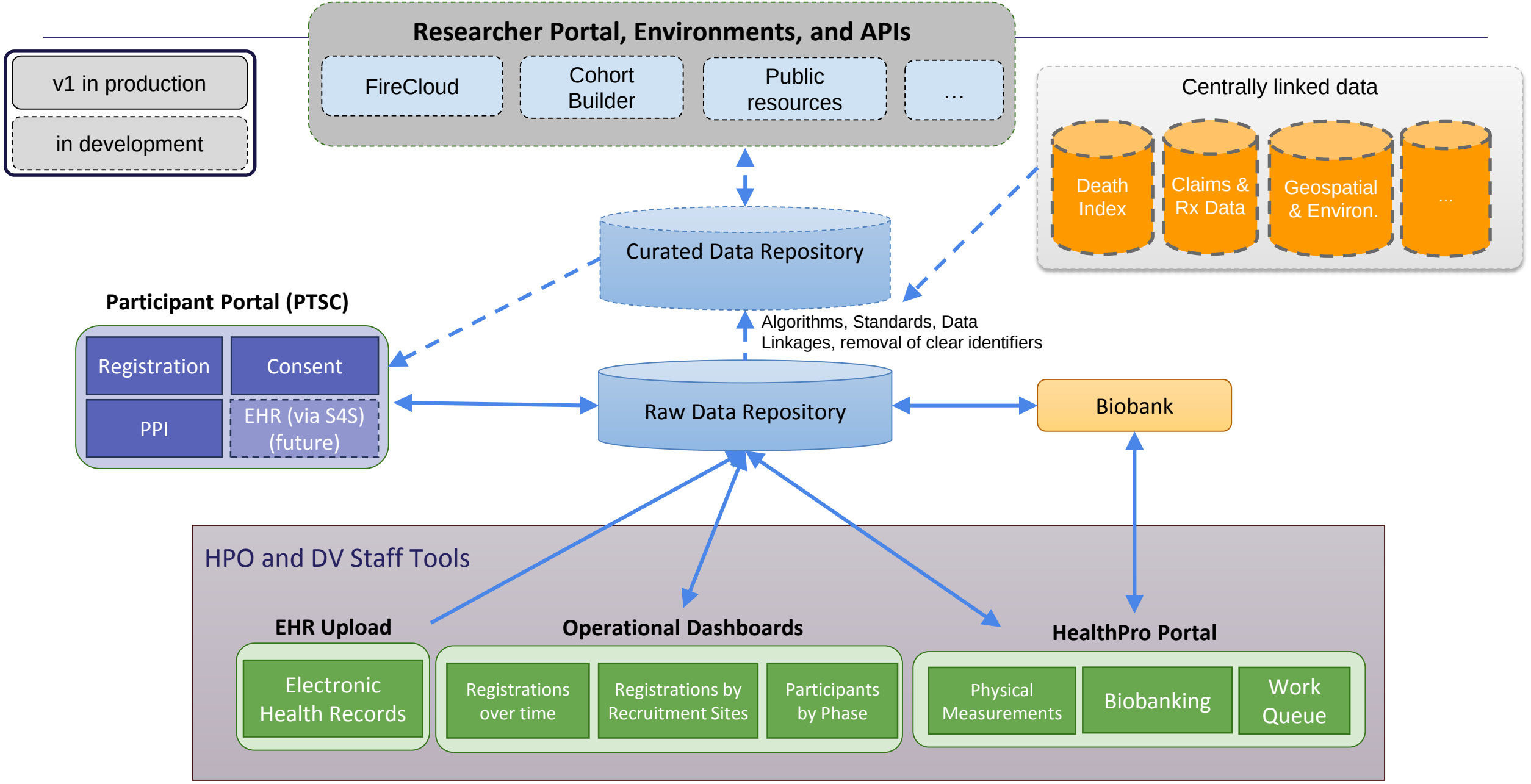
- manage & organize *All of Us* data
- build tools to enable data entry, perform quality control, and monitoring
- enforce data access policies and security
- make the data useful



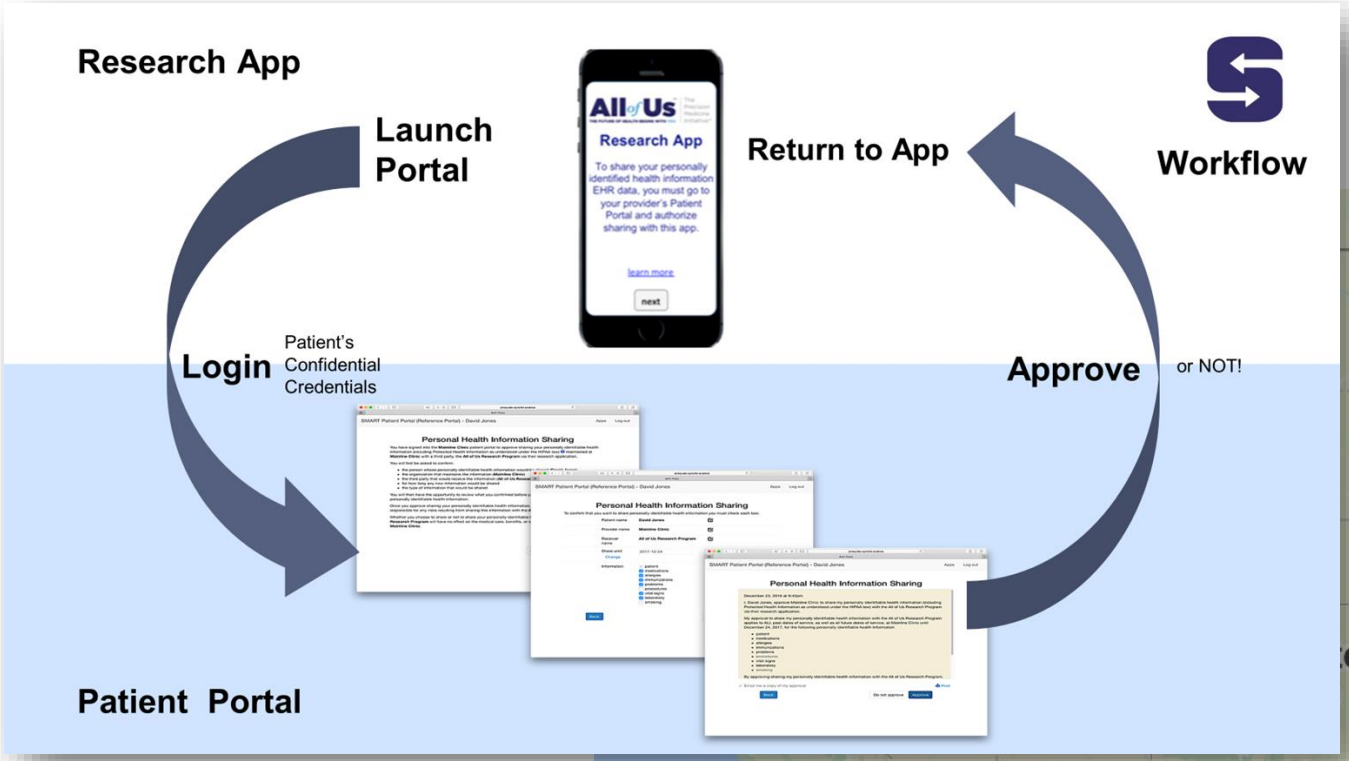


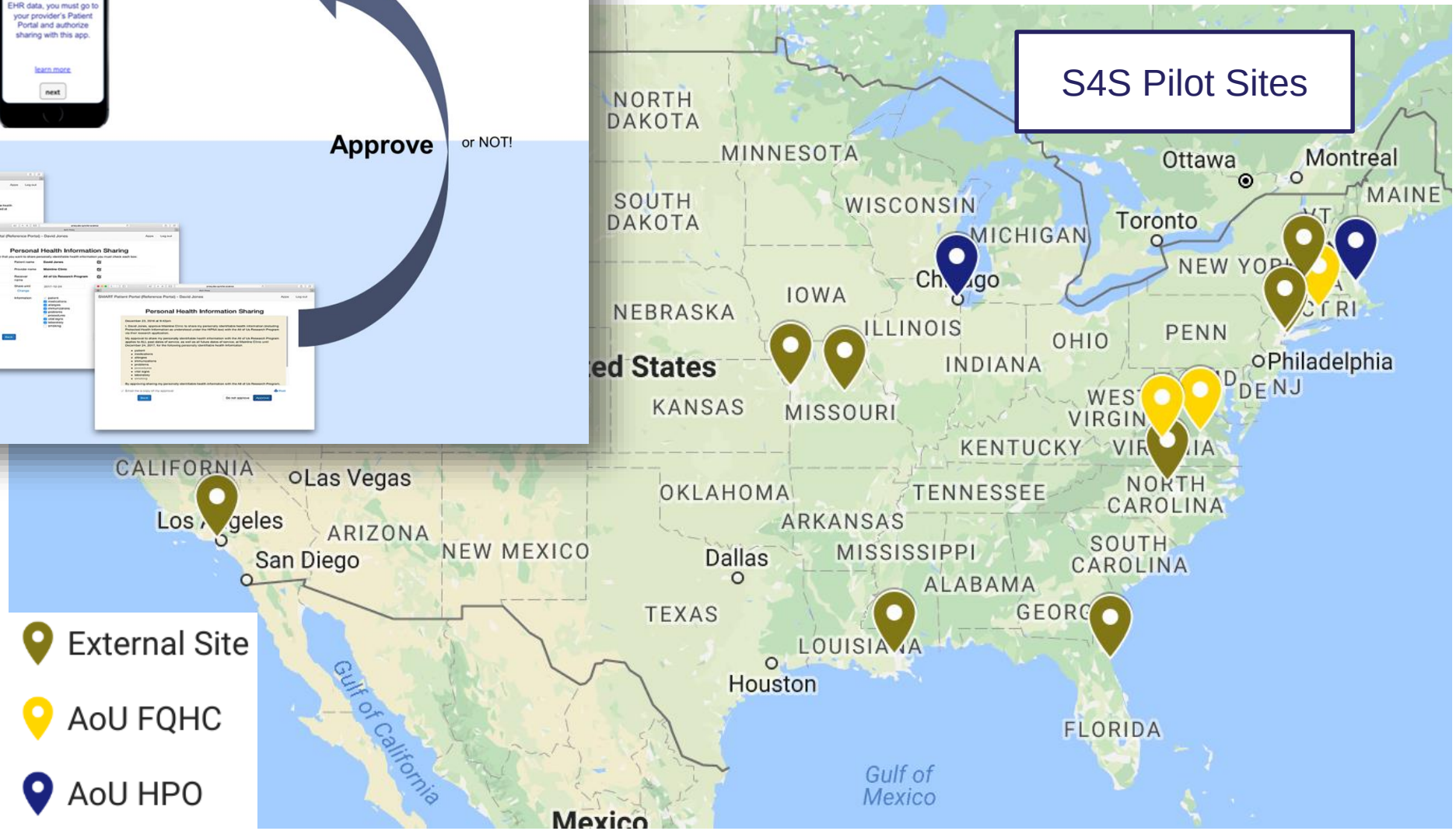


Key DRC Products for Data Ingestions, Curation, and Dissemination



Sync 4 Science (S4S) – a technology to share health data





Collecting Health Record/EHR data from *All of Us* Participants

From HPOs

Version 1 (2018)

Visits

Billing codes

Meds

Labs

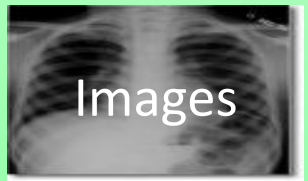
Version 2 (2019)

Clinical Notes & Reports

Clinical Messaging

Much longer term

Local Registries



Images

OMOP

Raw Data Repository

Curated Data Repository (OMOP)

FHIR, OMOP, ???

From Direct Volunteers

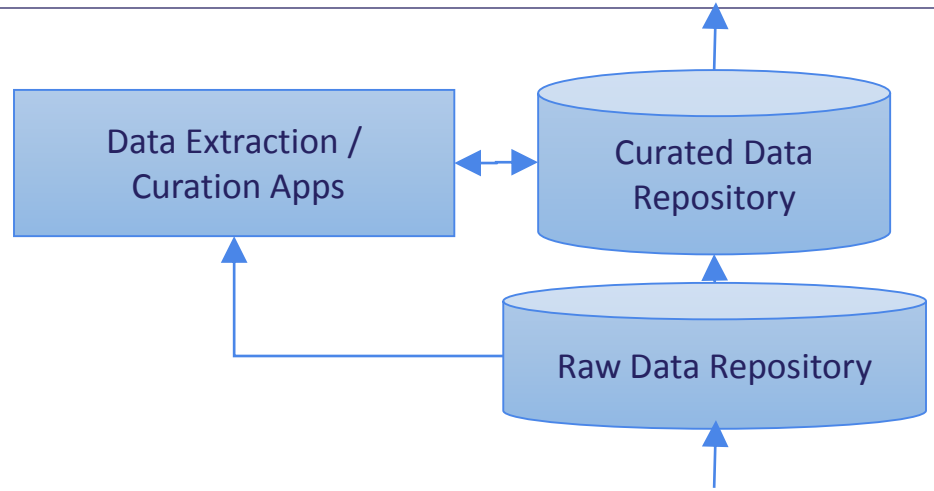
Sync for Science (2018-2019)

Health data aggregators (PicnicHealth)

Intelligent Curation

AOU data gets smarter over time.

- progressive mapping to common data models
- framework to allow anyone to add smart annotations
- plug in latest and greatest techniques



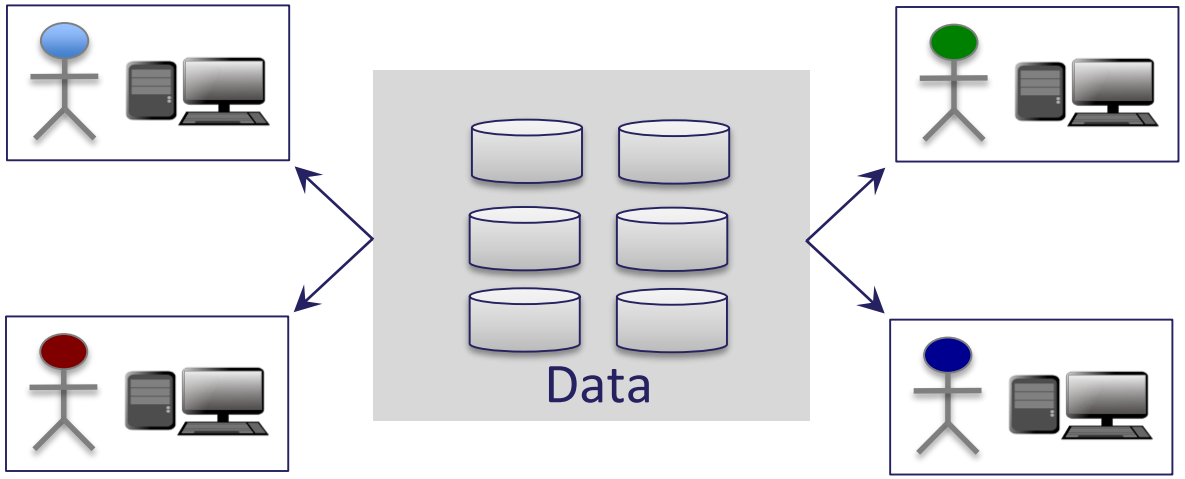
Automatically captioned:
"Two pizzas sitting on top of a stove top oven"

Leverage the "Web dividend" for Health.

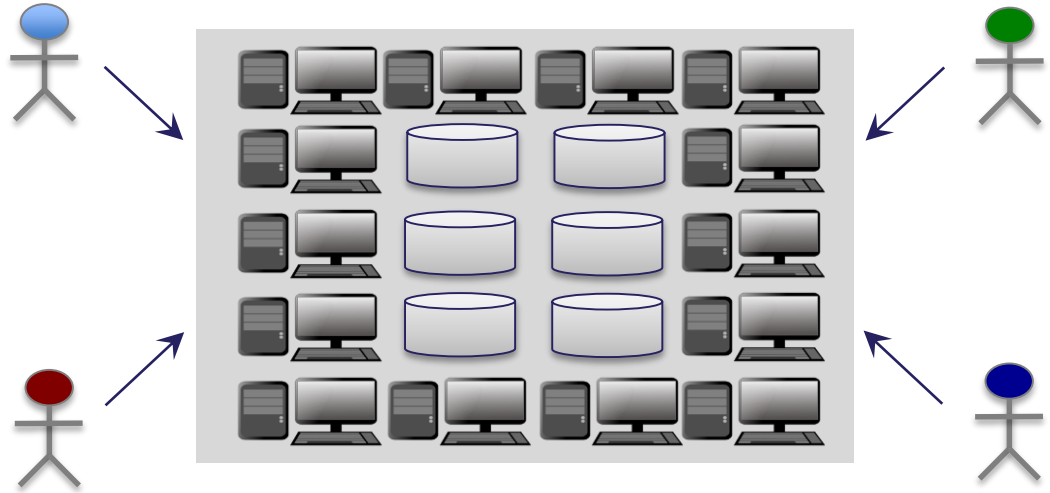
- Computable phenotypes
- Natural language processing
- Imputation
- Precomputed analyses
- Crowdsourcing
- Extracting geocoded environmental info

AOU centralizes data to **enhance security** and **improve usefulness**

Traditional Approach
Bring data to researchers



AoU Approach
Bring researchers to the data



Problems

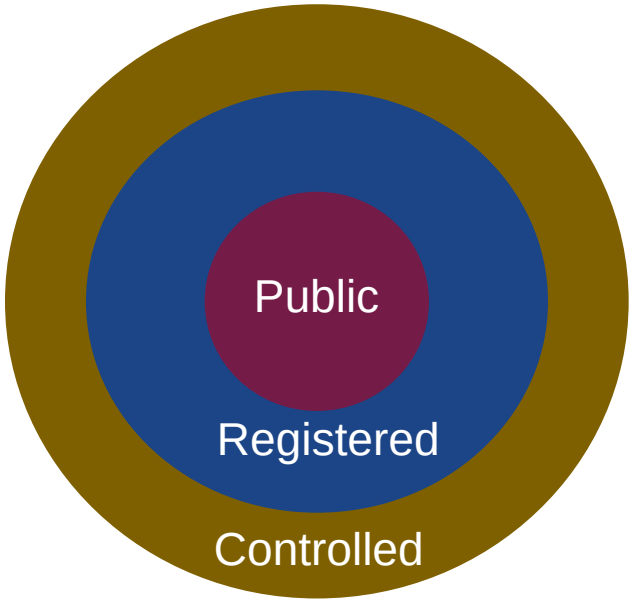
- *Data sharing = data copying*
- *Decreased security (data lots of places)*
- *Huge infrastructure needed*
- *Encourages siloed research*

Advantages

- *Improved security and auditing*
- *Increased accessibility to researchers*
- *Shared compute*
- *Facilitates collaboration*

Data Access Tiers (DRAFT)

Data Tiers



*All data tiers have **obvious identifiers removed***

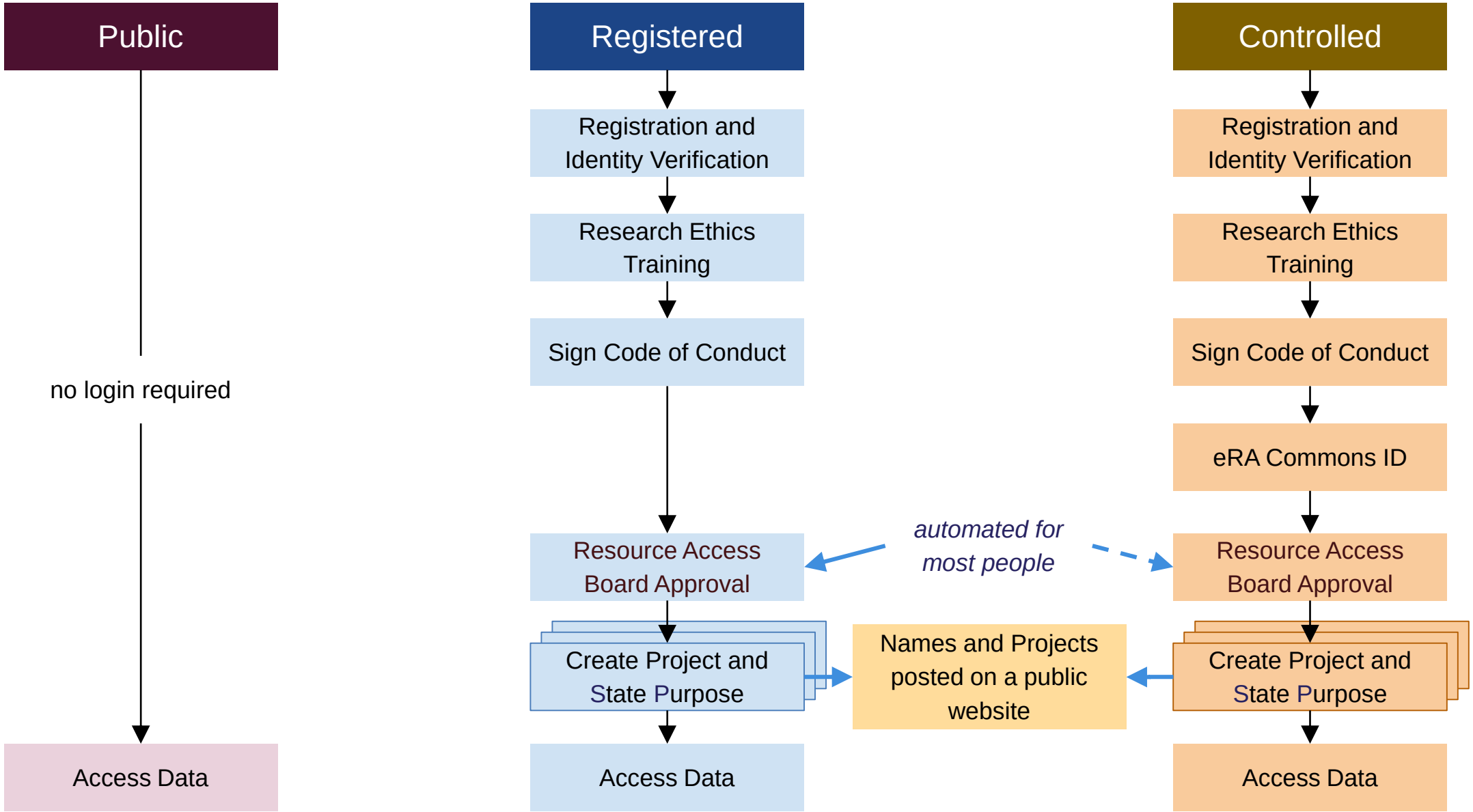
*Access to identifiers, recontact, or biospecimens **requires new IRB proposal***

1. **Public:** Data that poses minimal risks to the privacy of research participants. It can be accessed without logging into the *All of Us* Research Platform. *(e.g. aggregate statistics)*

2. **Registered:** Individual level data that has some risk to the privacy of research participants. It can only be accessed after logging into the *All of Us* Research Platform by **approved users**; all access will be logged and may be audited. *(e.g. PPI responses, EHR structured data)*

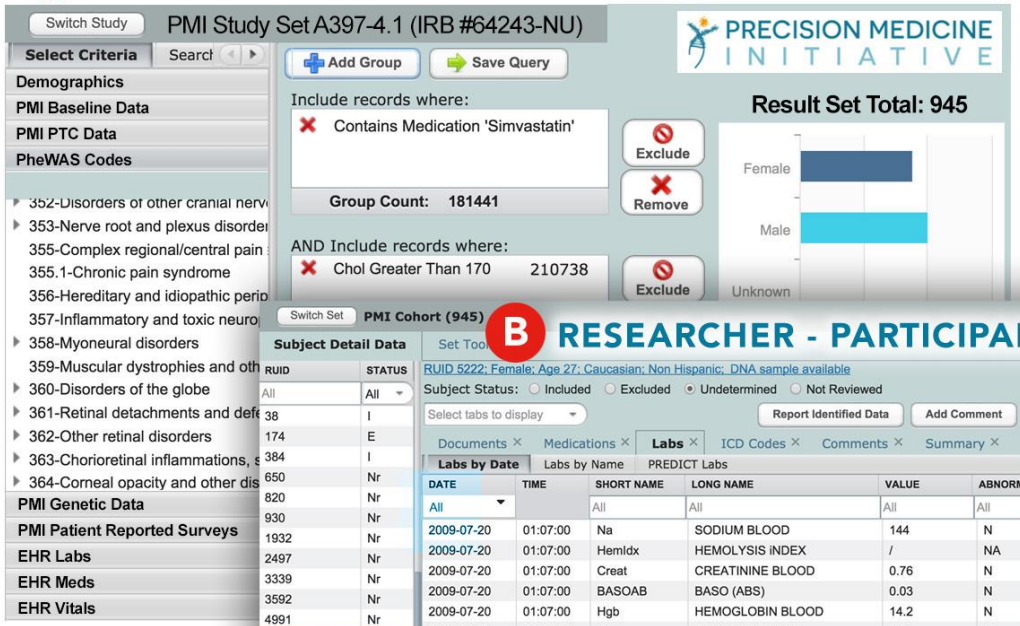
3. **Controlled:** Individual level data that poses the more significant risks to the privacy of research participants; needs additional approval step. *(e.g. EHR clinical notes)*

Data Access Protocol via a Passport model (**DRAFT**)

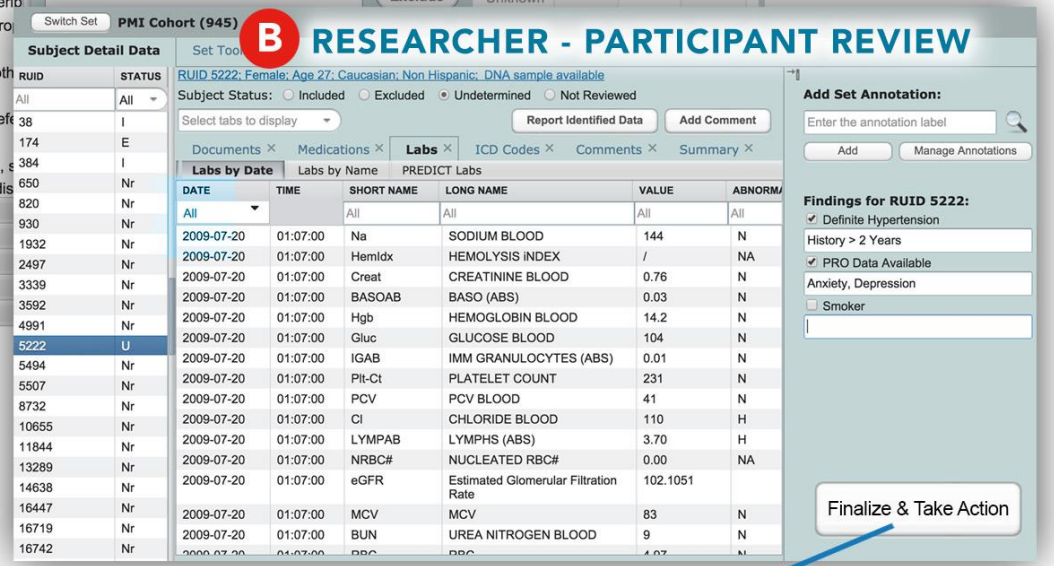


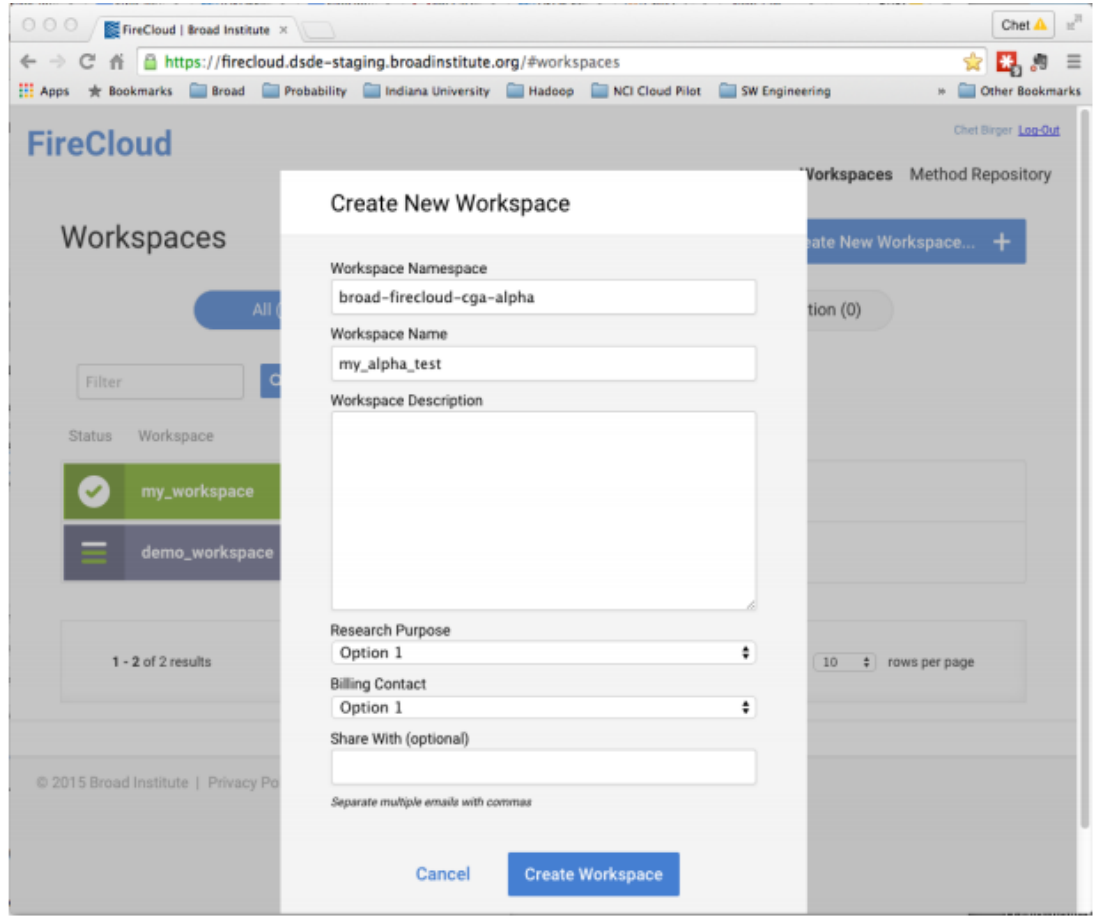
Building tools to facilitate research

A RESEARCHER - COHORT FILTER TOOL

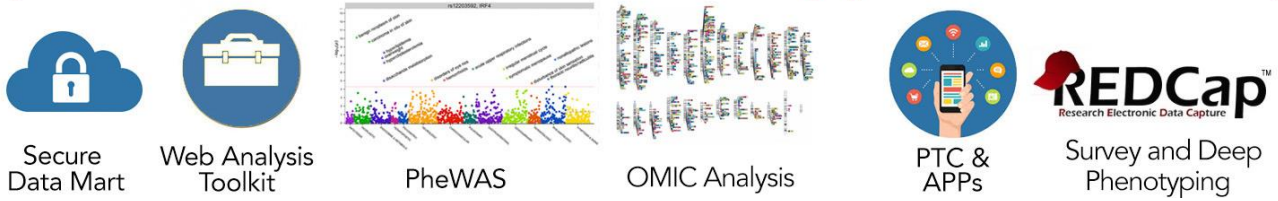


B RESEARCHER - PARTICIPANT REVIEW

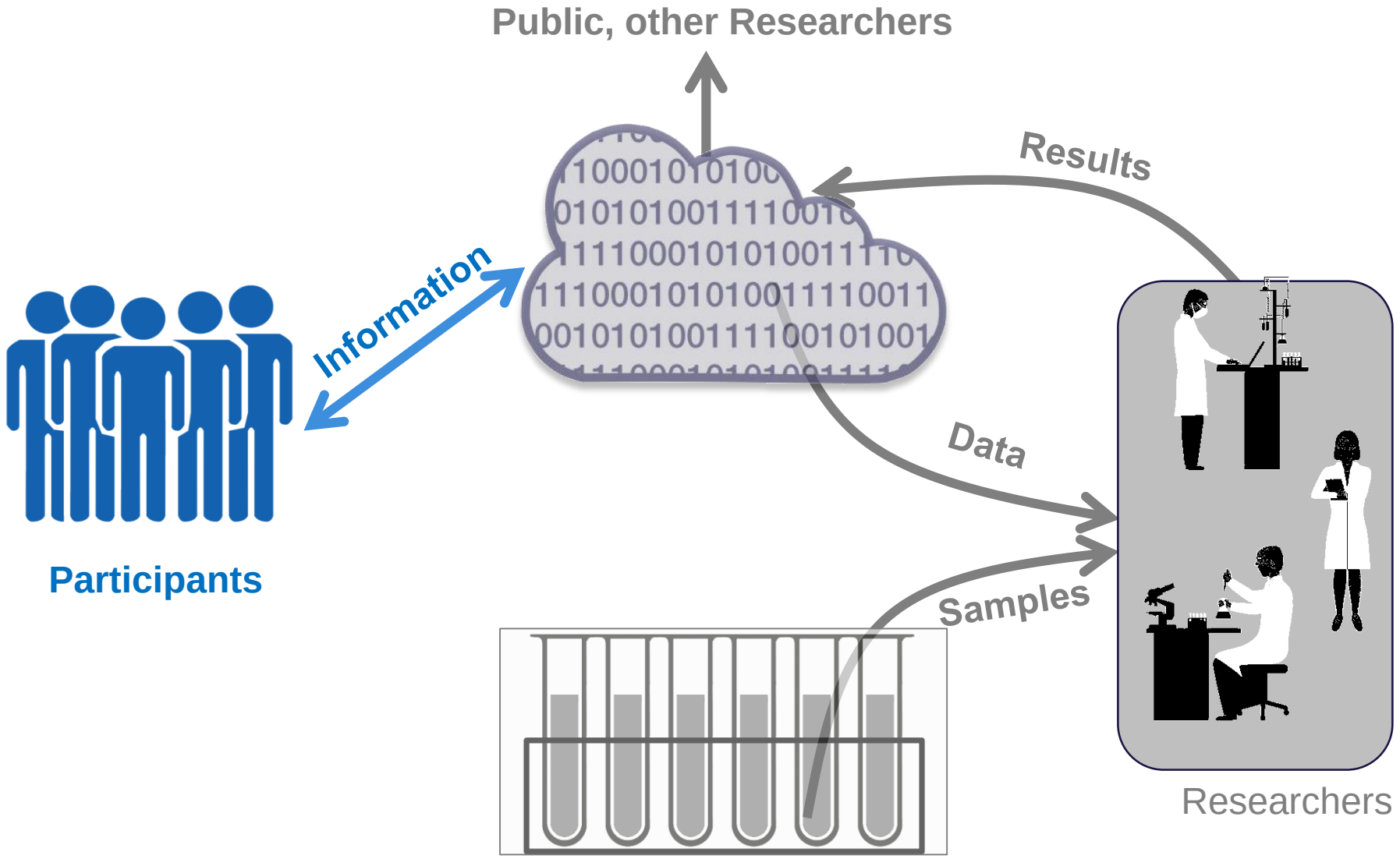




C CC ACTIVATED DATA SHARING & PARTICIPANT ENGAGEMENT



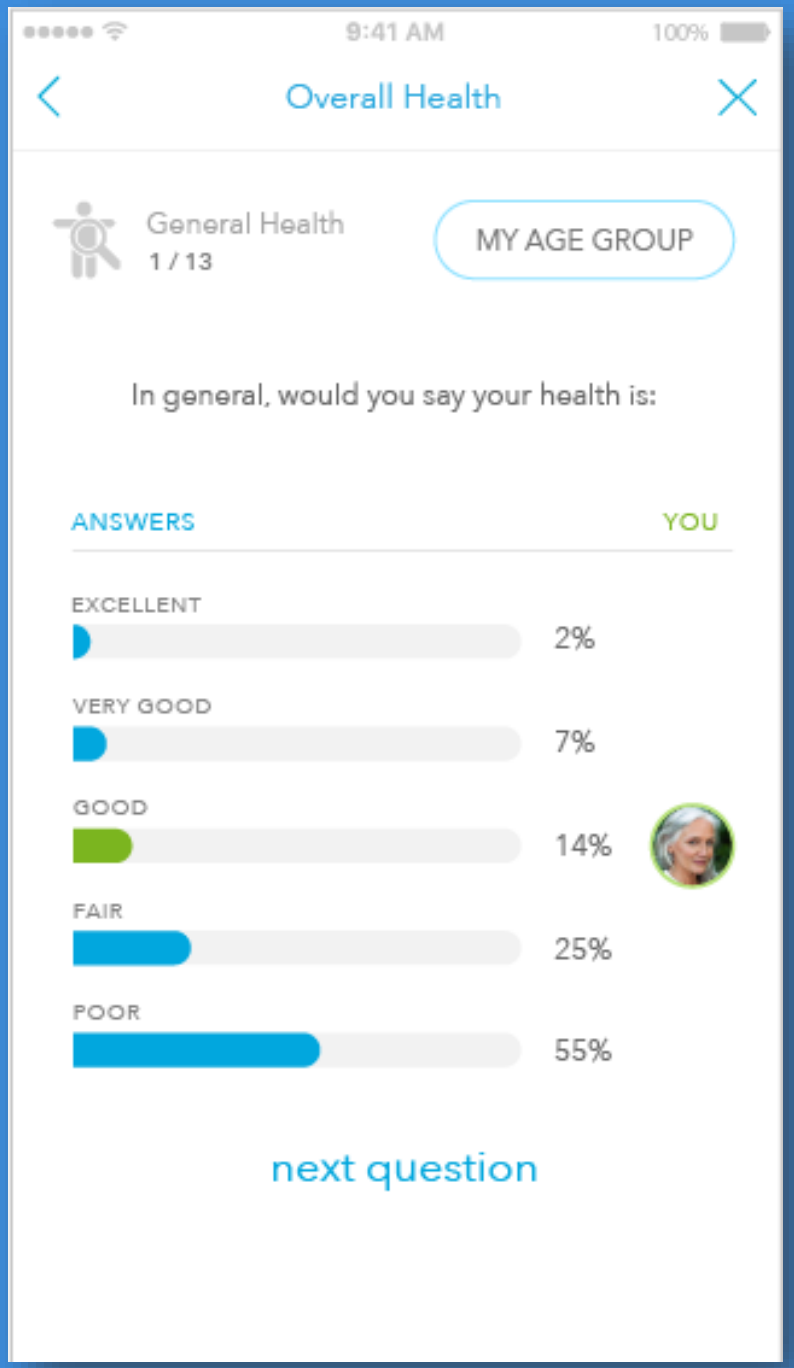
A key tenet of *All of Us*: participants will have access to their information



Return of Information

Participants may receive, depending on their preferences:

- Individual health information
- Survey data (comparative)
- EHR data, claims data
- Research results
- Ongoing study updates
- Aggregated results
- Scientific findings
- Opportunities to be contacted for other research opportunities



RESEARCH ARTICLE

HUMAN GENETICS

Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study

Frederick E. Dewey,[1*] Michael F. Murray,[2] John D. Overton,[1] Lukas Habegger,[1] Joseph B. Leader,[2] Samantha N. Fetterolf,[2] Colm O'Dushlaine,[1] Cristopher V. Van Hout,[1] Jeffrey Staples,[1] Claudia Gonzaga-Jauregui,[1] Raghu Metpally,[2] Sarah A. Pendergrass,[2] Monica A. Giovanni,[2] H. Lester Kirchner,[2] Suganthi Balasubramanian,[1] Noura S. Abul-Husn,[1] Dustin N. Hartzel,[2] Daniel R. Lavage,[2] Korey A. Kost,[2] Jonathan S. Packer,[1] Alexander E. Lopez,[1] John Penn,[1] Semanti Mukherjee,[1] Nehal Gosalia,[1] Manoj Kanagaraj,[1] Alexander H. Li,[1] Lyndon J. Mitnaul,[1] Lance J. Adams,[2] Thomas N. Person,[2] Kavita Praveen,[1] Anthony Marcketta,[1] Matthew S. Lebo,[3] Christina A. Austin-Tse,[3] Heather M. Mason-Suares,[3] Shannon Bruse,[1] Scott Mellis,[4] Robert Phillips,[4] Neil Stahl,[4] Andrew Murphy,[4] Aris Economides,[1] Kimberly A. Skelding,[2] Christopher D. Still,[2] James R. Elmore,[2] Ingrid B. Borecki,[1] George D. Yancopoulos,[4] F. Daniel Davis,[2] William A. Faucett,[2] Omri Gottesman,[1] Marylyn D. Ritchie,[2] Alan R. Shuldiner,[1] Jeffrey G. Reid,[1] David H. Ledbetter,[2] Aris Baras,[1] David J. Carey[2*]

The DiscovEHR collaboration between the Regeneron Genetics Center and Geisinger Health System couples high-throughput sequencing to an integrated health care system using longitudinal electronic health records (EHRs). We sequenced the exomes of 50,726 adult participants in the DiscovEHR study to identify ~4.2 million rare single-nucleotide variants and insertion/deletion events, of which ~176,000 are predicted to result in a loss of gene function. Linking these data to EHR-derived clinical phenotypes, we find clinical associations supporting therapeutic targets, including genes encoding drug targets for lipid lowering, and identify previously unidentified rare alleles associated with lipid levels and other blood level traits. About 3.5% of individuals harbor deleterious variants in 76 clinically actionable genes. The DiscovEHR data set provides a blueprint for large-scale precision medicine initiatives and genomics-guided therapeutic discovery.

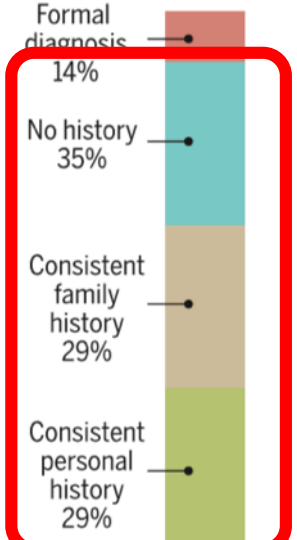
The application of high-throughput DNA | precision medicine requires further investigation

and DNA samples for a system-wide biorepository for broad research purposes, including genomic analyses, and linking to data in the GHS electronic health record (EHR). MyCode participants agree to be recontacted for additional phenotyping and return of clinically actionable results to inform their health care. The DiscovEHR cohort has clinical phenotypes recorded in the GHS EHR

B

	Variant positive/ total	Estimated prevalence
One clinically actionable genetic variant in G76	46/1415	1:31 (3.3%)
Two clinically actionable genetic variants in G76	3/1415	1:472 (0.2%)
Total	49/1415	1:29 (3.5%)

C



sociations between predicted LoF variants in lipid drug target genes ct, in SD units; whiskers denote 95% confidence intervals for effect. rriers. (**B** and **C**) Prevalence and expressivity of clinically actionable

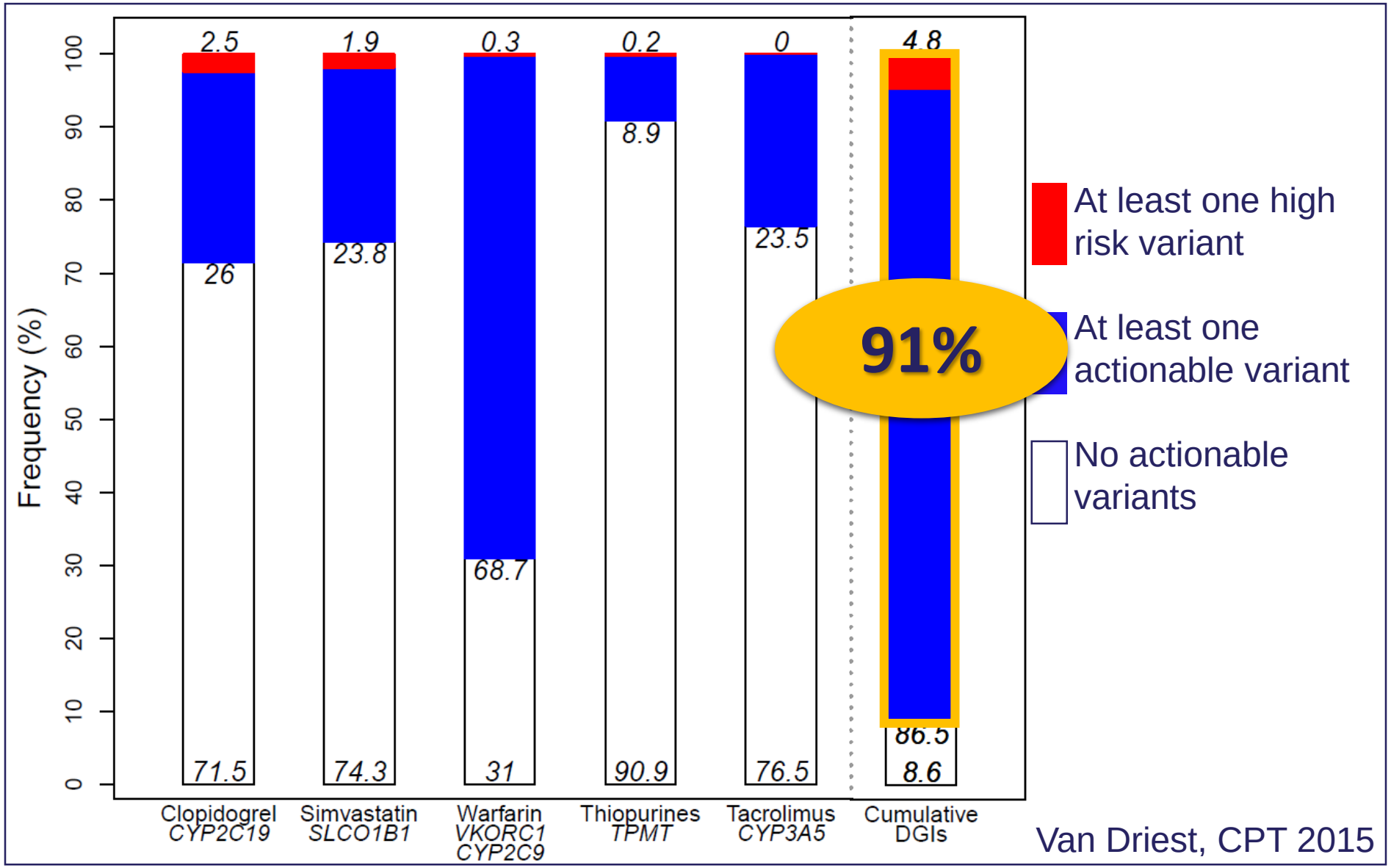
ilial relationships in this stable regional health care population. We therefore examined the extent of these relationships inferred from whole-exome sequence data using Pedigree Reconstruction and Identification of the Maximally Unrelated Set (PRIMUS) (*26*). Forty-eight percent of sequenced

3.5% of all tested had an actionable result

86% were new diagnoses

Dewey et al. *Science*. 2016

And… most people have an variant that would effect drug prescribing



All of Us timeline

January 2015	President's State of the Union Address announcing PMI
September 2015	NIH Advisory Committee to the Director (ACD) PMI Working Group report
July 2016	Initial awards
May 2017	IRB & FISMA security approvals
May 31, 2017	**Beta testing begins with initial participants**
	first set of participant provided information (PPI) surveys, baseline measures, biospecimen collections, EHR data
Now	**>13,400 people, first real initial EHR uploads, English and Spanish**
Through 2018	**Expand to national network, test & revise protocol, EHR uploads, develop additional survey modules, plans for genomics & pediatrics**
Spring 2018	**National launch**
Early 2019	**Initial researcher site launch**
Next 4-5 years	**Enroll 1M+ volunteers**
Next 10+ years	**Platform releases growing the data, tools, focus areas**

Scientific Priorities Workshop

The *All of Us* Research Priorities Workshop

- **Date and Location:** March 21-23, 2018, in Bethesda, MD
- **Purpose**: Identify key research priorities that will capitalize on the *All of Us* Research Program's one million or more participants to help ensure optimal value for advancing precision medicine.
- **Planning Committee:** Senior leaders across NIH's Institutes and Centers.
- **Workshop Participants:** A broad array of stakeholders (e.g., researchers, participants, professional societies, advocacy groups); by invitation.
- **Outreach Plans:** Obtain substantial input on research questions and requirements prior to the workshop.
- **Updates:** You can subscribe at https://www.joinallofus.org/news-and-events

All of Us RESEARCH PROGRAM | The Precision Medicine Initiative

Three Big Questions for the workshop

1. **Near Term**

   What are low-hanging fruit questions/measures for which the scale of *All of Us* could help accelerate knowledge & breakthroughs in precision medicine?

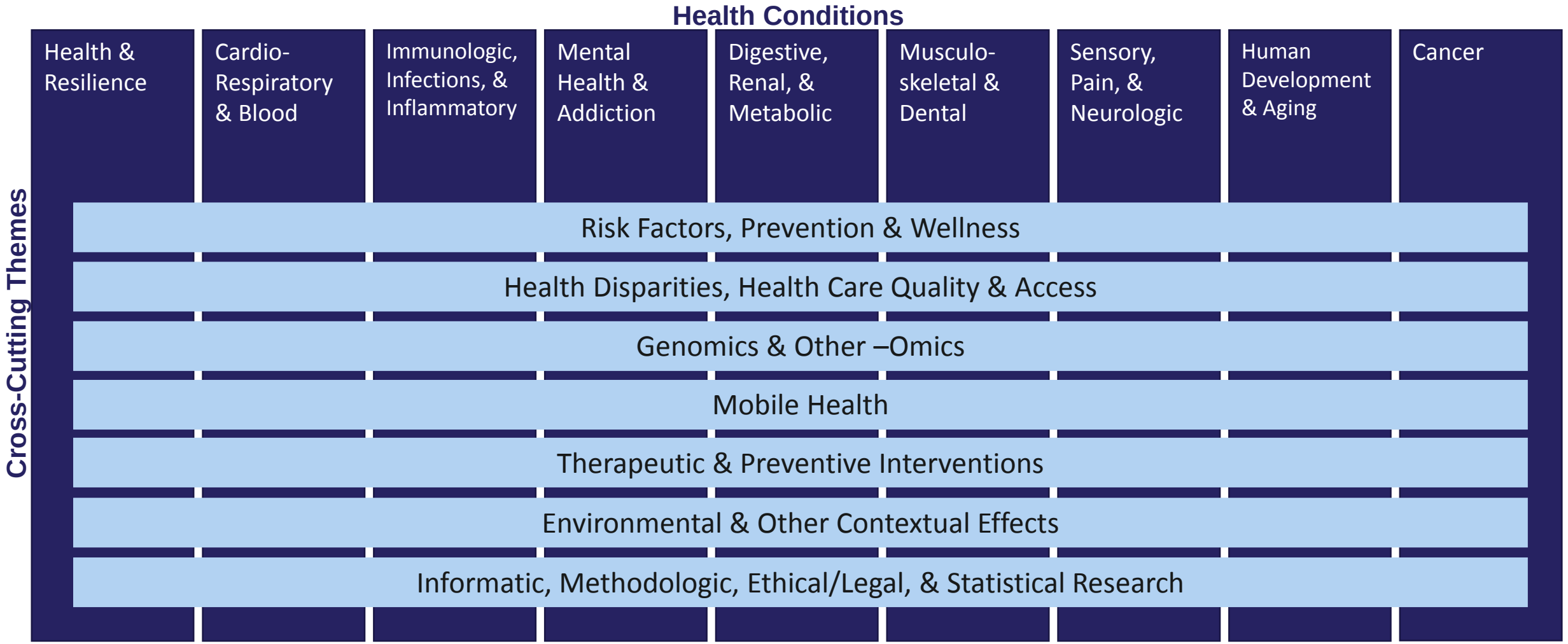
2. **Mid-Term**

   What kinds of questions might this Program answer where additional work selecting among measures/instruments is needed?

3. **Long Term**

   What kinds of questions are ripe for a program of this size but for which we need fundamental science & tech to develop the instruments and methods?

Scientific Framework for the *All of Us* March 2018 Workshop

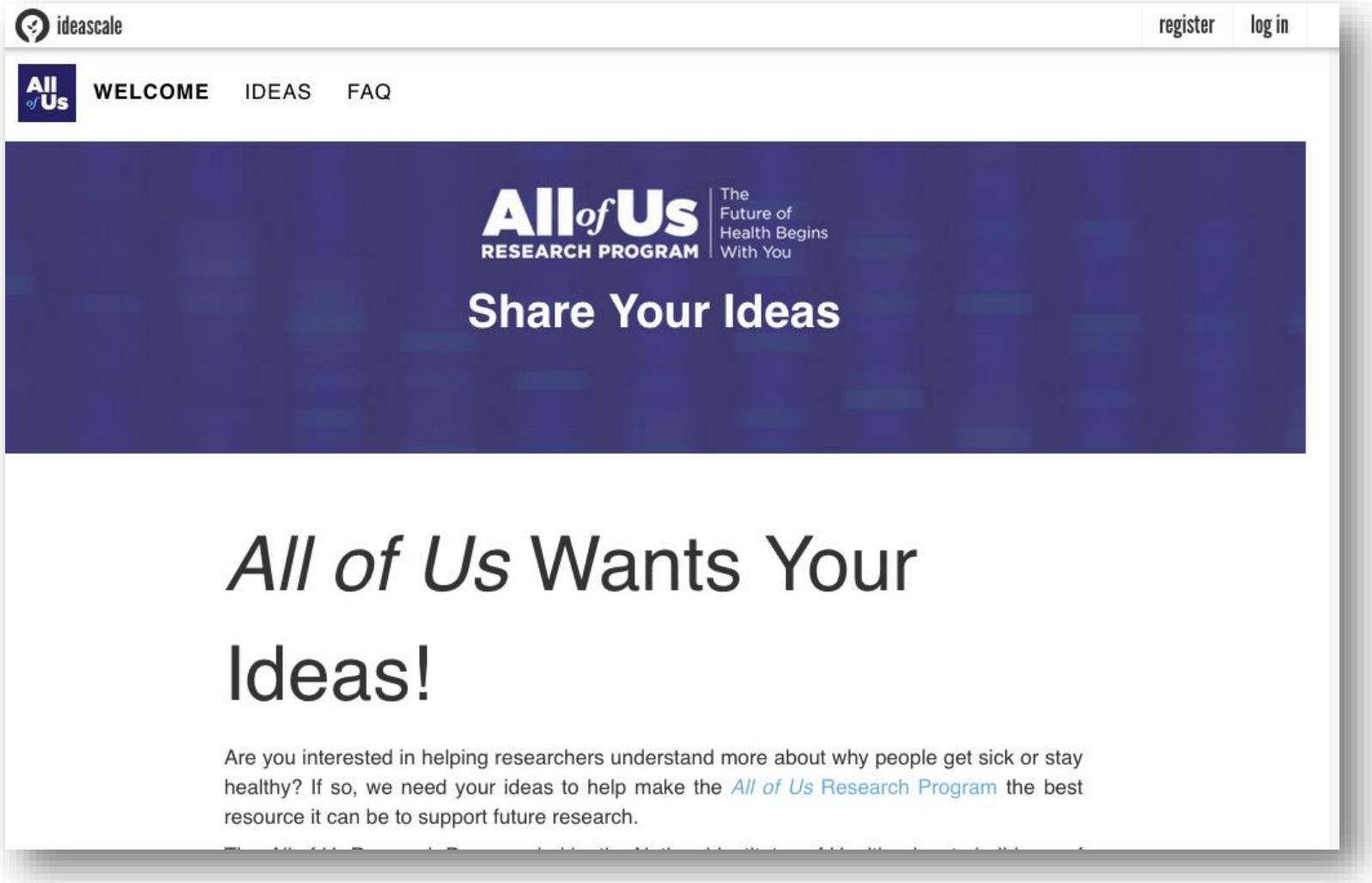


Developed with input from NIH Institutes & Centers

We need your input! Crowdsourcing Use Cases

https://allofusresearchpriorities.ideascale.com/

- Register (easy!)
- Submit use cases
- Read use cases
- Comment on use cases
- Vote on use cases



- **NIH *All of Us* website**: https://allofus.nih.gov
- **Enrollment site**: JoinAllofUs.org
- **Follow us on social media**: @AllofUsResearch, #JoinAllofUs

Thank you