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.

- Healthy
  - Drug beneficial

- Disease risk, self-limited
  - Drug beneficial, mild toxicity

- Severe disease
  - Drug toxic

- Atypical disease
  - Drug does nothing
Framingham Heart Study

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

“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

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

**Network formed – July 2016**

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

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

September 17, 2015


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.

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.

Our mission
To accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care for all of us.
Some of the *All of Us* research goals

- Studying exposures and habits
- Improving drug action
- New biomarkers (and their interactions)
- Engaging diverse participants
- Machine learning on Big Data
- New precision therapies & targeted clinical trials
Why Diversity?

PERSISTENT BIAS
Over the past seven years, the proportion of participants in genome-wide association studies (GWAS) that are of Asian ancestry has increased. Groups of other ancestries continue to be very poorly represented.

2009
373 studies
1.7 million samples
96% European ancestry
4% Non-European ancestry

2016
2,511 studies
35 million samples
81% European ancestry
19% Non-European ancestry

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

Illinois Precision Medicine Consortium

University of Pittsburgh

New York City Precision Medicine Consortium

California Precision Medicine Consortium

University of Arizona (w/Banner Health)

New England Precision Medicine Consortium

San Ysidro Health Center

Federal Partners: White House, HHS, NIH, ONC, HRSA, VA, USDS

Jackson-Hinds Comprehensive Health Center

Vibrent (Participant Technology Systems Center)

Southern All of Us Network

SouthEast Enrollment Center

Vanderbilt Univ. Medical Center, with Broad & Verily (Data and Research Center)

Wondros

Federal Partners: White House, HHS, NIH, ONC, HRSA, VA, USDS

California Precision Medicine Consortium

Scripps Translational Science Institute (Participant Center)

Community Health Center, Inc.
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
- **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
**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

- 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
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

**Researcher Portal, Environments, and APIs**
- FireCloud
- Cohort Builder
- Public resources
- ... (in development)

**Centrally linked data**
- Death Index
- Claims & Rx Data
- Geospatial & Environ.
- ... (in development)

**Curated Data Repository**
- Aligned with algorithms, standards, data linkages, removal of clear identifiers

**Raw Data Repository**
- EHR (via S4S) (in development)

**Participant Portal (PTSC)**
- Registration
- Consent
- PPI

**HPO and DV Staff Tools**
- EHR Upload
- Operational Dashboards
- HealthPro Portal
  - Physical Measurements
  - Biobanking
  - Work Queue

**v1 in production**
- in development
Sync 4 Science (S4S) – a technology to share health data

S4S Pilot Sites
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

**From Direct Volunteers**
- Sync for Science (2018-2019)
  - FHIR, OMOP, ???

**Raw Data Repository**

**Curated Data Repository (OMOP)**

**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

Leverage the “Web dividend” for Health.
- Computable phenotypes
- Natural language processing
- Imputation
- Precomputed analyses
- Crowdsourcing
- Extracting geocoded environmental info

Automatically captioned: “Two pizzas sitting on top of a stove top oven”
AOU centralizes data to **enhance security** and **improve usefulness**

### Traditional Approach

*Bring data to researchers*

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

### AoU Approach

*Bring researchers to the data*

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

- **Public**
  - Access Data
  - no login required

- **Registered**
  - Registration and Identity Verification
  - Research Ethics Training
  - Sign Code of Conduct
  - Resource Access Board Approval
  - Create Project and State Purpose
  - Access Data
  - Names and Projects posted on a public website

- **Controlled**
  - Registration and Identity Verification
  - Research Ethics Training
  - Sign Code of Conduct
  - eRA Commons ID
  - Resource Access Board Approval
  - Create Project and State Purpose
  - Access Data
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
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 Lucas Habegger,1 Joseph B. Leader,2 Samantha N. Fetterolf,3 Colm Ó'Dusháilíne,1
Cristeropher V. Van Hout,1 Jeffrey Staples,1 Claudia Gonzalez-Jarqueg,1 Raghu Mettapa,2 Sarah A. Pendergrass,1 Monica A. Giovann,3 H. Lester Kirchner,3
Suganthi Balasubramaniam,1 Noura S. Abu-Hus,1 Dustin N. Hartzel,2
Daniel R. Lavage,6 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. Milsnau,1 Lance J. Adams1, Thomas N. Person,1 Kavita Praveen,1
Anthony Mareketa,1 Matthew S. Lebo,1 Christina A. Austen-Tse,1
Heather M. Mason-Saures,7 Shannon Bruse,1 Scott Mellis,1 Robert Phillips,8
Nell Stahl,4 Andrew Murphy,4 Aris Economides, Kimberly A. Skelding,8
Christopher D. Still,9 James R. Elmore,1 Ingrid B. Boirecki,9 George D. Yancopoulos,9
F. Daniel Davis,9 William A. Faucett,8 Omri Gottesman,1 Marylyn D. Ritchle,1
Alan R. Shulidner,1 Jeffrey G. Reid,1 David H. Ledbetter,1 Aris Baras,1 David J. Carey2*

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.

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 phenotype and return of clinically actionable results to inform their health care. The DiscovEHR cohort has clinical phenotypes recorded in the GHS EHR. A median of 14 years with a median of 12 years.

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

3.5% of all tested had an actionable result
86% were new diagnoses

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

Van Driest, CPT 2015
### All of Us timeline

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

### 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

<table>
<thead>
<tr>
<th>Cross-Cutting Themes</th>
<th>Health Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health &amp; Resilience</td>
<td>Risk Factors, Prevention &amp; Wellness</td>
</tr>
<tr>
<td>Cardio-Respiratory &amp; Blood</td>
<td>Health Disparities, Health Care Quality &amp; Access</td>
</tr>
<tr>
<td>Immunologic, Infections, &amp; Inflammatory</td>
<td>Genomics &amp; Other –Omics</td>
</tr>
<tr>
<td>Mental Health &amp; Addiction</td>
<td>Mobile Health</td>
</tr>
<tr>
<td>Digestive, Renal, &amp; Metabolic</td>
<td>Therapeutic &amp; Preventive Interventions</td>
</tr>
<tr>
<td>Musculo-skeletal &amp; Dental</td>
<td>Environmental &amp; Other Contextual Effects</td>
</tr>
<tr>
<td>Sensory, Pain, &amp; Neurologic</td>
<td>Informatic, Methodologic, Ethical/Legal, &amp; Statistical Research</td>
</tr>
<tr>
<td>Human Development &amp; Aging</td>
<td>Cancer</td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
</tr>
</tbody>
</table>

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

All of Us Wants Your Ideas!

Are you interested in helping researchers understand more about why people get sick or stay healthy? If so, we need your ideas to help make the All of Us Research Program the best resource it can be to support future research.
NIH All of Us website: https://allofus.nih.gov

Enrollment site: JoinAllofUs.org

Follow us on social media: @AllofUsResearch, #JoinAllofUs

Thank you