The *All of Us* Research Program

Sheri Schully, Ph.D. (she/her)
Deputy Chief Medical and Scientific Officer
*All of Us* Research Program

April 5, 2024
The **All of Us** Research Program is a historic, longitudinal effort to **gather data from one million or more people** living in the United States to **accelerate research and improve health**. By taking into account individual differences in **lifestyle, socioeconomics, environment, and biology**, we hope that researchers will one day uncover paths toward delivering **precision medicine** – or individualized prevention, treatment, and care – for all of us.

The **All of Us** Research Program is part of the broader **Precision Medicine Initiative**.
We set core values for the program at its inception. Our Core Values fuel our purpose, shape our vision, and guide the implementation of our mission.

- Participation is **open** to all.
- Participants have **access** to their information.
- Participants reflect the rich **diversity** of the U.S.
- Data will be accessed **broadly** for research purposes.
- Participants are **partners**.
- Security and privacy will be of highest importance.
- Trust will be earned through **transparency**.
- The program will be a catalyst for **positive change** in research.
The All of Us Research Program Mission

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

- **Nurture partnerships** for decades with at least a million participants who reflect the diversity of the U.S.
- **Deliver one of the largest, richest biomedical datasets** that is broadly available and secure
- **Catalyze an ecosystem of communities, researchers, and funders** who make All of Us an indispensable part of health research
Enrolled 775K+ Participants With Continued Growth

Numbers current as of March 11, 2024

Participant Enrollment

- 775,000+ Participants
- 431,000+ Electronic Health Records
- 533,000+ Participants who have completed initial steps of the program
- 550,000+ Biosamples

Map of Consented Participants

Plus >4,200 consented participants across U.S. territories
Prioritizing Intersectionality & Social Determinants of Health

Underrepresented in Biomedical Research (UBR)

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/Ethnicity (non-white race)</td>
<td>47.4%</td>
</tr>
<tr>
<td>Healthcare Access (inadequate access)</td>
<td>27.6%</td>
</tr>
<tr>
<td>Age at Consent (&lt;18 and 65+)</td>
<td>26.3%</td>
</tr>
<tr>
<td>Income (annual income ≤ $25k)</td>
<td>23.3%</td>
</tr>
<tr>
<td>Disability</td>
<td>19.9%</td>
</tr>
<tr>
<td>Sexual and Gender Minorities</td>
<td>10.3%</td>
</tr>
<tr>
<td>Sexual Orientation (not straight)</td>
<td>10.0%</td>
</tr>
<tr>
<td>Educational Attainment (less than GED)</td>
<td>8.7%</td>
</tr>
<tr>
<td>Geography (rural area residents)</td>
<td>7.7%</td>
</tr>
<tr>
<td>Gender Identity (not man or woman)</td>
<td>1.2%</td>
</tr>
<tr>
<td>Sex Assigned at Birth (intersex)</td>
<td>0.02%</td>
</tr>
</tbody>
</table>

Source: All of Us Research Program Data Snapshots (Data current as of March 11, 2024)

All of Us Research Program

- Asian 3.2%
- BAA 16.8%
- H/L 16%
- MENA, NHPI, or None fully describe me 2.7%
- White 52.5%
- Prefer not to say 0.5%
- More than one race 6.9%

Source: All of Us Research Program Data Snapshots, Race and Ethnicity (Updated February 6, 2024)

Other Genomic Studies

- Asian 3.72%
- African American or Afro-Caribbean 0.5%
- African 0.2%
- H/L 0.35%
- Other/More than one 0.67%
- White/European Descent 94.56%

Source: Global Genome Wide Association Studies (GWAS) (Updated January 2024)
<table>
<thead>
<tr>
<th>Electronic Health Records</th>
<th>Data types collected from EHR include:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Demographics</td>
</tr>
<tr>
<td></td>
<td>• Vital signs</td>
</tr>
<tr>
<td></td>
<td>• Diagnoses</td>
</tr>
<tr>
<td></td>
<td>• Procedures</td>
</tr>
<tr>
<td></td>
<td>• Medications</td>
</tr>
<tr>
<td></td>
<td>• Doctor and Laboratory Visits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant Surveys</th>
<th>The Basics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Health Care Access &amp; Utilization</td>
</tr>
<tr>
<td></td>
<td>Overall Health</td>
</tr>
<tr>
<td></td>
<td>Personal and Family Medical History</td>
</tr>
<tr>
<td></td>
<td>Lifestyle</td>
</tr>
<tr>
<td></td>
<td>Social Determinants of Health</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical Measurements</th>
<th>• Blood pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Heart rate</td>
</tr>
<tr>
<td></td>
<td>• Height</td>
</tr>
<tr>
<td></td>
<td>• Weight</td>
</tr>
<tr>
<td></td>
<td>• BMI</td>
</tr>
<tr>
<td></td>
<td>• Hip circumference</td>
</tr>
<tr>
<td></td>
<td>• Waist circumference</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biosamples</th>
<th>• Blood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Saliva</td>
</tr>
<tr>
<td></td>
<td>• Urine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wearable Data</th>
<th>Fitbit data, including:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Heart Rate</td>
</tr>
<tr>
<td></td>
<td>• Activity (Daily Summary)</td>
</tr>
<tr>
<td></td>
<td>• Activity Intraday Steps</td>
</tr>
<tr>
<td></td>
<td>• Sleep data</td>
</tr>
</tbody>
</table>

Mental Health and Well-Being
Data Collected and Return of Value to All of Us Participants

Consent and Electronic Health Records

Participant Surveys

Physical Measurements

Biosamples

Mobile/Wearable Tech

Return of Value for Participants

Participants may receive:

- **Genetic information**
- Survey data (comparative)
- EHR and claims data
- Ongoing study updates
- Aggregate results
- Scientific findings
- Opportunities to be contacted for other research opportunities
Participants Can Receive Four Types of Genetic Research Results

- **Genetic ancestry and traits results**
  - 7 regions (20 subregions) and 4 traits
  - Sub-Saharan Africa
  - Europe
  - Oceania
  - Southern Asia
  - Eastern and northern Asia
  - The Middle East and North Africa
  - The Americas
  - Ear wax
  - Bitter taste perception
  - Cilantro preference
  - Lactose intolerance

- **Medicine and your DNA Report**
  - 7 pharmacogenomics (PGx) genes and 50+ medications
  - CYP2C19
  - DPYD
  - G6PD
  - SLCO1B1
  - NUDT15
  - TPMT
  - UGT1A1

- **Hereditary Disease Risk (HDR) Report**
  - 59 genes (SNVs + indels, e.g. no SVs)
  - Breast cancer
  - Ovarian cancer
  - Uterine cancer
  - Colorectal cancer
  - Prostate cancer
  - Melanoma
  - Brain cancer
  - Pancreatic cancer
  - Stomach cancer
  - Familial hypercholesterolemia
  - Cardiomyopathies
  - Arrhythmias
  - Arteriopathies
Genomic Health-Related Return of Results (as of March 20, 2024)

Hereditary Disease Risk

All of Us currently looks for genetic variants in 59 genes associated with serious health conditions.

- **Your result:**
  - Something very important for your health was found in your BRCA1 gene.

  
- **What does this mean?**
  - If confirmed by a clinical DNA test, this result means that you are more likely to get some types of cancers than other people.
  - It does not mean that you have some types of cancers.
  - It does not mean that you will definitely get some types of cancers.
  - This result is important and should not be ignored.

- **Important!**
  - This report comes from a research program, so it is a research result. Your doctor will need to confirm these results with a clinical DNA test before using them in your care.
  - Do not change your medical care before this result is confirmed by your doctor.
  - Results provided are from an investigational

- **222k+ offered choice**
  - 122.5k+ (53%) said “yes”
  - 99.5k+ viewed results
  - 2.9% with actionable result

- **222k+ offered choice**
  - 116k+ (52%) said “yes”
  - 95.4k+ viewed results
  - >96% with actionable result

Medicine and Your DNA

All of Us analyzes seven genes that can affect how bodies metabolize medicines.

- **Medicine and your DNA**
  - They do that in many different ways. Some genes help move medicines to the right part of the body.
  - This test looked at a few of the genes in your DNA that can affect how medicines are used. The technical term for this kind of information is “pharmacogenetics.”
  - Doctors and pharmacists use this kind of information when they consider why medicines work differently for different people.
  - But doctors and pharmacists don’t make decisions based on just DNA. Some other important considerations can be age, weight, health, diet, and other medicines you are taking at the same time.

- **290k+ offered choice**
  - 171k+ (63%) said “yes”
  - 168k+ viewed results

Genetic Ancestry and Traits

All of Us provides genetic ancestry details for 7 regions, and information on four genetic traits.

- **DNA Results**
  - You’ll see all of your DNA results here when they’re ready. See options for your DNA results.
  - **Genetic ancestry and trait results**
  - **Genetic ancestry**
  - Genetic ancestry can be very interesting, but you may also learn information you didn’t expect.
  - [Learn more]
Leveraging the *All of Us* Data
Registered researchers can access in-depth data and a variety of research tools to conduct a wide range of studies.

**PUBLIC TIER**

Public resources include:

- **Data Snapshots**: Aggregated, public-facing overviews of participant characteristics and data types
- **Data Browser**: Interactive preview into the All of Us dataset through public-facing aggregate data
  - Currently includes participant-provided survey responses, physical measurements, data from EHRs and wearables, and genomic data
- **Survey Explorer**: Details the questions included in each of the surveys
- **Research Projects Directory**: Descriptions of each research project within the Researcher Workbench

**REGISTERED TIER**

Registered researchers with amended institutional agreements can access all of the data in the Registered Tier plus additional and expanded data types, including genomic data, real dates of health events, ICD codes, granular demographic data, and more.

**CONTROLLED TIER**

Registered researchers can access all of the data in the Registered Tier plus additional and expanded data types, including genomic data, real dates of health events, ICD codes, granular demographic data, and more.

Data have been processed to protect participant privacy
Nearly 250,000 Whole Genome Sequences Available to Advance Precision Medicine

The All of Us Researcher Workbench contains the one of the largest sets of whole genome sequences widely available for research.

- **413,350+** Survey Responses
- **337,500+** Physical Measurements
- **312,900+** Genotyping Arrays
- **287,000+** Electronic Health Records
- **245,350+** Whole Genome Sequences
- **11,350+** Structural Variants
- **1,000+** Long-Read Sequences
- **15,600+** Fitbit Records

NEW! Sleep Data

The whole genome sequence dataset includes variation at more than **1 billion** locations, which is nearly **one-third** of the entire human genome.
Summary statistics of:

- EHR Data (Conditions, Drug Exposures, Lab & Measurements, Procedures)
- Genomic Variants
- Survey Questions (including COVID-19 surveys)
- Physical Measurements
- Open Access (no login required)
**Summary statistics of:**

- EHR Data (Conditions, Drug Exposures, Lab & Measurements, Procedures)
- Genomic Variants
- Survey Questions (including COVID-19 surveys)
- Physical Measurements
- **Open Access (no login required)**
### Summary statistics of:

- **EHR Data (Conditions, Exposures, Lab & Measurements, Procedures)**
- **Genomic Variants**
- **Survey Questions (including COVID-19 surveys)**
- **Physical Measurements**
- **Open Access (no login required)**

#### Data Browser

<table>
<thead>
<tr>
<th>Question</th>
<th>Concept Code</th>
<th>Participant Count</th>
<th>% Answered out of 185240</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you or anyone in your family ever been diagnosed with the following cancer conditions? Think only of the people you are related to by blood. - Breast cancer</td>
<td>43528499</td>
<td>19,300</td>
<td>10.42%</td>
</tr>
<tr>
<td>Have you or anyone in your family ever been diagnosed with the following cancer conditions? Think only of the people you are related to by blood. - Skin cancer</td>
<td>43529816</td>
<td>15,700</td>
<td>8.48%</td>
</tr>
<tr>
<td>Have you or anyone in your family ever been diagnosed with the following cancer conditions? Think only of the people you are related to by blood. - Lung cancer</td>
<td>43529183</td>
<td>11,600</td>
<td>6.26%</td>
</tr>
<tr>
<td>Have you or anyone in your family ever been diagnosed with the following cancer conditions? Think only of the people you are related to by blood. - Colon cancer/Rectal cancer</td>
<td>43528564</td>
<td>11,460</td>
<td>6.19%</td>
</tr>
<tr>
<td>Have you or anyone in your family ever been diagnosed with the following cancer conditions? Think only of the people you are related to by blood. - Prostate cancer</td>
<td>43529732</td>
<td>10,260</td>
<td>5.54%</td>
</tr>
</tbody>
</table>
All of Us Researcher Workbench: Access to Row-Level Data for Analysis

**Researcher Workbench**

- Cloud-based central resource for broad data accessibility
- Graphical tools for cohort selection and data set building
- R and Python in Jupyter notebooks; SAS available soon
- **Passport access model for approved researchers** - just create, describe your workspace, and get to work! **No separate IRB approval needed.**
- Currently open to academic, health care and not-for-profit organizations

[ResearchAllofUs.org/Data-Tools/Workbench/](ResearchAllofUs.org/Data-Tools/Workbench/)
### All of Us Data Includes Range of Diseases

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Domain</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Disease</td>
<td>Heart</td>
<td>89,180</td>
</tr>
<tr>
<td>Obesity</td>
<td>Endocrine</td>
<td>80,980</td>
</tr>
<tr>
<td>Type 1 Diabetes</td>
<td>Endocrine</td>
<td>6,740</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>Endocrine</td>
<td>56,340</td>
</tr>
<tr>
<td>Any cancer</td>
<td>Cancer</td>
<td>50,720</td>
</tr>
<tr>
<td>Asthma</td>
<td>Pulmonary</td>
<td>50,160</td>
</tr>
<tr>
<td>Chronic Obstructive Pulmonary Disorder</td>
<td>Pulmonary</td>
<td>24,940</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>Neuro</td>
<td>9,320</td>
</tr>
<tr>
<td>Stroke</td>
<td>Neuro</td>
<td>940</td>
</tr>
<tr>
<td>Rheumatoid Arthritis</td>
<td>Autoimmune</td>
<td>8,660</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>Autoimmune</td>
<td>99,880</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Domain</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressive Disorder</td>
<td>Mental Health</td>
<td>82,660</td>
</tr>
<tr>
<td>Bipolar Disorder</td>
<td>Mental Health</td>
<td>16,200</td>
</tr>
<tr>
<td>Dementia</td>
<td>Mental Health</td>
<td>6,020</td>
</tr>
<tr>
<td>Human Immunodeficiency Virus</td>
<td>Infectious Disease</td>
<td>5,500</td>
</tr>
<tr>
<td>COVID-19*</td>
<td>Infectious Disease</td>
<td>58,000*</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>Abuse</td>
<td>20,380</td>
</tr>
<tr>
<td>Opioid Usage</td>
<td>Medication</td>
<td>155,000</td>
</tr>
<tr>
<td>Age-Related Macular Degeneration</td>
<td>Eye</td>
<td>4,740</td>
</tr>
<tr>
<td>Hearing loss</td>
<td>Hearing</td>
<td>30,400</td>
</tr>
<tr>
<td>Falls</td>
<td>Aging/Nursing</td>
<td>4,860</td>
</tr>
</tbody>
</table>

*using combination of diagnosis code, lab test, and COVID survey answers
EHR Data on the Researcher Workbench is Diverse and Longitudinal

- **Surveys**: 79 million survey answers
- **Procedures**: 32 million procedures
- **Conditions**: 84 million conditions
- **Measurements**: 333 million labs and measurements
- **Drugs**: 78 million drug exposures
- **Fitbit**: 42 billion Fitbit data points
Doubled DHT and EHR Connections in 2023

as of 12/11/2023
What Kind of Research Can *All of Us* Support?

**Example studies in *All of Us***
- Associations between diseases, medications, behaviors, SDOH, genomics
- Health disparities
- Historically underrepresented populations
- Genomics and PGx
- Drug target discovery
- Early disease detection
- Geospatial linkages (future)
- *Insert your topic here*

**Modalities of research *All of Us* supports***
- AI/ML
- Risk stratification
- Predictive analytics
- Phenotype algorithms & cohort development
- Novel method development
- Basic EHR investigations
- Validation of other studies

**All of Us is not**
- A representative US sample
- A study with uniform follow-up of all variables
All of Us Researcher Workbench: Access to Row-Level Data for Analysis

**STEP 1**
CONFIRM YOUR INSTITUTION’S AGREEMENT
Before you can create an account, your institution must have a Data Use and Registration Agreement (DURA) in place with All of Us. Confirm DURA.

**STEP 2**
CREATE AN ACCOUNT AND VERIFY IDENTITY
After creating your Researcher Workbench account, you will be asked to verify your identity through login.gov. Learn more.

**STEP 3**
COMPLETE THE MANDATORY TRAINING
The training focuses on conducting responsible and ethical research using the Researcher Workbench. Additional training is required to access the Controlled Tier. Learn more.

**STEP 4**
SIGN THE DATA USER CODE OF CONDUCT (DUCO)
This agreement outlines the program’s expectations for researchers who use the Researcher Workbench and describes how program data may be used. View the DUCO.
All of Us Research Program’s Commitment to Researcher Diversity (data as of March 1, 2024)

- Creating a diverse researcher cohort that promotes responsible and ethical use of data, returns value to participant communities, and accelerates research impact.

- Encouraging student assemblies and early-stage investigators to bring fresh, creative perspectives & innovative research outcomes.

- Ensuring access for researchers from various institutions/organizations to establish a truly equitable resource for all.

Over 85% of our researchers are underrepresented in the biomedical workforce - including over 30% diverse by race and ethnicity
Growing Scientific Impact

The association of anxiety with granuloma annulare: a case–control study of the National Institutes of Health ‘All of Us’ research programme

Health disparities in the treatment of bipolar disorder

Vladimir Trhlikrov, Mark E. Ladner, Felicia V. Caples, Mitzi Morris, Hailey Spillers, Christina D. Jordan, Joyce E. Bulbs-Berry, Monica J. Taylor-Desir, Mark A. Frye, Eric J. Vallender

Perspective
“All of Us” and the Promise of Precision Medicine: Achieving Equitable Access for Federally Qualified Health Center Patients

Carolyn F. Neuhaus, Danielle M. Pacia, Johanna T. Crane, Karen J. Maschke, and Nancy Berlinger

Evaluating Discrepancies in Self-Reported Glaucoma and Electronic Health Records in the National Institutes of Health All of Us Database

March 20, 2023

Daily Step Counts Before and After the COVID-19 Pandemic Among All of Us Research Participants

researchallofus.org/publications/
The Future and Opportunities for Researchers
The Future of *All of Us*

**Pediatrics**

More Genomics in late 2024
About 400,000 whole genome sequences
About 1,500 long read whole genome sequences

More Digital Health Technology
Wearable data from diverse participants

Collecting New Data
Linkages to additional data sources: Environmental data, Claims, Cancer Registries and more

Ancillary Studies
Incorporating complex questions about how genomic, environmental, and social interactions influence human health

Learn More about the Data
Making Health Discoveries Possible

The *All of Us* Program wouldn't be possible without the generosity of our participants and the dedication of our researchers to enable health discoveries.
Thank You!

AllofUs.nih.gov
https://www.researchallofus.org/explore