The *All of Us* Research Study - Data Types, Data Access, and Data Analysis

SITC-NCI Computational Immuno-Oncology Webinar Series
December 7, 2022

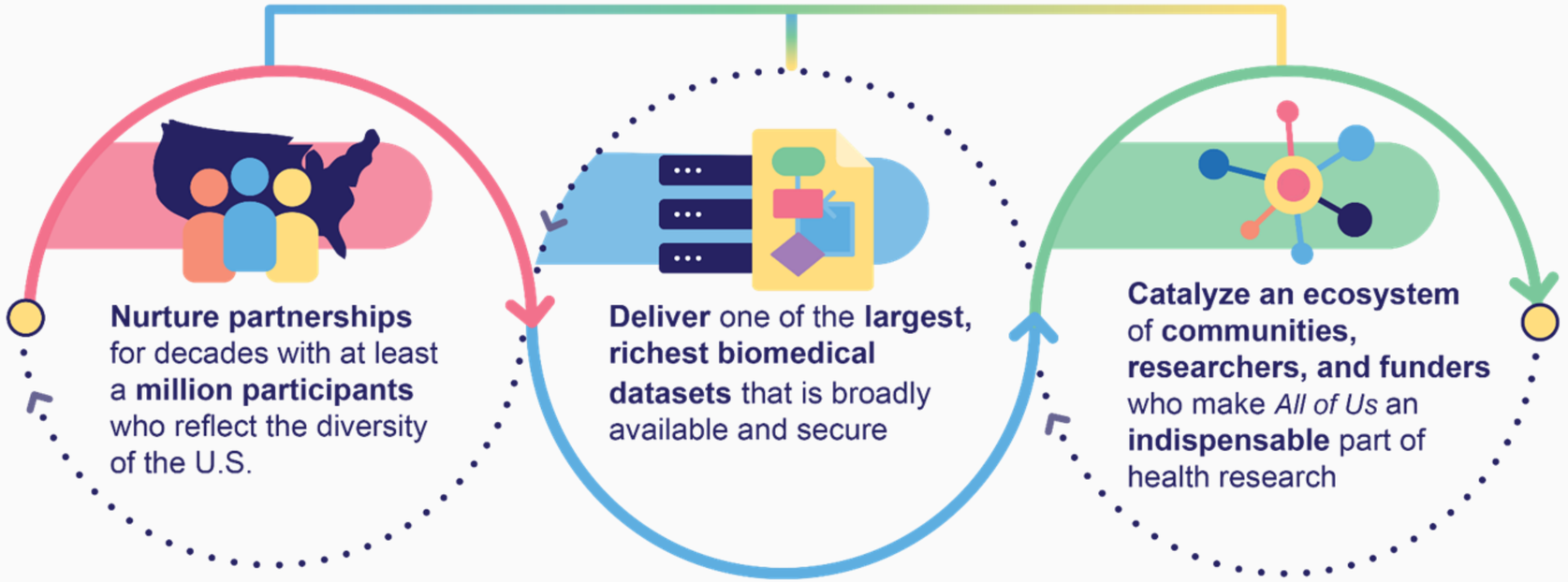
Sheri D. Schully, Ph.D.
Deputy Chief Medical and Scientific Officer
***All of Us* Research Program**

National Institutes of Health

All of Us Research Program Mission

Our Mission

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



Made possible by a team that maintains a culture built around the program's core values

Value

WHAT is the *All of Us* Research Program?	It will be **one of the largest, most diverse health databases** in history
WHY should I care?	It combines **biological factors and social determinants** on a large, inclusive scale
WHO benefits?	It equips researchers to make discoveries that **enable more precise approaches to care**, informing providers' recommendations & individuals' choices
HOW do I access it**?**	It is now **available to researchers with institutional signoff** across a wide range of settings and career stages with just a computer and internet connection

Current protocol

Enroll, Consent and Authorize EHR

- Recruiting 18+ years old initially; plan to include children in future
- Online, interactive consent
- Includes authorization to share EHR data

Answer Surveys

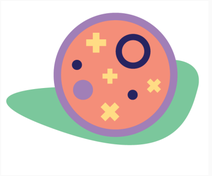
- The Basics
- Overall Health
- Lifestyle
- Health Care Access & Utilization
- Personal and Family Medical History
- Social Determinants of Health

Additional surveys will be released on an ongoing basis

Provide Physical Measurements

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

Based on diverse sampling and capacity



Provide Biosamples

- Blood (or saliva)
- Urine specimen
- Biosamples will be stored at the program's biobank

Based on diverse sampling and capacity

Share data from Wearables/Digital Apps

- Share data from wearable fitness devices, starting with Fitbit

Coming soon:

- Integrated apps to track mood & cardio-respiratory fitness

Consent and EHR Authorization

- Consent and EHR Authorization
- Participant Surveys
- Physical Measurements
- Biosamples
- Mobile/Wearable Tech

- Participants must be 18 years or older
- Online video consent
- Consent process includes authorization to share EHR data with researchers

Initial Data Types	Expanded Data Types (May Include)
• Demographics • Visits • Diagnoses • Procedures • Medications • Laboratory visits • Vital signs	• Clinical notes • Radiology, cardiology, and other reports • Mental health reports • Substance abuse, alcohol use, and tobacco use • More laboratory results, potentially including genomics

Participant surveys

Consent and EHR Authorization

Participant Surveys

Physical Measurements

Biosamples

Mobile/Wearable Tech

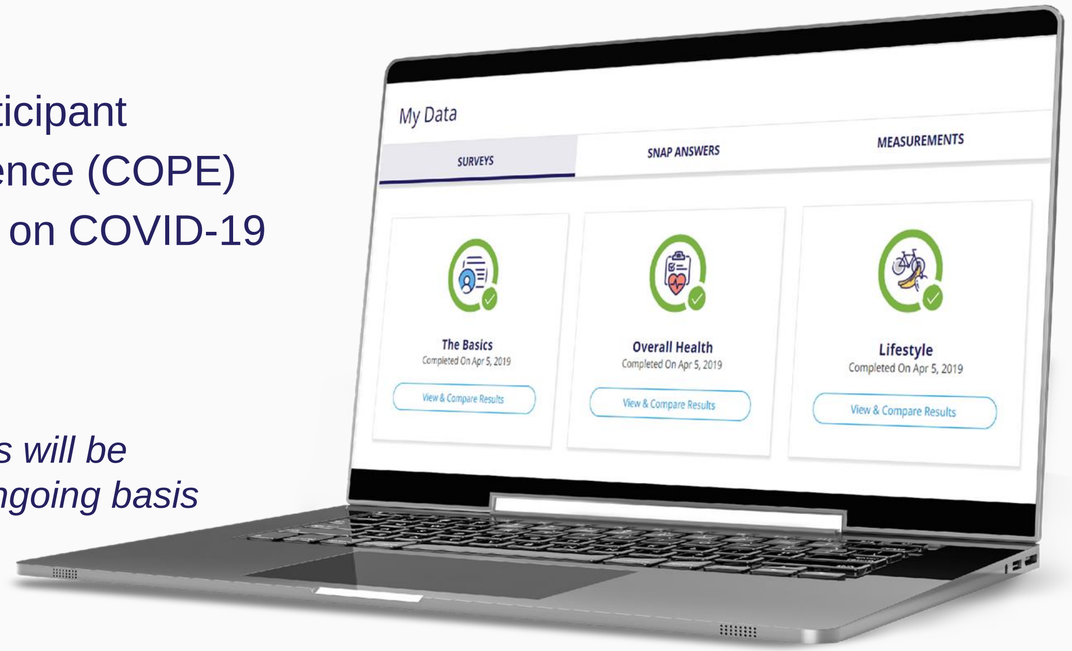
Open Surveys:

- The Basics
- Overall Health
- Lifestyle
- Health Care Access & Utilization
- Personal & Family Health History
- Social Determinants of Health

Closed Surveys:

- COVID-19 Participant Survey Experience (COPE)
- Minute Survey on COVID-19 Vaccines

Additional surveys will be released on an ongoing basis

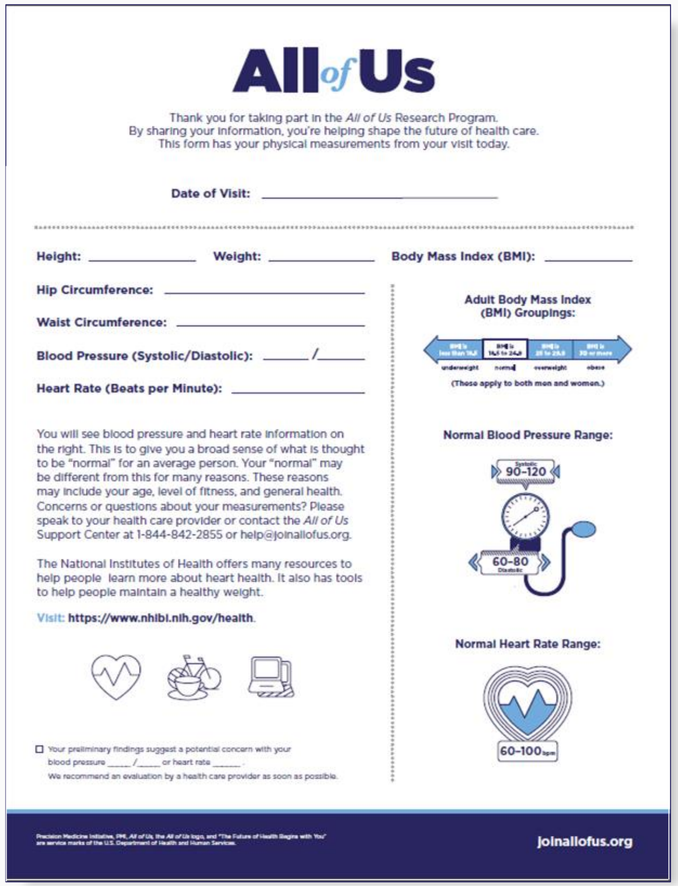


Physical measurements

- Consent and EHR Authorization
- Participant Surveys
- **Physical Measurements**
- Biosamples
- Mobile/Wearable Tech

Current Measurements:

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



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:

underweight | normal | overweight | obese

(These apply to both men and women.)

Normal Blood Pressure Range:

Systolic 90–120
Diastolic 60–80

Normal Heart Rate Range:

60–100 bpm

Precision Medicine Initiative, PMI, All of Us, the All of Us logo, and "The Future of Health Begins with You" are service marks of the U.S. Department of Health and Human Services.

joinallofus.org

Biosamples

- Consent and EHR Authorization
- Participant Surveys
- Physical Measurements
- **Biosamples**
- Mobile/Wearable Tech

Examples:

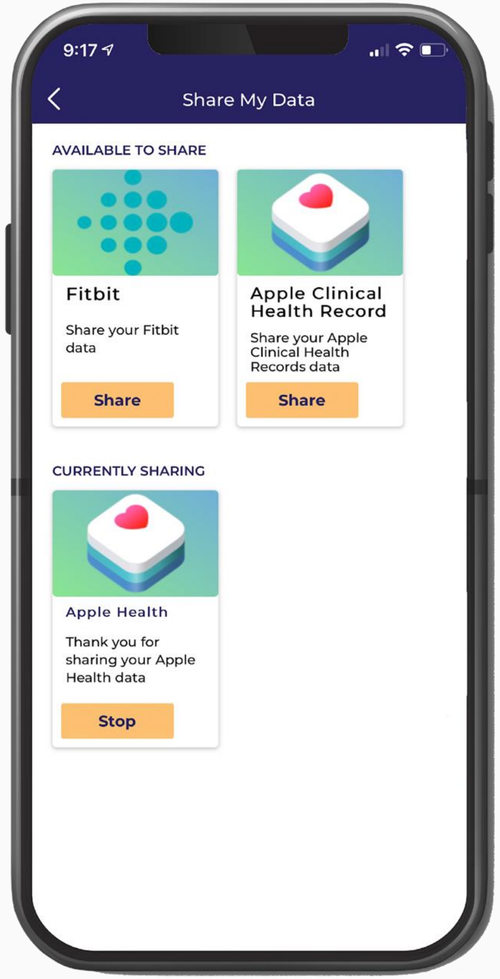
- Blood
- Saliva
- Urine

Mobile/Wearable technology

- Consent and EHR Authorization
- Participant Surveys
- Physical Measurements
- Biosamples
- **Mobile/Wearable Tech**

Current Data:

- Data from wearable fitness devices, including Fitbit and Apple HealthKit
- Additional integrations in the future
- Data includes:
  - Heart Rate by Zone Summary
  - Heart Rate (Minute-Level)
  - Activity (Daily Summary)
  - Activity Intraday Steps (Minute-Level)



All of Us Data

Different data types available

Surveys

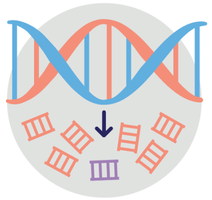
from 372,000+ participants
Including lifestyle, access to care, medical history, and **data from nearly 100,000 participants on their experiences during the COVID-19 Pandemic**

Physical Measurements

from 306,000+ participants
Blood pressure, heart rate, BMI and more

Electronic Health Records

from 258,000+ participants
Standardized to OMOP common data model



Genomics

98,600+ whole genome sequences
Only available via the Controlled Tier

Genomics

165,000+ genotyping arrays
Only available via the Controlled Tier

Wearables

(from 12,800+ participants)
Physical activity and heart rate from participants who have connected their Fitbit devices

Participant biosamples are forthcoming

Including the first batch of genomic data

Genomics

Only available via the Controlled Tier

98,600+
Whole Genome Sequences

165,000+
Genotyping Arrays

593,500,000+
Unique Variants
Represented and searchable through public genomic data browser

Genomics Analysis Tools
Hail and PLINK
in addition to R, Python, and Jupyter Notebooks

Genomic Data is Paired with Rich Phenotypic Data

77,000+
Have Whole Genome Sequences + Electronic Health Records + Physical Measurements + Survey Responses

95,000+
Have Whole Genome Sequences + Physical Measurements + Survey Responses

78,200+
Have Whole Genome Sequences + Electronic Health Records

3,500+
Have Whole Genome Sequences + Fitbit Records
Representing >30% of all participant Fitbit records

Selected Data Snapshots: Enrollment Progress

(Updated 12/4/22)

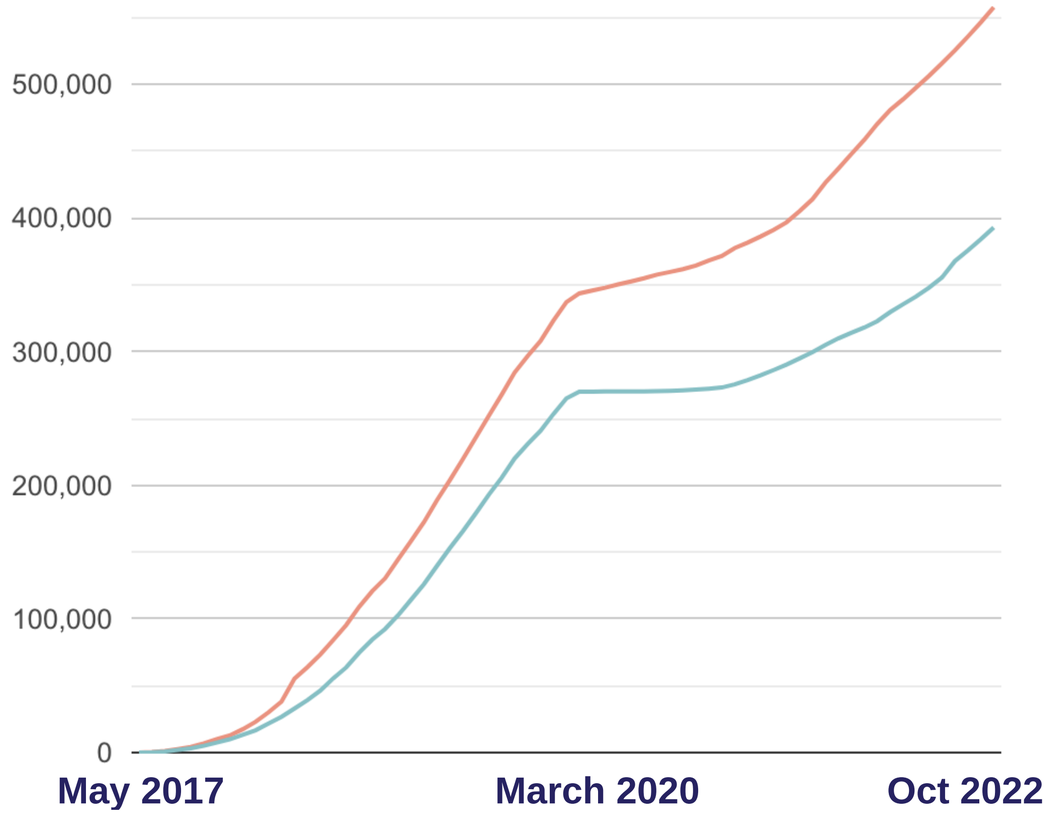
558,000+
Participants

339,000+
Electronic Health Records

393,000+
Participants who have completed initial steps of the program

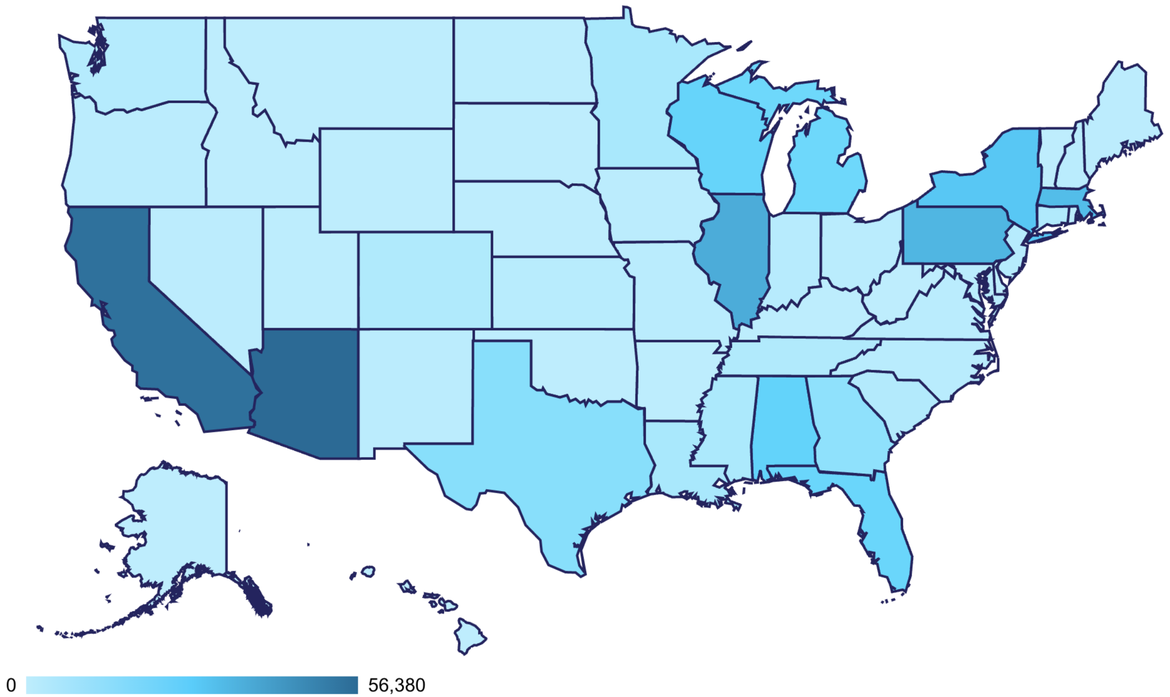
411,000+
Biosamples

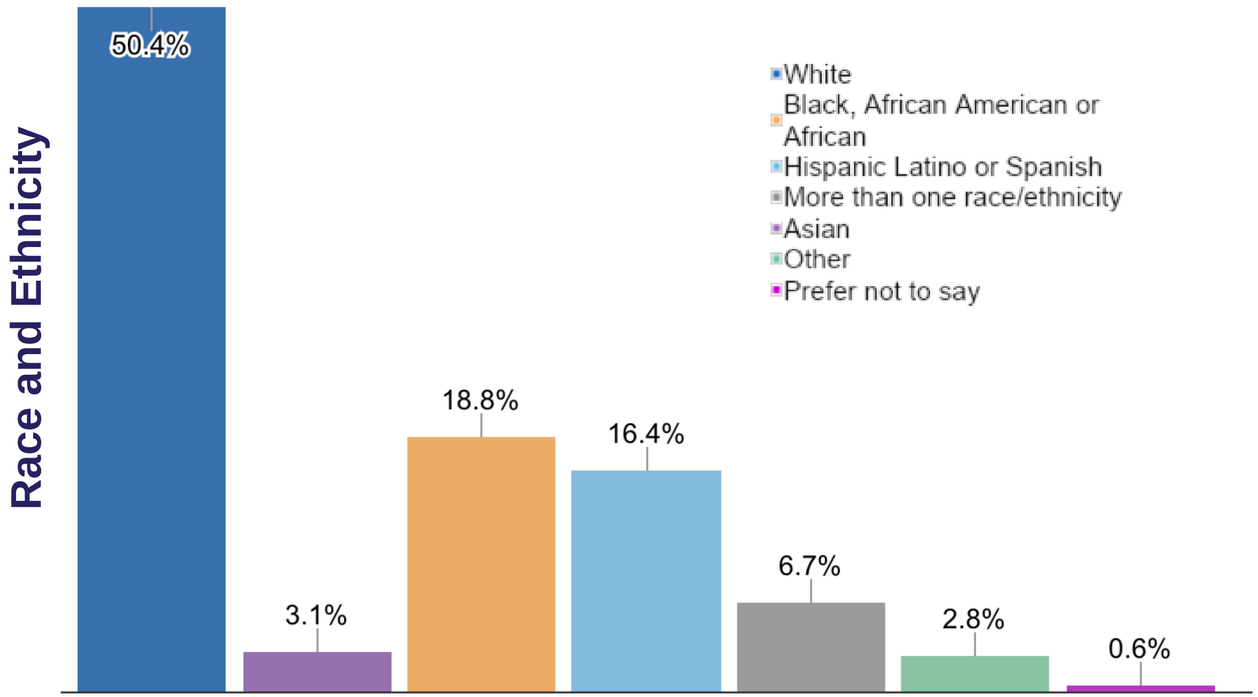
Enrollment Numbers



Commitment *to All of Us* Participant Diversity

Geographic Diversity





Data as of 12/4/22

Diversity

Includes racial and ethnic minorities as well as sexual and gender minorities, people with low income or limited education, and other groups.

50%+ Racial and Ethnic Minorities

80%+ Underrepresented in Biomedical Research

Accessing Data and Research Tools

DATA: Access

Three Tiers

Public Tier

Contains only anonymized, aggregate data which is available to anyone through the Data Browser and Data Snapshots. Public Tier tools include — Data Browser, Research Projects Directory, Publications, Data Snapshots, and the Survey Explorer

Registered Tier

Contains curated, anonymized, individual-level data which is available to registered researchers on the Researcher Workbench. Registered Tier tools include — Cohort Builder, Dataset Builder, Workspaces, Notebooks, and the Support Hub

Controlled Tier

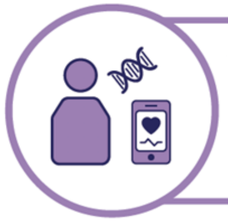
Available to approved researchers who have taken additional steps and training to access these data. Controlled Tier data includes: Genomic data, additional clinical fields in electronic health records, and additional demographic data from surveys that are suppressed or generalized in the Registered Tier

**Counts between all three data tiers and Data Snapshots may vary because of the lag time associated with the data curation process.*

DATA: Journey

How the Research Hub Works

All of Us RESEARCH PROGRAM



1. **Participants share their data** with the *All of Us* Research Program through multiple sources. These data are sent to a secure cloud environment, managed by the Data and Research Center.

2. **Participant data** is received and funneled through a curation pipeline within a secure repository that connects to the Research Hub tools.

3. **Anyone can visit the Research Hub** to learn more about the types of data *All of Us* makes available for research. The Survey Explorer and Data Browser offer more information about the unique data elements and let visitors browse aggregated participant data.

4. **Researchers register** and apply for access to the Researcher Workbench to analyze data.

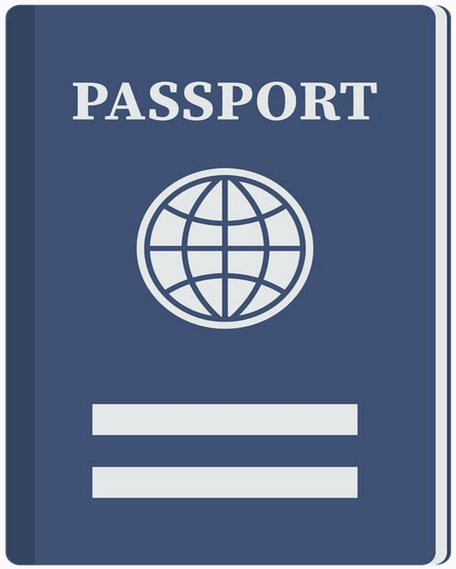
5. **Registered researchers** in the Researcher Workbench can create research projects using collaborative workspaces, cohort-building tools, interactive notebooks, and more.

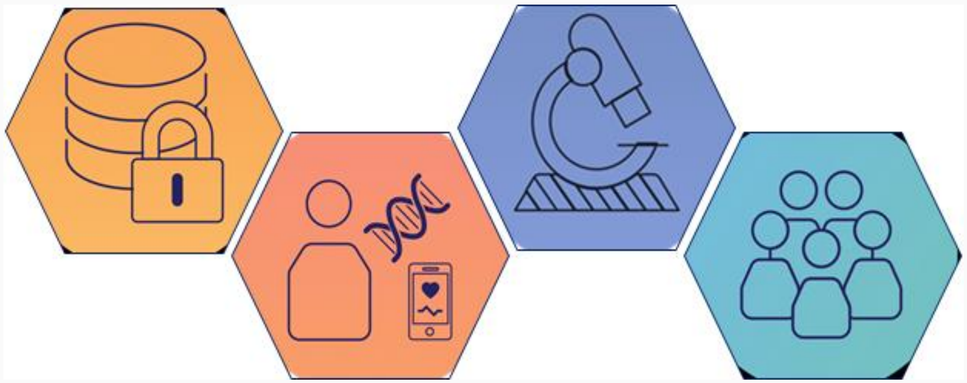
6. **Research underway** can be viewed on the Projects Directory. Publications related to *All of Us* Research Program data are posted on the Publications page.

Data passport model

- *All of Us* employs a “data passport” model for the Registered and Controlled tiers, to grant researchers **broad permission to explore the data for a wide range of studies**

- To facilitate collaboration and keep participant data secure, the Researcher Workbench is **hosted in a Cloud-based system**





Data release timeline

2019

Data Browser

(launched May 2019)

Interactive tool available to the public that provides summary statistics from the program's growing database, allowing researchers to understand the characteristics of our participant population, **explore the data types available and plan research questions**

2020

Researcher Workbench

(launched May 2020)

Released of the first version of the *All of Us* dataset and the first set of researcher tools for beta testing

- **Survey** data
- **Physical measurement** data
- **EHR** data

2021

Expansion of data and tools

(throughout 2021)

Expanded of the *All of Us* dataset with additional survey, measurement, and EHR data.

- **Start of genotyping** and whole genome sequencing
- **Initial return of results** to participants

>1,000K researchers registered

2022

Controlled Tier & Genomics

(ongoing)

Launch our Controlled tier, available to approved researchers who have taken additional steps and training to access:

- **Genomics** data
- **Individual and granular-level** information

Available in Registered and Controlled tiers, with basic information available via our public data browser:

- **ACS data linkage**
- **New surveys**

2023 and *beyond*

Additional data types and tools

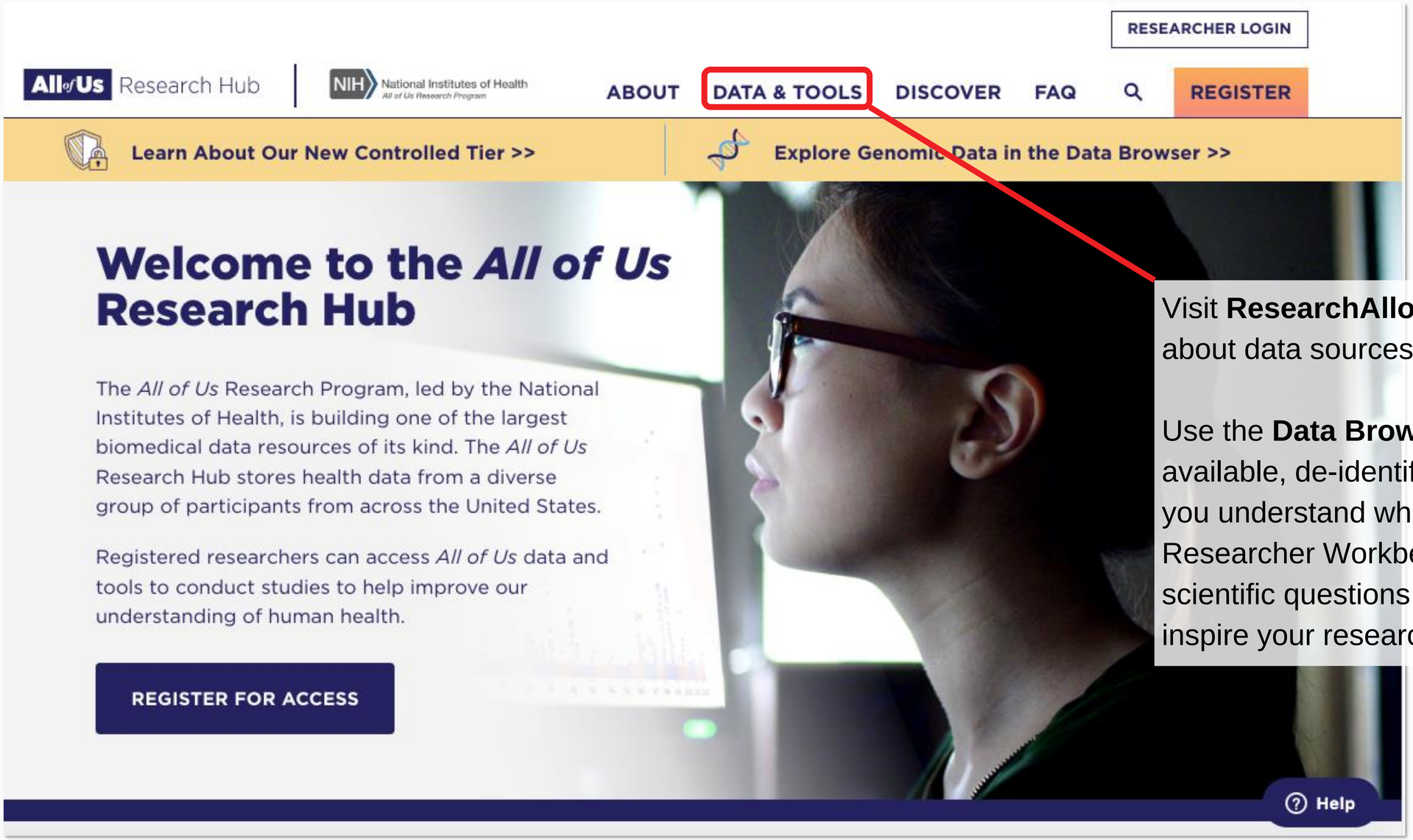
(ongoing)

Expect to enroll 1M participants within 5-7 years of launch and continue to expand the *All of Us* dataset with new data releases, including over time:

- **Assay data**
- **Additional wearables data**
  - Fitbit sleep data
  - Apple Healthkit
- **New surveys** (e.g., Mental Health & Wellbeing)
- **New Linkages to external data sources**
- **Expanded researcher access**
- **Ancillary study data**

Release of new versions of the Researcher Workbench

DATA: Learn more



Visit **ResearchAllofUs.org** to learn more about data sources and methodologies.

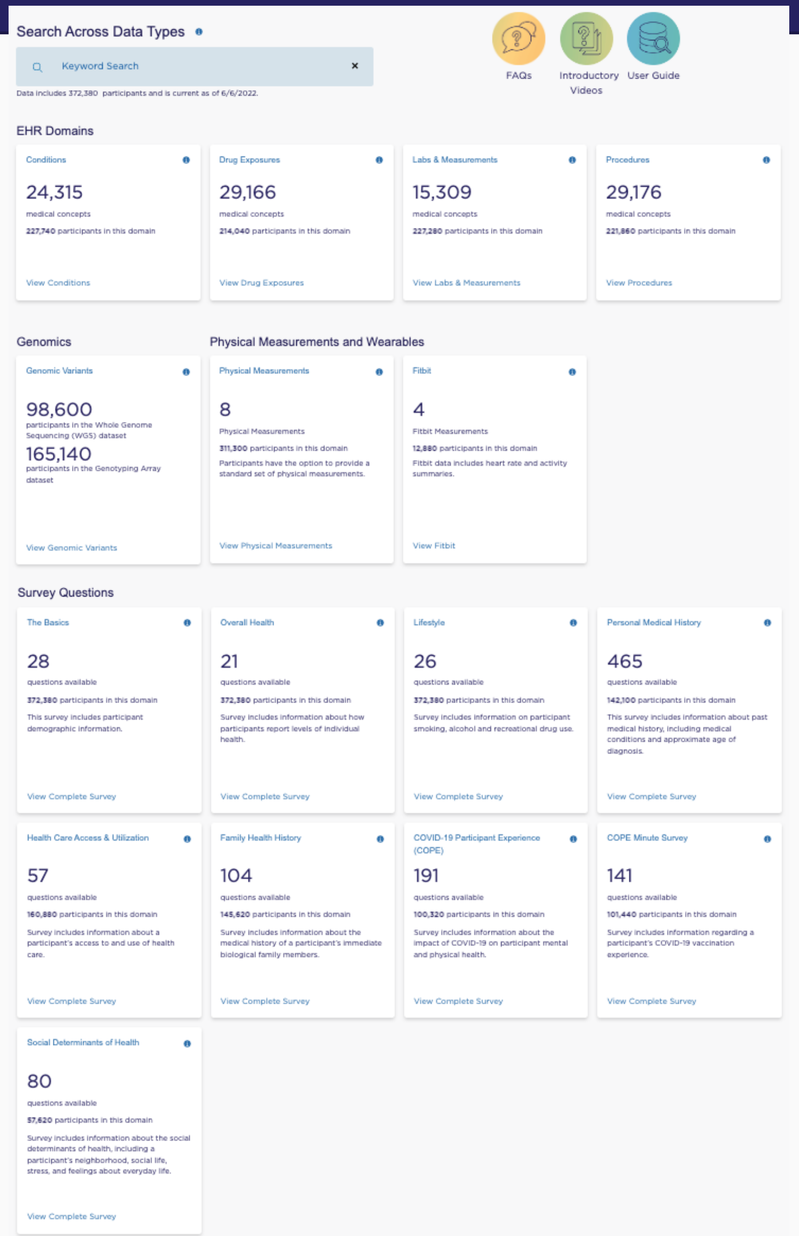
Use the **Data Browser** to view publicly available, de-identified data that can help you understand what's available within the Researcher Workbench and what kind of scientific questions you can develop to inspire your research.

All of Us Data Browser: Introduction

DataBrowser.ResearchAllofUs.org

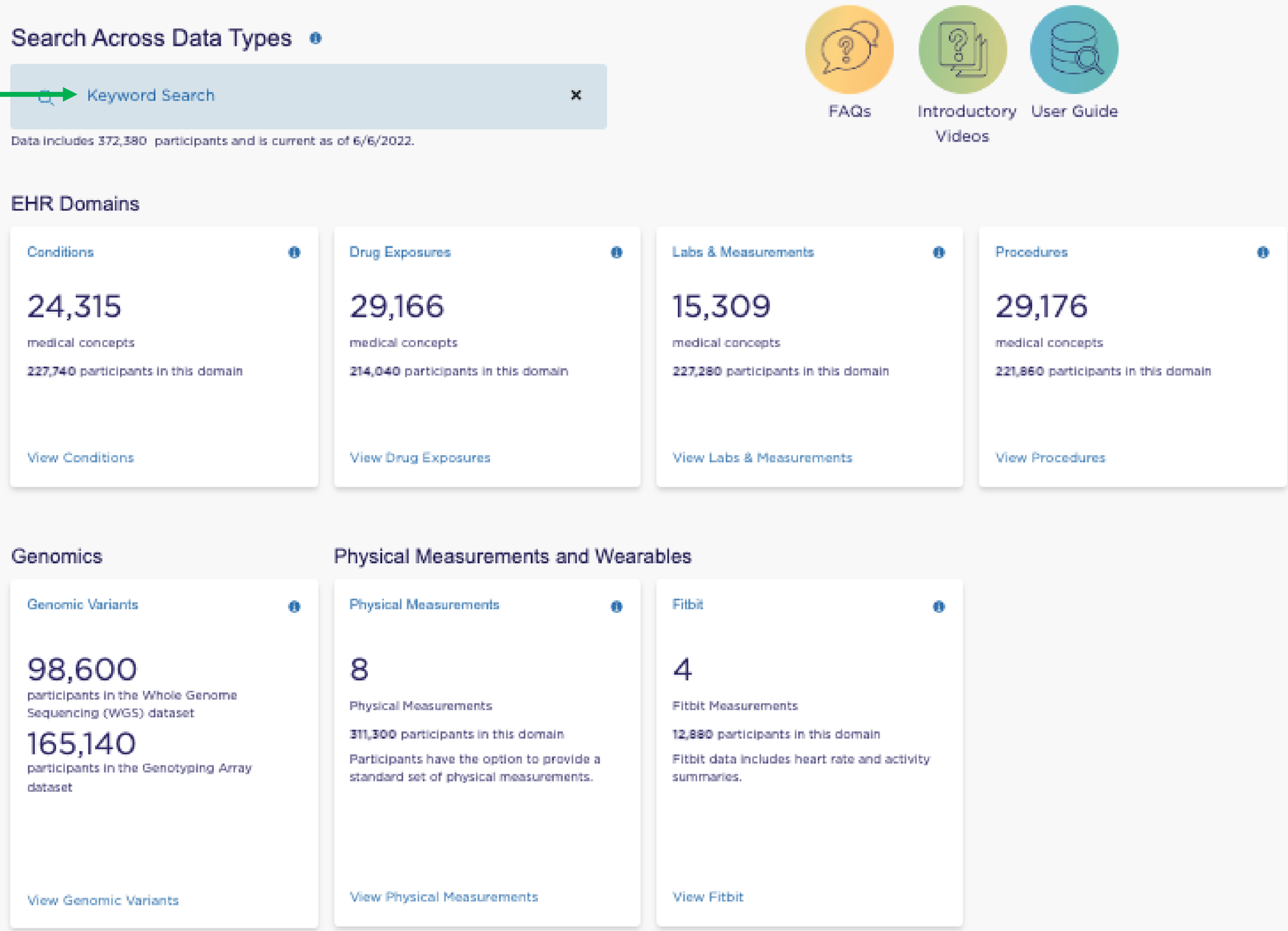
- Provides **summary statistics** from the program's growing database
- **Open to everyone** – no login!
- Allows participants to **understand the makeup of the cohort**
- Allows researchers to understand the characteristics of our participant population, **explore the data types available and plan research questions**

Interactive tool launched May, 2019



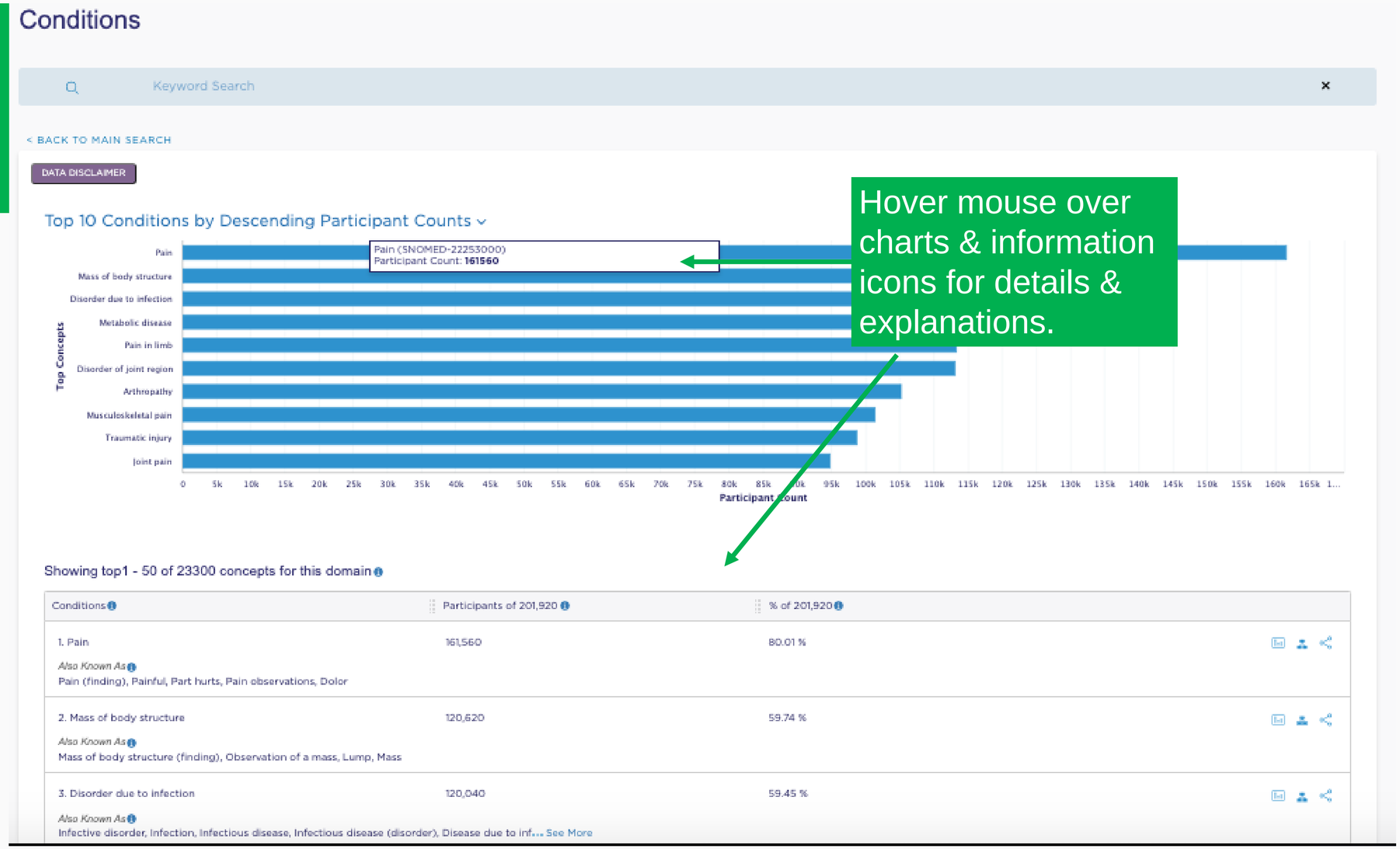
How to use the *All of Us* Data Browser

Search for specific keywords or browse using the different options underneath.

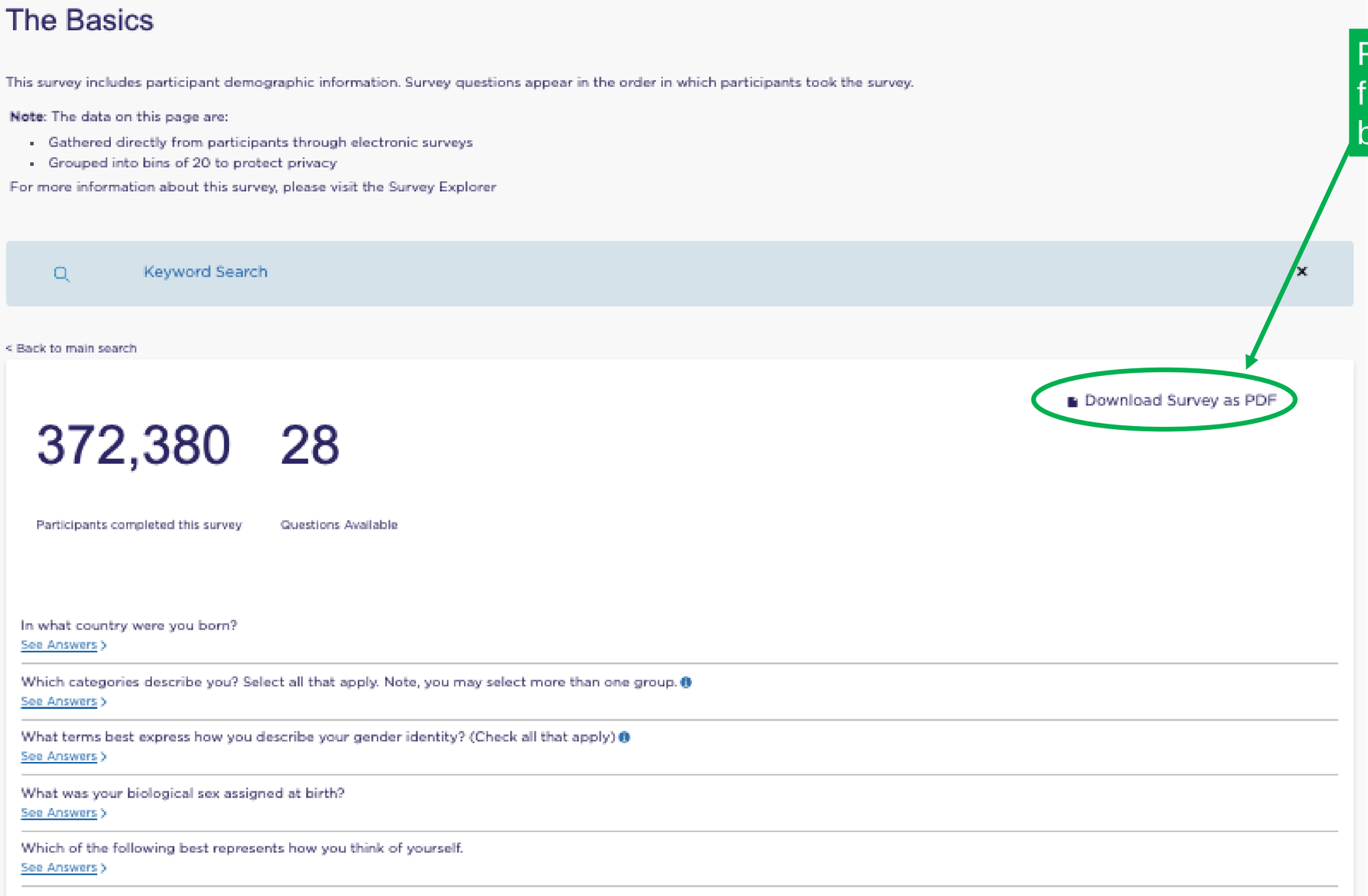


All of Us Data Browser: Dive in and spark your imagination

Available data gives insight into the participant cohort & research opportunities.

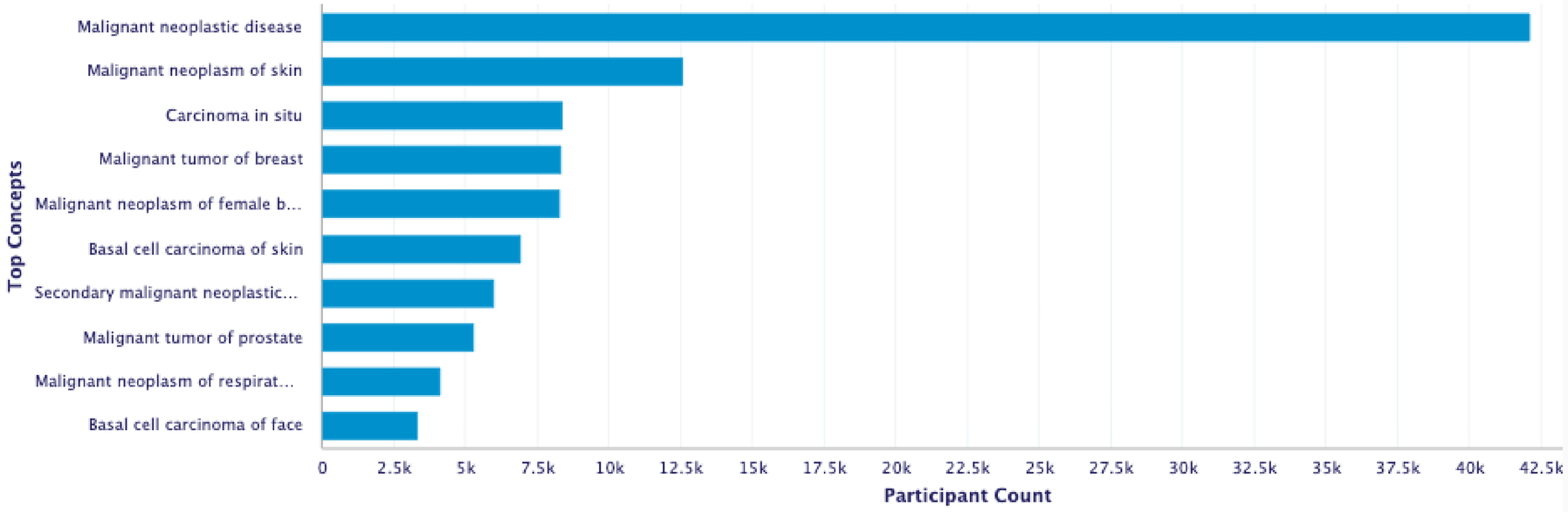


All of Us Data Browser: A quick look



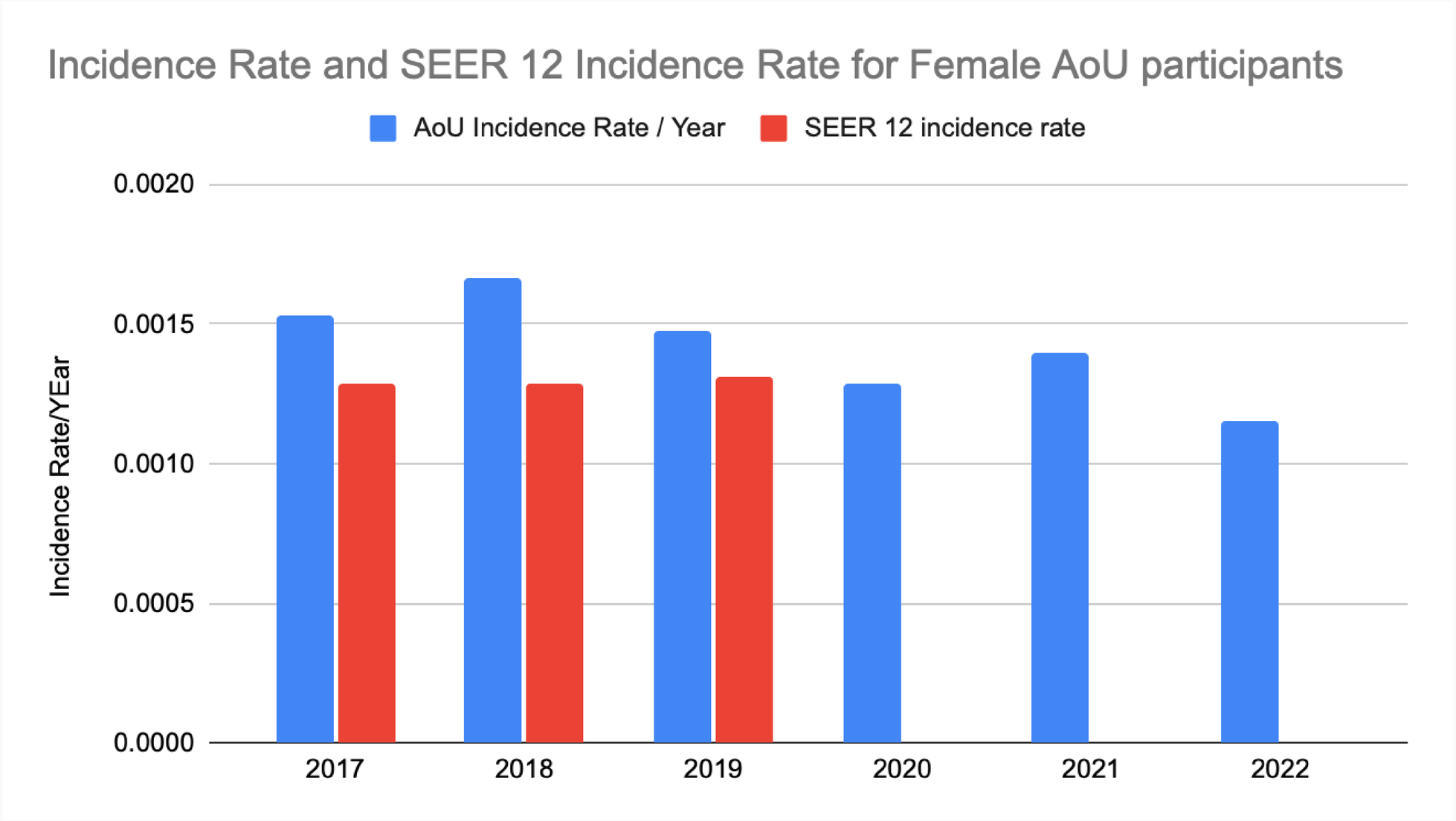
Researchers can view the full surveys, including branching logic.

Cancer Counts (EHR N= >42,000)

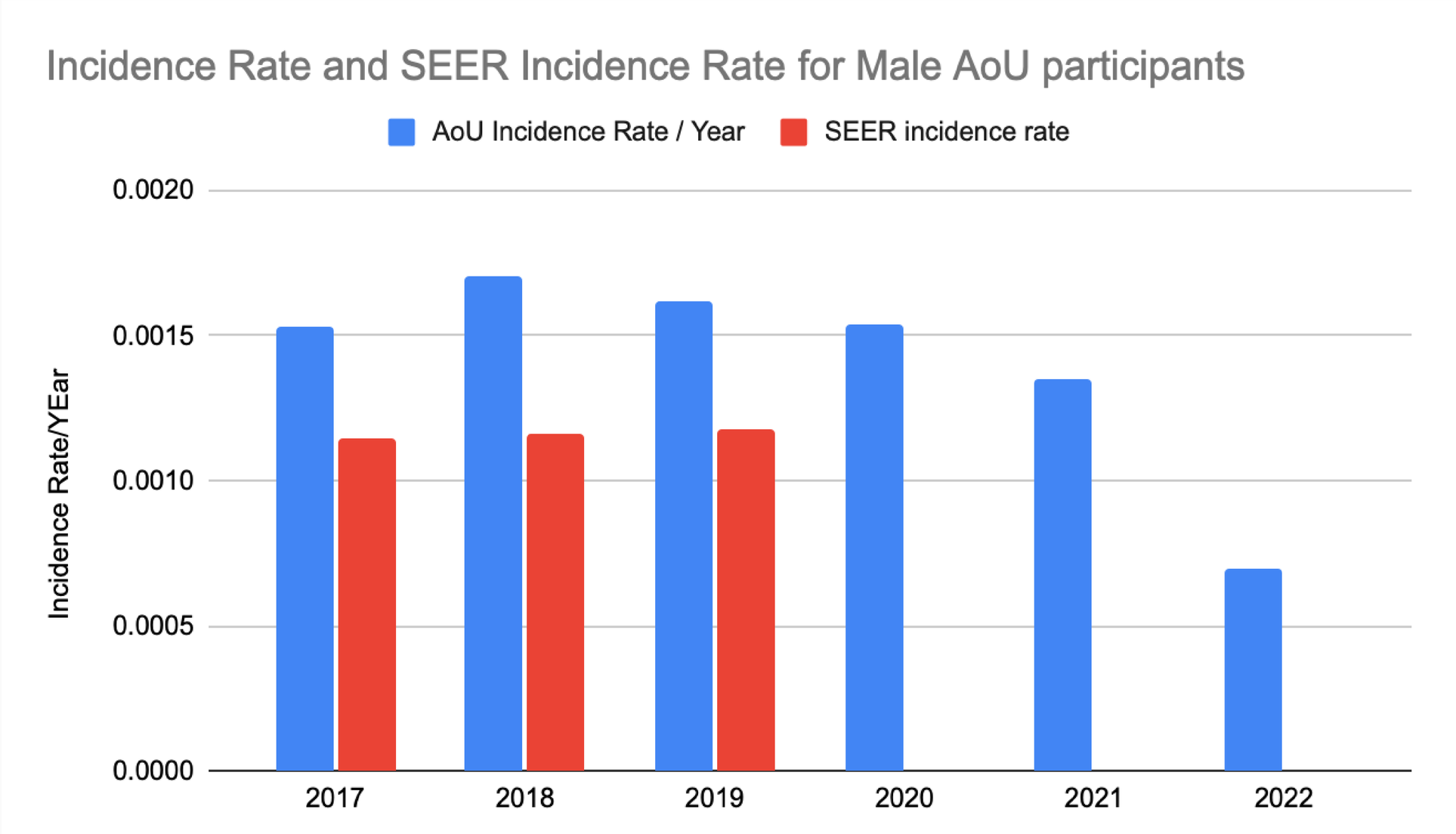


Conditions	Participants of 227,740	% of 227,740
1. Malignant neoplastic disease	42,080	18.48 %
Also Known As Malignant tumour, CA - **Cancer**, Malignant neoplasm, Malignant tumor, Malignant neoplastic disease (di... See More		
2. Malignant neoplasm of skin	12,560	5.52 %
Also Known As Malignant neoplasm of skin (disorder), Skin **cancer**, **Cancer** of skin		
3. Carcinoma in situ	8,360	3.67 %
Also Known As Carcinoma in situ (disorder), **Cancer** in situ		

Comparing Female Breast Cancer Incidence Rate in AoU to US Incidence Rate



Comparing Male Prostate Cancer Incidence Rate in AoU to US Incidence Rate



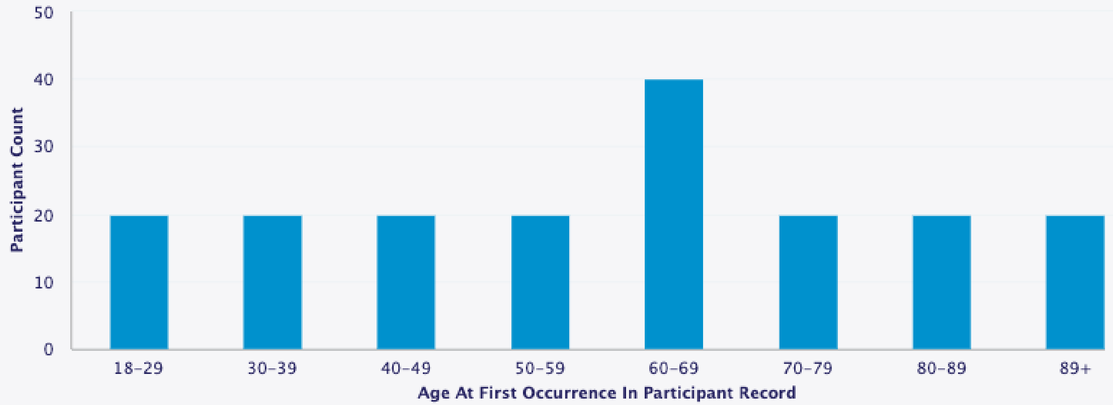
Immuno-Oncology use in AoU

1. **nivolumab** 60 0.03 %

Found in these commercially branded products
Nivolumab *BMS, Opdivo*

Nivolumab (N=60)

Sex Assigned at Birth | Age | Sources



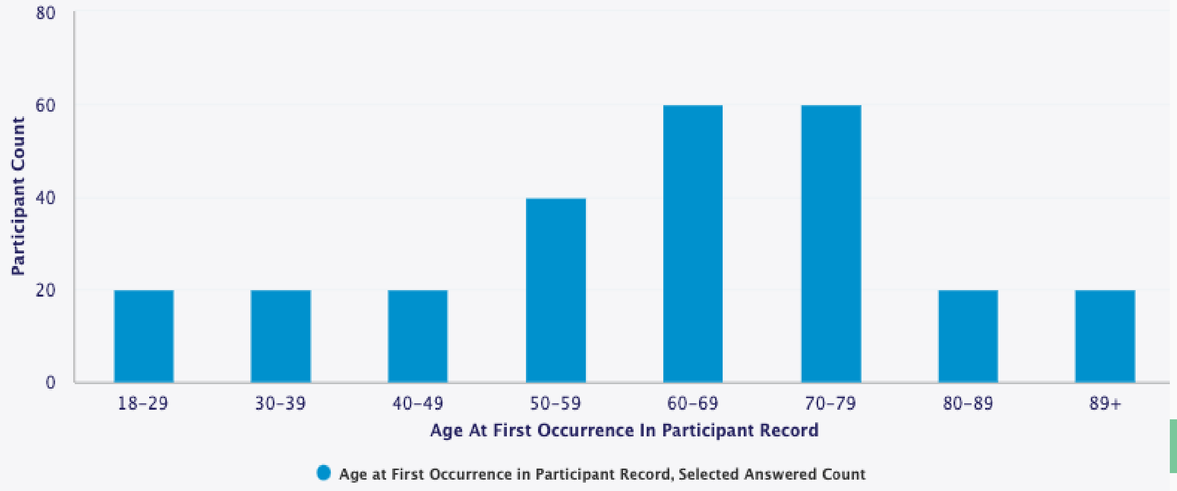
Atezolizumab, avelumab, durvalumab, and cemiplimab all in database but in <20 participants

1. 4 ML **pembrolizumab** 25 MG/ML Injection 160 0.07 %

Also Known As
pembrolizumab 100 MG per 4 ML Injection

Pembrolizumab (N=160)

Sex Assigned at Birth | Age | Sources



All of Us Data Browser: Genomics

Use the Variant Search to explore allele frequencies for a gene or genomic region.

BRCA2

Examples:
Gene: BRCA2, **Variant:** 13-32355250-T-C,
Genomic Region: chr13:32355000-32375000

13,056 variants found

Variant ID	Gene	Consequence	Protein Change	Clinical Significance	Allele Count	Allele Number	Allele F

Variant ID: 13-32320347-G-A

Consequence: intron_variant

Protein Change:

DNA Change: ENST00000380152.7:c.316+1022G>A

Transcript: ENST00000380152.7

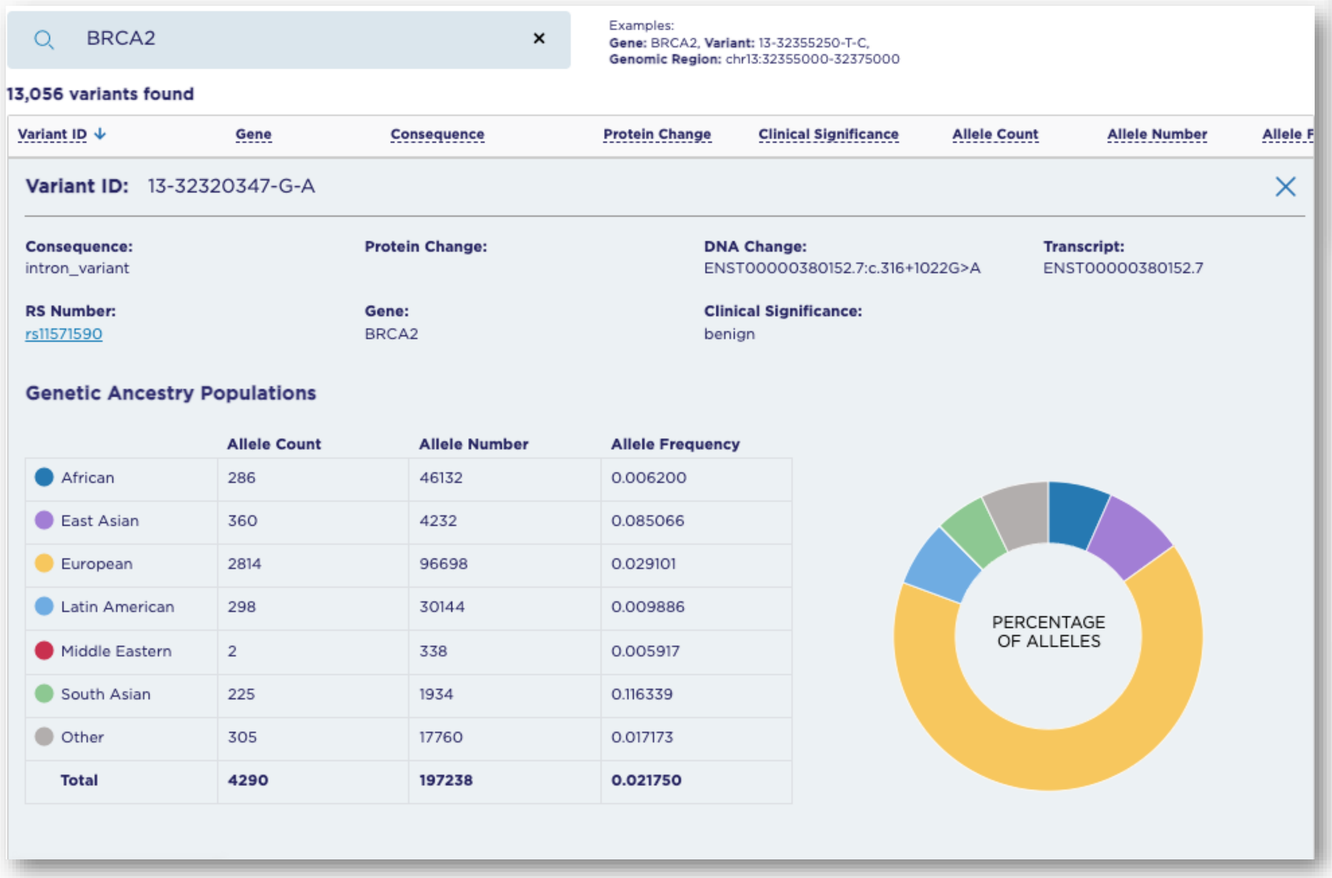
RS Number: rs11571590

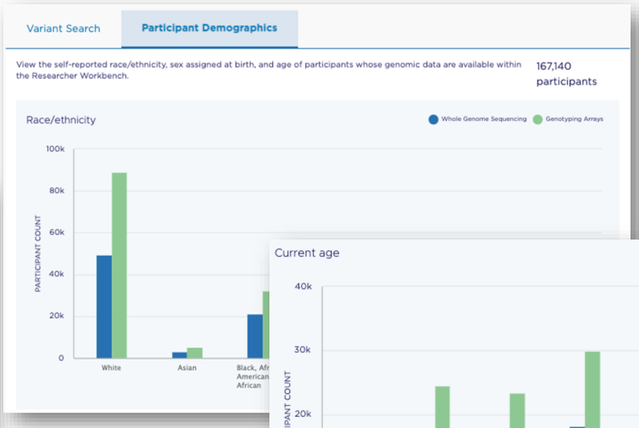
Gene: BRCA2

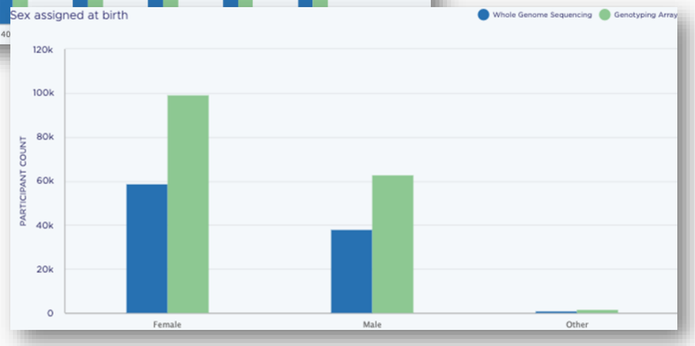
Clinical Significance: benign

Genetic Ancestry Populations

	Allele Count	Allele Number	Allele Frequency
African	286	46132	0.006200
East Asian	360	4232	0.085066
European	2814	96698	0.029101
Latin American	298	30144	0.009886
Middle Eastern	2	338	0.005917
South Asian	225	1934	0.116339
Other	305	17760	0.017173
Total	**4290**	**197238**	**0.021750**







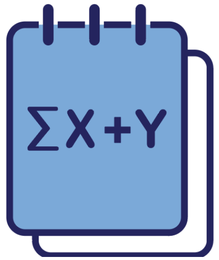
Explore data breakdowns by sex assigned at birth & age.

Drill down into specific variants to view select annotations and genetic ancestry associations.

Powerful analytical tools & user support

WORKSPACES

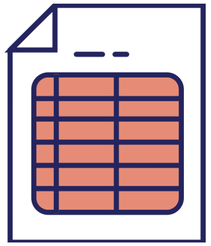
USES:
- Organizing research projects
- Collaboration
- Transparency



NOTEBOOKS

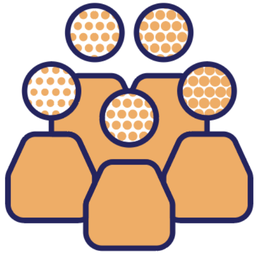
USES:
- Analyses
- Queries

Uses programming languages R and Python



DATASET BUILDER

USES:
- Pre-populated analyses
- Dataset previews



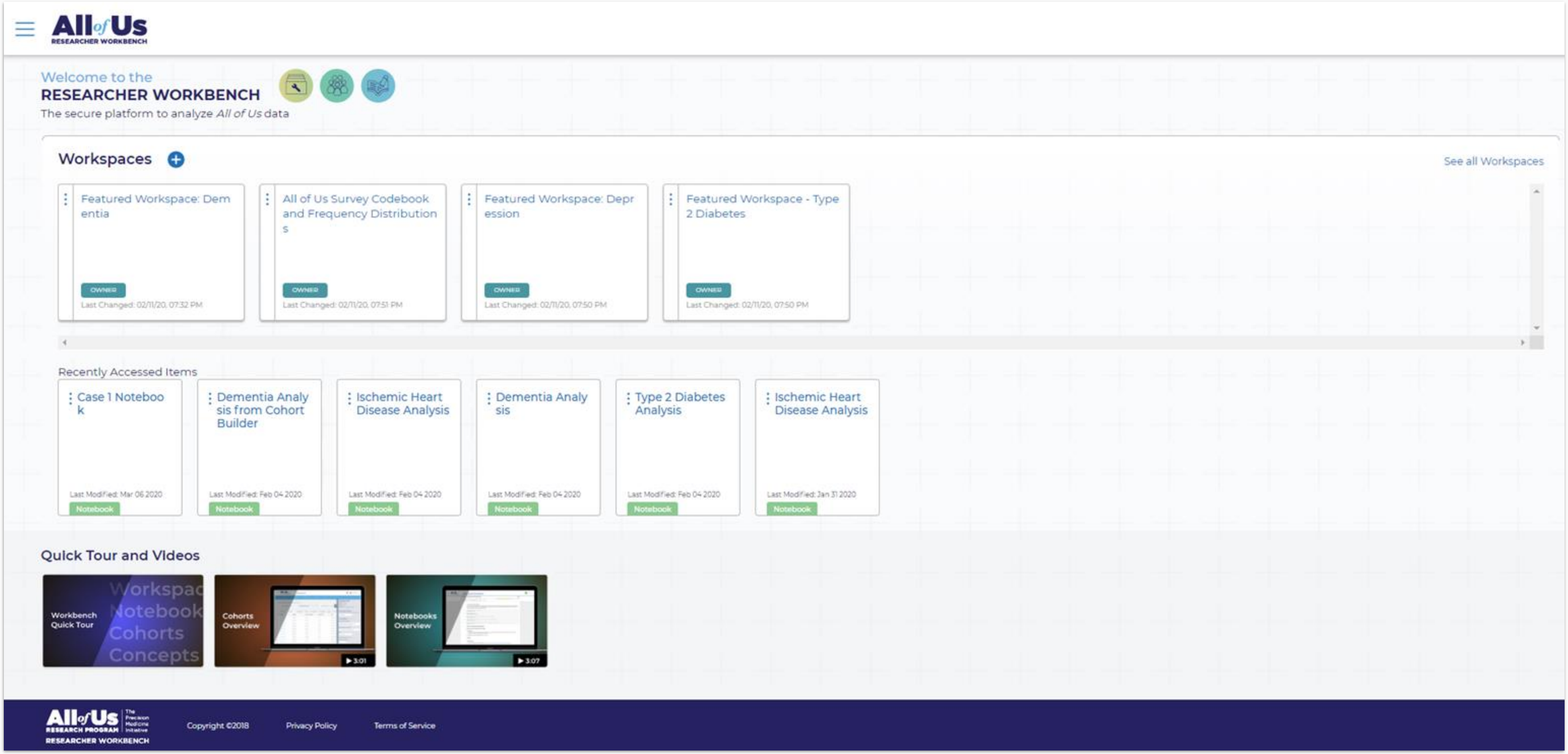
COHORT BUILDER

USES:
- Cohort creation

USER SUPPORT HUB

USES:
- Workbench training and resources
- Articles on Data and Tools
- Live office hours

Workspaces

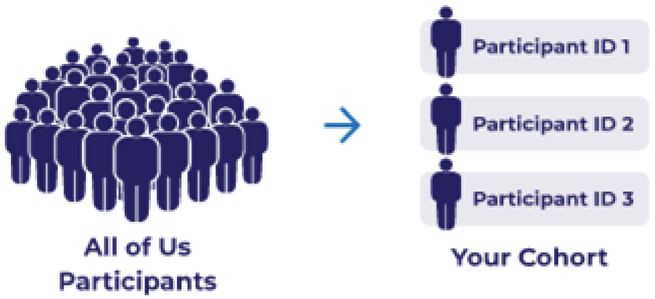

All of Us
RESEARCHER WORKBENCH
Welcome to the
RESEARCHER WORKBENCH
The secure platform to analyze All of Us data
Workspaces
See all Workspaces
Featured Workspace: Dementia
OWNER
Last Changed: 02/11/20, 07:32 PM
All of Us Survey Codebook and Frequency Distributions
OWNER
Last Changed: 02/11/20, 07:51 PM
Featured Workspace: Depression
OWNER
Last Changed: 02/11/20, 07:50 PM
Featured Workspace - Type 2 Diabetes
OWNER
Last Changed: 02/11/20, 07:50 PM
Recently Accessed Items
Case 1 Notebook
Last Modified: Mar 06 2020
Notebook
Dementia Analysis from Cohort Builder
Last Modified: Feb 04 2020
Notebook
Ischemic Heart Disease Analysis
Last Modified: Feb 04 2020
Notebook
Dementia Analysis
Last Modified: Feb 04 2020
Notebook
Type 2 Diabetes Analysis
Last Modified: Feb 04 2020
Notebook
Ischemic Heart Disease Analysis
Last Modified: Jan 31 2020
Notebook
Quick Tour and Videos
Workbench Quick Tour
Workspaces
Notebooks
Cohorts
Concepts
Cohorts Overview
3:01
Notebooks Overview
3:07
All of Us RESEARCH PROGRAM
The Precision Medicine Initiative
RESEARCHER WORKBENCH
Copyright ©2018
Privacy Policy
Terms of Service

Cohorts and Datasets

DATA | ANALYSIS | ABOUT

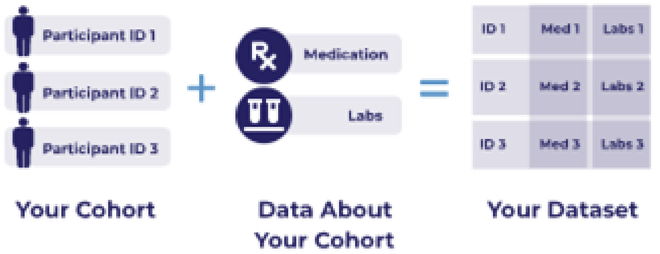
Cohorts

A cohort is a group of participants based on specific criteria.



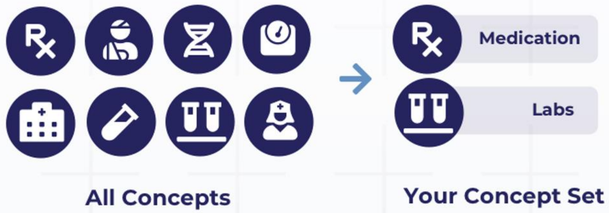
Datasets

A dataset is a table containing data about a Cohort that can be exported for analysis.



Concept Sets

Concept sets are a set of medical concepts (e.g., blood pressure) that are of interest for a research study



All of Us Data Dictionary

The Data Dictionary provides the following:

- A description for each data field;
- Information on whether the data in each field come from participant health records or from information the participants provide themselves, like survey data;
- Versioning data so you can see what has been changed, added, or removed since the previous curated dataset.

Create an *All of Us* account

Register to be an *All of Us* Researcher

STEP 1

LEARN MORE ABOUT THE DATA AVAILABLE

STEP 2

CHECK FOR YOUR INSTITUTION'S AGREEMENT

STEP 3

REGISTER AS A RESEARCHER

STEP 4

VERIFY YOUR IDENTITY

STEP 5

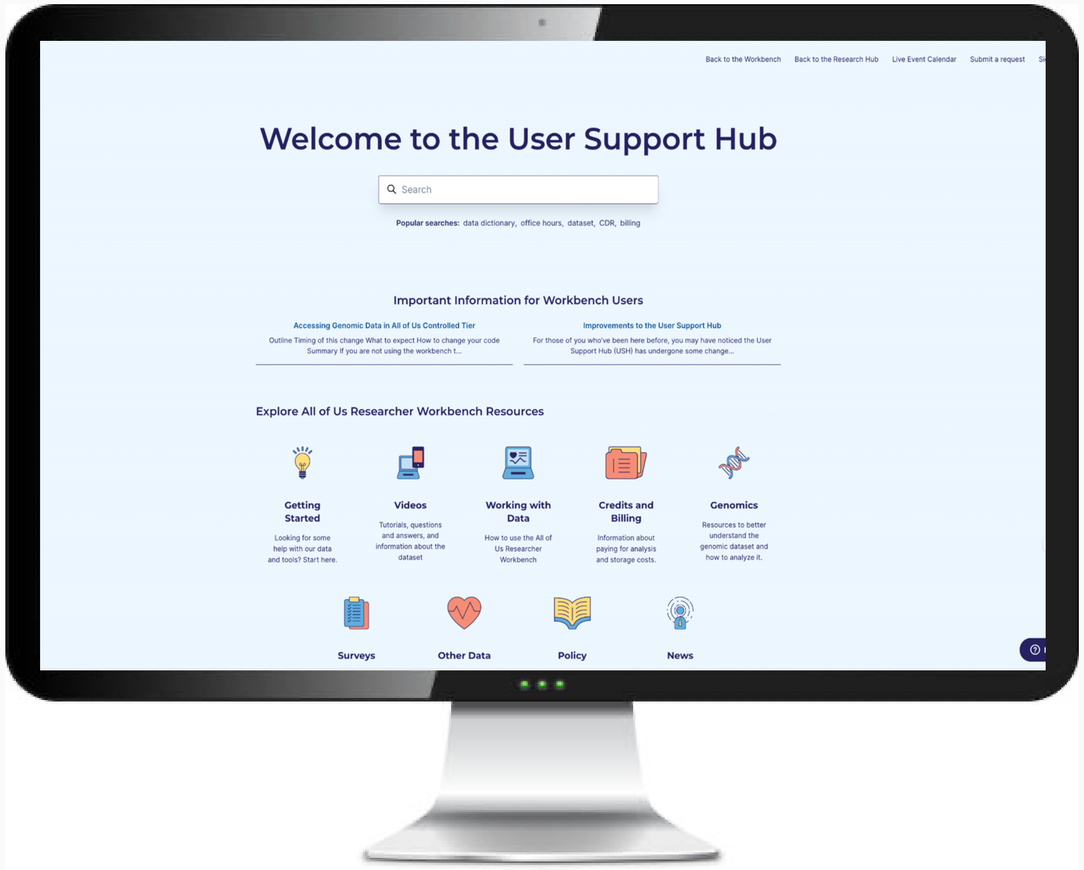
COMPLETE *ALL OF US* RESPONSIBLE CONDUCT OF RESEARCH TRAINING

STEP 6

SIGN DATA USER CODE OF CONDUCT

USER SUPPORT HUB: Resources to Support Your Research

No Login Required! Visit the Support Hub today to find answers to common questions about becoming a registered researcher, using the *All of Us* dataset, and more.



Find instructional materials about the *All of Us* Researcher Workbench

- Includes video tutorials, educational resources, release notes, and more

Search >250 useful articles

- Find information about data dictionaries, the Controlled Tier directory, how WGS and Array data are organized, and more

Connect with experts during weekly LIVE office hours

- Explore our calendar of dedicated office hours where researchers can talk about data types and tools with experts

researchallofus.org/support

35

Stay in touch to learn more

Sign up for our bimonthly newsletter

Subscribe to Research Roundup

Stay up-to-date on the latest news and insights from the *All of Us* Research Hub through our bimonthly email newsletter.

allof-us.org/RRSignup

Advantages and limitations

Advantages

- **DIVERSITY** OF PARTICIPANTS AND DATA
- **GRANULARITY** OF PARTICIPANT-PROVIDED DATA
- **SPEED** OF DATA AVAILABILITY
- **ACCESSIBILITY** TO BROAD GROUP OF RESEARCHERS
- POWERFUL **ANALYTIC CAPABILITIES**
- **TEAM SCIENCE** APPROACH
- AVAILABILITY OF **SAMPLE COHORTS**

Limitations

- **NOT NATIONALLY REPRESENTATIVE**
  (But reflective of the broad diversity of the U.S.)
- DATA IS **NOT YET COMPREHENSIVE**
  (It is continuing to expand)
- **NOT ALL PARTICIPANTS HAVE ALL DATA TYPES** AVAILABLE
- LIMITED SET OF **ENVIRONMENTAL/SOCIAL DETERMINANT DATA**
  (But more is coming through new surveys and linkages)
- SOME KNOWLEDGE OF **CODING REQUIRED**

All of Us Community and Provider Partner Network *(as of August 2022)*

Note: These are not approved lockups and should not be repurposed on assets.

All of Us Consortium Members *(beyond community partners, as of August 2022)*

The Participant Center

Scripps Research Translational Institute · BlueCross BlueShield · SENSIS · SageBionetworks · owaves · Walgreens · MAYO CLINIC · Dedalus · Medscape · Bloodworks Northwest · CVS Health · Blood Assurance · Gulf Coast Regional Blood Center Commit for Life. · careevolution · WebMD · active minds · patientslikeme · South Texas Blood & Tissue · Cascade Regional Blood Services · ACHA American College Health Association · SYNC FOR SCIENCE · Quest Diagnostics · San Diego Blood Bank · National Blood Collaborative · fitbit · Mary's Center

Communications & Engagement

WONDROS · PYXIS PARTNERS · RTI INTERNATIONAL · THE UNIVERSITY OF UTAH

HPO Network

(Health Care Provider Organizations)

HPO Lite

UTHealth · VCU Health · Washington University in St. Louis

RMCs

All of Us California

UC San Diego Health · UCI Health · UC Davis Health · UCSF · ECRMC · Cedars Sinai · Keck School of Medicine of USC

Illinois Precision Medicine Consortium

Northwestern Medicine · The University of Chicago · NorthShore · RUSH · The University of Illinois College of Medicine · UI Health · Cook County Health

All of Us New England

Brigham and Women's Hospital (Founding Member, Mass General Brigham) · Massachusetts General Hospital (Founding Member, Mass General Brigham) · Mass General Brigham · Boston Medical Center

Trans America Consortium

Henry Ford Health System · Spectrum Health · Essentia Health · BaylorScott&White Health · RELIANT MEDICAL GROUP (Part of OptumCare) · HealthPartners Institute

New York City Consortium

Columbia · Weill Cornell Medicine · NYC Health+ Hospitals | Harlem · NewYork-Presbyterian

All of Us Southern Network

UAB Medicine · USA Health · UAB Heersink · Cooper Green Mercy · Tuskegee University · Medical Center · LSU Health New Orleans · Tulane University

All of Us Southeast Enrollment Center

UHealth University of Miami Miller School of Medicine · Emory University · Morehouse School of Medicine · UF University of Florida · The University of North Carolina at Chapel Hill

All of Us Wisconsin

Marshfield Clinic Health System · Wisconsin · Froedtert & Medical College of Wisconsin · Gundersen Health System

All of Us Pennsylvania

University of Pittsburgh

University of Arizona and Banner Health

The University of Arizona · Banner Health · Mariposa Community Health Center

FQHCs (Federally Qualified Health Centers)

Cherokee · Jackson-Hinds Comprehensive Health Center · San Ysidro Health · Community Health Center, Inc. · Sun River Health

VA Medical Centers

VA · U.S. Department of Veterans Affairs · Veterans Health Administration

Participant Technology Systems Center (PTSC)

vibrent

Biobank

Mayo Clinic

Data & Research Center (DRC)

Vanderbilt University Medical Center · Broad Institute · verily · Columbia · School of Public Health, University of Michigan · Northwestern Medicine · UTHealth The University of Texas Health Science Center at Houston

Genomics Partners

Baylor College of Medicine Department of Molecular and Human Genetics · Johns Hopkins Medicine · UTHealth The University of Texas Health Science Center at Houston · HudsonAlpha · Broad Institute · color · Mass General Brigham · University of Washington

Note: These are not approved lockups and should not be repurposed on assets.

Thank You!

All of Us RESEARCH PROGRAM

joinallofus.org

NIH National Institutes of Health

AllofUs.nih.gov

@AllofUsResearch
#JoinAllofUs