CMSS Presents:

Deploying Cloud-based Platforms and Analytic Tools to Support Covid-19 and Beyond

August 6, 2020 | 12:00 – 1:30 pm ET

CMSS WEBINAR SERIES
Advancing Clinical Registries to Support Pandemic Treatment and Response

CMSS

CMSS WEBINAR SERIES

Advancing Clinical Registries to Support Pandemic Treatment and Response

The series will address key questions related to the rapid development, deployment and implementation of Covid-19 focused clinical registries and clinical repositories by specialty societies and academia.

SUMMER 2020 | FREE TO ATTEND

About the Series:

- Made possible with funding from the Gordon and Betty Moore Foundation
- To foster collaboration between specialty societies and academia, we are grateful to collaborate with the Association of Academic Medical Colleges

Continue the Conversation:

- Use #COVIDRegistries when tweeting about the webinar series
- Follow @CMSSMed and visit CMSS.org for frequent updates

CMSS WEBINAR SERIES

Advancing Clinical Registries to Support Pandemic Treatment and Response

Today's Webinar:

Deploying Cloud-based Platforms and Analytic Tools to Support Covid-19 and Beyond

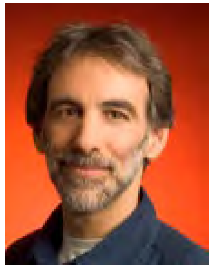
Moderator:

William J. Marks, Jr
MD, MS-HCM
Head of Clinical Science & Head of Neurology, Verily Life Sciences; Adjunct Clinical Professor of Neurology & Neurological Sciences, Stanford University School of Medicine

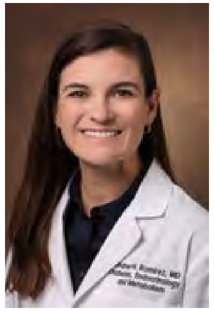
Host:

Helen Burstin, MD, MPH, MACP
Chief Executive Officer
Council of Medical Specialty Societies (CMSS)

Panelists:



David Glazer
Engineering Director
Verily Life Sciences



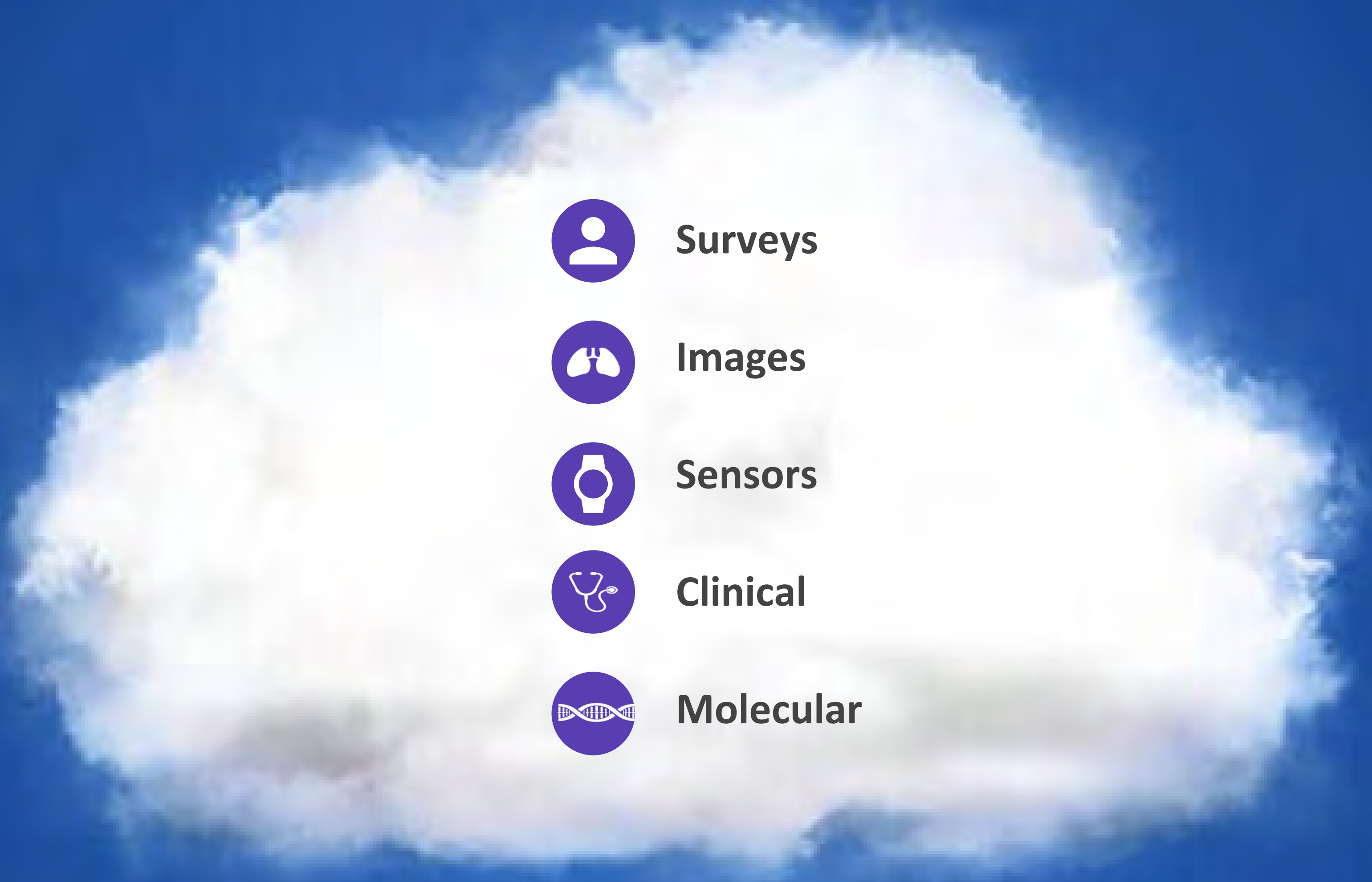
Andrea Ramirez, MD
Assistant Professor of Medicine, Vanderbilt University of Medicine

Chris Treml
Director of Operations, Data Science Institute, American College of Radiology

Biomedical Analysis in the Cloud

multi-modal, multi-source, multi-tool, multi-use

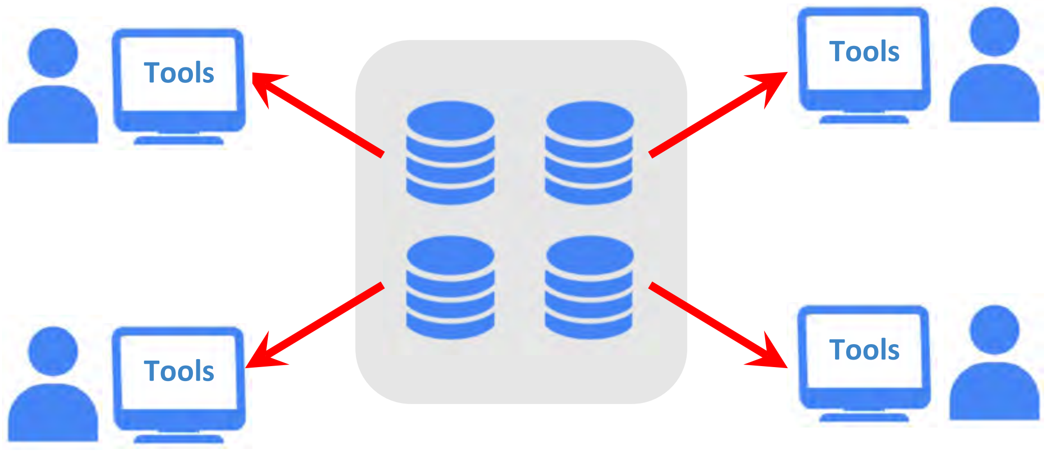
David Glazer, Verily Life Sciences
CMSS Webinar, 6-aug-2020



- Surveys
- Images
- Sensors
- Clinical
- Molecular

Traditional approach

Bring data to researchers



Discourages shared research

"Weakest link" security

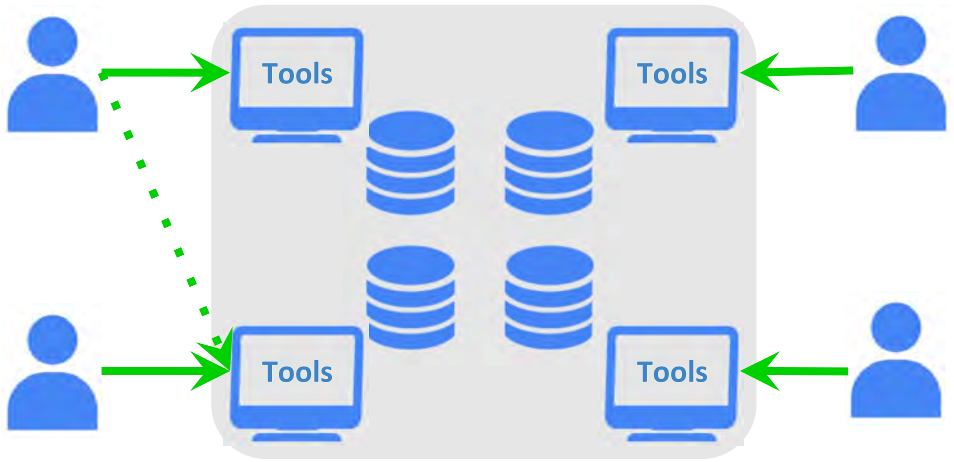
Huge infrastructure needed

Pay for multiple copies

Bespoke & unsupported tools

Cloud-centric approach

Bring researchers to data



Facilitates collaboration

Centralized security controls

Accessible to all researchers

Decreased cost of storage

Shared tool ecosystem

Uniting the biomedical ecosystem

DATA BIOSPHERE

VANDERBILT

verily

UNIVERSITY OF CALIFORNIA SANTA CRUZ

THE UNIVERSITY OF CHICAGO

BROAD INSTITUTE

DataBiosphere.org

Vision

Enable the next generation of collaborative biomedical research

Principles

Modular

composed of functional components with well-specified interface

Community driven

created by many groups to foster a diversity of ideas

Open

open-source licenses, software, arch to enable extensibility & reuse

Standards based

consistent with standards developed by coalitions such as GA4GH

Terra is a scalable
& secure
platform for
biomedical researchers
to access data,
run analysis tools,
and collaborate.

Terra's Value Proposition

Integrated data

Platform enables the integration and secure access of disparate data types and datasets, reducing data latency, costs, quality issues, and inefficiencies associated with fragmentation

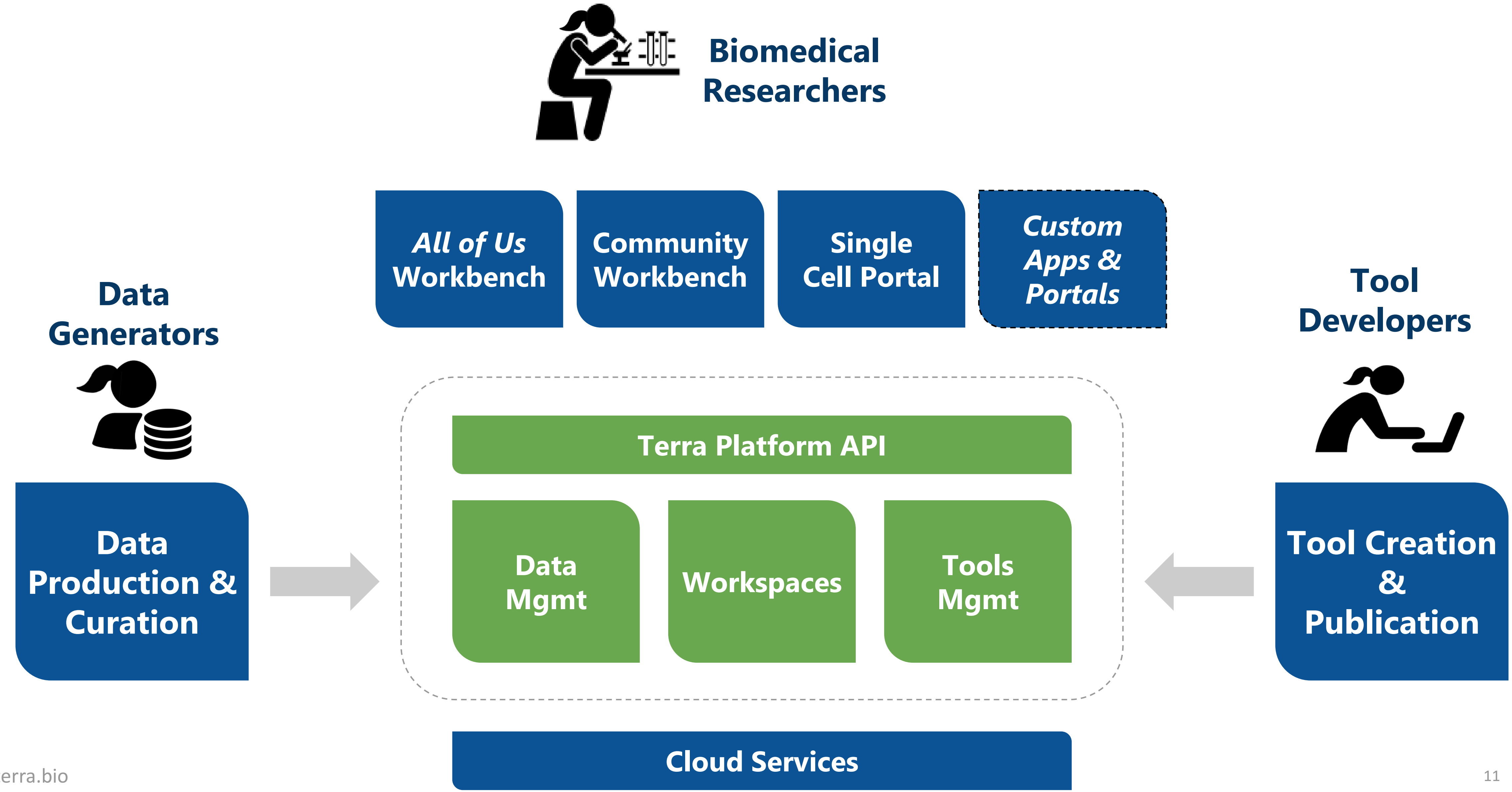
Integrated tools

Platform integrates proprietary and third-party tools to run leading batch and interactive analysis, reducing the challenges associated with multiple systems and frameworks

Collaborative workspaces

Platform enables rich collaboration with sophisticated provenance, security controls, and compliance for faster and more effective analysis within and across organizations

Terra's architecture primarily serves 3 roles



Biomedical Researchers

All of Us Workbench

Community Workbench

Single Cell Portal

Custom Apps & Portals

Data Generators

Data Production & Curation

Terra Platform API

Data Mgmt

Workspaces

Tools Mgmt

Tool Developers

Tool Creation & Publication

Cloud Services

Partners & Datasets

All of Us
RESEARCH PROGRAM

HUMAN
CELL
ATLAS

AMP PD

NHGRI AnVIL

NHLBI
TOPMed

NIH
National Heart, Lung, and Blood Institute
BioData
CATALYST

Nurses' Health Study

NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

biobank uk
Improving the health of future generations

Single Cell
PORTAL

NIH
NATIONAL CANCER INSTITUTE

Project Baseline

ENCODE

Tools currently available in Terra

Cromwell

jupyter

hail

R Studio

TensorFlow

K Keras

Bokeh

APACHE Spark

seaborn

biopython

scikit learn

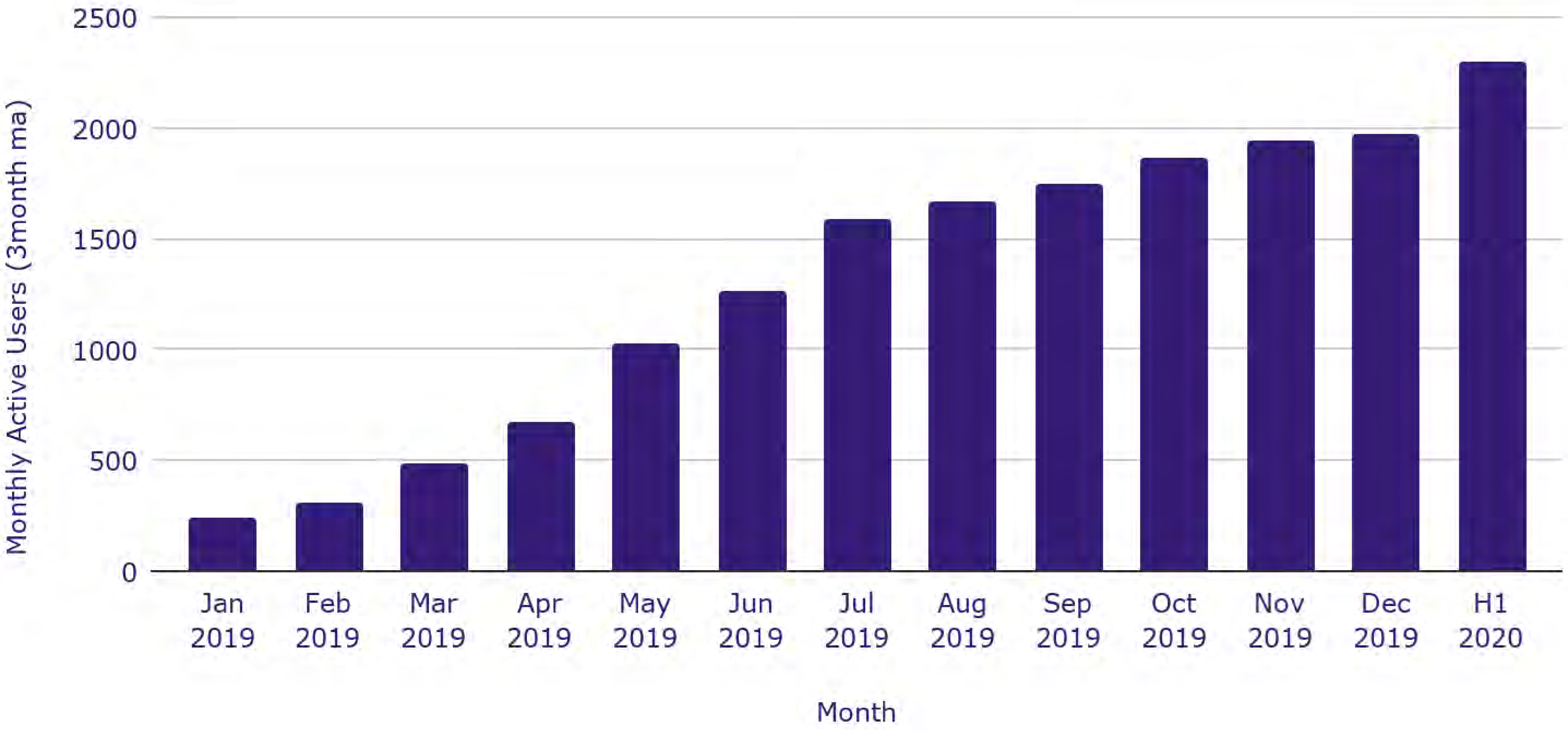
NumPy

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

pandas
$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$

Dockstore
Create, Share, Use

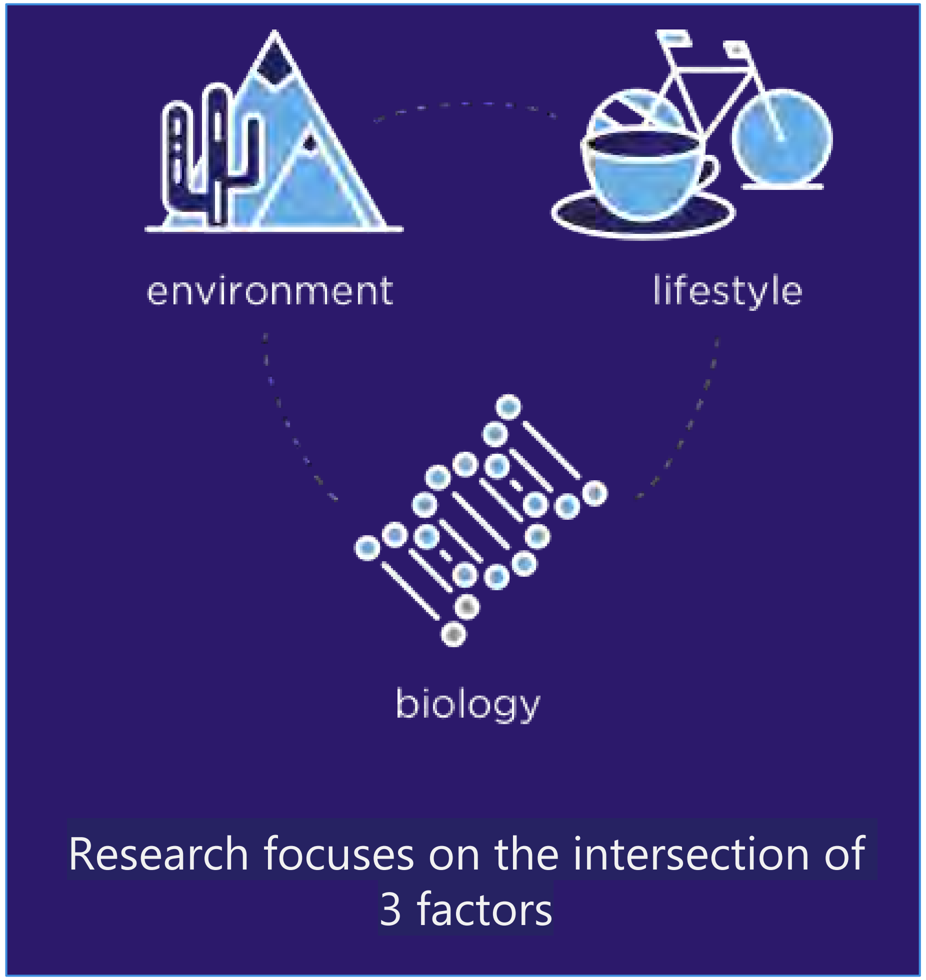
Thousands of researchers use Terra every month to analyze biomedical data



Terra use case:
All of Us Research Program

All of Us Research Program

- Engage **1,000,000 or more** U.S. research participants
- Share **tissue** samples**, genetic** data**, lifestyle** information**,** electronic **health records**
- Pioneer a new model of research that emphasizes **engaged research participants**, **responsible data sharing**, and **privacy protection**

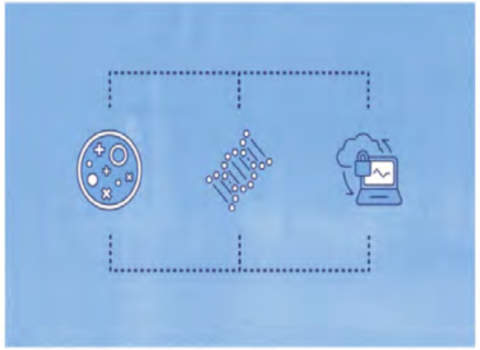


Research focuses on the intersection of 3 factors

All of Us and the All of Us logo are service marks of the U.S. Department of Health and Human Services (HHS).

All of Us Program Components

Health Care Provider Organizations



Genome Centers

Data and Research Center

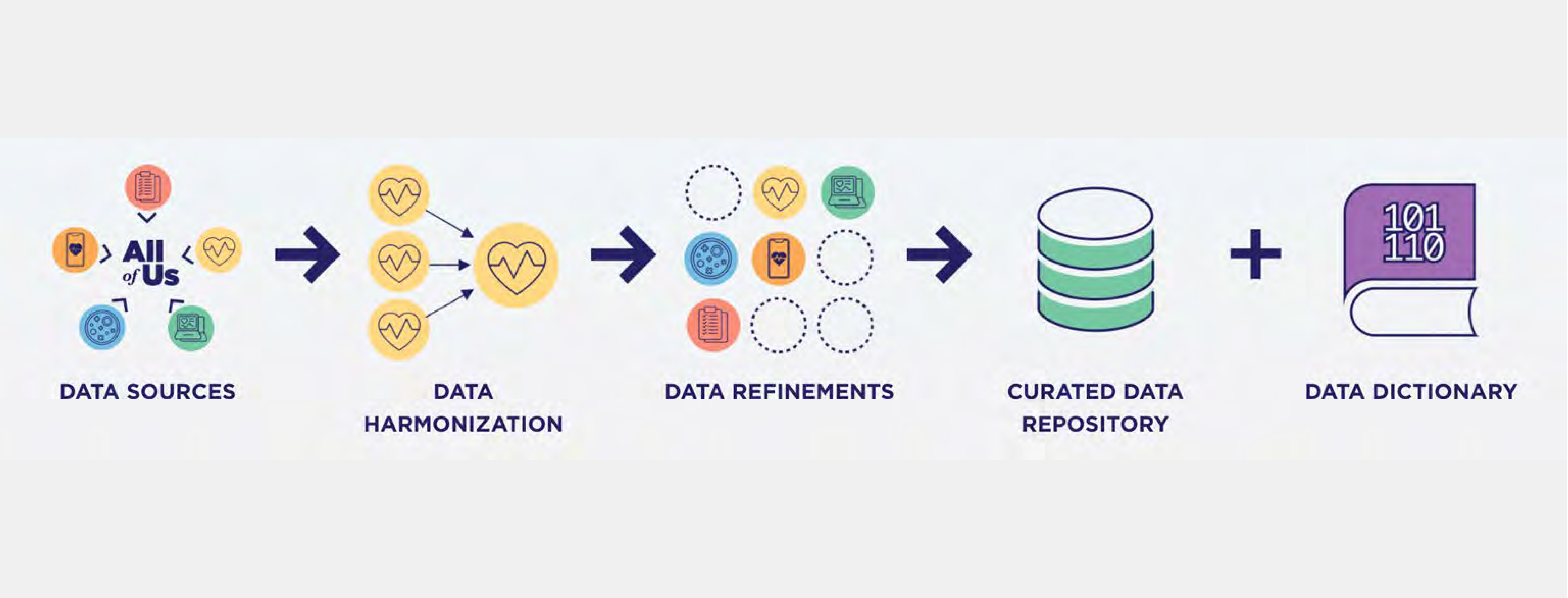
VANDERBILT UNIVERSITY
MEDICAL CENTER

BROAD INSTITUTE

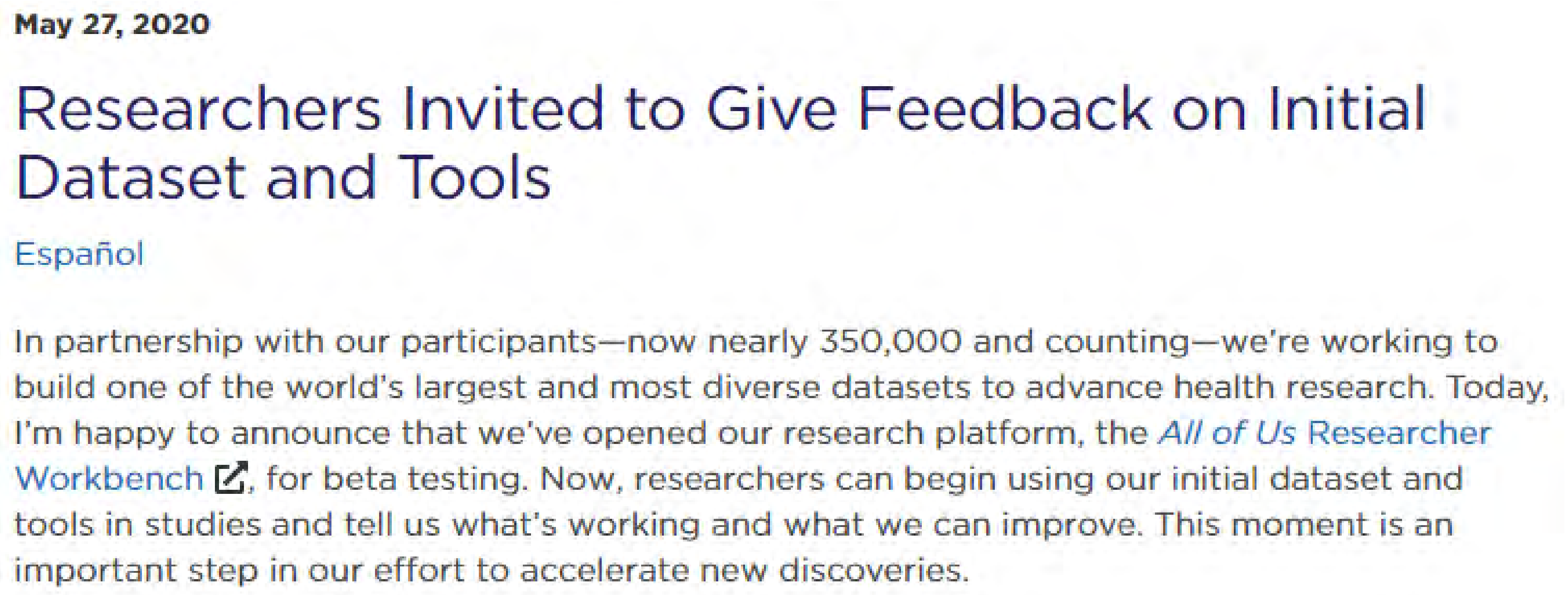
Biobank

Participant Technology Systems Center

Data Curation Process



All of Us Research Program Begins Beta Testing



May 27, 2020

Researchers Invited to Give Feedback on Initial Dataset and Tools

Español

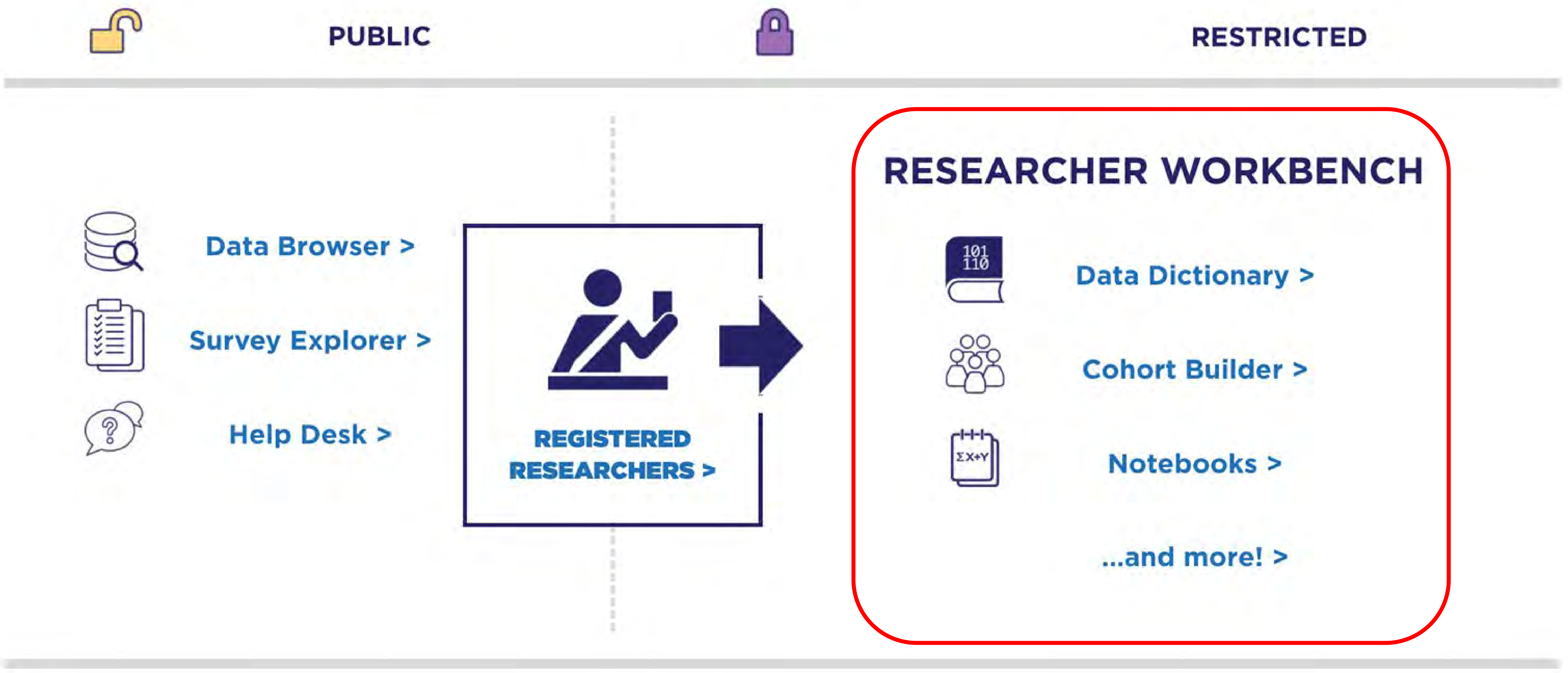
In partnership with our participants—now nearly 350,000 and counting—we're working to build one of the world's largest and most diverse datasets to advance health research. Today, I'm happy to announce that we've opened our research platform, the *All of Us* Researcher Workbench, for beta testing. Now, researchers can begin using our initial dataset and tools in studies and tell us what's working and what we can improve. This moment is an important step in our effort to accelerate new discoveries.

350,000+
Participants

211,000+
Electronic Health Records

277,000+
Biosamples Received

Research Hub → Researcher Workbench



PUBLIC

RESTRICTED

Data Browser >

Survey Explorer >

Help Desk >

REGISTERED RESEARCHERS >

RESEARCHER WORKBENCH

Data Dictionary >

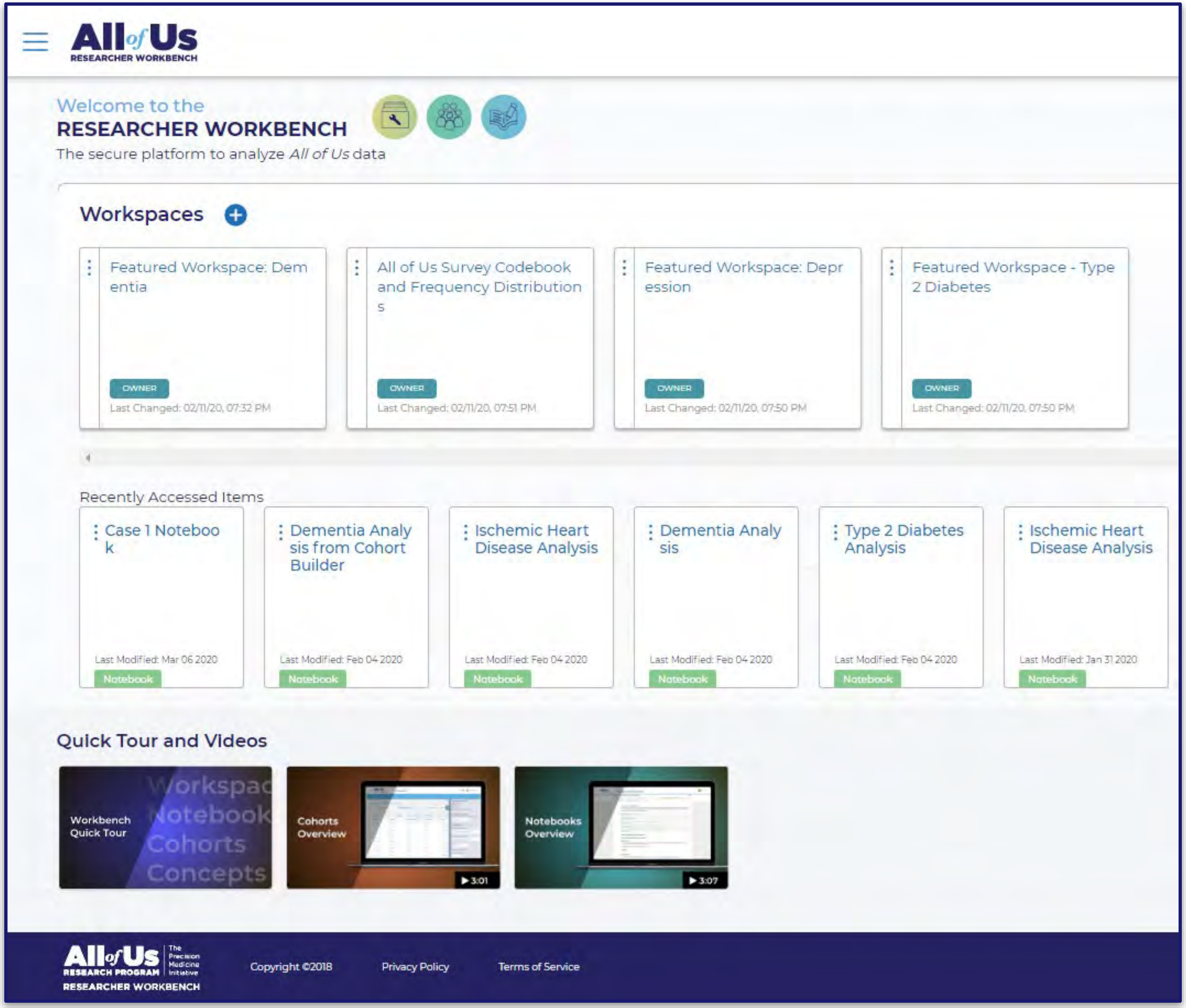
Cohort Builder >

Notebooks >

...and more! >

All of Us
DATA & RESEARCH CENTER

Welcome to the Workbench



All of Us
RESEARCHER WORKBENCH

Welcome to the
RESEARCHER WORKBENCH
The secure platform to analyze *All of Us* data

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

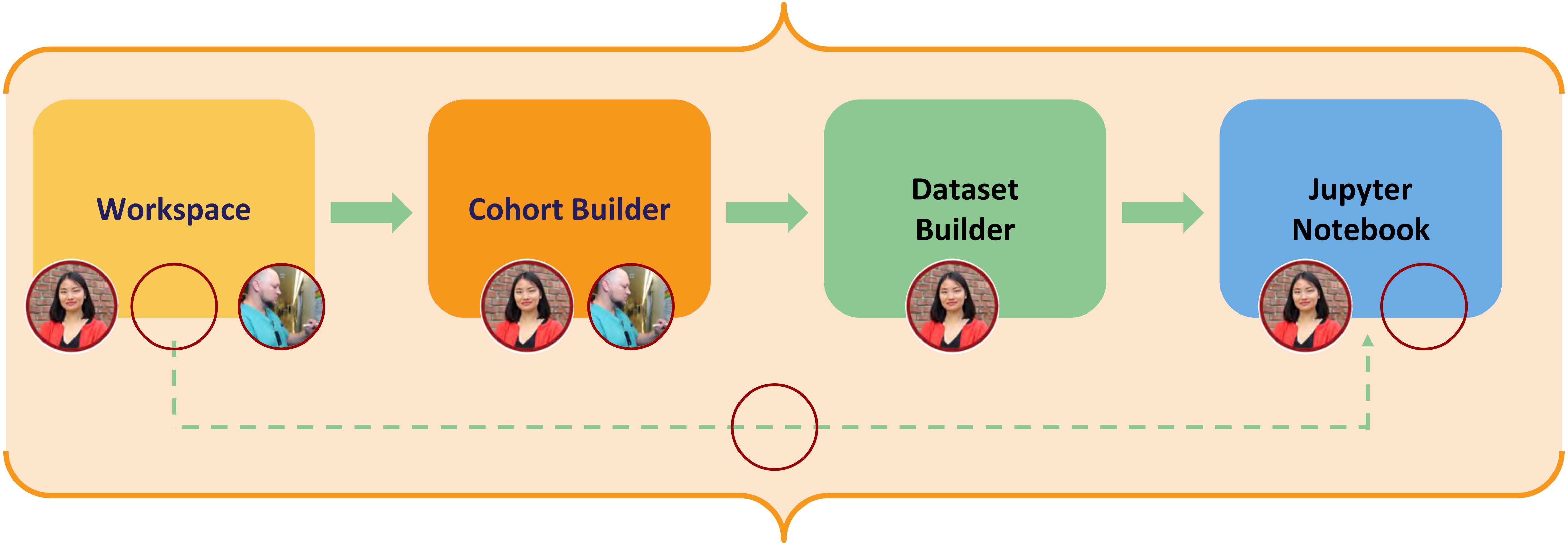
Quick Tour and Videos

All of Us RESEARCH PROGRAM
The Precision Medicine Initiative
RESEARCHER WORKBENCH
Copyright ©2018
Privacy Policy
Terms of Service

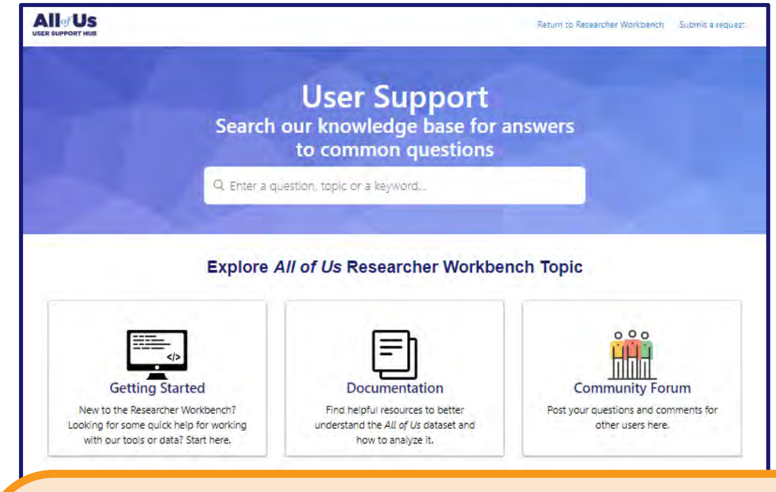
The Workbench *contains* the tools researchers need to learn about, access and analyze *All of Us* data.

Collaborative Workspaces

Within the Researcher Workbench, researchers run analyses, individually or collaboratively, in "Workspaces."



Researchers have access to integrated support



User Support

Search our knowledge base for answers to common questions

Explore *All of Us* Researcher Workbench Topic

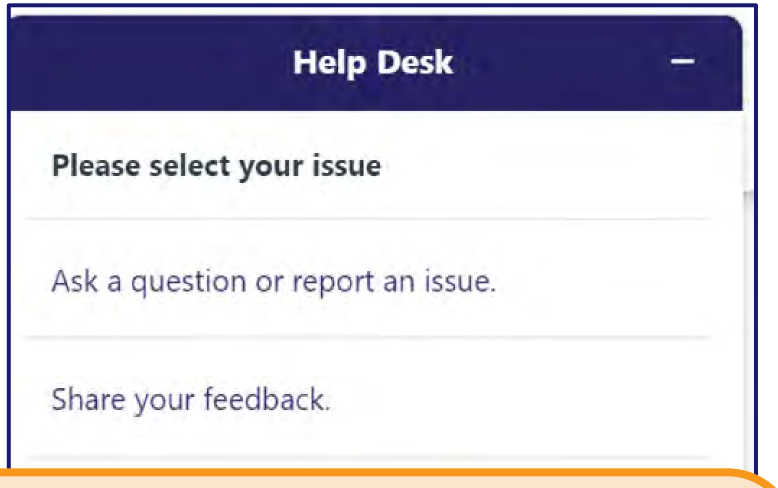
Getting Started
New to the Researcher Workbench? Looking for some quick help for working with our tools or data? Start here.

Documentation
Find helpful resources to better understand the *All of Us* dataset and how to analyze it.

Community Forum
Post your questions and comments for other users here.

All *of* Us

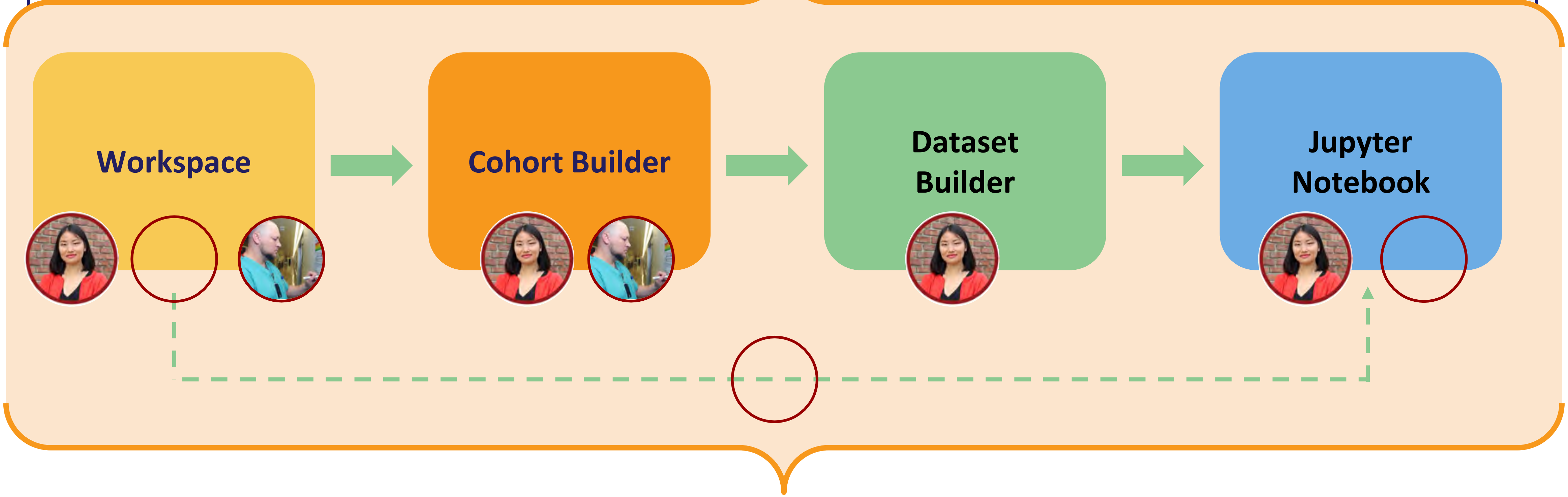
USER SUPPORT HUB



Please select your issue

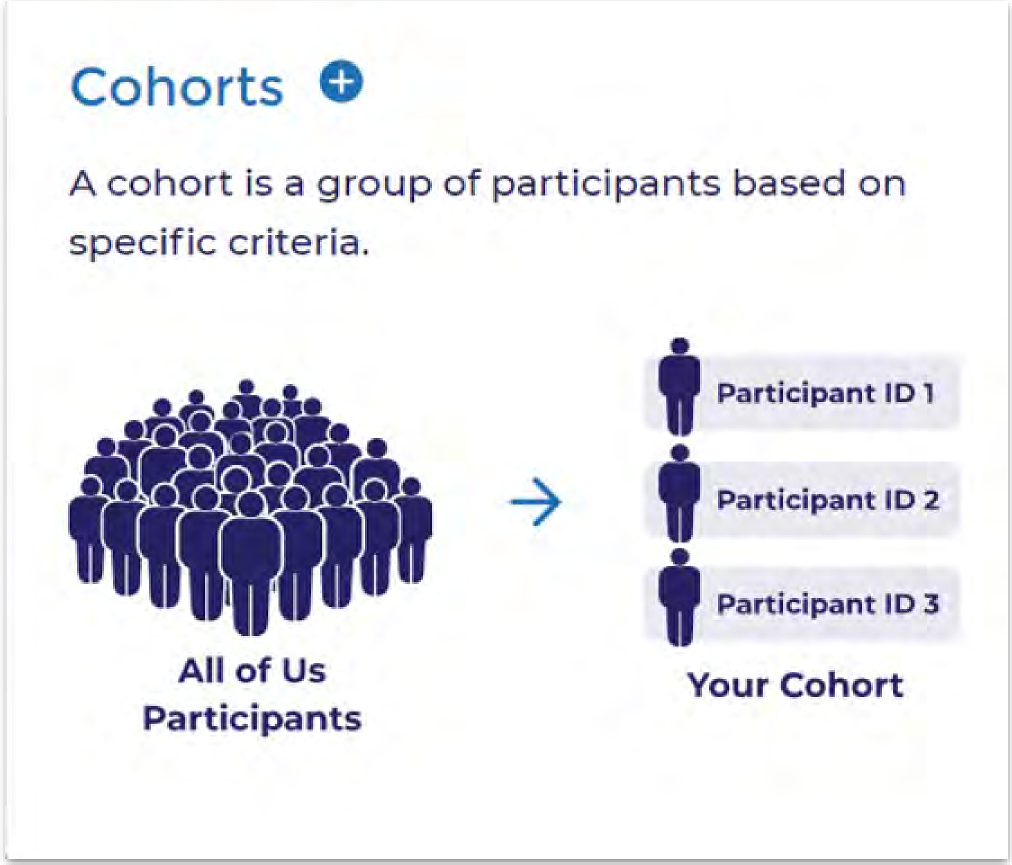
Ask a question or report an issue.

Share your feedback.



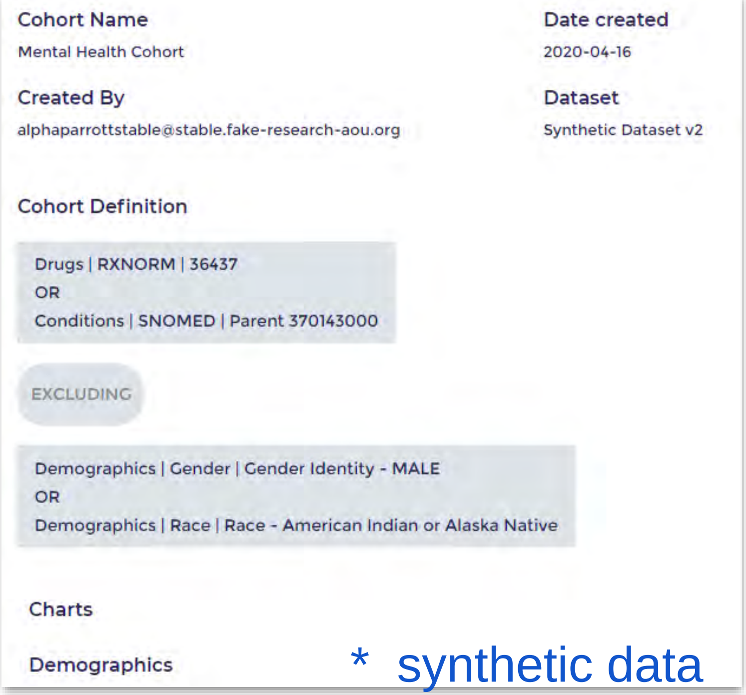
Cohort Builder: Select Participant Groups

- Once you create your workspace, you can create a cohort for your study using Cohort Builder, where you will select participants to include or exclude from your study based on criteria of interest for your research



Cohorts

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



Cohort Name
Mental Health Cohort

Date created
2020-04-16

Created By
alphaparrottstable@stable.fake-research-aou.org

Dataset
Synthetic Dataset v2

Cohort Definition

Drugs | RXNORM | 36437
OR
Conditions | SNOMED | Parent 370143000

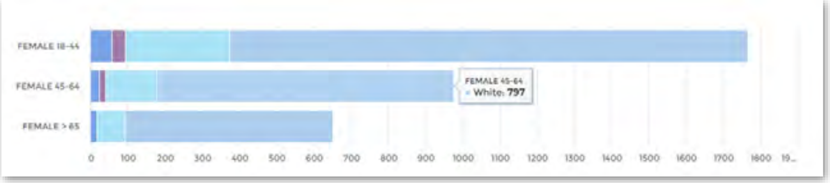
EXCLUDING

Demographics | Gender | Gender Identity - MALE
OR
Demographics | Race | Race - American Indian or Alaska Native

Charts

Demographics

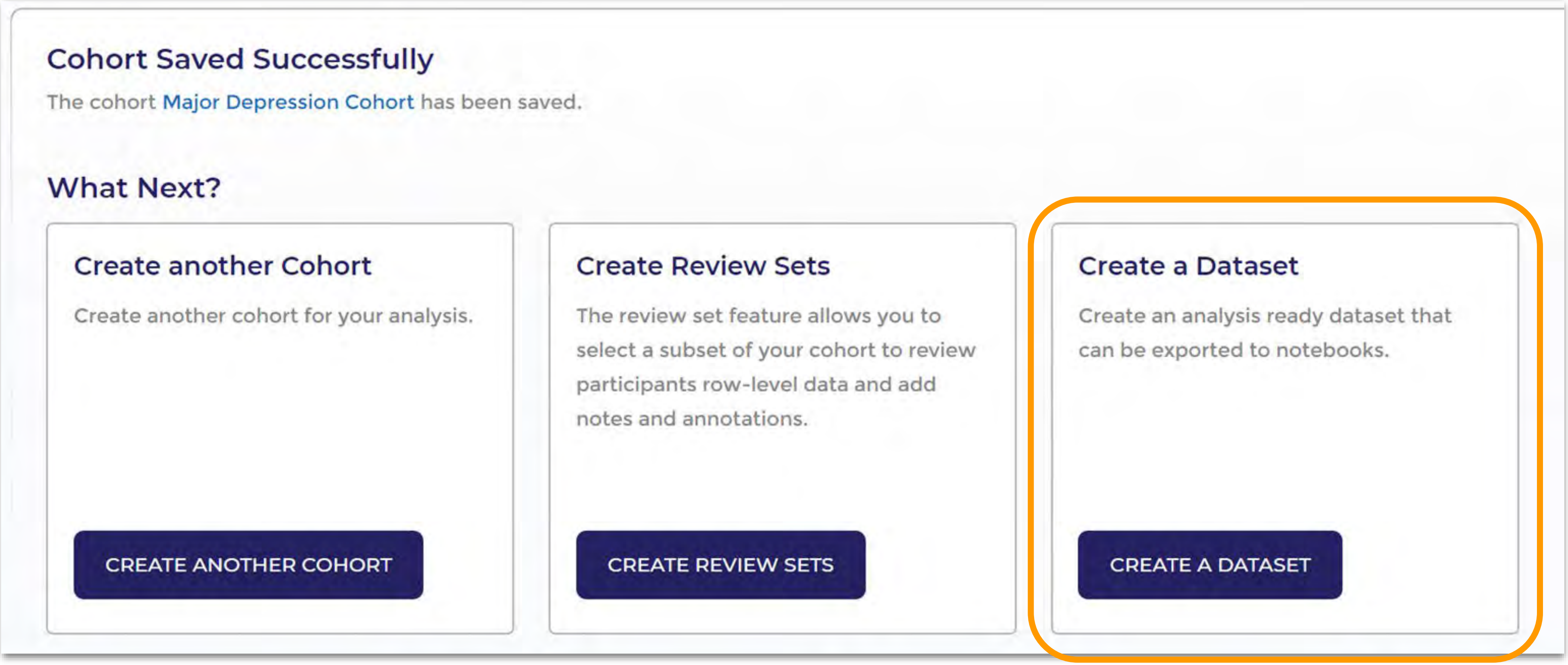
* synthetic data



Workspace → **Cohort Builder**

Cohort Review -> Dataset Creation

- After creating your cohort, you can check your work using a Cohort Review Set



Cohort Saved Successfully

The cohort Major Depression Cohort has been saved.

What Next?

Create another Cohort

Create another cohort for your analysis.

Create Review Sets

The review set feature allows you to select a subset of your cohort to review participants row-level data and add notes and annotations.

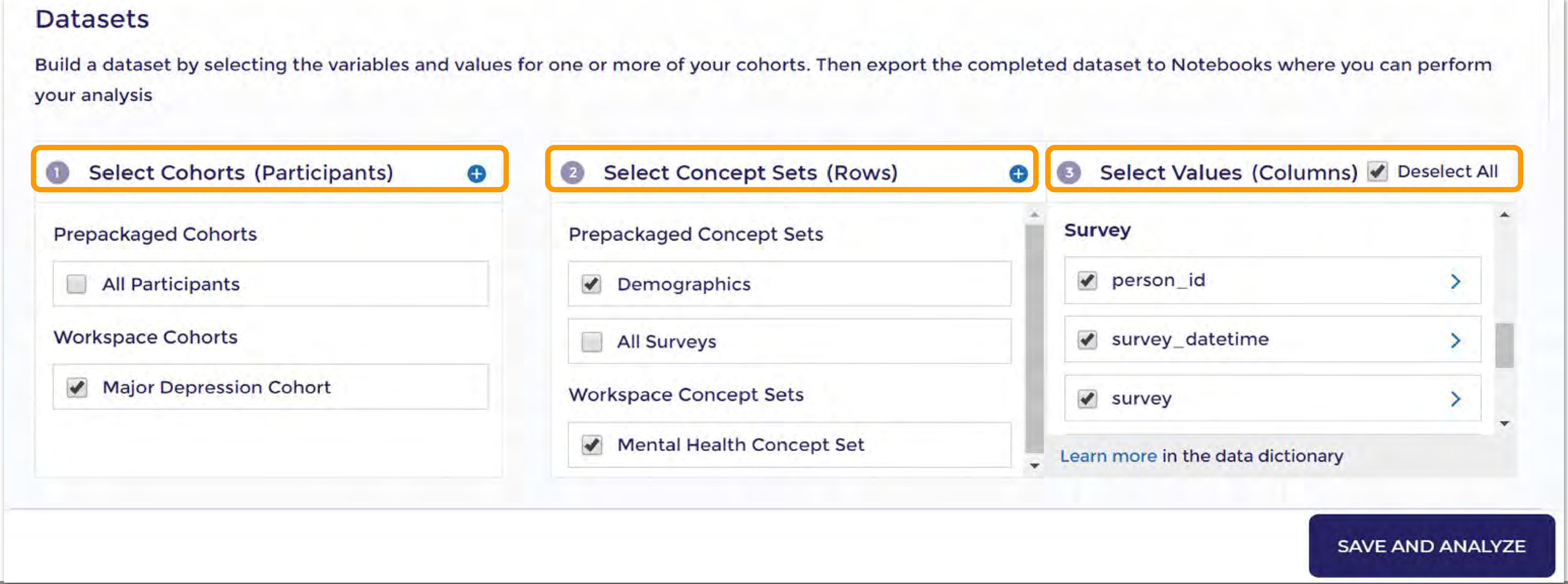
Create a Dataset

Create an analysis ready dataset that can be exported to notebooks.

Workspace → **Cohort Builder**

Cohorts + Concept Sets + Values = Datasets

- Once you have your cohort, you can start to put together your dataset.
- Your dataset is made up of a cohort, a concept set, and selected values.



Datasets

Build a dataset by selecting the variables and values for one or more of your cohorts. Then export the completed dataset to Notebooks where you can perform your analysis

1 Select Cohorts (Participants)

Prepackaged Cohorts

☐ All Participants

Workspace Cohorts

☑ Major Depression Cohort

2 Select Concept Sets (Rows)

Prepackaged Concept Sets

☑ Demographics

☐ All Surveys

Workspace Concept Sets

☑ Mental Health Concept Set

3 Select Values (Columns) ☑ Deselect All

Survey

☑ person_id

☑ survey_datetime

☑ survey

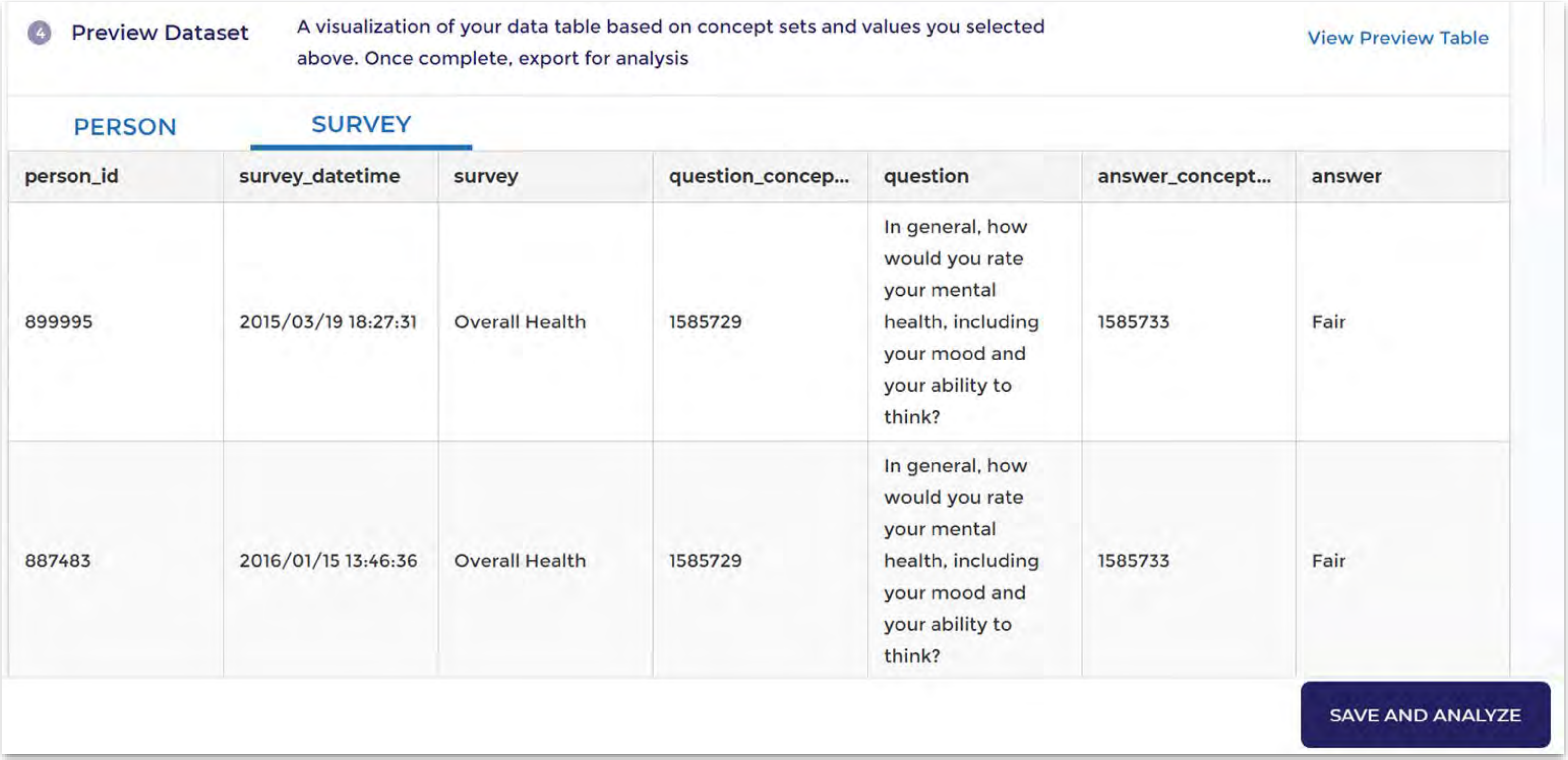
Learn more in the data dictionary

SAVE AND ANALYZE

Workspace → Cohort Builder → Dataset Builder

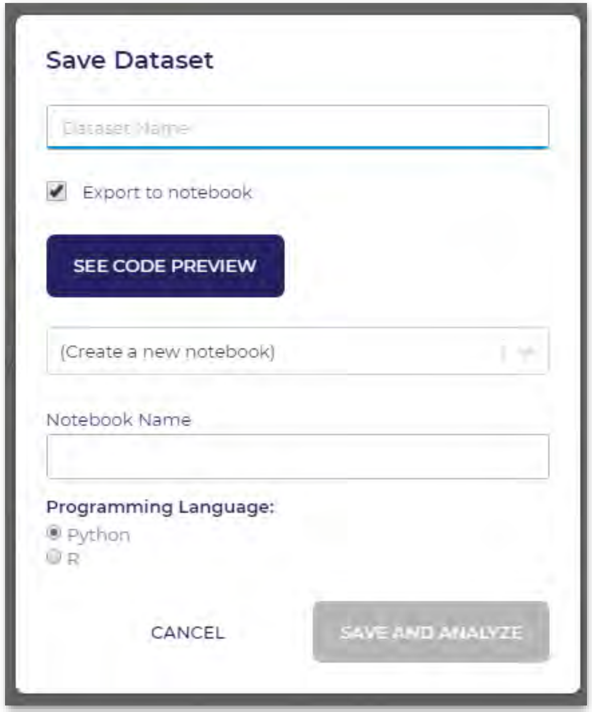
Dataset Builder: Preview and Export

- Once you select your cohort, your concept set(s), and values, you now have your dataset
- Now you can preview your dataset before saving it and exporting it to a notebook for analysis



Preview Dataset A visualization of your data table based on concept sets and values you selected above. Once complete, export for analysis

person_id	survey_datetime	survey	question_concep...	question	answer_concept...	answer
899995	2015/03/19 18:27:31	Overall Health	1585729	In general, how would you rate your mental health, including your mood and your ability to think?	1585733	Fair
887483	2016/01/15 13:46:36	Overall Health	1585729	In general, how would you rate your mental health, including your mood and your ability to think?	1585733	Fair



Save Dataset

Dataset Name

☑ Export to notebook

(Create a new notebook)

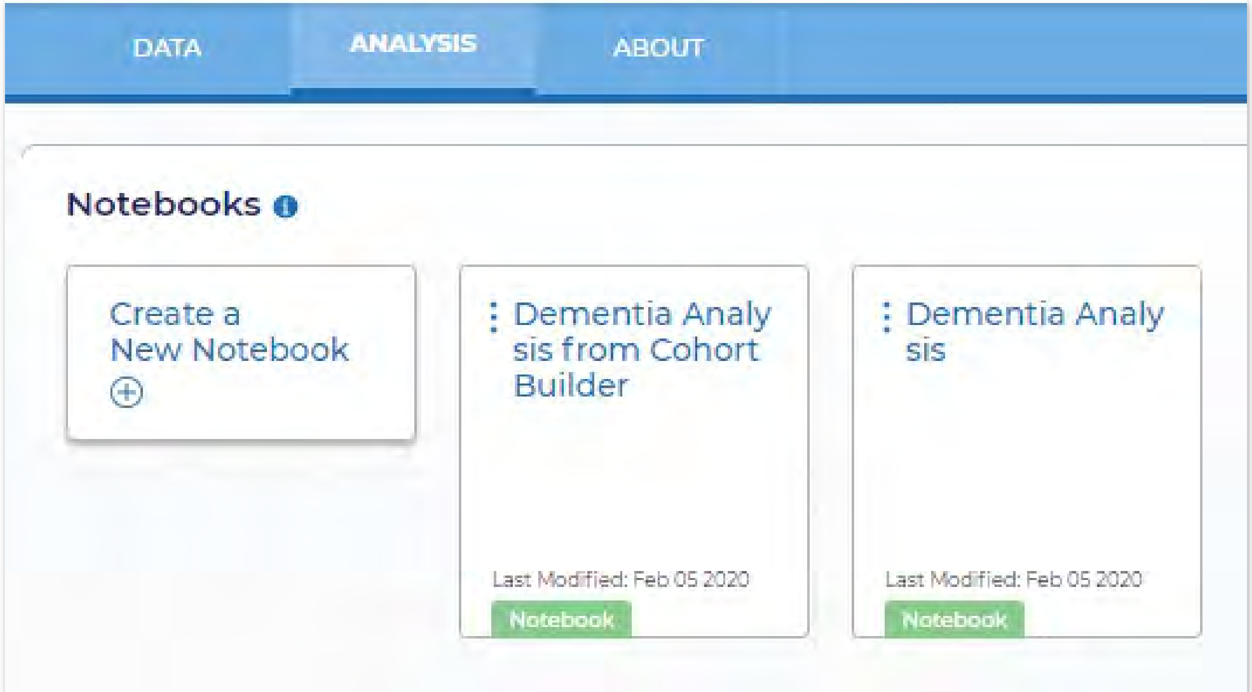
Notebook Name

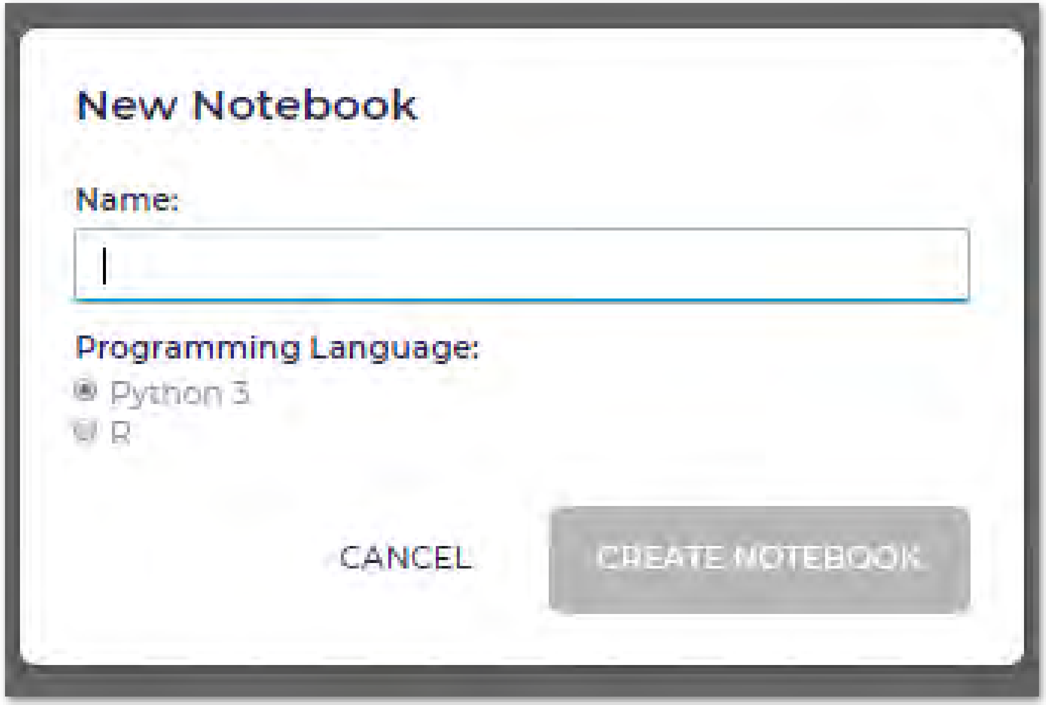
Programming Language:
● Python
○ R

* synthetic data

Workspace → Cohort Builder → Dataset Builder

Data Analysis: Jupyter Notebooks





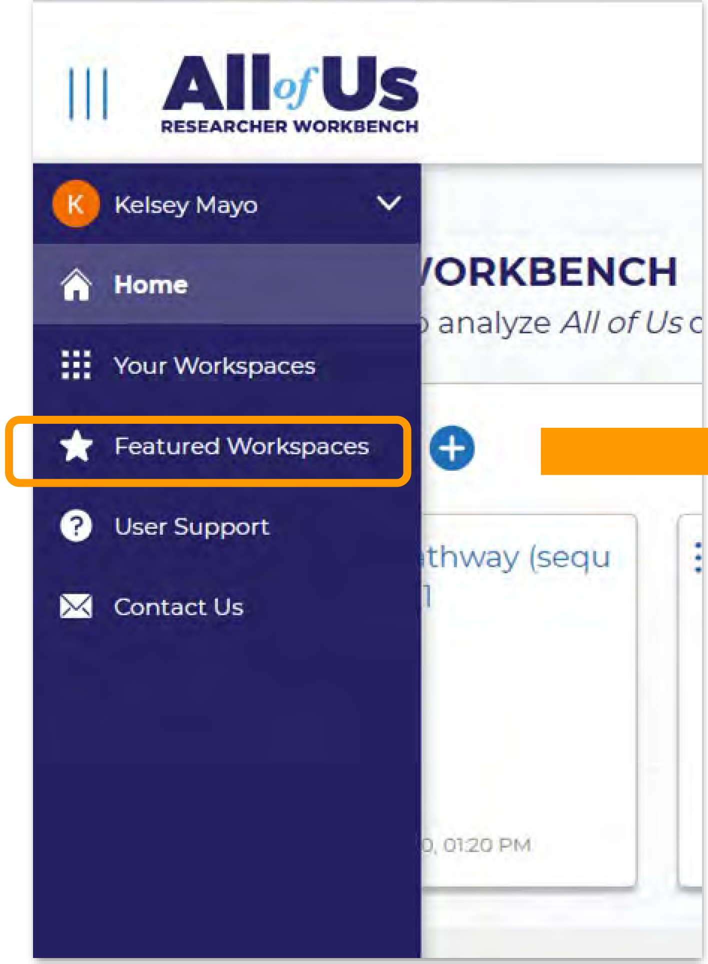
Name:

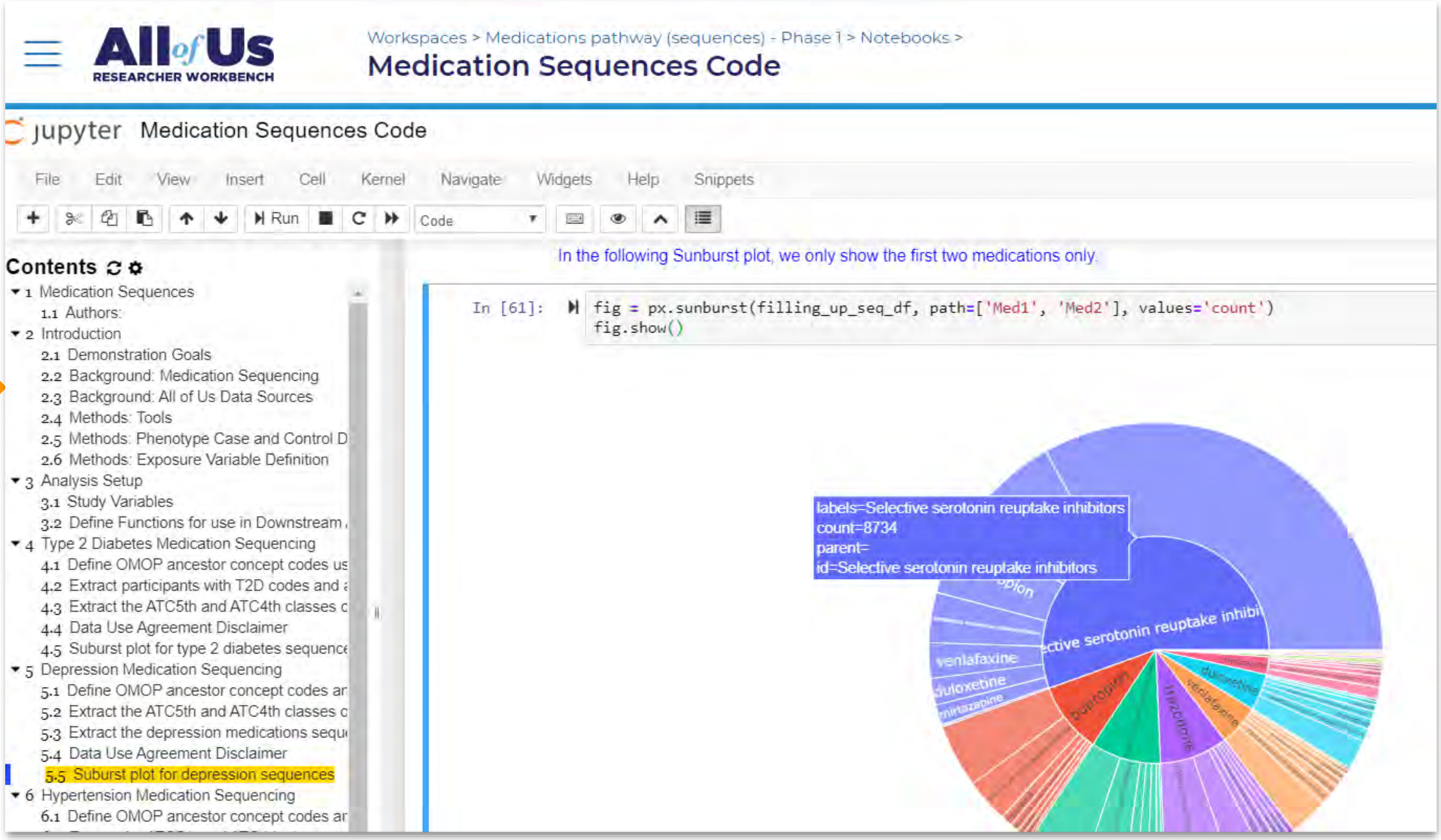
Programming Language:
● Python 3
○ R

- ➢ Jupyter Notebooks can be used for analysis in **R and Python**
- ➢ Notebooks come with ready-made code that can be inserted called “Snippets”
- ➢ Each user gets a virtual machine (computing system) per workspace

Workspace → **Cohort Builder** → **Dataset Builder** → **Jupyter Notebook**

Tools@Beta: Powerful, Flexible Tools for Reproducible Science





- 1 Medication Sequences
  - 1.1 Authors:
- 2 Introduction
  - 2.1 Demonstration Goals
  - 2.2 Background: Medication Sequencing
  - 2.3 Background: All of Us Data Sources
  - 2.4 Methods: Tools
  - 2.5 Methods: Phenotype Case and Control D
  - 2.6 Methods: Exposure Variable Definition
- 3 Analysis Setup
  - 3.1 Study Variables
  - 3.2 Define Functions for use in Downstream
- 4 Type 2 Diabetes Medication Sequencing
  - 4.1 Define OMOP ancestor concept codes us
  - 4.2 Extract participants with T2D codes and
  - 4.3 Extract the ATC5th and ATC4th classes c
  - 4.4 Data Use Agreement Disclaimer
  - 4.5 Suburst plot for type 2 diabetes sequence
- 5 Depression Medication Sequencing
  - 5.1 Define OMOP ancestor concept codes ar
  - 5.2 Extract the ATC5th and ATC4th classes c
  - 5.3 Extract the depression medications sequ
  - 5.4 Data Use Agreement Disclaimer
  - 5.5 Suburst plot for depression sequences
- 6 Hypertension Medication Sequencing
  - 6.1 Define OMOP ancestor concept codes ar

In the following Sunburst plot, we only show the first two medications only.

```
In [61]:  fig = px.sunburst(filling_up_seq_df, path=['Med1', 'Med2'], values='count')
          fig.show()
```

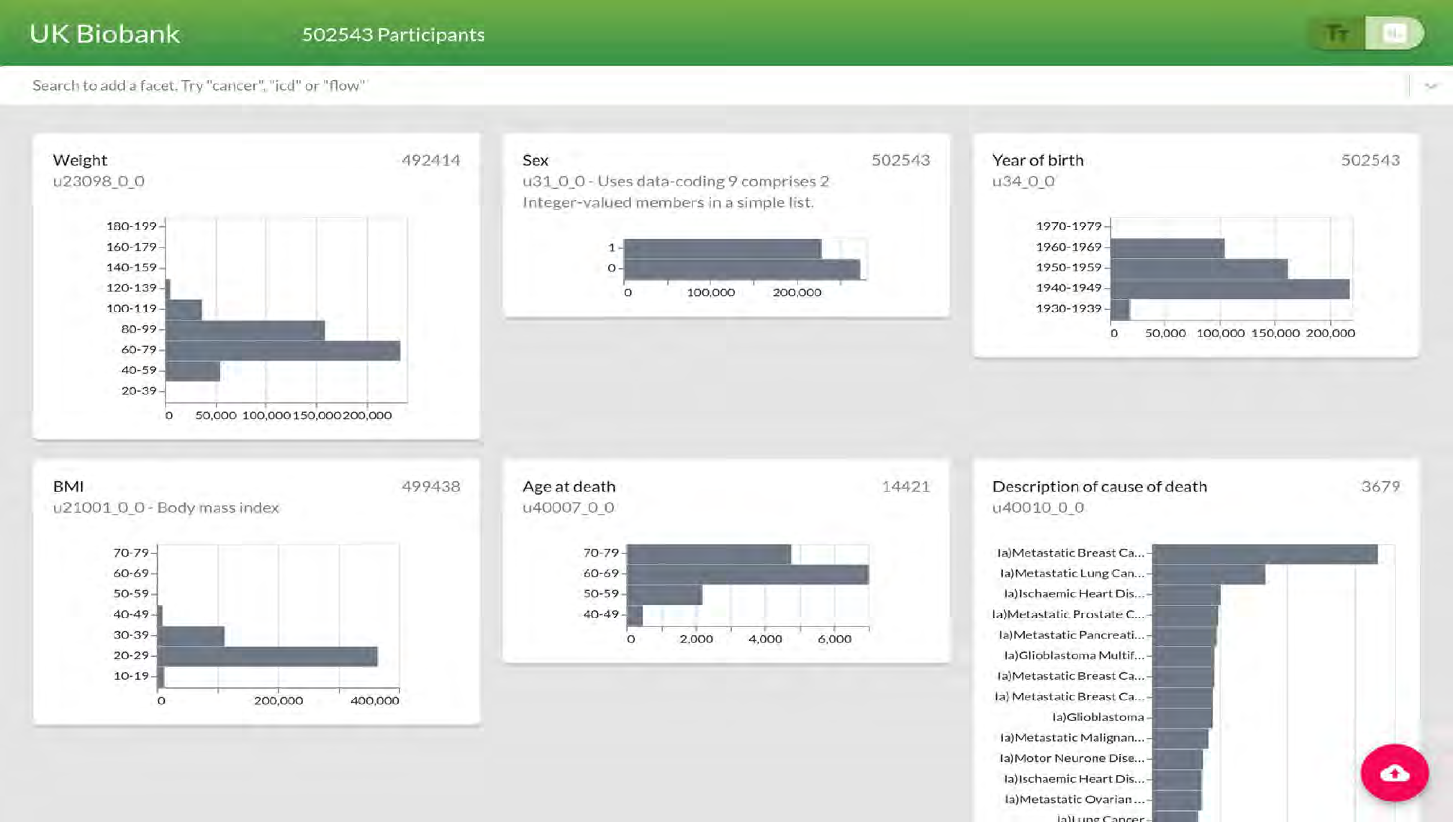
```
labels=Selective serotonin reuptake inhibitors
count=8734
parent=
id=Selective serotonin reuptake inhibitors
```

Workspace → **Cohort Builder** → **Dataset Builder** → **Jupyter Notebook**

All of Us
DATA & RESEARCH CENTER

Presented on Behalf of the Data and Research Center (DRC)

Terra use case:
UK Biobank



UK Biobank

502543 Participants

Search to add a facet. Try "cancer", "icd" or "flow"

Weight 492414
u23098_0_0

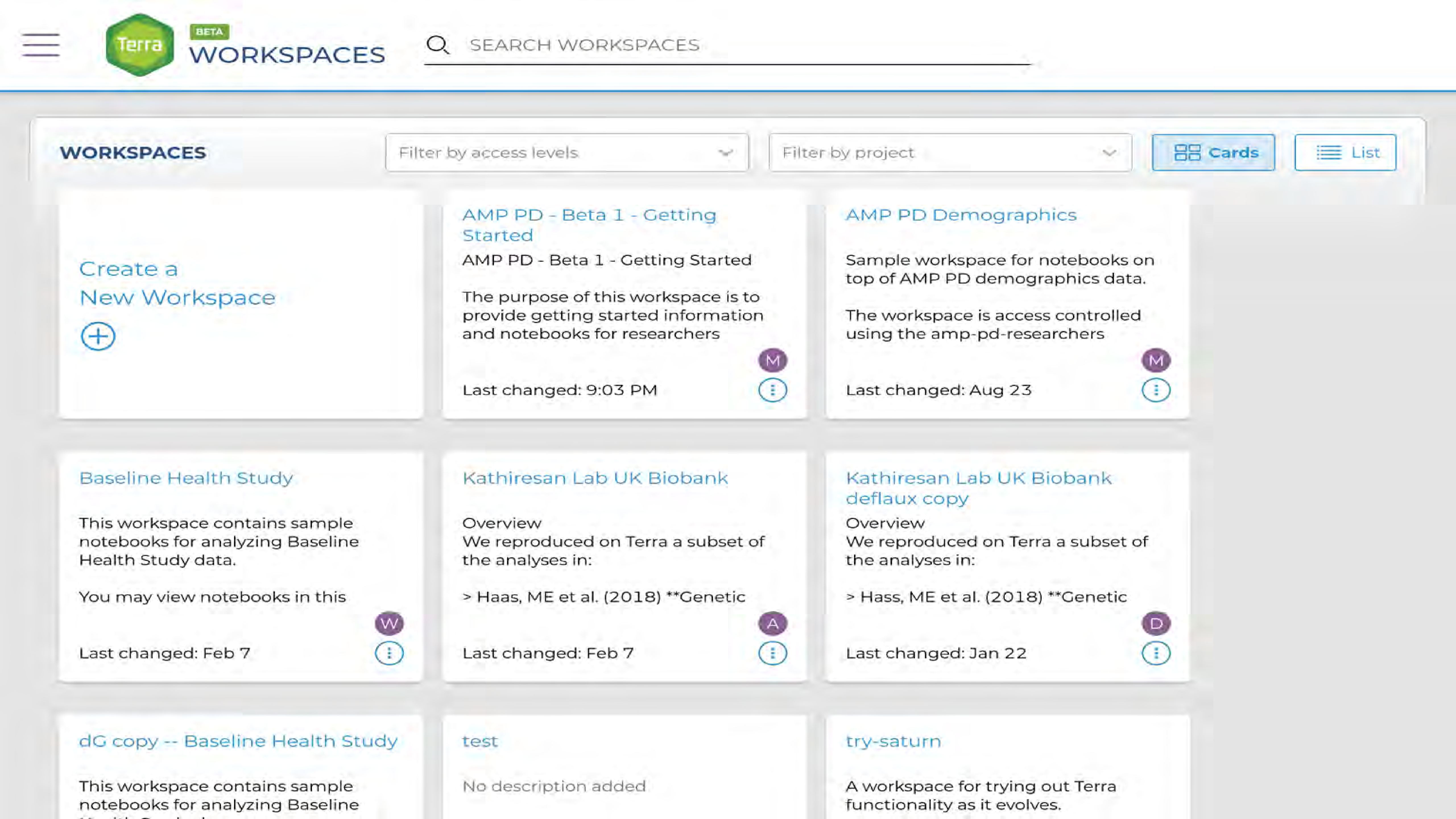
Sex 502543
u31_0_0 - Uses data-coding 9 comprises 2 Integer-valued members in a simple list.

Year of birth 502543
u34_0_0

BMI 499438
u21001_0_0 - Body mass index

Age at death 14421
u40007_0_0

Description of cause of death 3679
u40010_0_0



Create a New Workspace

AMP PD - Beta 1 - Getting Started

AMP PD - Beta 1 - Getting Started

The purpose of this workspace is to provide getting started information and notebooks for researchers

Last changed: 9:03 PM

AMP PD Demographics

Sample workspace for notebooks on top of AMP PD demographics data.

The workspace is access controlled using the amp-pd-researchers

Last changed: Aug 23

Baseline Health Study

This workspace contains sample notebooks for analyzing Baseline Health Study data.

You may view notebooks in this

Last changed: Feb 7

Kathiresan Lab UK Biobank

Overview
We reproduced on Terra a subset of the analyses in:

> Haas, ME et al. (2018) **Genetic

Last changed: Feb 7

Kathiresan Lab UK Biobank deflaux copy

Overview
We reproduced on Terra a subset of the analyses in:

> Hass, ME et al. (2018) **Genetic

Last changed: Jan 22

dG copy -- Baseline Health Study

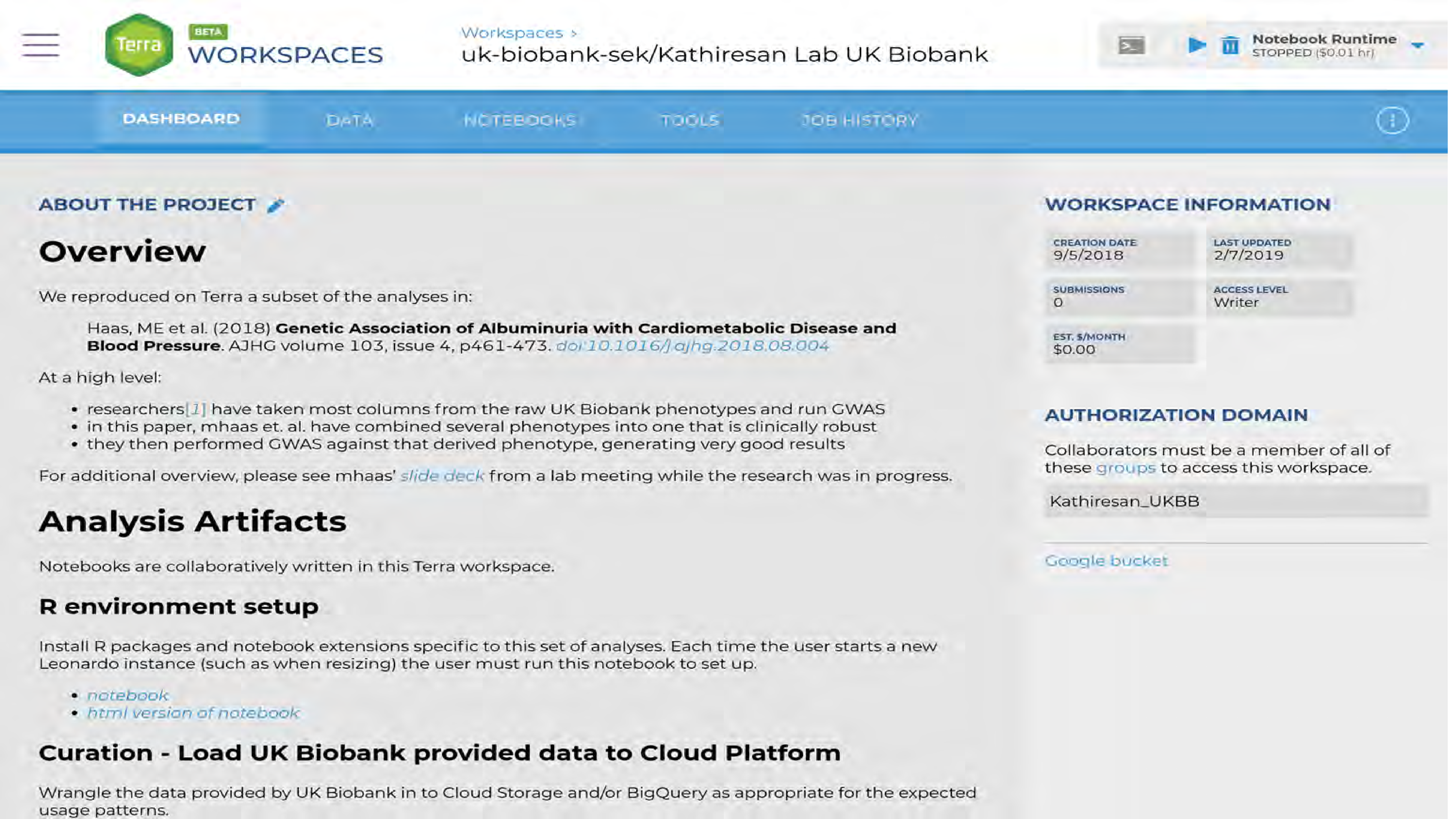
This workspace contains sample notebooks for analyzing Baseline

test

No description added

try-saturn

A workspace for trying out Terra functionality as it evolves.



CREATION DATE	LAST UPDATED
9/5/2018	2/7/2019
SUBMISSIONS	ACCESS LEVEL
0	Writer
EST. $/MONTH	
$0.00	

AUTHORIZATION DOMAIN

Collaborators must be a member of all of these groups to access this workspace.

Kathiresan_UKBB

Google bucket

Overview

We reproduced on Terra a subset of the analyses in:

> Haas, ME et al. (2018) **Genetic Association of Albuminuria with Cardiometabolic Disease and Blood Pressure**. AJHG volume 103, issue 4, p461-473. *doi:10.1016/j.ajhg.2018.08.004*

At a high level:

- researchers*[1]* have taken most columns from the raw UK Biobank phenotypes and run GWAS
- in this paper, mhaas et. al. have combined several phenotypes into one that is clinically robust
- they then performed GWAS against that derived phenotype, generating very good results

For additional overview, please see mhaas' *slide deck* from a lab meeting while the research was in progress.

Analysis Artifacts

Notebooks are collaboratively written in this Terra workspace.

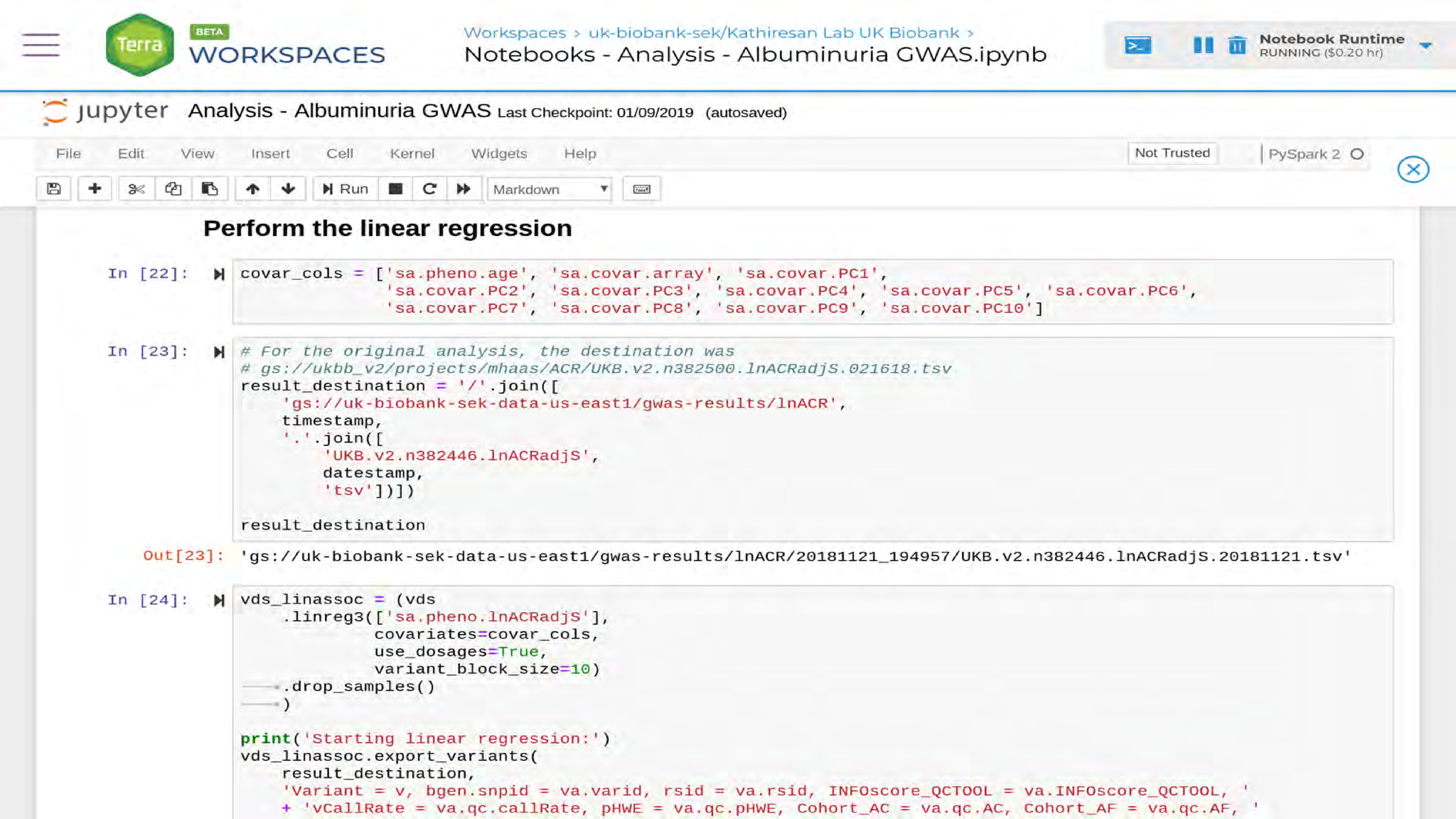
R environment setup

Install R packages and notebook extensions specific to this set of analyses. Each time the user starts a new Leonardo instance (such as when resizing) the user must run this notebook to set up.

- *notebook*
- *html version of notebook*

Curation - Load UK Biobank provided data to Cloud Platform

Wrangle the data provided by UK Biobank in to Cloud Storage and/or BigQuery as appropriate for the expected usage patterns.



Perform the linear regression

In [22]:

```
covar_cols = ['sa.pheno.age', 'sa.covar.array', 'sa.covar.PC1',
              'sa.covar.PC2', 'sa.covar.PC3', 'sa.covar.PC4', 'sa.covar.PC5', 'sa.covar.PC6',
              'sa.covar.PC7', 'sa.covar.PC8', 'sa.covar.PC9', 'sa.covar.PC10']
```

In [23]:

```
# For the original analysis, the destination was
# gs://ukbb_v2/projects/mhaas/ACR/UKB.v2.n382500.lnACRadjS.021618.tsv
result_destination = '/'.join([
    'gs://uk-biobank-sek-data-us-east1/gwas-results/lnACR',
    timestamp,
    '.'.join([
        'UKB.v2.n382446.lnACRadjS',
        datestamp,
        'tsv'])])

result_destination
```

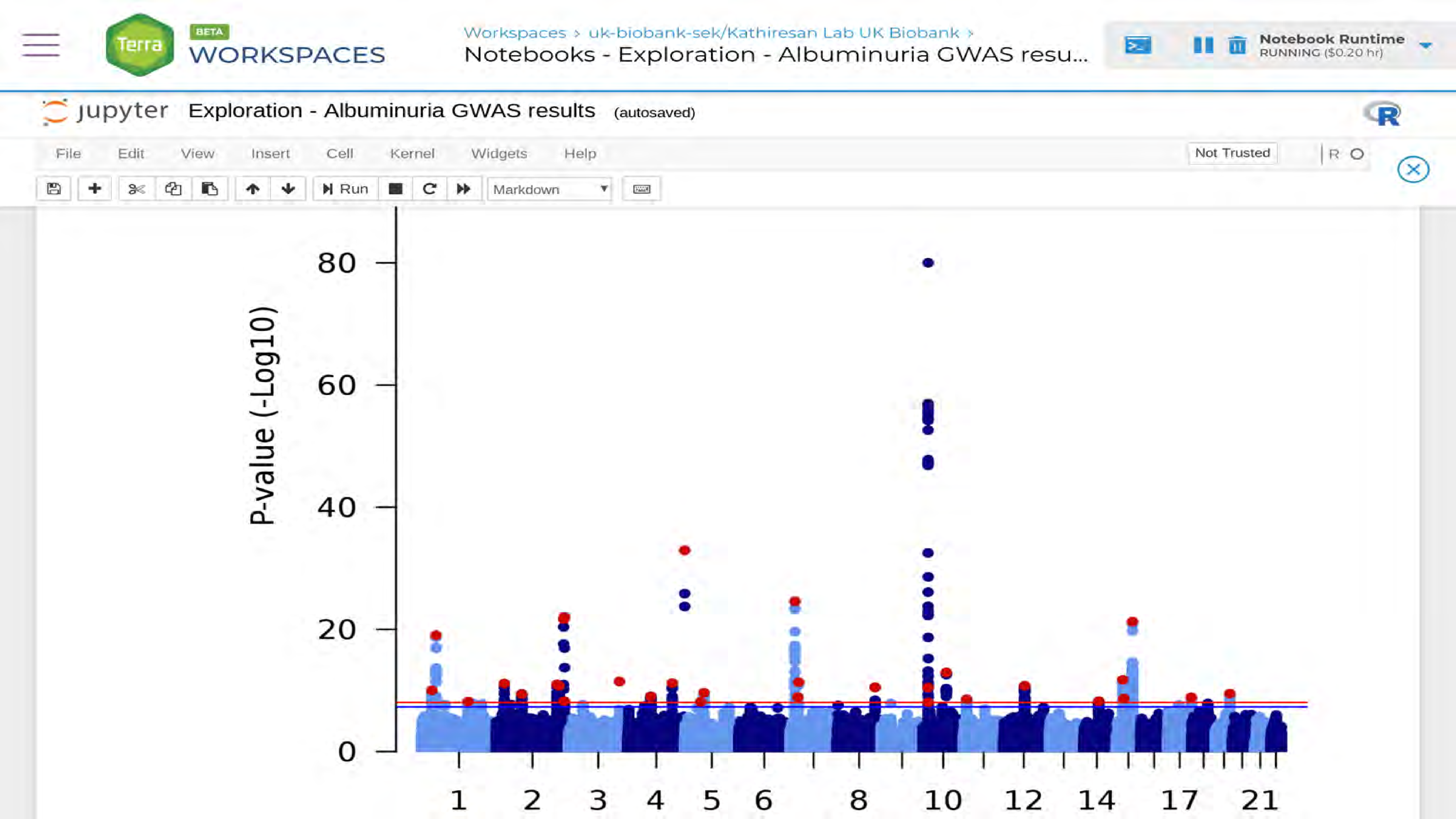
Out[23]:

```
'gs://uk-biobank-sek-data-us-east1/gwas-results/lnACR/20181121_194957/UKB.v2.n382446.lnACRadjS.20181121.tsv'
```

In [24]:

```
vds_linassoc = (vds
    .linreg3(['sa.pheno.lnACRadjS'],
             covariates=covar_cols,
             use_dosages=True,
             variant_block_size=10)
    .drop_samples()
    )

print('Starting linear regression:')
vds_linassoc.export_variants(
    result_destination,
    'Variant = v, bgen.snpid = va.varid, rsid = va.rsid, INFOscore_QCTOOL = va.INFOscore_QCTOOL, '
    + 'vCallRate = va.qc.callRate, pHWE = va.qc.pHWE, Cohort_AC = va.qc.AC, Cohort_AF = va.qc.AF, '
```



Workspaces › uk-biobank-sek/Kathiresan Lab UK Biobank ›
Notebooks - Exploration - Albuminuria GWAS resu...

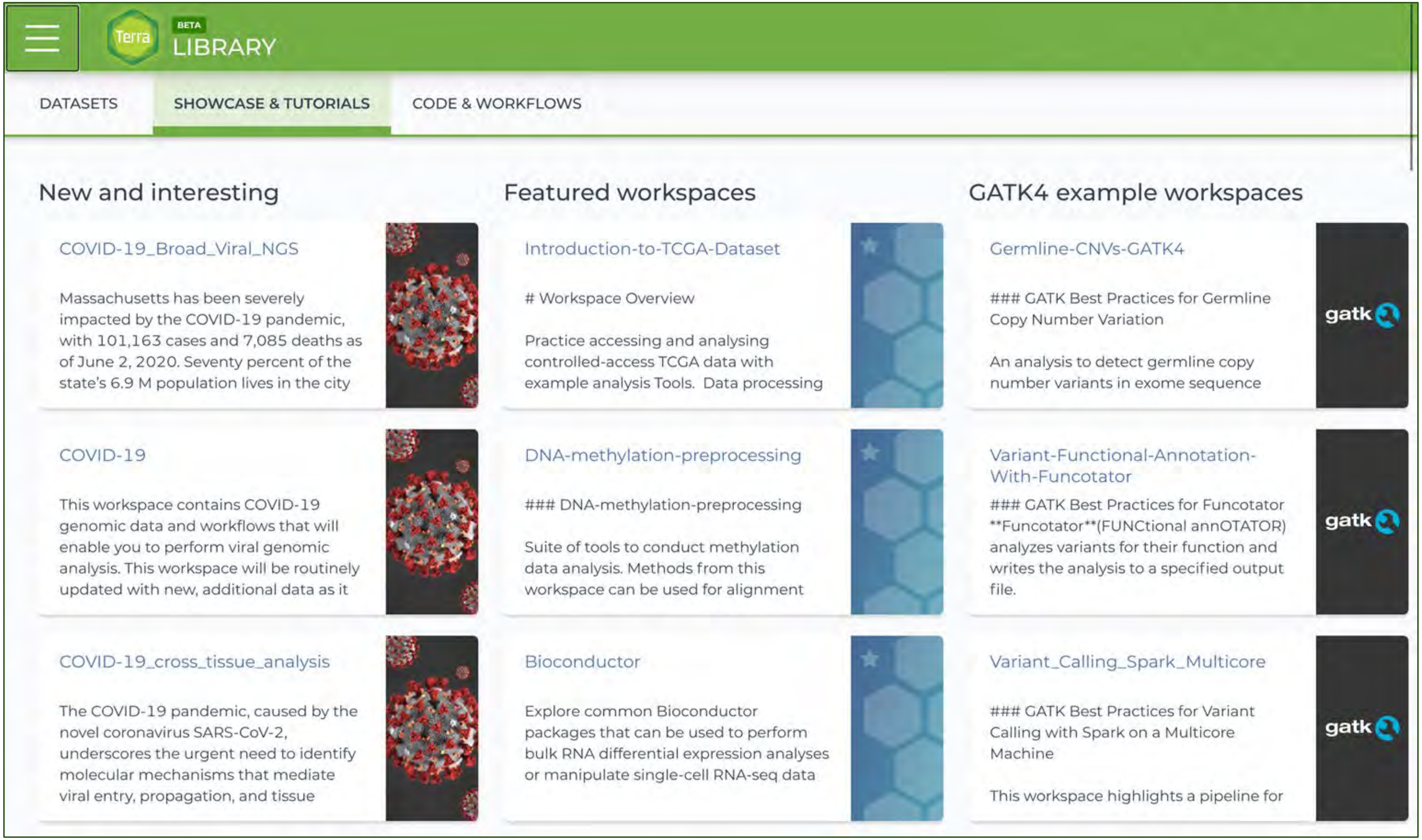
Exploration - Albuminuria GWAS results (autosaved)

Terra use case:
COVID-19 Research

Data & Tools for COVID-19/SARS CoV2 analysis

See this article for a summary of available resources.

Terra showcase



New and interesting

COVID-19_Broad_Viral_NGS

Massachusetts has been severely impacted by the COVID-19 pandemic, with 101,163 cases and 7,085 deaths as of June 2, 2020. Seventy percent of the state's 6.9 M population lives in the city

COVID-19

This workspace contains COVID-19 genomic data and workflows that will enable you to perform viral genomic analysis. This workspace will be routinely updated with new, additional data as it

COVID-19_cross_tissue_analysis

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, underscores the urgent need to identify molecular mechanisms that mediate viral entry, propagation, and tissue

Featured workspaces

Introduction-to-TCGA-Dataset

Workspace Overview

Practice accessing and analysing controlled-access TCGA data with example analysis Tools. Data processing

DNA-methylation-preprocessing

DNA-methylation-preprocessing

Suite of tools to conduct methylation data analysis. Methods from this workspace can be used for alignment

Bioconductor

Explore common Bioconductor packages that can be used to perform bulk RNA differential expression analyses or manipulate single-cell RNA-seq data

GATK4 example workspaces

Germline-CNVs-GATK4

GATK Best Practices for Germline Copy Number Variation

An analysis to detect germline copy number variants in exome sequence

Variant-Functional-Annotation-With-Funcotator

GATK Best Practices for Funcotator **Funcotator**(FUNCtional annOTATOR) analyzes variants for their function and writes the analysis to a specified output file.

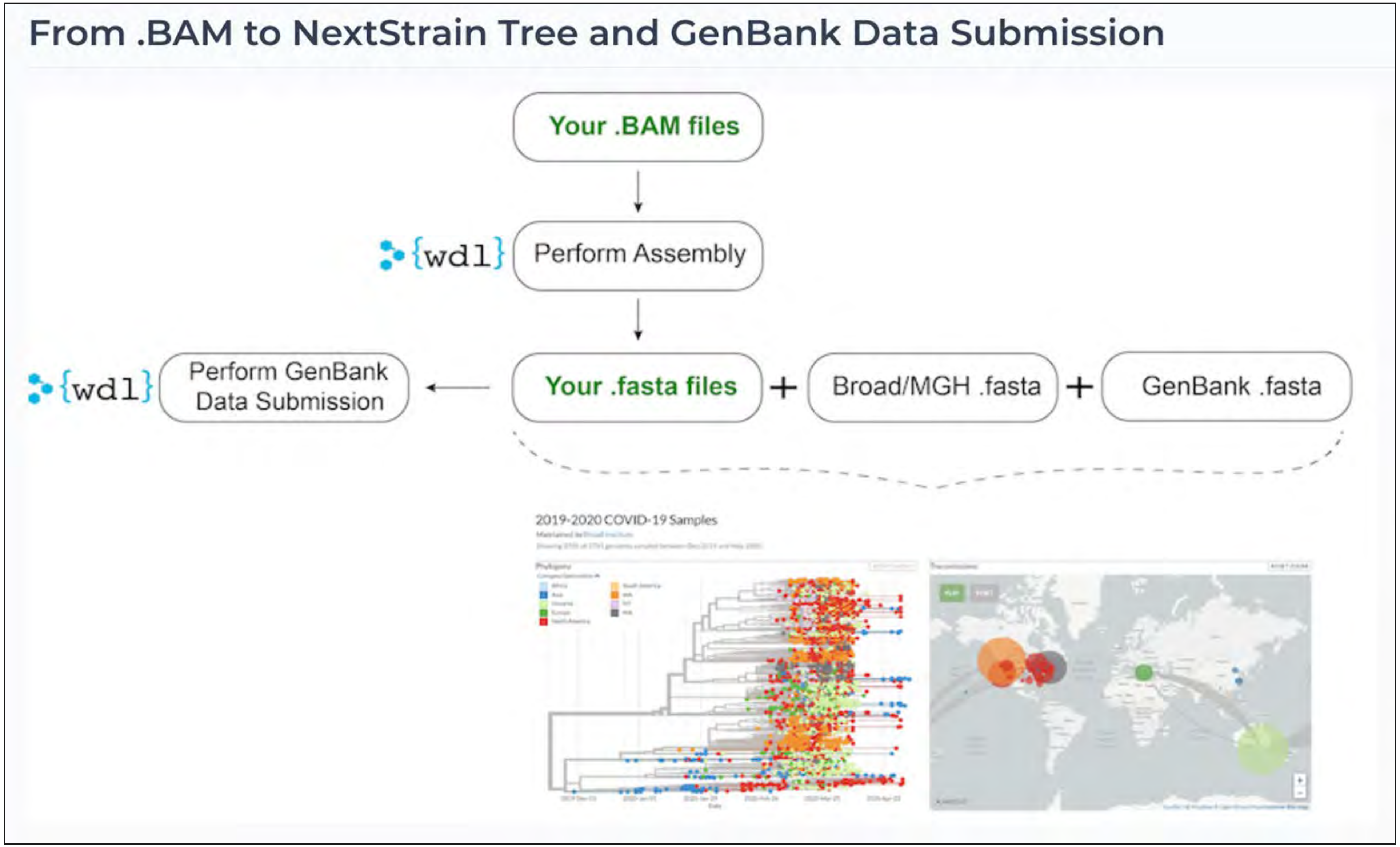
Variant_Calling_Spark_Multicore

GATK Best Practices for Variant Calling with Spark on a Multicore Machine

This workspace highlights a pipeline for

terra.bio

COVID-19 Viral Genomics



From .BAM to NextStrain Tree and GenBank Data Submission

Your .BAM files

{wdl} Perform Assembly

Your .fasta files + Broad/MGH .fasta + GenBank .fasta

{wdl} Perform GenBank Data Submission

COVID-19 Cross Tissue Analysis

Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells

Muus et al., bioRxiv 2020.04.19.049254; doi: https://doi.org/10.1101/2020.04.19.049254

METHODS: Code availability

Data and an interactive analysis examining the co-expression of genes across datasets can be accessed via the open-source data platform, Terra at https://app.terra.bio/#workspaces/kco-incubator/COVID-19_cross_tissue_analysis.

ABOUT THE WORKSPACE

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, underscores the urgent need to identify molecular mechanisms that mediate viral entry, propagation, and tissue pathology in distinct cell types across organs. The surface receptor angiotensin-converting enzyme 2 (ACE2) and the associated proteases, transmembrane protease serine 2 (TMPRSS2) and Cathepsin L (CTSL), were previously identified mediators of SARS-CoV cellular entry.

Experimental Overview

We use single-cell RNA-seq (scRNA-seq) across diverse tissues to assess the cell-type-specific expression of ACE2, TMPRSS2, and CTSL. We identify specific subsets of respiratory epithelial cells as putative targets of viral infection, including subsets of epithelial cells in the nasal passages, lung and airways. Additionally, we detect expression in other tissues that may serve as routes of viral transmission, including the gut and corneal epithelia, and in cells potentially associated with COVID-19 clinical pathology including cardiomyocytes, olfactory sustentacular cells, and renal epithelial cells.

For more details about the resources presented in this Terra Workspace, please use this link to refer to the official manuscript hosted on Biorxiv.

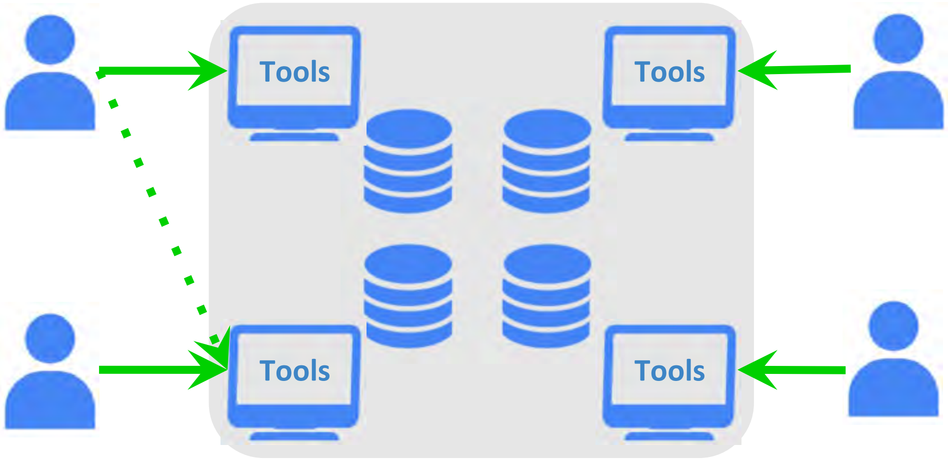
Platforms ⇒ Rapid Response

How do we prepare for the unexpected?

- Collaborative platforms allow fast sharing of new tools, knowledge, and data.
    - Example: COVID-19 viral genomics and tissue expression workspaces
    - Example: *All of Us* COPE surveys
- General-purpose platforms allow fast special-purpose analysis.
    - Example: UK Biobank surveyed their participants about COVID-19, and provided some of the first strong evidence about genetic factors in disease susceptibility and severity.
    - Example: *All of Us* is running serology assays on biosamples collected in early 2020, to look for early COVID-19 exposure patterns.

Cloud-centric approach

Bring researchers to data



Facilitates collaboration

Centralized security controls

Accessible to all researchers

Decreased cost of storage

Shared tool ecosystem

Thank you!

Novel Partnerships and Research Tools to Support Covid-19

Chris Treml

American College
Radiology

Who We Are

- Member based organization
  - ~40,000 members
- Center for Research and Innovation
  - Over 140 staff
  - 2 million+ images process annually

Operating Partnership Types

SINGLE PARTY

PRIMARY /SECONDARY

FEDERATED

Single Party

SIMPLE

WIDELY USED

CONTROL

Primary/Secondary

“SUB-CONTRACTOR”

COMPLEMENTARY EXPERTISE

MID COMPLEXITY

Federated

MULTI-ENTITY

SCALABLE

HIGH COMPLEXITY

MIDRC

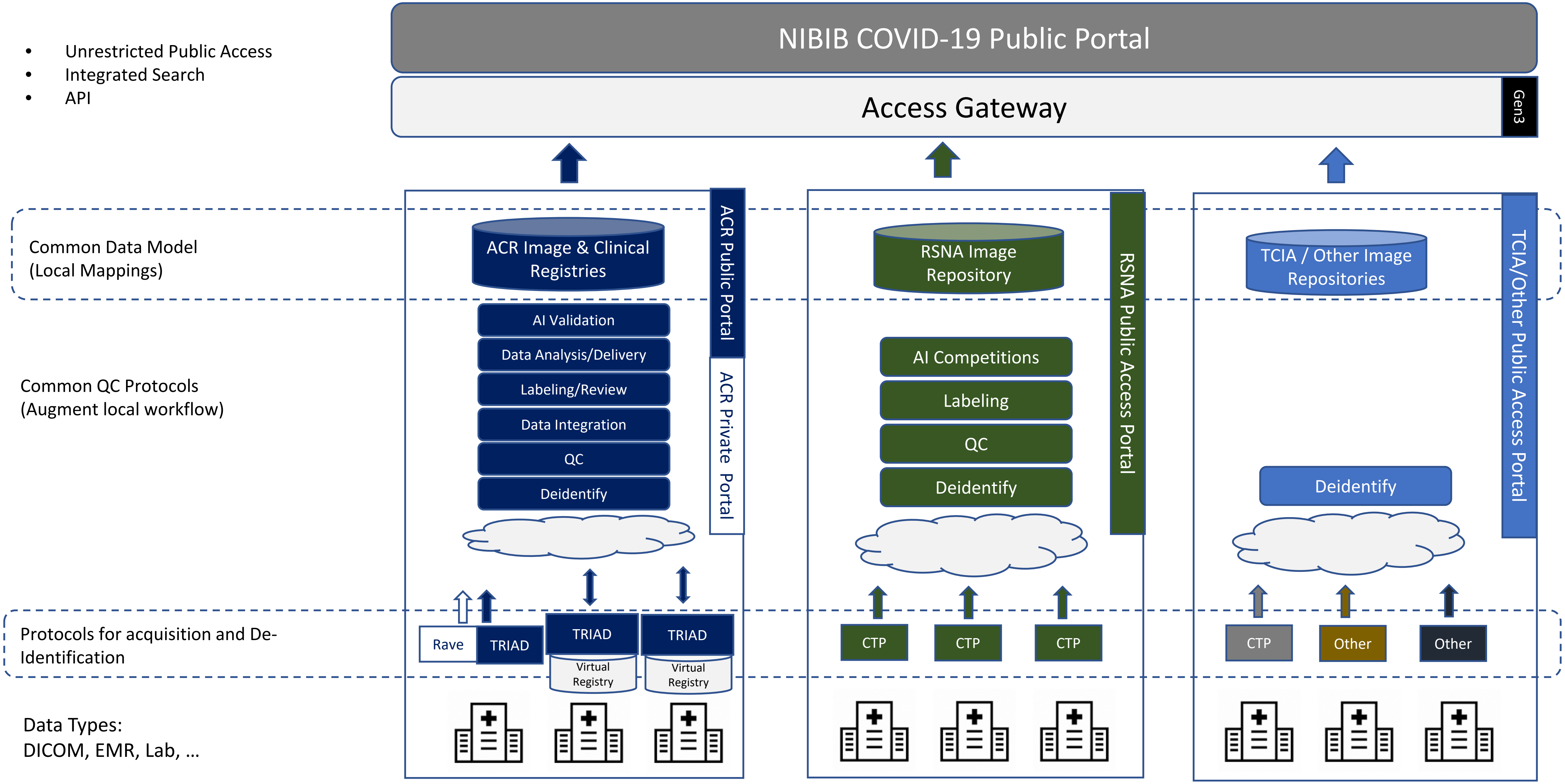
MIDRC

- Virtual registry
- >40 data elements collected
- Groups
  - NIBIB
  - Gen3
  - ACR
  - RSNA
  - AAPM
  - TCIA

MIDRC

- Independence on collection
- Top level abstraction layer
- Governance
  - Unifying data elements
  - Unifying subject IDs
  - Site enrollment
  - Timelines
- NIBIB

- Unrestricted Public Access
- Integrated Search
- API



NIBIB COVID-19 Public Portal

Access Gateway

Gen3

Common Data Model
(Local Mappings)

Common QC Protocols
(Augment local workflow)

Protocols for acquisition and De-Identification

Data Types:
DICOM, EMR, Lab, …

ACR Image & Clinical Registries

AI Validation

Data Analysis/Delivery

Labeling/Review

Data Integration

QC

Deidentify

ACR Public Portal

ACR Private Portal

Rave

TRIAD

TRIAD

Virtual Registry

TRIAD

Virtual Registry

RSNA Image Repository

AI Competitions

Labeling

QC

Deidentify

RSNA Public Access Portal

CTP

CTP

CTP

TCIA / Other Image Repositories

Deidentify

TCIA/Other Public Access Portal

CTP

Other

Other

- Integrated Registry (images and clinical data)
- Real-time Upload & Federated Queries
- Real-world Testing of AI

- Image Repository and limited clinical data
- Retrospective Batch Upload
- AI Competitions

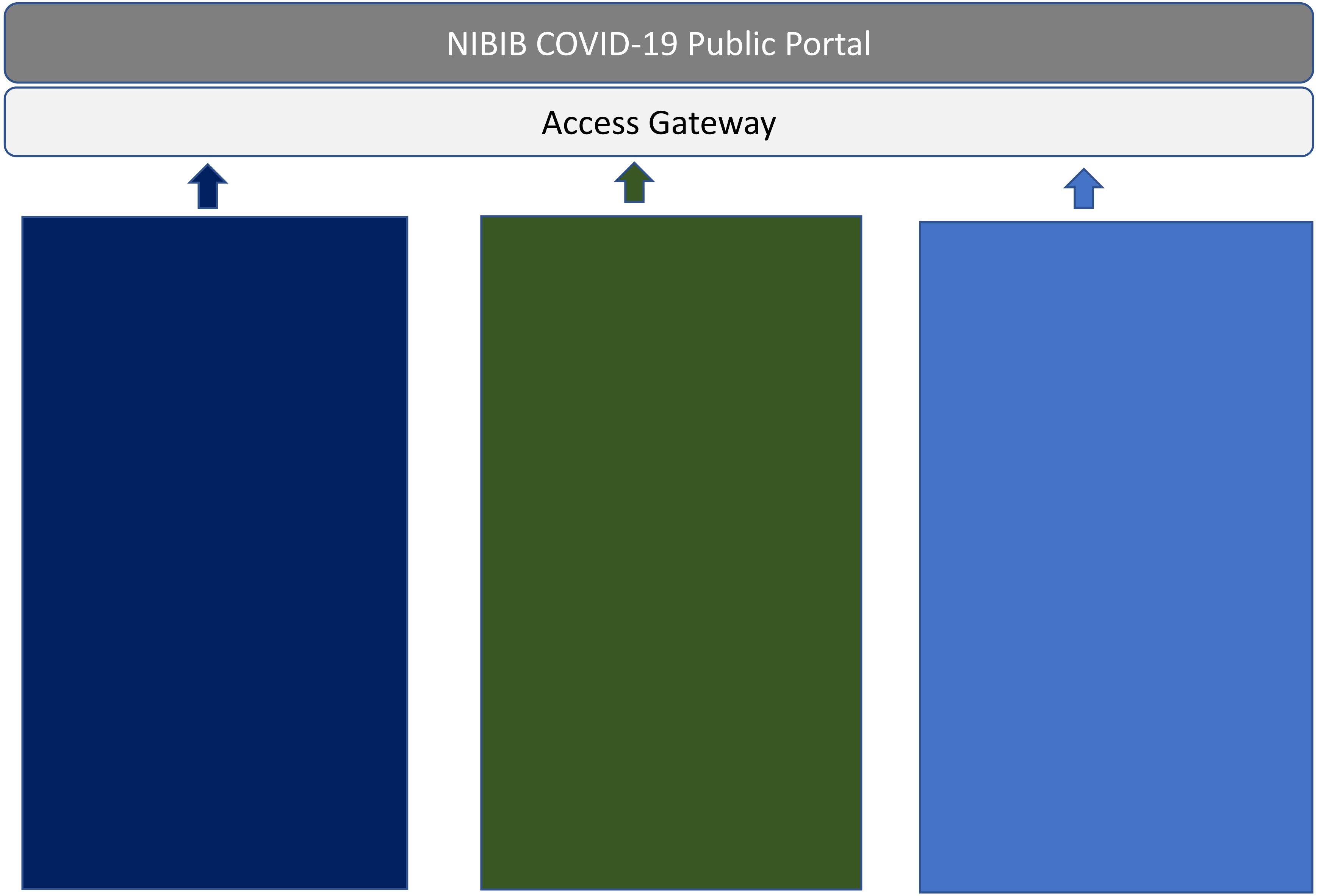
- Image Repository for NCI-funded trial activity

MIDRC

- >40 data elements collected
- Demographics
- Vitals
- Labs
- Images
- Diagnosis
- Repeatable
- Single Instance

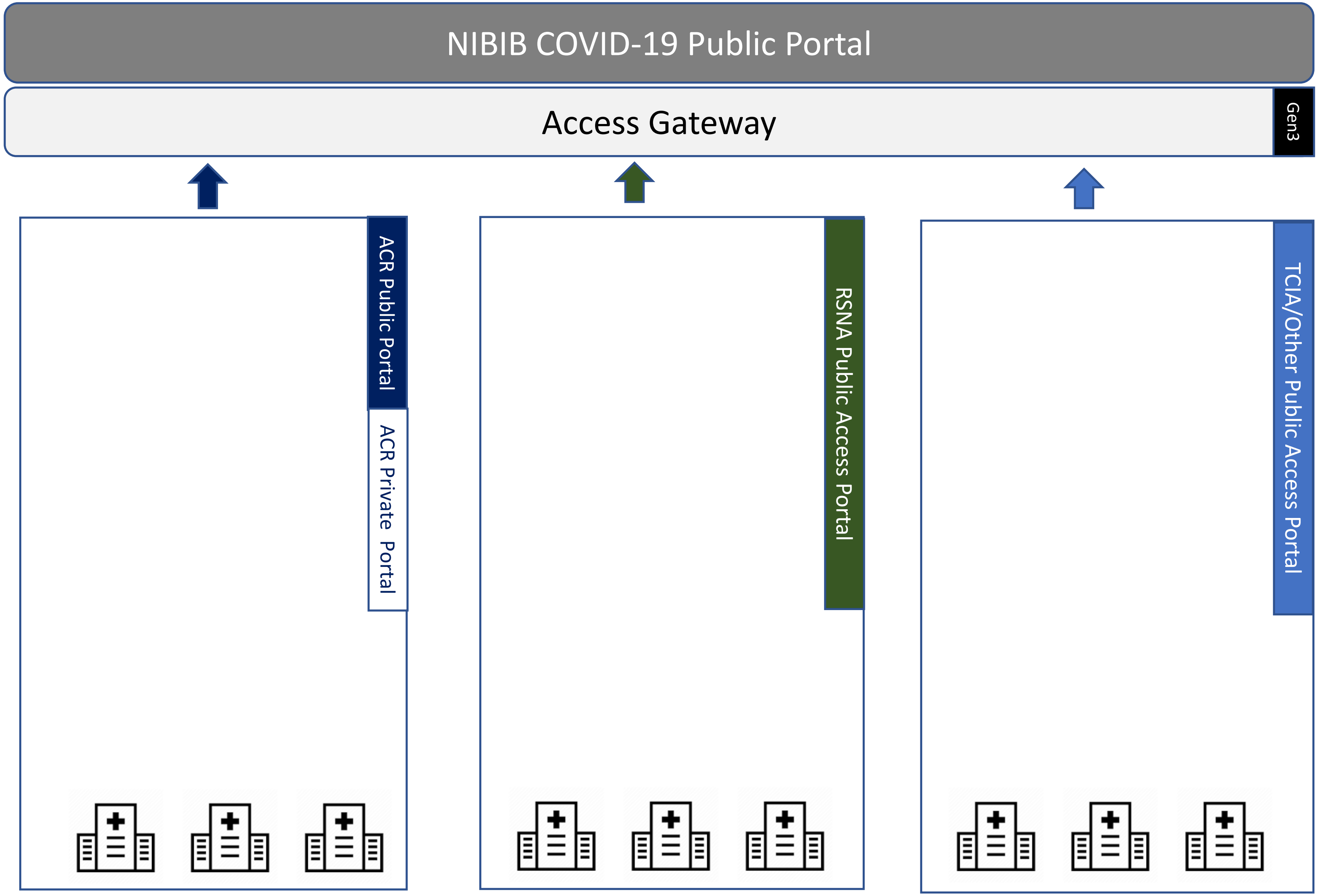
C

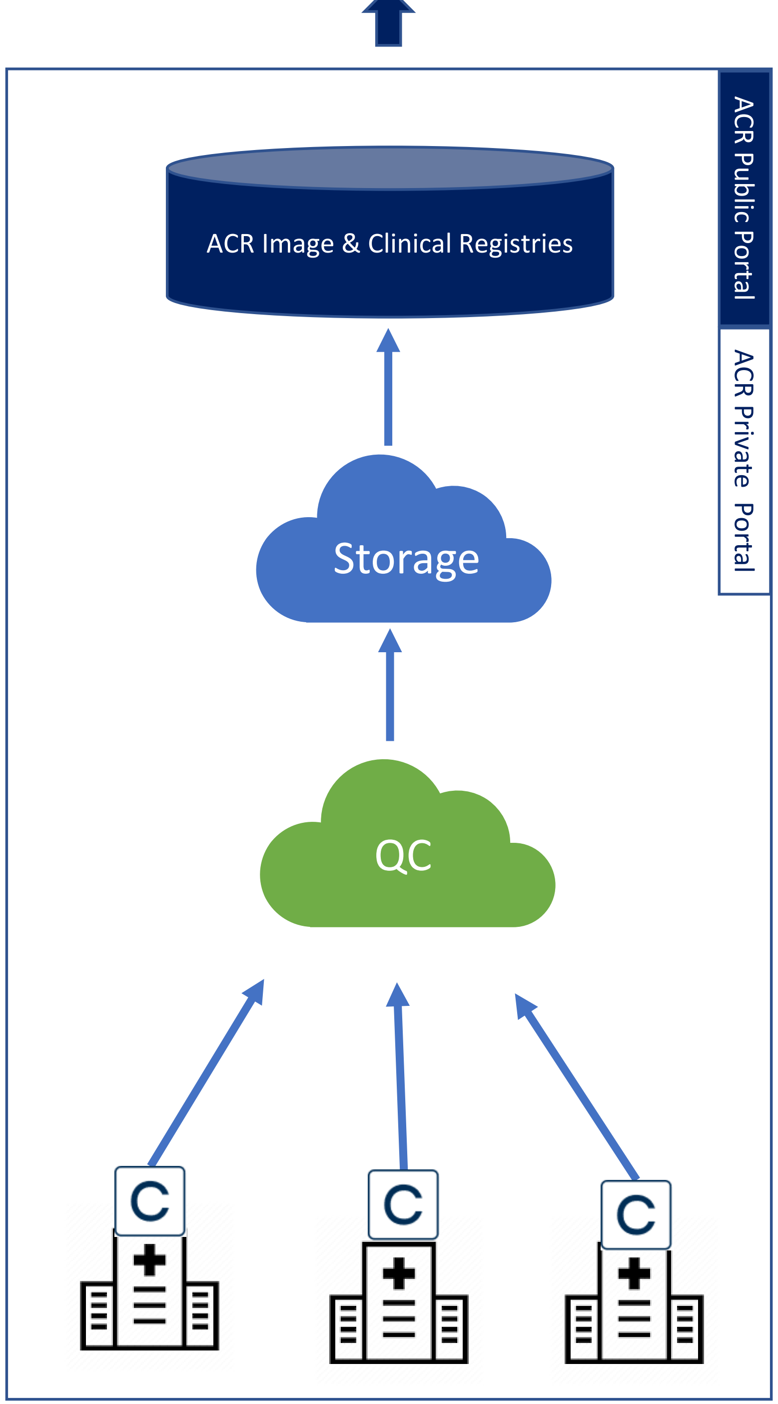
CONNECT



NIBIB COVID-19 Public Portal

Access Gateway


NIBIB COVID-19 Public Portal
Access Gateway
Gen3
ACR Public Portal
ACR Private Portal
RSNA Public Access Portal
TCIA/Other Public Access Portal



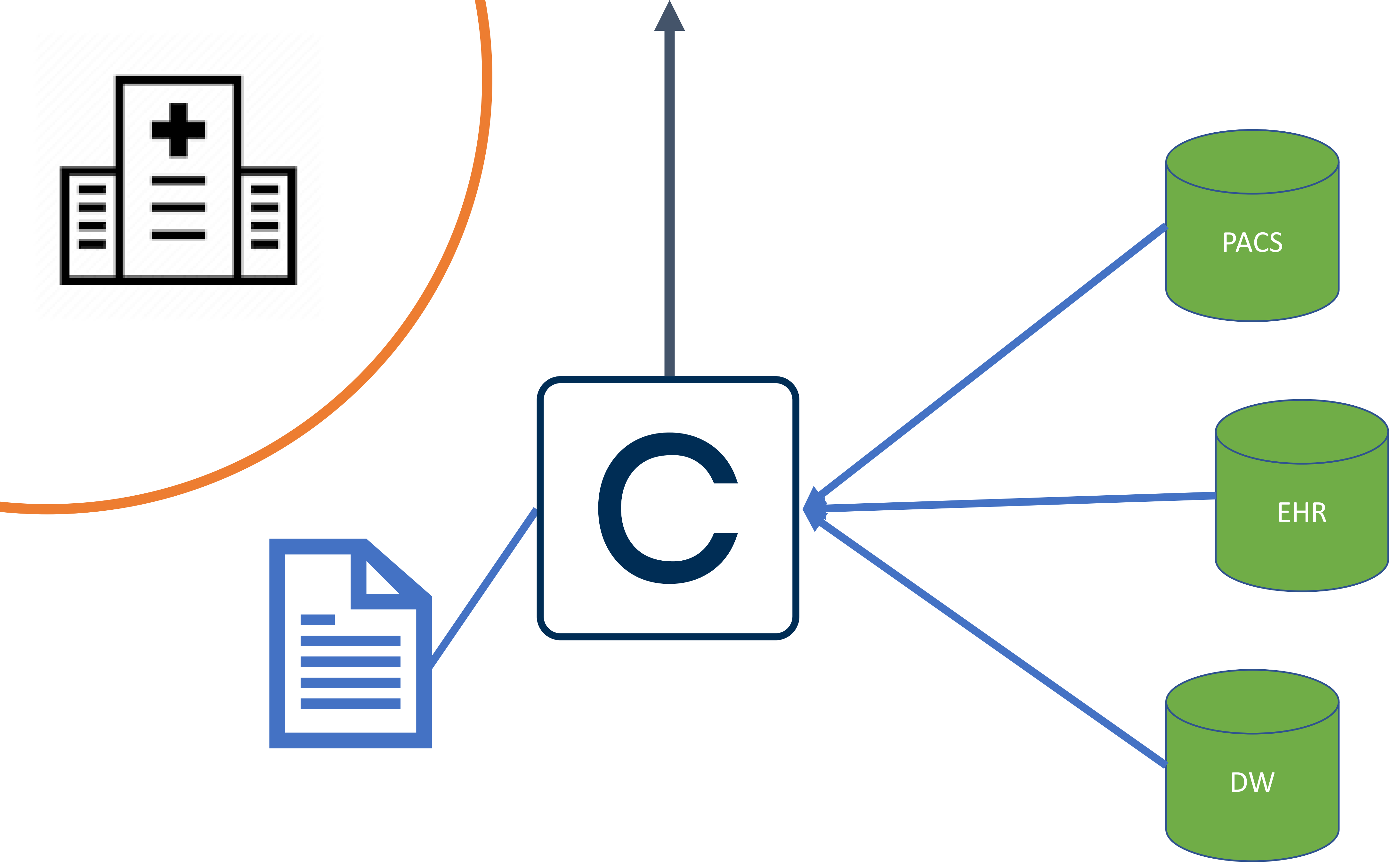
ACR Image & Clinical Registries

Storage

QC

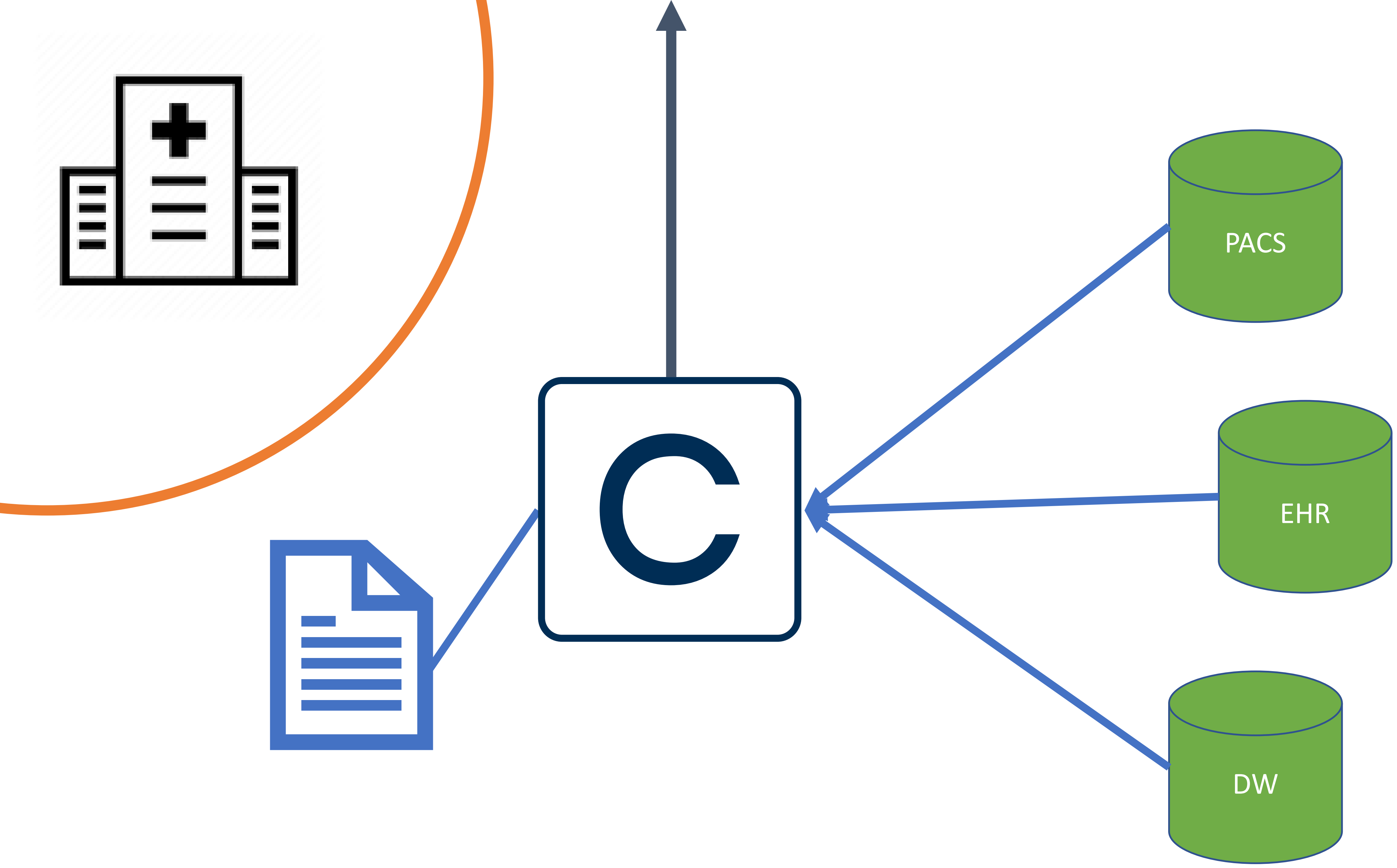
ACR Public Portal

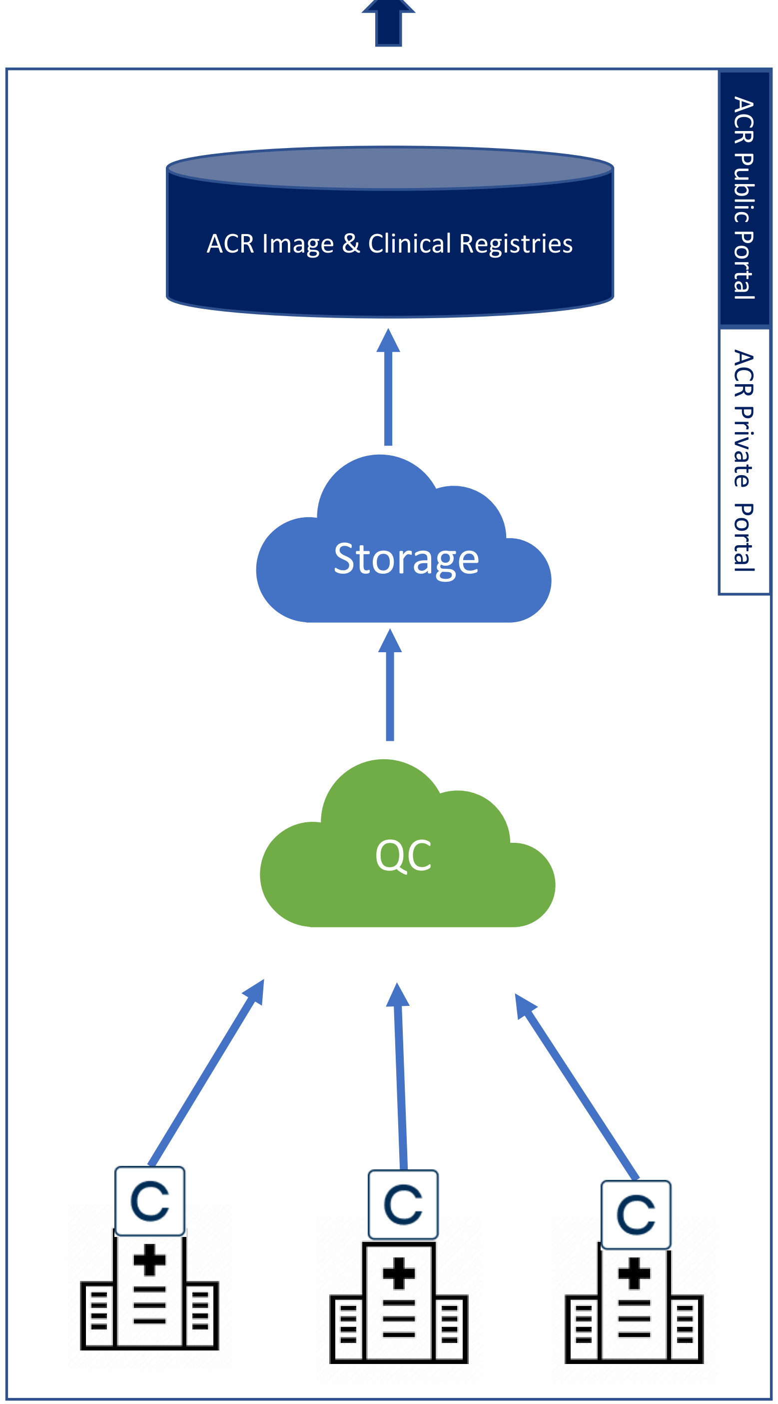
ACR Private Portal

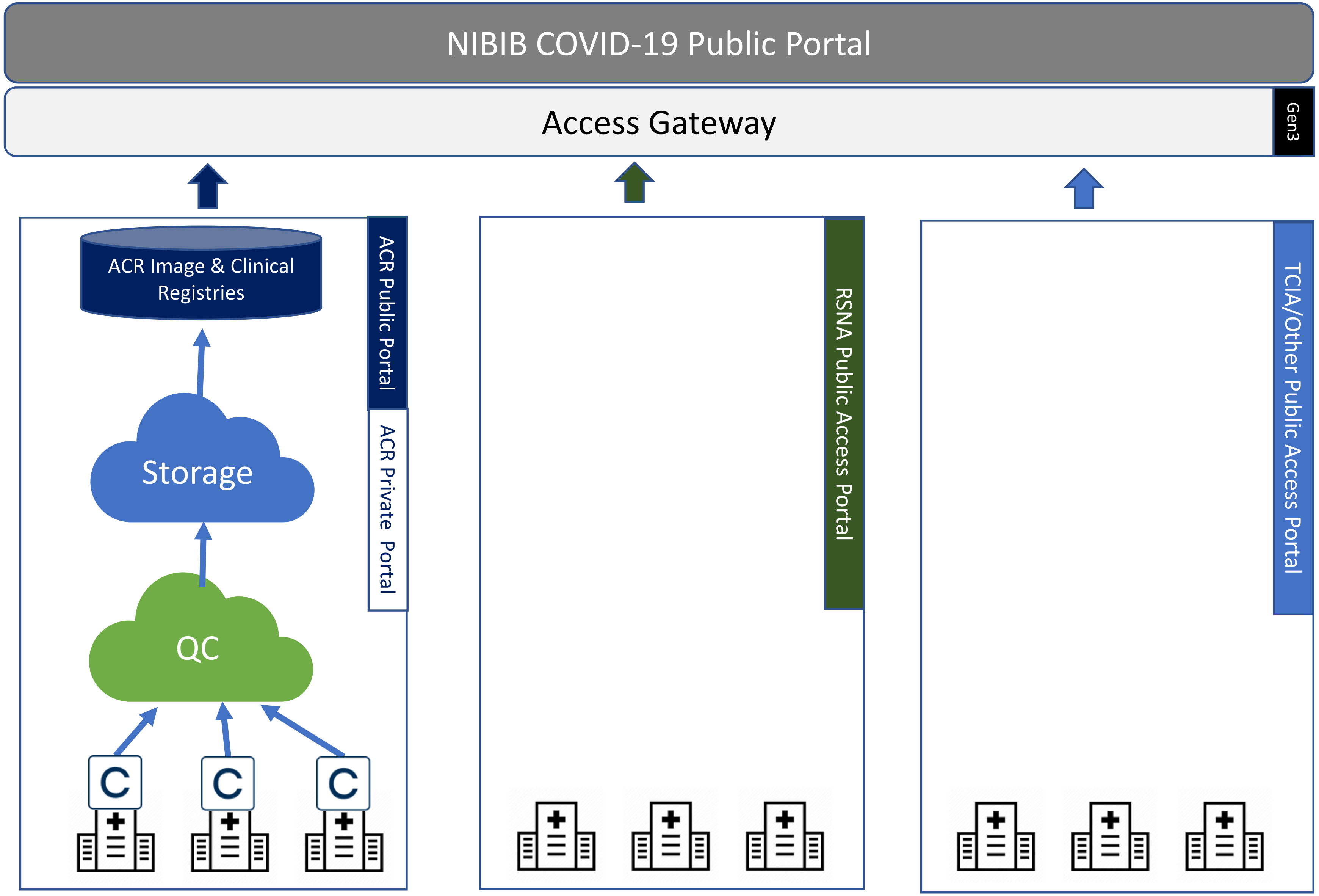

C
PACS
EHR
DW

C

- Anonymization
- QC
- Mapping
- Automatic abstraction
- Local control


C
PACS
EHR
DW


ACR Image & Clinical Registries
Storage
QC
C
C
C
ACR Public Portal
ACR Private Portal

NIBIB COVID-19 Public Portal
Access Gateway
Gen3
ACR Image & Clinical Registries
ACR Public Portal
Storage
ACR Private Portal
QC
C
C
C
RSNA Public Access Portal
TCIA/Other Public Access Portal

Wrapping up

- Different partnership models, different benefits
  - Single, Primary/Secondary, Federated
- COVID-19 presenting areas for high scalability
  - MIDRC example
- ACR Connect
  - Last mile platform for data gathering

All of Us Researcher Workbench

CMSS Webinar: Deploying Cloud-based Platforms and Analytic Tools to Support Covid-19 and Beyond

August 6, 2020

Andrea H Ramirez, MD, MS
Vanderbilt University Medical Center
andrea.h.ramirez@vumc.org

@AndreaRamirezMD

All of Us
DATA & RESEARCH CENTER

What is the NIH *All of Us* Research Program?

All *of* Us RESEARCH PROGRAM | The Future of Health Begins With You

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**, researchers will uncover paths toward delivering **precision medicine – or individualized prevention, treatment, and care – for all of us**.

"*All of Us* is among the most ambitious research efforts that our nation has undertaken!"

NIH Director Francis Collins, M.D., Ph.D.

The *All of Us* Research Program is part of the broader Precision Medicine Initiative.

On behalf of the Data and Research Center team!

VANDERBILT UNIVERSITY MEDICAL CENTER

verily

BROAD INSTITUTE

COLUMBIA UNIVERSITY MEDICAL CENTER

NIH National Institutes of Health
All of Us Research Program

All of Us
DATA & RESEARCH CENTER

The *All of Us* Research Program

Nurture relationships
with **one million or more** participant partners, from all walks of life, for decades

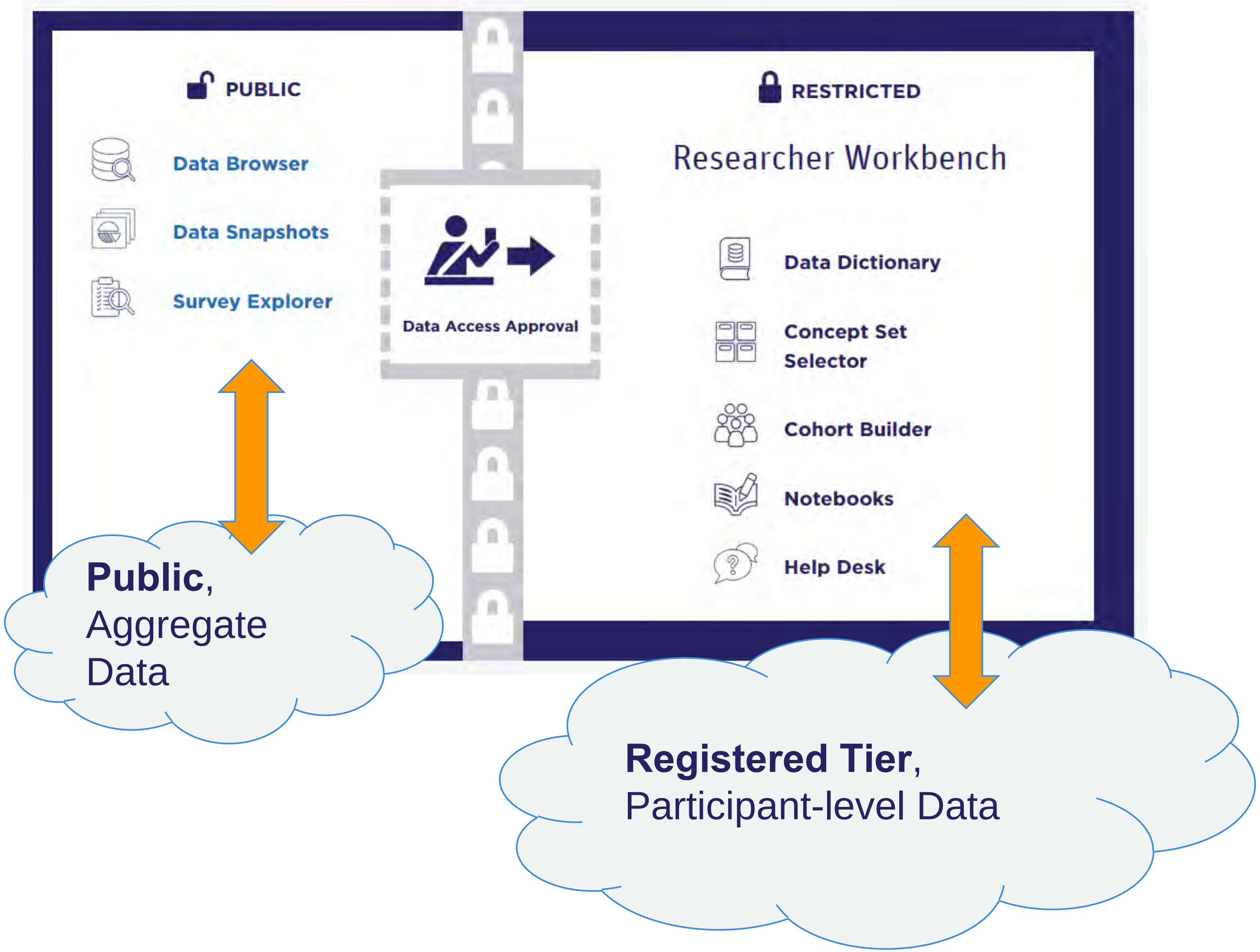
Our mission
To accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care for all of us

Deliver the largest, richest biomedical dataset
that is easy, safe, and free to access

Catalyze a robust ecosystem
of researchers and funders hungry to use and support it

All of Us
DATA & RESEARCH CENTER

The Research Hub: Access & Analyze *All of Us* Data



PUBLIC

- Data Browser
- Data Snapshots
- Survey Explorer

Data Access Approval

RESTRICTED

Researcher Workbench

- Data Dictionary
- Concept Set Selector
- Cohort Builder
- Notebooks
- Help Desk

Public, Aggregate Data

Registered Tier, Participant-level Data

Research Hub: Public Tools for Browsing Aggregate Data

Search Across Data Types

Keyword Search

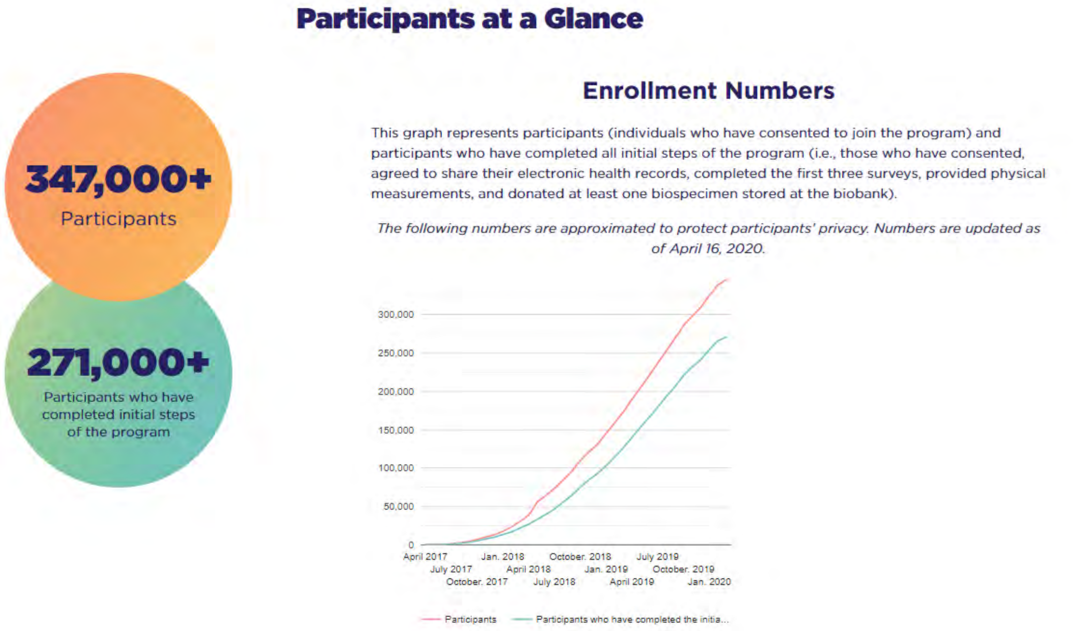
Data based on Curated Data Repository (CDR) dated 2/11/2020 with 225,360 total participants.

EHR Domains:

Conditions	Drug Exposures	Labs & Measurements	Procedures
20,782 medical concepts	20,955 medical concepts	10,050 medical concepts	20,549 medical concepts
113,280 participants in this domain	104,580 participants in this domain	109,180 participants in this domain	102,220 participants in this domain
View Top Conditions	View Top Drug Exposures	View Top Labs & Measurements	View Top Procedures

Data Browser

Data Snapshots



Participants at a Glance

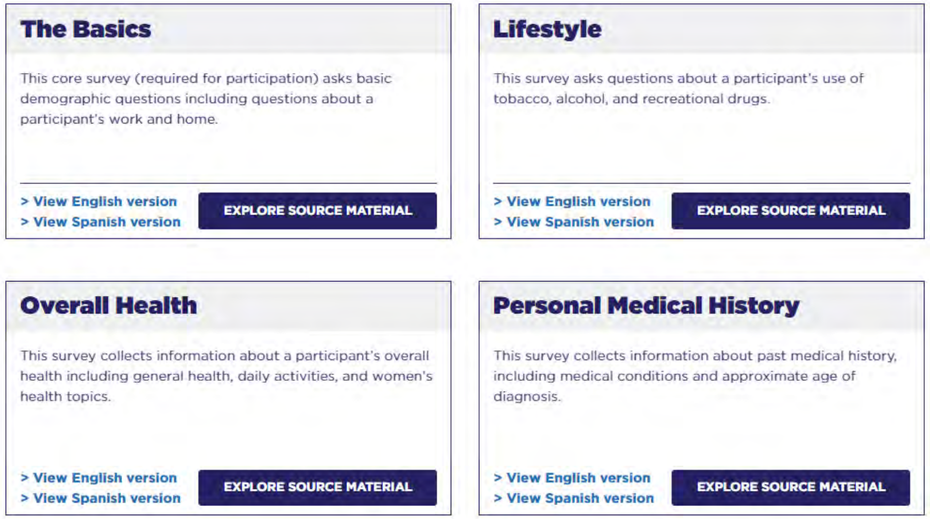
347,000+ Participants

271,000+ Participants who have completed initial steps of the program

Enrollment Numbers

This graph represents participants (individuals who have consented to join the program) and participants who have completed all initial steps of the program (i.e., those who have consented, agreed to share their electronic health records, completed the first three surveys, provided physical measurements, and donated at least one biospecimen stored at the biobank).

The following numbers are approximated to protect participants' privacy. Numbers are updated as of April 16, 2020.



The Basics

This core survey (required for participation) asks basic demographic questions including questions about a participant's work and home.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Lifestyle

This survey asks questions about a participant's use of tobacco, alcohol, and recreational drugs.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Overall Health

This survey collects information about a participant's overall health including general health, daily activities, and women's health topics.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

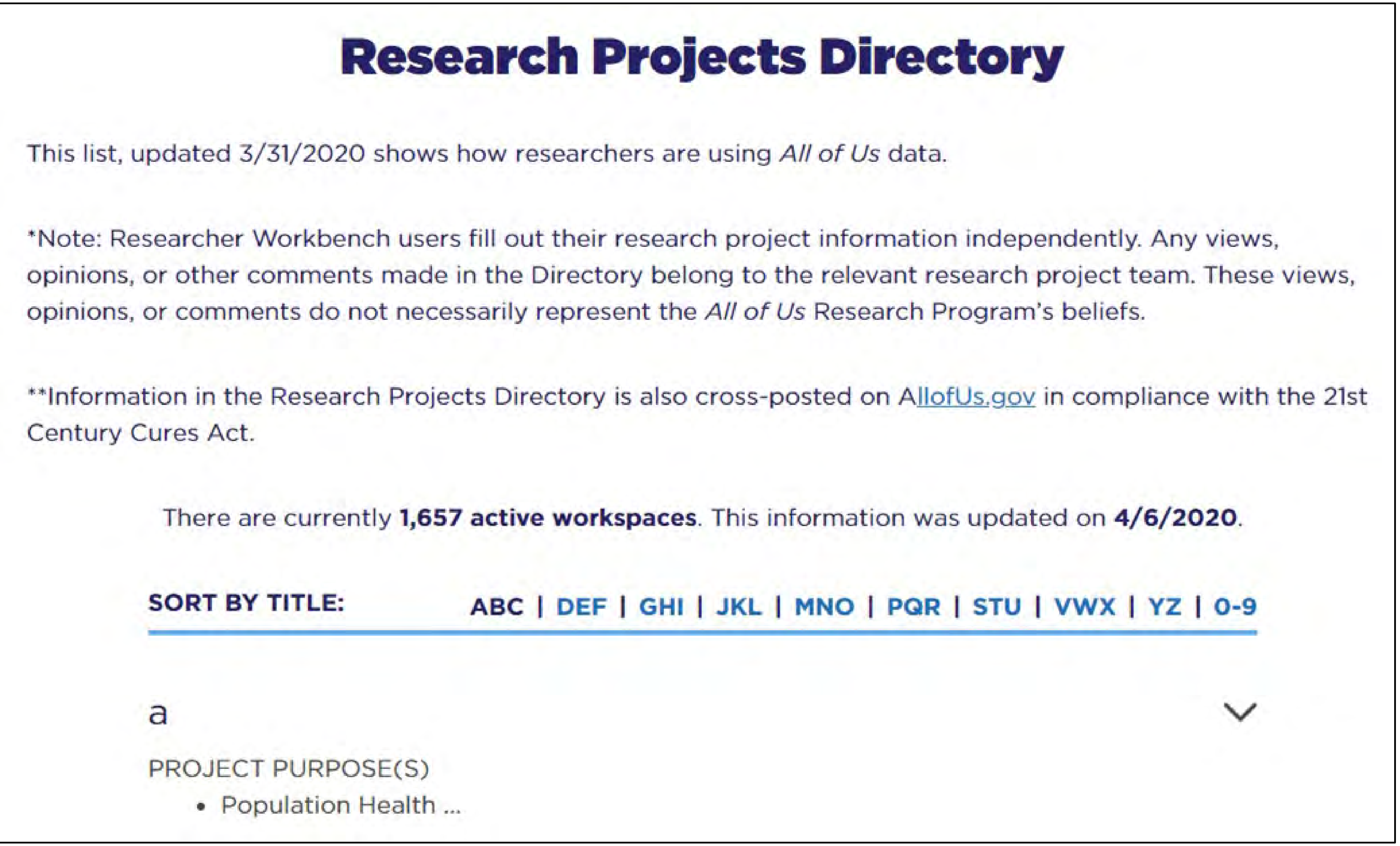
Personal Medical History

This survey collects information about past medical history, including medical conditions and approximate age of diagnosis.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Survey Explorer

Research Projects Directory Mock



Research Projects Directory

This list, updated 3/31/2020 shows how researchers are using *All of Us* data.

*Note: Researcher Workbench users fill out their research project information independently. Any views, opinions, or other comments made in the Directory belong to the relevant research project team. These views, opinions, or comments do not necessarily represent the *All of Us* Research Program's beliefs.

**Information in the Research Projects Directory is also cross-posted on AllofUs.gov in compliance with the 21st Century Cures Act.

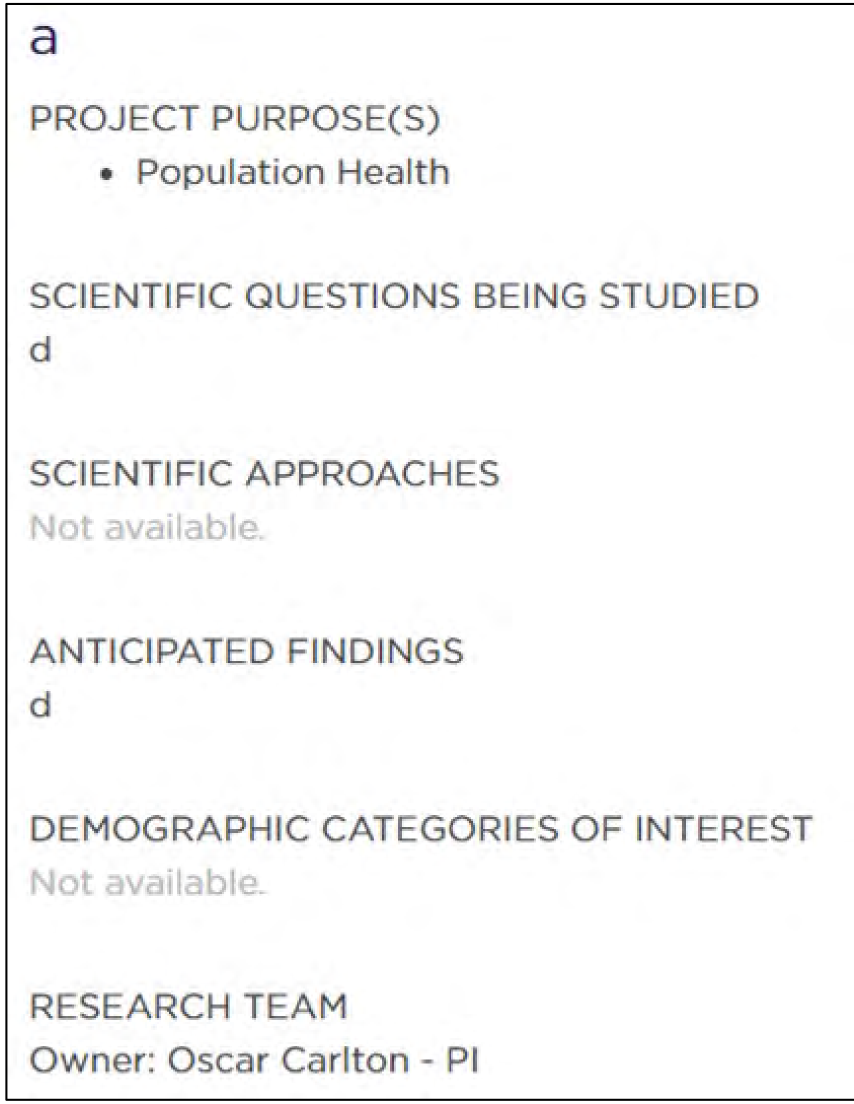
There are currently **1,657 active workspaces**. This information was updated on **4/6/2020**.

a

PROJECT PURPOSE(S)

- Population Health ...

https://www.researchallofus.org/research-projects-directory/?section=stu



a

PROJECT PURPOSE(S)

- Population Health

SCIENTIFIC QUESTIONS BEING STUDIED

d

SCIENTIFIC APPROACHES

Not available.

ANTICIPATED FINDINGS

d

DEMOGRAPHIC CATEGORIES OF INTEREST

Not available.

RESEARCH TEAM

Owner: Oscar Carlton - PI

Research Hub → (Beta) Researcher Workbench

What to Expect as a Beta Researcher

1. Currently, an institutional agreement & eRA Commons account must be in place.
2. Feedback is welcome.
3. The tools are continuing to evolve.
4. The program cohort is actively growing, and so is our data.
5. The goal is to be a true game changer for understanding health.

What data are available now?

Our Data Is Growing. Here are the Current Data Types

Enroll, Consent and Authorize EHR

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

Answering Surveys

- Initial surveys: The Basics, Overall Health, Lifestyle, Health Care Access & Utilization, Family Medical History, Personal Health History
- Additional surveys will be released on an ongoing basis.

Physical Measurements*

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

**Based on diverse sampling and capacity*

Provide Biosamples*

- Blood (or saliva, if blood draw is unsuccessful)
- Urine specimen
- Biosamples will be stored at the program's biobank

**Based on diverse sampling and capacity*

Wearables and Digital Apps

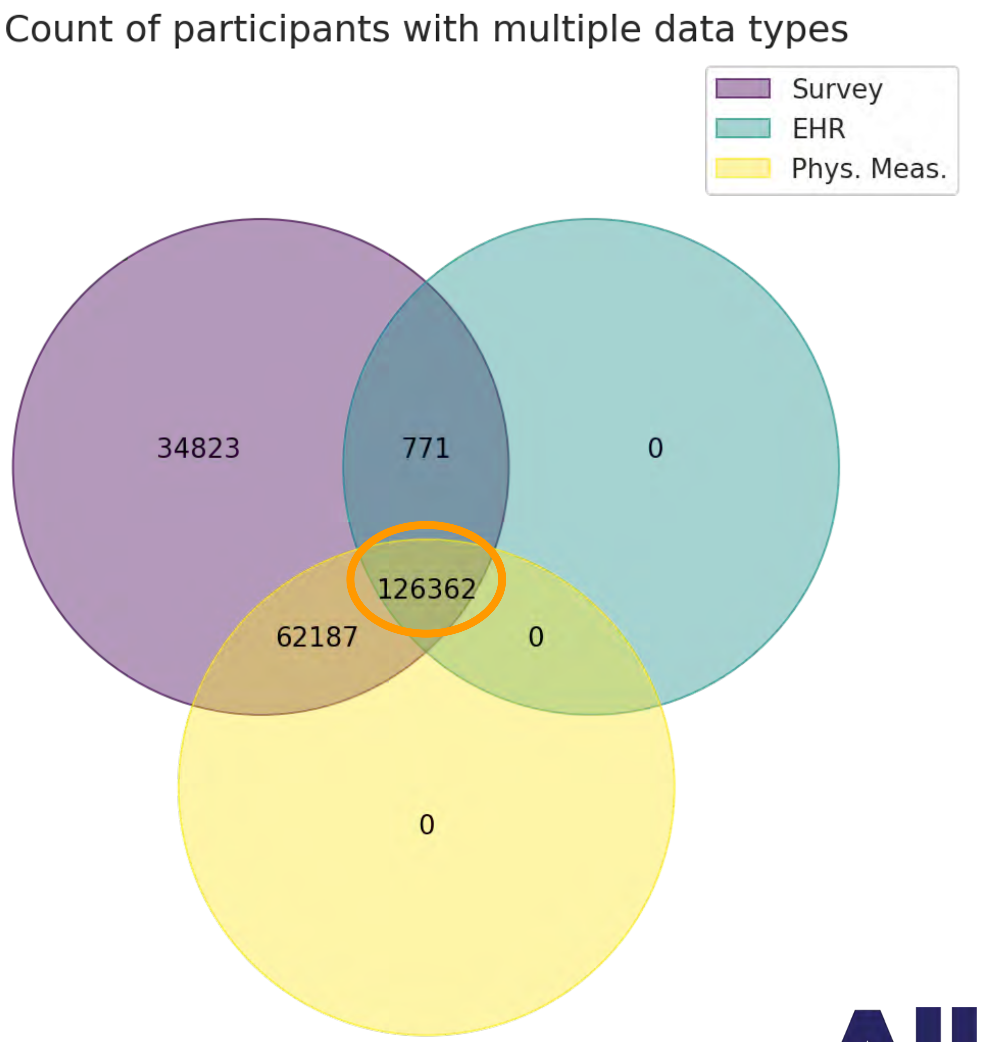
- Share data from wearable fitness devices, starting with Fitbit
- More integrations to come. E.g., integrated apps to track mood & cardio-respiratory fitness

Available in the Current Dataset

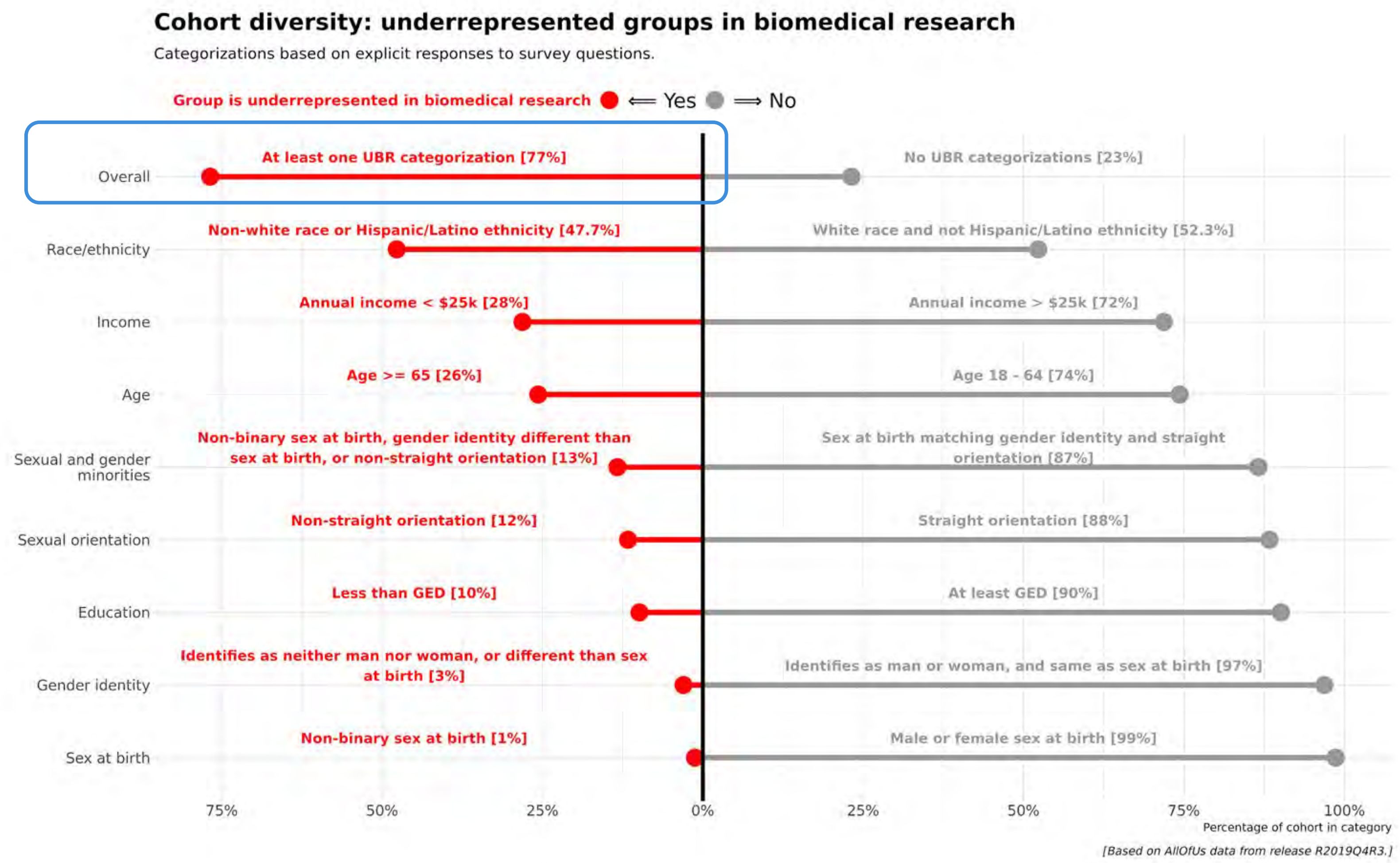
All of Us
DATA & RESEARCH CENTER

Research Data Available Now

Data Type	Participant Count
Survey	>224,000
Physical Measurement	>188,000
Electronic Health Record	>127,000



Data@Beta: Cohort Diversity



Cohort diversity: underrepresented groups in biomedical research

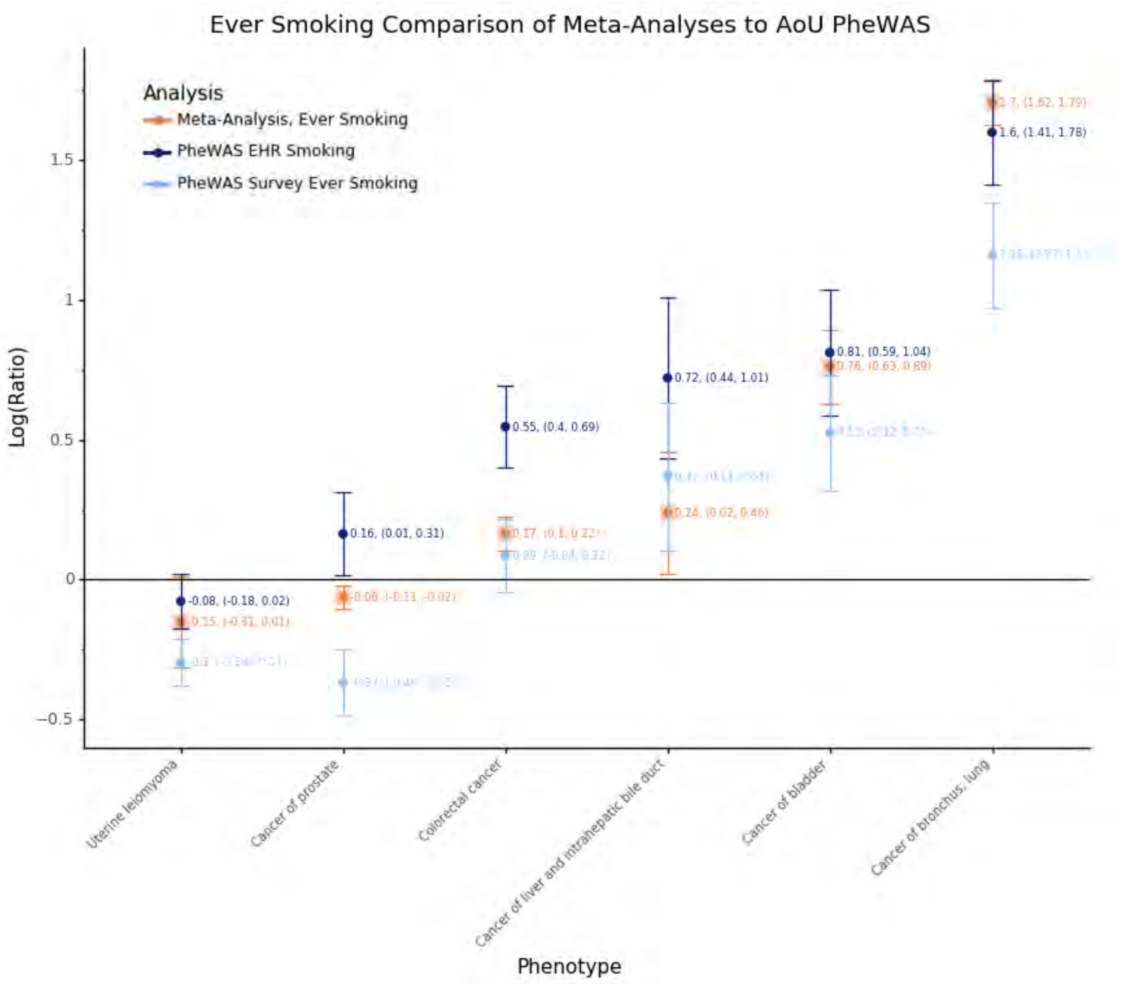
Categorizations based on explicit responses to survey questions.

[Based on AllOfUs data from release R2019Q4R3.]

A. Ramirez, L. Suleiman, D. Schlueter, et al., *The All of Us Research Program: data quality, utility, and diversity*, medRxiv 2020.05.29.20116905; doi:

Data@Beta: Data Quality

Description	EHR Ever Smoking OR (95% CI)	Survey Ever Smoking OR (95% CI)
Top 3 Increased risk effects		
Cancer of the bronchus; lung	4.94 (4.11, 5.95)	3.19 (2.65, 3.84)
Cancer within the respiratory system	4.94 (4.12, 5.92)	3.15 (2.62, 3.78)
Malignant neoplasm of bladder	2.36 (1.87, 2.98)	1.76 (1.42, 2.18)
Top 3 Decreased risk effects		
Vascular hamartomas and non-neoplastic nevi	0.51 (0.42, 0.62)	0.55 (0.48, 0.64)
Nevus, non-neoplastic	0.52 (0.43, 0.64)	0.57 (0.49, 0.66)
Benign neoplasm of skin	0.53 (0.49, 0.58)	0.62 (0.58, 0.66)



A. Ramirez, L. Suleiman, D. Schlueter, et al., *The All of Us Research Program: data quality, utility, and diversity*, medRxiv 2020.05.29.20116905; doi:

Want to learn more?

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

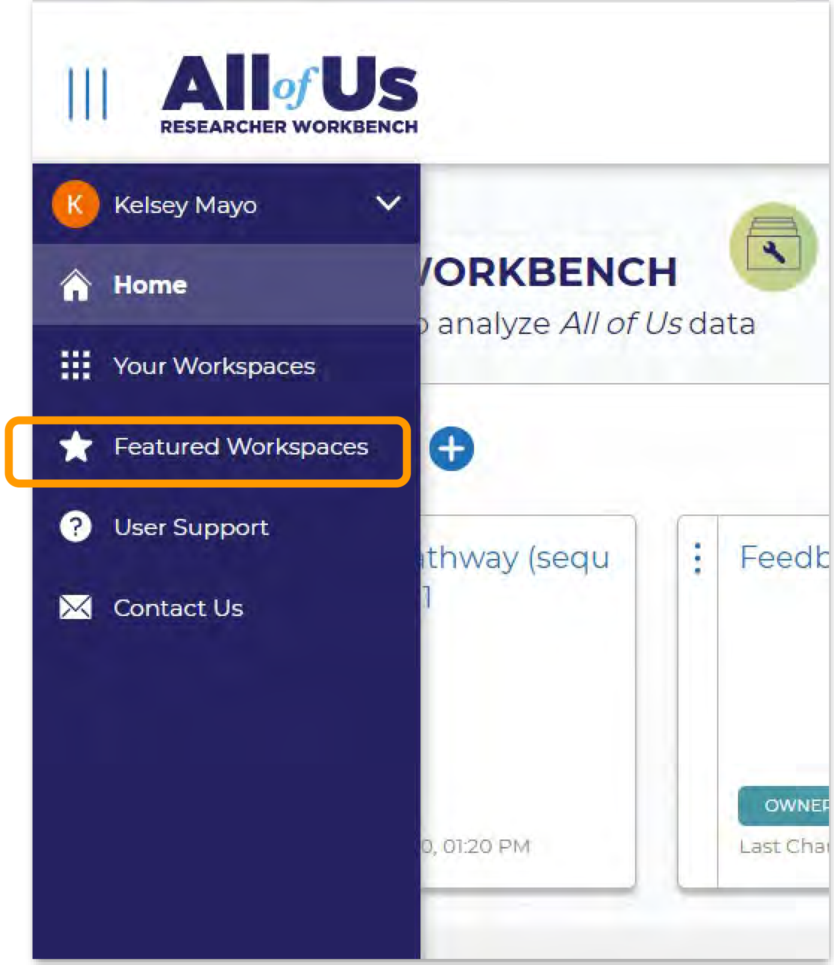
The All of Us Research Program: data quality, utility, and diversity

Andrea H Ramirez, Lina Sulieman, David J Schlueter, Alese Halvorson, Jun Qian, Francis Ratsimbazafy, Roxana Loperena, Kelsey Mayo, Melissa Basford, Nicole Deflaux, Karthik N Muthuraman, Karthik Natarajan, Abel Kho, Hua Xu, Consuelo Wilkins, Hoda Anton-Culver, Eric Boerwinkle, Mine Cicek, Cheryl R Clark, Elizabeth Cohn, Lucila Ohno-Machado, Sheri Schully, Brian K Ahmedani, Maria Argos, Robert M Cronin, Christopher O'Donnell, Mona Fouad, David B Goldstein, Philip Greenland, Scott J Hebbring, Elizabeth W Karlson, Parinda Khatri, Bruce Korf, Jordan W Smoller, Stephen Sodeke, John Wilbanks, Justin Hentges, Christopher Lunt, Stephanie A Devaney, Kelly Gebo, Joshua C Denny, Robert J Carroll, David Glazer, Paul A Harris, George Hripcsak, Anthony Philippakis, Dan M Roden, On behalf of the All of Us Research Program

doi: https://doi.org/10.1101/2020.05.29.20116905

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20116905.this version posted June 3, 2020. The copyright holder for this preprint

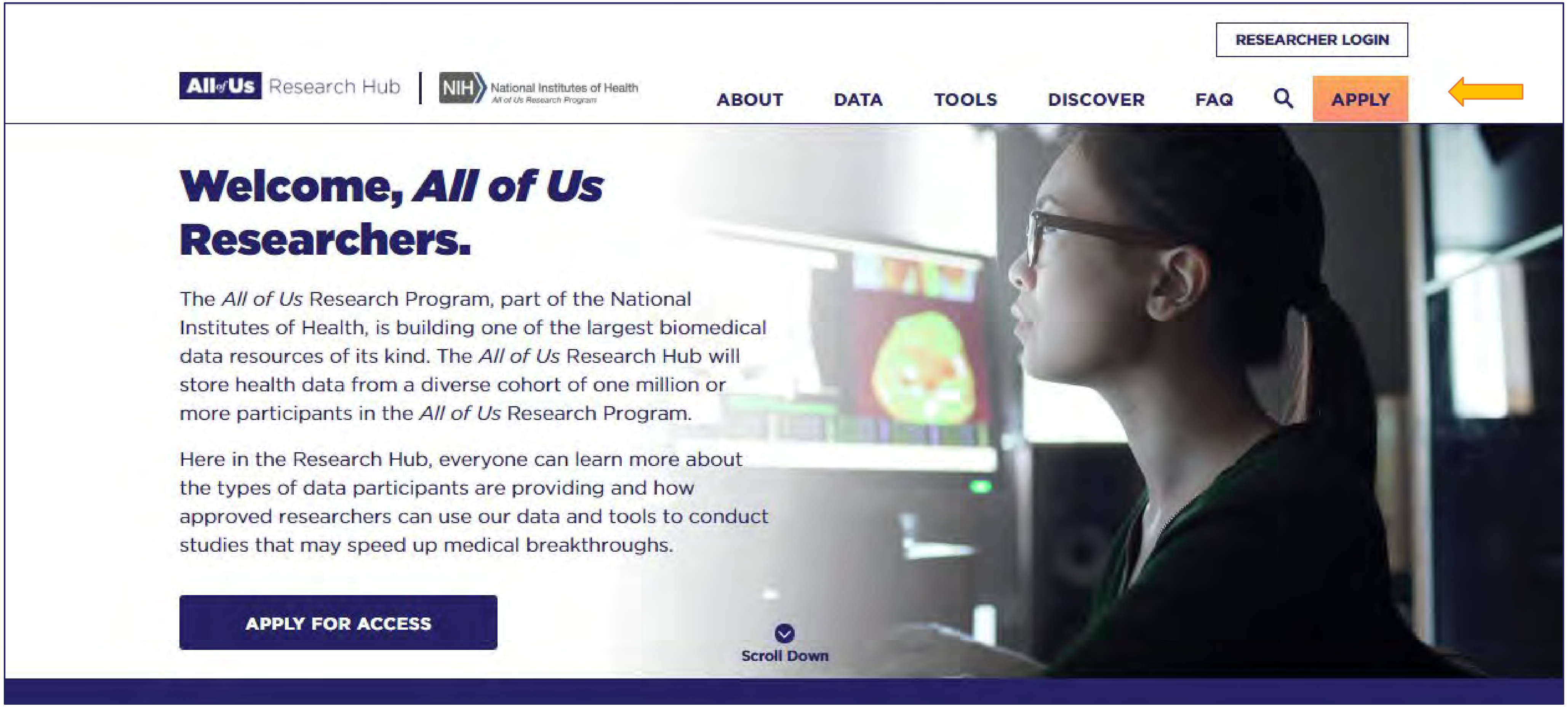


A. Ramirez, L. Suleiman, D. Schlueter, et al., *The All of Us Research Program: data quality, utility, and diversity*, medRxiv 2020.05.29.20116905; doi:

How do you get access?

Interested in becoming a Beta Researcher?

Visit researchallofus.org to apply!



Welcome, *All of Us* Researchers.

The *All of Us* Research Program, part of the National Institutes of Health, is building one of the largest biomedical data resources of its kind. The *All of Us* Research Hub will store health data from a diverse cohort of one million or more participants in the *All of Us* Research Program.

Here in the Research Hub, everyone can learn more about the types of data participants are providing and how approved researchers can use our data and tools to conduct studies that may speed up medical breakthroughs.

How the Application Process Works

STEP
1

CHECK FOR YOUR INSTITUTION'S AGREEMENT

Check that your institution has signed the Data Use and Registration Agreement. If you do not see your institution listed, please fill out the form to initiate the process.

STEP
2

REGISTER AS A RESEARCHER

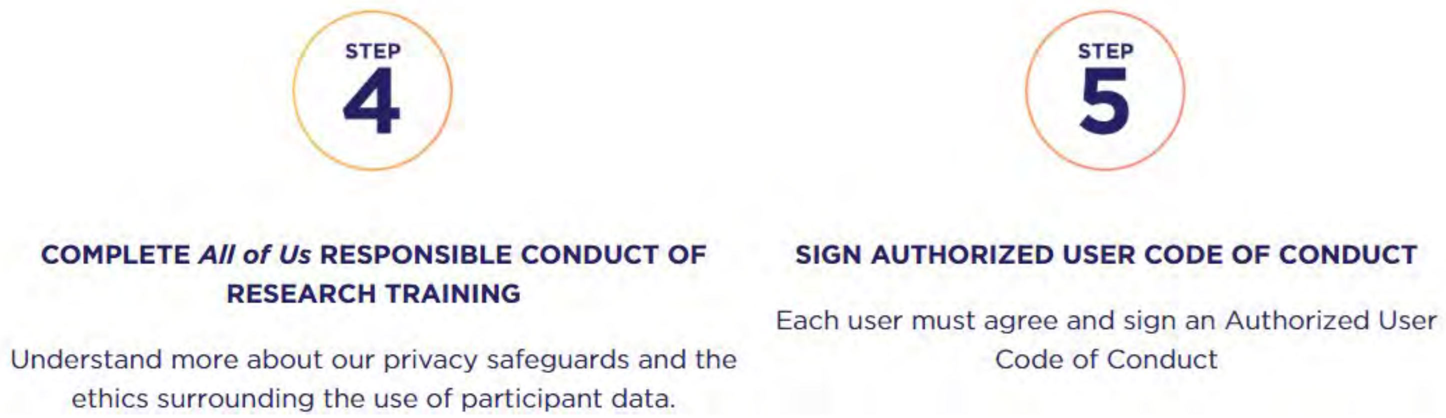
Complete your researcher profile, sign Terms of Services, and agree to the the Privacy Policy. Please note, a Data Use and Registration Agreement must be in place, and you must have an eRA Commons account.

STEP
3

CONNECT eRA COMMONS

Connect your eRA Commons account upon applying to the Researcher Workbench.

https://www.researchallofus.org/institutional-agreements/



STEP
4

COMPLETE *All of Us* RESPONSIBLE CONDUCT OF RESEARCH TRAINING

Understand more about our privacy safeguards and the ethics surrounding the use of participant data.

STEP
5

SIGN AUTHORIZED USER CODE OF CONDUCT

Each user must agree and sign an Authorized User Code of Conduct

Institutional Agreements

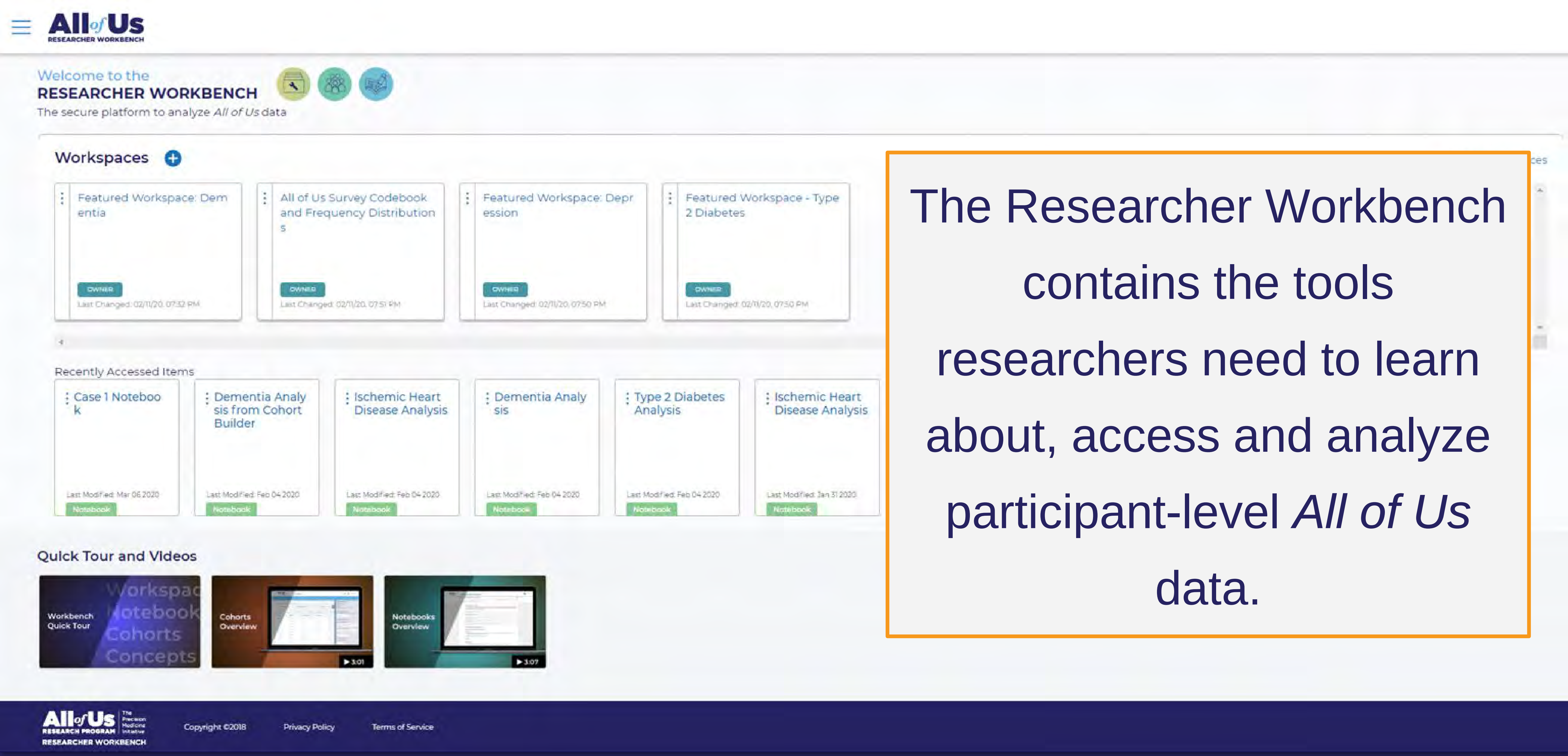
Currently, 165 institutions are in the access pipeline with 110 having completed agreements!

Institutions are Diverse:

- Non-profit Orgs
- Federally Qualified Health Centers
- Historically Black Colleges & Universities
- R2 Public Universities
- Major Academic Medical Centers
- Disease Advocacy Groups
- *And growing….*

Ok now what?

Welcome to the Researcher Workbench



Welcome to the
RESEARCHER WORKBENCH
The secure platform to analyze *All of Us* data

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

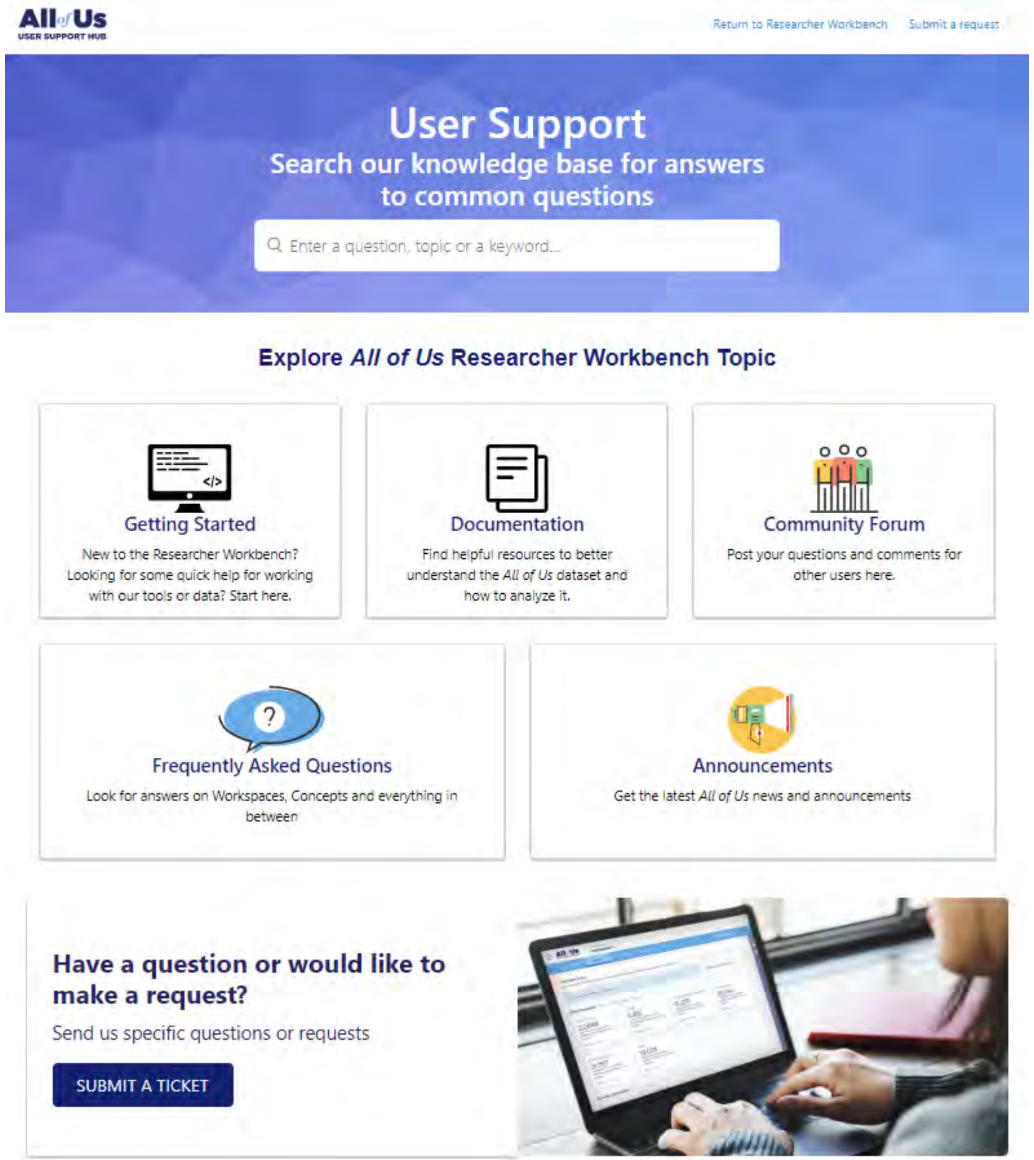
Recently Accessed Items

- 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

The Researcher Workbench contains the tools researchers need to learn about, access and analyze participant-level *All of Us* data.

Searchable Knowledge Base, Community Forums, and Help Desk



All of Us
USER SUPPORT HUB

Return to Researcher Workbench Submit a request

User Support

Search our knowledge base for answers to common questions

Enter a question, topic or a keyword...

Explore *All of Us* Researcher Workbench Topic

Getting Started
New to the Researcher Workbench? Looking for some quick help for working with our tools or data? Start here.

Documentation
Find helpful resources to better understand the *All of Us* dataset and how to analyze it.

Community Forum
Post your questions and comments for other users here.

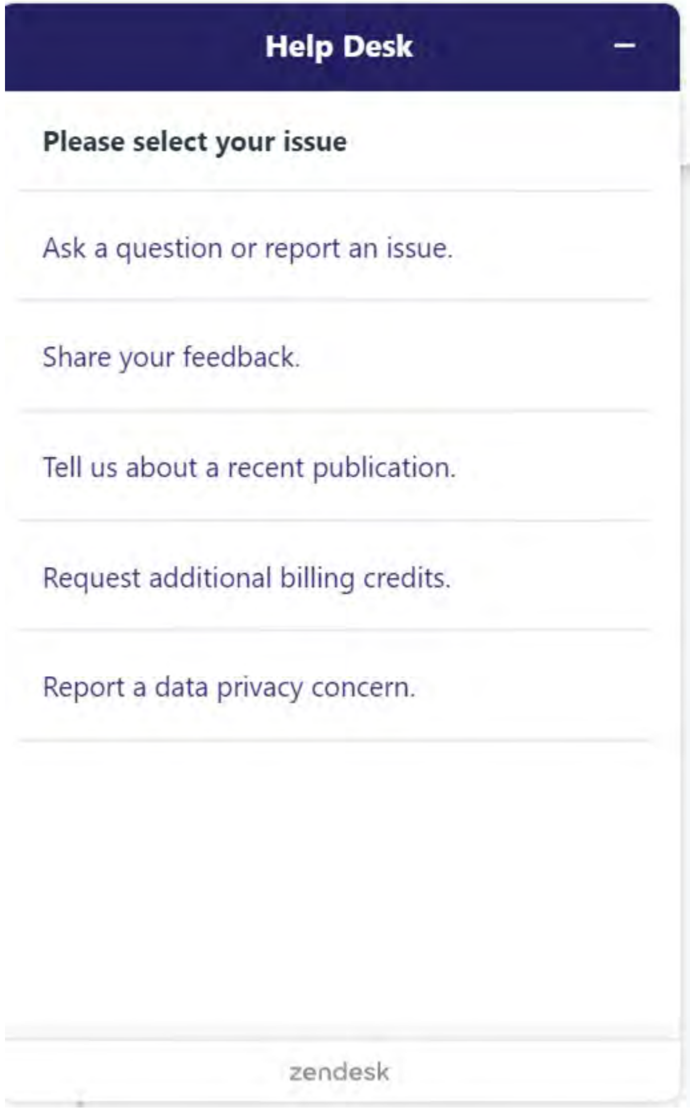
Frequently Asked Questions
Look for answers on Workspaces, Concepts and everything in between

Announcements
Get the latest *All of Us* news and announcements

Have a question or would like to make a request?

Send us specific questions or requests

SUBMIT A TICKET



Help Desk

Please select your issue

Ask a question or report an issue.

Share your feedback.

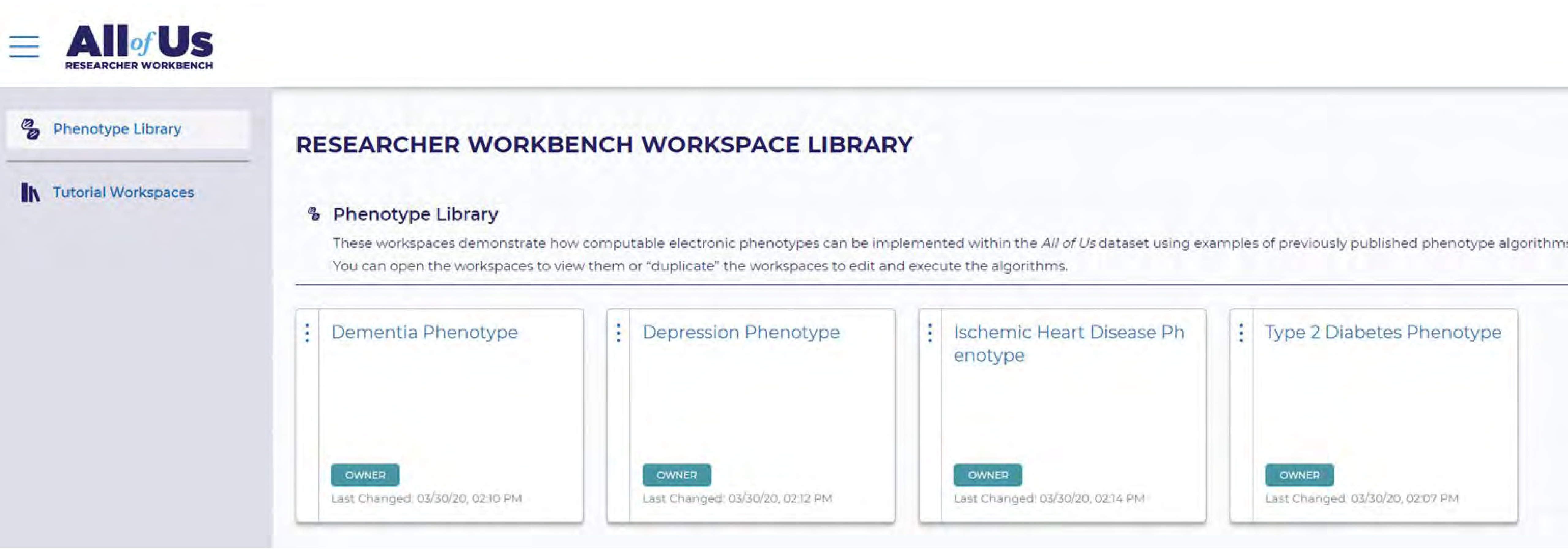
Tell us about a recent publication.

Request additional billing credits.

Report a data privacy concern.

zendesk

All of Us Phenotype Library



RESEARCHER WORKBENCH WORKSPACE LIBRARY

Phenotype Library

These workspaces demonstrate how computable electronic phenotypes can be implemented within the *All of Us* dataset using examples of previously published phenotype algorithm
You can open the workspaces to view them or “duplicate” the workspaces to edit and execute the algorithms.

Dementia Phenotype

Last Changed: 03/30/20, 02:10 PM

Depression Phenotype

Last Changed: 03/30/20, 02:12 PM

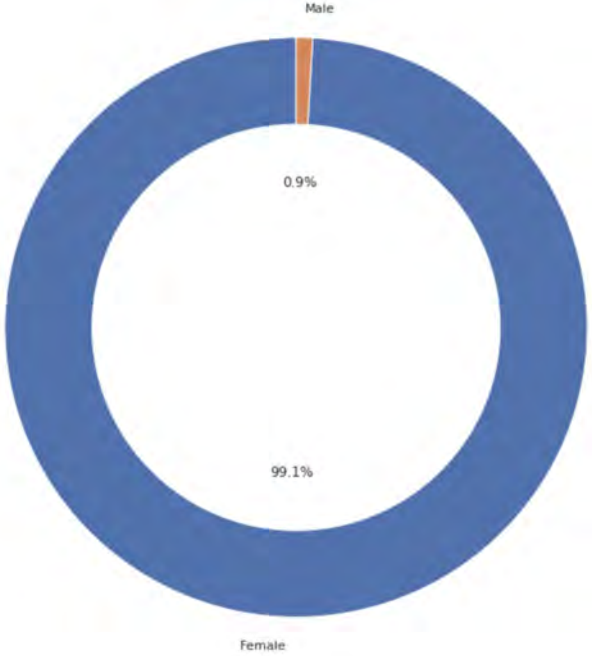
Ischemic Heart Disease Phenotype

Last Changed: 03/30/20, 02:14 PM

Type 2 Diabetes Phenotype

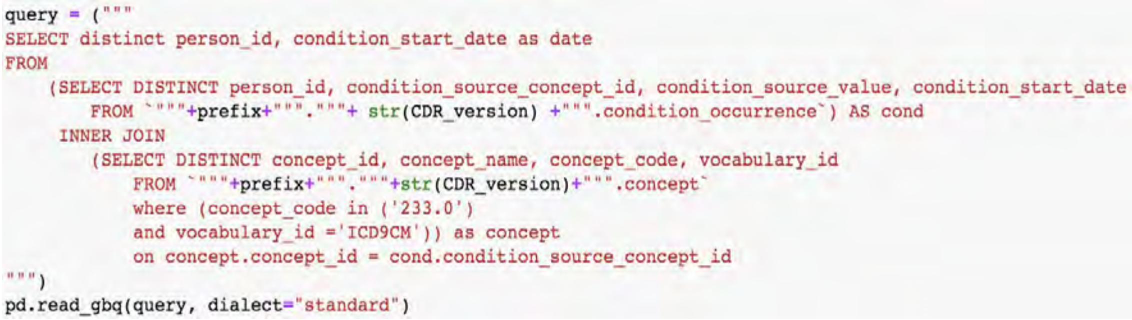
Last Changed: 03/30/20, 02:07 PM

Phenotype Library Example: Breast Cancer



Breast Cancer Cases

Example Query of ICD Codes in notebook:



```
query = ("""
SELECT distinct person_id, condition_start_date as date
FROM
    (SELECT DISTINCT person_id, condition_source_concept_id, condition_source_value, condition_start_date
        FROM `"""+prefix+"""."""+ str(CDR_version) +""".condition_occurrence`) AS cond
     INNER JOIN
        (SELECT DISTINCT concept_id, concept_name, concept_code, vocabulary_id
            FROM `"""+prefix+"""."""+str(CDR_version)+""".concept`
            where (concept_code in ('233.0')
            and vocabulary_id ='ICD9CM')) as concept
            on concept.concept_id = cond.condition_source_concept_id
""")
pd.read_gbq(query, dialect="standard")
```

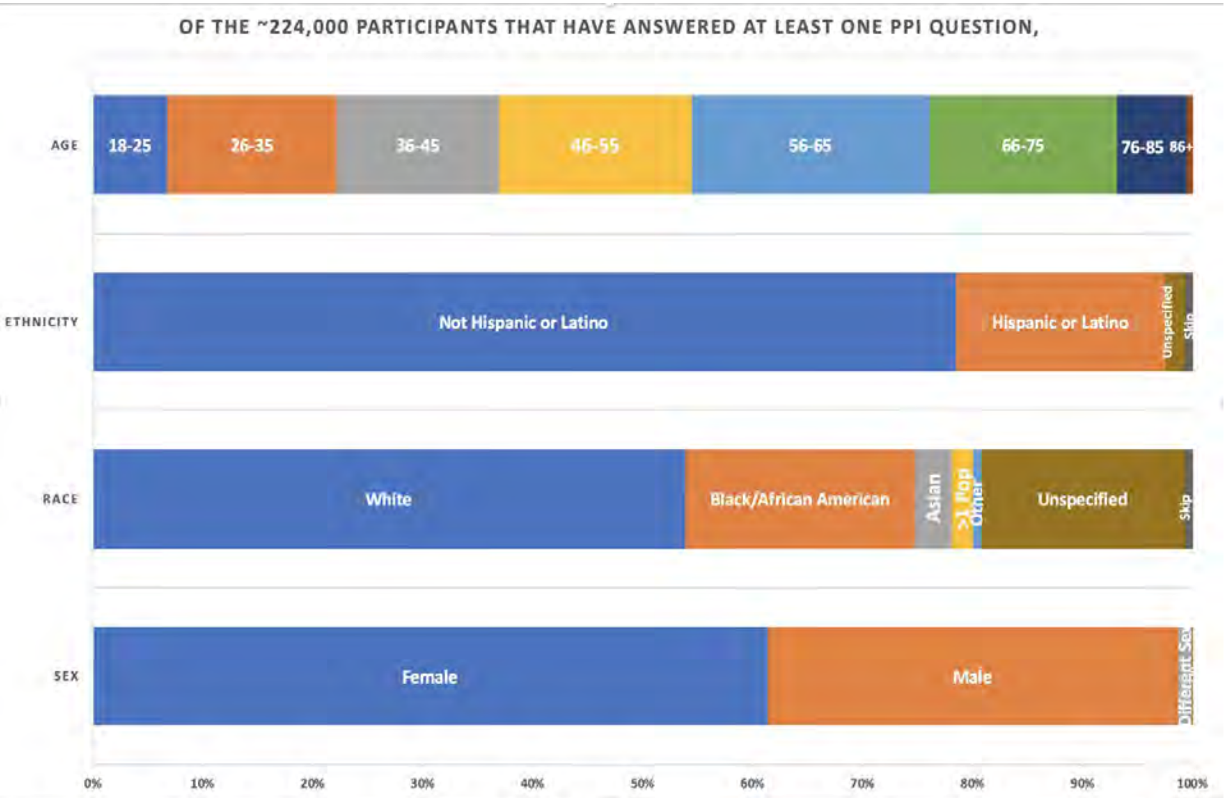
We implemented a validated phenotyping algorithm for breast cancer from eMERGE network.

1. We found 3748 total breast cancer cases
   a. 3714 female cases (99.1%)
   b. 34 male cases (0.9%)
2. Algorithm utilized diagnosis/history codes in the OMOP observation and condition occurrence tables
3. This algorithm will be made available to researchers as part of the Phenotype library
   a. Cohort builder implementation and notebook query implementation match

Issues encountered:

1. Be careful about how you specify certain codes in cohort builder vs. notebooks

Tutorial Workspace Example: Working with Survey Data



Ex. How to query survey data

1. View a pdf of the **Overall Health** Survey using [Survey Explorer]
2. Identify question of interest:

   <u>How confident are you filling out medical forms by yourself?</u>

   Extremely / Quite a bit / Somewhat / A little bit / Not at all / Skip

1. Go to [OHDSI ATHENA] to find the associated ID, Code, and Name

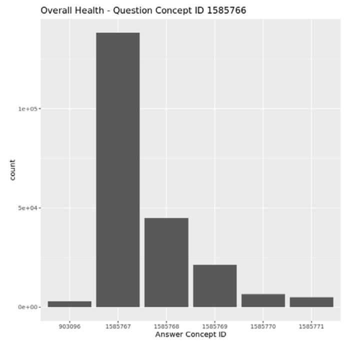
Question	Question ID (ID)	Question Code (CODE)	Question Name (NAME)	Data Type (CLASS)	Data Sourve (VOCAB)
How confident are you filling out medical forms by yourself?	1585766	OverallHealth_MedicalFormConfidence	Overall Health: Medical Form Confidence	Question	PPI

1. In a Notebook, query the *All of Us* custom **DS_SURVEY** table
   WHERE question_concept_id = 1585766

1. Output frequency table of answers

6. Plot counts of responses

Medical Form Confidence: Extremely	138210	63.1
Medical Form Confidence: Quite A Bit	44894	20.5
Medical Form Confidence: Somewhat	21420	9.8
Medical Form Confidence: A Little Bit	6586	3.0
Medical Form Confidence: Not At All	5040	2.3
PMI: Skip	2865	1.3
Total	219015	100.0

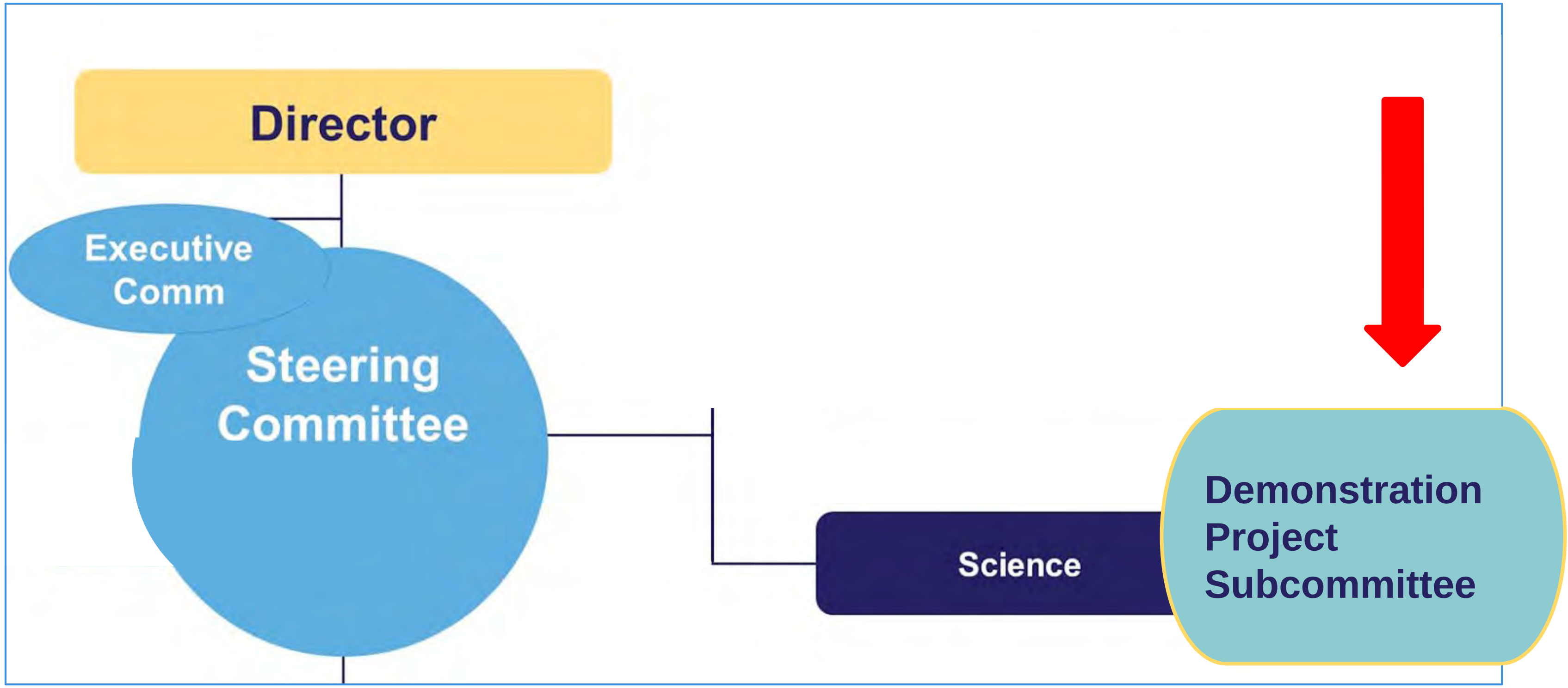


Issues Encountered

- Data is formatted **long**
- "**Child questions**" are not always answered by the intended participants
- Several tables contain survey information [`DS_SURVEY, CONCEPT, OBSERVATION, OBSERVATION_EXT`] and **field names may vary**
- May be joined together, but that takes **computational time**

All of Us Demonstration Project Subcommittee

The DRC AoU Demonstration Projects will use Research Hub tools to characterize and validate the AoU cohort and data in order to demonstrate the quality, utility, and diversity of AoU data and tools. AoU Demonstration Projects will replicate known, previously published findings, and will not aim to discover new disease classifications, relationships, or risk associations -- ***our role is to support the community's use of data, not to make first discoveries****.*



Demonstrate Scientific Utility

Members: Andrea Ramirez (Chair), Kelly Gebo, Cheryl Clark, Elizabeth Cohn, Lucila Ohno-Machado, Mine Cicek, Eric Boerwinkle, Sheri Schully, Steve Mockrin

AoU Science Committee: Demonstration Projects Subcommittee

NOVEMBER 2019

Goal: Fully executed research projects demonstrating the utility and validity of AoU data timed to publish at data platform launch, *not* novel discovery work.

Phase 1
[DRC] Description, Replication, Utility Assessment

Phase 2
[Consortium] Expanded Description, Replication, Utility Assessment

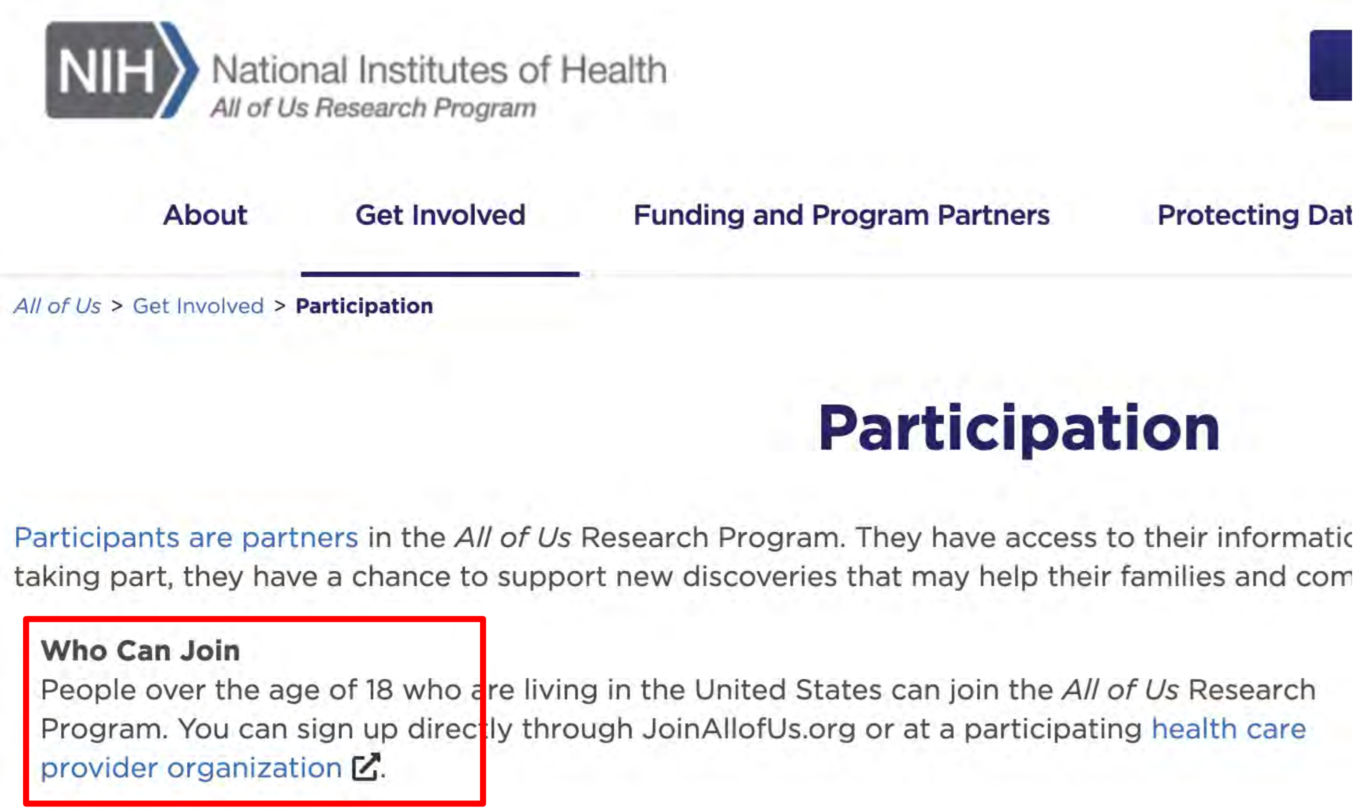
Phase 3
[Consortium] Future Preceding new data types

Engaging a large consortium, covering broad health areas

***All of Us* Awardee**	**No. of Approved Projects**
Trans-American Consortium for the Health Care Systems Research Network (TACH)	4
NYC Precision Medicine Consortium	4
The Participant Center	5
Asian Health Coalition	2
University of Arizona	4
New England Precision Medicine Consortium	4
California Precision Medicine	2
Stanford University	1
Illinois Precision Medicine Consortium	5
South East Enrollment Center (SEEC)	4
Data and Research Center (DRC)	3

Health Areas	**No. Projects in Health Area**
Cardiovascular Disease	7
Cancer	5 (2 overlaps with other area)
Diabetes and Obesity	6 (3 overlap with other areas)
Mental Health	2
Wellness	2
Opioids and Pain	2
Chronic Kidney Disease	1
Chronic lower respiratory disease	1
Neurodegenerative condition and cognition	1
Other areas: Environmental exposures, Health disparities, Infectious disease, Body temperature, Women’s health, Pediatric data description, Hematologic disorders	13
Gap addressing projects: Ophthalmology Compare Census and CDC data in AoU Rare diseases	3

D38: Exploration of Pediatric Data from *All of Us* -Is there pediatric data in AoU?



National Institutes of Health
All of Us Research Program

All of Us > Get Involved > **Participation**

Participation

Participants are partners in the *All of Us* Research Program. They have access to their informatic
taking part, they have a chance to support new discoveries that may help their families and com

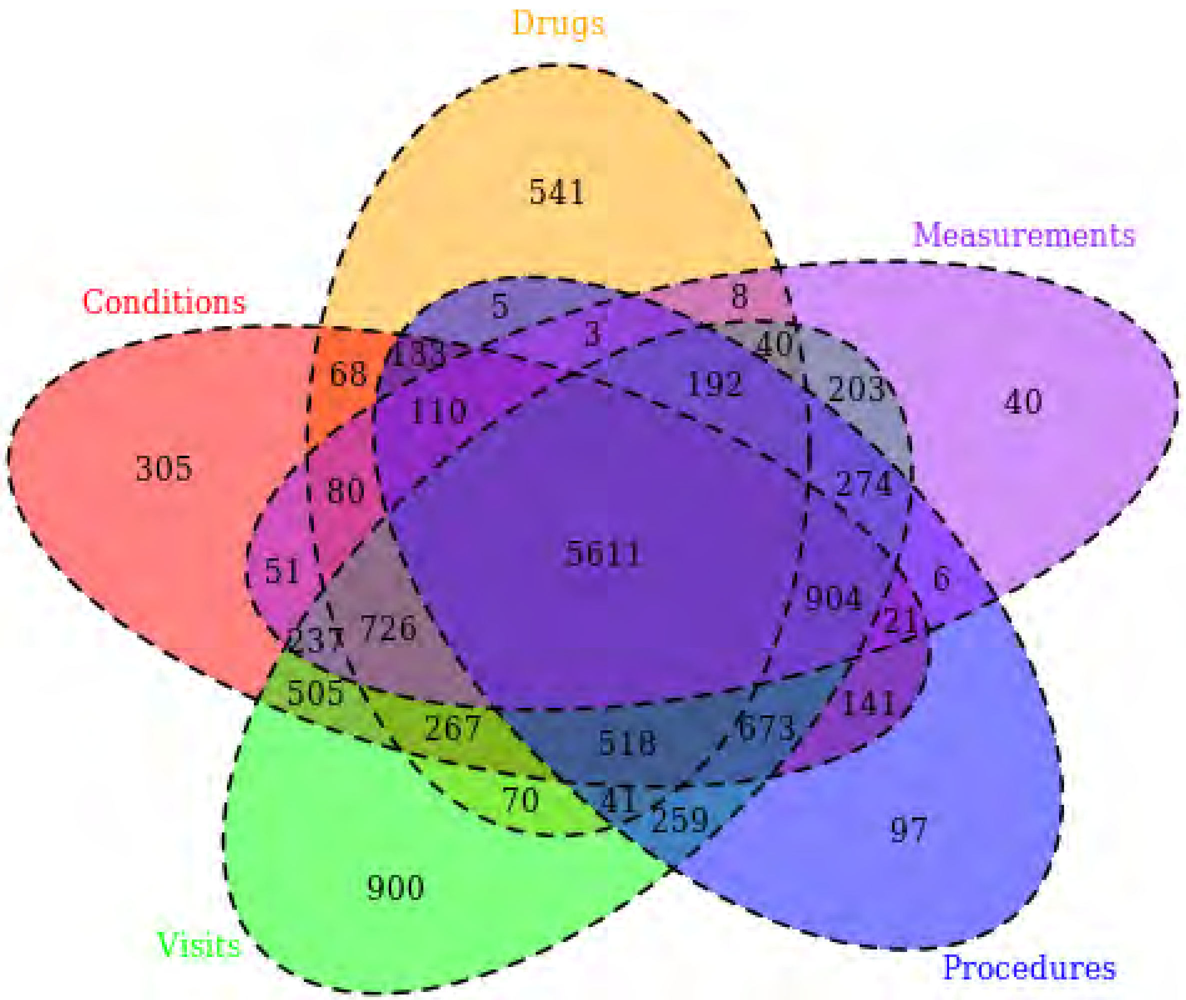
Who Can Join
People over the age of 18 who are living in the United States can join the *All of Us* Research Program. You can sign up directly through JoinAllofUs.org or at a participating health care provider organization.

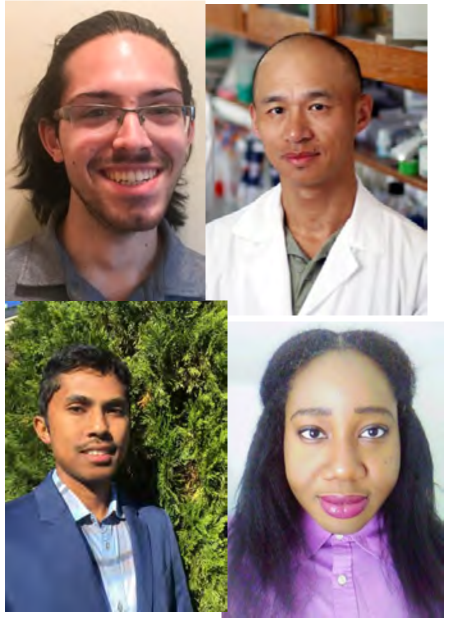
- When patients enroll, EHRs from childhood may come along
- Questions:
  - How much data is there?
  - What data types are there?
  - Can the data be used for pediatric research?

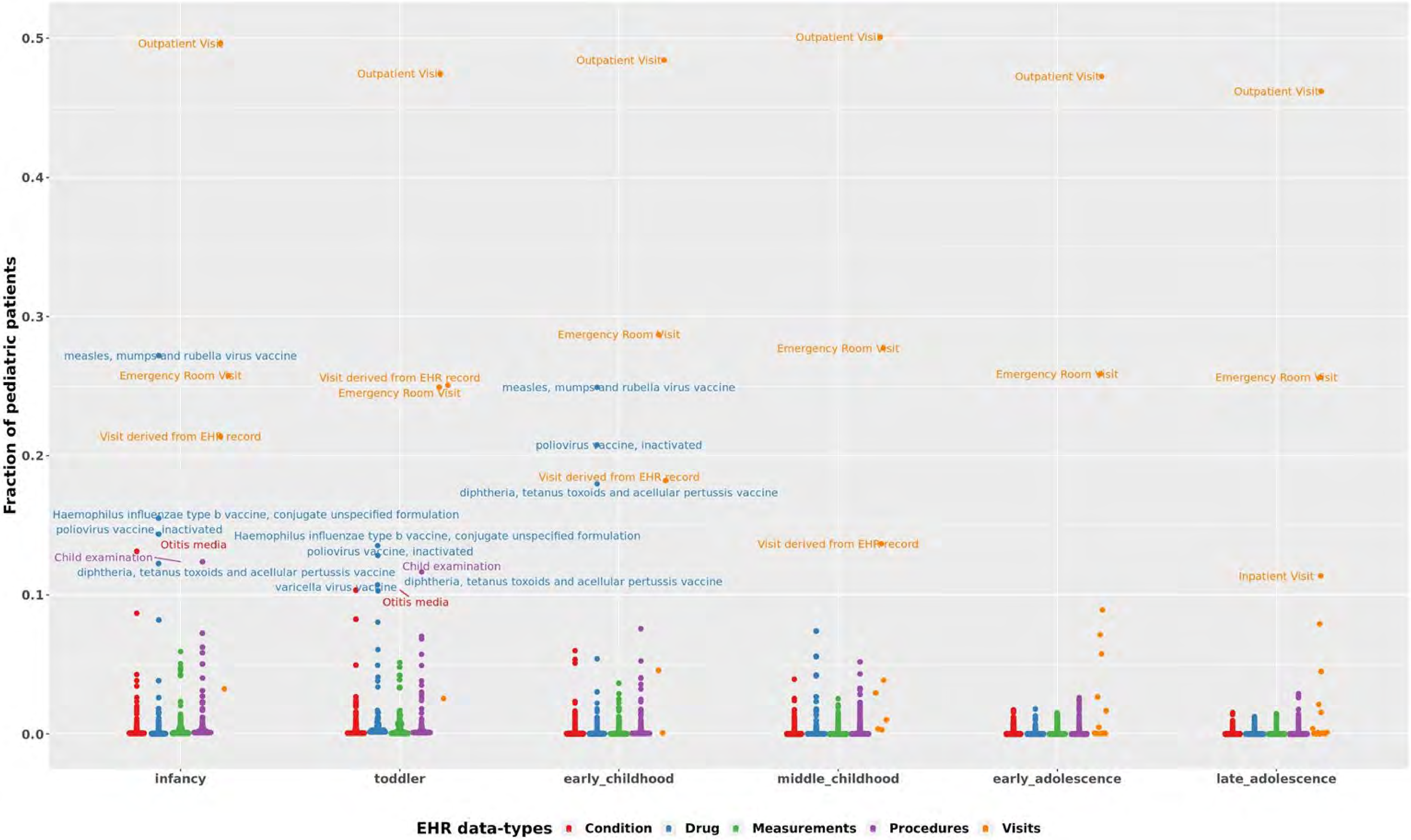
D38: Exploration of Pediatric Data from *All of Us*

Patients with:

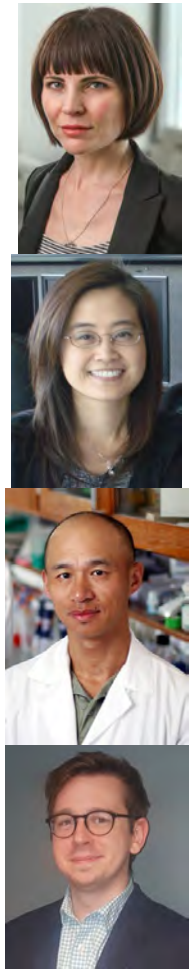
- Drugs: 4070
- Conditions: 4865
- Visits: 5718
- Procedures: 4537
- Measurements: 3633






Fraction of pediatric patients
0.5
0.4
0.3
0.2
0.1
0.0
Outpatient Visit
Outpatient Visit
Outpatient Visit
Outpatient Visit
Outpatient Visit
Outpatient Visit
Emergency Room Visit
measles, mumps and rubella virus vaccine
Emergency Room Visit
Visit derived from EHR record
Emergency Room Visit
Emergency Room Visit
Emergency Room Visit
Emergency Room Visit
measles, mumps and rubella virus vaccine
Emergency Room Visit
Visit derived from EHR record
poliovirus vaccine, inactivated
Visit derived from EHR record
diphtheria, tetanus toxoids and acellular pertussis vaccine
Haemophilus influenzae type b vaccine, conjugate unspecified formulation
poliovirus vaccine, inactivated
Haemophilus influenzae type b vaccine, conjugate unspecified formulation
Visit derived from EHR record
Otitis media
poliovirus vaccine, inactivated
Child examination
diphtheria, tetanus toxoids and acellular pertussis vaccine
Child examination
varicella virus vaccine
diphtheria, tetanus toxoids and acellular pertussis vaccine
Inpatient Visit
Otitis media
infancy
toddler
early_childhood
middle_childhood
early_adolescence
late_adolescence
EHR data-types
Condition
Drug
Measurements
Procedures
Visits

D43: Phenotype Risk Score (PheRS) Implementation



Goal: To replicate PheRS approach for three mendelian diseases including cystic fibrosis (CF), hereditary hemochromatosis (HH) and sickle cell (SC) anemia in allofus (AOU) cohort.

Contributors:

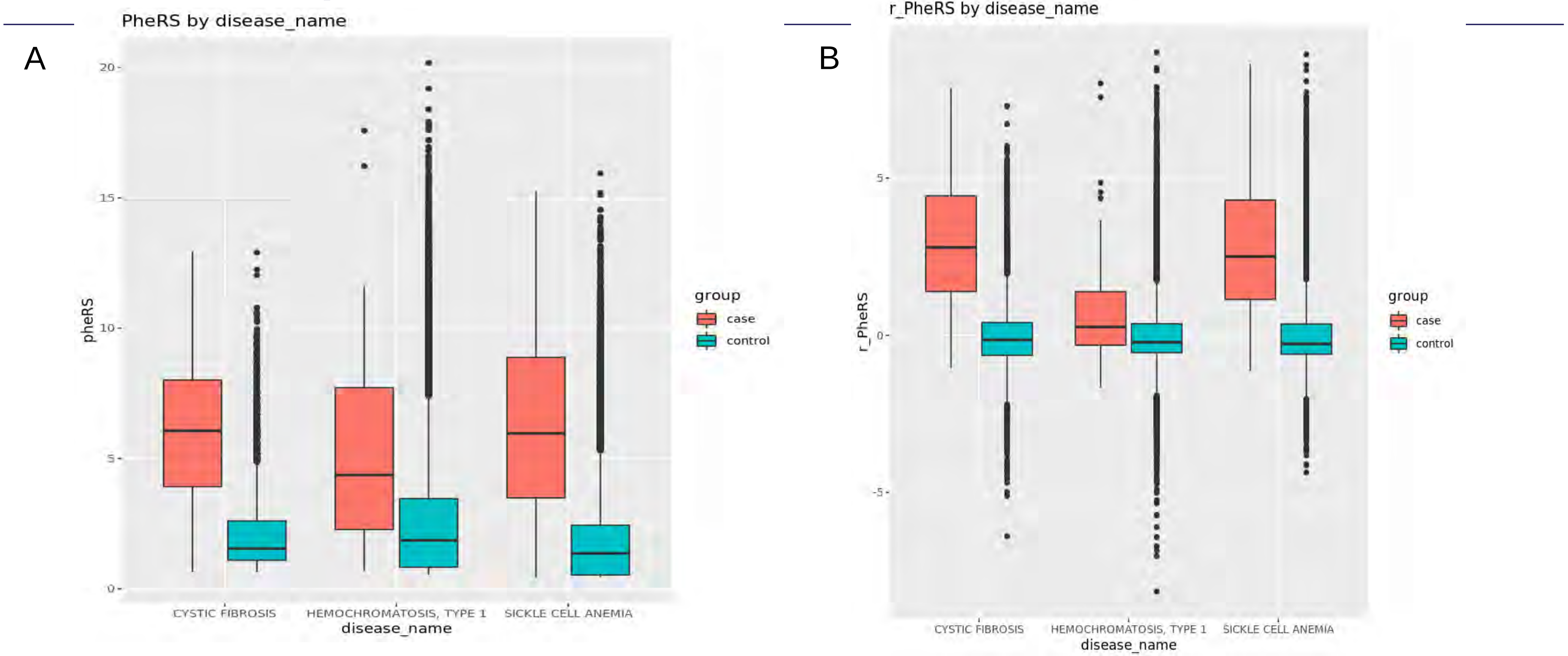
Jun Qian, Lisa Bastarache, David Schlueter, Janey Wang, Andrea Ramirez
Vanderbilt University Medical Center

Ref: Lisa Bastarache, et al. **Phenotype risk scores identify patients with unrecognized Mendelian disease patterns**. ***Science*** 2018, 359 (6381), 1233-1239. DOI: 10.1126/science.aal4043

Improving the phenotype risk score as a scalable approach to identifying patients with Mendelian disease.J Am Med Inform Assoc. 2019 Dec 1;26(12):1437-1447.

D43 PheRS Implementation

*Initial results, analysis still in progress



n=100 40503 74 71546 109 70040

P<2.2e-16 P<2.2e-16 P<2.2e-16

n=100 40503 74 71546 109 70040

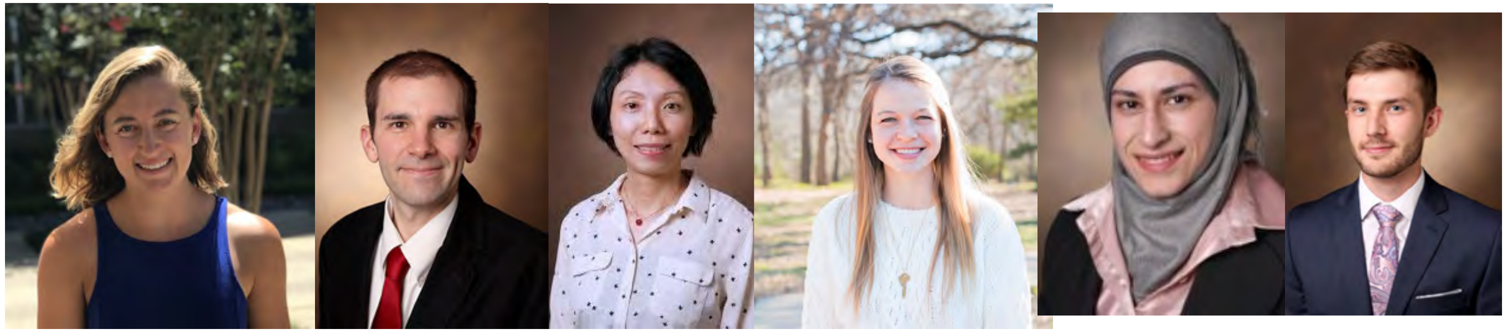
P<2.2e-16 P=6.51e-12 P<2.2e-16

Raw PheRS (A) and studentized residuals of the raw PheRS (B) for three diseases in AOU EHR data. The regression model, PheRS~bs(age)+sex+race+unique_encounter_years was used for the r_PheRS calculation (B). P values were calculated using Wilcoxon rank sum test.

D44: Comparison of family health history gathered in survey and electronic health records in *All of Us* Registered Tier Data

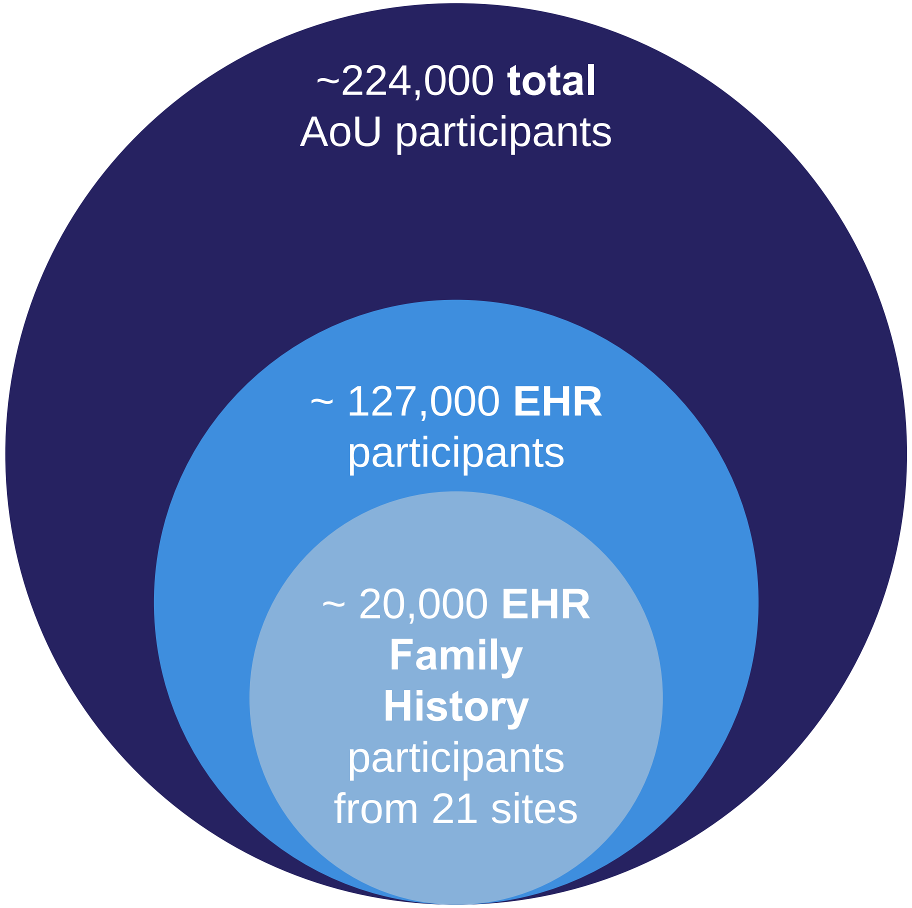
- **Background and Methods**
  - **Family health history** is integral to clinical care and important to integrated precision medicine care
  - Prior studies show gaps in data from Electronic Health Record **(EHR)** structured fields
  - Prior studies have been assisted by **free text extraction**
  - Surveys have shown great promise using common data models

All of Us has **both** data sources for family health history information.



Data and Research Center – Family History in EHR vs. Surveys

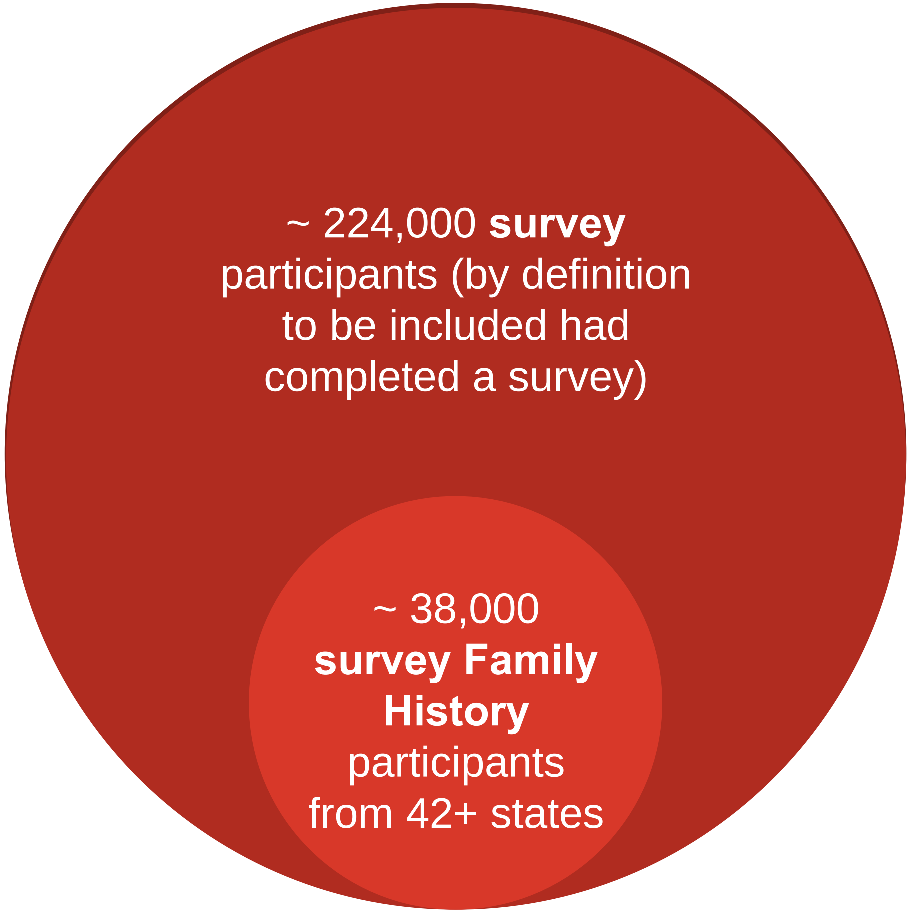
EHR


~224,000 **total** AoU participants

~ 127,000 **EHR** participants

~ 20,000 **EHR Family History** participants from 21 sites

Surveys


~ 224,000 **survey** participants (by definition to be included had completed a survey)

~ 38,000 **survey Family History** participants from 42+ states

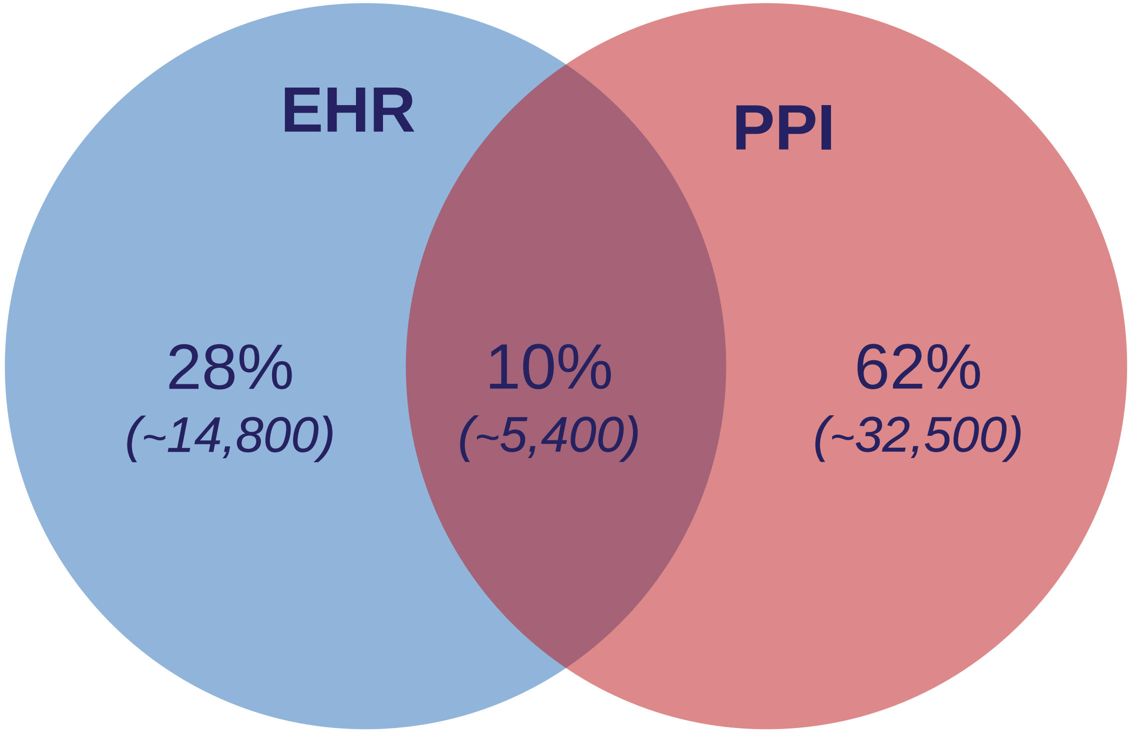
Is EHR confirming what we already knew from suryeing participants? Or are we getting new information?

Data and Research Center – Family History in EHR vs. PPI

- **Results and Conclusions**

Participants with any Family History data

Total: ~52,700

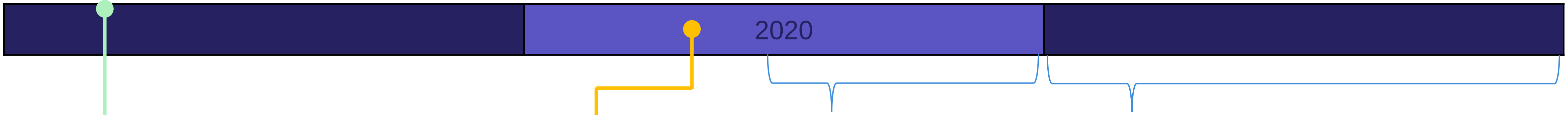


Phase II Demo Project Culmination Meeting

JANUARY 2020

- 50 overall attendees, 26 demo project team members
- 9 out of 11 Consortium Awardees represented
- 22 different projects presented both on-site and remotely (of 38 total)
- Discussions around feedback mechanism and new beta features
- Discussions around authorship and alpha access next steps and publication of projects internally in Featured Workspaces

Data Release Timeline



Data Browser Launch

May 6, 2019

Launch of an interactive tool available to the public that provides summary statistics from the program's growing database.

Researcher Workbench Beta Launch

May 2020

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

The dataset includes:

- Survey data
- Measurement data
- Electronic health record (EHR) data

Dataset Expansion

Expansion 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

Additional Data Types and Tools

2021 & beyond

Expect to enroll 1M participants within 5-7 years of launch.

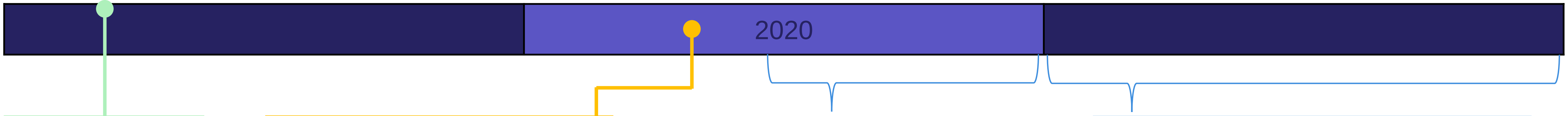
Continue to expand the *All of Us* dataset with new data releases, including over time:

- genomics data
- wearables data (such as Fitbit)
- new surveys
- assay data
- linkages to external data sources, etc.

Release new versions of Researcher Workbench with more tools.

**All of Us* is targeting a bold timeline. Projected dates are subject to change and dependent on security and usability testing.

Data Release Timeline



Data Browser Launch
May 6, 2019

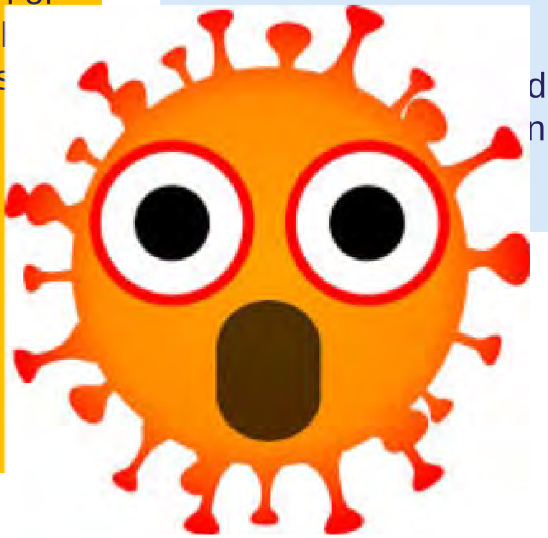
Launch of an interactive tool available to the public that provides summary statistics from the program's growing database.

Researcher Workbench Beta Launch
May 2020

Release of the first version of the *All of Us* dataset and t
first set of researcher tools
beta testing.

The dataset includes:

- Survey data
- Measurement data
- Electronic health record
  (EHR) data



Dataset Expansion

Expansion of the *All of Us* dataset with additional survey, measurement,

d whole
nitial return of

Additional Data Types and Tools
2021 & beyond

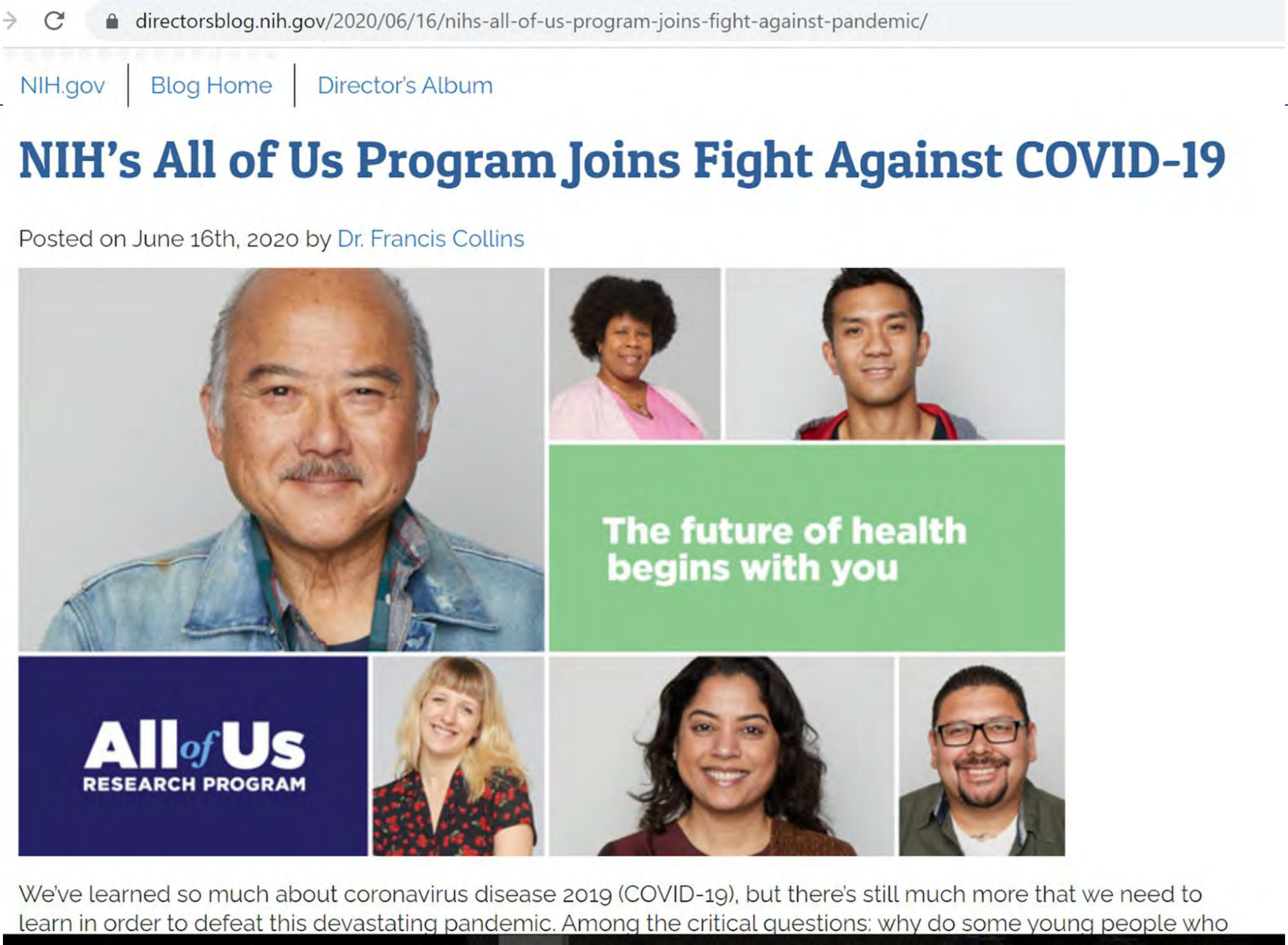
Expect to enroll 1M participants within 5-7 years of launch.

Continue to expand the *All of Us* dataset with new data releases, including over time:

- genomics data
- wearables data (such as Fitbit)
- new surveys
- assay data
- linkages to external data sources, etc.

Release new versions of Researcher Workbench with more tools.

**All of Us* is targeting a bold timeline. Projected dates are subject to change and dependent on security and usability testing.



NIH's All of Us Program Joins Fight Against COVID-19

Posted on June 16th, 2020 by Dr. Francis Collins

We've learned so much about coronavirus disease 2019 (COVID-19), but there's still much more that we need to learn in order to defeat this devastating pandemic. Among the critical questions: why do some young people who

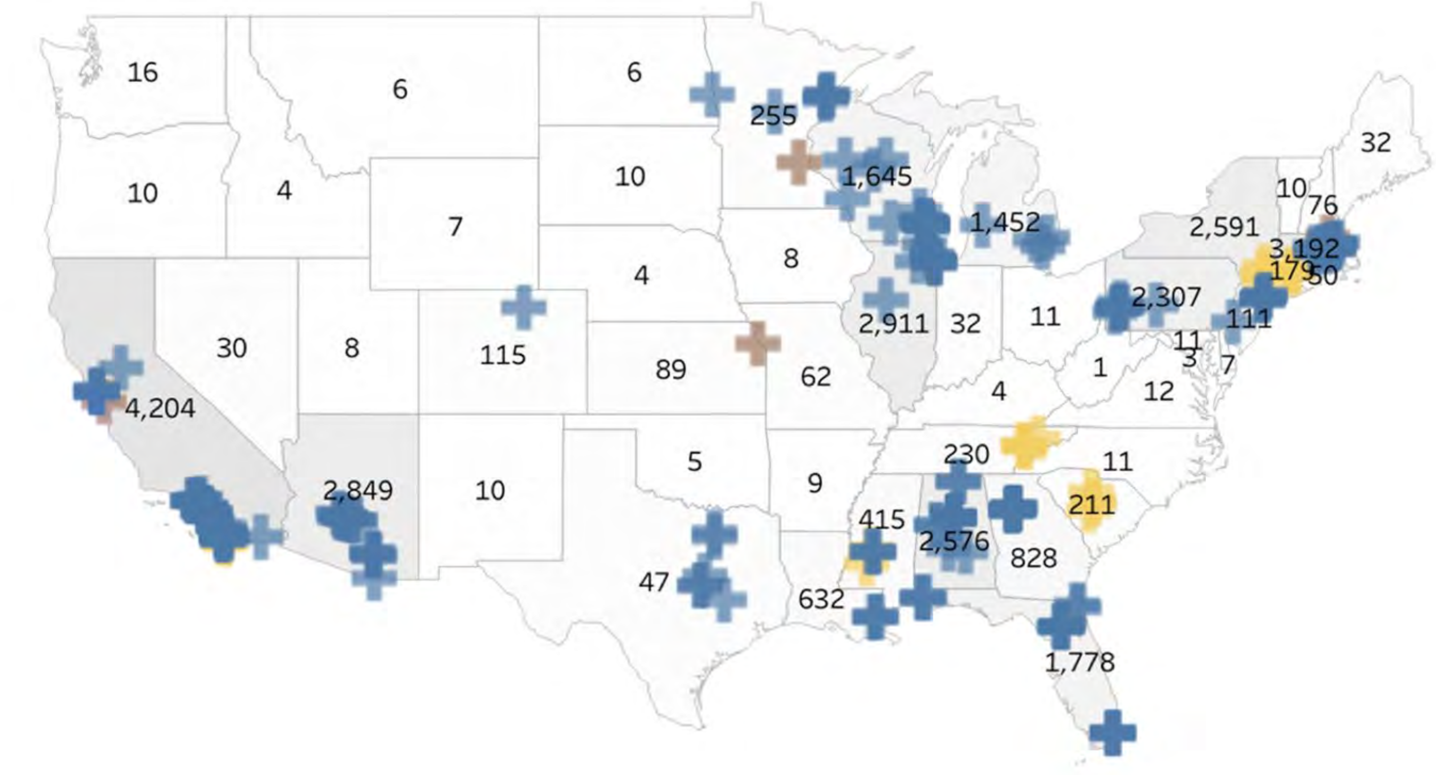
Pivot to COVID-19

0. Enrollment paused in March

1. Serology study (see map)

2. EHR curation effort

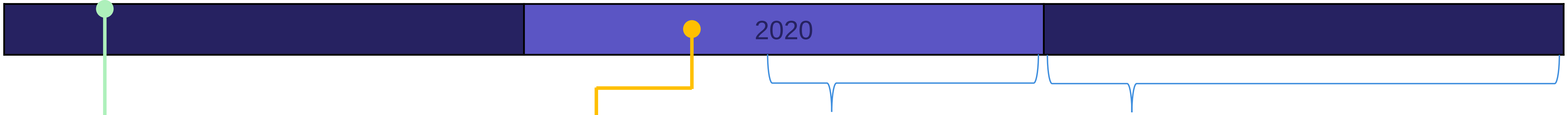
3. New survey: longitudinal collection on health and well being



"Aside from physical health concerns, the pandemic may be stressful to many of us. It can be difficult to cope with fear and anxiety, changing daily routines, a general sense of uncertainty, financial and economic hardships, social isolation, and/or stigmatization. I join you in acknowledging there are other unforeseen effects stemming from this global health pandemic.

Francis Collins, M.D., Ph.D., Director

Data Release Timeline



Data Browser Launch
May 6, 2019

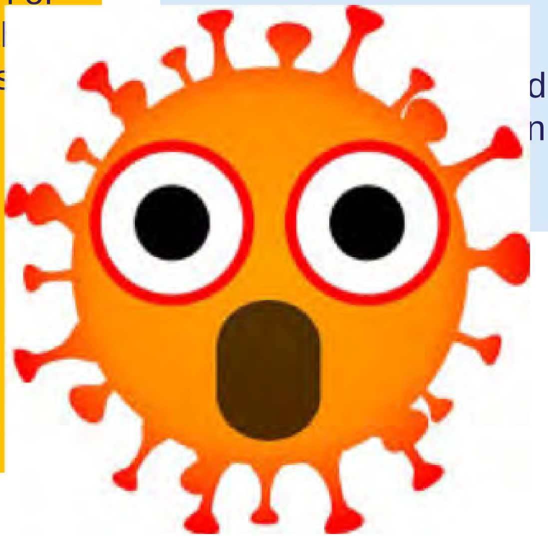
Launch of an interactive tool available to the public that provides summary statistics from the program's growing database.

Researcher Workbench Beta Launch
May 2020

Release of the first version of the *All of Us* dataset and t
first set of researcher tools
beta testing.

The dataset includes:

- Survey data
- Measurement data
- Electronic health record (EHR) data



Dataset Expansion

Expansion of the *All of Us* dataset with additional survey, measurement,

d whole
nitial return of

Additional Data Types and Tools
2021 & beyond

Expect to enroll 1M participants within 5-7 years of launch.

Continue to expand the *All of Us* dataset with new data releases, including over time:

- genomics data
- wearables data (such as Fitbit)
- new surveys
- assay data
- linkages to external data sources etc.

Release new versions of Researcher Workbench with more tools.

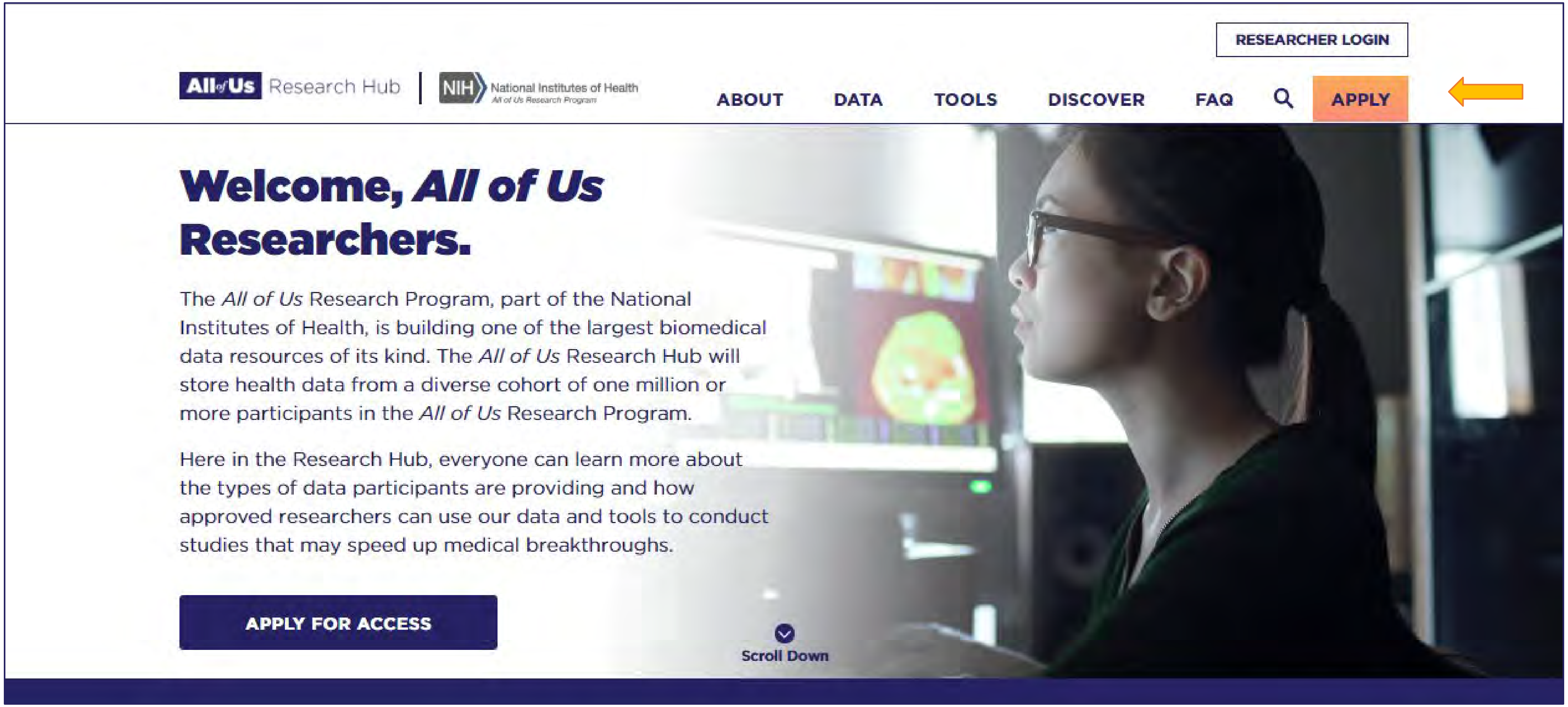
+ COVID-19
DATA

All of Us RESEARCH PROGRAM

**All of Us* is targeting a bold timeline. Projected dates are subject to change and dependent on security and usability testing.

Interested in becoming a Beta Researcher?

Visit researchallofus.org to apply!



Welcome, *All of Us* Researchers.

The *All of Us* Research Program, part of the National Institutes of Health, is building one of the largest biomedical data resources of its kind. The *All of Us* Research Hub will store health data from a diverse cohort of one million or more participants in the *All of Us* Research Program.

Here in the Research Hub, everyone can learn more about the types of data participants are providing and how approved researchers can use our data and tools to conduct studies that may speed up medical breakthroughs.

All of Us Community and Provider Partner Network *(as of January 2020)*

American Association on Health and Disability

AAFP

AAHIVM American Academy of HIV Medicine

LULAC

American Association of Colleges of Nursing

ΦΙΛΑΥΟΜΕΝ ΥΠΗΡΕΤΕΙΝ

APHA American Public Health Association For science. For action. For health.

ACCESS assisting. improving. empowering.

APIAHF Asian & Pacific Islander American Health Forum

ANAC Association of Nurses in AIDS Care

Asian Health Coalition

ΙΦΘ

W. Montague Cobb NMA Health Institute

the Y YMCA

ΖΦΒ 1920

Stanford Medicine

Founded ΣΓΡ 1922

Delta Research and Educational Foundation Creating possibilities improving lives

ΩΨΦ

Fifty Forward

NACCHO National Association of County & City Health Officials The National Connection for Local Public Health

CenterLink The Community of LGBT Centers

National Alliance for Hispanic Health

National Pan-Hellenic Council

NAHN National Association of Hispanic Nurses

NBNA National Black Nurses Association, Inc.

NHMA National Hispanic Medical Association

National Minority Quality Forum

NIH U.S. National Library of Medicine

Black Women's Health Imperative

NRHA National Rural Health Association

Phi Beta Sigma Fraternity, Inc. ΦΒΣ Founded 1914 Brotherhood · Scholarship · Service

Delta Sigma Theta ΔΣΘ

NHCOA National Hispanic Council on Aging

UnidosUS Stronger Communities. Stronger America.

ΑΦΑ

National Baptist Convention, U.S.A. Incorporated

Community Council Advancing Solutions...Empowering Lives

All of Us Consortium Members *(beyond community partners, as of March 2020)*

The Participant Center

Scripps Research Translational Institute
BlueCross BlueShield
Community Blood Center of the Carolinas
DXC.technology
EMSI Health
Montage Marketing
ACHA American College Health Association
Medscape
PicnicHealth
New York Blood Center
Gulf Coast Regional Blood Center Commit for Life.
The Behavioural Insights Team
Walgreens
National Blood Collaborative
patientslikeme
Urban One
fitbit
Sage Bionetworks
QTC
WebMD
Bloodworks Northwest
Qualcomm
Sync for Science
San Diego Blood Bank
Quest Diagnostics
careevolution Healthcare Technology

Communications & Engagement

WONDROS
The University of Utah
HCM Strategists

HPO Network

(Health Care Provider Organizations)

RMCs

All of Us California
- UC San Diego Health
- UCI Health
- UCSF University of California San Francisco
- UC Davis Health
- Cedars-Sinai
- USC University of Southern California

Illinois Precision Medicine Consortium
- Northwestern Medicine
- The University of Chicago
- NorthShore University HealthSystem
- Rush
- Cook County Health
- The University of Illinois College of Medicine
- UI Health UIC

All of Us New England
- Brigham Health Brigham and Women's Hospital
- Partners HealthCare
- Massachusetts General Hospital
- Boston Medical Center

Trans America Consortium
- Henry Ford Health System
- Essentia Health
- Baylor Scott & White Health
- Spectrum Health
- Reliant Medical Group Part of OptumCare
- HealthPartners

New York City Consortium
- Columbia University Medical Center
- NYC Health + Hospitals
- NewYork-Presbyterian
- Weill Cornell Medicine

All of Us Southern Network
- UAB School of Medicine The University of Alabama at Birmingham Selma Family Medicine
- UAB School of Medicine The University of Alabama at Birmingham Montgomery Regional Medical Campus
- The University of Mississippi Medical Center
- UAB Medicine
- Tulane University
- LSU Health New Orleans
- UAB School of Medicine The University of Alabama at Birmingham Huntsville Regional Medical Campus
- The University of Alabama University Medical Center
- USA Health
- Cooper Green Mercy Health Services

All of Us SouthEast Enrollment Center
- University of Miami Miller School of Medicine
- Emory University
- Morehouse School of Medicine
- UF University of Florida
- Jackson Health System Public Health Trust

All of Us Wisconsin
- Marshfield Clinic Health System
- Wisconsin University of Wisconsin–Madison
- Froedtert & Medical College of Wisconsin

All of Us Pennsylvania
- University of Pittsburgh
- Temple University

University of Arizona and Banner Health
- The University of Arizona
- Banner Health

FQHCs (Federally Qualified Health Centers)
- Cherokee Health Systems
- Jackson-Hinds Comprehensive Health Center
- San Ysidro Health
- Community Health Center, Inc.
- Cooperative Health
- HRHCare Community Health

VA Medical Centers
- VA U.S. Department of Veterans Affairs

Participant Technology Systems Center (PTSC)

vibrent

Biobank

Mayo Clinic

Data & Research Center (DRC)

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

Genomics Partners

Baylor College of Medicine
Johns Hopkins Medicine
UTHealth The University of Texas Health Science Center at Houston
HudsonAlpha Institute for Biotechnology
Broad Institute
color
Partners HealthCare
University of Washington

All of Us Research Program

It takes *All of Us*….

For more information…

Thank you!

andrea.h.ramirez@vumc.org
@AndreaRamirezMD

ResearchAllofUs.org
(to apply as a researcher!)

JoinAllofUs.org

@AllofUsResearch
#JoinAllofUs

AllofUs.nih.gov

Additional Slides

Screenshot Walk Thru

Underrepresented in Biomedical Research (UBR) Definitions

Category	*All of Us* UBR definition
Ancestry (Race/Ethnicity)	Participant identifies as any race or ethnicity other than White, including Multi-Ancestry (any combination of 2 or more races)
Age	Participant is under 18 or 65 or over at time of consent
Sexual & Gender Minorities	**Biological sex at birth:** Participant self-reports intersex as their sex assigned at birth **Gender Identity (GI):** Participant identifies as anything other than man or woman (e.g., non-binary, transgender, etc.) **Sexual Orientation (SO):** Participant identifies as anything other than straight (e.g., gay, lesbian, bisexual, etc.)
Income	Participant with household income equal to or below two times the Federal Poverty Level
Educational Attainment	Participant is currently without a high school diploma or GED
Geography	Participant's zip code is within the list of Rural/Non-Metropolitan zip codes (per HRSA's Federal Office of Rural Health Policy)
Access to Care	Participant cannot readily use the health care system or pay for needed care
Disability	Participant has a physical and/or cognitive disability

Questions & Answers

Please submit all questions through the question box.

Summary & Evaluation

- Thank you to all our panelists.
- A recording of the webinar will be available on the CMSS website in the coming weeks.
- Please compete a short evaluation following the webinar.
- For more information, contact info@cmss.org.

CMSS

CMSS WEBINAR SERIES

Advancing Clinical Registries to Support Pandemic Treatment and Response

The series will address key questions related to the rapid development, deployment and implementation of Covid-19 focused clinical registries and clinical repositories by specialty societies and academia.

SUMMER 2020 | FREE TO ATTEND

Upcoming Webinars:

Aug. 12	Prioritizing Patient Engagement and Inclusion of Patient-generated Data
Sept. 1	Using Clinical Registries to Address Disparities in Covid-19

CMSS

CMSS WEBINAR SERIES

Advancing Clinical Registries to Support Pandemic Treatment and Response

Upcoming Webinar:

Prioritizing Patient Engagement and Inclusion of Patient-Generated Covid-19 Data

Aug. 12 | 1:00 - 2:30 pm ET

Moderator:

Susannah Fox
Principal, Internet Geologist LLC;
Former Chief Technology Officer,
US Department of Health and
Human Services

Host:

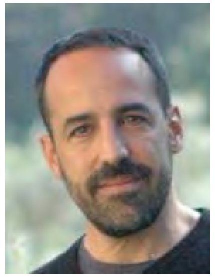
Helen Burstin, MD, MPH, MACP
Chief Executive Officer
Council of Medical Specialty
Societies (CMSS)

Panelists:

Gina Assaf
Independent digital design consultant, Leading the Patient-Led Research Group within Body Politic

Hannah Davis
Researcher with the Patient-Led Research Group within Body Politic

Emily Sirotich
COVID-19 Global Rheumatology Alliance; Canadian Arthritis Patient Alliance; Department of Health Research Methods, Evidence, and Impact, McMaster University



Gary Wolf
Co-Founder, Quantified Self