

UCI All of Us Research Program Researcher Newsletter

This week, the *All of Us* Research Program hit a new milestone. We released our first set of DNA information from participants into the *All of Us* Researcher Workbench. The data includes **nearly 100,000 whole genome sequences**, which involves a process that looks at participant DNA. Nearly half of this new genomic information comes from people who self-identify with a racial or ethnic minority group.

Our dataset will challenge what we know about health and disease. With the amount of data and diversity we have in our Researcher Workbench, there are new diseases waiting to be uncovered. There are genetic variants to identify that may better predict disease risk. And in the future, it's likely that *All of Us* data will contribute to the discovery of new treatments.

-Josh Denny, M.D., M.S., CEO of the *All of Us* Research Program

All of Us Research Program Releases First Genomic Data, Creating Unprecedented Platform for Health Discoveries

Precision medicine effort leverages diverse participant base and innovative technology to drive new research

The National Institutes of Health's *All of Us* Research Program has made its initial genomic data available, bringing nearly 100,000 whole genome sequences onto its data analysis platform, the Researcher Workbench. UCI *All of Us* has played a major role in this milestone, working to nurture relationships with diverse participants from throughout Orange County region to ensure strong diversity in the dataset.

In the past, many communities, including racial and ethnic minorities, those who live in rural areas, and LGBTQ+ people, have often been left out of biomedical research. As a result, we know less about the health of those historically underrepresented in medical research and ways to provide them with the best care, resulting in potentially less effective prevention and treatment strategies.

This lack of diversity has been especially evident in genomic studies, where more than 90% of participants of genome-wide studies come from European ancestry. The genomic data in the *All of Us* dataset represents an unprecedented level of diversity, including nearly 50% of participants self-identifying with minority racial and ethnic groups.



APPLY FOR DATA ACCESS

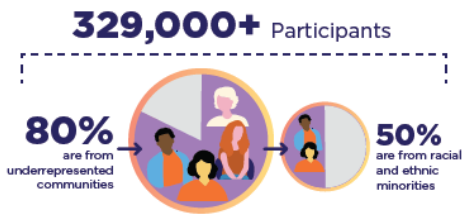
Join hundreds of researchers who are leveraging the *All of Us* data and tools to drive their work.

APPLY NOW

UCI Health

A Member of **All of Us California**

One of the largest, richest, most diverse biomedical data sets ever built



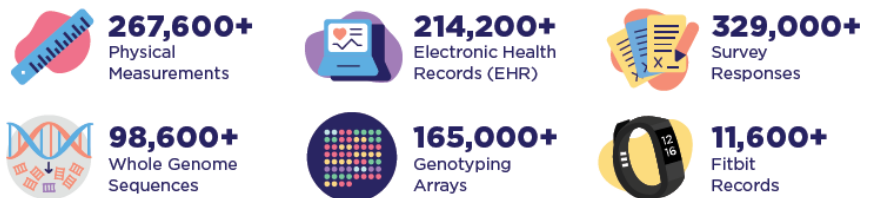
The *All of Us* Research Program is working to enroll one million or more participants across the U.S. to share data that can fuel new insights into human health. By engaging the participation of people and communities who have been left out of medical research in the past, and making their data available to a diverse cohort of researchers, the program seeks to accelerate health and medical research to enable individualized prevention, treatment, and care.

The genomic data from *All of Us* participants is available to approved researchers alongside clinical, lifestyle, and wearable data. This combination of data allows researchers to better understand how genes can cause or influence diseases in the context of other health determinants to ultimately enable more precise approaches to care for all populations, something that has not been possible in the past.

The program's [data browser](#) offers a summary view of genetic data from participants. Using this tool, anyone can search for specific genes or variants and see aggregate counts of their frequency in the *All of Us* dataset and the genetic ancestry of participants with each variant.

Detailed genomic data is available in the Researcher Workbench's Controlled Tier, a heightened access level that also includes more granular demographic data and additional electronic health record data. As with all data in the Workbench, no direct participant identifiers are included, to protect participant privacy. To get access, researchers must get sign-off from their institutions, complete training, and sign agreements for responsible data use. Once registered, researchers can use the data for a wide range of studies.

DATA AVAILABLE IN THE RESEARCHER WORKBENCH



Health is made up of more than just the DNA we inherit, or the care we receive. By bringing many different data types together with environmental and lifestyle data, at a scale and level of diversity not previously available, researchers can begin to see how all of these pieces of health fit together and better understand and address health inequities.

This progress could not have been possible without the generous contributions of *All of Us* participant partners. A benefit of being an *All of Us* participant is the choice to receive individual DNA results. Many have already received genetic ancestry and traits information, and in the coming year, health-related DNA results will begin to be made available to interested participants.

To learn more about the *All of Us* Researcher Workbench, or to register for access, go to www.ResearchAllofUs.org. To learn more about getting involved as a research participant, visit JoinAllofUs.org.

Transformative research potential

- Conduct **prospective, retrospective, and cross-sectional** analyses
- Develop improved **risk assessment and prevention strategies** to preempt disease
- **Investigate health disparities** and find new approaches to **improve** health equity
- Provide **earlier and more accurate diagnoses** to reduce illness burden
- Increase **wellness and resilience**, and promote healthy living
- Create a line of sight to enable new precision **treatments and interventions**

Researcher Workbench

Controlled Tier and Genomic Data

The *All of Us* Research Program is building one of the most robust databases to transform the future of health research with extensive health data generously contributed by our diverse cohort. Data is accessible with tiered access, including Public, Registered, and Controlled tiers.



Within the Controlled Tier

Genomic Data and Tools

Includes all of the data available in the Registered Tier as well as genomic data, additional clinical fields in EHRs, and more granular demographic data both from surveys and EHRs.



98,600+
Whole Genome Sequences



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



165,000+
Genotyping Arrays



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

More Individual-Level Information

- COVID-19 testing and diagnosis data
- Real dates of health events
- Residential location (first 3 digits of ZIP code)
- ICD codes



More Granular-Level Information

Non-generalized, demographic data provided by participants

- Race and ethnicity
- Sexual orientation
- Sex at birth
- Education
- Gender identity
- Employment status

Data From

329,000+ Participants



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

UCI Health

A Member of *All of Us* California

Follow us on social media:

