All of Us Research Program
American Society of Human Genetics–*All of Us* Educational Session
Meeting Summary
October 27, 2020

Welcome

Joshua Denny, M.D., M.S., began the meeting at 5:01 p.m. ET.

Opening Remarks

Dr. Denny welcomed attendees and provided an overview of the agenda and speakers. He opened the session with an overview of how large cohort studies have transformed disease treatment. He gave the example of the Framingham Heart Study, a multigenerational, longitudinal cohort study that analyzes familial patterns of cardiovascular disease and other diseases, leading to scientific discoveries related to disease prevention. The *All of Us* Research Program, as Dr. Denny described, aims to build a large, diverse, longitudinal cohort study to provide transformational advances in precision medicine over the next 40 years. Dr. Denny noted that a goal of the program is to improve diversity in genome science, as the lack of diversity in genomic research limits the understanding of health disparities, the discovery of new drugs, the accuracy of polygenic risk scores, and the classification of the pathogenicity of variants.

Dr. Denny asserted that the *All of Us* mission is to accelerate health research and enable individualized disease prevention, treatment, and health care. Participants enroll in the program, provide consent online, authorize the use of their electronic health record (EHR) data, and complete participant surveys on the web or via smartphone-based applications. The surveys encompass demographic, lifestyle, personal and family medical history, and health care access questions. A new COVID-19 Participant Experience (COPE) survey, launched by the program in summer 2020, encompasses questions about COVID-19–related symptoms and treatment, as well as social distancing and mental health factors, to understand the impact of COVID-19 based on individual experiences. In addition, the program collects blood, saliva, urine samples, and physical measurements during in-person appointments to generate a complete picture of human health across many conditions. The program is incorporating mobile and wearable technologies, starting with Fitbit and Apple HealthKit and later expanding to other applications and devices.

The program paused in-person appointments and events in March 2020 due to COVID-19. Clinics are beginning to re-open, with about 10 of approximately 350 clinics open. The program has enrolled more than 360,000 participants, has EHRs for about 230,000 participants, and has about 271,000 participants who have completed the initial steps of the program. The program has collected data from diverse populations, including sexual and gender minorities (SGM), rural participants, different races and ethnicities, and different socioeconomic statuses. More than 80% of the program population meets at least one of these criteria. The program plans to return genetic results to participants beginning in late 2021.

Part I: *All of Us* Research Program Overview

Engaging Underrepresented Research Participant Communities in Genomic Research

Consuelo H. Wilkins, M.D., M.S.C.I., discussed how the program has been engaging participants and communities who are underrepresented in biomedical research (UBR). This engagement is complex and, when looking at race, as Dr. Wilkins described, requires focusing

on the underpinnings of health disparities, which are primarily driven by social determinants of health and not genetics. Dr. Wilkins reviewed the program's engagement strategy, teams, partners, and funding recipients. She said that engagement is a bidirectional relationship built on trust, while recruitment has the goal of enrolling individuals in research. *All of Us* would like the engagement relationship to shape the program. Authentic engagement has not typically been done in genomic or precision medicine research at the scale that *All of Us* is attempting.

Dr. Wilkins said that race is a social identity, and it is one of the most imprecise variables in human research. In the U.S., race has been used primarily as a means of oppression. A key piece of the engagement strategy must acknowledge that there are valid reasons that individuals from UBR groups do not trust research, including the history of eugenics, the economic risks of health conditions, and the connection between DNA and the criminal justice system. She said that many who work in research and health fields are not as familiar with these historical atrocities and concerns. She said that the program must prove itself trustworthy and ensure these communities are not further marginalized. Dr. Wilkins said that she is proud of the work that *All of* Us has done to debiologize race. In the participant surveys launched by *All of Us,* participants are asked which racial and ethnic groups they identify with and can choose multiple options and provide more detailed ethnic categories if they wish.

Before the program launched, the program conducted more than 75 community engagement studios with UBR groups across the country, including SGM, racial and ethnic groups, and rural populations. Nearly half of those who participated in these groups identified as racial or ethnic minorities, and 10% identified as SGM. Later, the engagement core worked with many of these individuals and community organizations to review engagement approaches. They invited three individuals to serve on the *All of Us* Steering Committee. During Dr. Wilkins' presentation, she shared slides introducing the team members of the program's engagement core and the nearly 30 participant ambassadors. She said that the inclusion of diverse participants in the participant ambassador roster required a different approach; for example, the program had no requirements in terms of participant educational attainment or access to specific resources. Individuals who were interested in becoming participant ambassadors were required to answer a few questions about why they wanted to be involved in the program, and the engagement core conducted a blind review of applications. The program has tried to minimize barriers for participant ambassadors to attend in-person meetings and has provided travel assistance and other ways to make it easier for them to participate in programmatic activities. The participant ambassadors are part of multiple program governance boards and committees within *All of Us.*

Innovative Sequencing and Array Technologies at *All of Us*

Richard Gibbs, AC, Ph.D., discussed genome sequencing components of the program. As described by Dr. Gibbs, the program is conducting both array genotyping and whole genome sequencing on participant samples. The array genotypes will be run first; they are inexpensive and provide a predictable, though limited, dataset. Whole genome sequences are much more comprehensive and contain more complicated data that require deeper analyses. Together, the arrays and whole genome sequencing will complement each other. As the program works toward sequencing, program teams have been building components to address challenges of scaling, while maintaining cost structures and ensuring accuracy, as well as being mindful of security and compliance with Food and Drug Administration (FDA) requirements.

In 2021, the goal of the program is to produce 150,000 arrays and 100,000 whole genome sequences using participant samples. The arrays will be used to provide ancestry and trait data, and whole genome sequences will be used to look at hereditary disease risks and pharmacogenomics (PGx). During Dr. Gibbs' presentation, he shared slides introducing principal investigators and senior staff from the Broad Institute, Color, the Human Genome

Sequencing Center, the Center for Inherited Disease Research, and the University of Washington, as well as the working groups that have been building infrastructure, establishing standards, and dealing with compliance associated with genotyping and sequencing technologies.

The Illumina Infinium Global Diversity Array was designed with input from the consortium. It can process eight samples at a time and is relatively inexpensive; this device provides maximalist content and addresses diversity and clinical issues. Dr. Gibbs said that the array data are processed through fairly standard methodologies at the individual centers, and the data will be used for quality control and validation of whole genome sequencing, as well as variant calling.

The whole genome sequencing uses a standardized NovaSeq platform and conducts short-read sequences. Some of the analyses have been tuned to the consortium's needs. The process has been refined for the FDA Investigational Device Exemption (IDE). The process has taken into consideration DNA quality metrics, the coverage and call-rate minimums, genome references for research or clinical purposes, the potential for new references, and the methods for variant calling. Dr. Gibbs said that variant calling has been a controversial topic in the past, but now there is more emphasis on harmonized performance. As a result, the program chose the DRAGEN pipeline, which will be used at all participating *All of Us* Genome Centers. The participating centers include Baylor College of Medicine (Houston), with Johns Hopkins University (Baltimore) and the University of Texas Health Science Center at Houston (UTHealth); the Broad Institute (Cambridge, Massachusetts), with Color (Burlingame, California) and the Laboratory for Molecular Medicine at Partners HealthCare (Cambridge, Massachusetts); and Northwest Genomics Center at the University of Washington, Seattle. This pipeline involves programmable chip infrastructure, provides good harmonization, and can be run locally or in the cloud. The decision on single-variant (SV) calling is still under discussion. Alternative data types, such as long reads to better resolve SVs, may be considered. The program has enabled the HudsonAlpha Institute for Biotechnology to run a pilot study on long reads. So far, the pilot study has enabled the validation of the platform with different sample types and determined sensitivity and precision for three variant classes. Dr. Gibbs said that he suspected a hybrid model would be used in the future.

Dr. Gibbs said because of FDA's IDE approval process and COVID-19 prioritization efforts of the program, the whole genome sequencing work began in August 2020. Dr. Gibbs hoped to have the first 50,000 participant samples processed by the beginning of 2021. Returning genetic data to participants requires variant interpretation, which has required coordination among the Genome Centers conducting the sequencing. Dr. Gibbs said that the groups that are interpreting the sequences send their findings to a central site for processing; any nonconcordant results go before an adjudication committee. Dr. Gibbs also described the challenges of data security and the processes used to ensure the security of this sensitive data.

The program has designed, built, and deployed new arrays, and short-read whole genome sequencing is underway. FDA cleared genetic health data return to participants. The program is evaluating novel long-read genome sequencing and has achieved variant interpretation and harmonization at baseline.

Returning Genetic Results at Population Scale

Alicia Zhou, Ph.D., said that the program wants to communicate genetic results to interested participants, in particular, genetic variants that can affect disease risk. The program has the goal of engaging the population longitudinally and is thinking about how to retain participants and drive long-term participation. Providing genetic results is one strategy toward this goal.

The genotyping array data will enable the return of nonclinical genetic traits and ancestry to research participants. Whole genome sequencing will provide health-related information from 59 genes and findings from 7 PGx genes. The nonclinical traits will be launched for participants to view in the *All of Us* Participant Portal in late 2020. In order for participants to view their results, they will be required to complete a set of informing loops that collect consent for the participant to opt in to see their research results. Later in the presentation, Dr. Zhou said that there are informing loops for each result type (i.e., ancestry, traits, health-related, or PGx); she said that participants can opt in to receive each type of result. In the Participant Portal, the participant will see a series of screens that discuss the risks, benefits, and limitations of these results. The participant then selects whether they would like to see their results; selecting "yes" takes the participant to a dashboard of ancestry and trait results. Dr. Zhou said that there will be information about concepts around genetics and ancestry and possible sensitivities around receiving this type of information. Dr. Zhou said the program plans to start with returning four traits and to add new traits to these results intermittently.

Dr. Zhou said that the format for returning health-related genetic results and PGx results is still under development. She showed some concepts for communicating these results, including an example of a *BRCA1*-positive result. The results would include next steps for participants and information about their disease risks based on National Comprehensive Cancer Network or other society guidelines. The communications would clarify that participants need to confirm these results with a health care provider. Dr. Zhou's team worked on appropriate language with the *All of Us* Institutional Review Board (IRB) and FDA to address this. From the Participant Portal, a participant can choose to schedule an appointment with a genetic counselor, download a PDF of their report, and have the program share the results with a health care provider of the participant's choosing.

The program has been considering the ethical and regulatory implications of the return of genetic results. The IRB and FDA both influence how the program returns health-related results. The genetic sequencing is performed at the Genome Centers following CLIA/CAP (Clinical Laboratory Improvement Amendments of 1988/College of American Pathologists) standards, and the interpretation and secondary confirmation is performed at a CLIA-certified lab. Although there are many steps and facilities involved in this process, it will generate a single report of research results for participants. As the results derive from research and not clinical testing, there is no supervising physician and no clinical test requisition. The program has a Genetic Counseling Resource (GCR), which includes a network of genetic counseling professionals who can provide computer and phone counseling services.

All participants will be notified when their research results are ready, and they will receive an online refresher on the types of results that may have been found. Individuals who did not have an actionable variant will be able to see their results without talking to a genetic counselor first. Participants with actionable results will receive an invitation to schedule an appointment with a genetic counselor with the GCR; their results will be released to them during the appointment. The counselor will connect participants with local resources for follow-up care. The GCR aims to provide both technological infrastructure and participant support, as well as educating the *All of Us* clinical network so that providers are prepared to discuss results with participants.

PGx Framework and Approach for PGx Return of Results

Philip Empey, Pharm.D., Ph.D., discussed why the program is focused on PGx. Dr. Empey said that the program believes that a large portion of patients will have a finding that may have lifelong clinical value for participants as well as scientific value. In early discussions with participants, participants were very interested in learning how genetics affected medication, so this became a priority for the program. One unique aspect of this effort is that the program is

looking for specific, prevalidated targets. FDA approved an IDE for this research study, a process that took 18 months.

The program chose a subset of seven genes for the initial return of PGx data in 2018. The selection process focused on value to participants and actionability. The program selected alleles with known functional consequences, based on national guidelines, under the auspices of clinical guidance, and based on marketplace data. This list of genes will be expanded based on the program's experience and the emergence of new PGx data. Dr. Empey provided a list of the selected genes and their drug associations.

Each Genome Center that is part of the project has already been CLIA validated, but each needed to achieve FDA standards. Dr. Empey presented data that showed the centers' strong performance and very high standards of producing accurate results. Interpretation followed standard pipelines and interpretation tables; this was a key discussion with FDA and experts in the field.

Dr. Empey said that the participant return of results report design was critical so that participants have a comprehensive understanding that their PGx results were investigational. Participants should continue to take their medications and discuss their PGx results with a health care provider and pharmacist. Dr. Empey emphasizes that genetic information is just one aspect of health. Each report provides a list of relevant drugs for the particular PGx finding. The report includes a statement advising that a participant check with a health care provider to determine whether ordering a clinical PGx test is appropriate. The reports have been tested for comprehension with a non–*All of Us* population. The consent process includes a short video that provides education and support; additional educational tools will be built in the future. Dr. Empey reported that FDA approved the IDE for PGx return of results in July 2020 and stated that the program has planned content for new guidelines, has planned PGx targets, and is identifying new controls and adding additional targets.

Panel and Q&A

Karyn Onyeneho, M.S., thanked the panelists and started the first question-and-answer (Q&A) portion of the session. Ms. Onyeneho reminded attendees to stay tuned to Part II of the session following Q&A, which comprises a demonstration of the *All of Us* Researcher Workbench led by Andrea Ramirez, M.D., M.S. Ms. Onyeneho read questions that had been asked prior to the session, based on questions submitted by registered attendees, and afterward addressed questions presented to panelists during the session. The panelists responded to questions as follows:

- Dr. Denny said that the program surveys cover basic health habits and do not yet cover complementary therapies and activities such as the Sound Health Initiative. There is a program subcommittee that is considering future survey types. The next survey that will be available will look at social determinants of health. Dr. Denny said that this is an important element to complement the data the program is already collecting.
- Dr. Wilkins said that the program has used a number of different engagement strategies, starting with the community engagement studios. The program also engaged community organizations that already had a presence in UBR communities. She said that an important piece of building trust is working through trusted entities. Some of the barriers to engagement relate to historical research abuses, which the program needs to continue to address. Dr. Wilkins said that research team members have little experience engaging with communities, so the program is also working to prepare them for engagement.
- Dr. Zhao said that the Participant Portal has a module that allows individuals to share results with their health care providers by entering their providers' contact information. Participants who have actionable results will receive those results while on the phone with a genetic

counselor; much of the call will be dedicated to determining a participant's next steps. Dr. Zhao said that not all providers are familiar with genetic results, so the program is working on educational resources that could help educate providers, as well.
- Dr. Gibbs said that the program is performing polymerase chain reaction–free whole genome sequencing.
- Dr. Gibbs said that one positive aspect of the program's pipeline is its flexibility to adapt to new software or new genome references.
- Dr. Empey said that the *All of Us* data repository is excellent and can be linked with genetic tools on the research side. On the clinical side, there are certain requirements as to what can be placed on the report. The participant report can provide links to other, tangential resources, such as educational links and programs.

Part II: Interactive Panel Discussion

Researcher Workbench Demonstration and Developing Tools for Genomic Analyses

Dr. Ramirez said that the website and tool that *All of Us* has developed to deliver program data is called the Research Hub (available at ResearchAllofUs.org). All participant-collected data is stored in a cloud format; some of these data are available to the public in the *All of Us* public Data Browser. More extensive datasets are available in the *All of Us* Researcher Workbench for authorized researchers, who would have access to Registered Tier data, workspaces, the Cohort Builder tool, interactive Jupyter Notebooks, and more. Researchers must go through an approval process overseen by the program to become authorized to use more extensive datasets. Dr. Ramirez said that, traditionally, research programs have brought data to researchers, which discourages shared research workspaces. The cloud-based data of *All of Us* allows researchers to come to the data, which facilitates collaboration, centralizes data security, and increases access while lowering costs.

Dr. Ramirez provided an overview of the Research Hub. The data curation pipeline brings in participant data from multiple sources and is then harmonized on the phenotype side. The DRC makes refinements to the data and then presents the data either into the public Data Browser or the Researcher Workbench. In the public Data Browser, anyone who uses this tool can look at data snapshots to learn more about participants without creating an account. Users can view aggregate, deidentified data, which are rounded to groups of 20. Dr. Ramirez provided an example of using the Data Browser by searching for "diabetes" as a keyword in the "search across data types" field; from there, she displayed real-time updates of relevant EHR domains and survey questions that are publicly available. Dr. Ramirez also demonstrated how this tool can be used to explore gender and age distributions. Researchers can access the public data to analyze the types of data available and determine whether they want to apply for authorized access to the Researcher Workbench Registered Tier for more extensive exploration and analyses.

Access to the Researcher Workbench Registered Tier requires an individual researcher to be approved by the program based on identity verification, completion of a signed Data Use and Registration Agreement (at the institutional level), and additional requirements. Dr. Ramirez said that 217 institutions have started the institutional agreement process, and 169 have completed it, which averaged 24 days to complete. There is a form available for researchers to request that their institutions start the application process. Institutions would be required to have an electronic Research Administration (eRA) Commons account administered by the NIH Office of Extramural Research and complete the *All of Us* responsible conduct of research training (which include a signed data user code of conduct). Onboarding for authorized researchers takes about two hours.

Dr. Ramirez said that the data currently available to researchers in the Researcher Workbench include EHRs, participant surveys, and physical measurements. More than 126,000 participants currently have all three of these data types available in the workbench. An article describing diversity of the *All of Us* dataset, "The *All of Us* Research Program: Data Quality, Utility, and Diversity," was published with MedRxiv in June 2020 (Ramirez et al., 2020).[1] As described in the paper, approximately 77% of participants meet at least one category of diversity in research; the paper also looked for known associations in the dataset. For example, the researchers found associations with smoking in the EHR and in surveys in which participants reported ever smoking; these participants had increased risk effects. Dr. Ramirez said that all of the code for this work, and for other demonstration projects, is available in the Researcher Workbench.

The *All of Us* Research Program launched the Researcher Workbench in May 2020. The program plans to refresh the data in winter 2020–2021. The data refresh will involve increased amounts of data from participant surveys, EHRs, physical measurements, the new COPE survey, and Fitbit. In 2021, the program will add another access tier to the Researcher Workbench, the Controlled Tier, which will include genomic data and other data types.

Dr. Ramirez provided a demonstration of the Researcher Workbench Registered Tier (available to authorized researchers). The landing page provided several quick video tours and an overview of workbench features. One central aspect is the workspace, where collaboration between researchers and teams can take place. To create a workspace, registered researchers must provide a description of their research project, which the public can view through the Research Hub. Once created, these workspaces are enabled with cohort builder tools via a Jupyter Notebook platform. *All of Us* offers authorized users with support services, including help desk support. The workbench also has a phenotype library, and researchers can use prepopulated phenotypes to build a cohort. Dr. Ramirez invited all who were interested in becoming beta researchers to an *All of Us* workshop on November 12; the workshop will include an in-depth demonstration of the workbench and open the discussion to current workbench users to discuss their experiences with the workbench. Registration for this event is on a first-come, first-served basis. Those interested should email Adrienne Roman, Ph.D. (adrienne.s.roman@vumc.org).

Panel and Q&A

Ms. Onyeneho thanked Dr. Ramirez and started the second Q&A portion of the session. The panelists responded to questions as follows:

- Dr. Ramirez said that the program has looked into integrating external tools, such as Blockchain, into the infrastructure of the Researcher Workbench, although they are not enabled right now. She and the workbench team hope to enable researchers with new tools in the future.
- Dr. Ramirez said that researchers will not be able to access individual-level research results through the Researcher Workbench, as research results are available in the Researcher Workbench in aggregate.
- Dr. Gibbs said that, in terms of the data model, there is conformity within the established standards as far as file types. He and Dr. Zhou said that the program will be using the human reference genome GRCH38 and will conform with clinical validation standards; they will consider new references moving forward.

[1] Ramirez, A., Sulieman, L., Schlueter, D., Halvorson, A., Qian, J., Ratsimbazafy, F., … On behalf of the *All of Us* Research Program. (2020, June 3). The *All of Us* Research Program: Data quality, utility, and diversity. Retrieved October 30, 2020, from https://www.medrxiv.org/content/10.1101/2020.05.29.20116905v1

- Dr. Empey said that, because this is a research study, the program is not going to bring in EHR data alongside PGx for a clinical report or recommendation. He said that there will be opportunities to bring PGx data to the Researcher Workbench and perform analyses. On the clinical side, PGx results can trigger conversations between participants and health care providers.
- Dr. Zhou said that the return of genetic ancestry results currently map to 21 subregions. As finer data become available, the program may update this module.
- Dr. Denny said that the Nutrition for Precision Health project will be the program's first ancillary study, and there is excitement about using *All of Us* data to research this problem, which is foundational to other diseases. The program is currently developing a diet module that will be available in the future as well.
- Dr. Ramirez said that the program is putting effort into making sure that the research community, as well as the participant population, is diverse. Researchers can review demonstration projects, papers, and the code in the workbench to see examples of how to use the variables. There is also support available through the help desk, as well as hosted events and office hours, to review research approaches and technical aspects of the workbench. Dr. Ramirez said that the program is looking to the new social determinants of health module, which will include additional variables that can generate more diverse research.

Closing Remarks

Ms. Onyeneho thanked the panelists and the attendees and said that the program plans to answer all questions that were received during this presentation post-session and make the questions and answers available publicly. Dr. Denny appreciated the interest in and questions about this session and encouraged all to visit both the JoinAllofUs.org site and the ResearchAllofUs.org site for more information about the *All of Us* Research Program.

Adjournment

Dr. Denny adjourned the meeting at 7:01 p.m.