



All of Us Program Plans to Return Disease Variants, PGx Results, Primary Genomic Data

Jun 05, 2018 | [Julia Karow](#)

NEW YORK (GenomeWeb) – The National Institutes of Health's All of Us Research Program plans to initially return disease-associated variants in dozens of genes and pharmacogenomic results to participants, according to project organizers. In addition, study subjects will have the option to obtain their primary data files. The plan is to genotype participants first and to conduct whole-genome sequencing on all of them over the course of the program.

In early May, All of Us, which plans to enroll at least 1 million people in the US and conduct longitudinal health-related research on them based on a variety of data types, started [enrolling participants nationally](#).

Later that month, the NIH put out a [funding announcement](#) for genome centers that will generate data for the project. In a webinar last week to provide additional information for prospective applicants, NIH representatives discussed details of how data for the project will be generated, and what types of results participants can expect to obtain. "The ultimate goal is to create one of the world's largest and most comprehensive research platforms for precision medicine," said Brad Ozenberger, program director for the All of Us Data and Research Center.

The All of Us genome center, or centers — the program intends to make either one or two awards — will serve two different users: researchers, who will study the data, and participants, who will get access to certain results. Applications are due July 12, will be reviewed in early August, and awards will be made by the end of September, Ozenberger said. The goal is for the centers to begin generating data in late 2018, and to start returning results to participants in 2019.

The genome centers will receive DNA from the All of Us biobank at the Mayo Clinic, which will isolate DNA from participants' blood samples and, in rare cases, saliva samples. The goal is to recruit 70 to 75 percent of the 1 million participants from groups that are currently underrepresented in biomedical research.

The genome centers will generate both genotyping and whole-genome sequencing data, call variants, annotate variants in certain predefined regions, and upload the data to the Data and Research Center.

In addition, one of the genome centers will operate a Clinical Validation Laboratory (CVL), which will run validated clinical tests, such as Sanger sequencing or Taqman assays, to confirm pathogenic or likely pathogenic variants in 59 genes recommended for return of results by the American College of Medical Genetics and Genomics (ACMG). The program might add other genes to the ACMG-59 list later on, Ozenberger added.

The CVL, which will not have access to any phenotypic data of the participants, will then generate a clinical report, which it will forward to a Genetic Counseling Resource that the All of Us program plans to establish before the end of the year. Based on the known frequency

of pathogenic or likely pathogenic variants in ACMG-59 genes in the general population, the lab is expected to validate about 3,000 variants in the first year and about 6,000 per year after that.

Participants with a positive result in one of the ACMG-59 genes will be assigned a genetic counselor, who will communicate the result to them. NIH's vision for the Genetic Counseling Resource is that it will not only cater to participants with positive findings but also offer a call service to participants or their healthcare providers who have questions, either electronically or by phone. More information on what the resource might look like and what its objectives are will be coming soon, Ozenberger said.

Study subjects will also obtain a genome report from the Data and Research Center — generated in close collaboration with the genome centers — that will contain pharmacogenomics results with the highest level of evidence, traits, "and other genetic information that is still to be determined," he said.

Finally, participants will have the opportunity to obtain their primary data, though it has not been determined yet whether that will be a CRAM file or BAM file, which contains aligned sequence reads, or a VCF variant file.

What type of raw data is returned depends on whether the program needs to apply for an investigational device exemption (IDE) with the US Food and Drug Administration for its whole-genome sequencing assay, Ozenberger explained. "We believe this program likely will be considered high risk – we're predominantly evaluating healthy people – so it is likely we will have to file with the FDA for an [IDE]," he said. For such an application, the program would draw on the expertise of the genome centers. Without an IDE, All of Us would return BAM or CRAM files, and with an IDE, it would probably return VCF files, which are considered interpreted data files.

The current plan, which genome center applicants should budget for, is to start with 100,000 genotyping assays in the first year — the approximate number of participants the program plans to enroll during that year — and 10,000 whole genomes. In years two to five, the number of genotyping assays will double to 200,000 per year. Meantime, the number of WGS assays will rise to somewhere between 25,000 and 100,000 in year two, and between 50,000 and 200,000 in each of years three to five.

Applicants can ask for up to \$15 million in direct costs for the first year and need to propose their own budget for the later years, using three different scenarios for whole-genome sequencing throughput. They can also ask for a small amount of funding for innovation research.

"We expect that the primary data will really come from the genotyping platform in years one and two, maybe beyond that," Ozenberger said. However, as whole-genome sequencing is fully deployed, the project might discontinue genotyping at some point, he added.

The ultimate goal is to sequence all 1 million participants, though it is unclear whether that will happen over the first five years of the program. All of Us may prioritize certain groups for whole-genome sequencing initially, he said.

Asked by a webinar participant why the program does not focus exclusively on whole-genome sequencing from the get-go, he said that not many laboratories are currently able to run clinical-grade whole-genome sequencing assays, and that clinical validation still requires genotyping arrays. Also, he said, the program is unsure of the exact amount of funding it will have available over the next few years, and since WGS is more expensive than genotyping, it wants to remain flexible.

"So, we are moving rapidly on genotyping, realizing that down the road, the ultimate genomic data type is WGS, and we may shift entirely [to that]," he said. "But we think there is at least several years of great value in the genotyping platform for All of Us."

Ozenberger also provided information on additional requirements for the genome centers. Their scale may be larger than that of any existing center in the US today, he said, so one criterion will be existing expertise with large-scale genomics.

All sequencing for the program needs to be carried out in a CLIA-certified laboratory, and all labs need to be physically located in the US. However, labs do not need to be CLIA-certified at the time of application, only when data generation starts, and not all assays need to be "fully approved for the return of specific information by the medical director at the local site," he said, as long as they are conducted in a CLIA lab environment.

In addition, labs do not need to be certified by the New York State Department of Health initially, but are expected to work towards New York state certification over the course of the award. "We want the data to be compliant across all 50 states," Ozenberger said.

The program does not specify what genotyping array or sequencing technology labs should use, what coverage they need to provide for whole-genome sequencing, and what read alignment and variant calling approaches they should use. Rather, applicants should propose specific technologies and metrics, keeping in mind the goals of the program.

However, Ozenberger said, the project does not plan to use custom arrays, which take a long time to design and validate, but would like genome centers to go with one of the commercial arrays that are already widely used in research and add some custom content to cover clinically important variants and different ethnic groups. Also, if two genome centers win awards, they will need to agree on a single genotyping platform.

In terms of WGS, given the scale of the project and the available funding, All of Us believes that "the only technologies that are capable of achieving that scale at this cost are the short-read instruments," Ozenberger said. Nevertheless, the project is interested in generating haplotype and structural variant information that short-read sequencing may not be able to provide, so it may add long-read sequencing if sufficient funding becomes available, he said.

In addition, not all assays within a genome center need to be run by a single laboratory or even by a single organization. "We do encourage consortia," Ozenberger said.

Centers will be required to retain the data they generate for months but not years, and details about security and privacy requirements for data storage will become available after the awards have been made.

Genome centers should also be prepared to generate other types of omics data, for example, RNA-seq or microbiome data. "All of Us has aspirations to extend omics research data generation beyond the genome," Ozenberger said, if budget permits. Just in case, the program is already collecting participants' blood in a way that allows for the analysis of RNA and cell-free DNA. Additional omics data may be collected for part but not all of the cohort, he added.

Finally, genome centers might need to reanalyze data from participants later on using updated methods, or might need to generate additional data for parts of the All of Us cohort. "All that is a possibility," he said. "This is not a contract. We can't predict exactly where we will be two, three years from now."

"All of Us is expected to be around for 10 years," Ozenberger said. "We hope to be done with the sequencing long before that, but there could very well be opportunity for new activity by the genome centers in later years."