All of Us Research Program: Considerations Toward a Comprehensive Genomics Strategy (NIH Final Report)

The Policy

Synopsis

Considerations Toward a Comprehensive Genomics Strategy [10] is the final report of the Genomics Working Group (GWG) [11] of the All of Us Research Program Advisory Panel housed under the National Institutes of Health (NIH). In the report, the GWG evaluated strategies and provided recommendations to implement a large-scale strategy for the All of Us program, an integral part of the NIH's Precision Medicine Initiative [12], to collect and analyze genetic information from participants. This program aims to build a database of health-related information from a diverse set of one million or more individuals from...
across the nation with the ultimate goal of accelerating health research and precision medicine.

The database will include information about the lifestyle, environment, and genomes of each participant. As such, the GWG reviewed and suggested strategies for analyzing and evaluating genetic data (specifically, DNA sequences) of the participants. Importantly, one tenant of the All of Us program is to include participants “who reflect the diversity of the United States.” Historically, genetic testing has excluded [13] large population groups (e.g. Africans, Latin Americans, and native or indigenous people) from analyses which has led to researchers potentially overlooking the role of some genetic factors in different populations.

In their report, the GWG presented the pros and cons of the available options in processing genomic data fitted to the specific goals of the All of Us program. Ultimately, the report asserted the need for a pilot study to optimize the strategy for producing genomic data prior to scaling up to the full cohort of one million participants.

The GWG addressed four primary questions in the report:

1. What strategies exist for detecting variation in the human genome?

The GWG considered the three main technologies used to generate genomic data: whole-genome genotyping [14] (WGG), whole exome sequencing [15] (WES), and whole genome sequencing [16] (WGS). WGG identifies common variations in DNA sequence based on current knowledge, but it cannot identify rare variants in the DNA. However, WGG can be supplemented with custom variant content if there are specific variants of interest to the All of Us study. WES can identify both common and rare sequence variants, but only for DNA sequences that code for protein, which only account for only ~1% of the genome. Finally, WGS is the most information-rich technique as it can identify both common and rare genomic variants across the complete (or nearly complete) DNA sequence including both protein coding and non-coding regions.

In addition to information content, the GWG also considered cost when evaluating genomic techniques. WGG is the least expensive technique by a large margin, ranging from $30-100 per participant. WES and WGS are more expensive, ranging from $350-1,000 and $1,000-2,000 per participant, respectively. Although WGS is currently more expensive, the GWG expects that the cost of WGS will decrease over the next few years. Based on this evaluation, the GWG concludes that WGG and WGS are the most valuable techniques to test in the pilot study for All of Us.

2. What are the options for designing a custom genotyping array?

Although WGG can struggle to identify rare DNA sequence variants, it still has many advantages over other methods. These advantages include a lower cost, faster turnaround time, and lower informatics requirements. Additionally, there are WGG platforms that are optimized for multiethnic populations, which could be used for the diverse set of participants expected for All of Us. It is also possible to design custom content, but these add additional costs and increase the timeframe. The GWG recommends using existing platforms for a pilot study while noting that customization may be useful for the full
3. What factors need to be considered for returning results to patients?

One of the commitments of the *All of Us* program is to return results to patients when clinically actionable DNA variations are identified. To determine which variants are clinically actionable, the American College of Medical Genetics and Genomics [17] (ACMG) has published recommendations [18] based on information from ClinVar [19], an NIH-funded public archive of DNA variants with reported phenotypes. Prior studies found a 1-3% frequency of identifying variants in a set of 59 genes [18] determined to be clinically actionable by the ACMG. The ACMG recommendations only include variations that are known to be pathogenic, or disease causing. However, there are also variants designated as “likely pathogenic”, which the *All of Us* program may consider returning to participants, with the trade-off of potentially alerting participants to variants that may not turn out to be harmful.

Another consideration is whether to return only certain variants, such as those designated as most actionable by the ACMG, or variants that are fully reviewed and confirmed by ClinVar. Only returning fully reviewed variants would decrease the burden of manual variant review but could result in fewer returned results, especially for minority populations whose genetic profile of diseases may be less characterized.

Additionally, the working group considered whether to use a Clinical Laboratory Improvement Amendments (CLIA [20]) platform. CLIA-certified labs adhere to federal regulatory standards that allow them to return patient results for the purpose of health assessment. Using a CLIA platform would expedite the return process for *All of Us* and potentially avoid costly validations. Finally, validation approaches are well established for the sequencing techniques used in WES and WGS, but the accuracy of the array technology in WGG is not well established, especially for rare variants. Performing independent validation of the arrays will require additional time and money.

**Context**

In 2015, President Obama announced the Precision Medicine Initiative with the ultimate goal of improving healthcare treatments for everyone by enabling researchers and medical professionals to better understand health on an individual level and to tailor treatments and preventions specifically for the patients. The mission of the *All of Us* research program is to collect data on the lifestyle, environment, and health of one million or more Americans to produce the largest and most diverse health resource in history. Congress has approved a budget of $1.5 billion over ten years for the *All of Us* program.

**The Science**

**Science Synopsis**

Health depends on many factors spanning across biology, environment, and lifestyle that are specific to each individual person. In an ideal world, healthcare would also be individualized for each person, rather than the one-size-fits-all approach we currently use. The practice of tailoring medical treatment to an individual is called *precision medicine* [21]. Although still in its infancy, precision medicine has been
incorporated into some treatments. For example, some cancer patients undergo genetic testing [22] to
determine the exact cause of their cancer so that their treatment may be personalized.

An improved understanding of how genetics is linked to health will directly improve the ability to provide
individualized preventative measures and treatments to patients. This is because many diseases or pre-
dispositions are governed by variations in our genetic code, or DNA sequence. The variations in our DNA
can be small, such as a single-letter change in the code, called a single nucleotide polymorphism [23] (SNP),
or they can be larger changes such as insertions, deletions, rearrangements, or duplications of code, called
structural variants [24] (SVs). There are both common and rare occurrences of SNPs and SVs, with rare
mutations more difficult to identify with genetic testing.

Since the DNA code is the template for making RNA and proteins that perform specific functions in the cell,
variations in the DNA can affect how the downstream products function. This can lead to specific
phenotypes, including disease. Variations in the DNA can be passed down from parents, or they can arise
through errors in the process of genome copying. Finding links between specific DNA variations and
diseases can help researchers better understand disease processes, which can in turn lead to better
preventative medicine, drug design, and treatment strategies.

There are a variety of technologies and techniques that exist to obtain genetic data from biological
samples. In studies intended to describe the genetic contributors to disease, it is critical that researchers
utilize the technique to generate the most useful genetic data. The three techniques for generating genetic
data considered in this report are whole genome genotyping, whole exome sequencing, and whole genome
sequencing.

Whole genome genotyping is a technique that analyzes an individual’s DNA to return information about
the genetic variants in the genome. Unlike sequencing techniques, whole genome genotyping does not
generate a DNA sequence but rather targets SNPs, or areas of DNA which are known to differ between
individuals. Whole genome genotyping analyzes both coding [25] and non-coding [26] regions of an
individual’s DNA. This technique is the least expensive option with an estimated cost of $30 to $100 per
sample. However, the data generated from whole genome genotyping can only give researchers
information about common, not rare, SNPs. These data can also partially identify small, common SVs and
large SVs but cannot identify small, rare SVs.

Whole exome sequencing produces a dataset with the DNA sequence of all coding regions of the genome,
otherwise known as the exome [27]. This technique costs between $350 and $1,000 for each sample
analyzed. Data produced from whole exome sequencing can be used to identify both common and rare,
coding SNPs. However, as whole exome sequencing does not analyze non-coding portions of the genome,
these data cannot be used to identify rare, non-coding SNPs. This technique can also be used to partially
identify small, common SVs and large SVs but cannot identify small, rare SVs.

Whole genome sequencing is the most expensive of the three techniques with an estimated cost of $1,000
to $2,000 per sample. Unlike whole genome genotyping and whole exome sequencing, whole genome
sequencing generates the DNA sequence of an individual’s entire genome. It therefore provides the most
information to the researcher, including identifying both common and rare (coding and non-coding) SNPs, large SVs, and partial identification of both common and rare small SVs.

**Relevant Experts**

**Lori A. Orlando, MD** [28], is an Associate Professor of Medicine and an Associate Director of the Precision Medicine Program in the Duke Center for Applied Genomics and Precision Medicine. Her work is currently focused on incorporating patient decision-making behaviors into clinical practices.

- “The *All of Us* study will change the face of clinical care in the U.S. By understanding the interaction of genetics, lifestyle, and environment in a large, and importantly diverse, group of people from across the country we will finally begin to understand how to personalize the care of our patients.”

**The Debate**

**Endorsements & Opposition**

**Endorsements**

- Dr. Francis Collins, *Interview* [29], May 1, 2017. “The scientific opportunities with a research platform of this scale, diversity, and data are enormous.”
- Dr. Sekar Kathiresan, *New York Times Article*[30], March 19, 2018. “I think what the U.S. project adds is that it reflects the diversity of the U.S.”

**Opposition**

- Dr. George D. Yancopoulos, *New York Times Article*[30], March 19, 2018. “I think someone needs to ask tough questions about whether this is the best use of precious N.I.H. resources [...] Should the funding instead go to individual researchers who are doing truly basic and innovative science?”

**Status**

The *All of Us* Research Program’s Genomic Plan was released on December 21, 2017. Enrollment for the program opened on May 6, 2018 with community events focused on expanding the number of participants. Since then, the Department of Health and Human Services awarded $21 million to health centers involved in the *All of Us* Research Program on September 11, 2018.

**Recommended Citation**


**License Info**

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0). Please distribute widely but give credit to Duke SciPol, linking back to this