All of Us Research Program
Genome Centers Funding Announcement (FA) Webinar
Questions & Answers (Q&A)
May 31, 2018

Program/Procedures

Question: How can we find out more about the new project and how participants can be involved with the whole genome sequencing (WGS)?

Answer (Dr. Ozenberger): You can find out more information about the program through joinallofus.org.

Question: What is the desired deliverable from the Genome Centers?

Answer (Dr. Ozenberger): Genome Centers awarded under this FA will be responsible for (1) developing the capacity and rigor necessary to achieve unprecedented scale of genomic data generation, up to 200,000 assays per year; (2) generating high-quality genotype and genome sequence data, with state-of-the-science variant calling, as a crucial data element for the *All of Us* data resource; (3) providing limited variant pathogenicity interpretation as an initial step in workflows for return of information to participants in the program; (4) deploying a fully compliant clinical validation workflow, including generation of a clinical report for certain pathogenic variants defined by the program; (5) innovating for the advancement of technologies and approaches for population-scale genome analysis; (6) establishing strong collaborative relationships with other awardees supporting the *All of Us* genomics platform; and (7) contributing to strategic planning for the program as members of the *All of Us* consortium.

Question: What is the expected data retention period for the *All of Us* Research Program consortium data generated?

Answer (Dr. Ozenberger): That is to be determined. It will be months, not years. This is a very important question in dealing with very large data volumes. There will be an early coordination with the Data and Research Center (DRC) on how to manage these large data and how long the data will be retained by the Genome Centers. A final decision will be negotiated post-award.

Question: What factors are being used to determine the genetic counseling resource? Have you consulted with NSGC for input regarding the genetic counseling resource?

Answer (Dr. Ozenberger): Yes. Other organizations interested in genetic counseling have also been consulted. The return of genetic information to participants is not trivial, especially outside their health care providers. The genetic counseling resource will not only take each case of ACMG pathogenic variant observation but also provide a call service. Participants and health care providers will be able to contact *All of Us* Research Program, either electronically or via phone (to be determined), and ask an *All of Us* genetic counselor about results. There will be more information forthcoming on what the resource will look like and what the objectives are. The program will post a notice of intent to publish a funding announcement for a genetic counseling resource later this year.

Question: Are there any online forums for finding consortia partners, such as SEQanswers or others?

Answer (Dr. Ozenberger): None are supported by the program.

Question: Do you have some preliminary information about the "diversity" of the sample, with respect to ethnic groups sampled, relative to the proportion of the U.S. population of each ethnicity?

Answer (Dr. Rutter): An emphasis of the *All of Us* Research Program is to enroll participants who have been previously underrepresented in biomedical research. Our goal—and some of the numbers we have are trended towards—is a relatively high number of individuals who were underrepresented in biomedical research. Our goal ultimately is to have 70% to 75% of the 1 million participants be in the underrepresented in biomedical research category; however, this may be reevaluated in the future. Therefore, there would be an interest in applicants to show a way to capture genetic information based on diversity, and in terms of genotyping arrays, this would be one key element that we would be looking for.

Question: Is it appropriate to suggest an alternative orthologous method for variant confirmation (i.e., not Sanger), or is Sanger required?

Answer (Dr. Ozenberger): Any approved clinical assay can be proposed. It could be targeted sequencing as the most common choice. For some variants, there may be a better alternative, such as TaqMan™. Applicants should propose a sequencing pipeline to give autologous information about variants discovered on the genotyping array. In Year 1, there will be rapid genotyping and discovery of ACMG pathogenic variants, and this should be directed quickly to the clinical validation laboratory. For the technical descriptions beyond Year 1, there might be a case to be made for using sequencing to validate genotyping and vice versa.

Question: Are there turnaround expectations for any component, starting with the date of sample receipt?

Answer (Dr. Ozenberger): Yes. There will be performance milestones that could include turnaround times. It is not specified in the FA. As part of the awards, specific performance and production milestones will be established.

Question: Will researchers inside the Genome Centers have access to the preliminary data for development of alternative analysis methodologies?

Answer (Dr. Ozenberger): The answer is no, for now. A case could be made for later. As described in the FA, research is not allowed on data that are sitting transiently within the Genome Centers. Genome Center investigators would be required to access the data research hub for the *All of Us* Research Program. However, cases could be considered individually to determine whether earlier access to the data would be required. For now, there is no privileged access to data by Genome Center investigators.

Reinterpretation: Awardees may require gender, ethnicity, or other information as part of the analysis workflow, for sample identity checking or for selecting an optimized reference genome. Such requirements for participant data elements can be included in applications.

Application Process

Question: There is a statement in the FA that states that it is Step 2 but mentions a date to submit possibly before Step 1, which isn't due until July 12. Should this read "after July 9," or is it supposed to be July 12? "Step 2: Submit electronically after July 9 via ASSIST (*(https://public.era.nih.gov/assist)* using funding opportunity announcement number."

Answer (Dr. Ozenberger): Applications cannot be submitted prior to July 9. The deadline for submission is 5:00 local time, July 12.

Question: Funding will come from an Other Transaction Award, and thus the scope of the project could change drastically. Will there be any guarantees around timelines/numbers of samples for awardees that would allow them to scale up without too much of a financial risk?

Answer (Dr. Ozenberger): The program and awardees will establish a collaborative relationship that will result in awardees having all available information on enrollment numbers and timelines so that they can plan accordingly.

Question: Is this funding announcement only for Genome Centers, or can private companies apply as well?

Answer (Dr. Ozenberger): Private companies can apply as well.

Question: Is a separate data sharing plan document required?

Answer (Dr. Ozenberger): The evaluation process in the FA comments that a history of data sharing is an aspect *All of Us* Research Program is looking for. However, unlike most NIH large data grant applications, this FA does not ask for a separate data sharing plan. See the "Ownership of Data, Software, and Other Products" section for more information about rights to products developed under an award.

Question: Do you require an approval letter for applications, since the funding request is for greater than $500,000, like in other awards?

Answer (Ms. Fleisher): No, this is not required for this FA.

Question: Is a letter of intent required?

Answer (Ms. Fleisher): No, a letter of intent is not required for this FA.

Question: Do you have any specific requirements for the clinical report?

Answer (Dr. Ozenberger): We have not specified the structure of the clinical report; however, it should meet community norms. We do not allow appendices in the application; however, it should be described as part of the technical description for the CVL.

Question: Can multiple technical options or approaches be presented in a single proposal?

Answer (Dr. Ozenberger): Yes, but options should be clearly described. Reviewers will consider advantages and disadvantages.

Question: Should the main contractor and subcontractor register for eRA Commons? If submitting with a subcontractor, should the business and workflow relationship be detailed?

Answer (Ms. Fleisher): No. Only the main organization that is applying for this award needs to register with eRA Commons. It would be appropriate to detail that relationship for the reviewers to understand the connection and collaboration.

Question: Can the subcontractor appear on more than one proposal from several different primary proposals? Related: Can the same company submit a proposal as the main contractor and be a subcontractor on another company's main proposal?

Answer (Ms. Fleisher): The answer to both questions would be yes; however, that will be taken into consideration during the review and the award since we wouldn't want to have the same company be a prime on both awards we are giving out, because of the volume of work that will be done by them.

Question: Will there be an option for follow-up questions later?

Answer (Dr. Ozenberger, Ms. Fleisher): Yes. Dr. Ozenberger will be available. He can be contacted directly via email. In the FA, the email address is listed for the central mailbox. The central mailbox should be used for all questions about the submissions process or all other questions.

Question: The ASSIST system includes a place for budget forms, but on the webinar, it was stated that no budget forms should accompany this application and all cost information should be contained in the 5-page cost section. Please confirm this is correct, and that in the ASSIST form we can leave the budget form field blank.

Answer (Ms. Fleisher): The ASSIST OTA interface, does not require budget forms. Item #7 allows the user to enter an estimate for project funding and the completion of this field is requested.

Budget

Question: Are large equipment purchases allowed under the grant mechanism?

Answer (Dr. Ozenberger): Equipment amortization may be included in the models for genotyping or sequencing unit costs. No separate equipment request is permitted.

Question: Will there be a cap on F&A costs on the first $75,000 of genomic array costs?

Answer (Ms. Fleisher): The *All of Us* Research Program will follow the NIH precedent for limiting genomic array F&A cost. The same NIH policy that applies to grants and cooperative agreements will be followed. Therefore, F&A costs will be limited to the first $75,000 of direct costs spent on genomic arrays.

Question: Do you require "full costs" of analyses to be accounted for in the proposal, or do you expect discounted costs (i.e., for institutions that will repurpose existing instrumentation)? The latter formula would put smaller "shops" at a competitive disadvantage.

Answer (Dr. Ozenberger): We want actual costs as well as projected costs for out-years.

Question: Can you speak to the total funding envelope over the five-year award?

Answer (Dr. Ozenberger): The *All of Us* Research Program hopes to provide funds for whole genome sequencing on 200,000 people a year. It is not known what this means exactly in terms of funds. The *All of Us* Research Program has long-term support from the government as part of the 21st Century Cures Act. It is also funded through the Office of the Director and receives an annual allocation from NIH. This is a very large funding component for *All of Us* Research Program; no one has attempted a million genomes before.

Question: Are full budget forms required for each sample count scenario for each year in this initial submission? If there are no budget forms, do all cost information and tables still need to be included in the five-page limit?

Answer (Dr. Ozenberger): No. Only the genomic sequencing piece requires the descriptions of the budget required for those levels.

Answer (Ms. Fleisher): There are no standard budget forms for other transaction (OT) applications. However, when providing the budget documentation, it is critical that applicants be very clear and transparent on the cost model and how it relates to sequencing. All cost information and tables must be within the five-page limit.

Question: Do we incorporate the cost of clinical validation in the per-genome cost or propose a different cost structure just for the clinical validation (per variant validated or per sample)?

Answer (Dr. Ozenberger): The clinical validation activities should be budgeted separately. The *All of Us* Research Program provides some number guidance on the scale. It will be approximately 3,000 assays for Year 1 and 6,000 for years after that. This is based on the frequency of ACMG variants based on primary data. The clinical validation piece should be budgeted separately, as it is not based on the primary laboratory activities.

Question: Previously funded *All of Us* Research Program grants do not seem to have full indirect costs (IDC). Is that due to (1) equipment and other expenditures that do not have IDC or (2) negotiation with NIH to reduce IDC?

Answer (Ms. Fleisher): In general, this could be related to large costs that are excluded from the indirect costs base and negotiation with NIH for either a reduced IDC or determination of an agreed-upon IDC for organizations that may not have an approved rate, or a combination of both.

Question: Can you speak to budgeting using specialized service facilities or cores and its impact on the direct cost cap specified for the funds available in Year 1? Can the per-unit costs utilized by our SSF split direct costs from the indirect costs for purposes of calculating the total direct costs for Year 1?

Answer (Dr. Ozenberger): NIH staff (and reviewers) are expecting variations in cost models, and these will be considered in the review process. Applicants should use the $15 million direct-cost limit as basic guidance but should not be overly concerned if the total direct costs are clouded by an SSF budget approach. Transparency in cost categories and projections will be key for the reviewers.

Question: Can subcontract IDC be excluded from direct cost totals for the sake of computing $15M direct cost limit?

Answer (Ms. Fleisher): Yes, subcontract IDC is excluded.

Regulatory Compliance

Question: On page 9 of the announcement, it states, "limited variant interpretation for medical utility, specifically pathogenic and likely pathogenic variants in the genes on the ACMG list and common pharmacogenomic alleles (e.g., CPIC Level 1)." Can you please clarify the CPIC level designation? CPIC levels are designated by A–D. Is CPIC Level 1 in reference to the PharmGKB Annotation level? For reference: https://cpicpgx.org/prioritization/#flowchart.

Answer (Dr. Ozenberger): The intent was to reference CPIC Level A variants as an element for return of genetic information in the *All of Us* Research Program.

Question: Does the NIH policy on budgeting for genomic arrays apply to this application for the purposes of estimating per-genome costs (i.e., a limit of $75,000 from genomic arrays eligible for indirect costs)?

Answer (Dr. Ozenberger): Yes, for the genotyping objective.

Question: Must the CLIA lab be physically located in the United States, even if the bidding entity is U.S.A. registered and located?

Answer (Dr. Ozenberger): Yes. Samples will not go outside of the United States, and all work must be completed in the United States. Nothing can be sourced to foreign organizations.

Question: Are the assays required to be CLIA-validated assays, or can they be performed in a CLIA lab? Is a CLIA report required?

Answer (Dr. Ozenberger): The facility must be CLIA certified, but high-throughput assays (except those in the CVL) are not required to be fully implemented as clinical grade. Applicants should describe procedures for fully authorized CLIA assays at the *All of Us* Research Program scale, although implementation of such assays is not expected in Year 1. This will be a dynamic situation, and we will be working on it as we go along.

Question: The RFA requests a description of the lab's "CLIA certification and schedules." In this context, does "schedules" refer to inspection dates and/or proficiency testing?

Answer (Dr. Ozenberger): Yes, it refers to both. The *All of Us* Research Program likes to have an idea of the applicant's CLIA process and the frequency of testing and inspection.

Question: For applicants with CLIA certification at the time of application, would the *All of Us* Research Program want to see any documentation as part of the application package?

Answer (Dr. Ozenberger): No. Applicants do not need to use extra pages to provide CLIA documentation. This should be described within the text. There will be plenty of just-in-time information at the time of award.

Question: Do you require both the SNP genotyping and the WGS to be fulfilled in CLIA?

Answer (Dr. Ozenberger): Yes, we require that both genotyping and whole genome sequencing be done in a CLIA environment. We will be returning primary data files to participants, so all data needs to be done in a CLIA lab.

Question: If the lab is not CLIA by July 12 but it will be by the end of September (which is the time the award is given), is that still acceptable?

Answer (Dr. Ozenberger): Yes, this is acceptable. CLIA certification is not initially required, but the FA asks that, if the lab is not currently CLIA certified, a route and path to CLIA certification be provided.

Question: Is New York State CLIA certification required to be considered a CLIA environment? Do you require NY license for CLIA? Will NY CLIA certification be required by September 30?

Answer (Dr. Ozenberger): Yes. Awardees will be asked to work toward NY certification. Data and processes should be compliant across all 50 states. This will not be required by September 30, 2018. This will take considerable effort for applicants who have not already engaged the NY regulatory bodies. It is not a requirement for this application, but it is understood that any awardee will have to get this certification prior to initiating work.

Question: In the RFA, it states that" [t]he CVL within a Genome Center award will receive a test request and a DNA sample with a unique identifier, run a validation assay, and transmit a clinical report to an *All of Us* genetic counseling resource for communication of results to the participant." In New York State (NYS), clinical tests must be ordered by authorized individuals (e.g., physicians, other medical practitioners, law enforcement officers) and the results returned to those individuals. A patient also may request a copy of the results. Can you specify what the *All of Us* genetic counseling resource consists of to help us prepare to be compliant with NYS requirements?

Answer (Dr. Ozenberger): The genetic counseling resource is still under development, but we anticipate a case study approach with licensed genetic counselors being assigned each case in which an ACMG gene pathogenic variant is validated. They will contact and provide initial counseling of that participant and ensure that the participant is directed to appropriate health care.

Technical Approach

Question: Would you require that both genotyping and whole genome sequencing be performed in one facility (or one for genotyping and one for WGS)?

Answer (Dr. Ozenberger): We encourage consortia, so no, they both don't have to be performed in one facility; there can be multiple facilities.

Question: Can we bid for data analysis solutions only?

Answer (Dr. Ozenberger): No. The Genome Centers are to be fully functional and able to fully cover all the objectives described in the FA. However, consortia are encouraged. The activities do not have to occur within a single organization or facility.

Question: Do you require the Genome Center to interpret variant pathogenicity and clinical significance for both sequencing and genotyping?

Answer (Dr. Ozenberger): Yes, but only for those genes/variants to be potentially returned to participants. At present time, those genes are those contained in the ACMG incidental findings list and major pharmacogenomic loci defined by CPIC.

Question: Will all data be funneled into a central database run by the *All of Us* Research Program?

Answer (Dr. Ozenberger): Yes, all data are collected in the Data and Research Center (DRC) of *All of Us*.

Question: Is there information available on sample DNA quality or quantity to be expected? For example, is DNA expected to be high molecular length? Also, what will the starting material be for the Genome Centers?

Answer (Dr. Ozenberger): Quantitated and QCed (OD metrics) DNA will be provided from the *All of Us* Biobank. This data will be available and they will be able to provide the concentration of DNA at the centers.

Question: Will the NIH ship DNA samples pre-normalized specific to the manufacture guidelines for genotyping, or will the contractor normalize the samples in house?

Answer (Dr. Ozenberger): DNA samples can be provided by the Biobank to center specifications.

Question: What type of sample container (i.e. cryovial, matrix tube, plate) will the DNA samples be shipped in?

Answer (Dr. Ozenberger): Samples are stored in cryovials but can be provided in containers best suited to center work practices. The Biobank will work closely with the centers post-award.

Question: Do cloud storage instances for temporary genomic data have to be GovCloud/ FEDRAMP compliant?

Answer (Dr. Ozenberger): There have been discussions of whether it would be possible to immediately upload data to the DRC, which is in a Google Cloud platform. The privacy and security of the participant's data is a very serious issue. All applicants are asked to agree to the principles of privacy and security established by the Precision Medicine Initiative® (PMI) and the *All of Us* Research Program. There will be post-award discussions with the security team about the specific requirements.

Question: What is the timeframe for clinical validation of the array/WGS platform?

Answer (Dr. Ozenberger): Assuming this refers to the high-throughput platforms, this timeframe will be determined in collaboration with NIH staff.

Question: Does the NIH require just the workflow (analytical) clinically validated or validation of each gene of interest, or just the capabilities to expand as the NIH sees fit to validate certain markers?

Answer (Dr. Ozenberger): NIH requires the capabilities to expand as the NIH sees fit to validate certain markers.

Question: Will ACMG variants be classified according to a particular ClinVar release? How important is the inclusion of ACMG variants with conflicting pathogenicity classifications on the array platform?

Answer (Dr. Ozenberger): We are in the process of fully defining and documenting our return of information platform/polices/procedures for the *All of Us* Research Program. However, we are likely to take a conservative approach, so variants that don't have high evidence of pathogenicity and very high penetrance rate are probably going to be set aside. Over time, we may add variants. This will be a dynamic process, but for the purposes of the FA, just consider ACMG59 "pathogenic" and "likely pathogenic."

Question: Which data (BAM or VCF) will be made available to the participants, and will this be done by the Genome Centers or the Data Resource Center? If it is done by the Genome Centers, will this just be the ability to download or visualize or the capability for further analysis?

Answer (Dr. Ozenberger): The Genome Centers will not provide data directly to participants. They will provide all data to the DRC. In the case of the clinical validation laboratory, these data will go directly to the genetic counseling resource. The FDA advises that unless there is an investigational device exemption (IDE), which may be required anyway for this program, the variant call format (VCF) could not be returned. VCF is considered an interpreted data file. NIH works closely with HHS, such as with CMS in terms of CLIA requirements and with FDA in terms of IDE requirements. The program will likely be considered high risk, since predominantly healthy people are being evaluated. There will likely be a need to file with the FDA for an IDE. This will be done in close collaboration with NIH staff and the Genome Center awardees. Only Genome Centers will have the technical expertise to file such a request with FDA.

Question: What is the objective in devoting significant resources to genotyping when it may be phased out when WGS is fully implemented? Why not focus exclusively on WGS?

Answer (Dr. Ozenberger): We feel it will be a leap for many organizations to return results with CLIA assays with these variants all validated for ACMG and PGx to be complaint with FDA regulations for WGS. There may be organizations that do have that capability, but many do not yet. The regulatory compliance has pushed us toward a genotyping approach, and then a clinical validation pathway offers the genotyping arrays. Also, the program is looking at costs and how the additional cost of WGS over genotyping will affect funds at this scale. This is one reason we are asking for multiple cost models in the out-years. We are currently unsure of funding allocations available for the Genome Centers. The *All of Us* Research Program does have stable funding in the federal government, but the amounts allocated may change, so we are trying to push genotyping right away so we can at least return some genetic information to participants as soon as possible. The goal is WGS, and we may shift entirely down the road; however, we have several years of great value in genotyping in the beginning.

Question: Is the expectation that the contractor will work with the *All of Us* Research Program consortium on finalizing a clinical report to be released to participants for both genotyping and WGS?

Answer (Dr. Ozenberger): Yes, there will be considerable collaboration on development of the genome report to be offered to participants, particularly between the Genome Center awardees and the *All of Us* DRC.

Question: Will phenotypes be provided? If so, at what stage?

Answer (Dr. Ozenberger): There will be very rich phenotypic data. The *All of Us* Research Program is collecting EHRs, survey instruments, and simple physical measurements of each core participant. Those are all within the research platform. The work of the Genome Centers will not entail linking the genome data with the phenotypic data. If the question relates to family history and how to counsel someone who has an ACMG variant, this must be thought about more. The Genome Centers should consider their work to be devoid of phenotypic data. *Added post-webinar:* Phenotypic data (e.g., gender, ethnicity) can be provided to contribute to quality assurance and for optimization of analysis pipelines.

Question: Will basic metadata (gender, race, etc.) be made available for quality control purposes at any stage?

Answer (Dr. Ozenberger): As it makes sense in the protocol. If it is part of the quality control, data could be made available.

Question: Imputation was mentioned in the prefunding notice for this contract. Is this still a desired capability?

Answer (Dr. Ozenberger): Applicants should consider *All of Us* Research Program goals and determine whether imputations should be included in the analysis proposal. This work could be done at the Genome Center or at the DRC. We ask the applicants for their best judgment on whether to include this in the analysis.

Question: The RFA asks us to describe capabilities and experience in analysis of other omic data types such as RNA and microbiome. Will there be a component for these data types?

Answer (Dr. Ozenberger): Yes. The *All of Us* Research Program aspires to extend omics research and data generation beyond the genome. As the budget permits and as the protocol evolves, additional measures could be added (e.g., collecting blood in a way that would enable RNA or cell-free DNA analysis). There is a lot of advice to provide microbiome data. Some of these data may not be on the whole cohort but on certain elements of the cohort. We ask applicants to provide their expertise on other elements in the application. This is not a major part of the technical procedures element, but we appreciate information on other omics that could be provided for the program.

Whole Genome Sequencing

Question: Can you clarify if the centers will genotype and sequence the same samples? If so, will the genotyping and sequencing be requested from the same aliquot? If not, will the samples be annotated in such a way that the center would know that they are the same? We typically genotype all our WGS samples (on Exome arrays) prior to sequencing, and use the genotyping data for fingerprinting and QC, as well as check final concordance with the sequencing data. If

we already had genotyping data from the array used for All of Us Research Program, we could skip the genotyping we typically do.

Answer (Dr. Ozenberger): Correct. High-density genotyping results from a sample can be used for concordance checking of same sample in the WGS pipeline.

Question: Should applicants provide a path to sample selection from genotyping data? Who decides which samples get whole genome sequenced?

Answer (Dr. Ozenberger): The intent is to whole-genome-sequence every participant. The program wants to do this for 1 million participants. It remains to be seen whether this is achievable within the five-year award period. In the meantime, the intent is to begin genotyping every participant. Therefore, genotyping will come first. As whole genome sequencing is ramped up, it will catch up. There will not be any selection process, although certain groups may be prioritized for whole genome sequencing early on. This is being intensely discussed in the consortium in terms of scientific benefits. The awardees on this program will have an opportunity for input on this question.

Question: Do you have an expectation of sequencing coverage and sequence quality each sample needs to meet?

Answer (Dr. Ozenberger): The applicant should propose these metrics, taking into consideration the goal to optimize use of the data by researchers. The intent of the program is to return some findings to the participant.

Question: Can you clarify if you truly mean “assembly” in addition to “alignment” when talking about whole genome analysis?

Answer (Dr. Ozenberger): Yes, assembly is desired.

Question: What sequence coverage and sequence quality metrics are expected?

Answer (Dr. Ozenberger): In the FA, specific coverage numbers or quality metrics for the sequencing data were purposefully not specified. The program is asking applicants to propose these metrics, taking into consideration the goal to optimize use of the data by researchers and the intent to return findings to participants, as well as costs that would be required for additional coverage.

Question: Do you have a list of clinically actionable variants you would like reported, or would the 59 medically actionable genes recommended by the ACMG for return in clinical genomic sequencing be acceptable?

Answer (Dr. Ozenberger): The *All of Us* consortium has decided to return pathogenic variants from the ACMG59 list which are measured effectively. This list of genes may become larger or smaller as the program gains experience. Procedures for returning CPIC Level A pharmacogenomic markers also are being developed.

Question: Throughout the RFA, there are comments suggesting there will be up to 100,000 genomes per year (“Projections should extend to very large scale, up to 100,000 WGS assays in a year,” page 10; “requirements to expand these activities to achieve scale (up to 100,000

WGS assays per year) for the *All of Us* program," page 12). However, in the table on page 4 of the RFA, Option C calls for 200,000 in years 2003–2005. Which might be correct?

Answer (Dr. Ozenberger): The program aspires to achieve a scale of 200,000 WGS assays per year. The decision whether to make awards to one or two Genome Centers will be made following receipt and review of applications. Applicants should emphasize in the technical descriptions the procedures for attaining a throughput of 100,000 WGS assays per year. However, should the program arrive at a decision to make a single award, budget information for a scale of 200,000 assays per year is requested.

Question: Will there be an expectation of realignment/reanalysis over time, or will this be done by a data repository? Will there be expectations of additional validation work?

Answer (Dr. Ozenberger): The *All of Us* Research Program works on the edge of technology and its capabilities. It is feasible that new analysis approaches and methods may come online during the award. Then reanalysis or realignment could be considered. There could even be deeper dives with additional data generation for certain parts of the cohort. There may be certain haplotypes or certain racial groups of special interest. Enthusiasm for the program should be generated by this flexibility. The program will work closely with the awardees to get the best data possible over the long term. The *All of Us* Research Program is expected to be around for 10 years; therefore, opportunities for new activities in the Genome Centers could be formed in later years.

Question: Can you expand on the "haplotype elucidation" request? The RFA specifies "haplotype elucidation" as part of the required standard analysis of short-read WGS. Do you have an opinion on the possibility that a Genome Center would propose to come back to some samples later to add long-read? Can we add this in the application?

Answer (Dr. Ozenberger): For whole genome sequencing, the only technologies that can achieve this scale are short-read instruments. However, the *All of Us* Research Program would like to assay structural variants that may not be detected by short-read technology as much as possible. We would like to know the haplotype structure of each of those genomes. Given the limitations of short-read technology, long-read technologies would be desirable as an add-on. This could be done if funds were unlimited. There are opportunities in the application to propose ways to have more power to detect structural variants and haplotype structure. These need to be balanced with the cost requirements for the application. A primary objective of the program is to optimize the price of WGS to be able to afford 1 million participants. However, applicants may propose long-read technologies to supplement lower-cost short-read technology and present benefits and costs.

Question: If a center already has clinical-grade WGS at the time of submission, do they still need to propose as a clinical validation laboratory?

Answer (Dr. Ozenberger): Yes. There will be a slow ramp-up of whole genome sequencing. The program expects to start later this year with genotyping, and pathogenic variants detected by genotype assay will require clinical validation.

Question: Would it be appropriate to provide some different models for the percentages of arrays and WGS to be performed in each year?

Answer (Dr. Ozenberger): The table in the FA under "Genotyping" is predicted to be close to the enrollment pace. It will be at 100,000 by the end of this year, and then it is hoped that 200,000 or more will be added each year after that. The models will be part of the review of the technical proposal. It may be considered if it is justified, this is based on the cost of the whole genome sequence assay. If there are data suggesting that costs might drop quickly and there might be rapid scale-up, different models could be proposed. However, the maximum per year is what is found in the table.

Question: Is there an expectation to sequence and return results from the pilot participants that were based on saliva (spit)?

Answer (Dr. Ozenberger): The intent is that the Genome Centers will receive some DNA samples isolated from saliva; this is something to consider in the WGS data generation and analysis pipelines. Even though most DNA will be coming from blood, we would suggest that applicants consider that there will be some DNA coming from saliva when describing those pipelines in their applications.

Question: What is the depth of coverage required by the *All of Us* Research Program for whole genome sequencing?

Answer (Dr. Ozenberger): Applicants should propose their best solution, considering the dual recipients (researchers and participants) of the data and costs.

Genotyping

Question: On page 9 of the FA, it says, "The *All of Us* Research Program consortium has discussed with major manufacturers the possible content of a high-density array…. Awardees are expected to collaborate with the vendors and the *All of Us* Research Program consortium post-award to develop the specific array to be employed for *All of Us*." Has the *All of Us* Research Program consortium relayed their marker lists to the microarray manufactures to build this custom array, or will the development of the array take place after the award is made?

Answer (Dr. Ozenberger): Final selection of an array platform and the specific content of the array will be determined post-award.

Question: Is an annual throughput of 100,000 high-density genotyping arrays immediately following award a requirement for all centers applying?

Answer (Dr. Ozenberger): Yes. However, "immediately" means after award negotiation and the final determination of genotyping platform in collaboration with the program.

Question: Can we offer multiple platforms for the genotyping? There are two major players in the genome world; are we allowed to present both as a solution?

Answer (Dr. Ozenberger): Yes, within the page limits, applicants are allowed to provide this information. Applicants should propose instrumentation that best suits their technical and budget plans. The reviewers will be looking at the pros and cons, determining what is the best approach that has been presented in the application.

Question: Is there a specified genotyping assay?

Answer (Dr. Ozenberger): The *All of Us* consortium has been working over the past year to develop the return-of-results policy and genomics strategy. This work has included discussion with providers, manufacturers of the genotyping assays, and outside experts in the Genomics Working Group. The corresponding publication can be found on the *All of Us* Research Program website (joinallofus.org). It has been determined that it is best to wait for the awardees under this program to participate in the decisions. From other NIH large-scale genotyping programs, it is known that it takes a long time and a lot of work to custom-design a genotyping array from scratch. Therefore, the *All of Us* Research Program will likely start with one of the major products that is currently used broadly in the community, with some custom content added in clinically useful variants and to cover the broad racial and ethnic diversity of the cohort. We ask the applicants to provide what they think would be the best platform to use. In the end, at the time of awards, there will be one platform for the *All of Us* Research Program. If there are multiple awards, it will be necessary to normalize the genotyping platform across awards.

Dr. Ozenberger asked applicants to provide *All of Us* Research Program with the best answer that they can.

Question: Is there a preferred reference for mapping and variant calling?

Answer (Dr. Ozenberger): No, we have not prescribed any methods we are asking each applicant to provide their best solutions to mapping and variant calling; however, we would advise that if we make multiple awards, we will want to normalize/harmonize the pipelines of the two centers. Ultimately, this is a collaboration among NIH and awardees, so submit your best proposal on methods of harmonization.

Question: What are the minimum requirements for phased genome blocks?

Answer (Dr. Ozenberger): Applicants should provide the best information on the data that are being provided.

Question: In addition to the ACMG Secondary Findings list of markers, does the *All of Us* Research Program consortium have specific content they wish added to the array (i.e. GWAS grid, CNV markers, eQTLs, PGx, etc.)?

Answer (Dr. Ozenberger): The *All of Us* Research Program does not expect to build an array from scratch but rather will adopt a standard array already in wide use and then add custom content to better suit programmatic goals.