Pre-Application Webinar for RFA-CA-20-053
Genomic Data Analysis Network: Genomic Data Analysis Center
(U24 Clinical Trial Not Allowed)
October 20, 2020
Questions and Answers

1. Q: Do you have a sense of the scale (i.e., number of cases to expect, for budget calculation and computing) that applicants need to prepare for?

   A: Over the next 5 years, we anticipate 5000–8000 samples that will be characterized. These will have whole-genome sequencing (WGS) low pass, whole-exome sequencing (WES), total RNA, microRNA (miRNA), some will have methylation. There might be extra projects like ATAC-seq that used The Cancer Genome Atlas (TCGA) cases, too. We are currently piloting an effort to do single-cell sequencing, both ATAC-seq, expression and Copy Number (CN) using TCGA samples; there will be 300 cases. The network would be participating in this project too.

2. Q: Do you expect some of the cases to have multiple samples from the same patient?

   A: Yes. For example, the Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trials (ALCHEMIST) project has some patients with recurrence and we have characterized the recurrence. However, this does not mean that every sample will have recurrences and metastasis.

3. Q: What fraction of the anticipated cases will have some outcome to be correlated? This is not overall survival so much, rather response to a therapeutic.

   A: The vast majority should have outcome data and rich clinical data. We have moved away from landscape-analysis-programs. The vast majority are clinical trials that have rich clinical information and outcomes. For the Carcinomas of Unknown Primaries Program (CUPP), even though it will be a landscape analysis program, it will have a significant amount of clinical data associated with the cases.

4. Q: Will all Therapeutically Applicable Research to Generate Effective Treatments (TARGET) samples be analyzed through the Genomic Data Analysis Network (GDAN)?

   A: We are considering doing some secondary analysis of the TARGET data and that may involve the network. TARGET will not be expanding the number of cases.

5. Q: Are most of the samples expected to be from metastatic cases?

   A: No, it will be mixed and depend on the clinical trial. At this time, we do not have a program that will specifically characterize metastasis. Some projects may have

metastasis.

6. Q: Will the clinical data be made available through the Genomic Data Commons (GDC)?

   A: Yes, it will be made available through the same portal at the GDC, similar to the molecular data. Once we publish the initial paper on the data, the dataset will be released to the public, through the GDC.

7. Q: It was mentioned earlier that we can have at least one core competency and might extend beyond the list noted on the RFA. Will the reviewers be made aware that besides the core competencies, there can be other platforms not listed on the RFA? Will you provide that feedback to the reviewers?

   A: This will be part of the instructions. One must however respond to one of the competencies. One may claim that other things that they are proposing fall under the specialized genomics category.

   Before the review session, the Program Office meets with the Scientific Review Officer (SRO) to present the RFA, goals, what is in scope, or out of scope. There will also be people from the Program Office at the review, although we do not get to participate unless asked questions. The SROs turns to the Program Office as needed for clarification.

8. Q: There are a set of core technologies that will be applied to every sample, and additional technologies (cell-free DNA, or single-cell analysis) that could be applied in certain instances. If your core competency is just cell-free DNA, is that correct that you would not be expected to participate in all analysis?

   A: If you have something very specific and we are not doing it, we will not call on your team to participate on the Analysis Working Group (AWG); however, we always welcome volunteers.

9. Q: Is there going to be prioritization on certain core competencies over others?

   A: There are a number of core competencies that we must have; these will be prioritized.

   We can select the highest scoring for that specific competency.

   Programmatic needs have a significant weight in RFAs.

10. Q: How are the analysis results within the AWG going to be produced?

A: The GDC has 5 different pipelines for exome sequencing, one for WGS, one for RNA-seq, etc. The GDC processes the data for all projects the same way. We will be incorporating pipelines for other data platforms—CN and WGS. All the data produced is stored at the GDC and after publication made available to the public.

11. Q: How important is single-cell data in this project?

    A: It is not immensely important right now, but single-cell technology is evolving.

    There will be cases for which single-cell data will be produced. At this time, we have budget limitations, but in the future it will likely be a standard platform.

12. Q: There was deeper WGS done in TCGA cases in the past. Is there a plan to do samples with deeper WGS?

    A: The intention is to sequence all TCGA cases with deeper WGS. We do not expect to have an AWG for this effort, thus will not call upon the centers to do the analysis of these data. We anticipate releasing the data as we get them processed so the community may use them. We will continue doing this as funds are made available.

    There will be no coordinated analysis of the WGS data because 1) we are not sure when the entire set will be done; 2) many tools that might be needed to look at WGS are not there. We want to provide this data for the tool developers to have a good dataset to work on.

    These WGS cases are in addition to the 5000–8000 initially mentioned.

13. Q: For the WGS data, will there be additional data on other platforms such as long reads?

    A: We are not planning on doing long read sequencing due to pricing. If pricing comes down, it may be considered in the future.

    This is a secondary competency in the RFA, because there is already some long read data available.

14. Q: Is there any potential for single-cell DNA sequencing?

    A: We have a pilot that includes single-cell RNA sequencing, single-cell DNA sequencing and single-cell ATAC-seq.

All the core competencies are at a minimum, one can add analysis not already included. It is value added when an entity is able to perform additional analyses.

15. Q: The RFA does not make a distinction between core versus secondary competencies.

    A: This is correct. People should try to address more than one competency.

    There is no secondary competency that does not have a paired, related, core competency.

16. Q: Is there value in claiming more than one of the 4 major competencies? Or would this be considered too defuse?

    A: You could focus on the competencies that are specific to your core abilities, then identify other aspects/competencies you are also able to contribute.

    In our AWGs, we have one group carry out analyses specific to what they are funded to do; then they also offer additional analyses they could work on.

17. Q: Will all data be stored at the GDC?

    A: The AWG Portal, Synapse, Wiki are the spaces where the data will be posted.

    The data used for the publication will be private. This will be stored at the AWG Portal.

18. Q: The RFA notes that any results of the GDACs will go back to the GDC. Does the GDC accept any data?

    A: Yes, the GDC will accept the data and if they do not have the pipeline to process it they will present it on the publication page. Example: We do not have a pipeline for ATAC-seq. The ATAC-seq data is found on the publication page specific to that paper.

    The GDC continues to evolve. For example, the GDC is about to release a single-cell RNA sequencing pipeline. Moving forward, single-cell RNA sequencing data will be available in the GDC as something that is searchable and indexed, after it has run through the pipeline. Other platforms will become available at the GDC in the future in the same way.

19. Q: What is the publication strategy going to be?

    A: The AWG will choose between the named authorship model or having everyone listed under the network. It will depend on the paper and discussions will be held accordingly.

20. Q: Will the slides from this webinar be shared?

    A: The slides as well as the Q&A document will be posted on the CCG website, under the Funding Opportunities section.

21. Q: The Human Tumor Atlas Network (HTAN) data is not part of this?

    A: This is correct, HTAN is not related to this RFA.

    We may use some of their data, similarly as we have done with the Genotype-Tissue Expression (GTEX) data in the past.

    There may be opportunities to collaborate with other NCI programs, though this network is mainly for CCG projects.

22. Q: Do we have plans to hold twice a year in-person meeting? Should we budget for meetings?

    A: We do not expect to have a set, twice a year in-person meeting. Rather we expect to have an in-person face to face meeting for each program when we get closer to the end of the analysis. (Only when it is safe to do so given the current pandemic.)

    It is recommended that one may add a small amount for travel as well as around $3000 for the publication fee per year. The latter is important because we have to pay the publication fee for the papers that are produced by the AWG. This typically falls under the grant of the co-chair who is funded by us. Each group will have a co-chair who comes from the molecular side and is funded by us. Note: If the money for these two line-items is not used for a particular year, it may be reallocated.

23. Q: What are your thoughts on computational resources aside from storage?

    A: These must be accounted for in the award as we do not provide you with computational resources. As a reminder, the alignment and calling will be provided at the GDC. If you know that you may need to run your particular pipeline that it is not at the GDC, you may calculate how many computer hours will be needed for all the samples to be run through the pipeline. This should be included in the budget.

24. Q: Are you guaranteeing that all data in the GDC will be available on cloud—either Amazon or Google?

    A: It will depend on the particular negotiations with Amazon and Google.

The GDC has a bucket with the data. However, we do not have the AWG data on a bucket; it is behind closed doors until publication. You could take the data from the AWG Portal and put it in your bucket in the cloud.