National Cancer Institute

The Genomic Data Analysis Network (GDAN)

RFA-CA-15-018/019/020

Pre-application Teleconference

December 15th, 2015

U.S. DEPARTMENT
OF HEALTH AND
HUMAN SERVICES

National Institutes
of Health

Background

- TCGA molecularly characterized 33 tumor types and as of 2015 has generated approximately 2.5 petabytes (PB) of data on over 11,500 cases of human cancer.
- “Lessons learned” from the TCGA :
  - Need for high quality molecular analytes
  - Utility of strict standardized protocols
  - Value of depositing the results in structured formats in publicly accessible databases
- The unique aspect of TCGA Project was the development and function of an integrated research network. Due to the success of TCGA, it is envisioned that all future CCG-approved genomics projects will utilize a similar model.

Background (continued)

- The intent of CCG is to conduct a coordinated and comprehensive genome-wide analysis of cancer-relevant alterations by simultaneously applying several technologies to interrogate the genome, epigenome or transcriptome in large collections of quality controlled cancer biospecimens derived from specific cancer types.
- Ultimately, in collaboration with and in support of the NCI’s extensive program of individual projects in cancer research, such efforts are expected to accelerate the identification of markers for disease prevention and diagnosis, novel targets for the development of therapeutic drugs, and provide the basis for a refined clinical understanding of patient stratification in therapy.

NCI's Center for Cancer Genomics Pipeline

- **Biospecimen Core Resource (BCR, awarded through a contract mechanism):** The BCR serves as the tissue processing center and provides the molecular biomolecules for all CCG-approved projects.
- **Genome Characterization Centers (GCCs, awarded through a contract mechanism):** GCCs conduct high-throughput comprehensive genome-wide analyses.
- **Genomic Data Analysis Network (GDAN, present concept):** The aggregated capabilities of the awardees from the present concept will produce the bulk of the analysis required to interpret the data generated by the GCCs. This group will work closely and collaboratively with all other components of the pipeline and be responsive to the necessities of the analysis requests posed by the Analysis Working Groups (AWGs) that will be formed for each CCG-approved project.
- **Data Management, Storage and Public Access (awarded through a contract mechanism):** The whole of the data generated by CCG-approved projects is presented to the scientific community though publically-available databases.

Projects to be Served by Pipeline

- The proposed analytical network will serve to understand the data generated by CCG programs as well as collaborations with other divisions. Examples (but not an exhaustive list) are:
  - The Cancer Genome Atlas (TCGA)
  - Therapeutically Applicable Research To Generate Effective Treatments (TARGET)
  - Cancer Driver Discovery Program (CDDP)
  - The Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trials (ALCHEMIST)
  - Exceptional Responders (ER, in collaboration with DCTD)
  - Clinical Trials Sequencing Program (CTSP, in collaboration with DCTD)
  - Environment and Genetics in Lung Cancer Etiology (EAGLE, in collaboration with DCEG)
  - Human Cancer Models Initiative (HCMI)

Platforms to be Assayed and Case Numbers

- All the platforms mentioned in RFA-CA-15-020 *could* be used for each and every project.
- Number of cases used for each project vary, but we expect that up to 10,000 cases could be characterized in the first two years.

Composition of GDAN

- **Processing GDAC**: Develop and implement appropriate bioinformatic systems for rapid high-throughput processing, integration and standardization of genome-wide datasets generated by the GCCs. A single such center will be awarded.
- **Visualization GDACs**: Provide user-friendly bioinformatics tools and data portals for the integration and exploration of the results generated by the first and third class of GDACs. It is expected that up to two awards of this class will be made.
- **Specialized GDACs**: Analyze the data generated by the processing and visualization GDACs as well as raw data generated by the GCCs and other collaborators for any and all CCG-approved projects. Will take advantage of highly skilled researchers in a single facet of the genomic landscape to produce results from analysis of a limited set of platforms. Eleven awards are envisioned.

RFA-CA-15-018
Processing GDAC

- Will process the data from the GDC (all platforms) to produce first pass analysis that specialized GDACs will use as starting point. Will have to re-run analyses as needed and directed by the AWGs.
- Pipeline needs to be in place at the start of the award.
- Pipeline should be deployable outside the generating institution with no dependencies. Use of commercial components is discouraged.

RFA-CA-15-019
Visualization GDAC

- Will produce web-based interfaces to display higher level data (not raw sequence) in an user-friendly, multi-platform, multi-angle way. Need to be able to respond to custom queries (although this might be a later development).
- Pipeline should be deployable outside the generating institution with no dependencies. Use of commercial components is discouraged.
- System must exist at time of award.

RFA-CA-15-020
Specialized GDAC

- Specialized centers will support analysis on a discrete number of platforms to respond to the data analysis needs of the AWGs.
- They must respond to at least one core competency, although they might extend beyond the list in the RFA.
- Although tools might be further developed to serve the functions, they must exist in some form at time of award.
- Personnel must be able to respond to queries by program at a moments notice.

General Considerations

- All awards will be done under the Cooperative Agreement mechanism, which means NCI will have substantial input in the guidance of activities.
- All awardees work for a network, thus it is expected that they will work in a collaborative manner with all components of the CCG pipeline.
- No person can be the PI of record in more than one award. They can act as co-PIs.
- All costs must be included in the proposals, no centrally funded resources will be available.

Out of scope

- Proposals that have a narrow goal and respond only to interests by applicant to use the generated data (R01 style).
- Informatic tool development without analysis.
- Experimental validation of the large-scale data

Questions?

- Questions and answers about this presentation and the applications will be posted in conjunction with this presentation.