Attachment 2 - Statement of Work

Title: Sequencing for The Cancer Genome Atlas Program (TCGA)

Background

The Cancer Genome Atlas (TCGA) is a comprehensive, collaborative effort led by the National Institutes of Health (NIH) to map the genomic changes that occur in major types and subtypes of cancer. Researchers throughout the nation are using various genome analysis technologies, including high through-put DNA sequencing, to carry out this effort.

A pilot project was initiated in 2006 established the scientific infrastructure and demonstrated the "proof of concept" needed to mount a large-scale cancer genome mapping project. Based on this success, TCGA announced in September 2009 that it will map the genomes of at least 20 cancers over the next five years, focusing on brain, lung, and ovarian cancers.

Purpose of the Project

The purpose of this project is to perform genomic DNA sequencing and initial analysis for the comprehensive characterization of sequence from the genomes of cancer specimens and associated normal tissues in support of The Cancer Genome Atlas (TCGA) program. This project will catalog comprehensively somatic DNA changes that are associated with cancer. Under this project, the contractor shall utilize the most advanced DNA sequencing technologies and genomic analysis computational methods to investigate at least 400 and as many as 1000 cancer cases (specific types to be identified by the NCI) over the next two years. A case includes DNA from tumor and case-matched normal tissue. In addition, the contractor shall sequence entire exomes, deposit the sequenced data according the TCG rapid data release policies, efficiently and accurately analyze the data to initially detect >80% of true mutations and polymorphisms and at the same time minimize the false positive discovery rate including complete analysis, and have an established and integrated bioinformatics capability sufficient to support production, including systems and database administration, laboratory information management, data handling and deposition.

Description of the Technical Requirements

Independently and not as an agent of the Government, the Contractor shall furnish all the necessary services, qualified personnel, material, equipment, and facilities, not otherwise provided by the Government as needed to perform the following:

The Contractor shall perform genomic DNA sequencing and initial analysis for the comprehensive characterization of sequence from the genomes of cancer specimens and associated normal tissues in support of The Cancer Genome Atlas (TCGA) program. The Contractor shall utilize the most advanced state-of-the-art DNA sequencing technologies and genomic analysis computational methods to investigate at least 400 and as many as 1000 cancer cases (specific types to be identified by the NCI) over the first 2 years of the contract. A case includes DNA from tumor and case-matched normal tissue.

Sequencing. The Contractor shall possess the requisite experience and capacity at the time of contract award to sequence entire exomes from at least ten (10) cases per month and whole genomes from at least two (2) cases per month. The Contractor shall produce one hundred (100) gigabases alignable DNA sequence per month and generate and analyzed at least two (2) whole human genome sequence data sets. The Contractor shall deposit the sequence data according to TCGA rapid data release policies that specify primary data are to be released as soon as possible, but within one (1) week of being generated (see http://www.genome.gov/10506376). At the end of the first contract year, the Contractor shall achieve sequencing and analysis capacity to characterize comprehensively the exomes of at least twenty (20) cases per month and whole genomes of at least three (3) per month. The Contractor shall possess the necessary instrumentation, computational infrastructure, and personnel required to achieve this scale within the first contract year.

Analysis. The Contractor shall efficiently and accurately analyze the data to initially detect >80% of true mutations and polymorphisms and at the same time minimize the false positive discovery rate. The Contractor shall prioritize and validate putative mutations derived from primary analyses. The Contractor shall design the analysis pipeline. The Contractor shall complete the analysis, including putative variation calls, within four (4) weeks of data generation and the analyzed data must be released immediately, according to TCGA policy. The validation status of each called putative mutation shall be determined (see below) and variation data annotated within an additional four (4) weeks. Specifics will be developed in consultation with the NCI at the time of award.

Bioinformatics and Data Deposition. The Contractor shall use an established and integrated bioinformatics capable to support production, including systems and database administration, laboratory information management, data handling and deposition. The Contractor's data submission pipeline shall include: 1. the primary sequencing data files to be submitted to NCBI databases; 2. the processed variation files to be submitted to NCI's TCGA database; and 3. the integration of a Laboratory Information Management System (LIMS) with data submission protocols. The NCI will provide information about data formats to be used and assistance in identifying existing data elements and in curating new data elements, as needed.

Milestones. The Contractor shall meet the following milestones:

- Sequence information on a minimum of 95% of the exomic targets for a minimum of 95% of the cases sequenced by the Contractor.
- Sequencing data analyzed and aligned to a reference sequence and submitted to the public databases according to specifications provided by the Contractor within one (1) week of generation.
- A list of putative mutations that have been identified, prioritized (i.e. placed into bins relating to potential functional significance in whole genome studies) and submitted to the TCGA database within four (4) weeks of sequence generation.
    - Annotations of the analyzed data submitted to NCBI or NCI databases, including fields to be determined in consultation with NCI and other TCGA research groups.
- Verification of electronic submissions through NCBI reports, to be conducted by the

Contracting Officer's Technical Representative.

The Short Read Archive established by the NCBI is a NIH maintained database specifically designed for the deposition of data generated by second generation sequencing and the retrieval of those data. The Contractor will not develop, maintain or host a Federal Information System, but rather, will utilize the existing NCBI resources. The TCGA database is not a Federal Information System. It is a system maintained by a TCGA contractor specifically for the purposes of depositing, storing and distributing TCGA data.

The Contractor shall correct or re-work all unacceptable deliverables at no cost to the contract. The Government will not pay for the correction or re-work of any unacceptable deliverables.

Scale-Up Phase

Beginning on the effective date of the contract, the Contractor shall have the necessary staff and equipment to sequence a minimum of 100 gigabases of alignable DNA sequence per month. Within six (6) months of the effective date of the contract, the Contractor shall have the necessary staff and equipment to sequence a minimum of 300 gigabases of alignable DNA sequence per month and shall maintain this minimum and/or exceed this minimum throughout the entire contract period. Additional scale-ups may be required based on needs of the project