CONTRACT CONCEPT REVIEW

NTP Board of Scientific Counselors Meeting
April 19, 2022

Title: **Bioinformatics Support Contract for the Division of the National Toxicology Program (DNTP) and Division of Intramural Research (DIR), National Institute of Environmental Health Sciences (NIEHS)**

Presenter: **B. Alex Merrick, Mechanistic Toxicology Branch, Division of the National Toxicology Program, NIEHS**

Contract Purpose:

The objective of this contract is to provide bioinformatic support to the NIEHS' Division of the National Toxicology Program (DNTP) and the Division of Intramural Research (DIR) for their diverse research initiatives that involve large, complex, and multidimensional datasets. These groups have need for analyses of data from interrelated disciplines of toxicology, genomics, structural chemistry, high-throughput screening, data and text mining, and high-content-image analysis, and from files downloaded from large, public data repositories. The scope of the required activities and the space and personnel needed to conduct these studies, exceed the resources available at the NIEHS; therefore, we request support through contract mechanisms.

Background and Significance:

DNTP has a long history of conducting toxicology studies to characterize the effects of agents of public health concern. Current DNTP scientists conduct research using cutting-edge approaches and technologies to better understand how environmental exposures may impact human health. DNTP scientists work in multidisciplinary teams and may collaborate with other federal agencies, institutes, industry, and academia. DIR scientists conduct research on the role of environmental agents in human disease and dysfunction. Research initiatives for each division are further described at the following websites for the DNTP at
https://www.niehs.nih.gov/research/atniehs/dntp/index.cfm and for the DIR at
https://www.niehs.nih.gov/research/atniehs/dir/index.cfm.

The biostatistical analysis of toxicity data has played a major role in interpreting agent toxicity in terms of pathology, clinical chemistry, and biochemistry. With the introduction of 'omics' technologies and high-throughput screening, biomedical and toxicology datasets have become increasingly large, complex, and multidimensional. Over a decade ago, bioinformatics was first included in the NIEHS biostatistics contract to analyze toxicogenomics data. More recently, bioinformatics has emerged as a separate discipline to aid in the analysis and interpretation of toxicological effects. As a result, a separate bioinformatics contract was awarded in 2017 to support the more complex and multidimensional data being generated within NIEHS. Through this

recompeted bioinformatic contract that is in preparation, DNTP will explore using new approaches for the analysis of large-scale, experimental, and toxicological data that includes biological and technological advances in molecular and computational biology.

The DNTP and DIR anticipate the frequent need for bioinformatics support in the following areas:

- Data analyses of genomic and epigenomic sequencing data generated from newer next-generation-sequencing (NGS) platforms such as DNA-seq, exome-seq, DNA-methylation, and related strategies will be needed. Large data files would be analyzed for quality and then aligned to appropriate genomes obtained from public repositories. The biostatistical analyses will be guided by best practices and needs of DNTP and DIR research.

- Analytic support for data from transcriptomics, toxicogenomics, and high-throughput transcriptomic screening studies from high-dimensional platforms (e.g., microarray, RNA-seq, or BioSpyder) will be needed. The contractor will analyze gene-expression data from a variety of transcriptomic platforms, recognizing the strengths and weakness of each platform, and perform analyses with the necessary biostatistical rigor. Depending upon the size of the data files, they will be deposited in the Chemical Effects in Biological Systems (CEBS) database or other public repositories such as GEO (Gene Expression Omnibus) or SRA (Short Read Archive) and accompanied by appropriate metadata.

- Biostatistical, computational, and informatic methods will be needed to address toxicology problems using genetic, genomic, in-vitro or in-vivo data, and multi-scale modeling. Algorithms will need to be developed to efficiently extract and process data. The contractor will write source code to support analyses of complex, multi-dimensional data and should be able to integrate diverse data streams to solve different toxicology problems. These activities are expected to be highly interactive with DIR and DNTP staff and other contractors.

- Cheminformatic support for QSAR (Quantitative Structure Activity Relationships) and read-across efforts will be needed. This support requires the use and knowledge of chemical features, drug receptors, and biomolecular structures for construction of predictive toxicity models. Support also requires acquiring and combining data or data sets from various sources to construct read-across protocols.

- Image analysis and pattern recognition for in-vitro and high-throughput, cell-based screening studies will be needed. High-content data for subcellular features will be used to determine gradations of chemical cytotoxicity from cell-based screening studies.

- Informatics methods for information mining and scientific literature-based reviews will be needed. The contractor will provide computational support to develop approaches such as natural language processing, text mining, or machine learning for scientific literature-based reviews. Informatic methods will involve literature and data mining of the Internet (using large datasets to find patterns and correlations) to support systematic-review projects.

- Support will be needed for study designs and bioinformatic, biostatistical, and computational analyses that includes, but is not limited to, coding with Java, Python, C++, for programs used in text mining, machine learning, and natural language processing. Supported bioinformatics would include analyses of various 'omics' platforms, interactive data-interrogation tools, machine learning, and application of biostatistical methods to deploy new software tools needed for the research program.

- Support for data organization for slide presentations and peer-reviewed publications will be needed. Support activities will include constructing graphs, tables, and figures from data; providing text that specifies the methods and details the results for manuscripts, reports, or slide presentations; providing the citations and references for analytical procedures; reviewing draft technical reports, manuscripts, and related information; and participating in data presentations or in the development of study designs.

Proposed Changes to the Current Statement of Work:

The proposed Statement of Work (SOW) will include all current capabilities for carrying out bioinformatics support and, in addition, is being revised to include more contract abilities for computational support to the government for its use in accomplishing its research initiatives. Additional capabilities will include providing multi-scale analyses of different data types; building new bioinformatics tools that integrate large experimental or public data sets; using artificial intelligence and natural language processing approaches; and providing code for open-source software applications that address NIEHS' research needs. As a government research facility, the NIEHS will promote the sharing and distribution of data and source-code applications that it creates through this contract to address complex problems in toxicology data analysis.