CONCEPT CLEARANCE

THE NIH GENES AND ENVIRONMENT INITIATIVE (GEI): Whole Genome Association Genotyping

Cooperative Agreements for DNA Samples and Data Analysis

Contracts for Genotyping Centers and Coordinating Center

NHGRI Advisory Council, May 2006

Purpose

Staff seeks Council clearance for one RFA and two RFPs to support the whole genome association (WGA) genotyping component of a four-year NIH-wide initiative, known as the Genes and Environment Initiative (GEI). The long-range goal of GEI is to support research that will lead to the understanding of genetic contributions and gene-environment interactions in common disease. GEI is being developed and planned by an NIH-wide Coordinating Committee, administratively led by the National Human Genome Research Institute and the National Institute of Environmental Health Sciences (NIEHS).

Implementation of GEI will begin with a WGA component that will comprise three elements: 1) support as contracts for roughly three genotyping facilities, to perform high-throughput genotyping for a dozen or more WGA studies over four years; 2) support as a contract for a Data Coordinating Center (DCC) to manage the resulting data; and 3) support as cooperative agreements for a dozen or more investigative groups to submit samples from well-characterized subjects for WGA genotyping and/or replication studies and to analyze the resulting data. The WGA component of GEI will be administratively led by NHGRI on behalf of the NIH. The WGA genotype data will be assessed for quality and combined with phenotypic and environmental exposure data at the DCC. Curated data would be made available in a central database established by the National Center for Biotechnology Information (NCBI) for free and open research use.

Background

The Genes and Environment Initiative (GEI) was announced in February 2006 to determine the etiology of common diseases by focusing on genetic and environmental factors that increase the risk of these diseases. Funding for GEI is included in the President's budget proposal for fiscal year (FY) 2007. If approved by Congress, $26M per year in FY2007 through approximately FY2010 would be devoted to an NIH-wide program to identify major genetic susceptibility factors for diseases of substantial public health impact, and $14M per year during the same period would be devoted to the development of new technology for acquiring environmental exposure data.

Plans for the GEI are under development, but one of the initial decisions of the NIH-wide Coordinating Committee was that that the genetic component should not be limited only to the collection of WGA genotype data. In this way, the GEI genetics component will differ from the Genetic Association Information Network (GAIN, described at http://www.fnih.org/GAIN/GAIN_home.shtml). Support for data analysis, replication and fine mapping studies, sequencing, functional studies, database development, and clinical translation are all being considered for inclusion within GEI. The Committee recognizes, however, that the latter activities are all dependent on the data from high-throughput WGA genotyping and that acquisition and analysis of those data should be the focus of the early years of the GEI genetics program.

Contracts for Genotyping Facilities and Data Coordination Center

Research Scope and Objectives

The advent of low-cost, high-throughput genotyping has placed WGA genotyping within the reach of large-scale population-based studies of common diseases. Genotyping methods continue to improve in quality and decline in cost, making it feasible to assay 375,000 or more single nucleotide polymorphisms (SNPs), capturing 80% or more of the HapMap-defined genomic variation, in roughly 2,000 subjects for roughly $1.7 million per study. For these reasons, the first GEI solicitation for which Council clearance is being requested is an RFP for roughly three Genotyping Facilities. The second solicitation is an RFP for a Data Coordinating Center (DCC).

Genotyping methods are evolving rapidly, and methods that are state of the art when this solicitation is released may well be superseded or supplemented by better methods by the time the awards are made. Offerors for the Genotyping Facilities would thus be required to document not only an outstanding track record of timely and high-quality genotype production using currently state-of-the-art technology, but also to describe their past experience in adopting evolving technology and how they would propose to incorporate new methods in the future. High-throughput genotyping would likely focus on single nucleotide polymorphism (SNP) markers, but the specific technology or combination of technologies would be at the offerors' discretion. The specific platform to be used would be determined as close in time as possible to the actual performance.

The Genotyping Facilities would each: 1) receive DNA samples and assess their quality; 2) collectively, perform whole-genome SNP genotyping of roughly 2,000 samples for a dozen or more common complex diseases or traits, and as appropriate, replication studies on a subset of associated SNPs; 3) clean and assess quality of genotype data; 4) transfer completed and cleaned data sets to the Data Coordinating Center; and 5) work with analysts, the DCC, and NIH to resolve any issues with the data. More than one Genotyping Facility, potentially using more than one genotyping technology, would be sought to provide competition within the program to help ensure continued improvement in genotyping methods, and to provide the ability to compensate for possible poor performance by one contractor or take advantage of an improved platform. Offerors would be expected to demonstrate comparability of their proposed method(s) to other

available technologies and to provide approaches for assessing and ensuring comparability as methods continue to improve. Genotyping projects would be matched to Genotyping Facilities by NIH staff based on appropriate methodological considerations (if any), available capacity, and performance to date.

The Data Coordinating Center would: 1) receive, review, and manage genotype data, including quality control data, and request verification or further genotyping as needed; 2) coordinate efforts of multiple Genotyping Facilities and ensure data comparability among them; 3) receive, review, and manage phenotypic and environmental data from submitted studies; 4) transfer completed and merged data to NCBI; 5) work with the genotypers, the analysts, and NIH to resolve any issues with the data; and 6) conduct basic data analyses for individual studies or program-wide in collaboration with the submitting investigators. Recognizing the need for similar coordinating functions for other GEI components, such as replication/fine mapping, sequencing, and functional studies, and the desirability of including them in a single DCC for the entire genetics component, the statement of work for the DCC will include providing such functions where appropriate.

The Genotyping Facilities and DCC would be expected to cooperate closely with each other, with investigators submitting data and samples, and with NIH.

Mechanism of Support

This initiative will use the NIH Research and Development Contract award mechanism. While NHGRI will provide lead this WGA aspect of the NIH-wide cooperative GEI effort administratively, management of the program will be done on a trans-NIH basis.

Funds Available

NIH will commit approximately $44M in GEI funds over a four-year period to these contracts, with roughly three-quarters of that funding concentrated in the first two years Approximately 10% of this amount will be committed to the DCC. NIH anticipates funding up to four awards (one DCC, three genotyping facilities) in response to these solicitations.

Cooperative Agreements for Samples and Data Analysis

Research Scope and Objectives

The third GEI solicitation for which Council clearance is being requested is an RFA that would solicit applications from a dozen or more investigative groups, to submit samples and data for WGA genotyping, replication genotyping, or both, and to conduct analyses of the resulting data. It is expected that the work of an investigative group could be completed in a two-year period.

Responsibilities of investigators receiving support would include: 1) providing DNA from an adequate number of samples in sufficient quantity and quality for WGA or replication genotyping; 2) providing detailed phenotypic and environmental data, collected according to widely-accepted and validated protocols, to the study's Data Coordinating Center; 3) providing high-quality documentation, including study protocols, manuals, coding, and other information that will be sufficient for outside users to understand and use the data with minimal assistance; 4) participating in a study-wide, WGA Steering Committee to share design and analysis techniques and promote comparability across genotyped studies wherever possible; 5) analyzing and publishing WGA findings and pursuing replication or other follow-up studies as needed (but under separate support). Investigators would be permitted to request initial WGA genotyping, replication genotyping of a subset of associated SNPs in additional participants, or both. Applications from individual investigators or consortia would be permitted.

Studies to be genotyped would be selected by peer review, based on: the public health significance of the proposed trait(s); the need for a high-throughput WGA study; the strength of evidence for a genetic component for the trait, anticipated size of a genetic effect, and the power to detect it; the strength of evidence for an environmental component, anticipated size of an environmental effect, and the power to detect it; the appropriateness of the study design and population(s); the quality and completeness of phenotyping and exposure measures; strength of plans for replication and follow-up studies; appropriate sex and race/ethnic representation; and the programmatic balance among diseases and across multiple NIH WGA initiatives.

The investigative groups would be expected to cooperate closely with each other, the Genotyping Facilities, the DCC, and NIH, to ensure maximum progress in generating a high-quality resource of genotype and phenotype data and disease associations.

Mechanism of Support

This initiative will use the NIH U01 (Cooperative Agreement) award mechanism. While NHGRI will provide the administrative lead for this aspect of the NIH-wide cooperative GEI effort, the management of the program will be done on a trans-NIH basis.

Funds Available

NIH will commit approximately $4.0M in GEI funds over a three-year period to these two-year cooperative agreements, and expects to make roughly six awards in FY 2007 and roughly six additional awards in FY 2008. This funding will support analysis for each study; the funding for genotyping will be in the genotyping facility contracts. NIH anticipates funding a dozen or more investigative groups studying roughly 2,000 participants each in response to this solicitation.

Funding for the three solicitations, for the genotyping facilities, DCC, and submitting investigators, will total approximately $48M from FY2007-FY2010, representing approximately 46% of the total four-year GEI genetics budget of $104M.