

**National Advisory Council for Human Genome Research
May 22, 2006**

Concept Clearance for RFAs

Creating the Encyclopedia of DNA Elements (ENCODE) in the Human Genome

**A Data Coordination Center for the Encyclopedia of DNA Elements
(ENCODE) Project**

**Technology Development for the Exhaustive Determination of Functional
Elements in the Human Genome**

Background: The long-term goal of the Encyclopedia of DNA Elements (ENCODE) Project is to find all of the sequence-based functional elements in the human genome. This Project was launched in September 2003 with the funding of a three-year pilot effort to analyze 1% (30 Mb) of the human genome sequence. A parallel technology development effort, with the purpose of expanding the repertoire of high-throughput methods available for identifying functional elements, was supported in two rounds of funding, in 2003 and 2004.

In January 2006, the NHGRI held a meeting of the ENCODE Project's Scientific Advisory Panel and additional experts to assess the progress of the ENCODE pilot project and to discuss whether the Project was ready to expand to the next phase, a production-level effort to analyze the entire human genome. The meeting attendees were impressed with the progress that has been made in ENCODE to date, and thought that the pilot project had by and large been successful. In considering the next phase of ENCODE, they emphasized that conducting a genome-wide, comprehensive study is a complex undertaking that requires that attention be paid to many issues in order to meet the goal of producing large amounts of interpretable, high quality, and useful data. The meeting participants were of the opinion that while one or two technologies might be ready to be applied at scale, there were a number of components of the pilot project that had not yet addressed the issues of scaling well enough to give them confidence that the Project as a whole was ready yet to scale to the analysis of the entire human genome. Therefore, they recommended that the pilot project period be extended by a year to allow the ENCODE Consortium and the NHGRI to continue to prepare for the next phase of the Project. They felt that this was a sufficient amount of time for at least some of the technologies to become ready to scale, since it was evident that investigators were beginning to grapple seriously with these issues. The advisors also suggested that once the ENCODE Project does scale to the whole human gene, some aspects of the study would continue to benefit from some additional work on the same 1% of the genome that was targeted during the current pilot project, such as more in-depth sequencing for comparative sequence analysis. Lastly, they recommended that NHGRI consider funding more data analysis efforts and continue to support technology development, including better methods for validation, especially technically independent methods.

To implement these recommendations, NHGRI staff, after consultation with the Council at its February 2006 meeting, decided to extend the pilot project for a fourth year to allow the issues of scaling to be addressed.

Proposed Research Scope and Objectives: This request is for Concept Clearance for the implementation of the next, expanded phase of the ENCODE Project. The NHGRI proposes to expand the ENCODE Project in two ways: 1) to initiate the scaling of the ENCODE Project to the entire human genome; and 2) to support further in-depth investigation of the initial 1% target regions. For some ongoing studies, more in-depth analysis can be expected to yield greater insight into the most effective approaches to uncovering functionally important regions of the genome. Additional studies on these well-annotated targets can also serve as a test bed for new and emerging technologies.

To facilitate the project, the storage, display and release of data generated by the ENCODE Consortium will be coordinated through a central database. NHGRI proposes to seek applications to support such a database in a separate solicitation. Lastly, as envisioned from the beginning, this is a long-term endeavor that encompasses many different technological approaches, some of which are much further along in development than others. The need still exists for further technology development, especially in the areas of finding functional elements in repetitive sequences and developing validation methods. The NHGRI proposes a third RFA to solicit applications to support this area of technology development.

RFA: Creating an Encyclopedia of DNA Elements (ENCODE) for the Human Genome

NHGRI proposes to issue an RFA to solicit applications for four-year research projects to participate in the ENCODE Research Consortium, focusing on the use of high-throughput, cost-efficient approaches to generate a comprehensive, high-quality catalog of functional elements in the human genome. This solicitation will be open to all investigators, both those who are currently participating in the ENCODE pilot project as well as those who are not. Both computational and experimental approaches will be encouraged, as will applications for analysis of the forthcoming data. These projects should address the entire human genome using methods for which a defined process for data generation and validation, including quantitative measures of data quality (both accuracy and comprehensiveness) has been, or can rapidly be, established, and for which it can be demonstrated that application to the whole genome achieves economies of scale and improvement in data quality. The specificity (i.e., false positive rate) of the technical approach needs to be defined. Furthermore, evidence is needed to validate the authenticity of the biochemical event, such as could be determined by a second, preferably independent experimental method. The sensitivity (i.e., false negative rate) of the technical approach is more difficult to obtain as there is no "gold-standard" against which experimental conclusions can be compared, but it will be necessary for applicants to estimate the fraction of all functional sites that can be identified by the experimental approach and to evaluate the comprehensiveness of the approach. In reality,

achieving a “complete” catalog of all functional elements will not be feasible in the next four years, given the practical limitations of current technologies and funding levels to query every tissue and developmental stage. Therefore, NHGRI will consider focusing these efforts on a particular biological system or directing the project in some other way as to realize “completeness” in a biologically constrained context.

As mentioned above, the generation of more data and more in-depth analysis of the 1% of the genome being studied in the current pilot effort can be expected to yield greater insight into the most effective approaches to uncovering functionally important regions of the genome. New and emerging technologies may also benefit from preliminary application to these well-annotated targets as they are being developed for whole genome analysis. Therefore, this solicitation will also seek four-year research projects to continue investigation of the 30 Mb of the human genome targeted in the pilot project. These projects may be 1) a continuation of ongoing pilot projects to allow for more in-depth analyses; 2) an expansion of the current ENCODE technology development grants to allow testing and comparison to other technologies at the pilot-scale level; and 3) tests of novel technologies that have not yet been used in the pilot project.

Mechanism of Support: The primary mechanism of support will be the cooperative agreement. It is anticipated that different mechanisms may be used, in addition, to support other activities necessary to meet the needs of this project, such as the development and distribution of common reagent sets. The precise level of support will be approximately \$23M in the first year, but will, of course, be dependent upon the amount of available funds and the quality of the applications received.

RFA: A Data Coordination Center for the Encyclopedia of DNA Elements (ENCODE) Project

The purpose of this RFA is to solicit applications to develop and implement a Data Coordination Center (DCC) to serve as a centralized database for the ENCODE Project. The DCC will be funded primarily to develop, house, and maintain databases to track, store, and provide access to the data generated as part of the ENCODE Project.

Mechanism of Support: The DCC will be supported through the U41 Cooperative Agreement mechanism. A total of \$1.5 million per year in total costs for four years will be set aside for this initiative. It is anticipated that only one award will be made.

RFA: Technology Development for the Exhaustive Determination of Functional Elements in the Human Genome

As mentioned above, NHGRI issued RFAs in 2003 and 2004 to support technology development for new and improved genome-wide methods to find functional elements, and funded six grants from each of those initiatives. Novel high-throughput technologies for the comprehensive identification of functional sequence elements continue to be needed. Although NHGRI supports a small number investigator-

initiated projects in this area, now is the time to further stimulate research projects in this area based on knowledge gained from the ENCODE Project thus far. NHGRI proposes, therefore, to issue an RFA to solicit applications for research projects to develop robust methods to identify comprehensively all sequence elements of a given type in the human genome. Both “wet bench” and computational approaches will be eligible. The intent of this solicitation is to support the development of novel approaches 1) to identify poorly characterized or previously unknown elements, 2) to identify functional elements in repetitive sequences, and 3) to validate the identity of functional elements using methods independent of the primary mode of detection.

Mechanism of Support: The mechanisms of support will be the R01 research project and the R21 pilot project grant awards. The total project period for applications submitted in response to this RFA may be three years for R01s and two years for R21s. NHGRI plans to set aside \$2 million per year for three years for this RFA with a maximum of \$350,000 direct costs per year for R01 applications and \$100,000 direct costs per year for R21 applications.