National Advisory Council for Human Genome Research

Concept Clearances for Functional Genomics

May 18, 2015

Overview

We propose the following four concepts based on the discussions at the March 10-11, 2015, NHGRI planning workshop: *From Genome Function to Biomedical Insights: ENCODE and Beyond*. The proposed initiatives are designed to meet the future goals of functional genomics to:

- expand the catalog of candidate functional elements;
- move beyond cataloging towards understanding the functional roles of genomic elements in specific biological contexts;
- develop strategies to apply these studies to disease;
- increase the number of scientists from the research community contributing to the creation of the encyclopedia of functional elements;
- develop analytical tools to enhance the utility of the data; and
- make the data, tools, analyses and assembled encyclopedia freely available to the research community.

This body of work will enhance our understanding of the role of genetic variation in human biology and disease.

1. **Functional Element Mapping Centers** – Purpose: to expand the catalog of candidate functional elements in the human and mouse genomes. **New goals include**: mapping disease samples to identify new elements, to identify new combinations of elements used in disease states, and to learn how to enable disease-oriented studies in the future; incorporating mapping samples provided by experts outside of the consortium; and developing a scientific basis for bounding the experimental space to be studied.
   ($20M TC/year; four years; 6-8 awards)
2. **Functional Element Characterization Centers (*NEW*) –** Purpose: to enhance the catalog of candidate functional elements by creating a set of well-characterized and validated functional elements in healthy and disease states, to investigate the role of genetic variation in human disease, and to obtain an understanding of the utility and generalizability of functional characterization approaches. ($5.9M TC/year; four years; 7-10 awards)
3. **Computational Analysis Research Projects** – Purpose: to maximize the utility of ENCODE while simultaneously engaging more of the research community in close collaboration, by bringing in additional computational analysis expertise to develop and apply computational and statistical methods to ENCODE data.
   ($3.0M TC/year; four years; 5-8 awards)
4. **ENCODE Data Coordination and Analysis Center (EDCAC)** – Purpose: to support an expanded ENCODE Data Coordination and Analysis Center to continue to provide community access to ENCODE data and resources and support analysis of ENCODE data necessary to create and make available a high quality encyclopedia; and to support **new activities** to enable direct submissions of data and metadata from the research community into the ENCODE portal, to support analysis of ENCODE and other data to develop a scientifically-based strategy for bounding the cell space to be studied, and to organize and facilitate Consortium activities.
   ($7.5M/year; five years; 2 awards)

Background

Creating a high quality, well-validated, and comprehensive encyclopedia of functional elements in the human and mouse genomes is central to meeting NHGRI's long term goals of enabling genomic and precision medicine. This fundamental resource is essential for interpreting genome sequences and understanding the consequences of sequence variation. ENCODE (Encyclopedia of DNA Elements) and its related model organism projects have identified many candidate functional elements in the human, mouse, worm and fly genomes and the data are made freely available to the research community in a rapid, pre-publication manner. The encyclopedia adds layers of predictive interpretation to the ENCODE data to increase the understanding of the functional role of these candidate genomic elements. (For more

information see https://www.encodeproject.org.) ENCODE has also supported the development of technologies and methods to generate, characterize and analyze functional genomics data. ENCODE data are being used by researchers for studies of the genetic basis of disease, basic biology and methods development; to date community researchers (those without ENCODE funding) have published more than 950 papers using ENCODE data and tools.

Funding for the current phase of ENCODE will end in July 2016. NHGRI convened a planning workshop "*From Genome Function to Biomedical Insight: ENCODE and Beyond*" on March 10-11, 2015, to consider what projects in functional genomics should be supported that build on the work of ENCODE. A workshop report summarizing the discussions and recommendations will be provided to the NACHGR as background prior to the May Council meeting.

It was clear from the workshop discussions that much work still remains to complete the encyclopedia and to maximize its utility. Many functional elements are manifest only in restricted cell contexts, yet only a small fraction of cell contexts (cell types, fates, states) have been studied to date. Since it is impractical to consider interrogation of every possibly cell context, efforts are needed to develop a scientifically-based strategy to bound the cell space to be examined, such that those cell states that are examined will sufficiently represent the universe of possibilities. Further, only a tiny fraction of the candidate functional elements have been experimentally characterized with respect to specific biological function(s). In addition, taxonomy used to define different classes of functional elements is still rudimentary, and further work is needed to refine these classes to provide a more detailed annotation of the genome. The emergence of new technologies, e.g., those for genome editing, provide an unprecedented opportunity for pursuing these activities in a cost-effective and high-throughput manner. It was also recognized at the workshop that there is a need to bring into this effort scientific expertise in specific biological systems, both to expand the sources of cells for analysis and to contribute relevant data. Functional genomics studies on samples related to disease were encouraged to expand the encyclopedia while simultaneously providing biological insights about the specific diseases and enabling the development of general strategies for applying these assays to disease. Finally, to meet the long term goals of functional genomics, continued technology development is needed.

The four initiatives described in these Concepts will address the future goals of functional genomics, moving beyond the current goals of ENCODE. Although it is clear that further technology development is needed, NHGRI will incorporate that into a broader technology development initiative because these technologies will have utility and implications significantly beyond the goals described here.

What's new:

These initiatives differ from the current ENCODE program in several key ways:

- Functional Element Characterization Centers (Initiative 2) is a new activity targeted at validating and characterizing candidate functional elements, which is currently only a very small part of the ENCODE effort. Importantly, this initiative will be directed at providing an understanding of the utility and generalizability of functional characterization approaches.
- Direct application of functional genomics assays to disease studies is a new activity. Both the Functional Element Mapping Centers and the Functional Element Characterization Centers (Initiatives 1 and 2) will incorporate disease samples. An explicit goal is to develop and test strategies for directing functional genomics data collection and analysis to enable understanding of the genetic basis of disease.
- Significantly increased community participation will be an important new element in Initiatives 1, 2, and 4.
    - Functional Element Mapping Centers (Initiative 1) will profile samples provided by experts in biology and/or human disease to take advantage of widely dispersed expertise while maintaining uniform data collection and processing.
    - Functional Element Characterization Centers (Initiative 2) will include experts in biology and/or human disease to provide input on candidate elements and cell sources for study.
    - ENCODE Data Coordination and Analysis Center (Initiative 4) will establish the standards and process to enable the community to submit samples for study under Initiatives 1 and 2 and also to directly submit data and metadata to ENCODE.
- Opportunities for co-funding by other NIH Institutes and Centers (ICs) to support projects focused on specific diseases or biology of interest will be encouraged by NHGRI for all initiatives.

- Projects can be funded wholly or in part by another IC; this is especially envisioned for Initiative 2 (Functional Element Characterization Centers) and possibly for Initiative 3 (Computational Analysis Research Projects) if the focus is on a particular disease or biological system of interest.
- Other ICs may contribute supplemental funding for studies of particular relevance to their mission, e.g., the study of specific disease tissues in Initiative 1 (Functional Element Mapping Centers), which would be one of a number of areas of emphasis for any given mapping center.
- Supplemental funding from other ICs could be provided to support submission of existing or planned datasets generated under other projects to the EDCAC (Initiative 4).
- Mechanisms for project selection, prioritization and co-funding will be established by NHGRI Program Staff.

Funding note

To optimize the mix of mapping assays, characterization assays, and biological systems/diseases, NHGRI will need the flexibility to build a balanced portfolio at the time of funding each RFA, which may entail funding across the score range.

1: <u>Functional Element Mapping Centers</u>

What's new: Mapping centers will engage in two new activities; mapping disease samples and working on community-provided samples at scale. All human samples to be assayed should be consented for open access data release.

Purpose

The purpose of this initiative is to expand the catalog of candidate functional elements in the human and mouse genomes, to provide a scientific basis for bounding the experimental space to be considered in creating this comprehensive resource, to conduct deep analyses of specific diseases that will inform how to direct disease-oriented studies in the long run, and to broaden the participation of researchers in creating this community resource.

Proposed Scope and Objectives

This initiative will support high-throughput pipelines of a range of assays capable of generating high quality data to map, comprehensively, biochemical activities exhibited by the human and mouse genomes that are associated with specific functional elements. The primary activity of the mapping centers will be unbiased data generation in an expanded number of cell contexts (cell types, fates, states) using multiple biochemical assays. Since many functional elements are manifest only in restricted cell contexts, many additional cell contexts will be needed to create a comprehensive catalog. Since there is a vast number of cell contexts, this initiative, in conjunction with activities supported within the Data Coordination and Analysis Center (see below), will support efforts to design and conduct experiments to provide a scientific basis for bounding the cell space to be thoroughly interrogated in the long run such that those cell states that are examined will sufficiently represent the universe of possibilities. Some samples will be provided by biology and disease experts from the community to enable the study of specific cell lineages, biologically relevant conditions and diseases. Approximately 25% of the production effort will be devoted to thorough interrogation of a range of diseases to serve as examples of how to apply these biochemical assays to disease, and in turn, to inform how disease studies could be conducted over the longer term. All human samples to be assayed should be consented for open access data sharing. New and improved methods will be applied at high-throughput to identify new elements or overcome previous technical limitations, for example, using smaller number of cells or single cells, or providing higher resolution or otherwise more informative data. Depending on the capabilities of the different assays, they may be applied across a relatively small set of common samples or, if possible, across a wide range of cell types.

The specific components of this initiative will include:

- Increased interrogation of cell space using assays/data types that have been widely applied in ENCODE plus newly-developed assays.
- Coordinated efforts to generate experimental data to provide a scientific basis for bounding the cell space to be thoroughly interrogated in the long run.
- Study of disease-related samples. Approximately 25% of the mapping center pipelines will be devoted to the study of samples from a range of diseases. These samples may be specialized cells of high relevance to a disease from unaffected individuals if not previously studied, and from affected individuals. Data from these cell sources will simultaneously help build the catalog of functional elements and inform about differences between normal and affected samples. Information from what is learned about disease from these studies will be used to develop a general strategy to apply these approaches to disease in the long run. Beyond the core disease-focused efforts funded through these mapping centers, additional disease-focused projects can be supported by collaborations with other ICs through supplemental funding.
- Exclusive use of samples consented for open access data release.
- Outreach to the research community to provide unique cell samples to mapping center pipelines.
    - Samples could be for unbiased data generation, biology focused (e.g., developmental or differentiation pathway) or disease-relevant tissues; all will help fill in the catalog.
    - Opportunity for co-funding from other ICs to add disease-relevant tissues for deep investigations of particular interest.
- Employing new assays that are highly informative but for which throughput may be more limited, on a matrix of common cell types for which significant amounts of ENCODE data already exist.

- Mapping binding sites of all known transcription factors and RNA binding proteins in at least two human cell types.
- Studies of the mouse genome focusing exclusively on adult tissues since much work will have already been conducted in embryonic development. Mouse studies will constitute approximately 10% of the overall data production effort. These data will be used primarily to further understand the human genome.
- Flexibility in the overall production pipeline built into the funding plan in order to adjust to the rapidly changing landscape. For example,
    - Scale of application of new methods, starting with common cell space then broadening out as feasibility is demonstrated.
    - Bringing in new assays over time.
    - New opportunities brought in by collaborators.

Applicants should propose which specific samples they will use in the first 18 months of the project, including the ability to share those samples with other funded groups. Once awards are made, the groups will work together to prioritize samples to be studied across all assays and to develop a process for obtaining new samples from collaborations with the broader research community.

Relationship to Ongoing Activities

The proposed initiative will be unique in its scientific exploration of the universe of non-coding DNA elements and in its emphasis of identifying those elements most informative to studies of disease and basic biology.

Related Projects

International Human Epigenome Consortium (IHEC) and Reference Epigenome Mapping Centers (REMCs) both provide resources of human epigenomic data, but the proposed program is unique in its focus on mapping transcription factor and RNA binding protein binding sites, and on extracting information from epigenomic data to identify putative non-coding DNA elements and assign them to probable functional categories. Funding for the REMCs has ended.

PsychENCODE aims to identify functional non-coding elements in human brain, towards elucidating their role in mental disorders. PsychENCODE may help demonstrate how targeted epigenomic tissue profiling can aide functional element identification and understanding.

Library of Integrated Network-based Cellular Signatures (LINCS) aims to computationally derive cellular signatures indicative of chemical or environmental perturbations, but does not catalog non-coding elements or explore their contributions to cellular signatures. In contrast, the proposed initiative focuses on identification of non-coding elements; it could leverage information gained from LINCS for targeted exploration of the relationships between perturbations and non-coding element activity and function.

Genomics of Gene Regulation (GGR) aims to elucidate a network-based understanding of gene regulation. This involves targeted collection of epigenomic data that could serve to inform small, specific sections of an immense epigenomic data matrix that will be explored much more thoroughly and systematically by the proposed initiative.

Genotype Tissue Expression (GTEx) collects gene expression data for ~30 tissues across hundreds of individuals, along with whole-genome sequencing data, to help elucidate relationships between genetic variation and tissue-specific gene expression. Additionally, the enhanced GTEx (eGTEx) project will collect epigenomic data, similar to some of the data proposed to be generated by Functional Element Mapping Centers, from a small number of individuals. Together these resources will provide essential context for understanding how genomic variation impacts gene expression and ultimately human health and disease.

Mechanism of Support

As in the current and previous phases of ENCODE, large-scale data production will be supported by U54 Centers. A cooperative agreement (“U”) funding mechanism will be required to provide the degree of staff involvement needed for

coordination, management and flexibility with respect to collaboration within the Consortium and with outside collaborators.

Funding Anticipated
$20M TC/year; four years (FY16-19); 6-8 awards

We anticipate funding 6-8 mapping centers in order to support the range of technologies needed to map functional elements in the human and mouse genomes. Centers may use more than one experimental assay. Sufficient funds are needed to establish high-throughput, cost-effective experimental and computational pipelines, and to have sufficient personnel to interact effectively with the community.

2: Functional Element Characterization Centers

What's new: This is a new activity. These centers will characterize elements that are relevant to the projects they propose (75% effort) as well as common elements studied by all characterization centers (25% effort). Projects will focus on characterizing elements within specific biological contexts and can focus on specific diseases. All human samples to be assayed should be consented for open access data release.

Purpose

The purpose of this initiative is to enhance the catalog of candidate functional elements by creating a set of well-characterized and validated functional elements under healthy and disease states, and to obtain an understanding of the utility and generalizability of functional characterization approaches.

Proposed Scope and Objectives

The objective is to employ multiple approaches to test candidate elements in specific biological contexts to validate and characterize functional activities. Functional characterization and validation of candidate functional elements is an inherently broad endeavor given multiple scales of function (e.g., molecular, organismal), the dependence of function on context (e.g., cell fate/state, organismal environment, disease) and the variety of functional elements (e.g., enhancer, promoter, splice site). Multiple approaches are likely needed because of their different abilities with respect to throughput, scale of function and types of function. Because of the need for highly specialized expertise in biological systems, this initiative lends itself well to participation by multiple groups and to collaborations with other ICs.

The focus will be to develop and test strategies that could be more broadly applied by the community, and to begin to ask how to make better predictions of function, rather than to cover the entire catalog of candidate functional elements in all cell contexts. Applicants will define the cell systems/context for in-depth studies. Studies of specific diseases using relevant cell sources are especially encouraged and information that is learned from applying different functional assays in disease studies will be used to develop a general strategy to apply these approaches to disease in the long run. Opportunities for co-funding and collaboration with other ICs are possible, especially when the focus is in a particular cell type/system/disease.

Features of Potential Projects:

- Projects can use high-throughput assays to broadly survey a significant number of candidate functional elements (e.g., massively parallel reporter assays) or use lower throughput assays to characterize candidate functional elements in more physiological systems (e.g., transgenic animals).
- Both human and mouse candidate functional elements can be characterized.
- Functional characterization assays can be conducted in a model organism(s) other than mouse if well-justified.
- Projects can be focused on specific biological questions, including disease.
    - Assays can be performed in cell sources from individuals with a specific disease and unaffected individuals to inform how elements may work differently in normal versus disease states.
    - Projects can focus on investigation of a biologically defined sub-set of candidate functional elements using multiple approaches.
- All human samples to be assayed should be consented for open access data release.
- To facilitate comparison between approaches, all characterization centers will reserve 25% of their capacity for analysis of a common set of candidate functional elements across projects. Coordination of common cell types and assays will be led by awardees with significant program involvement after projects are funded.
- All groups will be required to use existing ENCODE data, and can use data from other projects and anticipated new ENCODE data, as is feasible.
- Collaborations with other ICs/projects for deep interrogations into specific biological systems, cell types, diseases, networks, developmental/differentiation pathways are encouraged and will be facilitated by NHGRI.
- Data will be submitted to the EDCAC (see Initiative #4), which will store and provide access to files, but not necessarily conduct the same quality control and processing as for the data generated under Initiative #1 due to the anticipated non-standard nature of the data.
- Data will be used to enhance the quality of the encyclopedia.
- Data will be analyzed by awardees and the EDCAC to help determine how to better predict function.

- Awardees will participate in Consortium activities.

Relationship to Ongoing Activities
The proposed program will develop and test strategies for characterizing and validating functions of genomic elements, in the context of investigator-selected disease or model systems. Unlike Initiative 1, which aims to identify and catalog functional elements, this initiative will focus on developing and testing approaches for understanding element function and how it impacts biology and disease.

While the current phase of ENCODE has supported pilot functional characterization and validation efforts on a very limited scale, this initiative represents a major shift and will occupy a largely unexplored conceptual space.

Currently, no other coordinated effort exists to characterize (as opposed to catalog) non-coding element function. Much research in this arena has been investigator-initiated; the proposed program will enable studies by the broader research community by providing paradigms and use-cases for element validation and characterization methods.

The projects conceptually most similar to this proposed initiative include GGR and "FunVar" (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-13-013.html). However, those focus on computational elucidation of gene regulatory networks and non-coding variant interpretation, not experimental characterization and validation of element function as proposed here.

Mechanism of Support
U01s. A cooperative agreement ("U") funding mechanism will be required to provide the degree of staff involvement needed for coordination and management with within the Consortium and coordination with outside collaborators.

Funding Anticipated
$5.9M TC/year; four years; 7-10 awards (FY16-19)

Support 7-10 projects ($0.5M - $1M each, depending on scope/assay)
Opportunity for co-funding by other ICs for deep studies of specific biological systems and/or diseases.

3: Computational Analysis Research Projects

Purpose
The goals of this initiative are to maximize the utility of ENCODE and related functional genomics resources while simultaneously engaging more of the research community in close collaboration, by bringing in additional computational analysis expertise to develop and apply computational and statistical methods to the analysis of ENCODE data.

Proposed Scope and Objectives
This initiative will support investigator-initiated projects to develop and apply analytical and statistical tools to use and improve the ENCODE and related functional genomics resources. These activities can include developing new methods to improve on analysis, visualization and interpretation of ENCODE data, combining ENCODE data with related functional genomic data from other projects to derive new biological insights, or using the ENCODE data to improve on the analysis of disease mapping studies to identify causal variants. This broad scope will best be addressed by investigator-initiated projects.

Specific examples of projects that could be supported include:

- novel imputation methods to identify information-rich cell states/fates for experimental prioritization
- broadening element 'lexicon' through classification of novel functional subtypes and elucidating impacts on phenotype and disease
- predictive modeling of key gene/element networks and their contributions to disease states
- tools for integrative analyses employing disparate data or tissue types
- computational models to predict element function or gene-element relationships
- software for visualization or analysis of high-dimensionality data

Work supported under this initiative will be primarily computational. Limited data generation for the purposes of testing, validating and refining the methods will be allowed.

Awardees will be affiliated with ENCODE and will be encouraged to participate in Consortium activities but will not be voting members of the Consortium.

Co-funding by other ICs to support projects focused on the analysis of specific diseases or biology of interest will be encouraged by NHGRI.

Relationship to Ongoing Activities
The current phase of ENCODE supports several computational analysis groups which have pioneered methods for analysis and interpretation of ENCODE data. Because of the continued need for new and improved computational tools, the proposed program seeks to build on this success by funding a number of investigator-initiated projects that will develop novel methods for the research community to analyze, interpret, and interact with data arising from ENCODE and other community efforts.

Mechanism of Support
U01s. The "U" cooperative agreement funding mechanism is proposed because substantial NHGRI programmatic involvement is anticipated to facilitate close interactions between these awardees and those supported under the other proposed initiatives.

Funding Anticipated
$3.0M TC/year; 4-6 projects; four years (FY16-19)

Budget Justification
A critical mass of investigators is needed to support a range of projects. Each award needs a sufficient budget to develop methods and generate experimental data for validation of these methods.

4: <u>ENCODE Data Coordination and Analysis Center (EDCAC)</u>

What's new: The EDCAC will integrate functional characterization data with mapping data, enable direct submissions of data and metadata from the research community into the ENCODE portal, support analysis of ENCODE and other data to develop a scientifically-based strategy for bounding the cell space to be studied, and organize and facilitate Consortium activities.

Purpose

The purpose of this initiative is to support an ENCODE Data Coordination and Analysis Center (EDCAC) to provide community access to ENCODE and related functional genomics data and resources, to support analysis of ENCODE data necessary to create and make available a high quality encyclopedia, to support analysis of ENCODE and other data to develop a scientifically-based strategy for bounding the cell space to be studied, and to organize and facilitate Consortium activities.

Proposed Scope and Objectives

The EDCAC will consist of two components, a Data Coordination Center (DCC) and a Data Analysis Center (DAC). These will be solicited and funded separately (under the same FOA) due to the required differences in expertise, but they are expected to regularly communicate and coordinate their activities so as to function as a single entity, as is the case in the current EDCAC organization. The DCC activities will be significantly expanded relative to the current scope, while the DAC activities will be more tightly focused on creating and validating the encyclopedia and have a reduced focus on integrative analyses. While many of the activities will require the coordinated efforts of both the DCC and the DAC, each will have the lead in specific areas.

The DCC-led activities will include:

- Developing, housing, and maintaining databases to track, store, and provide access to the data, metadata and computational tools generated by the ENCODE Project (past and future).
- Developing, maintaining, and updating data processing pipelines.
- Maintaining and enhancing the ENCODE portal to ensure easy access to ENCODE data and resources.
- Providing community access to ENCODE data in state-of-the-art browser/visualization formats.
- Tracking and reporting on data submission and providing strategic data-based prioritization of areas in need of data generation.
- Importing data from investigators and projects outside of ENCODE, developing the needed infrastructure and methodology.
    - Data will need to meet certain quality and quantity thresholds.
    - Investigators will need to provide required level of bioinformatics expertise to interface with DCC.
    - Other ICs can provide co-funding for bringing in data from IC-supported large projects.
- Serving as the ENCODE Consortium's Coordinating Center to:
    - Facilitate communication and coordination within the Consortium
    - Organize and support annual Consortium meetings
    - Organize and support Consortium working groups
    - Support outreach activities to promote broad use of data, analyses and tools

The DAC-led activities will include:

- Specifying and updating data processing pipelines.
- Providing leadership and computational expertise to update and refine the encyclopedia, the major ENCODE deliverable to the research community.
- Providing leadership and computational expertise to analyze data from Initiative 1 and other available data to develop a scientific strategy to bound the cell space to be deeply interrogated.
- Providing leadership and computational expertise to analyze data from Initiatives 1 and 2 and analyses from Initiative 3, if relevant, to expand the lexicon of functional elements beyond the basic categories (e.g., promoters and enhancers).
- Providing leadership and expertise to update existing and develop new data quality metrics and standards for all data types.

- Identifying and bringing in additional analytical expertise as needed by the Consortium, and having reserved funds to support short-term, specialized analyses needed to maximize the quality and utility of the ENCODE resource.

Additional analysis activities, including methods to derive insights into element function and linkages to biological or disease processes will not be supported by this initiative, but may be appropriate for Initiative #3.

Relationship to Ongoing Activities
The current phase of ENCODE has a Data Coordinating Center (DCC) tasked with providing community access to ENCODE data, and a separate Data Analysis Center (DAC) that functions to establish guidelines for analysis of ENCODE data. These two centers work together as a single ENCODE Data Coordination and Analysis Center (EDCAC).

The proposed program will continue the current DCC's essential mission of providing community access to ENCODE data. Given the new emphasis on taking in data from community sources, increased capacity relative to current DCC activities will be required. Currently, the ENCODE DCC is incorporating data from several large community resource projects, i.e., modENCODE, REMC and GGR.

The proposed program will also incorporate essential activities of the current DAC, including specifying data analysis pipelines, data quality metrics, and curating the encyclopedia.

Mechanism of Support
The U41 mechanism will be used, as in the past. The "U" cooperative agreement funding mechanism is needed to allow for significant staff involvement in directing activities.

Funding Anticipated
$7.5 M; 2 projects; five years (FY16-20 - one year beyond the proposed data production period)

Budget justification: Current level of support of the EDCAC is $2M for the DAC and $2.7M for the DCC, for a total of $4.7M Total Costs per year. The current DCC has been under-funded to conduct its activities. It has or will receive several supplements to import data from other projects. The EDCAC activities will be greatly expanded to include integration, analysis and display of a more sophisticated and user friendly encyclopedia, creating an infrastructure to bring in community data (from many groups, not just a few large projects like REMC or modENCODE), and funding coordination activities beyond the current DCC's efforts. These additional activities include organizing and paying for an annual Consortium meeting with potentially >200 individuals, organizing working groups for multiple activities, such as establishing and updating quality metrics and data standards, setting agendas for Consortium and working group calls, and expanded outreach activities such as annual User's meetings and other tutorials.

In addition, $2.6M (75% current level) total costs in FY 16 funds are needed for an additional year of support for the existing ENCODE Data Coordination Center (DCC) to wrap up essential ENCODE 3 activities. This extension was suggested by and strongly endorsed by the ENCODE External Consultants Panel.

Summary of Request

Activity	FY16	FY17	FY18	FY19	FY20
Functional Element Mapping Centers (U54)	$20M	$20M	$20M	$20M	-0-
Functional Element Characterization Centers (U01)	$5.9M	$5.9M	$5.9M	$5.9M	-0-
Computational Analysis (U01)	$3.0M	$3.0M	$3.0M	$3.0M	-0-
ENCODE Data Coordination and Analysis Center* (U41)	$7.5M	$7.5M	$7.5M	$7.5M	$7.5M
New Initiatives Total	**$36.4M**	**$36.4M**	**$36.4M**	**$36.4M**	**$7.5M**
DCC 1-yr extension**	$2.6M	-0-	-0-	-0-	-0-
Total	**$39.0M**	**$36.4M**	**$36.4M**	**$36.4M**	**$7.5M**

*Build in extra year for DCC to take in and process data and analyses coming in at the end of the data production phase.

**$2.6M (75% current level) TC in FY 16 funds are needed for an additional year of support for the existing ENCODE Data Coordination Center (DCC) to wrap up essential ENCODE 3 activities. This extension was suggested by and strongly endorsed by the ENCODE External Consultants Panel.