2014/2015 Genome Sequencing Program Concepts

February 2015 Meeting of the National Advisory Council for Human Genome Research

OVERVIEW

The following concepts are based on the discussions at the July 28-29, 2014, NHGRI workshop: *Future Opportunities for Genome Sequencing and Beyond*.

At the September 8, 2014, meeting of the National Advisory Council for Human Genome Research (NACHGR), the Council approved Concept Clearances for Concepts I and II, which are included here in abridged form only for context.

Council requested changes to the previously-proposed Concept for a Genome Sequencing Program Coordinating Center, including splitting it into two concepts (here, III and IV) which will be considered during the February 2015 NACHGR meeting. In addition, at the February 2015 NACHGR meeting, we will request approval for two additional concepts related to the July 2014 workshop discussions.

I. **ALREADY APPROVED**: Centers for Common Disease Genomics (CCDG) —Purpose: to identify variants contributing to common diseases; to develop the means to do so comprehensively; to explore a range of example disease architectures and study designs; and to develop resources for multiple disease research communities and the wider biomedical research community.

Mechanism: Cooperative agreements; $60M of NHGRI funding per year for four years; seek funding partnerships to increase the number of example diseases studied; November 2015 funding start date.

Released: RFA-HG-015-001 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-15-001.html)

II. **ALREADY APPROVED**: Centers for Mendelian Genomics (CMG)—Purpose: to identify the genomic bases of at least ~300 Mendelian disorders that represent a broad range of phenotypes; to understand the genomic characteristics of Mendelian disorders as a class; to learn what it will take to identify the genomic basis for all Mendelian disorders; and to

develop and disseminate resources, methods, and tools to lay a foundation for finding the variants underlying all Mendelian disorders.

Mechanism: Cooperative agreements; $10M of NHGRI funding per year for four years; seeking co-funding from NHLBI to increase number of disorders studied to at least ~400; November 2015 funding start date.

Released: RFA-HG-015-002 (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-15-002.html)

Concepts III-VI are detailed in separate sections below

Fiscal Year 2015 estimated NHGRI Genome Sequencing and Related Program Funding

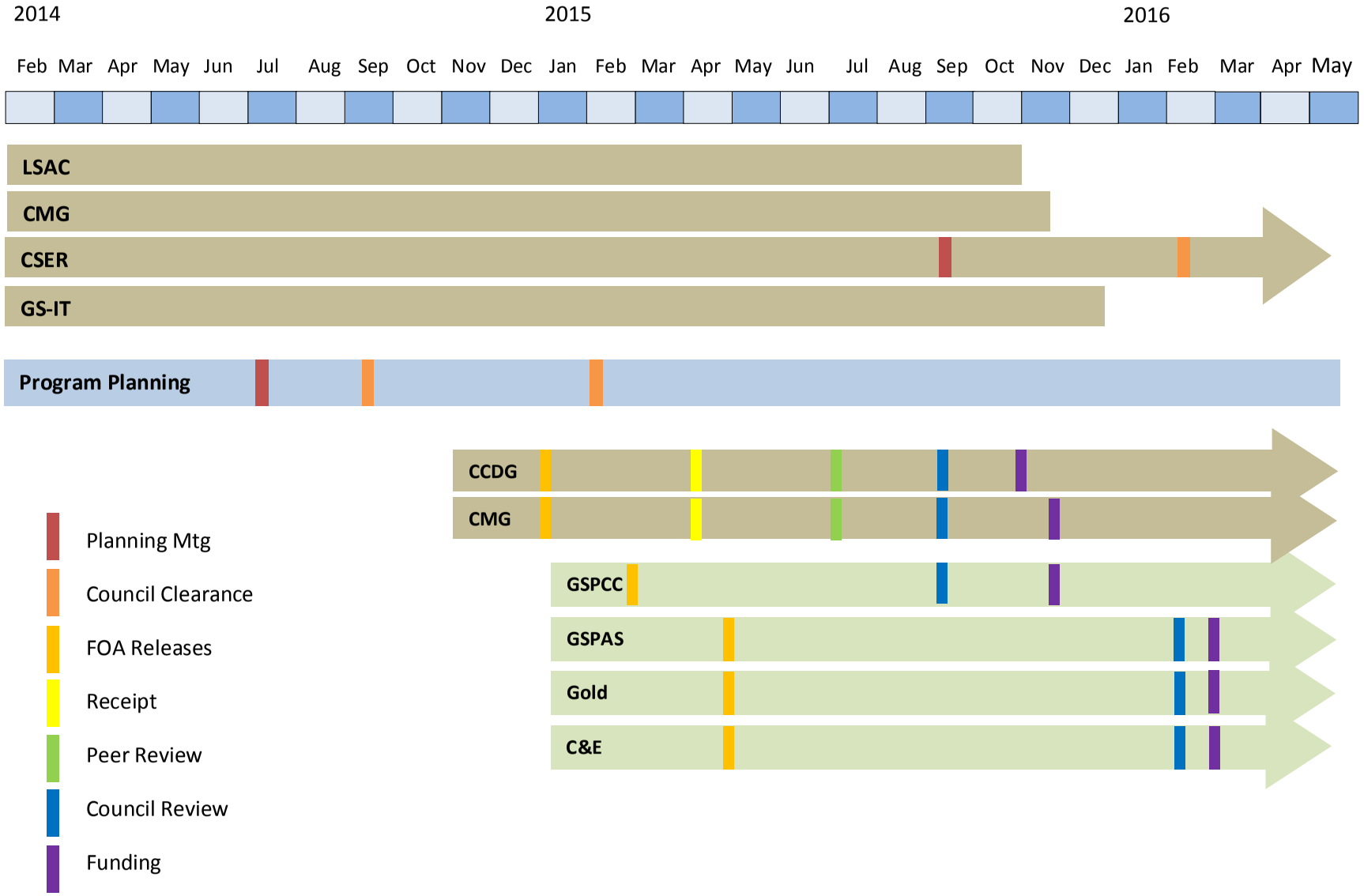
Program	$M per year (co-funding)	Cumulative FY 15 total (co-funding)
LSAC	**73**	**73**
CMG	**10 (+2)**	***83*** **(+2)**
With additional current GSP components:		
CSER	**15 (+2.7)**	**100 (+2.7)**
GS-IT	**4**	**102.3 (4.7)**

Estimated Year 1 (Fiscal Year 2016) NHGRI Funding of Proposed Concepts. All proposed funding is contingent on receipt of high-quality applications and availability of funds.

	Program	HG $M per year	Cumulative FY 16 Total
I.	**CCGD**	**60**	**60**
II.	**CMG**	**10**	**70**
III.	**Genome Seq. Analysis Satellites**	**3**	**73**
IV.	**Program Coordinating Center**	**1**	**74**
V.	**Gold Genome Production**	**2**	**76**
VI.	**Comparative and Evo. Genomics**	**(2)**	**(78)**

Please note that the Comparative and Evolutionary Genomics component is being proposed as a PAR, and so does not require set-aside funds.

With extension of CSER phase I sites:		
CSER	**17.7**	**95.7**


2014
2015
2016
Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May
LSAC
CMG
CSER
GS-IT
Program Planning
CCDG
CMG
GSPCC
GSPAS
Gold
C&E
Planning Mtg
Council Clearance
FOA Releases
Receipt
Peer Review
Council Review
Funding

Background documents relevant to these concepts:

- **Report of the July 28-29, 2014 Workshop "The Future of Genome Sequencing and Beyond"; video of the workshop is available at http://www.genome.gov/27558042**
- **Presentations about the current Genome Sequencing Program: http://www.genome.gov/27556133 has links to presentations about the LSAC program from the February 2014 NACHGR meeting; http://www.genome.gov/27557112 has links to the presentations about the CMG and CSER programs from the May 2014 NACHGR meeting**

III. Genome Sequencing Program Analysis Satellites (GSPAS)

Purpose

The GSPAS will conceive of and carry out innovative, creative analyses of the data produced by the Centers for Common Disease Genomics (CCDG) and Centers for Mendelian Genomics (CMG), especially analyses that will cut across individual projects, grants, or programs. The GSPAS will have considerably more freedom to define and propose their own analyses than will the other program components: for example, they will be encouraged to propose analyses that integrate data not just from the GSP sequence production groups (the CCDGs and CMGs), but also data from other NHGRI-funded or –related programs, including ENCODE, GTEx, eMERGE, and others, to understand variant function, leading to an increased likelihood of creative analyses on what will be an unprecedentedly large and diverse human sequence dataset. In addition, the GSPAS will work closely with the other GSP components, including the coordinating center, towards two specific overarching program goals: defining when a large common disease project is finished, and understanding the specifications for a set of common controls for such studies. The GSPAS may also assist with additional cross-cutting analyses that are identified as being of high priority for the program over time.

The GSPAS will be coordinated with the other program components and the projects that they are undertaking their results can quickly feed back to the production and coordinating centers.

GSPAS awards will be made to investigators outside of the data production centers.

Background

In early 2016, NHGRI will renew, with modifications, its Genome Sequencing Program (GSP), by funding substantial programs aimed at identifying and understanding the genomic variants underlying inherited disease. The CCDG will focus on rare variants underlying common, complex disease, with the aim of understanding general principles of common disease architecture. The CMG will focus on Mendelian disorders, with the aim of identifying causal variants for as many conditions as possible, and also developing more general conclusions about how to do this efficiently for all Mendelian conditions. Although both of these efforts will include extensive analyses, the volume and type of data produced represent an excellent opportunity for genome analysis investigators to be part of a larger, collaborative program (research network), and to propose creative analyses towards achieving the overall program goals, especially analyses that require the integration of data across different projects, or even across different programs (e.g., not just genome sequence data but also functional data). Moreover, at a higher level, such a program would represent a means to ensure that the data produced by the NHGRI GSP are rapidly subjected to creative analyses by those outside large production centers. This opportunity was recognized and encouraged by the NACHGR at its September 2014 discussion of concept clearances for the NHGRI GSP.

Proposed scope and objectives

Because the aim of the GSPAS program is to encourage creative computational analyses, it is appropriate to allow a wide scope for analyses to be proposed by applicants to this program. This includes not just the genome sequence data produced by the production centers, but any available data including, for example, data available from ENCODE, GTEx, eMERGE, and other NHGRI-funded initiatives. However, these investigators will also be operating in the context of a research network, and are also expected to help to address overarching challenges specific to the program that arise over time. Two that we can identify at the outset of the program are: determining when a large common disease project should be considered to be complete, and determining the specifications for a set of common controls for complex disease sequencing studies. Thus each GSPAS will work both on creative, outward-looking analyses (first two bullet points below), and also analyses that are directed towards some explicit goals of the overall program:

- NHGRI seeks to stimulate improved or novel analyses for all non-automated aspects of characterizing sequence variants in the data, after variant calling.
- There is particular interest in questions about association, analyses using existing functional data to leverage associations and/or make functional inferences; means to incorporate gene-by-environment or longitudinal data, means to improve study design to increase power; and other higher level analyses. These are only examples and not an exhaustive list.
- NHGRI is also interested in questions that span disease architectures; these include determining when a common disease study is “comprehensive” or complete; and characterization and specification of sample sets that could serve as common controls across many different diseases and architectures. These are questions that are also intrinsic to the goals of the proposed Genome Sequencing Program Coordinating Center (see below).
- The GSPAS would be encouraged to identify, and in collaboration with the overall program, carry out, analyses that would bridge across multiple grantees within one program or multiple programs. These could include design of arrays based on consortium data, quality assessments, and allele frequency analysis.
- The GSPAS will also be encouraged to address analytical challenges that are identified by the research network over the duration of the program.

The GSPAS will be expected to work collaboratively with one another, and with the other GSP components to achieve the GSP program goals.

NHGRI will seek GSPAS that will focus on different areas of analysis (for their more creative, outward-looking component), for example, one might focus on questions about improving association analyses, while another might focus on improving functional inferences. This preference will be reflected in any FOA that is issued.

NHGRI prefers that the GSPAS will not focus on improved processes or analyses that are already well-defined and funded though other NHGRI programs.

Relationship with ongoing activities

NHGRI funds a number of human genome sequencing analysis activities through initiatives and investigator initiated R01's. This includes every program that produces genome sequence across two extramural divisions, as well as initiatives such as "Interpreting Variation in Human Non-Coding Genomic Regions Using Computational Approaches and Experimental Assessment" (http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-13-013.html) and analysis groups funded in conjunction with the 1000 Genomes Project (http://1000genomes.org). Moreover, NHGRI also supports a program aimed at making sequence analysis software tools more usable and robust (http://www.genome.gov/27546195), although this specific program will be ending in late 2015. The GSPAS would be distinct, however. First, it would be integrated with the CCDG and CMG programs directly, both in order to ensure that the analysis investigators have direct and immediate access to the data, but also are able, over time, to contribute directly to the objectives of the CMG and CCDG programs, including developing improved study designs. Second, as an element of the research network, GSPAS will be encouraged to pursue activities that are not already ongoing. In particular, NHGRI has multiple other venues to fund basic sequencing data processing, SNP calling, and even SV calling.

Mechanism of support

We propose a cooperative agreement (U) mechanism for four years to allow the requisite degree of integration among program elements. It will be part of the research consortium with the programs in Concepts I and II (and probably IV), and will be managed by NHGRI staff with the help of an external scientific panel.

Because one aim of the GSPAS program involves dissemination of data and analyses beyond the production centers, NHGRI will exclude from eligibility investigators who are named PI's or key personnel on CCDG or CMG awards. This aim will be reflected in the funding criteria. However, GSPAS investigators could also be eligible to be PIs for or key personnel on the Genome Sequencing Program Coordinating Center (concept IV).

Funds anticipated

We propose to provide $3M per year for four years, and expect to make 3-4 awards with NHGRI funds. We propose that this effort begin as close as possible to the same time as the CCDG, CMG, and GSPCC begin.

IV. Genome Sequencing Program Coordinating Center (GSPCC)

Purpose

The GSPCC will help NHGRI coordinate across GSP activities (CCDG, CMG, GSPAS), and facilitate cross-study activities to increase the integration and efficiency of the program as a whole. It will also support administrative, logistical, and certain outreach functions. The GSPCC will work collaboratively with the GSP investigators, investigators in other relevant NHGRI programs (e.g., CSER and eMERGE) as scientific opportunities arise, as well as with NHGRI staff to facilitate a comprehensive program of research to promote discovery of rare disease-related variants and elucidation of their potential causal role. The GSPCC will also participate in certain analysis activities, including working with other GSP investigators to develop the criteria for determining when a large common disease project is finished, develop common controls, and possibly other analyses that span multiple NHGRI-supported sequencing-oriented programs. However, the GSPCC will in general not be responsible for analyses of individual projects undertaken by the CCDG and CMG. The GSPCC will not necessarily be required to include the specific scientific expertise for all tasks described in the concept, but rather will need to understand them to the extent that they can help coordinate and facilitate specific expertise that exists within the sequencing program as a whole.

In general, the GSPCC will have a complementary but distinct role from the GSPAS and any analysis done by the CCDG and the CMG. Except as specifically noted below, the GSPCC role will be to identify, facilitate, and lead the accomplishment of objectives that arise out of the entire program. In doing so, it is expected that the GSPCC will work collaboratively with those with analytical expertise from within other funded GSP components. This role is designed to acknowledge the overlap in interests and capabilities of the different components, and also the need for leadership to accomplish specific overarching objectives.

Administrative activities would include the following for the CMG and CCDG components: tracking costs, production, project completion; developing Web resources to facilitate program-wide communication and interaction and inform the community about program activities; planning meetings and generating and distributing relevant documents; and arranging/coordinating activities related to outreach and interactions with research and disease communities.

Background

The new GSP, as a whole, has the potential to achieve certain objectives that cut across the entire program (or significant parts of it). Although we expect that each program grantee will have an interest in achieving these goals, and that their efforts will be included, it is important that there be leadership and responsibility for coordinating these efforts.

Administratively, the current NHGRI Genome Sequencing Program is large and complex, requiring tracking of production data and extensive logistical coordination of large projects (e.g., multiple weekly conference calls, project documents, sequencing program or project meetings). The new program will be even more complex, and is likely to exceed the capacity of NHGRI staff (who will increasingly need to shift effort to manage other features of the new GSP). Many large programs at NHGRI employ coordinating centers to help manage these tasks. Moreover coordinating centers are well-positioned to

facilitate outreach to help ensure that the products and activities of the program are known to the community.

The GSPCC will, of necessity, need to work closely with other program grantees and with NHGRI program staff. A key criterion for this role will be success in leading large, diverse, and dispersed consortia.

Proposed scope and objectives

The GSPCC will:

- Help coordinate and provide leadership for analyses that cut across multiple program components (CMG, CCDG, GSPAS). In the implementation of the CCDG, for example, it is likely that some large disease projects will be split among awardees. A successful GSPCC will work with GSP investigators and staff to help ensure that the results are integrated including helping to reconcile differences in analyses (for example, different variant call sets on the same data), or looking for ways to synergize across sample sets with similar phenotypes. The GSPCC also will contribute scientific expertise and leadership as needed to facilitate cross-study activities in areas such as policy compliance, harmonization of phenotype and exposure measures, and facilitating consensus with regard to implications of disease causality of identified variants.

- As an essential program component, the GSPCC will coordinate and facilitate two cross-cutting tasks that are high-priority for the CCDG and the program as a whole: 1. Facilitate the development of a consensus definition of, or criteria for, when a comprehensive common disease rare variant study is finished; and 2. Aid in and coordinate the planning and specifications for development of a set of common controls for common disease rare variant studies. Consideration of such universal controls will require analysis across all the WGS data produced by the overall program (including any produced by the CMGs) and outside of it. The scope of this task could potentially be large; we therefore aim to begin with tasking the GSPCC with developing design considerations for a broadly generalizable control set, based on the projects that will be undertaken by the production centers.

- As an essential component, the GSPCC will carry out integrated allele frequency calling, the results of which will be publicly released, as a high priority for the CMG and the program as a whole.

- Be encouraged to identify other analyses that would cut across multiple grantees within one program or multiple programs. These could include design of arrays based on consortium data, quality assessments, etc.

- Not necessarily be required to include the specific scientific expertise, but rather will need to understand the requisite tasks to the extent that they can help coordinate and facilitate specific expertise that exists within the program as a whole.

- Have a role in many studies and publications that are produced by the consortium as a whole, while noting that some individual studies may be the sole responsibility of a single production center or analysis group.

- Undertake some key administrative tasks, including working with NHGRI staff to track throughput, cost, data deposition, and project completion information; maintaining websites that provide information to the community about the activities of the individual programs, including causal variant information, project completion status, etc.; helping to coordinate large projects being undertaken by multiple program awardees; organizing conference calls and program meetings required to coordinate complex consortia.

- Assist in certain outreach activities including workshops between the program grantees and the community (e.g., for outreach or project selection; see Concept I); assist in summarizing and communicating to the community general "lessons learned" about how to use genome sequencing to find rare variants; and help ensure that project data are available to the wider community.

Relationship with ongoing activities

The CSER program has already established a coordinating center, which is closely integrated in consortium activities and provides outstanding coordination for activities related to actionability and return of results, psychosocial implications and ethical issues, and integration of results into clinical care. That coordinating center will continue together with the CSER program. The current iteration of the CMG program also has a coordinating center function which is carried out by one of the CMG sequencing center grantees. Some of its current responsibilities, particularly analyses that cut across data types, as discussed above, will be separated out and will be undertaken by the GSPCC. However, other functions must remain very tightly integrated within the future CMG program, in particular, the coordination of identification of samples, so this function will likely be the responsibility of the future CMG grantees.

The analysis functions of the GSPCC have potential to overlap those of the Analysis grantees (GSPAS), which will be coordinated and delineated in the writing of the FOAs for this program, and in an ongoing way over the course of the program. Initially, the GSPCC will be restricted to analyses that bear on project completion criteria, and development of common controls. The solicitation will emphasize that these activities will need to be done in concert with the other program components.

Mechanism of support

We propose a cooperative agreement (U) mechanism for four years. The GSPCC will need a high level of flexibility to adapt rapidly to new projects and consortia, and to anticipate the range of projects that will be undertaken by the successful applicants to Concepts I and II.

The GSPCC will be part of the genome sequencing program research consortium with the programs in Concepts I - III, and will be managed by NHGRI staff with the help of an external scientific panel.

Principal and key investigators who are recipients of CMG and CCDG awards will not be eligible for funding as the GSPCC.

Funds anticipated

We propose to provide a total of $1 M in the first year. This amount may need to grow in out-years if, for example, creation of a common control set is shown to be feasible.

V. High-Quality Human and Primate Genomes as Foundational Resources

Purpose

NHGRI seeks to stimulate rapid development of a resource of 25-50 "Gold-quality" human genomes (i.e., the quality of the current human genome reference). These will be chosen so as to add diversity to the haplotype representation in human genome references to increase their utility for research and clinical applications. A second goal will be to refine the genome assemblies of multiple nonhuman primate species to improve their utility to the research community. These two sets of high-quality genomes will be used somewhat differently, but will be achieved using similar technical approaches, so are bundled together in this concept.

Background

The human genome reference sequence (current version: GRCh38; see http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/) is a very widely used resource for basic and clinical genomics applications. The reference has evolved from the original International Human Genome Project sequence, but still has several shortcomings—it contains gaps, it represents structural variation incompletely, it is a mosaic of contributions from multiple individuals, and it mostly does not represent actual haplotypes over long contiguous regions. Moreover, the sequenced samples were not broadly representative of world populations. In the past several years, each of these shortcomings has begun to be addressed. This includes the ongoing effort to represent multiple alternate genomic haplotypes as "alternate paths" in the reference, but this is not comprehensive, and hence investigators often find that an individual they have sequenced has some regions that will not map onto the reference. Moreover, producing multiple, high-quality human assemblies is a key step towards a complete account of structural variation.

Separately, NHGRI and others have funded the genomic sequencing of multiple nonhuman primates based on their utility as experimental systems (including disease and drug studies), and also to address basic questions in comparative genomics. While the quality of the present genome assemblies is reasonably sufficient for many applications in comparative genomics, improving the quality of these genomes is needed to facilitate additional comparative analyses, such as analyses of lineage-specific structural rearrangements and in general to fully enable the use of these organisms in biomedical research. There are at least a dozen primate genome assemblies that are potentially high priorities (Chimp, Bonobo, Gorilla, Orangutan, Gibbon, Baboon, Rhesus, African Green Monkey/Vervet, Marmoset, Bushbaby, Tarsier, Tree Shrew).

The cost of producing a genome of high quality is approximately $250,000 with established technology, using a combination of efficient short-read approaches and more expensive long-read platforms. (Refining existing primate genomes may be less expensive, but will require similar technical approaches.) Newer technologies may arise over time that more efficiently generate Gold genomes, and

it is expected that these will be incorporated into this effort when they are robust. However, the demand for higher quality data for basic and clinical applications is immediate.

Scope and objectives

The scope and objectives are straightforward:

- Develop a specific end-point for "Gold" human genomes, starting with the current standards and balancing practical and scientific considerations to define a high-quality product. Current high-quality genomes, within the state of the art, have the following general characteristics: all sequence data from a single individual; majority of assembly is haplotype-resolved over long intervals; very high contiguity/quality through multiple (several hundred) specific regions of the genome known to be high in structural variation and significant for human disease. Develop a similar endpoint for "refined" non-human primate genomes, which may or may not be the same, but must be based on added scientific utility balanced against cost.
- Develop such high quality assemblies from up to 50 individual humans chosen so as to maximize utility. Diversity of the populations of origin with respect to the current human genome reference sequence and its uses will be an important scientific factor.
- For applications proposing to improve nonhuman primate genome sequence assemblies, identify up to 10 primate genomes for improvement to Gold quality to be justified by the applicant in terms of added scientific or biomedical value (but final selection will be negotiated with NHGRI after award).
- Remain at the forefront of technology in order to maximize quality/quantity of products.
- Individual applications may propose either or both objectives (human and non-human primate genome sequence improvement).

Relationship with other activities

The Gold genome has direct potential overlap with the NHGRI-funded Genome Reference Consortium award, which aims to produce at least five Gold genomes and two "Platinum" genomes (end-to-end contiguity). Any new awards in response to this Concept/FOA will need to be tightly coordinated with the GRC efforts. It will not be difficult to avoid direct overlap in sample choice.

Mechanism of Support

Cooperative agreements for resources (e.g., U41). One release date; one or two awards.

Funds anticipated

$2M/year for three years would be sufficient for roughly 20 Gold genomes, plus an additional 5-10 refined primate genomes. This is not sufficient without additional cost savings to reach the goal of 50 "Gold" human genomes. However, it is reasonable to anticipate that costs of, e.g., long sequence reads, will decrease over time, and/or that other funders can be identified.

VI. Comparative and Evolutionary Genomics

Purpose

This proposed program will fund creative, investigator-initiated projects in comparative and evolutionary genomic sequencing and analysis. Although the primary purpose is to address fundamental, significant scientific questions, the resulting data are also intended to become a long-term resource for the scientific community. Investigator-initiated applications to this program would be driven by major scientific themes and questions that can be addressed by comparative sequencing and analysis of multiple species. The program will encourage questions that are of wide importance, and of interest to multiple NIH institutes.

Background

Comparative genomic sequence data have been foundational for a number of investigations covering a very wide range of biological questions that cut across all of biological science. NHGRI has, in the past, funded many such projects through the NHGRI Large-Scale Genome Sequencing and Analysis Program.

More recently, NHGRI has not initiated new comparative genome projects because we reasoned that many remaining questions could be addressed by sequencing a few organisms, could be done with increasing efficiency by individual investigators, and were best selected through conventional, hypothesis-driven approaches, funded by diverse organizations. However, NHGRI has been recently advised that there are some comparative projects that are of a wide enough scope that they may not be attractive to other funders. Specifically, these are projects that entail sequencing and comparative analysis of multiple different species in order to inform basic biological questions (as listed in *Scope and Objectives,* below). Because of this, they are not as attractive to categorical institutes that may wish to focus on questions relating to a particular organ system or disease. Moreover, although sequencing costs have been significantly reduced over the last four years, these projects are usually somewhat larger than a typical R01 budget allows, making it harder for categorical institutes to see such projects as falling within their areas of interest.

Scope and objectives

NHGRI proposes to solicit applications for investigator-initiated comparative and evolutionary genomics projects that address fundamental questions about genome biology, evolution, and function, such as:

- Characterizing basic comparative features of genome structure leading to inferences about mechanisms of origin or function—e.g., ploidy, repeats, orthology, paralogy, loss and gain, horizontal transfer, mobile elements; also differences in evolutionary rates, selection, lineage-specific changes, etc.;

- Resolving conserved regulatory sequences in human and other sequenced genomes and inferences about function;
- Discovering true genomic innovation (e.g., origins of proteins, biochemical pathways, core metazoan developmental program, etc.);
- Defining the genomic basis of phenotypic innovations at different time-scales, ranging from major taxonomic innovations – e.g., multicellularity or adaptive immune system or lifespan -- to those occurring on a shorter time scale, for example, differences between closely related species or within species, including phenotypes of medical relevance such as disease resistance;
- Reconstructing ancestral genomes;
- Putting bounds on gene/protein sequence space; and
- Understanding the basis for speciation/hybridization.

These topics are chosen for their relevance to fundamental, highly significant questions about genome biology and evolution, and exclude topics that have specific biomedical aims likely to be of interest to another NIH institute (sequencing pathogens or disease vectors), or aims that fall outside of the NIH mission (e.g., plants or agricultural organisms). Priority will be given to questions that inform understanding of human health and disease.

We will encourage multi-species proposals for animal, fungal, and protist genomes, but not plant or prokaryotic genomes, which we believe are well-covered by other funders and also are less likely to inform a number of the rationales above. We will discourage individual species projects except in particular cases, such as those that provide the key remaining genomic information needed to address one of the questions above; or those proposing to sequence multiple individuals of the same species to understand genome evolution occurring at relatively short time intervals.

We will seek applications that have a rationale based on some aspect of the wider NIH mission to inform human health. However, justifications based solely on adding research value to experimental model organisms will not be accepted.

More generally, applications will be primarily justified based on one or more questions or hypotheses put forward by the applicant. The proposed program will not be primarily a resource-generating program (though data will be made available to the community).

Relationship with other activities

There is no overlap with any managed NHGRI programs, but investigator-initiated applications could partially overlap.

There is potential synergy with ENCODE or other functional genomics program.

Mechanism of Support

We will issue a PAR for investigator-initiated R01s; potential to re-release in 2017 based on response. A PAR does not require formal Council clearance, nor a set-aside of funds, though we have that option.

Funds anticipated

Although this concept does not propose to set aside funds, we estimate that the appropriate level of effort will require $2M per year for four years for up to 3 awards. All award decisions will be contingent on the quality of applications and availability of funds.