Future Opportunities for Genome Sequencing and Beyond

Concept Clearances for the NHGRI Genome Sequencing and Analysis Program Part 2

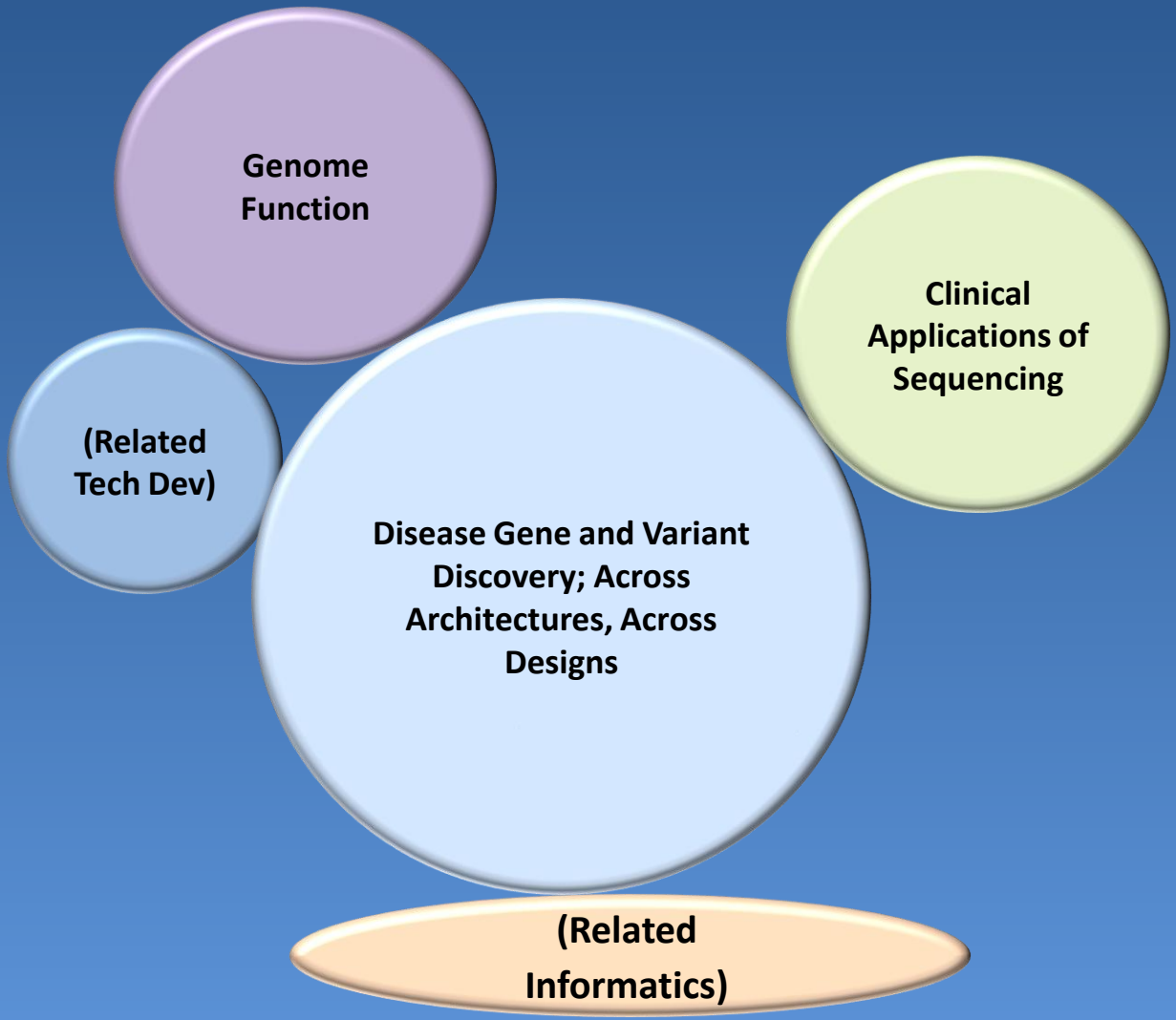
National Advisory Council on Human Genome Research

February 9, 2015

National Human Genome Research Institute

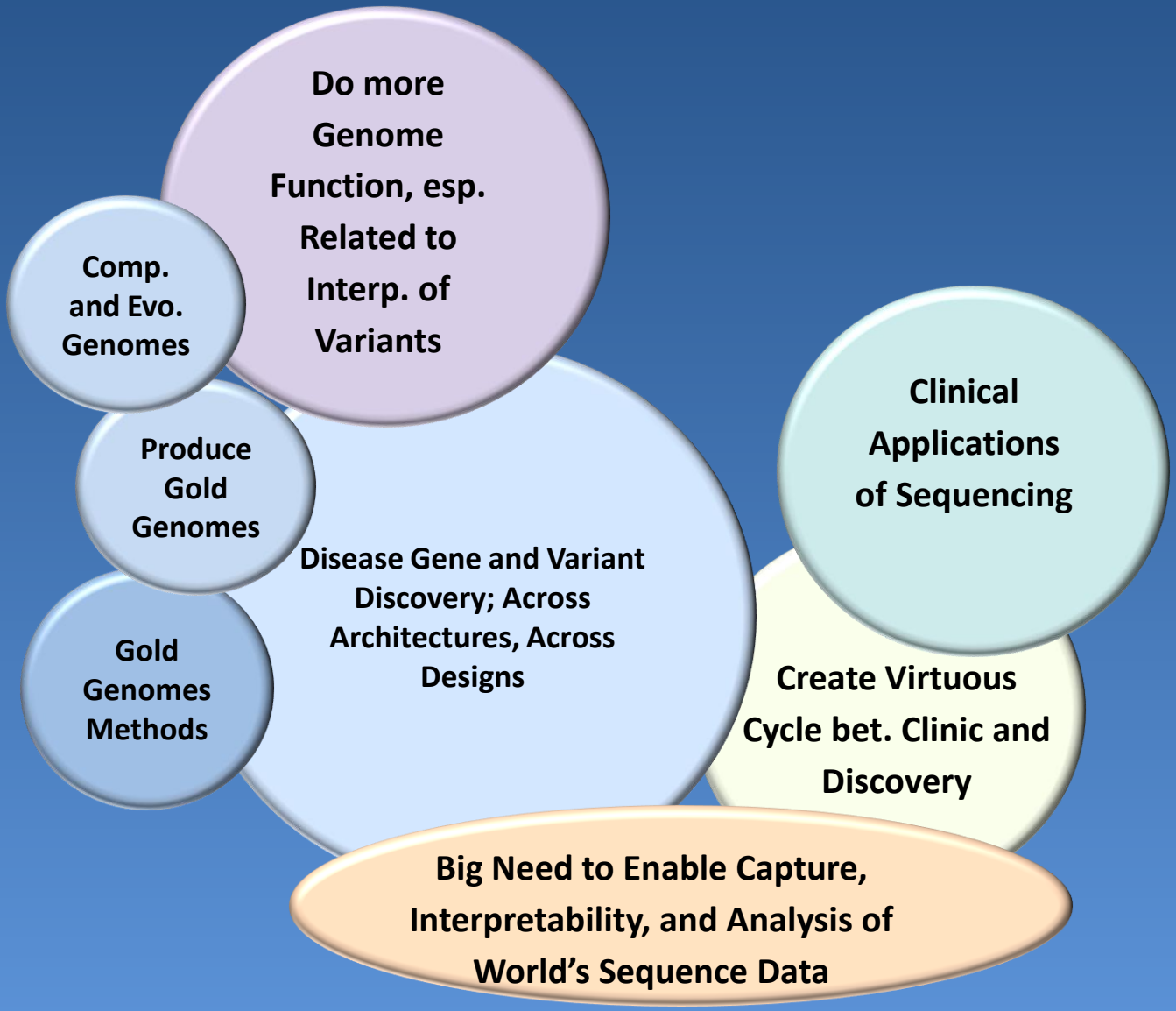
July 2014 Strategic Planning Workshop

Structure of Genomes	Biology of Genomes	Biology of Disease	Science of Medicine	Effectiveness of Healthcare



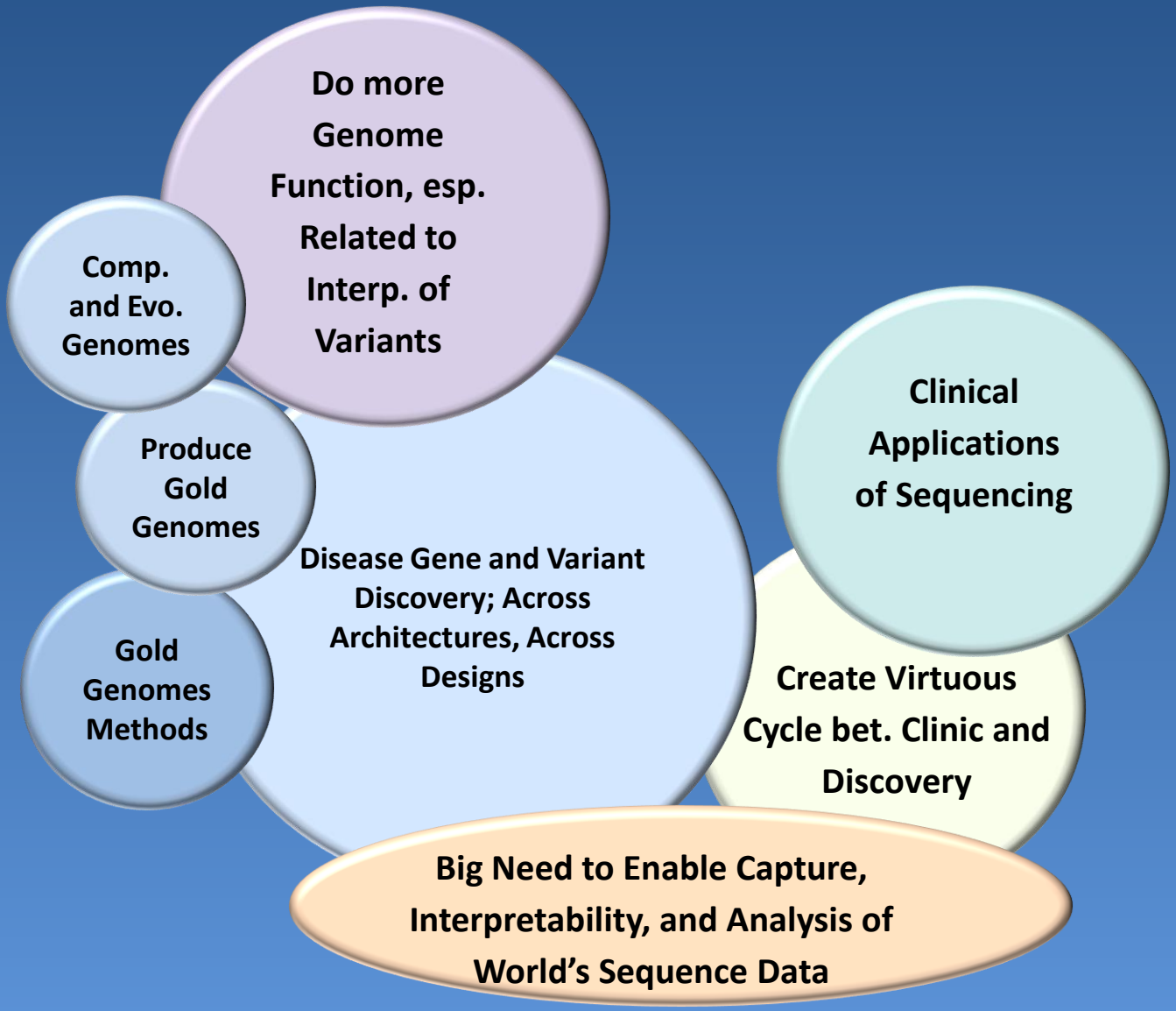
Workshop Wish-List

Structure of Genomes	Biology of Genomes	Biology of Disease	Science of Medicine	Effectiveness of Healthcare



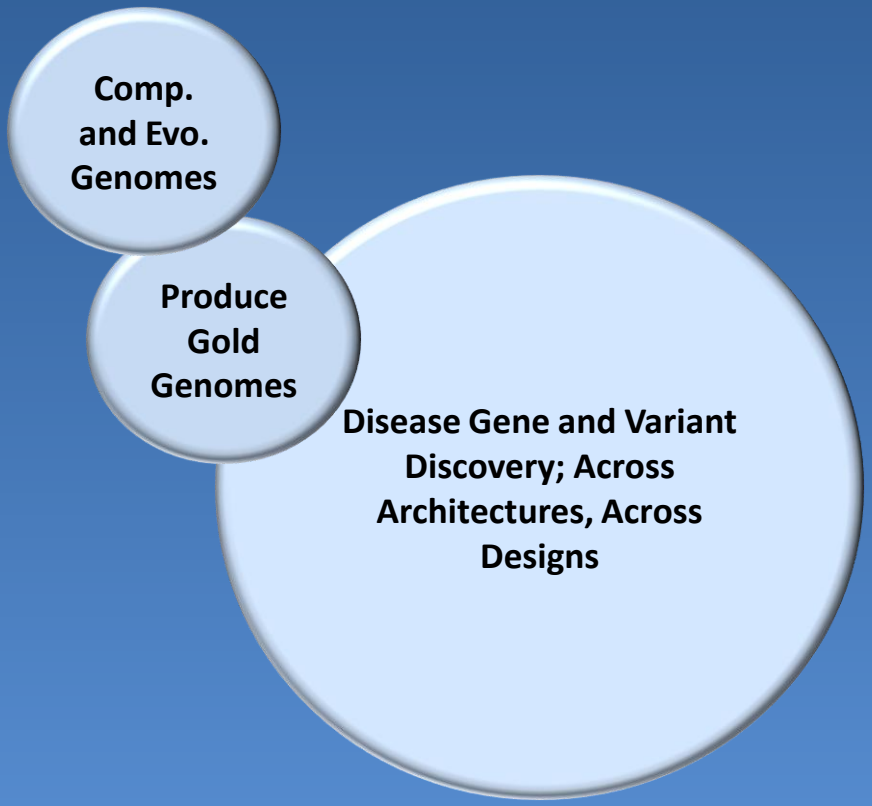
Concepts for Clearance

Structure of Genomes	Biology of Genomes	Biology of Disease	Science of Medicine	Effectiveness of Healthcare



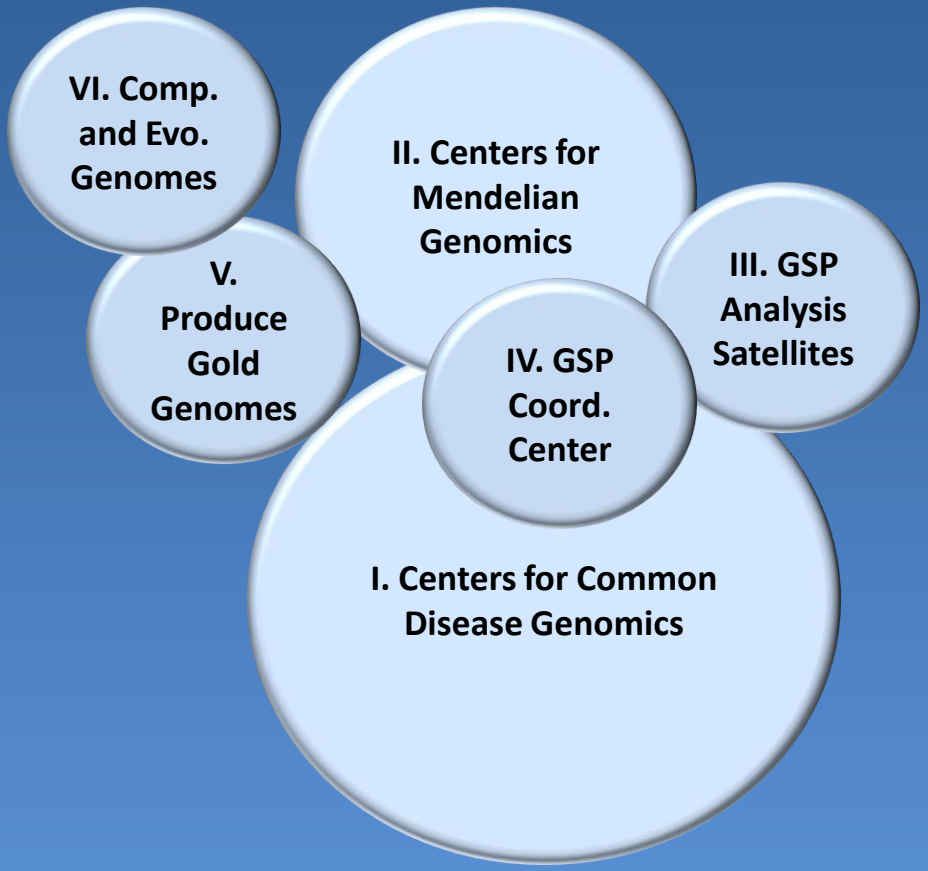
Concepts for Clearance

Structure of Genomes	Biology of Genomes	Biology of Disease	Science of Medicine	Effectiveness of Healthcare



Concepts for Clearance

Structure of Genomes	Biology of Genomes	Biology of Disease	Science of Medicine	Effectiveness of Healthcare



Multiple Concepts/RFA's

I. Centers for Common Disease Genomics (CCDG): RFA-HG-015-001

II. Centers for Mendelian Genomics (CMG): RFA-HG-015-002

*III. Genome Sequencing Program Analysis Satellites**

*IV. Genome Sequencing Program Coordinating Center**

V. "Gold Genome" Production

VI. Comparative and Evolutionary Genomics

The GSP: Analysis and Coordination

Considering both the CCDG and the CMG:

- ***A lot of data: Analysis Opportunities → Concept III***

- ***A complex structure: Coordination Needs → Concept IV***

III. GSP Analysis Satellites (GSPAS)

- Will propose and carry out novel, creative analyses of the data produced by the GSP, that will cut across individual projects, grants, and even programs (*not just GSP*)
- Will also help with cross-program analyses that we can define in advance or with the program
- Not routine data processing

III. GSP Analysis Satellites

- Improved or novel analyses for non-automated aspects of characterizing sequence variants in the data, after variant calling
- Particular interest in questions about association; analyses using existing functional data to help make associations and/or make functional inferences; means to improve study design to increase power; and other higher level analyses
- With overall GSP: when is a common disease study "comprehensive" or complete; characterization and specification of sample sets that could serve as common controls
- Will identify, and in collaboration with the GSP, carry out, other analyses that bridge across multiple grantees

III. GSP Analysis Satellites

Relationship to other HG activities:

- **NHGRI funds other sequence analysis activities both investigator-initiated and HG-initiated**
- **We will encourage proposals that will take best advantage of the GSP, and discourage those that would duplicate other ongoing or routine analysis efforts**
- **The GSPAS will be an integral part of the GSP**

III. GSP Analysis Satellites

Mechanism:

- Cooperative agreements to facilitate coordination with the GSP as a whole, while simultaneously encouraging creative proposals
- Investigators and key personnel from CMG or CCDG grants will not be eligible to be funded, in order to encourage dissemination
- We will write the FOA to increase the chances that each GSPAS will have a different area of focus

Funds:

- $3M per year for four years; 3-4 awards. Start as soon possible in time to other GSP elements

IV: GSP Coordinating Center (GSPCC)

Two Realities:

- **The current GSP is already a large, complex program; the new one will be more complex: More elements; more data; multiple large projects (and the collaborators that they bring)**

- **There are some cross-program "deliverables". We need scientific help spurring, leading, and tracking those**

IV: GSP Coordinating Center: Two Roles

1. **Leadership/coordination roles for cross-program objectives we can identify now:**

   - **Specification for common controls**
   - **When is a project "comprehensive"?**
   - **Joint allele frequency analysis for CMG**

   And others that may arise over time

Will bring leadership and a sense of mission to the items above; will need to work with other program components for bulk of analysis and to ensure consensus. Requires scientific expertise.

IV: GSP Coordinating Center

2. **Administrative/logistic/outreach:**

- Tracking progress (project status, costs, etc.)

- Logistical coordination within program and among collaborators (calls, meetings, hosting of documents, etc.)

- Policy coordination/dissemination (e.g., data access within the program)

- The program overall will need at least a minimum level of outreach: e.g.: what projects are underway, progress, collaborators, highlights, how the community can get the data

- Possible role in facilitating process for choosing new CCDG projects

In both leadership and administrative roles, the CC will work collaboratively with the rest of the program and with NHGRI. The CC will be co-equal with the other components.

IV. GSP Coordinating Center

Relationship to other HG activities:

- None

Mechanism:

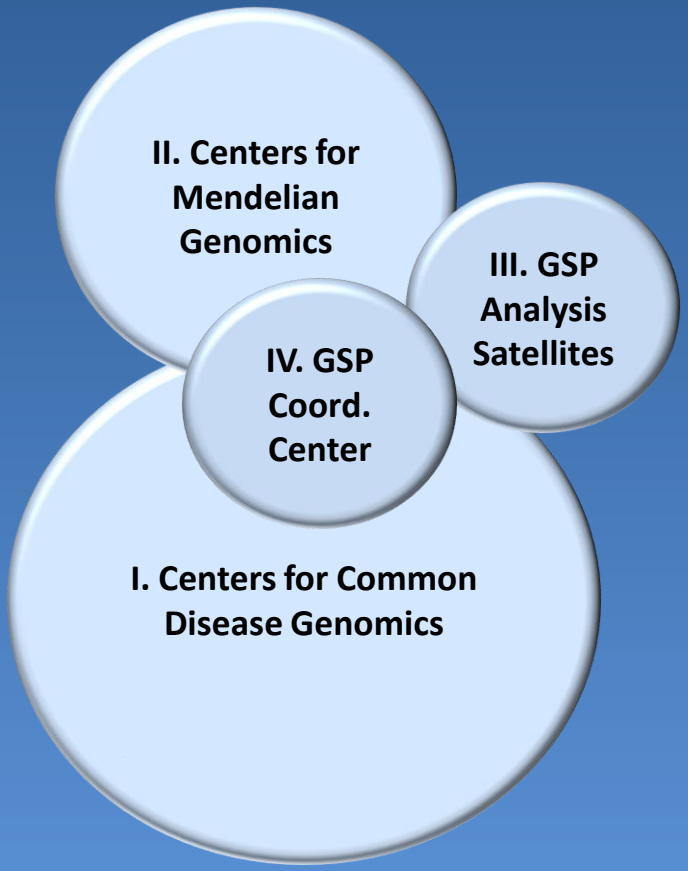
- Cooperative Agreement (e.g., U24) is required to ensure coordination with NHGRI staff and the GSP
- CCDG and CMG investigators will not be eligible for funding

Funds:

- $1M/year for four years; one award. FOA to be released as soon as possible to allow funding to begin at about the same time as the rest of the GSP

The Core GSP

Structure of Genomes	Biology of Genomes	Biology of Disease	Science of Medicine	Effectiveness of Healthcare



V: Gold Genomes

Purpose:

Produce high quality* finished genomes at “current costs” in the next three years

- **Human (25-50) and Select Non-Human Primate (10-12)**
- **Develop more specific definition of “high-quality finished”**
- **Use state-of-the-art methods to maximize quality/quantity**

V: Gold Genomes

Why?

Human:
The reference is good but can be made better:

- **gaps**
- **incomplete representation of SV**
- **move from a mosaic to a faithful haplotype representation**
- **can better represent world populations**

Non-human primates:
Primate genomes are key to understanding fundamental questions about human genome evolution, but existing assemblies mostly fall short of being reference quality

V: Gold Genomes

What is “high quality”?

- Well beyond what is achievable by current “short-read only” methods
- Can state in absolute terms, e.g., haplotype-resolved; contiguity, number of gaps, specific lists of known “difficult” regions, etc.
- Can state in terms of gain in biomedical utility
- Practical issues need to be considered (a moving state of the art; cost)
- FOA language will need to consider all, and also encourage creative proposals

V: Gold Genomes

Relationship to other activities:

- The NHGRI Genome Reference Consortium award already funds 5-7 very high quality human genomes; will need to coordinate effort with GRC (production AND integration into references)
- No overlap for primate genomes
- May need some community input re. selection of samples

Mechanism:

Cooperative Agreement for a resource

Funds:

$2M/year for three years; one or two awards

VI: Comparative and Evolutionary Genomics

Purpose:

Encourage *investigator-initiated* applications for comparative and evolutionary genomics projects, of wide scope and of high interest to the NIH research community, that address fundamental questions about genome biology, evolution, and function

VI: Comparative and Evolutionary Genomics

Some examples:

- Characterizing basic comparative features of genome structure leading to inferences about mechanisms of origin or function—e.g., whole genome duplication, 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 genomic innovation (e.g., origins of proteins, biochemical pathways, core metazoan developmental program, etc.)
- Defining the genomic basis of phenotypic innovations ranging from major taxonomic innovations – e.g., multicellularity or the adaptive immune system -- to those occurring on a shorter time scale, for example, differences between closely related species or within species

VI: Comparative and Evolutionary Genomics

- Animal, fungal, protist, but not others
- Multi-species or perhaps multiple individuals within a population
- De-prioritize those with a narrow focus more appropriate to another funder
- Prioritize those with some relevance to human biomedical science

VI: Comparative and Evolutionary Genomics

Relationship to other HG activities:

- No direct overlaps; NHGRI has no other announced mechanism to consider proposals for organismal genome sequencing
- Opportunity for synergy with Functional Genomics program

Mechanism:

- PAR for investigator-initiated R01 applications: this allows for a written FOA, a single review date; applications reviewed together. Second release in a year depending on response.

Funds:

- We estimate the activity will require ~$2M/year but will not set aside funds in advance; one or two awards; three years

NHGRI GSP

I. Centers for Common Disease Genomics:	$60M/year; 4 years
II. Centers for Mendelian Genomics:	$10 M/year; 4 years
	(+$2M/year from NHLBI)
I. GSP Analysis Satellites	*$3M/year; 4 years*
II. GSP Coordinating Center	*$1M/year; 4 years*
I. Gold Genome Production:	*$2M/year; 3 years*
II. Comparative and Evolutionary Genomics:	*$2M/year; 3 years*
Year 1 Total:	$78M (+$2M)

Thanks To Many

Shannon Biello
Larry Brody
Valentina di Francesco
Bettie Graham
Eric Green
Lucia Hindorff
Carolyn Hutter
Jon LoTempio
Teri Manolio
Rudy Pozzatti
Jeff Schloss
Mike Smith
Heidi Sofia
Lu Wang
Chris Wellington
Kris Wetterstrand