Concept Clearances for the NHGRI Genome Sequencing and Analysis Program

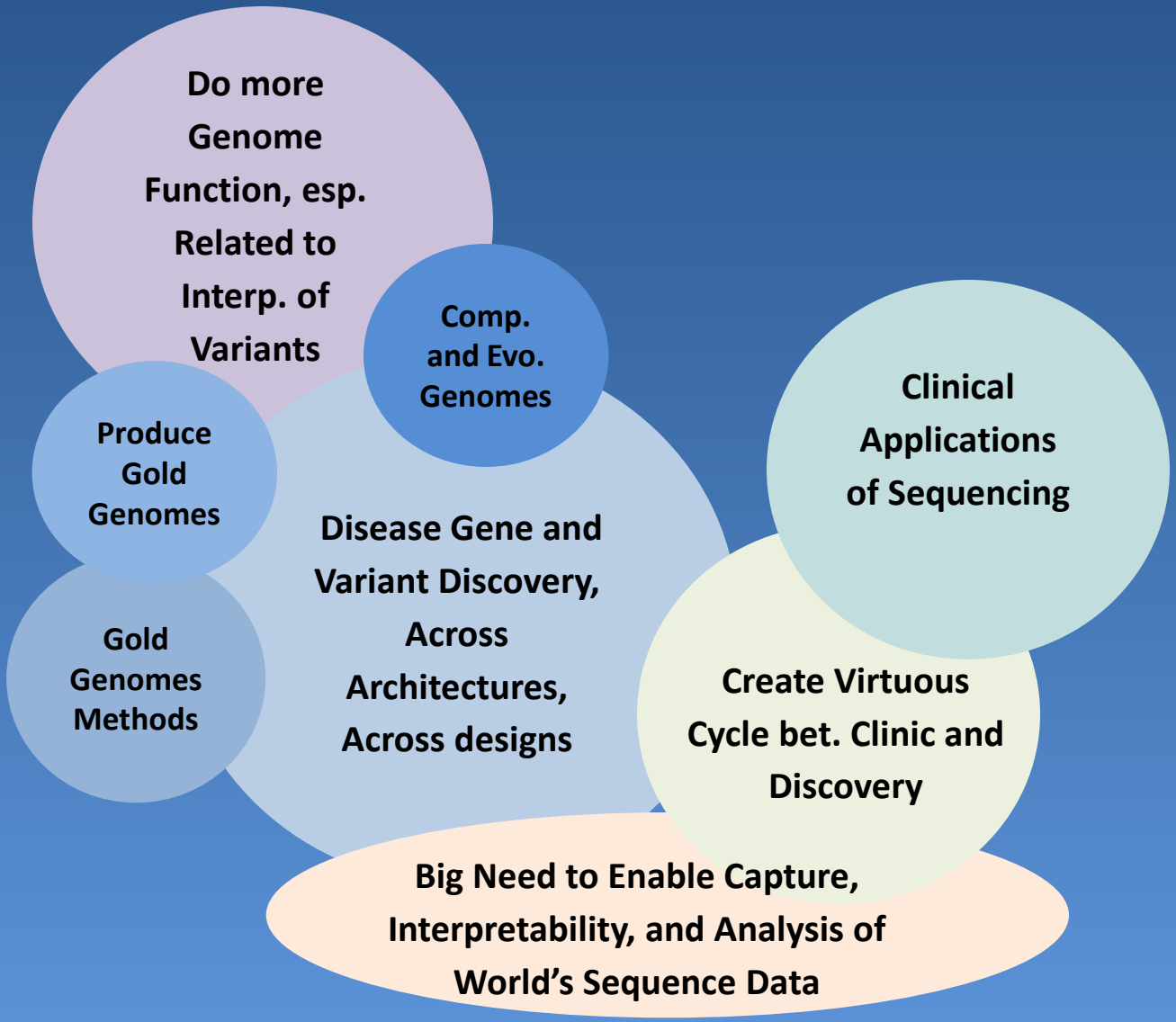
National Advisory Council for Human Genome Research

September 8, 2014

National Human Genome Research Institute

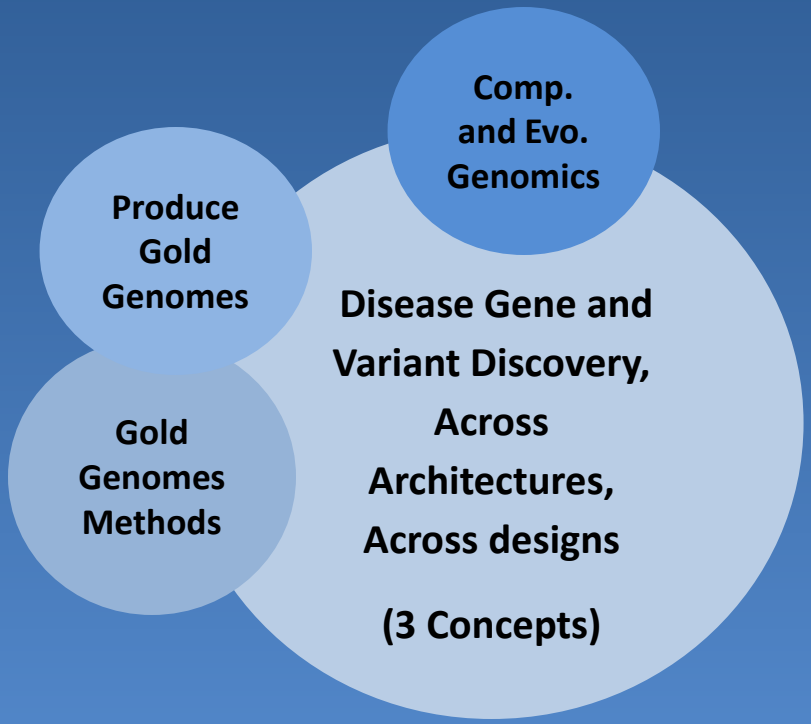
Workshop "Wish List"

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

I. Common Disease Variant Discovery (CDVD)

II. Centers for Mendelian Genomics (CMG)

III. Genome Sequencing Program Coordinating Center (GSPCC)

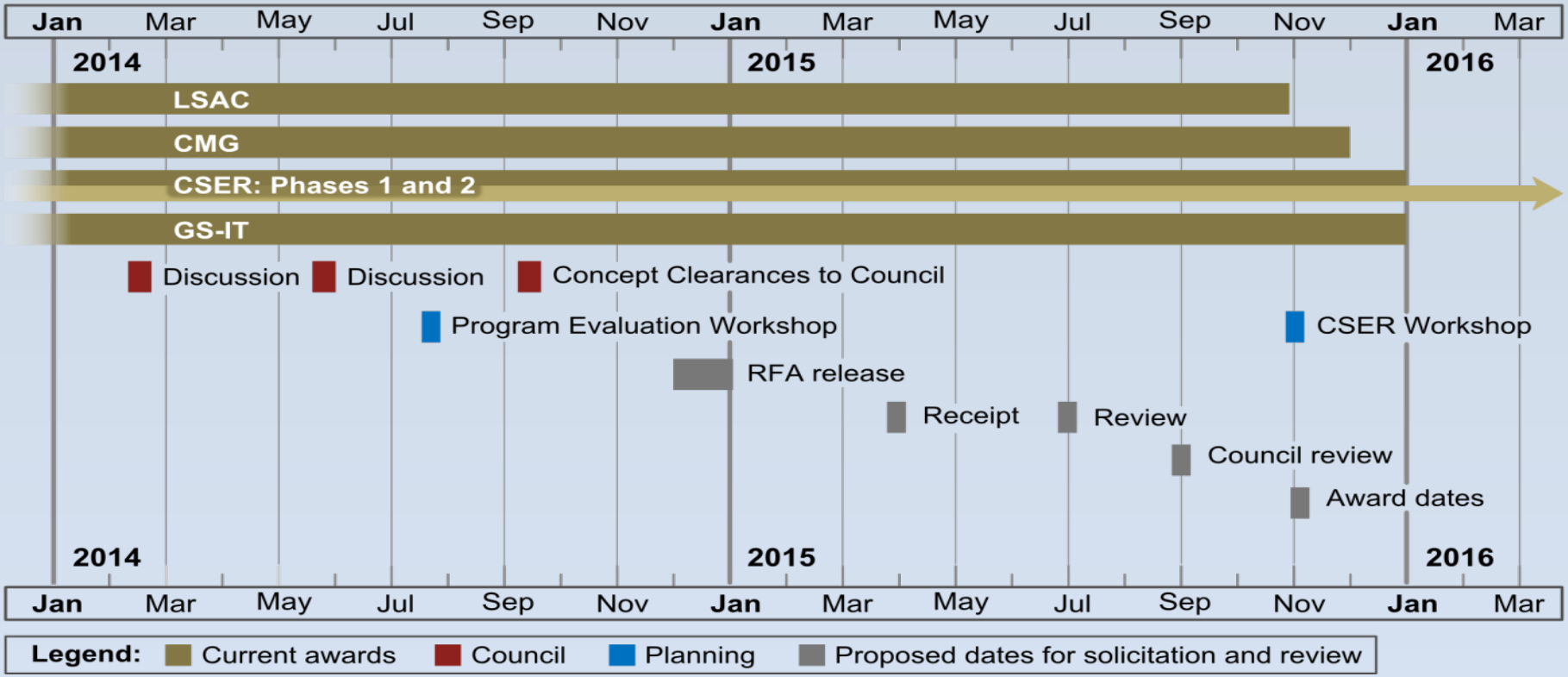
IV. [a. “Gold”-Quality Genome Production; b. “Gold”-Quality Genome Methods]

V. [Comparative and Evolutionary Genomics]

Total proposed here for I-III is $71.5M (FY 2016). *This is the amount needed to do all of these at least at credible level of implementation. It does not include potential co-funding.*

2015 estimated cost of *current* LSAC, CMG = $85M (includes $2M co-funding)

Timeline


Jan Mar May Jul Sep Nov Jan Mar May Jul Sep Nov Jan Mar
2014
2015
2016
LSAC
CMG
CSER: Phases 1 and 2
GS-IT
Discussion
Discussion
Concept Clearances to Council
Program Evaluation Workshop
CSER Workshop
RFA release
Receipt
Review
Council review
Award dates
2014
2015
2016
Jan Mar May Jul Sep Nov Jan Mar May Jul Sep Nov Jan Mar
Legend: Current awards Council Planning Proposed dates for solicitation and review

I. Common Disease Variant Discovery

What?

- **Establish a collaborative large-scale genome sequencing effort to identify genomic variants contributing to multiple common, complex disease phenotypes**
- **Explore *comprehensively* a range of disease types, architectures to learn general principles about how to approach these studies**
- **Undertake and compare a range of designs (e.g., yield and efficiency)**
- **At least one comprehensive whole genome sequencing study (cost vs number); plus WES**
- **Develop "foundational" deliverables, e.g. data resources for disease research communities, know-how for similar studies, technology innovation, data handling, standards, policies, common controls**

Common Disease Variant Discovery

Why?

- **Common diseases affect hundreds of millions of people**
- **Understanding the genomic variants influencing risk (or protection from) these will provide insight into the basis for important individual diseases, which can lead to better diagnosis, prognosis, treatment, and even identify targets for therapy**
- **Also provides general insight into the biology of disease and the relationship between genotype and phenotype.**
- **Requires genome sequencing esp. to discover rare variants systematically.**
- **Studies need to be *comprehensive:* well-powered; large sample numbers (up to 50K or more) are indicated by data so far. Workshop: "*Better a few comprehensive than many partial*"**

More….

Common Disease Variant Discovery

Why?

- It needs to be done at scale, esp. to compare across multiple diseases (large studies; the need to derive general lessons e.g. about disease architecture).
- A well-chosen set of comprehensive studies and datasets have high potential to provide a resource that will be catalytic for many studies by many investigators– specific disease communities as well as the general genomics community, including lists of disease variants, large WGS datasets for developing tools for interpreting noncoding function, etc.

Common Disease Variant Discovery: Scientific Considerations 1

How many studies will be enough to explore a range of disease types, architectures, and allow examining a range of designs?

- **We propose six to ten over four years as a minimum**
- **At least one will be WGS (vs WES)**

Common Disease Variant Discovery: Scientific Considerations 2

What is "comprehensive"?

- **Can define only partly in advance, e.g.,**
  - **-with reference to power (freq, effect size)**
  - **-keep going until discovery curve falls off**
  - **-with qualifiers for populations studied**
  - **-and practical limits**

Will need a starting point, then iterate and refine as a program goal.

Common Disease Variant Discovery: Some Desirable Features

1. Allow some production of non-genome seq data e.g., epi or transcriptomes (*BUT coordinate with Function program*)

2. Flexibility to do projects not directly related to a specific disease

3. Opportunities for outreach/liaison with other investigators/programs: "spokes" for collaboration with outside investigators on *pilot-level* efforts that link sequencing to function, or new analyses, or tech dev, etc.

Common Disease Variant Discovery: Resource/Budget Considerations

- Total for seven projects (incl. one WGS)= $292M (see *Concept* cost assumptions), or $73M/year; four years

- BUT we propose 80% of that, or $60M/year

- Several factors will reduce cost over time

  -Technology, also data storage (2X in two years?)

  -Study design (not all studies will require 50K samples; common controls?)

  -Co-funding and other funding collaborations

High confidence for seven studies as described. More possible.

Common Disease Variant Discovery: Organizational Considerations

1. Cooperative Agreements/Research Network. Open competition.
2. 2-4 awards– very large projects will need to be split up among grantees

Projects

1. Better peer review of individual projects—at least the initial year of work to be fully proposed in the application

2. Multiple methods to recruit new projects over time:
    - X01's
    - Community Workshops
    - IC-initiated projects (opportunity for co-funding)
    - HG-priority projects

Common Disease Variant Discovery: Leveraging Resources

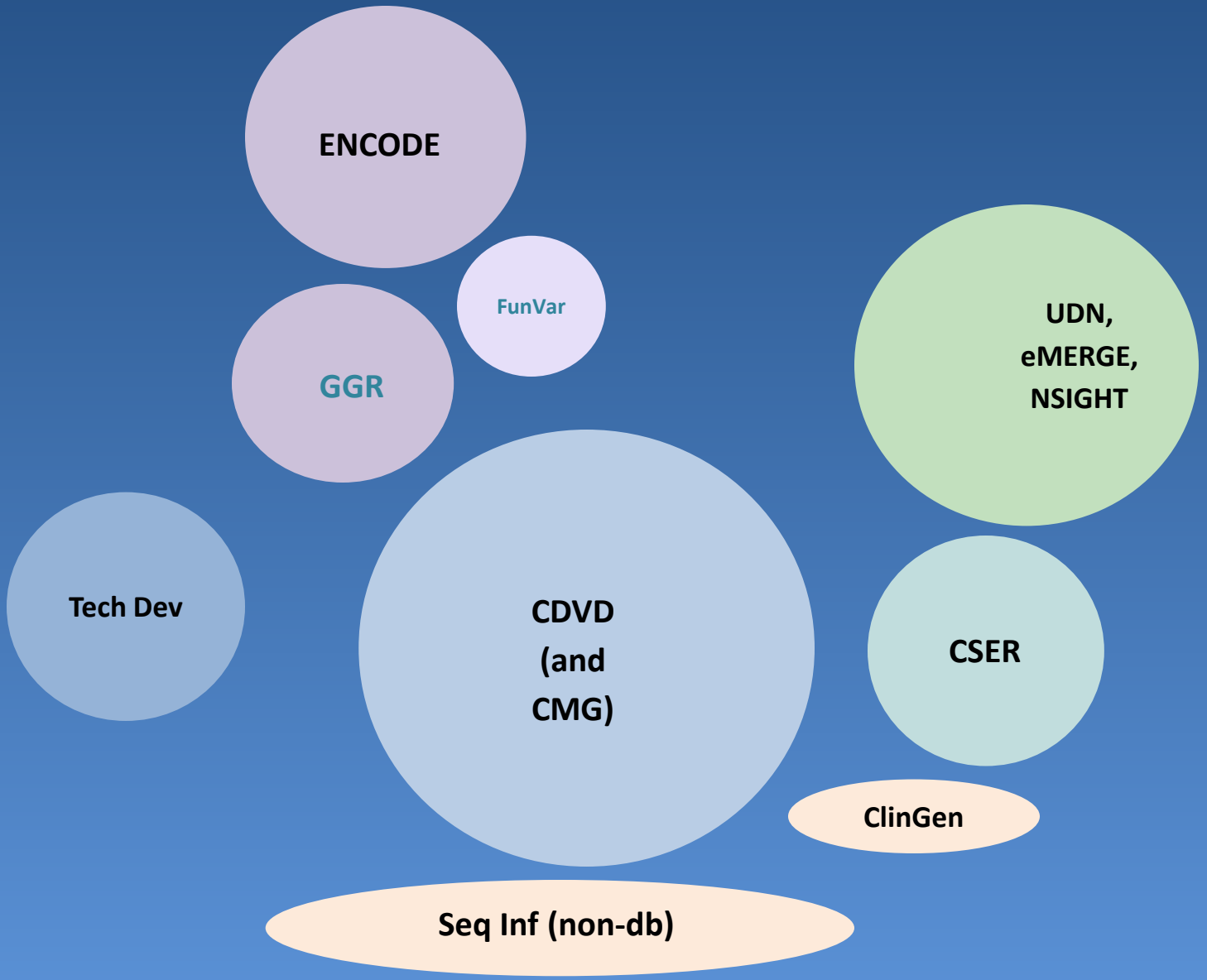
1. Mechanism must allow/incent co-funding; e.g. with X01's to allow route for participation of other ICs. NHGRI must reach out to other institutes to collaborate on these projects

2. Mechanism should incent applicants to seek other outside funding to add to the number of example diseases that can be explored. For example, NHGRI could make partial awards and then provide additional funds to grantees that are successful in identifying resources for more, or more comprehensive, projects.

3. Mechanism must accommodate potentially significant changes in capacity (for example, large increases due to identification of new opportunities, as well as changes due to the completion of projects).

Relationships to Other Programs

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



Relationships to Worldwide Efforts

- Multiple individual complex disease efforts
- Large cohort studies
- Will need to look for opportunities to collaborate/synergize

Common Disease Variant Discovery: Summary

- We have the technical capability to find comprehensive sets of genomic variants that contribute to common disease.
- The proposed program will attempt this for a representative set of common diseases in a four years.
- In so doing it will explore a range of genomic architectures and project designs, develop improved methods for efficiently obtaining and analyzing the data, and develop resources for communities of disease investigators and genomics in general.
- We believe that the recommended funding will be sufficient to have high confidence to achieve the minimum goals given reasonable projections. The program will work to exceed the minimum.

Common Disease Variant Discovery: Summary

- To increase the number of studies that we can do, and the likelihood of program success, we will seek co-funding.
- We will structure funding to incent institutional and other contributions. We will need to try several approaches before we know what works.
- The scientific goals and requirements of the program mean that it will need as much flexibility as possible built in to the solicitation, funding, and NHGRI management of the program.

Common Disease Variant Discovery: Cost Assumptions

- Next year: $2000 (data) plus $600 (automated analysis and storage) per WGS, *fully loaded*; $450 per WES.
- Assume as high as 25K case/25K control samples per disease will be required for power (could be less, could be more)
- WGS study =$130M; WES study =$23M; seven studies = ~$260M
- This does not include project management, sustained collaborations, bringing in samples, high-level analysis, many "foundational" elements. Historically, 20% (mostly WES) add ~$32M (or, ~$2-$4M/year/center)
- Total= $292M; or $73M/year

Huge sample collections available for RVAS

GWAS samples in 18 diseases: 400,000 cases

		Cases GWAS			Cases GWAS
Cardiovascular/Metabolic	Early Myocardial Infarction*	20,000	Psychiatric/Neurologic	Schizophrenia*	30,000
	Coronary Artery Disease*	64,000		Bipolar	10,000
	Type 2 Diabetes*	60,000		Autism	20,000
	Atrial Fib/Stroke	10,000		Alzheimer	10,000
Germline Cancer Risk	Breast Cancer	25,000	Autoimmune	Type 1 Diabetes	30,000
	Prostate Cancer	10,000		IBD/Crohn's	30,000
	Colon Cancer	13,000		Multiple Sclerosis	20,000
	Lung Cancer	20,000		Rheumatoid Arthritis	30,000
	Melanoma	13,000		Lupus	15,000