Concept Clearance: Understanding the Impact of Genomic Variation on Genome Function

Mike Pazin, Stephanie Morris, and Dan Gilchrist

Division of Genome Sciences, NHGRI, NIH 10-11 February 2020 Council





Thank you

NHGRI Leadership

NHGRI Function Team

Division of Genome Sciences

Genome to Phenotype Workshop participants



Overview

Background

Proposed Scope And Objectives

Relationship To Ongoing Activities

Discussion And Vote



Structure For Proposed Program

1) Functional Characterization Centers

Test Variant Impact 4) Predictive Modeling Projects

Predict Variant Impact

5) Data Coordination Center

2) Mapping Centers

Functional Activity Maps

3) Regulatory Network Projects

Network Maps





Background











Genome To Phenotype Workshop





Related Ideas From The Genome To Phenotype Workshop

Functional genomics at scale

Deep perturbation profiling, proximal molecular readout of variants

Single cell multi-omic technology

Large-scale perturbations in pools of cells followed by high-content, high-resolution data collection

Transferable gene regulatory or multi-scale models from multi-layered data across biological systems



Scientific Challenges To Address

- Understanding how genomic variation affects genome function
- Moving from association to a causal understanding of how variation impacts human health and disease
- Improving variant interpretation
- Predicting the function of untested variants



Understanding Genome Function Is At The Forefront of Genomics

- NHGRI is positioned to make a major impact in understanding genome function
- Value in continuing, and even potentially increasing, our commitment in this area
- Recognize this proposed concept addresses only a portion of the larger field of genome function
- Expect understanding function to be foundational in 2020 strategic plan



Proposed Scope And Objectives



0

Significance

Challenge/ Opportunity

Interpreting variants and candidate genotype/phenotype associations

Objective

Understanding how genetic variation impacts genome function, phenotypes (including disease)

Implementation

Data collection, community data resource, integrative data analysis, and predictive modeling



Weaving Threads from Other Projects

Systematic Genome Perturbation and Characterization

- ENCODE
- LINCS

Systematic Identification of Functional Elements

- ENCODE
- Roadmap Epigenomics Mapping Centers (REMC)
- GTEx

Predictive Gene Regulatory Networks

Genomics of Gene Regulation (GGR)

Predicting Causal Variants

Non-coding Variants Program (NoVa)



Program Objectives

Transform our understanding of how variation impacts function and leads to phenotypes in health and disease

- 1) Test impact of genomic variation on function
- 2) Determine where and when regulatory elements and genes are active
- 3) Explore role of sequences in regulating gene networks and establishing phenotypes
- 4) Model and predict relationships among genomic variation, function, and phenotype
- 5) Establish a resource to enable future studies



Structure For Proposed Program

1) Functional Characterization Centers

Test Variant Impact 4) Predictive Modeling Projects

Predict Variant Impact

5) Data Coordination Center

2) Mapping Centers

Functional Activity Maps

3) Regulatory Network Projects

Network Maps





Resource and Research

c g T

Resource

Research

Mapping Centers

Data Coord. Center

Functional Char. Centers

Regulatory Network Projects

Predictive Modeling Projects



Goal: Improve Understanding of the Relationships Between Variation and Genome Function

1) Functional Characterization Centers

Test Variant Impact

- Systematically apply high-throughput genomic perturbation methods
- Test non-coding and/or protein-coding functional elements
- Assay and catalog impact of perturbations
- Refine and develop novel generalizable approaches



Goal: Improve Ability to Determine Where and When Regulatory Elements and Genes Are Active

2) Mapping Centers

Functional Activity
Maps

- Obtain high-value biological samples
- Establish multi-omics pipelines for high-throughput mapping at single-cell resolution to:
 - Map functional regions while preserving information about biological and/or spatial context
 - Identify cell type-specificity and gene-specificity of regulatory elements



Goal: Explore the Roles of Sequence in Regulating Gene Networks and Establishing Phenotypes

3) Regulatory Network Projects

Network Maps

- Collect data with spatial/temporal resolution
- Measure differences in gene and regulatory element activity during biological changes
- Identify gene regulatory networks
- Place variants, elements, and phenotypes in networks



Goal: Develop Computational Approaches To Model Relationships Among Variation, Function, and Phenotype

4) Predictive Modeling Projects

Predict Variant Impact

- Develop, apply, and experimentally test predictive models
- Create tools to enable inferences about genome function
- Prioritize samples, approaches, and experiments for the consortium to conduct the most informative experiments
- Provide analytical expertise for consortium-wide efforts



Goal: Establish a Resource to Enable Future Studies of Variation, Function, and Phenotype

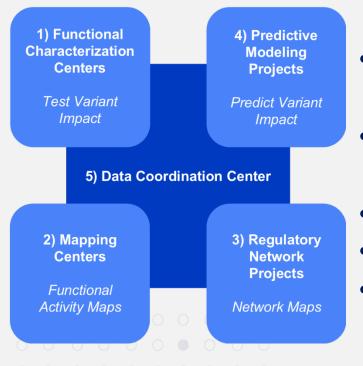
5) Data Coordination Center

Build Community
Resource

- Provide community access to data, software, and resources
- Make data available in a form ready for advanced machine-learning approaches
- Work with other groups to facilitate data integration, joint analysis, and sharing of standards and best practices
- Organize consortium activities



Shared Activities

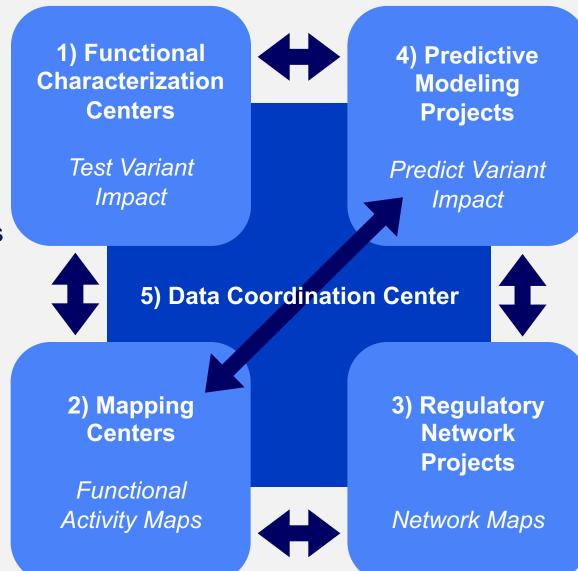


- Contribute data, metadata, analyses, and software to the Data Coordination Center and appropriate repositories
- Planning, implementation, and analysis of consortium-wide or component-wide projects
- Develop standards and establish data quality metrics
- Share best practices and lessons learned
- Contribute to outreach efforts



Synergies Within The Program

Characterization centers help prioritize samples to be mapped; element maps inform testing by systematic perturbation



Characterization and mapping data inform predictive models; models help prioritize variants to test, samples to map



Expected Outcomes

Key Outcome: Understanding how genomic variation impacts genome function

- Identify and characterize functional elements
- Model variants and elements in networks and pathways
- Improve variant interpretation

Key Outcome: Catalogs of variant impact

- Data resource
- Tools, models, methods, standards, technologies



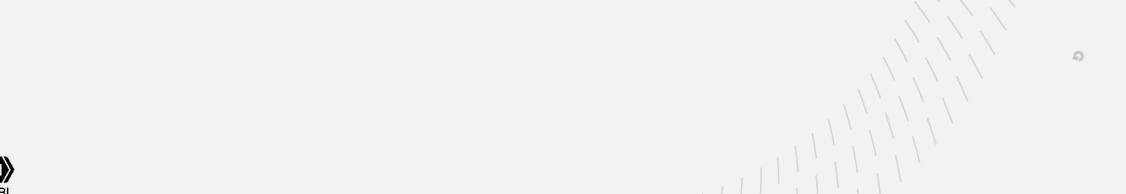
Overview of Program Components

| Component | Objective | # Awards | TC/year |
|----------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|----------|---------|
| 1) Functional Characterization Centers | To characterize the effects of genomic variation on genome function | 7-10 | \$15M |
| 2) Mapping Centers | To identify where and when regulatory elements and genes are active | 3-5 | \$8M |
| 3) Regulatory Network Projects | To explore the role of functional genomic elements in regulating networks and influencing phenotypes | 5-7 | \$7M |
| 4) Predictive Modeling Projects | To develop computational approaches to model and predict relationships among genomic variants, genome function, and phenotype | 6-8 | \$5M |
| 5) Data Coordinating Center | To enable future studies by organizing data, methods, and software and distributing them to the community | 1 | \$5M |
| Total | | | \$40M |





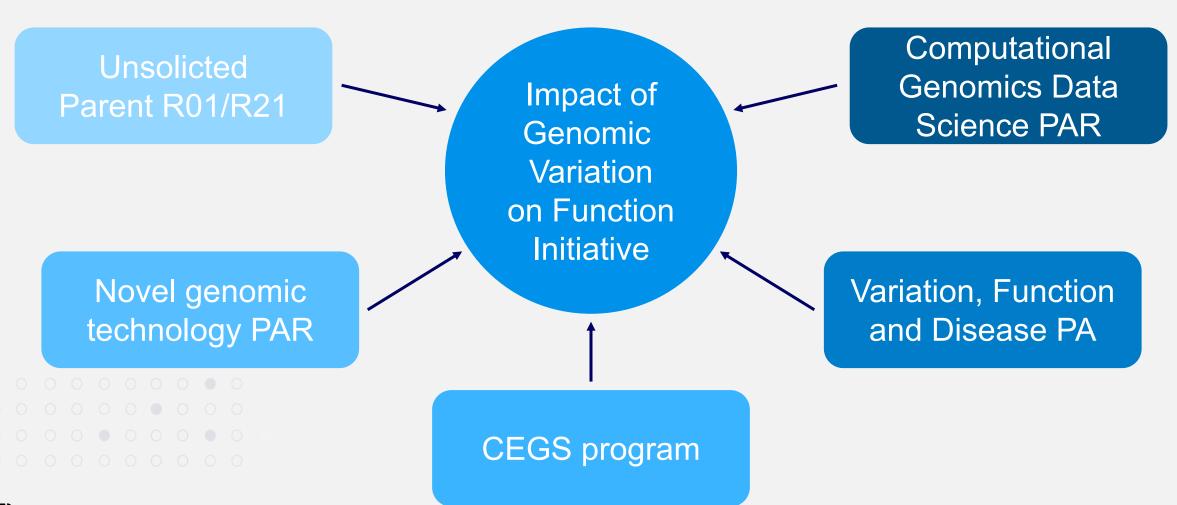
Relationship To Ongoing Activities





0

Other NHGRI Activities That Augment This Program





Program Interactions 4) Predictive 1) Functional Modeling Characterization Variant **Projects** Centers ClinGen Discovery **Projects Predict Variant** Test Variant **Impact Impact** 5) Data Coordination Center 3) Regulatory 2) Mapping **Network Centers Projects IHEC Functional** Activity Maps Network Maps **HUBMAP** and HCA

NHGRI

Understanding How Genomic Variation Affects Genome Function

- One of the central challenges in biology
- We think now is the time to bring new approaches to this problem



Discussion And Vote







Significance

Challenge/ Opportunity

Interpreting variants and candidate genotype/phenotype associations

Objective

Understanding how genetic variation impacts genome function, phenotypes (including disease)

Implementation

Data collection, community data resource, integrative data analysis, and predictive modeling

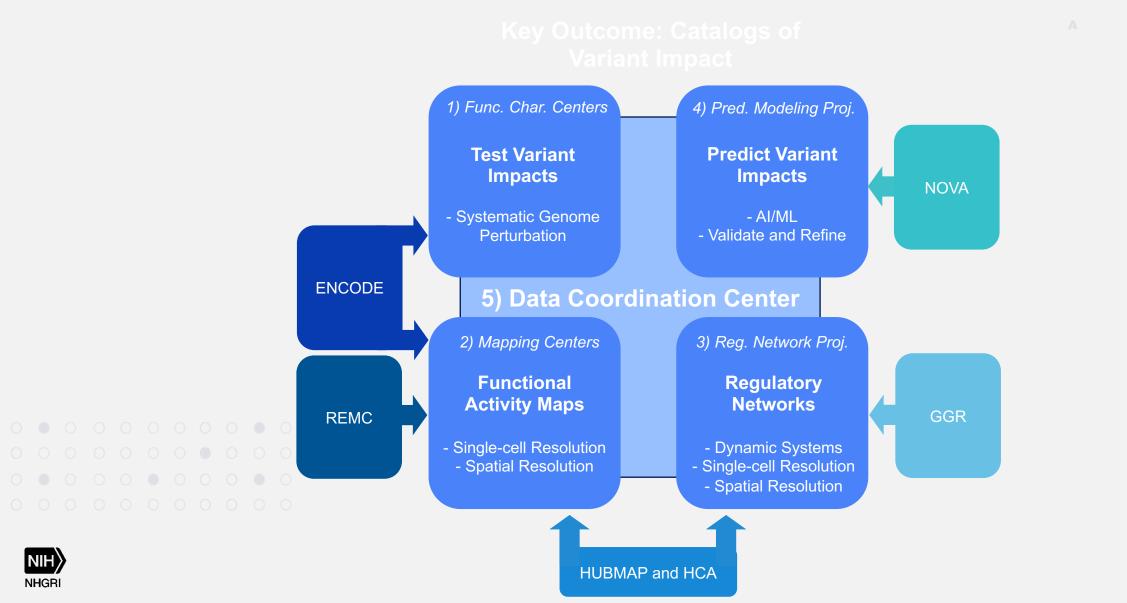


New Ideas In Proposed Initiative

- Single-cell resolution
- Spatial and/or temporal resolution
- Systematic genome-scale functional characterization of both protein-coding and non-coding regions
- Network-level and enhanced causality inferences
- Predictive models for untested variants



Foundational Projects



Recommendations from the Genome to Phenotype Workshop

- Discovery and Interpretation of Variation Associated with Human Health and Disease
 - Functional genomics at scale
- Address Basic Research Questions in Genome Variation...
 - •
- Predict and Characterize Functional Consequences of Genome Variation...
 - •
- Generate Data Resources, Methods, Technologies and Computational Capabilities
 - Large-scale perturbations in pools of cells followed by high-content, high-resolution data collection



Structure For Proposed Program

Production and Resource \$30 M

Research and Resource \$10 M

5) Data Resource \$5M 2) Mapping Centers

Identify where/when regulatory elements are active \$10M

1) Func. Char. Centers

Identify element functions and impacts of genetic variation

\$15M

3) Reg. Network Proj.

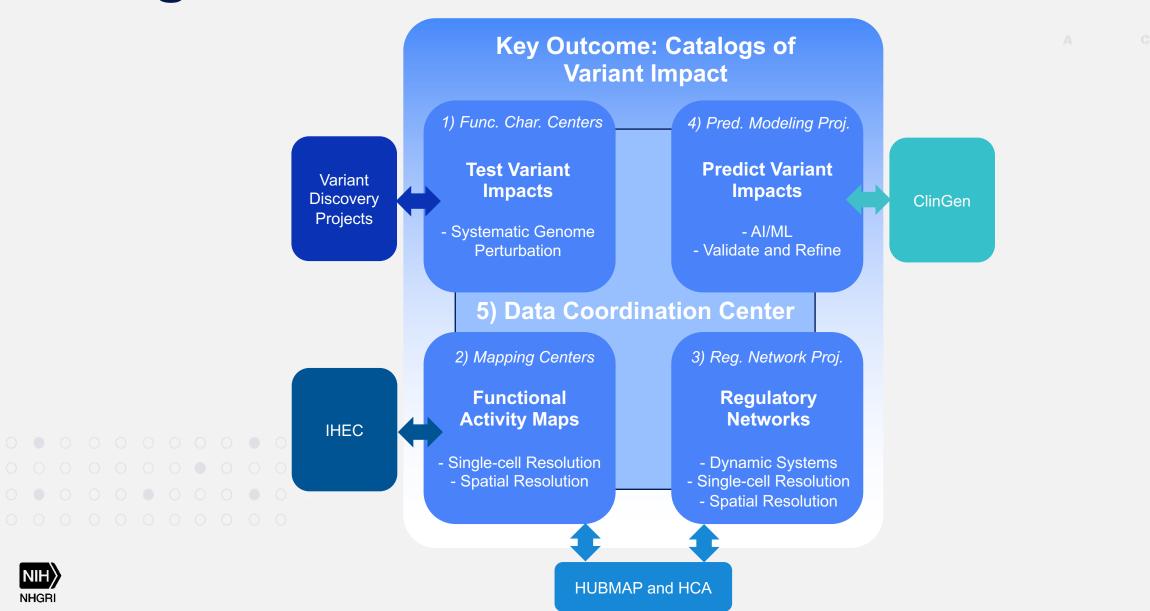
Networks and causality via spatial/temporal resolution \$5M

4) Pred. Modeling Proj.

Integrative analysis and predictive modeling \$5M



Program Interactions



Structure For Proposed Program

Production Research Resource 5) Data 1) Functional 3) Regulatory 4) Predictive 2) Mapping Centers Characterization Coord. Modeling Projects **Network Projects** Center Centers



Overview of Program Components

| Component | Objective | # Awards | TC/year |
|--------------------------|--------------------------------------------------|----------|---------|
| 1) Functional | To characterize the effects of genomic variation | 7-10 | \$15M |
| Characterization | on genome function | | |
| Centers | | | |
| 2) Mapping | To identify where and when regulatory elements | 3-5 | \$8M |
| Centers | and genes are active | | |
| 3) Regulatory | To explore the role of functional genomic | 5-7 | \$7M |
| Network Projects | elements in regulating networks and influencing | | |
| | phenotypes | | |
| 4) Predictive | To develop computational approaches to model | 6-8 | \$5M |
| Modeling Projects | and predict relationships among genomic | | |
| | variants, genome function, and phenotype | | |
| 5) Data | To enable future studies by organizing data, | 1 | \$5M |
| Coordinating | methods, and software and distributing them to | | |
| Center | the community | | |
| Total | | | \$40M |
| | | | |
| | | | |



Understanding How Genomic Variation Affects Genome Function

One of the central challenges in biology

We don't do this as well as we'd like



Variation And Phenotype

- Many genomic variants have been identified
 - Which affect phenotype?
 - What phenotypes do they affect?



Variation and GWAS

- Many genes and variants are associated with a disease
 - Which variants are causal?
 - Which combinations are causal?
 - What are the affected genes and cell types?



Variant Interpretation Has Not Kept Pace With Discovery Or Association

- Until recently relatively low throughput or low resolution approaches
- Current algorithms to predict the function of untested variants are not as accurate as we'd like

