Multi-omics for Health and Disease Pre-Application Webinar

RFA-HG-22-008 – Disease Study Sites (DSSs)
RFA-HG-22-009 – 'Omics Production Center(s) (OPCs)
RFA-HG-22-010 – Data Analysis and Coordination Center (DACC)

NIH Multi-Omics Team

NHGRI, NCI, NIEHS
September 26, 2022

National Human Genome Research Institute

The Forefront of Genomics®

Agenda

- Welcome and Participation Guidelines - Erin Ramos
- Program Overview – Joannella Morales
  - Background and Rationale
  - Consortium Components
  - Focus on Diversity
  - The AnVIL Eco-system
  - Data Sharing Expectations
  - IC-specific Interests – Kim McAllister, Leah Mechanic
- Q&A - Team

Zoom Best Practices & Participation Guidelines

- Your video and audio will be turned **OFF by default** to reduce noise and for the purpose of recording.
- You may **post questions** using the **Q&A** function on your screen.
  - The moderator will collect the questions and read them during the Q&A session.
- This webinar **will be recorded and posted** on the NHGRI website
- Contact us: **MultiOmicsProgram@mail.nih.gov**

NIH NHGRI

NIH Multi-Omics Team

- **NHGRI**
  - Jyoti Dayal
  - Iman Martin
  - Joannella Morales
  - Erin Ramos
  - Riley Wilson
- **NCI**
  - Leah Mechanic
  - Melissa Rotunno
- **NIEHS**
  - Kim McAllister








NIH NHGRI

Background

Advances in **high-throughput technologies**

⬇

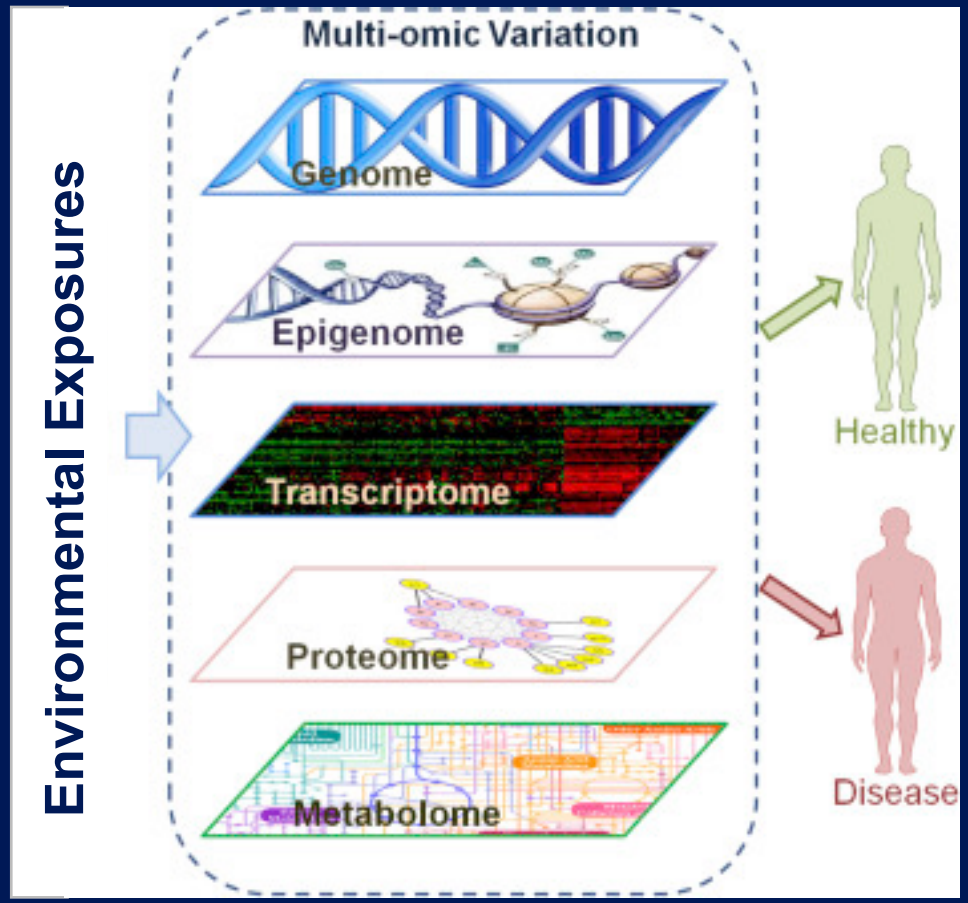
Increased **access** to distinct molecular data types **(‘omics data)**

Genomics	Epigenomics	Transcriptomics	Proteomics	Metabolomics
				
DNA (e.g. SNP or WGS)	Chromatin accessibility Chromatin structure DMA methylation	mRNA Non-coding RNA (e.g. miRNA, piRNA, lncRNA)	Secreted and intracellualr proteins	Endogenous circulating metabolites Xenobiotics

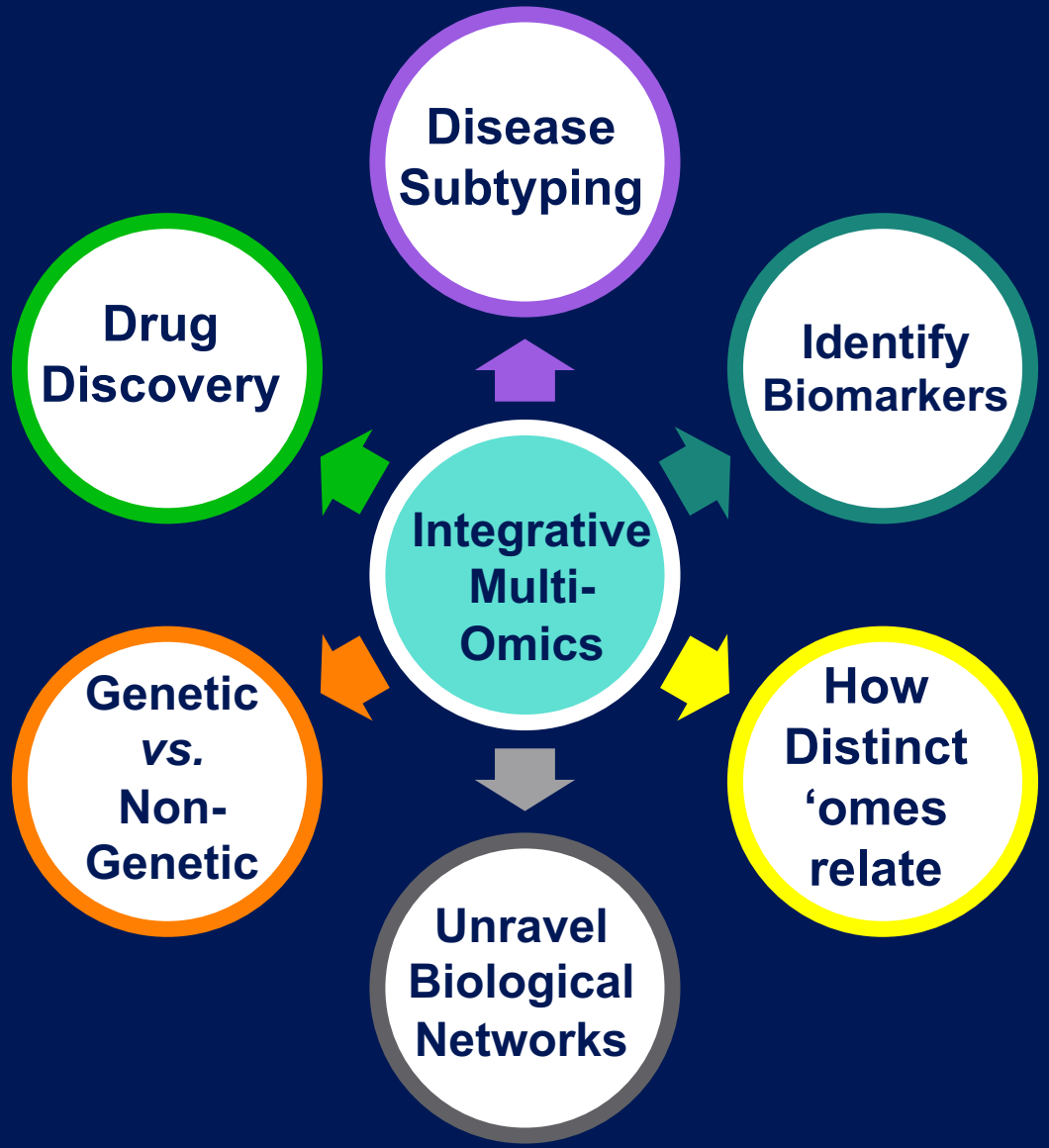
Adapted from Rhee *et al.* 2020

Multi-Omics Defined

- **Systems biology** approach
- **Multiple ‘omic** data types
- Integration → **increased insights**
- Comprehensive assessment
- High-throughput technologies
- “Big data”



Benefits of Multi-Omics Integration



NIH NHGRI

Program Objectives

Validate and enhance generalizable **multi-omics approaches** to identify meaningful biological **changes** related to health and disease

- **Explore the use** of multi-omics to **detect and assess molecular "profiles"** associated with healthy and disease states
- Leverage exploratory studies to **develop** **generalizable data harmonization, integration, and analysis** **methods, best practices, and standards**
- **Create a multi-dimensional dataset** **(and visualization portal)** that is available to the wider research community and is interoperable with existing resources (TOPMed, GTEx etc.)

NIH NHGRI

Program Structure: Multi-Omics Consortium

Data Analysis & Coordination Center (1)

RFA-HG-22-010

Steering Committee
Working Groups

Disease Study Sites (**Up to 6**)

RFA-HG-22-008

‘Omics Production Centers (**1-2**)

RFA-HG-22-009

NIH NHGRI

Disease Study Sites

RFA-HG-22-008

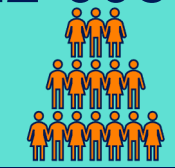
Disease Study Sites (**Up to 6**)

NIH NHGRI

Disease Study Sites: Clinical Conditions

- Focus on a disease area **with evidence** that integrative multi-omics would be impactful to:
  - Define molecular profiles associated with healthy and disease states
  - Detect changes to profiles over time
- **Detect** aspects of **disease progression** within program timeframe
  - Change from healthy to disease
  - Transitions across stages of a disease
- **Examples** of disease areas in scope
  - Diseases with distinct stages or transitions
  - Relapsing diseases, with exacerbations and remissions
  - Disease strongly impacted by environmental exposure

Disease Study Sites: Enrollment



- Each DSS should **enroll**:
  - 200 participants with disease
  - 100 generally healthy participants
  - Minimum of 75% of individuals → self-identified racial and/or ethnic communities expected to have genetic ancestries currently understudied
- Ensure appropriate recruitment, retention, and community engagement strategies
- Obtain **broad data sharing consent** for the collection of multiple ‘omic, phenotypic, and environmental exposure measures at multiple times

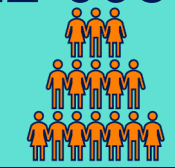
Disease Study Sites: Phenotypes, exposures, biosamples, and 'omes

RFA-HG-22-008

- Each DSS should:
  - Collect phenotypic and environmental exposure data, including SDOH, using **standard measures** when available (e.g., PhenX Toolkit)
  - Collect biosamples at **minimum of 3 timepoints**
    - blood, urine, or saliva
    - tissue and/or biopsy samples, if appropriate
  - Budget for **management of biosamples**, including costs for submitting to OPCs
  - **Propose** 'omic technologies to be applied, considering unique disease characteristics

Disease Study Sites: Analysis and Methods Development

RFA-HG-22-008

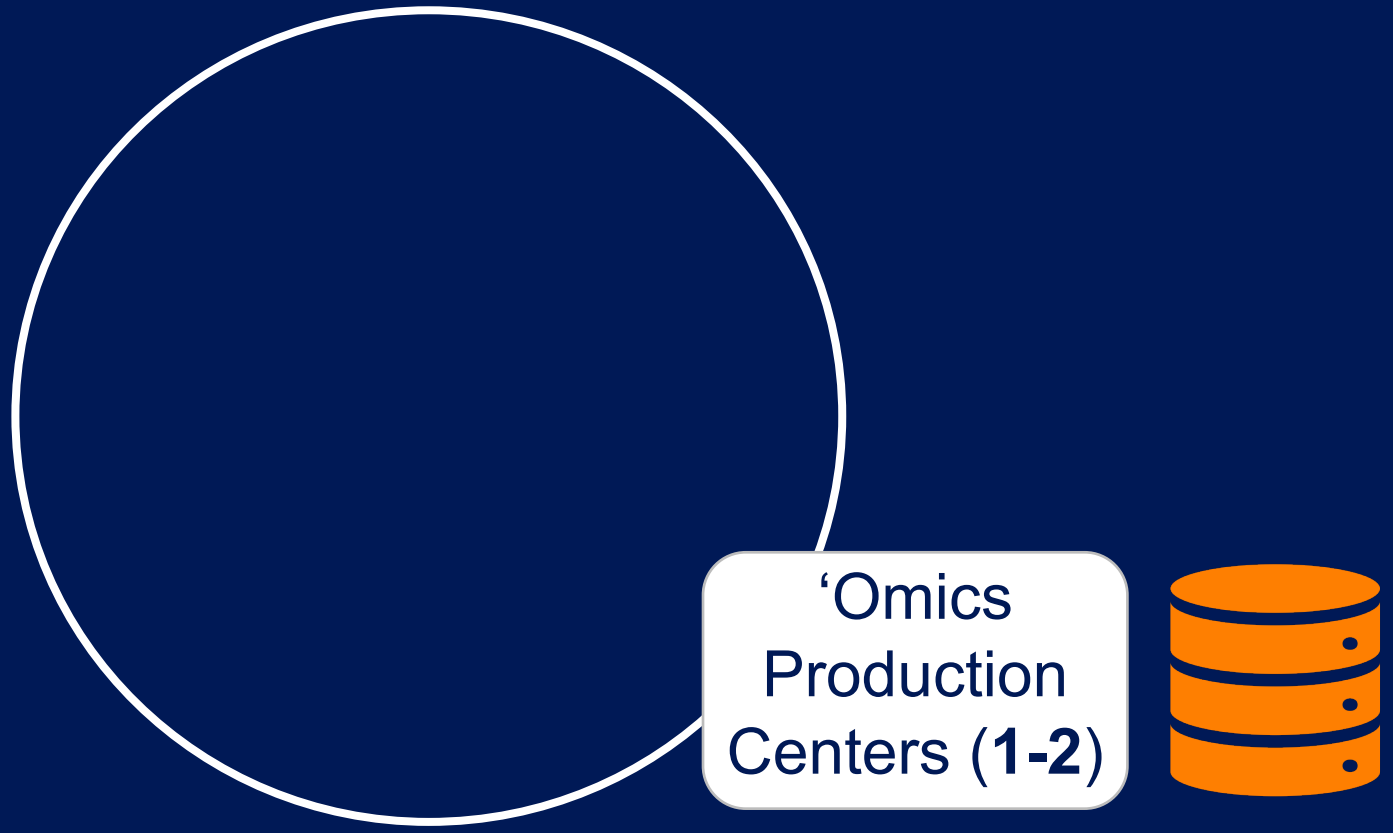


- Each DSS should:
  - Propose **technological and computational analytical** protocols appropriate for the disease and proposed ‘omic technologies
  - Propose methodological approaches to **address challenges** related to:
    - Data harmonization
    - Quality control and batch effects
    - Optimization of computational and statistical methods to integrate and analyze multiple modalities
    - Analysis of multiple ‘omic data measures from the same sample
  - Demonstrate that approaches and methods proposed are (or would be) **generalizable** across ancestrally diverse populations

NIH NHGRI

‘Omics Production Center(s)

RFA-HG-22-009



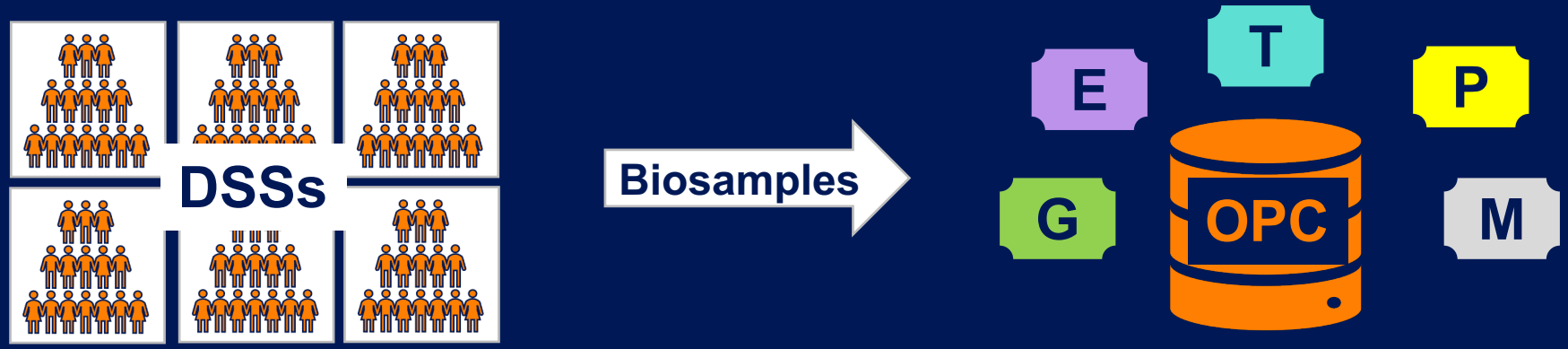
‘Omics Production Centers: ‘Omic Data Types

- The OPC(s) will use **high-throughput technologies** to produce:

‘Omic Data Type	Expected Data	Assays
Genomics	Genes, Regulatory regions, and variants	WGS
Epigenomics	Chemical modifications to DNA, RNA, and associated proteins	ATAC-Seq, ChIP-Seq or methylation arrays
Transcriptomics	Levels and expression patterns of transcript isoforms	RNA-Seq
Proteomics	Identification and quantification of proteins	Targeted or untargeted MS- or aptamer-based
Metabolomics	Identification and quantification of small metabolites	Targeted or untargeted MS- or NMR spectroscopy-based

‘Omics Production Centers

- Preferred Option:
  - 1 OPC
  - 5 ‘omics data types



- Will be considered:
  - 2 OPCs
  - Each produce a minimum of 3 ‘omics data types
  - **Genomics** + proteomics or metabolomics + Any of the remaining three

‘Omics Production Centers: Analysis and Methods Development

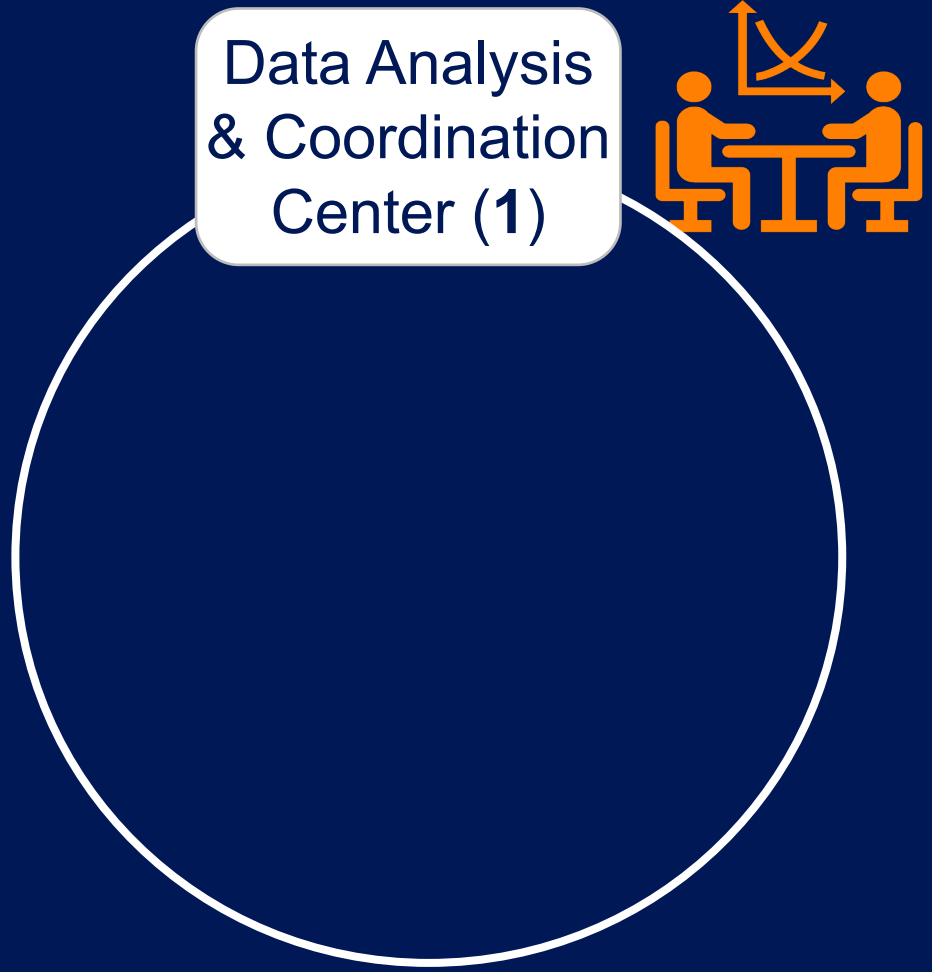
RFA-HG-22-009

- Each OPC should:
  - Propose **technological and computational analytical** protocols appropriate for the proposed ‘omic technologies
  - Propose methodological approaches to **address challenges** related to:
    - Data harmonization
    - Quality control and batch effects
    - Optimization of computational and statistical methods to integrate and analyze multiple modalities
    - Analysis of multiple ‘omic data measures from the same sample
  - Demonstrate that approaches and methods proposed are (or would be) **generalizable** across ancestrally diverse populations

NIH NHGRI

Data Analysis and Coordination Center

RFA-HG-22-010



NIH NHGRI

Data Analysis and Coordination Center: Coordinate Consortium Activities

RFA-HG-22-010

- **Coordinate Consortium activities and logistics**
  - Year-long, consortium-wide, **protocol development** process
  - **Collaborative data analysis** efforts
  - **Joint development and optimization of methods**
  - **Use of AnVIL workspaces** for analysis and methods development
  - Perform **outreach** and ensure appropriate **engagement** with all stakeholders
  - Produce a **user-friendly website** that highlights products of the Consortium
- **Facilitate submission** of data to the **AnVIL** platform
  - Define **submission process**
    - Phenotypic, environmental exposure data (DSSs)
    - Molecular data (OPC(s))
  - Define **data standards and data model**, utilizing widely available protocols
  - **Validate** all Consortium data prior to submission

NIH NHGRI

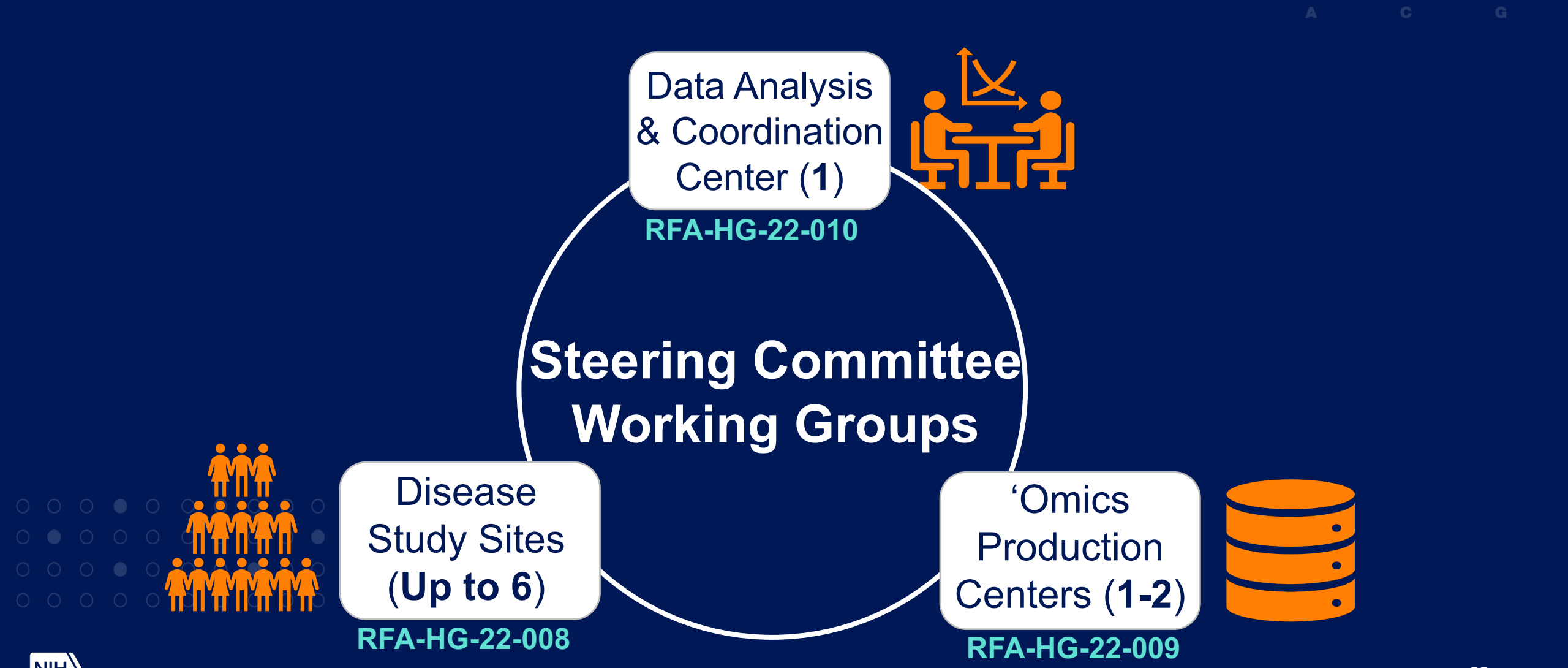
Data Analysis and Coordination Center: Create multi-dimensional Dataset

RFA-HG-22-010

- Create dataset that includes:
  - persons from **ancestrally diverse** populations
  - persons **with and without** specific **diseases**
  - harmonized and standardized **phenotypic and environmental exposure** data
  - harmonized and standardized data for **all or most 'omes** for each sample
  - data from **all time points**
  - **associated meta-data** to facilitate links across data types
- Liaise with the **AnVIL** to make dataset accessible as **controlled access data**
- Submit novel and previously known variant interpretations to **ClinVar**
- Create **open-access online data resource** for storing, cataloging, searching, and sharing aggregated **summary-level data** from the program

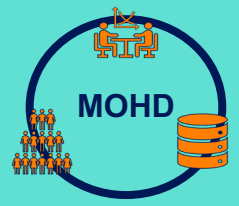
NIH NHGRI

Consortium-wide Activities



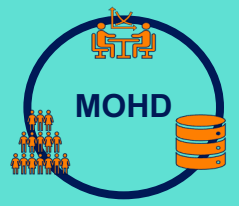
NIH NHGRI

Consortium-wide collaborative Planning



- **Year 1** will be devoted to **development of protocols, plans, and approaches**
  - Community engagement approaches
  - Recruitment strategy
    - characteristics of participants
    - consent language
  - Plan for defining, measuring, and collecting core phenotypic and environmental exposures
  - Biospecimen procurement and processing approaches
  - Selection of ‘omics assays,
  - Methodology
  - Common data models
  - Plan for utilizing the AnVIL platform

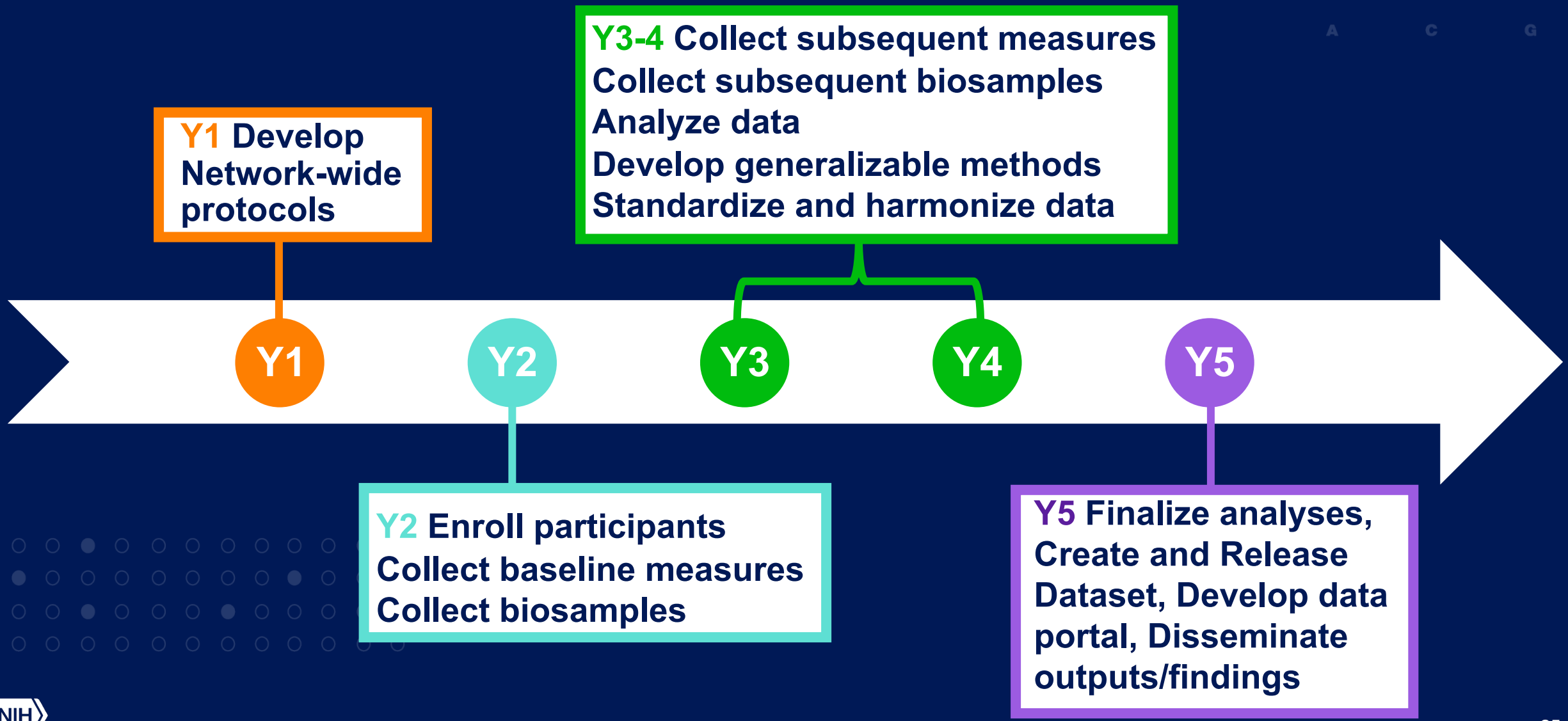
Collaborative Analysis & Methods Development



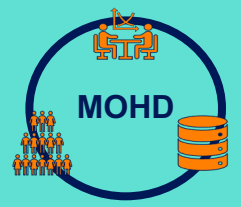
- Data Analysis
  - Integrate multiple data types → ‘omic, phenotypic, and environmental exposure
  - Define molecular profiles
  - Assess associations with healthy or disease states
  - Detect patterns of change in profiles and associations
  - Identify genetic variants, genes, and gene networks involved in disease
  - Determine how networks are impacted by phenotypic and environmental data
  - Explore potentially causal relationships that lead to changes in health status

- Methods Development
  - Develop generalizable methods, best practices, and standards
  - Address data harmonization, integration, and analysis challenges and gaps

Consortium Timeline



Consortium Governance



- Steering Committee (SC) → Oversee project goals and progress
  - PIs from DSSs, OPC(s), DACC
  - NIH program staff
- Working Groups (WG)
  - Facilitate collaborative work
  - Standardize approaches
  - Members of the Consortium
- External Scientific Panel (ESP)
  - Convened by NIH
  - Provide scientific expert recommendations

Consortium Diversity Goals

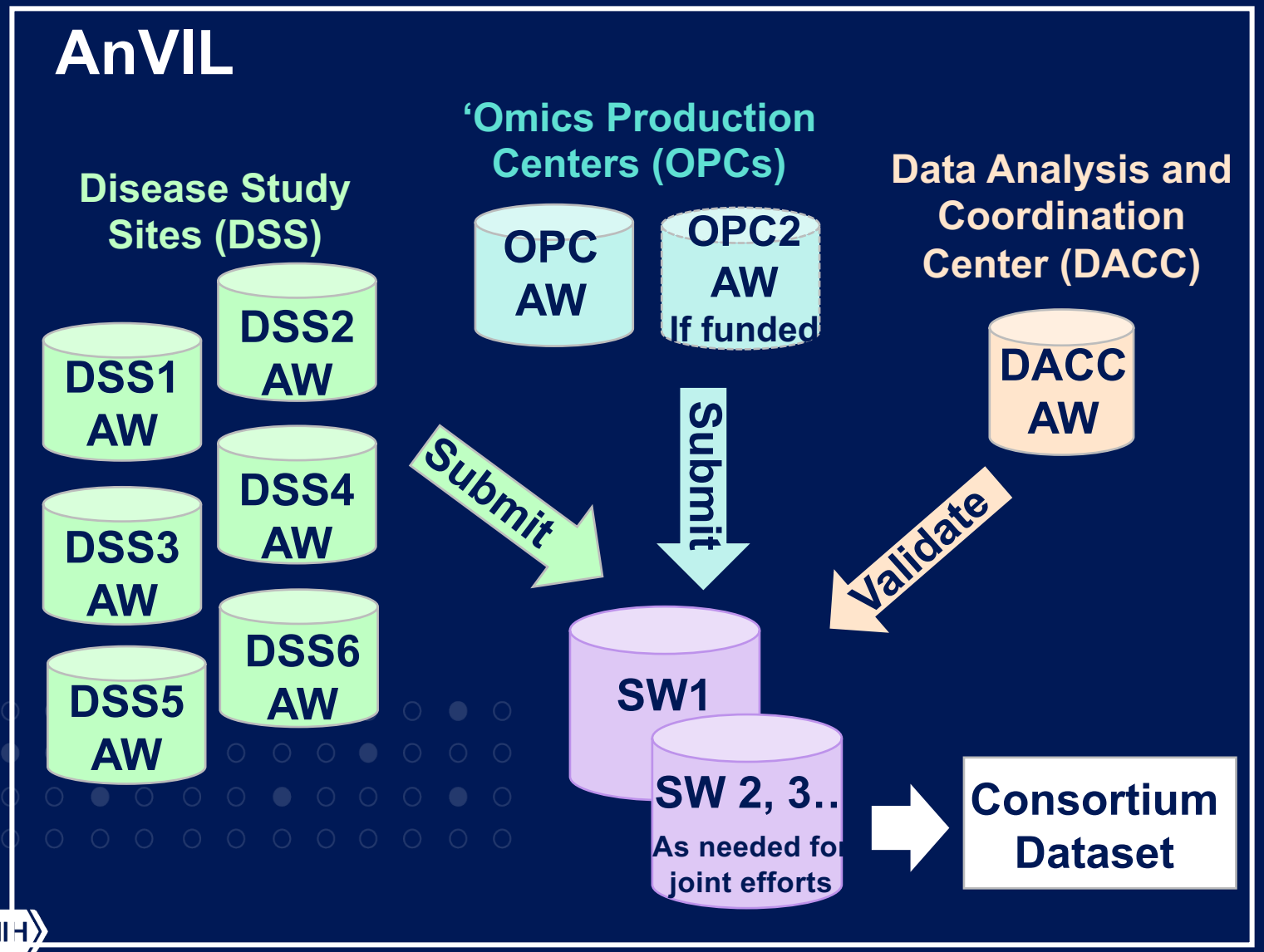
- To increase the diversity of genetic ancestries:
  - 75% of individuals from racial and ethnic communities expected to have genetic ancestries underrepresented in genomics
  - Establish recruitment and community engagement strategies, including outreach to racial and ethnic minority communities
- To enhance the excellence and inclusivity of the research environment:
  - Assemble richly diverse study teams

The AnVIL Eco-system

- AnVIL = Analysis, Visualization and Informatics Lab-space
- NHGRI-designated data repository
- Cloud storage and computing platform
- Provides secure environment for data storage and analysis
- Enables analysis by each site or center
- Enables management of Consortium data by the DACC

NIH
NHGRI

AnVIL and Multi-Omics Consortium



Analysis Workspace (AW)
- Perform Analyses
- Store Outputs
- Develop methods
- Harmonize data

Shared Workspace (SW)
- Submit data
- Validate data
- Produce dataset
- Joint efforts, as needed

NIH NHGRI

GDS Expectations for Multi-Omics Program

- Address a Data Sharing Plan in the Resource Sharing Plan
- Where to submit data:
  - Submit all controlled-access data to the AnVIL
  - Submit variant interpretations to ClinVar
- Consent
  - General Research Use consent for future research use and broad sharing
- Share comprehensive metadata and phenotypic, clinical, and environmental exposure data associated with the study;
  - Use standardized data collection protocols and survey instruments
  - Use standardized notation for metadata (e.g., controlled vocabularies or ontologies)

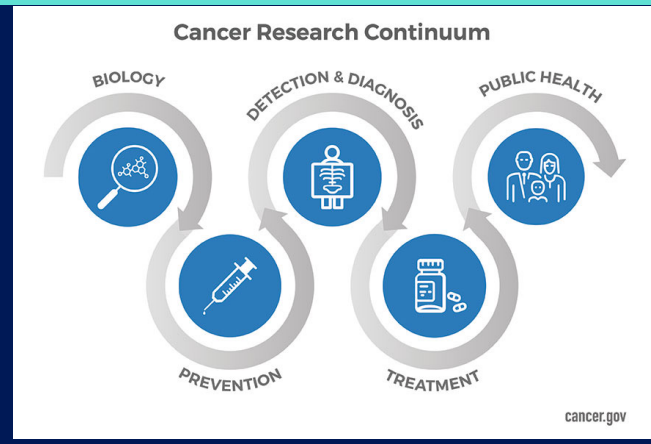
http://sharing.nih.gov
https://www.genome.gov/about-nhgri/Policies-Guidance/Genomic-Data-Sharing

National Institute of Environmental Health Sciences (NIEHS) Interests

- The incorporation of multi-dimensional exposure data with other ‘omics data to inform mechanistic understanding of diseases/health outcomes, particularly where exacerbation or progression of disease is impacted by environmental exposures
- Harmonization and integration of complex environmental health data with other multi-omics and phenotypic data
- Innovative methods and approaches for integrating various types of environmental data (including location-based data, personal monitoring, biomonitoring, and validated questionnaires, etc.) with other omics data
- Strong community engagement (especially in the context of exploring environmental health disparities and environmental justice issues)
- Exposures of interest include, but are not limited to: air pollution, extreme weather/climate, pesticides, metals, and other chemical stressors and biologically derived toxins

National Cancer Institute (NCI) Interests

NCI supports cancer research to advance scientific knowledge to prevent, detect, diagnose, and treat cancer and to help cancer survivors live longer, healthier lives.



- For this RFA:
  - NCI is interested in supporting a disease study site focused on cancer
  - Studies collecting both tumor and normal tissues from cancer cases are strongly encouraged and will be prioritized for funding
- Unique opportunities supported by this RFA include:
  - Informing how to scale to larger population-based studies
  - Supporting collection of longitudinal samples
  - Integrating environmental data with multi-omics data
  - Emphasis on social determinants of health and ancestrally diverse populations

Review Criteria

- **Significance** → Will the completion of the project result in a useful resource for the research community to drive the field forward?
- **Investigator(s)** → Are the PD(s)/PI(s), collaborators, and other researchers well suited for the project in terms of expertise and effort?
- **Innovation** → Are novel strategies employed to ensure success?
- **Approach** → Is the conceptual design and overall operating plan adequate?
- **Environment** → Are the institutional support, equipment and other physical resources available for the successful completion of the project?

RFA Key Dates

RFA-HG-22-008 – Disease Study Sites (DSSs)

RFA-HG-22-009 – 'Omics Production Center(s) (OPCs)

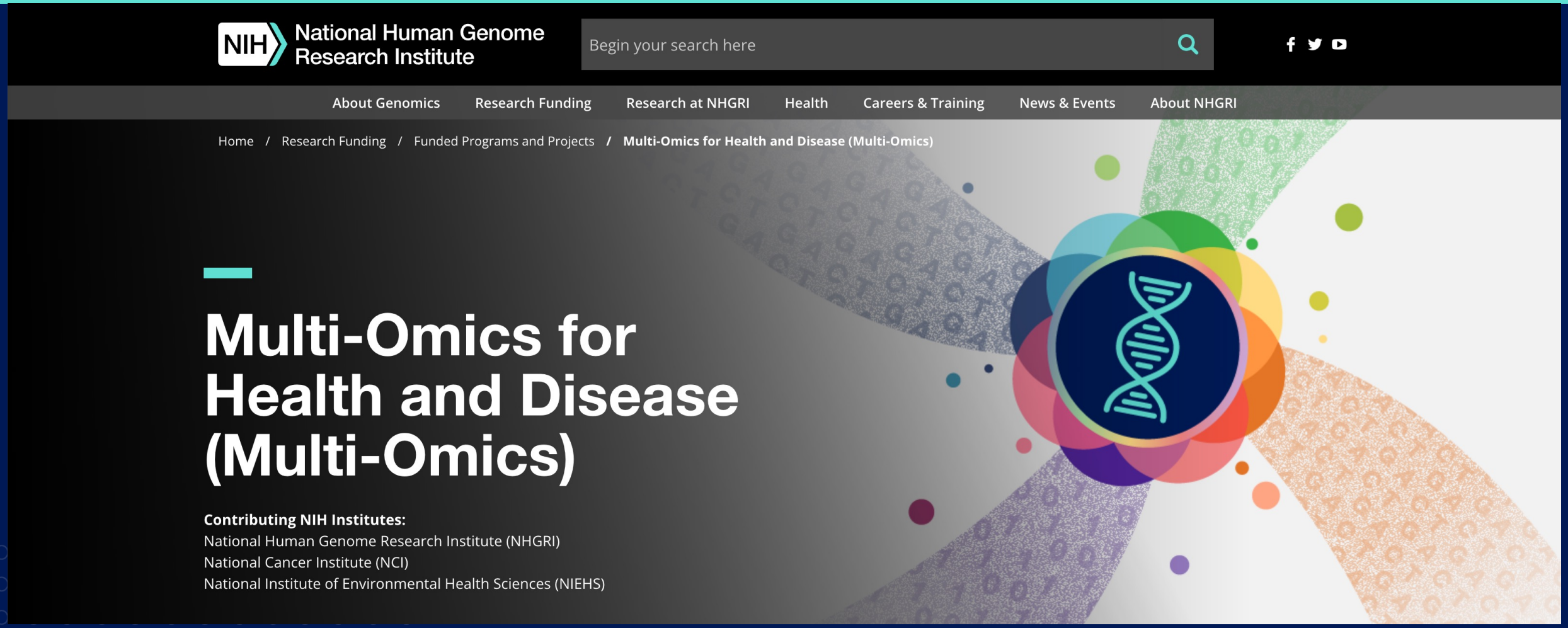
RFA-HG-22-010 – Data Analysis and Coordination Center (DACC)

- **Letters of Intent – Due October 18**
- **Applications – Due November 19**
- **Awards – Summer 2023**

Please send all questions to:
MultiOmicsProgram@mail.nih.gov

NIH NHGRI

Resources



https://www.genome.gov/research-funding/Funded-Programs-Projects/Multi-Omics-for-Health-and-Disease

Frequently Asked Questions

1. **What diseases are eligible?**

Frequently Asked Questions

1. What diseases are eligible?
2. Can all participants enrolled by a DSS be from one understudied population?

NIH NHGRI

Frequently Asked Questions

1. What diseases are eligible?
2. Can all participants enrolled by a DSS be from one understudied population?
3. Can existing cohorts be used?

Frequently Asked Questions

1. What diseases are eligible?
2. Can all participants enrolled by a DSS be from one understudied population?
3. Can existing cohorts be used?
4. How many ‘omes are expected?

Frequently Asked Questions

1. What diseases are eligible?
2. Can all participants enrolled by a DSS be from one understudied population?
3. Can existing cohorts be used?
4. How many ‘omes will be produced?
5. How well established should the environment-disease connection be?

Frequently Asked Questions

1. What diseases are eligible?
2. Can all participants enrolled by a DSS be from one understudied population?
3. Can existing cohorts be used?
4. How many ‘omes will be produced?
5. How well established should the environment-disease connection be?
6. What should the Year 1 budget be given the program timeline?

NIH NHGRI

Applicant Q&A


The Forefront
of Genomics®