

C G T A C G T A
A C G T A C G T

Concept Clearance: Multi-omics for Health and Disease

Joannella Morales, PhD

On behalf of the Multi-Omics Team: Erin Ramos and Teri Manolio

Division of Genomic Medicine

NHGRI Advisory Council Meeting
February 7, 2022



National Human Genome
Research Institute

—
The **Forefront**
of **Genomics**[®]
—

Outline

A C G

- Preliminary Council Feedback
- Background and Rationale
- Objectives and Scope
- Budget
- Summary



Outline

A C G

- **Preliminary Council Feedback**
- Background and Rationale
- Objectives and Scope
- Budget
- Summary



Recent NHGRI efforts on Multi-Omics

- **June 2021**: Hosted workshop to review state of the field and gather recommendations
- **Sept 2021**: Provided Report to Council on workshop and recommendations
- **Nov 2021**: **New approach**: Sought preliminary feedback from Council early in the concept development process



Preliminary Council Feedback

- **11/18**: H. Chang, I. Kullo, L. Pennachio, S. Rich, O. Troyanskaya;
11/29: H. Dietz
- **Outline version** of concept was presented
- Council **generally enthusiastic**; No follow up meeting requested
- Recommendations:
 - Clearly articulate the primary objective and desired outcomes
 - Increase linkage between sample collection and data production
 - Centralize data analysis while ensuring collaborative effort
 - Expand proposed list of 'omics assays

Outline

A C G

- Preliminary Council Feedback
- **Background and Rationale**
- Objectives and Scope
- Budget
- Summary



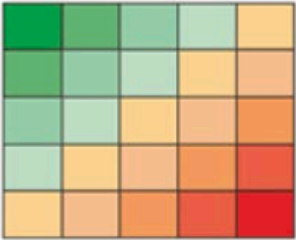
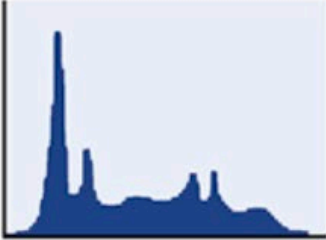



Background

Advances in **high-throughput technologies**

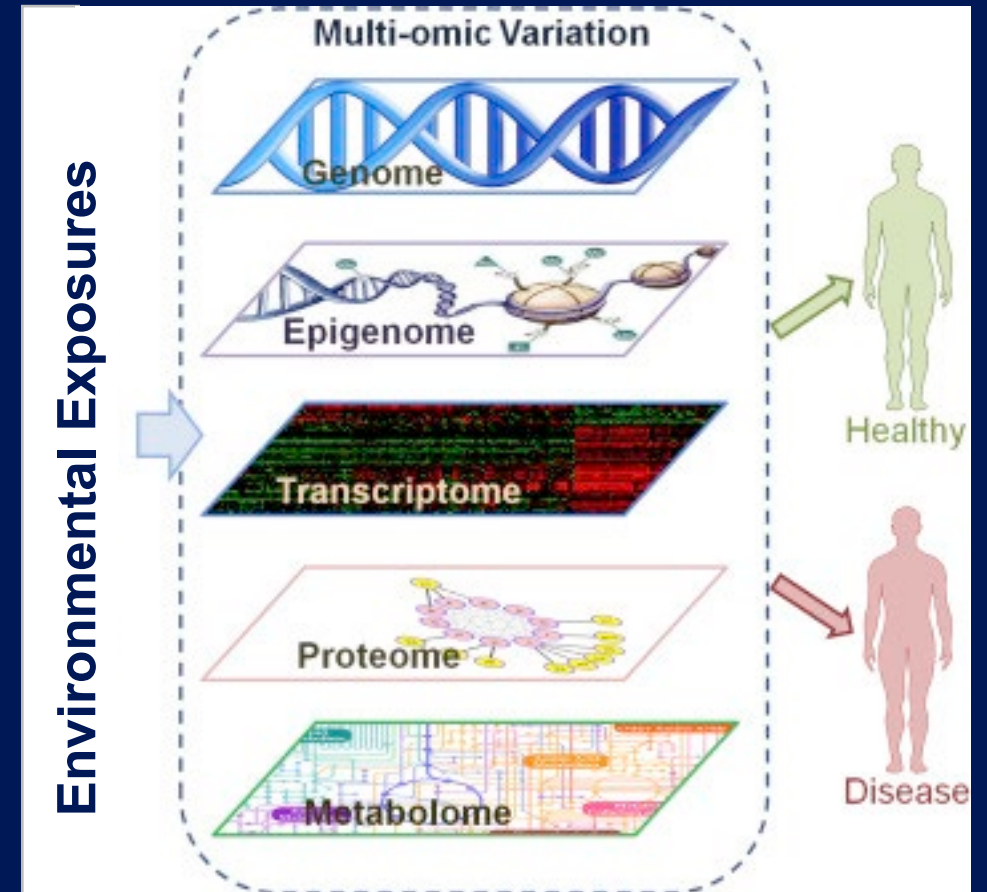


Increased **access** to distinct molecular data types (**'omics data**)

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 intracellular proteins	Endogenous circulating metabolites Xenobiotics

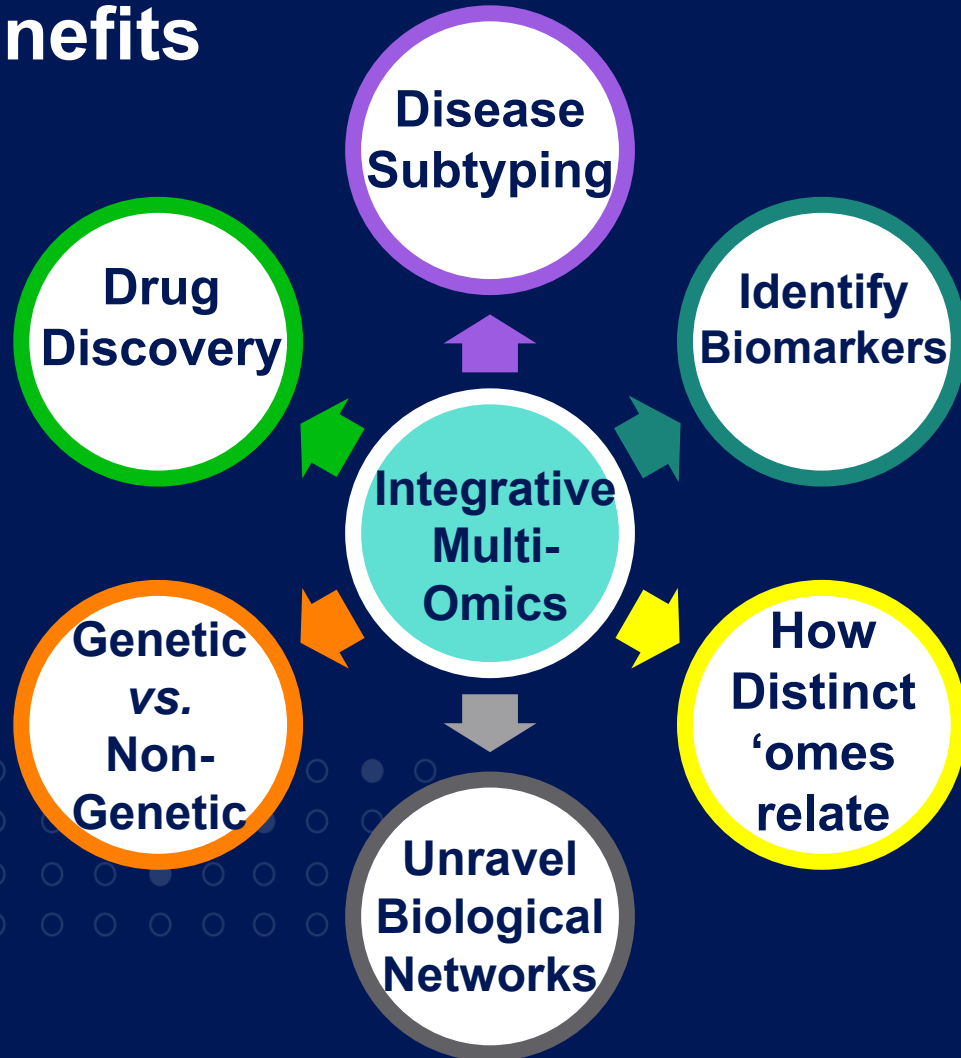
Multi-Omics Defined

- **Systems biology** approach
- Data sets are **multiple 'omes**
- Integration → **increased insights**
- Comprehensive assessment
- High-throughput technologies
- “Big data”
- Interdisciplinary expertise

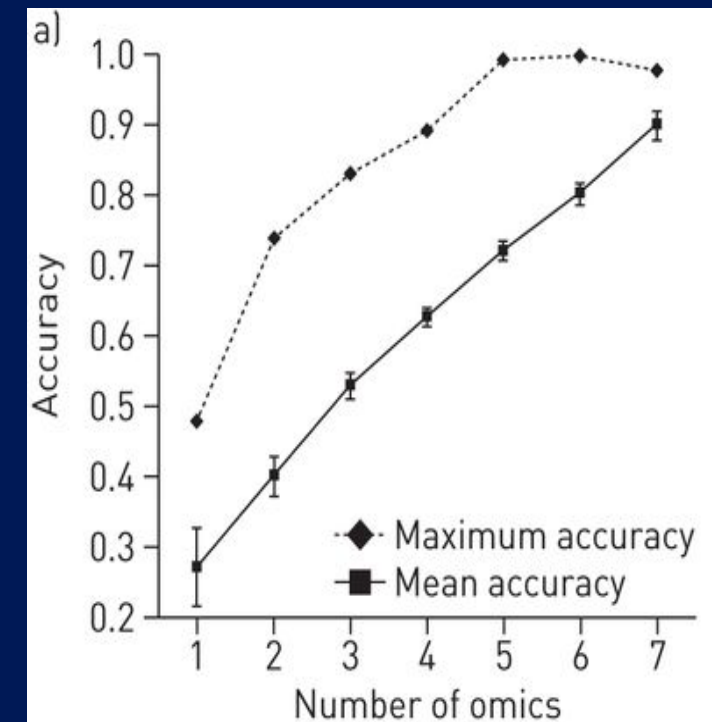


Multi-Omics Integration

Benefits



Some successes! e.g. Improved classification of COPD



Integration of multi-omics datasets enables molecular classification of COPD; Li *et. al.*

Gaps and Opportunities

Production of multiple 'omics data from **same sample**

- Inter and intra 'ome variability
- Non-uniform content across platforms and assays
- Lack of consensus approaches for QA and imputation

Computational methods to integrate, analyze, and interpret

- Multiple 'omes from the same sample
- Multi-omics + clinical + environmental exposure data
- Across diverse populations

Prospective data collections

- Informed consent for broad data sharing/ General Research Use
- Well-described and harmonized metadata
- Collection of major 'omics data types

Outline

A C G

- Preliminary Council Feedback
- Background and Rationale
- **Objectives and Scope**
- Budget
- Summary

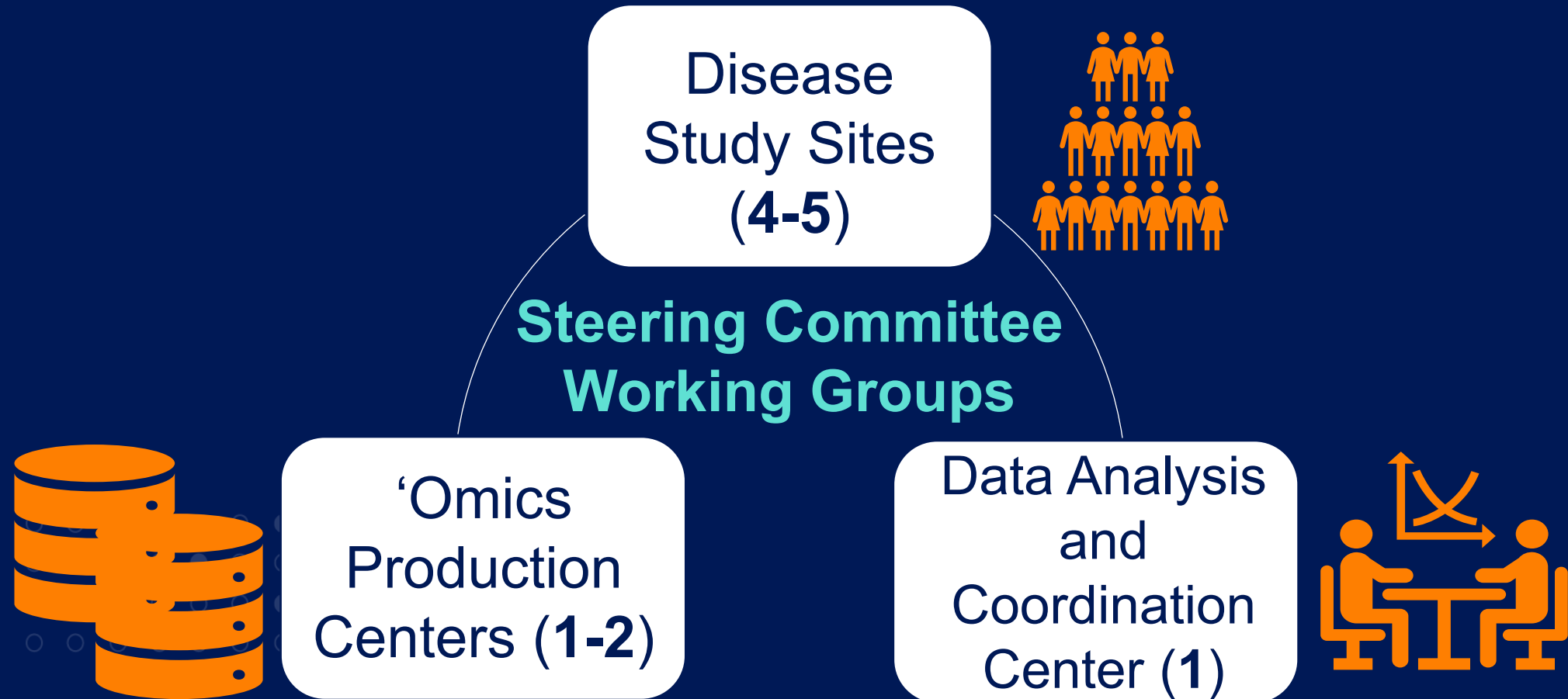


Purpose

- **Validate and enhance** generalizable multi-omics approaches to identify biological changes related to 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 portal** that is available to the wider research community and is interoperable with existing resources (TOPMed, GTEx etc.)

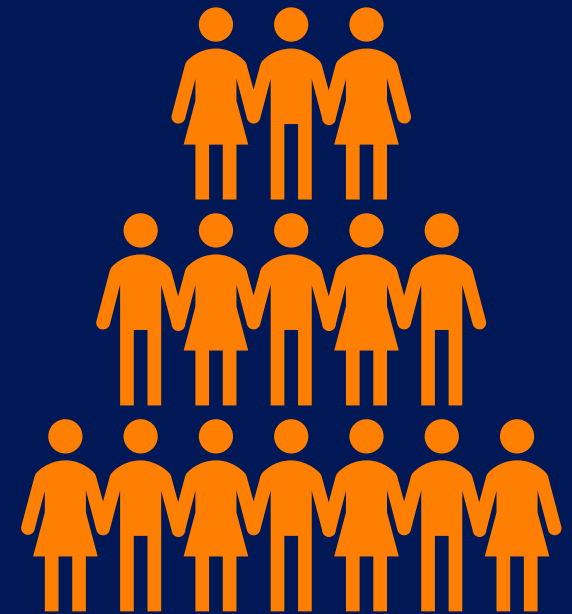
Program Structure: Multi-Omics Consortium

Three FOAs to support:



Disease Study Sites (DSS, 4-5)

- Focus on a disease area with evidence that integrative multi-omics would be impactful:
 - Relapsing diseases
 - Heterogeneous diseases
 - Diseases with distinct stages or transitions
- Recruit / re-consent 200-300 participants:
 - Appropriate consent and community engagement
 - 75% underrepresented ancestral backgrounds
 - Standard measures for phenotypes and environmental exposures, (e.g. SDOH)
- Collect specimens at minimum of 3 timepoints (tissues/cells, as needed)



'Omics Production Centers (OPC, 1-2)

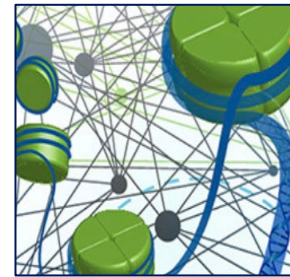
- Utilize state-of-the-art high-throughput molecular assays
- Minimum of 3 'omic data types, 1 should be non-nucleic acid-based



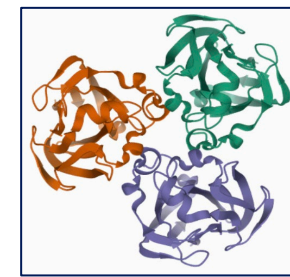
WGS



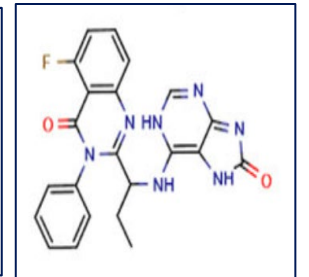
RNASeq



Methylation Arrays
ATAC-Seq
CHIP-Seq



Targeted; Untargeted
Mass Spectrometry (MS)
NMR Spectroscopy



Other 'omic data types will be considered

Data Analysis and Coordination Center (DACCC, 1)

- Receive, track, and catalog data from DSSs and OPCs
- Coordinate consortium-wide activities including:
 - Consensus on recruitment strategies and the choice of 'omic data types and assays
 - Developing the consortium-wide data analysis process
 - Liaising with the AnVIL to facilitate data sharing
 - Establishing working groups for methods development
 - Producing the standardized multi-dimensional data set
 - Diverse ancestral backgrounds
 - Participants with and without disease
 - Data from all or most 'omes **AND** all time points
 - Key meta-data
 - Developing the visualization portal
 - Rapidly disseminating consortium outputs



Consortium-wide Expertise, Activities, and Timeline

Network-wide protocol development period (Y1)

- Recruitment strategy: participants w/ disease + pooled controls
- Community engagement and informed consent
- Core phenotypic and environmental exposure measures
- ‘Omics data types and assays
- Biospecimen procurement, processing, and analysis method

Portal Development and Dissemination (Y4-5)

- Develop portal; FAIR principles
- Disseminate generalizable methods and findings

Y1

Y2

Y3

Y4

Y5

Activities (Y2-3)

- Integration and analysis
 - Apply computational approaches;
 - Interpret molecular “profile” associations;
 - Explore gene networks;
 - Assess causal relationships
- Develop generalizable methods, best practices, and standards
- Create standardized and harmonized multi-dimensional data set

Diversity and the Multi-Omics Consortium

- Overrepresentation of European ancestry in research:
 - Undiscovered genetic variation
 - Inaccurate risk prediction tools
 - Inequity in the distribution of benefits from research
- To increase the diversity of genetic ancestries:
 - Minimum of 75% individuals from underrepresented ancestral backgrounds
 - Establish recruitment, retention, and meaningful community engagement strategies, including outreach to racial and ethnic minority communities
- Promise of Genomics requires a diverse genomic workforce
- To enhance the excellence and inclusivity of the research environment:
 - Strongly encouraged to assemble study teams from diverse backgrounds, including individuals from underrepresented groups



Relationship to Ongoing Activities

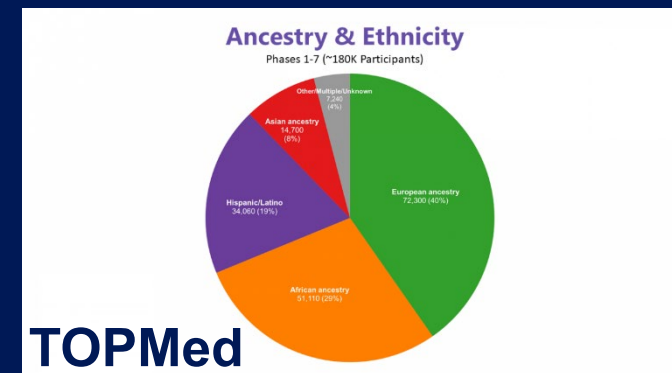
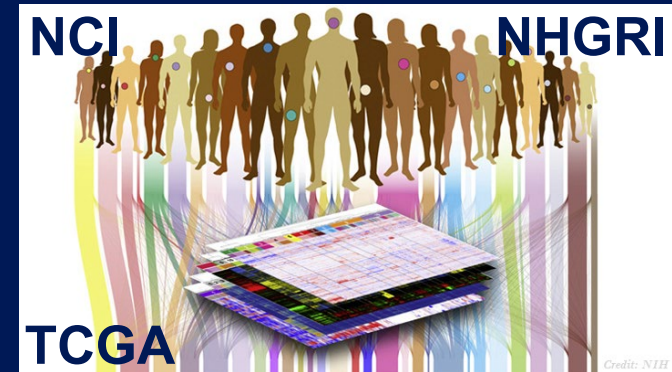
- Complements existing NIH investments
- Differs from other initiatives:
 - **Prospective** enrollment and study design
 - Multiple disease areas
 - Collection of specimens at **multiple time points**
 - Production of **major 'omics data types** from the same sample
 - Consent for future use and **broad data sharing** without data use limitations

NIMH: The Brain Initiative

NIDDK: Longitudinal 'omics in diabetes

NIAID: 'Omics for predictive modeling of infectious diseases

NIA: Integrative 'omics for aging and neurodegenerative diseases



Common Fund



Outline

A C G

- Preliminary Council Feedback
- Background and Rationale
- Objectives and Scope
- **Budget**
- Summary



Proposed Budget

- Approximately \$8M per year
- Duration: 5 Years
- Total for 5 Years: \$40M
- Total number of samples and sites will depend on available funds

A C G



Outline

A C G

- Preliminary Council Feedback
- Background and Rationale
- Objectives and Scope
- Budget
- **Summary**



In Summary, NHGRI Proposes...

- **Explore the use** of multi-omics to detect and assess molecular “profiles”
- Leverage these exploratory association studies to **develop generalizable methods, best practices, and standards** for the optimal application of multi-omics;
- **Create a multi-dimensional dataset** that is **available and interoperable**

The primary goal is to **validate and enhance generalizable multi-omic approaches** to identify meaningful biological changes related to health or disease.

Acknowledgements

A C G

Stream B

NHGRI

NIH Colleagues

Council Members

Thank You!



Questions/Comments





The **Forefront**
of **Genomics**[®]