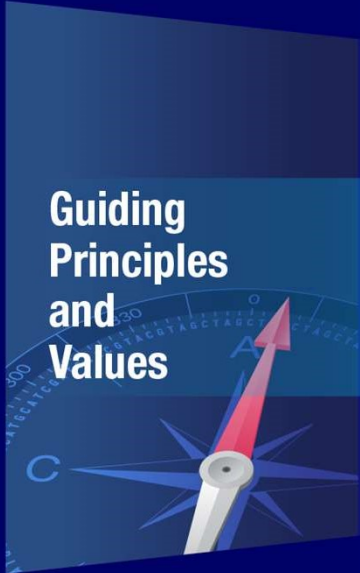# Data Science at the Forefront of Enhancing Diversity in Genomics

Future Directions of the NHGRI Analysis, Visualization, and Informatics Lab space (AnVIL)

## Vence L. Bonham, JD
**Acting Deputy Director, NHGRI**

Guiding Principles and Values

Robust Foundation for Genomics

Breaking Down Barriers

Compelling Genomics Research Projects

The **Forefront** of **Genomics**®

## Box 1

# Guiding principles and values for human genomics

• **Maintain an overarching focus on using genomics to understand biology, to enhance knowledge about disease, and to improve human health** — genomics is now foundational across the entire continuum of biomedical research, from deciphering fundamental principles of biology to translating that knowledge into disease prevention and medical advances.

• **Strive for global diversity in all aspects of genomics research, committing to the systematic inclusion of ancestrally diverse and underrepresented individuals in major genomic studies** — attention to diversity in genomics research is both socially just and scientifically essential, which includes meaningful, sustained partnerships with diverse communities in the design and implementation of research studies, the propagation of research findings, and the development and use of new technologies.

• **Maximize the usability of genomics for all members of the public, including the ability to access genomics in healthcare** — engagement, inclusion, and understanding the needs of diverse and medically underserved groups are required to ensure that all members of society benefit equitably from genomic advances, with particular attention given to the equitable use of genomics in healthcare that avoids exacerbating and strives towards reducing health disparities.

• **Champion a diverse genomics workforce** — the promise of genomics cannot be fully achieved without attracting, developing, and retaining a diverse workforce, which includes individuals from groups that are currently underrepresented in the genomics enterprise.

• **Provide a conceptual research framing that consistently examines the role of both genomic and non-genomic contributors to health and disease** — routinely considering the importance of social and environmental factors that influence human health (and the interactions among those components and genomics) will be important for the comprehensive understanding of most human diseases.

• **Promote robust and consistently applied standards in genomics research** — the use of carefully defined standards (for example, those for generating, analysing, storing, and sharing data) has benefited genomics in numerous ways, and this must include appropriate privacy and data-security protections for those participating in genomics research.

• **Embrace the interdisciplinary and team-oriented nature of genomics research** — starting with the Human Genome Project, some of the most challenging genomics endeavours have benefited from the creation and management of large, interdisciplinary research collaborations.

• **Adhere to the highest expectations and requirements related to open science, responsible data sharing, and rigor and reproducibility in genomics research** — the genomics enterprise has a well-respected history of leading in these areas, and that commitment must be built upon and continually reaffirmed.

• **Pursue advances in genomics as part of a vibrant global community of genomics researchers and funders** — the challenges in genomics require the collective energies and creativity of a collaborative international ecosystem that includes partnerships among researchers, funders, and other stakeholders from academia, government, and the commercial sector.

**Box 1**

> **· Champion a diverse genomics workforce** — the promise of genomics cannot be fully achieved without attracting, developing, and retaining a diverse workforce, which includes individuals from groups that are currently underrepresented in the genomics enterprise.

and underrepresented individuals in major genomic studies — attention to diversity in genomics research is both socially just and scientifically essential, which includes meaningful, sustained partnerships with diverse communities in the design and implementation of research studies, the propagation of research findings, and the development and use of new technologies.

**· Maximize the usability of genomics for all members of the public, including the ability to access genomics in healthcare** — engagement, inclusion, and understanding the needs of diverse and medically underserved groups are required to ensure that all members of society benefit equitably from genomic advances, with particular attention given to the equitable use of genomics in healthcare that avoids exacerbating and strives towards reducing health disparities.

**· Champion a diverse genomics workforce** — the promise of genomics cannot be fully achieved without attracting, developing, and retaining a diverse workforce, which includes individuals from groups that are currently underrepresented in the genomics enterprise.

**· Provide a conceptual research framing that consistently examines the role of both genomic and non-genomic contributors to health and disease** — routinely considering the

— the use of carefully defined standards (for example, those for generating, analysing, storing, and sharing data) has benefited genomics in numerous ways, and this must include appropriate privacy and data-security protections for those participating in genomics research.

**· Embrace the interdisciplinary and team-oriented nature of genomics research** — starting with the Human Genome Project, some of the most challenging genomics endeavours have benefited from the creation and management of large, interdisciplinary research collaborations.

**· Adhere to the highest expectations and requirements related to open science, responsible data sharing, and rigor and reproducibility in genomics research** — the genomics enterprise has a well-respected history of leading in these areas, and that commitment must be built upon and continually reaffirmed.

**· Pursue advances in genomics as part of a vibrant global community of genomics researchers and funders** — the challenges in genomics require the collective energies and creativity of a collaborative international ecosystem that includes partnerships among researchers, funders, and other stakeholders from academia, government, and the commercial sector.
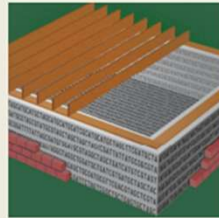
# Foundation for Genomics



Box 2

## Sustaining and improving a robust foundation for genomics

**Genome structure and function**
- Enable the routine generation and analysis of increasingly complex genomic data
- Use evolutionary and comparative genomic data to maximize understanding of genome function

**Genomic data science**
- Develop new methods and build sustainable data resources for genomics research
- Ensure facile storing, sharing, and computing on large genomic datasets
- Develop integrated knowledgebases and informatics methods for genomic medicine

**Genomics and society**
- Understand the interrelationships between genomics and the social and environmental factors that influence human health
- Empower people to make well-informed decisions about genomic data and develop data-stewardship systems that reinforce their choices
- Increase the genomic literacy of all sectors of society

**Training and genomics workforce development**
- Ensure that the next generation of genomic scientists are sufficiently trained in data science
- Train healthcare providers to integrate genomics into the clinical workflow
- Foster a diverse genomics workforce

**"Ensure that the next generation of genomic scientists are sufficiently trained in data science."**
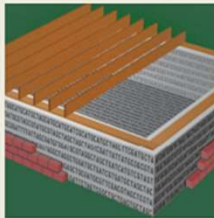Nature. 2020 Oct;586(7831):683-692

# Foundation for Genomics



Box 2

## Sustaining and improving a robust foundation for genomics

**Genome structure and function**
- Enable the routine generation and analysis of increasingly complex genomic data
- Use evolutionary and comparative genomic data to maximize understanding of genome function

**Genomic data science**
- Develop new methods and build sustainable data resources for genomics research
- Ensure facile storing, sharing, and computing on large genomic datasets
- Develop integrated knowledgebases and informatics methods for genomic medicine

**Genomics and society**
- Understand the interrelationships between genomics and the social and environmental factors that influence human health
- Empower people to make well-informed decisions about genomic data and develop data-stewardship systems that reinforce their choices
- Increase the genomic literacy of all sectors of society

**Training and genomics workforce development**
- Ensure that the next generation of genomic scientists are sufficiently trained in data science
- Train healthcare providers to integrate genomics into the clinical workflow
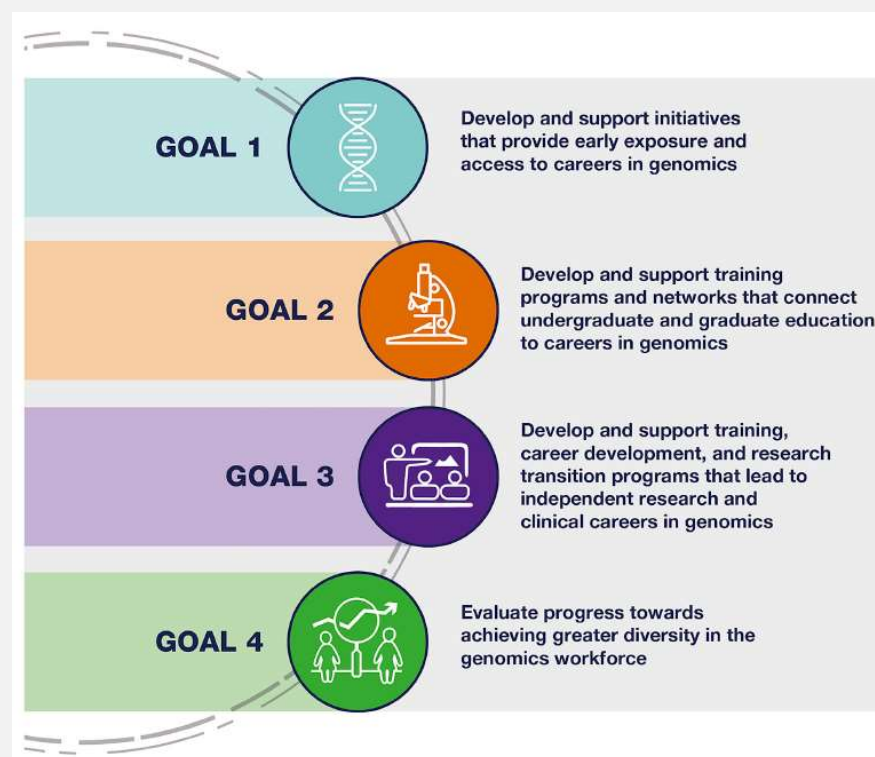- Foster a diverse genomics workforce

**"Foster a diverse genomics workforce."**
Nature. 2020Oct;586(7831):683-692

**genome.gov/workforcediversity**

# Genomic Data Science Community Network



To broaden the spectrum of diverse institutions active in bioinformatics and genomic data science, we seek partnerships with educators and researchers at Historically Black Colleges and Universities (HBCUs), Minority Serving Institutions (MSIs), Tribal Colleges and Universities (TCUs), and Community Colleges (CCs).

http://www.gdscn.org/

# Thank you

**bonhamv@nih.gov**