Capturing RNA Sequence and Transcript Diversity, From Technology Innovation to Clinical Application

Jennifer Strasburger

Program Director,
Division of Genome Sciences

Brenton Graveley

UConn Health Department of Genetics & Genome Sciences, Institute for Systems Genomics

Executive Committee Member

Presentation to NACHGR – September 19, 2022

National Human Genome Research Institute

The Forefront of Genomics®

NHGRI 2020 Strategic Vision

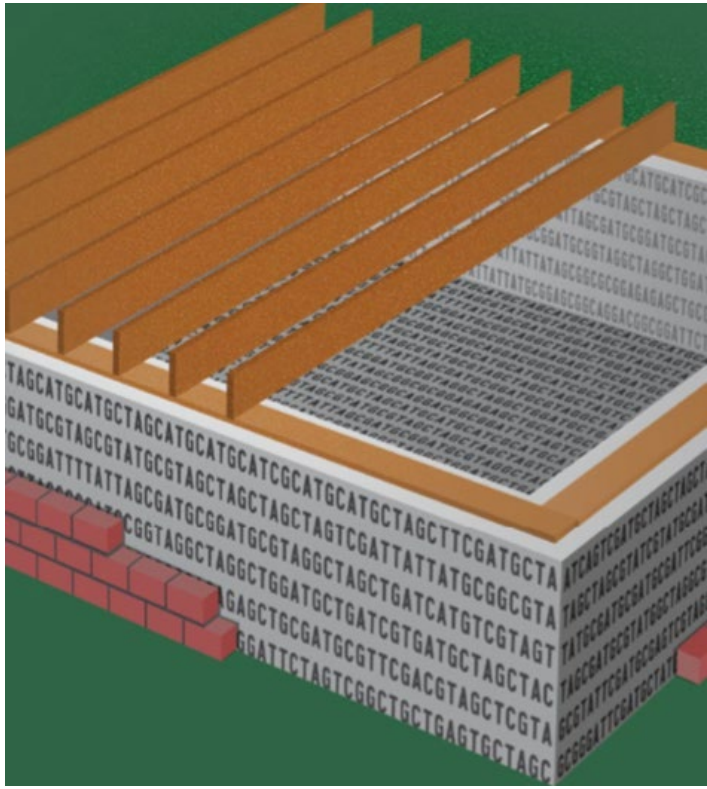
Green, E.D., Gunter, C., Biesecker, L.G. *et al.* Strategic vision for improving human health at The Forefront of Genomics. *Nature* 586, 683–692 (2020).

NIH NHGRI

Building a Robust Foundation

Genome Structure & Function:

*"**Technologies** for generating DNA sequence and other datatypes (for example, transcriptomic data, epigenetic data, and functional readouts of DNA sequences) need to be enabled at orders-of-magnitude **lower costs**, at **single-cell resolution**, at distinct **spatial locations** within tissues, and **longitudinally** over time."*



NHGRI's RNA Goals

- Direct RNA sequencing
- Modified RNA bases
- Transcript diversity
- Synthetic RNA with modifications
- Multi-omic approaches to disease and risk
- Functional genomics

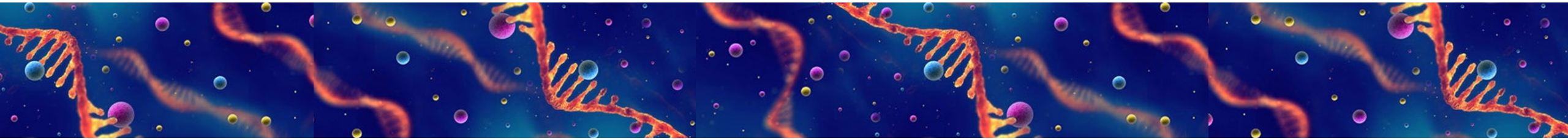
A call from the RNA community

- Technologies needed to sequence full-length RNA directly
- Address 140+ chemical RNA modifications
- Defects related to modifications account for 100+ human diseases
- Informatics to detect and identify all modifications & splice variants

NIH
NHGRI

Capturing RNA Sequence and Transcript Diversity, From Technology Innovation to Clinical Application

- Virtual workshop hosted by NHGRI & NIEHS
- May 24-26, 2022



www.niehs.nih.gov/news/events/pastmtg/2022/rnaworkshop2022/index.cfm

The Planning Team

Thanks to the many others who participated and helped with planning efforts!

Executive Committee:

- Vivian Cheung
- Pete Dedon
- Brenton Graveley
- Blanton Tolbert

Avanti Corporation:

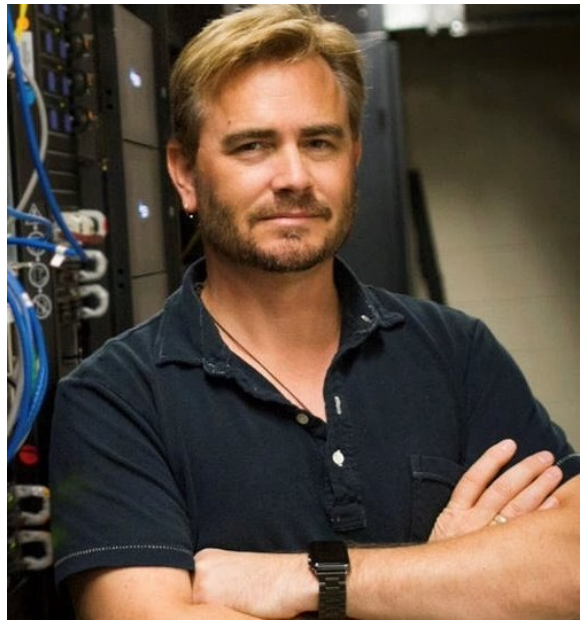
- Justin Crane
- Manashree Malpe
- Devan Smith-Brown

NHGRI & NIEHS Staff:

- Katie Bardsley
- Michelle Heacock
- Kim McAllister
- Alicia Ramsaran
- Dan Shaughnessy
- Mike Smith
- Jennifer Strasburger
- Fred Tyson
- Leroy Worth

Brenton Graveley

- UConn Health
- Chair of the Department of Genetics and Genome Sciences
- Associate Director of the Institute for Systems Genomics
- Workshop Executive Committee Member
- Former NACHGR Member



State of the Field

- Most RNAs involved in active gene expression are **very long** (>1kb), **fragile,** and **low abundance**
- Functionally **relevant information** within RNA is **not captured** by its encoded sequence
- Most RNA sequencing is done with **short reads, bulk samples,** and **inefficient enzymes**
- Impossible to **detect RNA modifications transcriptome-wide,** with known stoichiometry
- **Machine learning** approaches and **methods to interpret and manage the data** are lacking

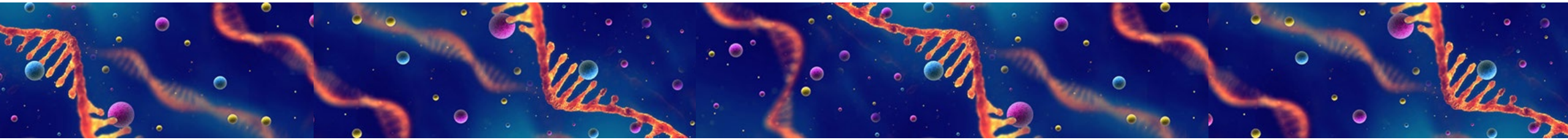
From Anna Pyle's keynote address

Workshop Objectives

- Identify **technologies** needed to **comprehensively characterize RNA**
- Determine **infrastructure**, **bioinformatics**, and other **resource needs**
- Consider the steps needed to **facilitate rapid adoption** of these technologies
- Identify how to best incorporate **public outreach** and **workforce development**

Process

- Participants discussed challenges and opportunities throughout the workshop.
- Key take-aways were synthesized by the planning team, reviewed by the Executive Committee, and appear in the Executive Summary on genome.gov.



www.genome.gov/event-calendar/97th-Meeting-of-National-Advisory-Council-for-Human-Genome-Research

NIH NHGRI

Priority Areas

- Diversity, Inclusion, & Training
- Technology Development
- Biological Standards & Centralized Resources
- Bioinformatics, Computation, & Data Science
- A Path to Clinical Implementation
- Environmental Exposure and Stress

Diversity, Inclusion, & Interdisciplinary Training

- **Additional funding** across career stages to **promote diversity.**
- **Outreach components** in large NIH-funded consortium and center projects.
- Support of **undergraduate and graduate-level training**.
- **Public Outreach**.
- **Interdisciplinary virtual and in-person training** in both wet and dry lab methods available and accessible to all professional levels.

Technology Development

- **New technologies** for full-length, direct RNA analysis.
- **New MS capabilities** as well as improvements of existing instrumentation.
- Focused efforts to harness the **full capability of enzymes** for RNA analysis.
- Creating a **library of long, synthetic RNAs** with specific modifications.
- Detecting, locating, and analyzing **interactions between DNA, RNA, proteins, and small molecules**.

Biological Standards & Centralized Resources

- "**Gold standard" reference set** of synthetic and cellular RNA splice isoforms with known chemical modifications.
- **Centralized resource** with banked tissues, cell lines, and exemplar viral strains.
- Consortium or coordination center-directed efforts to **stimulate collaborations** for both wet and dry lab advances in direct RNA analysis
- Centralized resource for **direct RNA sequencing**.

Bioinformatics Tools, Computational Resources, & Data Standards

- **RNA database** of different RNA types, includes normal and diseased tissue.
- **Standardized** nomenclature, file formats, pipelines, and software.
- **Machine learning/AI-based tools.**
- RNA **training sets**.
- RNA **secondary structure prediction** that considers chemical modifications.

Environmental Exposure and Stress

- Tools/reagents to study stress-, **environmental exposure-induced modification** in normal and disease states.
- Tools/reagents to map **temporal and spatial dynamics** of RNA modifications in different cellular and tissue contexts.
- Methods to interrogate impact of stress on RNA relative to **signaling, RNA-protein interactions, and trafficking.**

A Path to Clinical Implementation

- Small molecule libraries **targeting RNA structures**.
- RNA-modifying **enzymes** as potential **drug targets**.
- Improved **delivery technologies** to target specific tissues and cell types.
- New **therapeutic** development pipelines.
- High-grade **oligonucleotide** discovery and synthesis; **GLP manufacturing** for pre-clinical and clinical studies.

Conclusions

Identified opportunities in:

- Direct RNA sequencing with modification identification.
- The broader arena of RNA structure, RBPs, and biology.
- RNA as therapeutic and drug target.
- Developing the capabilities to obtain a truly comprehensive view of the transcriptome.

Discussion


The Forefront
of Genomics®