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

## May 24-26, 2022

## Executive Summary

NIEHS and NHGRI collaborated on a virtual workshop "Capturing RNA Sequence and Transcript Diversity, From Technology Innovation to Clinical Application." The goal was to determine current capabilities and needs for comprehensive characterization of the diversity of all RNAs and their modifications at a chemical and structural level in typical and disease states. The workshop addressed these 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 in this area

The workshop sessions included talks on the current state of the science and highlighted the challenges, needs, and opportunities relative to significance, innovation, and feasibility (see Appendix 1). Each scientific session was followed by breakout sessions focused on a set of compelling scientific questions. Key themes from the breakouts and the main sessions are synthesized below by topic area. A more detailed summary can be found in the full workshop report.

**Diversity and Inclusion**

Improved training and outreach opportunities will help build the diverse workforce needed to address emerging challenges and opportunities in the field of RNA biology.

- Additional funding is needed across career stages to promote diversity in the genomics workforce.
- Outreach components should be required of large NIH-funded consortia and center projects.
- Support of undergraduate and graduate-level training, particularly at minority-serving institutions, could help improve diversity within the genomics research community.
- Outreach to the public is needed to increase understanding of the importance and impact of RNA research on public health and medicine.

**Interdisciplinary Training**

Training in both experimental methods ("wet lab") and computational biology and bioinformatics ("dry lab") is critical for advancement and widespread adoption of RNA technologies.

- Training should incorporate virtual and hands-on instruction in both wet and dry lab methods.
- Easily accessible, quality, low (or no)-cost, online training videos on technology use and data analysis could be a considerable driver in technology adoption.
- Training resources should be available and accessible to all professional levels which requires a long-term commitment to funding and maintenance.

**Technology Development**

Technology development underlies many biomedical discoveries and applications. Advances in RNA science will require technological innovation, development, and standardization.

- New methods and technologies are needed for full-length and direct RNA analysis to comprehensively characterize the diversity of isoforms and modifications in sequence context.
- Investment in new mass spectrometry capabilities through modification of existing instrumentation and development of new devices focused on RNA is a critical opportunity.
- Focused enzymatic research is needed to develop useful molecular tools that overcome rate-limiting analytical barriers to molecular analysis in RNA biology.
- Focused technology innovation and development efforts that enable generation of long synthetic RNA with specific modifications would allow for multiple advances in the field.

- High-throughput and subcellular methods for detecting, locating, and analyzing interactions between RNAs, DNA, RNA-binding proteins, and small molecules are needed.

**Biological Standards and Centralized Resources**

Standards are a compelling, rate-limiting need in the field. Both data and experimental standards are important to facilitate and enable rapid progress.

- A community resource of "gold standard" reference sets of synthetic and cellular RNAs spanning the range of known modifications and isoforms is needed. This will serve as a ground truth for comparisons across a diversity of experimental approaches and for technology development, validation, and benchmarking.
- A centralized resource with banked tissues, cell lines, and an exemplar virus sample could provide common sources of RNA for testing of new RNA modifications and sequencing methods.
- A consortium or coordination center approach was proposed to stimulate collaborations for both wet and dry lab advances in direct RNA analysis. The effort would focus on technology innovation and development for the establishment of best practice guidelines and standards.
- A centralized mass spectrometry facility/resource for RNA sequencing and modifications (beyond single nucleotide detection) could be used to assemble and calibrate work on a pilot project using direct RNA sequencing methods.

**Bioinformatics Tools, Computational Resources, and Data Standards**

Bioinformatics tools, computational resources, and data standards are key to this developing field that encompasses a rich diversity of comprehensive RNA information.

- A comprehensive, centralized, interoperable, standardized, and searchable RNA database that includes different RNA types and both typical and diseased tissue will be critical.
- Nomenclature, file formats, pipelines, and software need to be standardized, updated, and developed to accommodate current knowledge and anticipated advances in RNA biology.
- More efficient data processing and development of machine learning/artificial intelligence-based tools and streamlined workflows will enable translation of RNA analysis efforts.
- RNA secondary structure prediction is an important opportunity. Very few methods of RNA structure prediction take chemical modifications into consideration.

**Environmental Exposure and Stress**

A broad diversity of RNA modifications and alternative RNA species can be induced by reactive oxygen species (ROS), oxidative stress, and other environmental exposures and toxicants.

- Identification of specific enzymes that can insert stress-induced modifications is required to understand the biological impact of distinct environmental exposures.
- Tools to study specific RNA modifications in a variety of different cell types are needed to connect these changes to environmentally induced or associated pathologies.
- There is a need to map temporal and spatial dynamics of RNA modifications inside a cell and tissue in relation to functional responses to stress or physiological conditions.

**A Path to Clinical Implementation**

Translation of fundamental knowledge to the clinic can be facilitated by focused efforts.

- Creation and curation of small molecule libraries for RNA structure screening that encompass a variety of binding efficiencies for specific RNAs will facilitate translational research.
- Exploration of RNA-modifying enzymes as potential drug targets could be beneficial.
- Capabilities in delivery technologies, for example lipid nanoparticles, are critical to this effort, particularly in targeting specific tissues and cell types.
- There is an opportunity to establish therapy development pipelines that encompass the range of expertise needed for all steps of the process including initial discovery, synthesis of high-grade oligonucleotides, and GLP manufacturing for pre-clinical and clinical studies.

## Appendix 1: Session Talks

**Setting the Stage**
Anna Pyle described the status of RNA knowledge and highlighted opportunities to advance the field.

**Impact of Nucleotide Sequence on RNA Structure and Biological Roles**
Yunsun Nam focused on the impact of RNA sequence and modifications on structure, folding, and function. The talk emphasized that structure more than sequence affects function.

**RNomics Applications from Research to Clinic**
Jeannie Lee discussed clinical and therapeutic applications. She addressed both RNA as the drug itself and RNA as the target of the drug. One focus was on how creation of curated, small molecule libraries targeting RNA structures would broadly facilitate research along with drug and diagnostic development.

**Technologies for Direct Sequencing**
Meni Wanunu covered the prospects and needs for technologies for direct RNA analysis by covering current and promising technologies including sequencing and mass spectrometry analysis.

**Infrastructure, Bioinformatics, and Critical Resources: What is Needed?**
Christopher Burge presented an overview of the current state of bioinformatic analysis of complex RNA data. The talk highlighted the need to recognize the origins of sequencing technology biases, develop robust statistical procedures to minimize their impact, and use insights to inform future efforts.

**Facilitating Technology Dissemination and Adoption**
Brenton Graveley discussed barriers to dissemination and adoption of the latest genomics technologies. Training, particularly in bioinformatics and computational biology, is key to improving adoption of new technologies. Widely accepted RNA data and experimental standards are also needed.

**Wrap-up**
The workshop concluded with a summary presented by Blanton Tolbert on behalf of the Executive Committee (also including Vivian Cheung, Peter Dedon, and Brenton Graveley). This summary highlighted key priority areas including: 1) methods to sequence RNA while capturing all chemical modifications; 2) RNA structure identification and prediction; and 3) development of RNA-based diagnostic tools and therapeutics to study regulation and impact on health and disease.