Twenty years of *RNA*

DONALD RIO

Division of Biochemistry, Biophysics and Structural Biology, California Institute for Quantitative Biosciences (QB3), Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-3204, USA

In the 20 years since the *RNA* journal was started there have been many exciting and unexpected findings made in the broad field of RNA biology, biochemistry, genetics and structural biology. Below, I will briefly describe a few of the many seminal discoveries.

Structural biology: catalytic RNAs – group I & group II introns, riboswitches, RNA biosensors, and ribosomes

Back in the mid 1990's, we had just been able to visualize the organization, guanosine cofactor-binding site, and catalytic core of the Tetrahymena group I intron ribozyme RNA. Given that no large RNA structures had been obtained since the 1970's when the tRNA structure was solved, this was a dramatic and far-reaching advance. This was followed by the discovery of small ligand-binding RNA riboswitches in bacteria and fungi, the X-ray structure determination of the pre-mRNA/spliceosome-related group II intron RNA from the extremophile *Oceanobacillus iheyensis* and many riboswitches bound to their cognate metabolite ligands. More recently, *in vitro* RNA selection methods were used to generate ligand binding-controlled RNAs that act as biosensors that fluoresce, much like green fluorescent protein (GFP) and related fluorochrome proteins, upon ligand binding and these have been developed into small molecule biosensors. Thus, the realization, through the analysis of new atomic-level RNA structures, that RNA can form well-organized ligand-binding sites for small molecules has been used to design and develop novel RNA molecules with defined functions.

Another major advance (which was recognized by the 2009 Nobel Prize) was the determination of the X-ray structure of the ribosome, the large RNA-protein machine that cells use to de-code messenger RNA and synthesize proteins. Long thought to be an RNA-based catalytic machine, based on biochemical studies from the Noller lab, the visualization of the locations of the large ribosomal RNAs in both subunits and the distant locations of any proteins from the peptidyl transferase center left little doubt that peptide bond formation is

RNA-catalyzed. More recently, exciting developments in high-resolution electron detectors have pushed the limits of cryo-electron microscopy of ribosomes almost to the resolution of X-ray crystallography. This technical breakthrough promises to impact many areas of RNA and RNA-protein structural biology. Electron microscopy had been extensively used to analyze ribosomes and spliceosomal small nuclear ribonucleoprotein (snRNP) complexes. Finally, breakthroughs in high-throughput cDNA sequencing methods have led to the development of quantitative genomic methods for transcriptome-wide “ribosome profiling,” which is already leading to important insights about the activity of previously unidentified upstream open reading frames (uORFs) and other modes of translational regulation.

Genome sequencing and alternative splicing

Advances in genome sequencing and annotation have led to an explosion of information about alternative pre-mRNA splicing. In the early 2000's microarrays were used to detect and quantitate splicing patterns, but from the late 2000's into the 21st century, the development and use of high-throughput next-generation cDNA sequencing coupled with corresponding improvements in genome/transcriptome mapping and analysis software has led to the realization that levels of alternative splicing increase with organismal complexity and that essentially all known human mRNA-encoding genes generate at least two and on average three alternatively spliced mature mRNAs from genes with an average of 7–8 introns. Of course, the transcript isoform profiles vary considerably from tissue to tissue with the nervous system exhibiting the most complex alternative splicing patterns. A challenge for the future will be to understand how the diverse splicing isoform patterns in the brain generate functional changes in connectivity and activity in different brain regions.

We know that RNA binding proteins play a role in determining alternative RNA splicing patterns through their interaction with splicing enhancer and silencer elements in the pre-mRNA. While methods have been developed to locate

Corresponding author: don_rio@berkeley.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.049817.115>. Freely available online through the *RNA* Open Access option.

© 2015 Rio This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

RNA binding sites for different proteins transcriptome-wide, the functional role of these mapped binding sites is not always clear. In fact, one puzzling finding related to these “RNA maps” is that the same factor can act as a repressor of splicing in one position and as an activator of splicing in another position. How this process of position-specific protein binding controlling splicing outcomes is still a mystery. Another important insight is that alternative pre-mRNA splicing patterns are altered in diseases, such as cancer, where mutations in splicing sites, exonic or intronic splicing silencers or enhancers and in genes for trans-acting 3' splice site recognition proteins can lead to aberrant splicing patterns. A challenge for the future is how to identify the causal or functional changes in splicing isoforms that lead to the disease states, since our ability to sensitively detect splicing pattern changes now exceeds our ability to test alternative isoform changes for direct disease connections. New, high-throughput methods of transcriptome-wide RNA structure determination (SHAPE-seq and DMS-seq) will be informative about the effects of disease mutations on RNA structure and how this could affect alternative splicing patterns.

Coupling of transcription and RNA processing

The study of RNA processing reactions, initially using biochemical assays and reconstitution *in vitro*, did not typically involve coupling of these reactions to RNA synthesis by eukaryotic RNA polymerases. However, studies over the past 10–15 years had led to the view that at least for pre-mRNA synthesis by RNA polymerase II, the RNA processing reactions are often coupled to transcription. Some indications of this are that changes in RNA polymerase elongation rates can alter pre-mRNA splicing patterns and that *in vitro* assays for either splicing or polyadenylation are stimulated by the addition of phosphorylated recombinant protein from the C-terminal heptad repeat domain of largest RNA polymerase II. More recent studies using high-throughput sequencing of transcriptionally engaged RNA polymerase II (Nacent-seq) have indicated that many introns are or can be spliced co-transcriptionally *in vivo*.

How these processes of gene expression are coordinated remains somewhat unexplained, but there are some recent studies on RNA binding proteins and RNA polymerase II that might go a long way to provide a basis for coupling of transcription and RNA processing through the action of RNA binding proteins. Several observations are important. It has long been known that there is a link between RNA binding proteins, such as FUS-TLS, TDP-43 and hnRNPA1 and neurodegenerative diseases, like ALS (amyotrophic lateral sclerosis) and FTLD (frontotemporal dementia). These RNA binding proteins, as well as the C-terminal heptad repeat domain of the largest subunit of RNA polymerase II (CTD) contain low complexity (LC) amino sequence domains of glycine (G), serine (S), aspartic acid (D), and tyrosine (Y). These low complexity domains resemble the “prion-

like domains or PrLD” found in well-documented disease-causing prion proteins. Recent studies have shown that the low complexity domain of FUS-TLS can interact with the CTD of pol II in an RNA-dependent manner. The RNA appears to allow a high local concentration of the FUS-TLS protein and promote formation of long protein fibers that can then strongly interact with the pol II CTD. Biochemical and functional characterization of several LC domains from FUS-TLS, hnRNPA1, hnRNPA2, and TDP-43 have shown that a common structural motif involving extended β -sheets can allow these LC domains to form fibers and be incorporated into semi-solid “hydrogels.” Similar findings about these fibers had emerged from studies of polyglutamine LC domains found in the Huntington’s disease protein and some transcription factor activation domains. Recent human genetic studies have linked single amino acid changes in the LC domains of hnRNPA1 and hnRNPA2/B1 to ALS disease-causing mutations. These LC domains are also found in the SR repeat domains of the SR protein splicing factor family (e.g., SRSF1) and many other hnRNP family members, known to be involved in alternative splicing decisions. Thus, these low complexity domains may be involved in interactions with RNA polymerase II to coordinate alternative splicing patterns and these interactions are likely perturbed by disease-causing variants causing significant disruption of the normal transcription/RNA processing programs.

Non-coding RNAs (small and large) and RNA-guided genome editing

Over the past 15 years there has been an explosion in our knowledge about the role of large and small non-coding RNAs. The discoveries of RNA silencing pathways in plants, animals, fungi, and bacteria have truly been astounding. Their normal biological roles in, for instance, silencing transposons in the germlines of mice, flies and worms, heterochromatin formation in fission yeast and in the primitive innate immune systems of bacteria are truly biological marvels. Moreover, the discovery of RNA interference pathways has led to novel genetic screening and assay tools. Biochemical dissection of the pathways of RNA interference, microRNAs, and the CRISPR systems has been exciting. More recently, biochemical characterization of the CRISPR/Cas9 system has led to rapid development of and improvements for RNA-guided genome editing as well as gene up- and down-regulation (CRISPRa/CRISPRi). This is just the beginning of a truly exciting era in genome editing and functional genomics using cas9.

Beginning with tiling microarrays and more recently with high-throughput sequencing, it became evident that much of the genome is transcribed beyond protein-coding genes. A variety of approaches identified a class of long, polyadenylated RNAs that lacked open reading frames, termed long non-coding RNAs or lncRNAs. Several lncRNAs have been linked to epigenetic effects on gene expression through the polycomb heterochromatin/histone methylation complex.

However, many of these lncRNAs have not been studied and we do not know their functions.

While RNA-guided genome editing certainly plays a role in the bacterial CRISPR immune systems, RNA guided genome editing has also been characterized in both *Tetrahymena* and *Oxytricha*. In these cases small RNA molecules guide or specify sequences of the genome for elimination or for re-assembly of an intact genome after “scrambling” catalyzed by a DNA

transposase. Here, RNA molecules are templating genome re-assembly, a truly bizarre process. However, recall that studies on the ciliated protozoan *Tetrahymena* led to the discovery of RNA catalysis by group I ribozymes and the RNP enzyme telomere reverse transcriptase, better known as telomerase.

It has been a remarkable 20 years of RNA and I am anxious to find out what discoveries over the next 20 years of research in RNA are in store.

RNA

A PUBLICATION OF THE RNA SOCIETY

Twenty years of *RNA*

Donald Rio

RNA 2015 21: 718-720

Open Access

Freely available online through the *RNA* Open Access option.

Creative Commons License

This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>