EDITORIAL

Special issue on RNA processing and regulation

Xinshu Xiao[1,*] and Chaolin Zhang[2,*]

[1] Department of Integrative Biology and Physiology, Bioinformatics Interdepartmental Program, Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA 90095, USA

[2] Department of Systems Biology, Department of Biochemistry and Molecular Biophysics, Center for Motor Neuron Biology and Disease, Columbia University, New York, NY 10032, USA

* Correspondence: gxxiao@ucla.edu, cz2294@columbia.edu

This year (2018) marks the 60th anniversary of the "central dogma", summarized as "DNA makes RNA makes protein", which was originally proposed by Francis Crick in 1958. Three years later, messenger RNA was identified as the template of protein synthesis. After 60 years of discovery, including discovery of the split nature of eukaryotic genes (*i.e.*, splicing), it becomes evident that messenger RNAs are not merely messengers, but a hub of co- and post-transcriptional regulation, which is fundamental to amplify the complexity encoded in the genome of higher eukaryotic organisms. The mature forms of RNA of protein-coding genes and their abundance have to be tightly regulated through multiple steps of sophisticated processing, including capping, splicing and polyadenylation. In addition, their function also critically depends on proper localization — sometimes trafficking to the remote parts of the cell such as dendrites and axons of neurons — and proper control of their stability. Furthermore, thousands of long and small noncoding RNAs are produced to play a wide range of roles in gene regulation. From our perspective, two overarching goals for RNA biology include (i) characterizing the spatial-temporal regulation of various RNA species and elucidating the underlying regulatory mechanisms; (ii) understanding the functional impact of such regulation on human physiology and disease.

We believe we are at an exciting time to fulfilling these goals, thanks to revolutionary technological advances, especially high-throughput sequencing that started about ten years ago. It is now routine, and even trivial, for individual labs to profile the whole transcriptome in specific conditions at single-nucleotide resolution, which allows us to catalogue and quantify all RNA species in specific cellular context. Many different types of genomic technologies were also developed to probe specific aspects of the transcriptome, *e.g.*, mapping of specific modifications of single nucleotides, polyadenylation sites, and pre-mRNA branch sites that mediate splicing. The regulation of RNA processing and homeostasis is largely mediated through RNA-RNA and RNA-protein interactions. Therefore, technologies have also been developed to probe the regulatory processes, such as mapping of *in vivo* RNA-binding protein (RBP) interaction sites and RNA-RNA secondary structures on a genome-wide scale. These exciting progresses have motivated us to put together this special issue focusing on RNA processing and regulation. As we indicated in our invitation letter for contributors, we tried to "cover a wide spectrum of topics in this area, ranging from basic mechanisms, to functional studies and evolution, and to applications that help to decipher disease relevance" from both experimental and computational sides.

This special issue features seven original research and review articles. Given the routine application of RNA sequencing (RNA-seq), Li and Li reviewed statistical models used in analysis of RNA-seq data. While most analyses have focused on gene quantification and differential gene expression, this review provided a

comprehensive overview of analyses at the sample, gene, transcript and exon levels, with the aim to provide a guide for users to select proper computational tools depending on specific biological questions.

We have two original studies focusing on specific steps of RNA processing and regulation. Guvenek and Tian's original study focused on quantification of alternative polyadenylation (APA) using RNA-seq data, as APA provides an important mechanism to alter the length of 3′ UTR where a plethora of regulatory sequence elements reside (*e.g.*, binding sites of RBPs and microRNAs). They observed similar preference for longer 3′ UTRs during neural differentiation and myogenesis, but the extent is larger in neurons. They also found neurons frequently have longer 3′ UTRs compared to non-neuronal cells in the brain using bulk and single-cell RNA-seq data. Taggart and Fairbrother developed a new computational method named ShapeShifter to detect and quantify intron lariat resulted from RNA splicing using total RNA-seq data. Previous methods to detect intron lariats rely on chimeric reads that are generated when reverse transcriptase traverses across the branch site, which are rare events with limited sensitivity and in general insufficient for quantification. Interestingly, they found stable intron lariats display a characteristic read coverage profile they can leverage to detect new lariats without relying on on chimeric reads.

Martinez *et al*. reviewed detection of various kinds of modifications at single nucleotides in pre-mRNA. Over 100 different types of modification were documented in the literature, but we are just at the beginning to understand their function. This review summarized experimental techniques to detect specific types of modifications and discussed the conceptual framework how they can affect RNA processing such as splicing, *e.g.*, by modifying protein-RNA interactions and RNA structures. Zhang *et al*. described an original study to detect RNA methylation sites from immunoprecipitation and sequencing data using a carefully designed Bayesian hierarchical model.

Finally, Hannigan *et al*. provided a comprehensive review of crosslinking and immunoprecipitation (CLIP) that is now the *de facto* method to map *in vivo* protein-RNA interactions on a genome-wide scale. This review provided an overview of the evolution and application of this powerful technology. The authors also emphasized the importance of integrating protein-RNA interactions with functional data, *e.g.*, alteration in transcriptomic output upon perturbation of RBPs. In parallel, Gong *et al*. reviewed genomic technologies to profile RNA-RNA interactions. RNA-RNA interactions are known to be fundamental for RNA processing such as splicing and translation at the very beginning, but their regulatory role in specific cellular context in general is poorly understood. These new technologies may provide new avenues for this important direction.

In sum, we are at a golden age to explore the RNA world.