https://doi.org/10.1038/s41592-022-01654-5

Decoding noncoding RNAs

Research interest in noncoding RNAs and their biological implications in a variety of cellular contexts has been growing. In this issue, we present a series of pieces discussing recent method advances and future directions for deciphering the regulatory roles of noncoding RNAs.

About 75% of the human genome can be actively transcribed, but less than 2% of the human genome encodes proteins. The unknown world of noncoding RNAs (ncRNAs) is vast. In-depth examinations of the noncoding regions have uncovered many regulatory ncRNAs essential to RNA processing and gene regulation. Yet we are still far from fully understanding the regulatory roles of ncRNA given the remarkable variety of biological processes in which they participate. We therefore believe that efforts in methodology development will help researchers not only to identify these regulatory ncRNAs from the large noncoding space but also to study their cellular functions in both health and disease.

ncRNAs include long noncoding RNA (lncRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and small RNAs such as microRNA (miRNA), piwi-interacting RNA (piRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA), which play distinct roles in diverse cellular processes. For example, rRNAs and tRNAs are well known to be involved in protein translation; lncRNAs and miRNAs can regulate gene expression. The diversity of both short and long forms of ncRNA, the chemical modifications embedded in transcripts, and their dynamic folding structures mean that characterizing these species is technically challenging. The advent of high-throughput sequencing tools have allowed investigation of previously unknown ncRNAs and their relative abundances across different cell lines. In 2012, the ENCODE project consortium characterized the landscape of transcription in human cells, which offers a transcript reference for previously unannotated and annotated RNA species.

lncRNAs are defined as transcripts comprised of more than 200 nucleotides. They have attracted increasing attention for their regulatory roles in gene expression and local chromatin structure. There are thousands of lncRNAs transcribed from the genome, but the functions of most remain unclear. In 2017, the FANTOM consortium built a catalog of human lncRNAs and their expression profiles across human primary cell types, providing a valuable resource for functional studies of nearly 20,000 lncRNAs. In this issue, a Comment from Ling-ling Chen highlights recent progress in deciphering lncRNA expression patterns and subcellular localizations enabled by RNA sequencing and fluorescence imaging tools, as well the use of mass spectrometry and sequencing to study lncRNA-protein complexes. Chen also explores opportunities for dissecting lncRNAs and their phenotypes in a spatial-temporal manner within different cellular contexts.

Despite recent efforts in charting the functional roles of ncRNA, most ncRNAs are functionally uncharacterized. A Comment from Jimmy K. Guo and Mitchell Guttman discusses how the identification of protein–RNA interactions can provide insight into ncRNA's functional roles. They cover developments in RNA-centric proteomic methods and protein-centric sequencing methods for exploring proteins and RNAs, respectively. They also propose a framework for determining ncRNA and protein functions.

Although understanding protein–ncRNA interactions is important, this only provides part of the picture. Another crucial aspect of understanding ncRNA function is knowing their structures: RNA molecules are dynamically folded and may form distinct structures under different conditions. Moreover, RNA structures themselves, such as those of catalytic RNAs (ribozymes), can possess critical functions. A Review from Qiangfeng Cliff Zhang and colleagues summarizes recent experimental and computational tools developed for RNA secondary and tertiary structure determination. The rapid growth of experimental data has facilitated deep learning and integrative modeling strategies to better refine RNA structures. We hope this review will inspire experimentalists and computational experts to learn from each other and fulfill the methodological needs to reveal structure-dependent RNA functions.

One relatively new class of ncRNA called circular RNA (circRNA) has attracted considerable attention. CircRNAs are formed by back-splicing, in which a 5′ splice site attaches to an upstream 3′ splice site to circularize one or more exons. Most circRNAs were thought to be the result of splicing error, yet recent progress in sequencing approaches and circRNA-specific software suggested circRNA can have tissue-specific and development-specific expression patterns. For example, a number of circRNAs are dynamically regulated in the mammalian brain and are enriched at synapses. CircRNA is experimentally challenging to study due to the difficulty of studying the circular form without affecting linear counterparts. A Perspective supported by the Marie Curie Innovative Training Network circRTrain proposes a series of experimental guidelines for purifying, profiling, expressing and deleting circRNA, and factors that should be considered when studying the regulatory mechanism of circRNAs. The authors also emphasize the need for careful optimization when generic RNA tools are adapted for circRNA studies. We hope that this Perspective can serve as a starting point for new researchers entering the circRNA field and inspire development of circRNA-specific tools.

Chemical modifications on rRNA and tRNA were first identified decades ago. In recent years, a growing number of modifications, such as N^6-methyladenosine (m6A), 5-methylcytidine (m5C), pseudouridine and 2′-*O*-methylated nucleotides (Nm), have been identified in other species of RNA molecules, including mRNA and lncRNA. These modifications endow an additional layer beyond sequence to further regulate gene expression through altering charge, RNA folding and RNA–protein interactions. In a Comment, Mark Akeson and colleagues highlight the advantages of nanopore native RNA sequencing to directly profile modifications on RNA strands. Despite the promise of direct RNA profiling, technical challenges associated with relatively low accuracy of base calling and the requirement for large input materials still exist. Furthermore, the authors suggest that independent validation is needed for new tools developed for quantitative identification of modifications. They also provide practical advice for new users who wish to apply direct

sequencing for mapping RNA and modification landscape.

Finally, in two Technology Features our journalist Vivien Marx presents several researchers' personal histories and views on how ncRNA field evolved over time and highlights the ncRNA community's excitement about the field's future. We also put together a collection of key Nature Portfolio articles published on ncRNA research.

Despite the increasing prominence of ncRNA, we remind readers that the presence of a ncRNA molecule does not always imply functionality. It is also possible that these transcripts are non-functional or products from, for example, splicing errors. We hope this Focus issue will provide researchers with practical advice for deciphering ncRNA's roles in biological processes.

Published online: 6 October 2022