

RNA: An Expanding View of Function and Evolution

■ Xinwei Han

Postdoctoral research scientist, Department of Systems Biology, Columbia University, New York, NY, USA.

■ Yuan Chen

VP of Research and Development, Aperiomics, Sterling, VA, USA.

■ Liuyang Wang

Senior Research Associate, Duke University School of Medicine, Durham, NC, USA.

■ Wenwen Fang

Damon-Runyon Postdoctoral Fellow, Whitehead Institute for Biomedical Research, Cambridge, MA, USA.

■ Ning Zhang

ORISE Fellow of Evolutionary Biology, US Food and Drug Administration, Silver Spring, MD, USA.

■ Qiyun Zhu

Postdoctoral Fellow of Genomic Medicine, J. Craig Venter Institute, Rockville, MD, USA.

Supplement Aims and Scope

Next generation sequencing (NGS) technologies have enabled unprecedentedly deep characterization of transcriptomes. Compared to the microarray technology, NGS has been a much more favorable method for transcriptome profiling, as it doesn't require any pre-existing knowledge of the transcriptome of any given species. By sequencing transcriptomes to enough depth, several studies have reported a remarkably large number of novel RNA species and/or previously undetected splice forms. An inclusive identification of genes for organisms without genome information is also within reach. With intricate designs, it is even possible to specifically sequence the double-stranded portion of RNAs, such that the secondary structure of these RNAs could be inferred. Such comprehensive elucidation of transcriptomes has opened many new venues for studies on the function and evolution of the diverse RNA repertoire. The current supplement aims to serve as a portal to report original studies and summarize progress in this fast-moving area.

■ Novel RNA species discovered by NGS and their unique evolutionary history

Genome-wide studies utilizing NGS technologies have uncovered novel types of RNAs, such as lincRNA, circular RNA and spliRNA. Many of these novel RNA species have been reported to play important regulatory roles. With the continuous advancement of sequencing technologies and experimental designs, it is expected that even more RNA species will be discovered. Surprisingly, many lincRNAs, including several lincRNAs with known functions, have low interspecific sequence conservation. Several studies

have suggested that structural conservation in these lincRNAs may have been retained, despite the apparent lack of sequence conservation. These recently identified non-coding RNAs represent an evolutionary history different from that of the protein coding genes, which remains to be explored.

- The function and evolution of alternative splicing
Alternative splicing tremendously diversifies transcriptomes among organisms, even when their repertoires of protein coding genes are similar. The fast improvement on read length of NGS technologies will render a more thorough and unambiguous identification of alternative splice forms. Such resources could serve as the basis for exploring conservation and divergence of splicing events among organisms. They could also facilitate studies on the function of splicing events specific to certain organisms.
- RNA-seq of non-model organisms for phylogenomic studies
RNA-seq approaches enabled de novo identification of genes from organisms of no assembled genome sequences. Analyzing such extensive list of genes will result in better resolution of organism phylogeny. In addition, by comparing gene sequences among a wide range of organisms, many intriguing evolutionary questions may be addressed. For instance, how frequently genome duplication has happened in a certain taxon? How frequently horizontal gene transfer has happened between symbiotic organisms or parasites and hosts?



- RNA sequencing as an approach to study RNA secondary structure
Specific enrichment and sequencing of double-stranded portion of RNAs have been proven to be a useful measure to study RNA secondary structures.

Comparison of RNA secondary structures among organisms will lead to identification of conserved structure motifs, and also may be used to infer their functions.

As Lord Kelvin put it, “nearly all the grandest discoveries of science have been but the rewards of accurate measurement and patient long-continued labor in the minute sifting of numerical results.” The impacts of RNA-seq technology on current functional and evolutionary studies clearly reflect the importance of accuracy and scalability. Compared to traditional methods of gene expression quantification, such as microarray, RNA-seq does not require prior knowledge of the reference genome to design probes and shows much greater dynamic range, achieving more precise characterization of genes with extremely high/low expression. On the other hand, as with other high-throughput approaches, state-of-the-art data analysis is always a necessity to turn measurements into biological insights. Motivated by the striking advantages of RNA-seq and the challenges in data analysis, the goal of this supplement is to showcase how RNA-seq experiments could be designed to accurately measure transcriptomic dynamics and how elegant analysis of large-scale data could address specific biological questions.

Since its inception, RNA-seq has been widely adopted in various areas of studies. The unprecedented deep investigation of transcriptome enabled by RNA-seq has led to discoveries of novel RNA species and new ways of regulation.^{1,2} In this direction, Yin et al reported the detection of 30 conserved miRNA families and 113 novel miRNAs in *Nicotiana benthamiana*, most of which are likely to be involved in the immune response to Tobacco Mosaic Virus.³ The study by Taguchi found microRNA-target-specific histone modifications occur during mammalian spermatogenesis, inferring a new level of regulation.⁴ With the development of *de novo* assembly methods, RNA-seq has also catalyzed the blossom of studies on non-model organisms that currently do not have reference genome sequences.⁵ In this supplement, Zhou et al identified 59 differentially expressed genes under drought treatment in a non-model tropic plant, *Bombax ceiba* L., of which genes are involved in multiple functional pathways, indicating complicated genetic responses to drought stress.⁶ Wang et al conducted *de novo* transcriptome assembly for an endangered herb, *Veratrum baillonii*, and discovered a large number of microsatellite markers and validated with experiments, providing valuable resources for the future population genetics studies and effective conservation management.⁷ Furthermore, RNA-seq provides previously inaccessible information for phylogenomic and evolutionary studies.^{8,9} Ma et al discovered 107 *WRKY* genes in desert poplar using a comparative genome approach, of which 10 genes were specific to *Populus euphratica*. By profiling

gene expression in both desert poplar and mesophytic poplar, they found this desert plant adapted to salt stress through *de novo* formation of new genes in *WRKY* gene family and acquisition of new functions for existing *WRKY* orthologs.¹⁰ Gao et al analyzed protein kinase superfamily and gene expression from 69 different conditions in Pacific oyster, and found strong positive correlation of sequence divergence and expression divergence within this gene family.¹¹

The sequencing capability behind RNA-seq is still progressing at an astonishing pace, with the cost of sequencing the same sample dropping exponentially. The scope of transcriptome characterization is continually expanding, especially fueled by large-scale tissue-specific and developmental-stage-specific studies.¹² We expect more such studies will emerge in the next few years to further push the limit. The accumulation of expression data, associated with fast increasing genotype information, will lead to a comprehensive identification of functional regulatory variants and deeper understanding of mechanisms of gene regulation. In this exciting era of high-throughput sequencing, we hope this supplement shares with you a sip of enthusiasm.

REFERENCES

1. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol.* 2014;32:453–61.
2. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell.* 2013;154:26–46.
3. Yin K, Tang Y, Zhao J. Genome-Wide Characterization of miRNAs Involved in N Gene-Mediated Immunity in Response to Tobacco Mosaic Virus in *Nicotiana benthamiana*. *Evolutionary Bioinformatics.* 2015;11(S1):1–11.
4. Taguchi Y-H. Apparent microRNA-Target-specific Histone Modification in Mammalian Spermatogenesis. *Evolutionary Bioinformatics.* 2015;11(S1):13–26.
5. Hornett EA, Wheat CW. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC Genomics.* 2012;13:361.
6. Zhou ZL, Ma HC, Lin KH et al. RNA-seq reveals complicated transcriptomic responses to drought stress in a non-model tropic plant, *Bombax ceiba* L. *Evolutionary Bioinformatics.* 2015; 11(S1):27–37.
7. Wang L, Wang ZK, Chen JB, et al. De novo transcriptome assembly and microsatellite markers development for the traditional Tibetan medicinal herb *Veratrum baillonii* Franch (Gentianaceae). *Evolutionary Bioinformatics.* 2015;11(S1):39–45.
8. Brawand D, Soumillon M, Necsulea A, et al. The evolution of gene expression levels in mammalian organs. *Nature.* 2011;478:343–8.
9. Gu X, Zou Y, Huang W, Shen L, Arendsee Z, Su Z. Phylogenomic distance method for analyzing transcriptome evolution based on RNA-seq data. *Genome Biol Evol.* 2013;5:1746–53.
10. Ma JC, Lu J, Xu JM, Duan BB, He XD, Liu JQ. Genome-wide Identification of *WRKY* Genes in the Desert Poplar *Populus euphratica* and Adaptive Evolution of the Genes in Response to Salt Stress. *Evolutionary Bioinformatics.* 2015;11(S1):47–55.
11. Gao DH, Ko DC, Tian XM, Yang G, Wang LY. Expression divergence of duplicate genes in the protein kinase superfamily in Pacific oyster. *Evolutionary Bioinformatics.* 2015;11(S1):57–65.
12. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;348(6235):648–60.



Lead Guest Editor **Dr Xinwei Han**

Dr Xinwei Han is a postdoctoral research scientist of computational genomics at Columbia University. He completed his PhD at The Pennsylvania State University and has previously worked at Duke University. He now works primarily in disease gene discovery for developmental diseases. Dr Han is the author or co-author of 4 published papers and has presented at 3 conferences.

xh2272@columbia.edu <https://scholar.google.com/citations?user=r2TEXRYAAAAJ&hl=en>

Guest Editors

YUAN CHEN

Dr. Yuan Chen is VP of Research and Development at Aperiomics. He completed his PhD at Chinese Academy of Sciences and has previously worked at Duke University Medical Center as Post-Doctoral Associate. He now works primarily in pathogen evolution and diagnostics. Dr. Chen is the author or co-author of papers published in Nature Communications, PLoS Genetics, MBio and Molecular Ecology.

ychenbioinfo@gmail.com <https://www.linkedin.com/pub/yuan-alvin-chen/19/673/138>

LIUYANG WANG

Dr. Liuyang Wang is a Senior Research Associate at Duke University School of Medicine. He completed his PhD at University of Chinese Academy of Sciences, and has worked at University of Michigan before he joined Duke University. He works primarily in molecular and genome evolution, human-pathogen co-evolution and genotype-phenotype interaction. His extensive studies involve the study of speciation, biogeography, meta-GWAS analysis and infectious disease.

liuyang.wang@duke.edu <http://orcid.org/0000-0001-9556-2361>

WENWEN FANG

Dr. Wenwen Fang is a Damon-Runyon Post-Doctoral Fellow at Whitehead Institute for Biomedical Research. She completed her PhD at Princeton University. She now works primarily in RNA biology. Dr Fang has co-authored six published papers and presented at eight conferences.

wfang@wi.mit.edu

NING ZHANG

Dr. Ning Zhang is an ORISE fellow of evolutionary biology at FDA. He completed his PhD at the Chinese Academy of Sciences and has previously worked at the National Museum of Natural History, Smithsonian Institution. He now works primarily in DNA barcoding of plant species used for foods and herbs. Dr Zhang is the author or co-author of 15 published papers and has presented at five conferences.

Ning.Zhang@fda.hhs.gov <https://scholar.google.com/citations?user=OPFsFScAAAAJ&hl=en>

QIYUN ZHU

Dr. Qiyun Zhu is a post-doctoral fellow of genomic medicine at the J. Craig Venter Institute. He completed his PhD at University at Buffalo, the State University of New York. He now works primarily in metagenomics of human gut microbiome. Dr Zhu is the author or co-author of 10 published papers and has presented at 5 conferences.

qzhu@jcvj.org <http://www.jcvj.org/cms/research/groups/genomic-medicine/>

SUPPLEMENT TITLE: RNA: An Expanding View of Function and Evolution

CITATION: Han et al. RNA: An Expanding View of Function and Evolution. *Evolutionary Bioinformatics* 2015:11(S1) 77–79 doi: 10.4137/EBO.S38105

TYPE: Editorial

FUNDING: Authors disclose no external funding sources.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC3.0 License.

CORRESPONDENCE: xh2272@columbia.edu

All editorial decisions were made by the independent academic editor. All authors have provided signed confirmation of their compliance with ethical and legal obligations including (but not limited to) use of any copyrighted material, compliance with ICMJE authorship and competing interests disclosure guidelines.