Editorial:

Transcription: the epicenter of gene expression

Jiannan GUO
Guest Editor
(*Department of Biochemistry, University of Iowa, Iowa City, IA 52242, USA*)
E-mail: jiannan-guo@uiowa.edu

The complexity of a living organism is not driven by gene number but gene regulation. Controlling which genes to express and to what extent dictates the subsequent cell identity. Transcription, the critical initial stage in gene expression, is regulated delicately to maintain the cell status. Recent developments in the genomic approaches provided unparalleled coverage of the study of transcription. Still, basic molecular biology and biochemistry are providing mechanistic insights into how the regulation is achieved. In this feature "Regulation of transcription: mechanisms and biological functions", the latest advances in epigenetics, mRNA processing, RNA quality control, and human immunodeficiency virus (HIV) transactivation are discussed.

Control transcription, control gene expression

A gene has been traditionally viewed as the basic molecular unit of heredity (Crick, 1958; 1970). In the form of DNA or RNA, it carries the raw genetic information that can be turned into functional products, usually proteins. However, the number of genes does not reflect the complexity of the organism. For example, a human has about 20000 protein-coding genes, which is ~6000 more than a fruit fly, ~2000 more than *Caenorhabditis elegans*, and ~14000 more than budding yeast, but ~10000 less than a lab mouse, ~5000 less than the model plant *Arabidopsis*, and ~17000 less than rice. Clearly, the level of complexity of the organism is achieved by regulating available genes, not simply by introducing more genes.

The central dogma of gene expression includes two sequential steps: transcription (DNA to RNA) and translation (RNA to protein). Transcription is the key step that controls the "on and off" of genes and subsequently underlines the identity and the status of the cell (Young, 2011; Lee and Young, 2013). For example, when I compare a string of my hair and my finger tip, they appear so different, as if they are made from different genetic materials. However, the fact is that all the different tissues and cells in my body contain the same DNA and it is the differential expression profiles that created the functional diversity. Understanding the mechanism of gene expression will help us to understand the formation and evolution of life and to find possible cures for diseases.

In this issue

We have invited scientists on the leading edge in their respective fields to share their expertise and perspectives. Ma *et al*. (2014) focused on the mechanism that opens and closes specific regions in the chromatin. The issue starts with the packaging of human DNA. If stretched out, the total length of DNA from a human cell is about 3 m. In a living cell, DNA is packaged into a highly compact form called chromatin. Genes are not active when they are in the compacted form. Factors involved in opening and closing specific regions control the accessibility of genes (Guertin and Lis, 2012). A set of factors called polycomb group (PcG) proteins is involved in modifying the chromatin structure and subsequently regulating a large number of genes. The dynamic nature of the multisubunit protein complex and its complex functional impact are described in detail.

Gene regulation does not stop at finding the gene. The transcription machinery finds an accessible promoter and assembles a multi-megadalton protein complex to initiate transcription (Kornberg, 2007). As soon as the nascent transcript emerges from the body of RNA polymerase II (Pol II), the enzyme that

transcribes protein coding genes, the RNA is protected by the addition of a 5' cap (Shatkin, 1976). A text-book would describe the capping to be a unanimous phenomenon that takes place on every molecule of the RNA transcripts. However, Zhai and Xiang (2014) summarize the recent discovery that improperly capped RNA is subjected to quality control and degradation. Previously, removal of the cap and subsequent RNA degradation are thought to only take place after the RNA is exported to the cytoplasm and translation has been initiated. This discovery changed the way that mRNA capping and de-capping are viewed in time and space.

RNA degradation or turnover has a significant place in gene expression (Houseley and Tollervey, 2009). Compared to DNA, RNA is a very short-lived species in cells. Its relatively rapid turnover provides the flexibility for organisms to react to the environment and adapt quickly to a change in gene expression profile. Historically, mRNA degradation is only thought to occur once at least a round of translation has taken place. Liu H. *et al.* (2014) provide an updated review of the many RNA degradation pathways taking place in the nucleus. Degrading nascent RNA transcripts in the nucleus has been viewed as a quality control mechanism to quickly get rid of undesired RNAs before they are exported to cytoplasm and waste the effort of ribosomes in translating wrong proteins.

The start of transcription relies on promoters. The termination signal typically relies on sequence information at the end of the genes (Richard and Manley, 2009; Guo *et al.*, 2011). Davis and Shi (2014) review the current understanding of the sequence requirements and the protein factors involved in this process called mRNA 3' end processing. It cuts the RNA off the moving Pol II and adds a stretch of adenines to the RNA. The sequence driving this process is called polyadenylation signal or polyA signal (Proudfoot, 1991). It has been known that the promiscuous nature of the polyA signal sequences leads to the selective use of multiple polyA signals (Tian and Manley, 2013). This feature, described as alternative polyadenylation, has been linked with many aspects of gene expression: pre-mature termination, RNA splicing, RNA stability and expression level (Mayr and Bartel, 2009). However, the exact mechanism that dictates alternative polyA signal usage is unclear. This article also provides a simple yet broadly applicable model to characterize how the choice among many polyA signals is made.

The process of making RNA nucleotide by nucleotide, transcription elongation, is tightly controlled and interlaced with all above events. Indeed, elongation control is a critical step in determining transcriptional output. Instead of looking for an open promoter to start transcription, an initiated and paused Pol II can be found on all expressed human genes (Core *et al.*, 2012; Kwak *et al.*, 2013). This paused Pol II population serves as a reserve for transcription to rapidly respond to activation signals (Guo and Price, 2013). Upon activation, positive transcription elongation factor P-TEFb triggers the fast-paced productive elongation that leads to mRNA production. P-TEFb is therefore seen as the key for rapid amplification of gene expression. Indeed, this is the exact factor virus hijack to promote the activation of themselves (Zhou *et al.*, 2012). Liu R.D. *et al.* (2014) provide the latest understanding of factors involved in the transition into productive elongation and how HIV finds its niche to achieve replication effectively and in a deadly way.

On the journal

The *Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology)* has been at the frontier of peer-reviewed scientific journals based in China. Since its establishment in 2005, the editorial office has been dedicated to publishing high quality and high impact research and review articles.

In this special issue, we invited primary authors of recent high impact research articles to provide the most up-to-date view of their area of expertise. The studies they led were groundbreaking and provided significant milestones for their field. Justly, the invitees have all recently started their own labs with generous funding. Demonstrating their dedication to nurturing future scientists, they have all invited their trainees and colleagues to join in the adventure of composing the articles. We believe these cutting edge articles will promote discussions among established scientists and benefit the training of junior researchers.

References

Core, L.J., Waterfall, J.J., Gilchrist, D.A., *et al.*, 2012. Defining the status of RNA polymerase at promoters. *Cell Rep.*, **2**(4):1025-1035. [doi:10.1016/j.celrep.2012.08.034]

Crick, F.H., 1958. On protein synthesis. *Symp. Soc. Exp. Biol.*, **12**:138-163.

Crick, F., 1970. Central dogma of molecular biology. *Nature*, **227**(5258):561-563. [doi:10.1038/227561a0]

Davis, R., Shi, Y., 2014. The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation. *J. Zhejiang Univ.-Sci. B (Biomed. & Biotechnol.)*, **15**(5):429-437. [doi:10.1631/jzus.B1400076]

Guertin, M.J., Lis, J.T., 2013. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr. Opin. Genet. Dev.*, **23**(2): 116-123. [doi:10.1016/j.gde.2012.11.008]

Guo, J., Garrett, M., Micklem, G., *et al.*, 2011. Poly(A) signals located near the 5' end of genes are silenced by a general mechanism that prevents premature 3'-end processing. *Mol. Cell. Biol.*, **31**(4):639-651. [doi:10.1128/MCB.00919-10]

Guo, J., Price, D.H., 2013. RNA polymerase II transcription elongation control. *Chem. Rev.*, **113**(11):8583-8603. [doi:10.1021/cr400105n]

Houseley, J., Tollervey, D., 2009. The many pathways of RNA degradation. *Cell*, **136**(4):763-776. [doi:10.1016/j.cell.2009.01.019]

Kornberg, R.D., 2007. The molecular basis of eukaryotic transcription. *PNAS*, **104**(32):12955-12961. [doi:10.1073/pnas.0704138104]

Kwak, H., Fuda, N.J., Core, L.J., *et al.*, 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, **339**(6122):950-953. [doi:10.1126/science.1229386]

Lee, T.I., Young, R.A., 2013. Transcriptional regulation and its misregulation in disease. *Cell*, **152**(6):1237-1251. [doi:10.1016/j.cell.2013.02.014]

Liu, H., Luo, M., Wen, J.K., 2014. mRNA stability in the nucleus. *J. Zhejiang Univ.-Sci. B (Biomed. & Biotechnol.)*, **15**(5):444-454. [doi:10.1631/jzus.B1400088]

Liu, R.D., Wu, J., Shao, R., *et al.*, 2014. Mechanism and factors that control HIV-1 transcription and latency activation. *J. Zhejiang Univ.-Sci. B (Biomed. & Biotechnol.)*, **15**(5):455-465. [doi:10.1631/jzus.B1400059]

Ma, R.G., Zhang, Y., Sun, T.T., *et al.*, 2014. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *J. Zhejiang Univ.-Sci. B (Biomed. & Biotechnol.)*, **15**(5):412-428. [doi:10.1631/jzus.B1400077]

Mayr, C., Bartel, D.P., 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**(4):673-684. [doi:10.1016/j.cell.2009.06.016]

Proudfoot, N., 1991. Poly(A) signals. *Cell*, **64**(4):671-674. [doi:10.1016/0092-8674(91)90495-K]

Richard, P., Manley, J.L., 2009. Transcription termination by nuclear RNA polymerases. *Genes Dev.*, **23**(11):1247-1269. [doi:10.1101/gad.1792809]

Shatkin, A.J., 1976. Capping of eucaryotic mRNAs. *Cell*, **9**(4): 645-653. [doi:10.1016/0092-8674(76)90128-8]

Tian, B., Manley, J.L., 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem. Sci.*, **38**(6):312-320. [doi:10.1016/j.tibs.2013.03.005]

Young, R.A., 2011. Control of the embryonic stem cell state. *Cell*, **144**(6):940-954. [doi:10.1016/j.cell.2011.01.032]

Zhai, L.T., Xiang, S., 2014. mRNA quality control at the 5' end. *J. Zhejiang Univ.-Sci. B (Biomed. & Biotechnol.)*, **15**(5): 438-443. [doi:10.1631/jzus.B1400070]

Zhou, Q., Li, T., Price, D.H., 2012. RNA polymerase II elongation control. *Annu. Rev. Biochem.*, **81**(1):119-143. [doi:10.1146/annurev-biochem-052610-095910]

Introducing the guest editor:

Jiannan GUO received his Bachelor's degree at Northeast Forestry University in China and his Master's degree at the University of Bath in the United Kingdom. He then worked with Dr. Saverio BROGNA at the University of Birmingham in the United Kingdom for his Ph.D. Currently he is a post-doctoral fellow in the Price lab at the University of Iowa in the United States. Generally he is interested in the regulation of gene expression, and his expertise includes developing *in vitro* transcription assays to characterize factors involved in Pol II transcription elongation.

中文概要：

本文题目：转录在基因表达中的重要性
Transcription: the epicenter of gene expression

本文概要： 基因被认为是遗传信息的载体，但是生物体的复杂程度并不是由其所含基因数量决定的。基因表达的调控机制掌控着哪些遗传信息被使用，以及使用到何种程度。高等生物的复杂性正是这样实现的：一个人有各种不同器官，而其中的细胞中却携带着相同的基因。特定的基因被不同程度地表达导致了某些细胞发展成头发，某些细胞发展成手指等。基因表达的调控中，转录是初始的一个步骤，也是最关键的步骤之一。近二三十年来，分子生物学和生物化学的发展对研究转录的调控机制起到了重要的推动作用。调控的机理、调控的对象以及促使调控产生的信号，都在进行更深入和广泛的研究。

关键词组： 基因表达；转录；DNA；RNA