Cell

# Transcription, Splicing, and Release: Are We There Yet?

Ranjan Sen[1,*] and Sebastian D. Fugmann[1]
[1]Laboratory of Molecular Biology and Immunology, National Institute on Aging, 251 Bayview Boulevard, Baltimore, MD 21224, USA
*Correspondence: ranjan.sen@nih.gov
http://dx.doi.org/10.1016/j.cell.2012.07.004

**A systematic analysis of LPS-induced gene expression in macrophages by Bhatt et al. demonstrates that inflammatory responses are governed primarily at the level of transcription initiation. Unexpectedly, full-length nascent RNAs that contain introns appear to accumulate on chromatin, presumably to complete processing, prior to release of functional mRNA for export to the cytoplasm.**

Interaction of lipopolysaccharide (LPS) of Gram-negative bacteria with Toll-like receptor 4 (TLR4) on the mammalian cell surface induces a transcriptional cascade that leads to acute inflammation. Key mediators of this response include proinflammatory cytokines, such as interleukins 1, 6, and 12 and TNF$\alpha$, as well as chemokines and their receptors that are used to mobilize cells to sites of infection and injury. In this issue of *Cell*, Bhatt et al. (2012) now provide a comprehensive picture of LPS-induced transcriptional responses in macrophages using deep RNA sequencing technology. In an unexpected twist, the authors uncover evidence that challenges the widely accepted model of cotranscriptional RNA splicing.

Transcriptional responses to LPS have been best characterized in macrophages and dendritic cells of the innate immune system. These studies identified primary and secondary LPS-responsive genes in macrophages (Ramirez-Carrozzi et al., 2009; Hargreaves et al., 2009). Primary response genes are rapidly induced in the absence of new protein synthesis, as they are "poised" for transcription. This poised state is characterized by activation-associated histone modifications, as well as promoter-bound RNA polymerase II prior to activation. In contrast, the induction of secondary response genes requires ATP-dependent chromatin remodeling and de novo protein synthesis of transcriptional activators. At the structural level, promoters of many primary response genes are enriched for CpG dinucleotides (commonly referred to as CpG island promoters), which prevent the formation of stable nucleosomes and allow for their easy displacement en route to transcription activation. Secondary response gene promoters, however, are low in CpG content (and therefore lack unstable nucleosomes).

These pioneering studies were limited in two ways. First, transcription responses were measured by assessing messenger RNA (mRNA) levels using microarrays. This technique scores for the cumulative outcome of several processes, including transcription, mRNA processing, and mRNA stability, rather than RNA-polymerase-dependent transcription per se. Second, a limited number of genes were selected to define the biochemical features associated with primary and secondary response genes.

Bhatt et al. (2012) analyze the kinetics of LPS-induced transcriptional response during a 2 hr time course. Specifically, they compare the profiles of newly transcribed (or nascent) RNA to nucleoplasmic and cytoplasmic RNAs, both by deep sequencing. To directly measure ongoing transcription, the authors use a cell fractionation scheme originally devised by Ueli Schibler, which isolates chromatin-associated RNA (considered to reflect newly transcribed RNA transcripts) as a source of nascent RNA. This method identifies 560 LPS-induced genes, which can be grouped into seven categories based on kinetics and temporal (transient versus continuous) patterns of gene activation. Interestingly, the expression profile of nascent RNAs closely resembles nucleoplasmic and cytosolic profiles in all categories, indicating that LPS-induced gene expression is regulated largely at the level of transcription initiation. Similar conclusions were reached in analyses of LPS-induced transcription in dendritic cells (Rabani et al., 2011).

This genome-wide transcriptional data permitted Bhatt et al. (2012) to re-evaluate the relative roles of CpG island promoters and low-CpG promoters in the LPS response. Consistent with the previous studies, CpG island promoters are more prevalent among primary response genes, whereas low-CpG promoters are also more prevalent among secondary response genes. However, many CpG promoters are present among secondary response genes. Moreover, both primary and secondary CpG promoters are poised by the criteria noted above, yet they have distinctive induction characteristics. Thus, CpG content of a promoter per se dictates neither the basal transcription level of a given gene nor its temporal responsiveness to the LPS stimulus. One interesting difference observed by Bhatt et al. (2012) is that CpG promoters, on average, have lower fold-induction response to LPS compared to low-CpG promoters. But again, there are many genes that do not follow this trend. Thus, it will be interesting to compare promoter architectures of the low- and high-responding CpG promoters, as well as the high-responding CpG promoters to the high-responding LCG promoters. The latter comparison will be particularly illuminating with regard to mechanisms that allow these genes to be "super inducible."

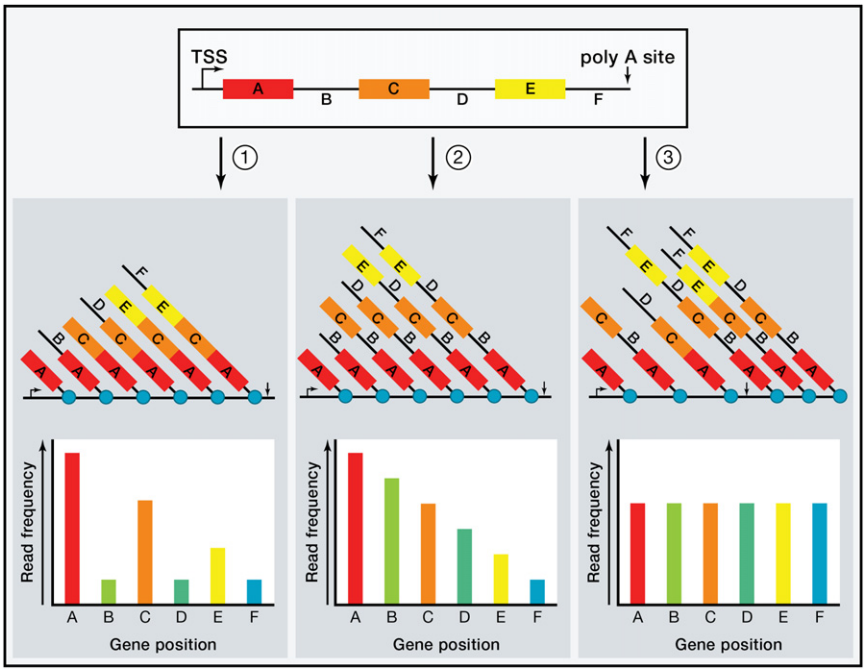


**Figure 1. Coordination of Transcription and RNA Processing**

(Top) A hypothetical gene contains exons A, C, and E and introns B, D, and F. Arrows at the left and right represent the sites for transcription initiation and cleavage/poly(A) addition, respectively. (Middle) Idealized views of cotranscriptional splicing (arrow 1), splicing following the completion of transcription (arrow 2), and partial cotranscriptional splicing and accumulation of partially spliced full-length transcripts, as described by Bhatt et al. (2012) (arrow 3). In model 1, essentially no introns are left on polymerase-associated nascent transcripts. All introns are left on the transcripts in model 2, and most introns are left on full-length transcripts that accumulate on chromatin after transcription is complete in model 3. (Bottom) The y axis represents idealized read frequencies of nascent RNA-seq expected from each model, whereas exons and introns are indicated along the x axis.

Bhatt and colleagues also find that nascent RNAs derived from genes at transcriptional steady state contain comparable levels of intron and exon sequences across the length of each gene and terminate at or close to the polyadenylation (poly(A)) site. Quantitative analyses of the constitutively expressed genes in macrophages show that a substantial numbers of genes fit this profile, leading them to propose that full-length, partially spliced transcripts remain tethered to chromatin for some time prior to being released for transport to the cytoplasm (Figure 1).

At steady state, transcription is viewed as a gene with multiple RNA polymerase complexes actively moving all along the length of the gene. In line with this model, genome-wide nuclear run-on (Gro-seq) (Core et al., 2008) experiments show approximately a uniform distribution of newly transcribed RNA across a gene. Viewed from this perspective, nascent RNAs should be enriched for sequences at the 5′ compared to the 3′ ends of genes (Figure 1A, models 1 and 2). In contrast, Bhatt et al. (2012) observe equal representation of sequences from all parts of genes. How does one reconcile this inconsistency? One possibility is that the shorter incomplete transcripts are lost during the purification procedure and are therefore underrepresented in nascent RNA. This seems unlikely because Bhatt and colleagues detect these shorter RNAs at inducible genes during the early response phase. Alternatively, at steady state, full-length RNA might accumulate to higher levels on chromatin than incomplete shorter transcripts in the gene body, and thus, these longer RNAs dominate the RNA sequencing (RNA-seq) data (model 3).

The apparent accumulation of full-length transcripts that retain introns necessitates reevaluation of our views on cotranscriptional splicing. The prevailing hypothesis is that introns are spliced out as splice donor and acceptor sites become available on newly transcribed RNA (Figure 1A, model 1). In this scenario, introns derived from the 5′ end should be completely absent, whereas some slowly spliced introns from the 3′ end may still be detectable in the nascent RNA. Evidence for this model in metazoans is largely based on analyses of a small set of individual genes (Singh and Padgett, 2009; Pandya-Jones and Black, 2009). Khodor et al. (2011), however, recently addressed this question on a genome-wide scale using a procedure very similar to that used by Bhatt et al. (2012). They found that most genes were spliced cotranscriptionally in *Drosophila* S2 cells (Khodor et al., 2011). Interestingly, they observed preferential retention of the first intron among the residual introns left in nascent RNA. In contrast, Bhatt et al. (2012) provocatively conclude that incomplete intron removal is frequent and widely distributed in nascent RNA.

These observations raise several questions. Given the increasing use of this method for isolating nascent RNA, it is important to know how the chromatin-associated RNA is tethered to chromatin. During ongoing transcription, RNA polymerase provides the physical link, but the full-length transcripts identified by Bhatt et al. (2012) have been cleaved at the poly(A) site and should no longer be associated with RNA polymerase. Thus, these transcripts are retained on chromatin by an unknown mechanism.

The working hypothesis is that additional time is required to complete RNA processing and to thereby ensure the selective release of functional mRNAs for export to the cytoplasm. From a mechanistic standpoint, it will be interesting to determine the temporal relationship between 3′ end processing, splicing, and transcript release from chromatin. Specifically, once RNA polymerase extends past the poly(A) site and the primary transcript is cleaved, is the poly(A) tail added immediately? Or does splicing occur first followed by poly(A) addition? Analysis of poly(A)+ and poly(A)− chromatin-associated RNAs from the same sample will provide insight into this question. Furthermore, it would be useful to know what proportion of the chromatin-associated RNA contains a poly(A) tail. In our current view, 3′ cleavage and poly(A) occur

rapidly; if so, the majority of cleaved full-length transcripts observed by Bhatt and colleagues should be poly(A)+. The presence of a significant number of introns in nascent RNA and their virtual absence in nucleoplasmic RNA suggests that release of RNA from chromatin is not a stochastic process but rather a regulated one occurring after splicing is complete. This predicts that a biochemical signal may initiate this release by "dissolving the glue" that holds RNA and chromatin together. It will be interesting to discover such a signal and the nature of this glue. Alternatively, the lag time of RNA on chromatin may be predetermined by characteristics of the primary transcript, such as transcript length or the size, number, and quality of introns.

One important aspect to keep in mind is that RNA-seq technology provides an ensemble average that precludes direct evaluation of individual transcription units. For example, in the nascent RNA-seq analysis described by Bhatt and colleagues, the question of how many full-length transcripts remain associated per transcribing allele is open to interpretation. Figure 1 portrays several full-length RNA molecules associated with one allele (to explain the dominance of full-length transcripts in RNA-seq); however, it is equally possible that there is only one full-length transcript per allele, but there are many more cells with this configuration than there are cells with partially transcribed genes. Thus, resolving the question of what actually happens on a transcribing gene will require combining ensemble studies with single-molecule techniques in the future.

## ACKNOWLEDGMENTS

R.S. thanks Dr. Michael Rosbash for valuable discussion. This work was supported entirely by the Intramural Research Program of the NIH National Institute on Aging.

## REFERENCES

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Cell *150*, this issue, 279–290.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Science *322*, 1845–1848.

Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Cell *138*, 129–145.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., II, and Rosbash, M. (2011). Genes Dev. *25*, 2502–2512.

Pandya-Jones, A., and Black, D.L. (2009). RNA *15*, 1896–1908.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Nat. Biotechnol. *29*, 436–442.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). Cell *138*, 114–128.

Singh, J., and Padgett, R.A. (2009). Nat. Struct. Mol. Biol. *16*, 1128–1133.

# Unfolding the Bridge between Transcription and Translation

**Vladimir Svetlov[1] and Evgeny Nudler[1,*]**
[1]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA
*Correspondence: evgeny.nudler@nyumc.org
http://dx.doi.org/10.1016/j.cell.2012.06.025

**Transcription antiterminator RfaH alternates between closed (inactive) and open (activated) conformation. In this issue of *Cell*, Burmann et al. show that opening is accompanied by dramatic all-α to all-β refolding of its C-terminal domain. Each of the folds has a distinct function: all-α-fold acts as a specificity determinant, directing RfaH to a small subset of operons, whereas the all-β-fold recruits ribosome, thereby coupling RfaH-stimulated transcription to translation.**

In bacterial cells, RNA polymerases (RNAPs) and ribosomes populate the same space, accessible by simple diffusion, which allows occupation of the same mRNA by both transcription and translation machineries. Cotranscriptional translation (or transcription-translation coupling) was known to play a role in such regulatory mechanisms as transcription attenuation and operon polarity; more recently, trailing ribosomes were shown to affect the rate of transcription by suppressing RNAP backtracking, harmonizing the rates of mRNA and protein synthesis (Proshkin et al., 2010). A direct physical link between the RNAP elongation complex and the trailing ribosome was discovered, wherein the general transcription factor NusG engaged the RNAP with its N-terminal domain (NTD) while interacting with ribosomal protein S10( = NusE) via the C-terminal domain (CTD) (Burmann et al., 2010). RNAP-NusG-S10 bridge complements ribosome binding to mRNA through the engagement of the start codon and Shine-Dalgarno sequence (SDS). In this issue of *Cell*,