Long Noncoding RNAs Add Another Layer to Pre-mRNA Splicing Regulation

Minna-Liisa Änkö[1] and Karla M. Neugebauer[1],*
[1]Max Planck Institute of Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
*Correspondence: neugebau@mpi-cbg.de
DOI 10.1016/j.molcel.2010.09.003

In this issue of *Molecular Cell*, Tripathi and coworkers (Tripathi et al., 2010) decode some of the functions of a long noncoding RNA MALAT1. They provide evidence that MALAT1 regulates alternative splicing by controlling the activity of the SR protein family of splicing factors.

Protein-coding genes cover only a small fraction of mammalian genomes, which raises a puzzling question about the function of the pervasive transcription of the noncoding genome. Noncoding RNAs (ncRNAs) can be arbitrarily divided into short ncRNAs, typically ~20–200 nt in length, and long ncRNAs ranging from ~200 nt to 100 kb (Mercer et al., 2009). Although much evidence on the importance of the different types of short ncRNAs in the regulation of gene expression has accumulated, the functions of long ncRNAs remain largely unexplored. It seems unlikely that the long ncRNAs are just transcriptional noise, because many of them are conserved, their expression is developmentally regulated, and they have tissue-specific expression patterns (Mercer et al., 2009). Furthermore, several long ncRNAs appear to be misregulated in diseases, such as cancer. Loci encoding long ncRNAs often overlap with or are interspersed between multiple protein-coding or noncoding genes in the genome, where they may regulate the expression of their neighbors. Moreover, long ncRNAs have been shown to act as chromatin modifiers, as transcriptional regulators that affect the activity of transcription factors or RNA polymerase II, and as posttranscriptional regulators that bind to complementary RNA sequences (Mercer et al., 2009). In other words, long ncRNAs seem to have the capacity to operate on every step of gene expression.

One of the most versatile processing steps in the life of an mRNA molecule is pre-mRNA splicing. Most eukaryotic pre-mRNAs contain introns, which have to be removed, and exons, which are ligated together, before mRNAs can be trans-lated into proteins. Although the spliceosome performs splicing catalysis, additional splicing factors are required to determine the splice sites. Pre-mRNAs can be alternatively spliced, providing a major source of transcriptome and proteome diversity in cells (Pan et al., 2008). SR proteins are a family of splicing factors involved in both constitutive and alternative splicing (Lin and Fu, 2007). They all share a modular structure with one or two N-terminal RNA recognition motifs (RRMs) and an RS domain composed of arginine-serine repeats. The RS domain is a target for extensive phosphorylation, and the phosphorylation status of the RS domain is important for the activity of SR proteins in splicing. SR proteins are primarily located in the nucleus where they are enriched in speckles. Although SR proteins are abundantly and ubiquitously expressed in cells and tissues, their cellular expression levels are tightly regulated. Indeed, even relatively small perturbations in the concentration of SR proteins can influence alternative splicing patterns (Lin and Fu, 2007). However, how the concentration or activity of SR proteins is regulated in cells is not fully understood.

Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1; also known as nuclear enriched autosomal transcript 2, NEAT2) is a highly conserved long ncRNA originally identified as a transcript overexpressed in many cancers (Ji et al., 2003). Interestingly, MALAT1 is retained in the nucleus and predominantly localizes to nuclear speckles. Because SR proteins and other splicing factors are also enriched in these nuclear subdomains, Tripathi and coworkers asked in this issue of *Molecular Cell* whether MALAT1 could play a role in the regulation of pre-mRNA metabolism and/or the integrity of the speckle per se (Tripathi et al., 2010). They first determined whether MALAT1 directly associates with SR proteins. Computational analysis identified predicted SR protein-binding motifs in MALAT1 RNA, and specific interactions between MALAT1 and a subset of SR protein family members were shown by RNA immunoprecipitation. One of the SR proteins, SRSF1, displayed specific binding to MALAT1 via its RRM.

Do SR proteins recruit MALAT1 to nuclear speckles? Depletion of SRSF1 suggested that this is not the case; however, it is formally possible that other SR protein family members play a redundant role in MALAT1 localization to speckles. Interestingly, another long ncRNA NEAT1 was recently shown to be critical for the integrity of para-speckles, nuclear subdomains implicated in the regulation of mRNA nuclear export (Bond and Fox, 2009; Chen and Carmichael, 2009). In agreement with a role for ncRNA in the establishment or maintenance of nuclear subdomains, MALAT1 was required for proper localization of SRSF1 as well as several other splicing factors to nuclear speckles (Tripathi et al., 2010). However, MALAT1 is apparently not required for speckle integrity, because poly A+ RNA and UAP56 continue to localize in a speckled pattern when MALAT1 is depleted. Thus, MALAT1 activity appears to be specifically linked to SR proteins rather than generally to speckles as nuclear subcompartments.

Links between the regulation of transcription and splicing are provided in a related paper, which showed that

MALAT1 localization was dependent on transcription; moreover, MALAT1 depletion compromised the recruitment of SR proteins from speckles to the sites of transcription, where splicing occurs (Bernard et al., 2010). Both studies show that depletion of MALAT1 leads to changes in alternative splicing of a subset of transcripts (Bernard et al., 2010; Tripathi et al., 2010). A possible mechanism for these effects is suggested by the observation that upon MALAT1 depletion, expression of SRSF1 protein is increased. Importantly, SRSF1 overexpression mimicked MALAT1 depletion in terms of alternative splicing phenotype. The consequences of an observed increase in dephoshorylated splicing-inactive SRSF1 awaits for further investigation. Taken together, MALAT1 regulates the expression level, localization, and activity SR proteins.

Tripathi and coworkers propose that MALAT1 could act as a "molecular sponge" by interacting with SR proteins in the nuclear speckles, and thereby modulate the concentration of splicing-competent SR proteins in cells (Figure 1A). This model has an interesting correlation to splicing-associated disease mechanisms. Microsatellite expansions in noncoding regions of genes can cause disease through an RNA gain-of-function mechanism (Ranum and Cooper, 2006). The best-characterized example of a microsatellite expansion disease is the myotonic dystrophy type 1 (DM1). The cause of DM1 is a CTG expansion in the 3′ untranslated region (UTR) of dystrophia myotonica protein kinase (DMPK) gene. The repeat-containing RNA accumulates in the nucleus and affects the activity of splicing factors. A predominant molecular feature of DM1 is a misregulation of alternative splicing of a subset of transcripts, caused by the altered splicing factor activity
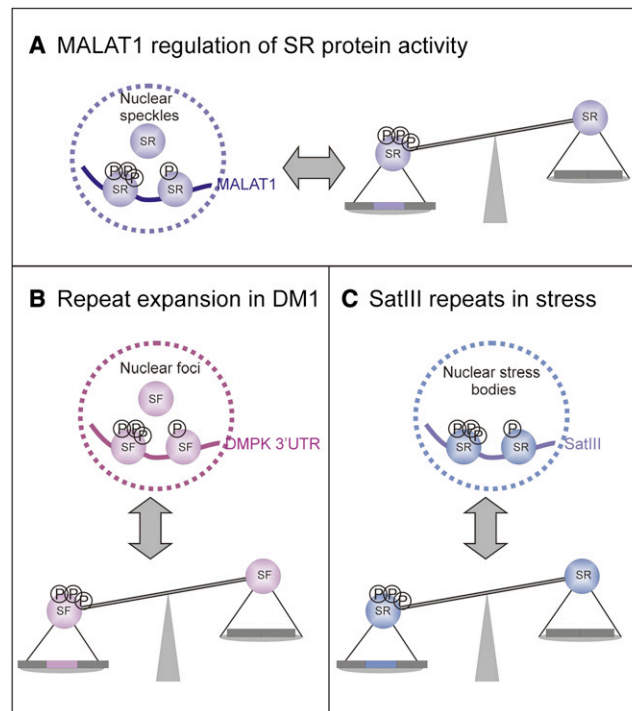

Figure 1. ncRNAs and Expanded RNA Repeats Have the Properties of Acting as a "Molecular Sponge" for Splicing Factors
(A) Association of SR proteins with MALAT1 in the nuclear speckles alters their activity by changing their phosphorylation status. This leads to changes in alternative splicing patterns.
(B) In DM1, a trinucleotide repeat expansion in the 3′UTR of DMPK gene leads to accumulation of the DMPK transcript in the nucleus. The presence of the repeat RNA alone leads to altered activity of splicing factors (muscleblind-like 1 and CUG repeat-binding protein 1) and further changes in alternative splicing.
(C) Upon cellular stress, the SatIII loci are transcribed, and this leads to accumulation of SR proteins into the nuclear stress bodies. SR, SR protein; SF, splicing factor; P, phosphate group.

(Figure 1B). Similarly, CGG repeats occur in the fragile X mental retardation gene 1 (FMR1) in the case of fragile X-associated tremor ataxia syndrome (FXTAS), and they recruit a set of splicing regulators into nuclear inclusions (Sellier et al., 2010).

Independent of mutations, constitutive heterochromatin contains various repeated DNA sequences that are transcribed, especially in cellular stress conditions (Denegri et al., 2001). These satellite RNA transcripts recruit splicing factors to nuclear stress bodies, thereby lowering the active concentration of the splicing factors in the nucleus (Figure 1C). Could the repeat RNAs resulting from a microsatellite expansion or cellular stress mimic the action of long ncRNAs,

thereby altering the localization and concentration of the active pool of splicing factors similar to MALAT1? Intriguingly, because ncRNAs do not encode proteins, they can be expressed by cells to elicit regulation, as in the sponge mechanism, without the potential "side effects" or restrictions imposed by protein expression. Regardless of the mechanism, these examples highlight the potential of long ncRNAs as important regulators of pre-mRNA splicing, which may be extended to further ncRNAs and/or splicing factors in the future.

REFERENCES

Bernard, D., Prasanth, K.V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourdren, L., Coulpier, F., et al. (2010). EMBO J. Published online August 20, 2010. 10.1038/emboj.2010.199.

Bond, C.S., and Fox, A.H. (2009). J. Cell Biol. 186, 637–644.

Chen, L.-L., and Carmichael, G.G. (2009). Mol. Cell 35, 467–478.

Denegri, M., Chiodi, I., Corioni, M., Cobianchi, F., Riva, S., and Biamonti, G. (2001). Mol. Biol. Cell 12, 3502–3514.

Ji, P., Diederichs, S., Wang, W., Boing, S., Metzger, R., Schneider, P.M., Tidow, N., Brandt, B., Buerger, H., Bulk, E., et al. (2003). Oncogene 22, 8031–8041.

Lin, S., and Fu, X.D. (2007). Adv. Exp. Med. Biol. 623, 107–122.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Nat. Rev. Genet. 10, 155–159.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Nat. Genet. 40, 1413–1415.

Ranum, L.P.W., and Cooper, T.A. (2006). Annu. Rev. Neurosci. 29, 259–277.

Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R.K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D.J., et al. (2010). EMBO J. 29, 1248–1261.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). Mol. Cell 39, this issue, 925–938.