No way out: when RNA elements promote nuclear retention

F. Agostini[1], J. Ule[1,2], J. A. Zagalak[1,2]

[1] The Francis Crick Institute, Midland Road 1, Kings Cross, London NW1 1AT

[2] Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK

Summary

The proper localization of RNA transcripts is a highly controlled and fine-tuned process. Indeed, regulation of RNA trafficking is mediated by both cis-acting and trans-acting factors, and defects in either mechanism have been associated with disease. Identifying the RNA sequence motifs that determine cellular localization for a given transcript therefore represents an important and challenging task. A new study from *Shukla et al.* in *The EMBO Journal* – along with related work from Lubelsky and Ulitsky published elsewhere – describes a new screen that uses hybrid RNAs with barcoded oligonucleotides to identify cis-acting elements that increase the propensity of RNAs to be retained in the nuclear compartment.

KEYWORDS

RNA localization, RNA import/export and retention, lncRNAs, Alu elements

Main body

The ability to sort RNAs into specific cellular compartments is a fundamental property of eukaryotic cells. The primary mechanisms guiding the delivery of RNA molecules involve a plethora of RNA-binding proteins (RBPs) that decorate each RNA. These RBPs bind to cis-acting elements on RNA molecules, which often consist of single-stranded motifs and RNA hairpin loops (Martin & Ephrussi, 2009). The quest to unravel the nature of cis-acting RNA elements involved in RNA localization began with genetic and microinjection studies of hybrid mRNA transcripts. These studies identified diverse RNA localization elements in cell types and species ranging from Drosophila oocytes to human neurons (Martin & Ephrussi, 2009). Notably, several RNA localization elements and their bound RBPs show conserved function in multiple cell types and species. While genetic screens have often been used to identify trans-acting factors, screens to comprehensively identify cis-acting elements have entered the scene more recently. The first studies to attempt this inserted a library of cis-acting sequences

into 3’ untranslated regions (UTRs) that were fused to reporter genes, which provided an indirect readout of the UTR activity (Zhao *et al*, 2014; Oikonomou *et al*, 2014).

While most mRNAs are efficiently exported, the export of long non-coding RNAs (lncRNAs) is generally less efficient, indicating that specific mechanisms can inhibit export and regulate nuclear RNA retention (Derrien *et al*, 2012). Most of the known sequences that regulate nuclear RNA export in mammals were originally identified in viral genomes (Giulietti *et al*, 2015). Two new studies now use elegant tiling-based approaches to screen for cis-acting RNA elements involved in nuclear retention. The first study (*Shukla et al 2018)* developed Massively Parallel RNA Assay (MPRNA) that is based on sequences from 38 human lncRNAs, which were chosen to cover a spectrum of nuclear and cytosolic localization patterns. Briefly, 11.969 unique, barcoded oligos of 153 nt length were designed to tile the lncRNA sequences in a partially overlapping manner (Figure 1). The oligos were fused to a reporter, a frame-shifted Sox2 mutant (fsSox2) which does not encode a full-length protein. This reporter, known for its native cytoplasmic localization, was transfected into HeLa cells, which were then fractionated into nuclear and cytosolic fraction, and the barcodes in the reporter were sequenced. The authors implemented and applied a statistical method to rank and aggregate the tile-derived results into longer regions. By this means, 109 unique RNA differential regions (DRs) were identified as significantly shifting the localization of the fusion transcripts to the nuclear compartment. Interestingly, a cytosine-rich motif was enriched within these regulatory sequences.

The second study (*Lubelsky and Ulitsky 2018)* set out to identify nuclear localization signals using a similar tiling-based screening approach covering exonic sequences of 37 human lncRNAs, 13 3' UTRs of mRNAs enriched in the nucleus and four homologs of MALAT1. The screen was carried out by the use of 5.511 unique, partially overlapping, barcoded oligonucleotides (oligos) of 109 nt length. The oligos were cloned into 5' and 3' UTRs of AcGFP mRNA and transfected into human MCF-7 cells. Subcellular fractionation followed by sequencing revealed 19 regions from 14 genes that increased nuclear enrichment by >30%. The three lncRNA regions showing the highest levels of nuclear enrichment overlapped Alu repeat sequences, representing a 42 nt fragment that contained three cytosine-rich elements that matched a RCCTCCC (R=A/G) consensus sequence. This nuclear localization motif was therefore named “SINE-derived nuclear RNA LOcalizatIoN element”, or SIRLOIN. Thus, both studies identified C-rich motifs as mediating nuclear RNA retention. The researchers also searched for RBPs that may bind to the SIRLOIN motif by using the ENCODE eCLIP data, and found the motif to be most enriched at the binding sites of the hnRNPK protein.

Both studies went on to examine the enrichment of the identified motifs in retained RNAs. The study by *Lubelsky and Ulitsky* found that both the presence of SIRLOIN motifs and hnRNPK binding sites correlate with nuclear enrichment of lncRNAs and mRNAs. *Shukla et al*. performed a comparative analysis of ENCODE RNA-seq data produced from nuclear and cytoplasmic fractions of 11 cell lines and similarly found a correlation of DRs with nuclear retention, with the effect being more

pronounced for lncRNAs than mRNAs. Interestingly, the RNAs that are strictly nuclear tend to have many repeats of DRs; for example, 18 DRs are present in XIST and 10 in MALAT1 lncRNAs. This indicates that RBPs mediating nuclear enrichment might be more efficiently bound at multivalent motifs, reminiscent of the importance of such motifs for the regulation of RNA processing (Cereda *et al*, 2014). Indeed, many RBPs that regulate splicing were previously reported to be involved in mRNA export (Giulietti *et al*, 2015), and hnRNPK can also regulate splicing (Tsai *et al*, 2013), indicating a potential for common regulatory principles.

Lubelsky and Ulitsky find that the SIRLOIN motif is derived from Alu elements that are integrated in the antisense direction. These Alu elements are primate-specific, and are most prevalent in introns of human genes, since their exonization is normally suppressed due to binding of hnRNPC to antisense Alu elements (Zarnack *et al*, 2013). It has been shown that hnRNPC represents the first layer of quality control in the evolution of new Alu-derived exons and retained introns (Attig *et al*, 2016), and it could be speculated that the hnRNPK-SIRLOIN mechanism of nuclear retention represents the next layer of quality control for newly emerging Alu-containing transcripts. It remains to be seen if non-primate species contain other types of repetitive elements that are under similar types of regulation. Given the potential interplay between multiple SiRLOINs and DRs that can be present within each RNA, it is clear that further studies will be essential to fully understand their combinatorial effects on nuclear retention. Genome editing techniques that disrupt the cis-acting motifs within endogenous genes and quantify the resulting changes in a native cellular environment are likely to provide valuable insights into this interplay, as has been demonstrated by the recent studies of 3' UTRs (Wu *et al*, 2017; Zhao *et al*, 2017).

References

Attig J, Ruiz de Los Mozos I, Haberman N, Wang Z, Emmett W, Zarnack K, König J & Ule J (2016) Splicing repression allows the gradual emergence of new Alu-exons in primate evolution. *Elife* **5:** Available at: http://dx.doi.org/10.7554/eLife.19545

Cereda M, Pozzoli U, Rot G, Juvan P, Schweitzer A, Clark T & Ule J (2014) RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. *Genome Biol.* **15:** R20

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22:** 1775–1789

Giulietti M, Milantoni SA, Armeni T, Principato G & Piva F (2015) ExportAid: database of RNA elements regulating nuclear RNA export in mammals. *Bioinformatics* **31:** 246–251

Lubelsky Y, Ulitsky I. (2018) Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. Nature doi:10.1038/nature25757

Martin KC & Ephrussi A (2009) mRNA localization: gene expression in the spatial dimension. *Cell* **136:** 719–730

Oikonomou P, Goodarzi H & Tavazoie S (2014) Systematic identification of regulatory elements in conserved 3' UTRs of human transcripts. *Cell Rep.* **7:** 281–292

Shukla CJ, McCorkindale AL, Gerhardinger C, Korthauer KD, Cabili MN, Shechner DM, Irizarry RA, Maass PG, Rinn JL (2018) High–throughput identification of RNA nuclear enrichment sequences. The EMBO Journal DOI:10.15252/embj.201798452

Tsai P-L, Chiou N-T, Kuss S, García-Sastre A, Lynch KW & Fontoura BMA (2013) Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing. *PLoS Pathog.* **9:** e1003460

Wu Q, Ferry QRV, Baeumler TA, Michaels YS, Vitsios DM, Habib O, Arnold R, Jiang X, Maio S, Steinkraus BR, Tapia M, Piazza P, Xu N, Holländer GA, Milne TA, Kim J-S, Enright AJ, Bassett AR & Fulga TA (2017) In situ functional dissection of RNA cis-regulatory elements by multiplex CRISPR-Cas9 genome engineering. *Nat. Commun.* **8:** 2109

Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe NM & Ule J (2013) Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152:** 453–466

Zhao W, Pollack JL, Blagev DP, Zaitlen N, McManus MT & Erle DJ (2014) Massively parallel functional annotation of 3' untranslated regions. *Nat. Biotechnol.* **32:** 387–391

Zhao W, Siegel D, Biton A, Tonqueze OL, Zaitlen N, Ahituv N & Erle DJ (2017) CRISPR-Cas9-mediated functional dissection of 3'-UTRs. *Nucleic Acids Res.* **45:** 10800–10810

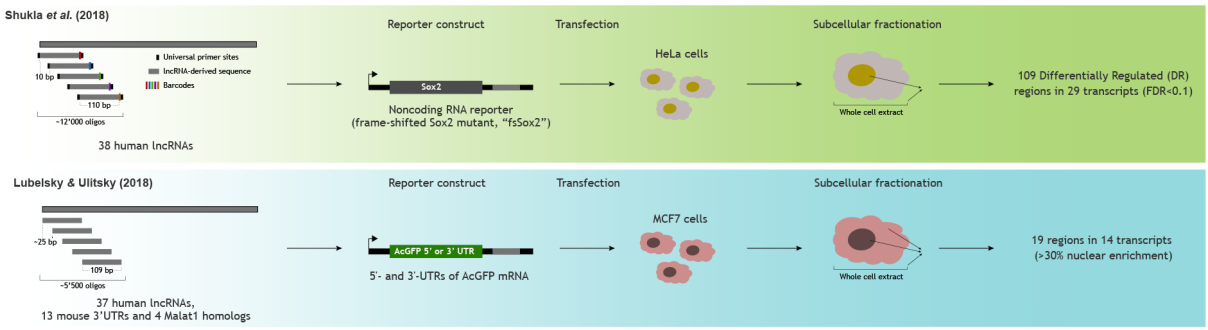

Fig 1: High-throughput screening approaches using oligos allow the identification of cis-acting RNA motifs. The barcoded oligos contain unique, and partially overlapping sequences that can be traced back to defined regions of reference RNAs. These oligos were cloned into reporter gene constructs, which contain either a frame-shifted Sox2 mutant (fsSox2) or the 5' and 3' UTRs of AcGFP mRNA, and then transfected into cancer-derived human cell lines. This was followed by assessing the subcellular localization with RNA-seq analysis of nuclear and cytoplasmic fractions, as well as whole cell lysate.