

Intronic retroelements

Not just “speed bumps” for RNA polymerase II

Kristel Kaer and Mart Speek*

Department of Gene Technology; Tallinn University of Technology; Tallinn, Estonia

Two well-known retroelements, L1 and Alu, comprise about one third of the human genome and are nearly equally distributed between the intergenic and intragenic regions. They carry different regulatory elements and contribute structurally and functionally to the expression of our genes. Recent data also suggest that hundreds of intronic L1s and Alus interfere with the transcription of human genes by inducing intron retention, forcing exonization and cryptic polyadenylation. These novel features can be explained with the RNA polymerase kinetic model and suggest that intronic L1s and Alus are not just “speed bumps” in regulation of RNA polymerase traffic. Here we discuss the complexity of the regulation of gene transcription imposed by intronic retroelements and predict that in addition to transcriptional activity, transcription factor binding and nucleosomal occupancy play a significant role in the transcriptional interference effects of the host genes.

Keywords: LINE, SINE, retrotransposon, transcriptional interference, exonization, intron retention, cryptic polyadenylation, mobile genetic elements

Abbreviations: RNA pol, RNA polymerase; TI, transcriptional interference; ASP, antisense promoter; TF, transcription factor

Submitted: 04/05/12

Revised: 05/14/12

Accepted: 05/15/12

<http://dx.doi.org/10.4161/mge.20774>

*Correspondence to: Mart Speek;
Email: Mart.Speek@ttu.ee

Commentary to: Kaer K, Branovets J, Hallikma A, Nigumann P, Speek M. Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. *PLoS One* 2011; 6:e26099; PMID:22022525; <http://dx.doi.org/10.1371/journal.pone.0026099>

Long and short interspersed nuclear elements, classified as LINEs (L1, L2, L3) and SINEs (Alus, SVA and other composite elements) are the two most abundant families of type I mobile genetic elements or retroelements, which have been generated by reverse transcription and transposition into mammalian genomes over the last ~170 million years. L1s and Alus are the representative members of these two classes, comprising about 17% and 10% of the human genome, respectively.¹ L1s are the only autonomous and retrotranspositionally competent elements (~100 active copies per individual genome²),

encoding functional reverse transcriptase. They are involved in trans-mobilization of other, passive or nonautonomous retroelements, such as Alus, SVAs and generation of processed pseudogenes (for a review see ref. 3). When inserted into the genome, retroelements could influence the transcription of the host genes by various ways. For example, they could provide alternative promoters, involved in alternative splicing and termination of transcription. Depending on the location, insertion into coding region (exon) breaking ORF sequence disrupts the normal gene expression. However, insertion into intron (or intergenic region) is generally less damaging and thus may or may not influence the transcription of the host gene.

About 7,000 full-length human L1s³ are nearly equally distributed between intragenic and intergenic regions. Most of them contain broken ORFs and have accumulated many mutations, as a result of which they are no longer able to retrotranspose. Nevertheless, perhaps thousands of them have retained transcriptional activity⁴⁻⁶ and/or contain binding sites for YY1, SOX family members, RUNX3 and other unknown transcription factors (TFs).⁷ Similarly to L1, about half of one million Alus are located in introns of genes. Only a small fraction of intra- or intergenic Alus are transcriptionally active. However, activation of many Alus depends on the cell stress (e.g. viral infection and heat shock). Alus are transcribed by RNA polymerase (RNA pol) III, and contain many functional TF binding sites for RNA pol III as well as RNA pol II.⁸⁻¹⁰ Despite their small size (~300 nt), compared with full-length L1s (~6.0 kb), Alus of different ages

(old-J, intermediate-S and young-Y) have accumulated many mutations which resulted in loss of TF binding in some elements or gain in other, overall expanding the number of TF binding sites.

Previous studies performed in Deininger and Boeke laboratories^{11,12} have shown that intronic L1s could act as molecular “rheostats” by inhibiting elongation of the RNA pol II and causing premature transcriptional termination and cryptic polyadenylation within L1 ORF1 and ORF2 sequences of the host gene transcripts. It was shown that L1 sense strand, containing 19 cryptic polyadenylation sites (vs. two of antisense strand) with respect to gene transcription and having A-rich bias (~40%), had a stronger effect than antisense strand. Earlier, it was suggested that evolutionary pressure acting against premature polyadenylation within the L1 sense strand, has contributed to the selection of L1 antisense strand, as this strand occurred about twice more frequently.¹³ Another mechanism, by which intronic L1s could interfere with normal gene expression involves exonization. Many intronic L1s contain functional donor and acceptor splice sites (e.g. positions 97 and 791/979 in sense strand, respectively) which in combination with intronic sequences could provide alternative transcripts.¹⁴ When translated, these transcripts, containing only a partial ORF sequences, could produce truncated proteins and thus may alter the normal gene expression. Similarly to L1s, Alus contain many splice sites, either present in consensus sequences, e.g. seven splice donor and 12 acceptor sites in antisense strand and only a few in sense strand, or created by mutations.¹⁵ Therefore, exonization of Alus could affect alternative splicing of the host gene. Indeed, about 5% of alternatively spliced exons are derived from Alu.¹⁵ Also, in some cases a single point mutation in Alu sequence could shift from alternative to constitutive splicing resulting in human disease.¹⁶

We have recently demonstrated that intronic L1s and possibly other retroelements in tandem orientation with respect to gene transcription could cause transcriptional interference (TI) of the host gene by inducing intron retention and forcing exonization and cryptic polyadenylation

(Fig. 1).¹⁷ Differently from previous studies,^{11,12} these effects were observed upstream to L1, depended on the activity of L1 sense and antisense (ASP) promoters^{18,19} and/or TF binding, distance between the observed TI effect and L1, and the presence or absence of cryptic splice/polyadenylation signals. It was proposed that intronic L1s could interfere with the elongation of RNA pol II by forcing it to slow down or dissociate from the template giving rise to prematurely terminated transcripts. While the RNA pol II complex bound to L1 sense promoter would act as a “sitting duck,” another RNA pol II complex operating on the L1 antisense strand (ASP)^{19,20} would cause head-on-collision with the host RNA pol II according to the mechanisms described for TI.²¹ Consistent with the latter, correlation between the L1 ASP activity and TI effect was found for nine genes in different human tissues (our unpublished results). We predicted that a minimum of 100 L1 copies from about one thousand full-length intronic tandemly oriented L1s may be involved in the TI of the host genes.¹⁷ Our preliminary bioinformatic analysis also revealed that from the total of 55 Alus analyzed in 123 genes, 39 tandemly orientated intronic Alus possibly caused intron retention in 31 cases and exonization in seven cases.¹⁷ These (39) Alus had the following age distribution: AluJ-23%, AluS-67% and AluY-10%. The remaining Alus (15) in convergent orientation showed seven and one events of intron retention and exonization, respectively. This result shows that tandemly arranged Alus, like L1s, are about five times more efficient in causing TI than convergent Alus. The exact reason for this strand bias is not known. We suspect that TF binding and nucleosomal structure may be somehow involved (see below). Interestingly, seven of the 27 single exon protein-coding nested genes²² in tandem orientation showed effects on TI, while only one of the 23 convergently oriented similar genes was effective, suggesting that orientation bias is characteristic also for protein-coding genes.¹⁷ It remains to be shown how many intronic Alus could potentially induce TI effects in the human genome. Despite the lack of

orientational bias,¹³ the extremely large number of intronic Alus and their high TF binding potential⁹ suggests that hundreds, if not thousands of Alus could participate in the TI of the host genes.

Sela et al.²³ suggested that insertion of the retroelements into introns can lead to activation of certain alternatively spliced exons. According to their definition, also originally proposed by Parma et al.,²⁴ exonization is a creation of new exon as a result of mutation(s) in intronic sequence. In evolutionary terms, it is believed that this important driving force could enhance the divergence and speciation. However, the question here is: Do we only need a point mutation or mutations to get the intronic sequences exonized? We proposed that retroelements could act as roadblocks or “speed bumps” by slowing down the elongating RNA pol II and forcing its action to be completed by transcription-coupled splicing and termination factors/processes (Fig. 1).¹⁷ Therefore, retroelements not only reduce the elongation speed and increase RNA pol II pausing, but also force exonization and polyadenylation at cryptic splice sites and polyadenylation signals. All these data are in accordance with the kinetic model of transcription originally proposed by Eperon et al.²⁵ and subsequently modified by Kornblihtt.²⁶ According to this model, the use of alternative splice sites depends on the elongation rate of RNA pol II. Under high elongation speed of RNA pol II and in the absence of internal stops, only strong acceptor splice sites are used and exons with weak splice sites are skipped. However, when elongation speed is reduced or inhibited, e.g., with mutant form of RNA pol II or artificial pause sites, then the weak acceptor splice sites become available and can be used.

The kinetics of RNA pol II elongation is influenced by nucleosome occupation, histone code and the action of chromatin remodelling factors.²⁷ With exons enriched in nucleosome structures because of their matching sizes (both ~150 nt) and higher GC content, compared with surrounding introns, RNA pol II has higher binding level toward exons, determined from chromatin immunoprecipitation experiments.²⁸ Therefore, nucleosomes could serve as transcription “speed bumps,”

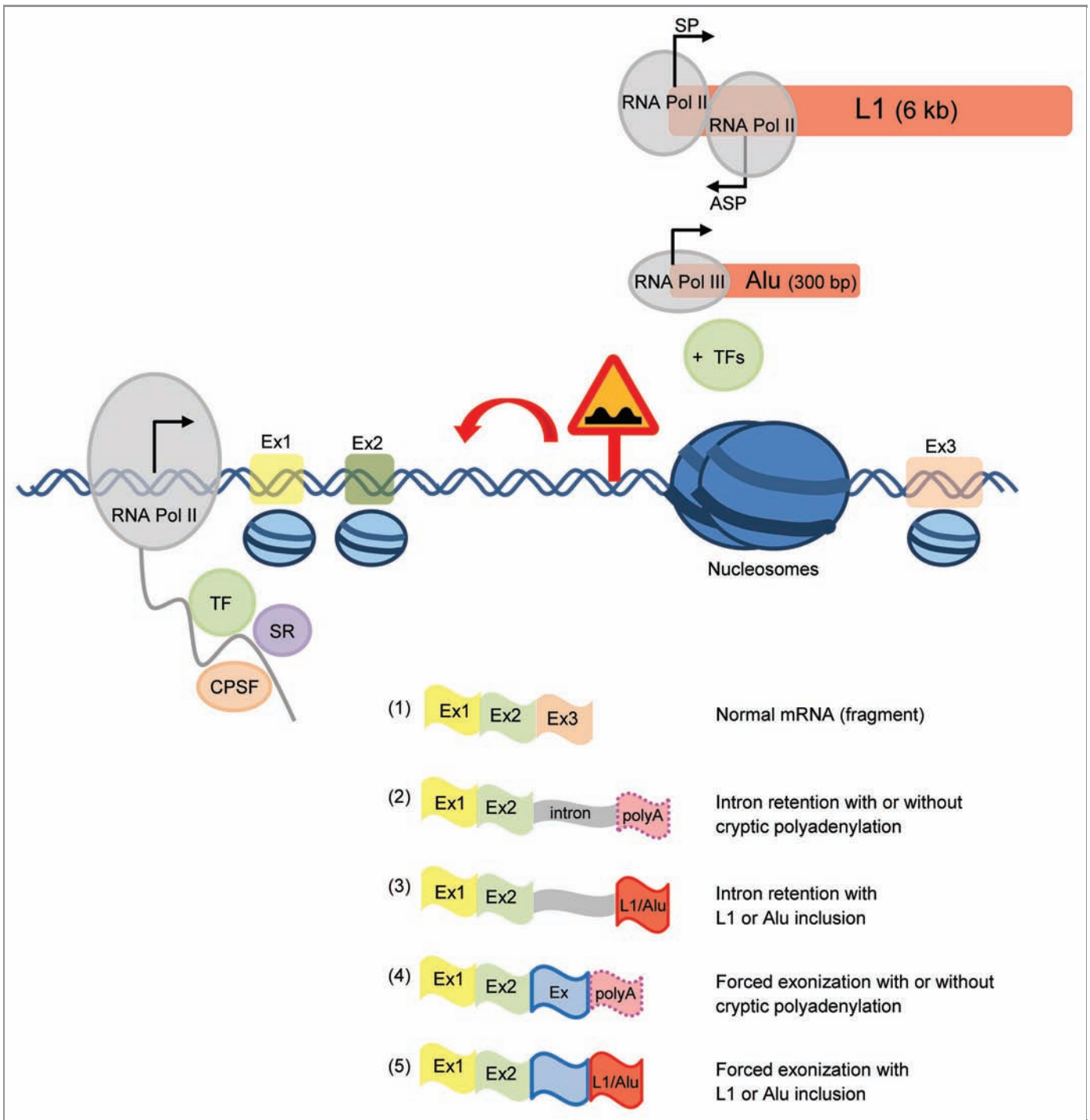


Figure 1. Transcriptional interference induced by human retroelements. L1 and Alu positioned in the nucleosomal structure located in the intron of the hypothetical gene fragment act not only as “speed bumps” for RNA pol II, but also interfere with the host gene transcription by causing intron retention, forcing exonization and cryptic polyadenylation in their upstream region (characteristic structures 2 and 4 are shown at bottom). Also, in some instances, intron retention and exonization could include part of the retroelement (structures 3 and 5, respectively) and prematurely terminated transcripts may or may not contain poly(A) tail (structures 2 and 4). The observed TI effects depend on the L1 SP-ASP and Alu transcriptional activity, TF binding and the presence or absence of cryptic splice sites and polyadenylation signals upstream to or within the retroelements. TI effect may be also influenced by the nucleosome occupation in the L1 or Alu region. For the sake of simplicity, only two nucleosomes aligned to the 5’ region of the L1 and Alu are shown. Exons 1–3 are shown as colored boxes and their preferential location in nucleosomal regions is marked with small nucleosomes. Red arrow refers to the position, where TI effects occur. For each RNA pol II complex direction of transcription is shown with arrow.

regulating RNA pol II traffic. Also, in support of this idea, Alexander et al.²⁹ have demonstrated transient accumulation of RNA pol II around 3' ends of introns, which in agreement with their experimental data was required for cotranscriptional splicing. It is not clear to what extent intronic retroelements, their transcriptional activity and TF binding are involved in the formation or positioning of nucleosomes. Interestingly, the high level of nucleosome occupancy, visualized by UCSC Genome Browser, was predicted for L1 promoter region (~0.9 kb) and Alus.^{30,31} Both these elements have high GC content (~60–65%). However, only

for Alus, but not for L1s, the tendency to form nucleosome structure has been experimentally proved by using DNase footprinting combined with primer extension³² and micrococcal digestion combined with sequencing.³³ The potential role of L1 or Alu nucleosomal structure in TI has not been studied. Recently, an interesting connection between TI and nucleosome occupancy was revealed by Thebault et al.³⁴ They showed that nucleosome deposition initiated by transcription from the upstream non-coding RNA gene (SRG1) mediates repression of the downstream tandemly oriented protein-coding gene (SER3) by a promoter

occlusion mechanism.²¹ It remains to be shown whether intronic retroelements are involved in the nucleosome occupation or positioning by similar or other mechanisms.

Acknowledgments

We apologize to authors whose work was not cited here because of space limitations. We thank anonymous reviewer for detailed comments on the manuscript. This work was supported by grants from the Estonian Science Foundation (ETF8381) and the Estonian Ministry of Education and Research (SF0140143s08).

References

- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001; 409:860-921; PMID:11237011; <http://dx.doi.org/10.1038/35057062>
- Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, et al. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* 2003; 100:5280-5; PMID:12682288; <http://dx.doi.org/10.1073/pnas.0831042100>
- Khan H, Smit A, Boissinot S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res* 2005; 16:78-87; PMID:16344559; <http://dx.doi.org/10.1101/gr.4001406>
- Skowronski J, Fanning TG, Singer MF. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol* 1988; 8:1385-97; PMID:2454389
- Rangwala SH, Zhang L, Kazazian HH, Jr. Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biol* 2009; 10:R100; PMID:19772661; <http://dx.doi.org/10.1186/gb-2009-10-9-r100>
- Faulkner GJ, Kimura Y, Daub CO, Wani S, Plesky C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 2009; 41:563-71; PMID:19377475; <http://dx.doi.org/10.1038/ng.368>
- Ostertag EM, Kazazian HH, Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 2001; 35:501-38; PMID:11700292; <http://dx.doi.org/10.1146/annurev.genet.35.102401.091032>
- Shankar R, Grover D, Brahmachari SK, Mukerji M. Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependant mobile Alu elements. *BMC Evol Biol* 2004; 4:37; PMID:15461819; <http://dx.doi.org/10.1186/1471-2148-4-37>
- Grover D, Kannan K, Brahmachari SK, Mukerji M. ALU-ring elements in the primate genomes. *Genetica* 2005; 124:273-89; PMID:16134339; <http://dx.doi.org/10.1007/s10709-005-3086-8>
- Polak P, Domany E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 2006; 7:133; PMID:16740159; <http://dx.doi.org/10.1186/1471-2164-7-133>
- Perepelitsa-Belancio V, Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet* 2003; 35:363-6; PMID:14625551; <http://dx.doi.org/10.1038/ng1269>
- Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 2004; 429:268-74; PMID:15152245; <http://dx.doi.org/10.1038/nature02536>
- Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 1999; 9:657-63; PMID:10607616; [http://dx.doi.org/10.1016/S0959-437X\(99\)00031-3](http://dx.doi.org/10.1016/S0959-437X(99)00031-3)
- Belancio VP, Hedges DJ, Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res* 2006; 34:1512-21; PMID:16554555; <http://dx.doi.org/10.1093/nar/gkl027>
- Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. *Genome Res* 2002; 12:1060-7; PMID:12097342; <http://dx.doi.org/10.1101/gr.229302>
- Deininger PL, Batzer MA. Alu repeats and human disease. *Mol Genet Metab* 1999; 67:183-93; PMID:10381326; <http://dx.doi.org/10.1006/mgme.1999.2864>
- Kaer K, Branovets J, Hallikma A, Nigumann P, Speek M. Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. *PLoS One* 2011; 6:e26099; PMID:22022525; <http://dx.doi.org/10.1371/journal.pone.0026099>
- Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol* 1990; 10:6718-29; PMID:1701022
- Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* 2001; 21:1973-85; PMID:11238933; <http://dx.doi.org/10.1128/MCB.21.6.1973-1985.2001>
- Nigumann P, Redik K, Mätlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* 2002; 79:628-34; PMID:11991712; <http://dx.doi.org/10.1006/geno.2002.6758>
- Shearwin KE, Callen BP, Egan JB. Transcriptional interference—a crash course. *Trends Genet* 2005; 21:339-45; PMID:15922833; <http://dx.doi.org/10.1016/j.tig.2005.04.009>
- Yu P, Ma D, Xu M. Nested genes in the human genome. *Genomics* 2005; 86:414-22; PMID:16084061; <http://dx.doi.org/10.1016/j.ygeno.2005.06.008>
- Sela N, Mersch B, Hottz-Wagenblatt A, Ast G. Characteristics of transposable element exonization within human and mouse. *PLoS One* 2010; 5:e10907; PMID:20532223; <http://dx.doi.org/10.1371/journal.pone.0010907>
- Parma J, Christophe D, Pohl V, Vassart G. Structural organization of the 5' region of the thyroglobulin gene. Evidence for intron loss and "exonization" during evolution. *J Mol Biol* 1987; 196:769-79; PMID:3681978; [http://dx.doi.org/10.1016/0022-2836\(87\)90403-7](http://dx.doi.org/10.1016/0022-2836(87)90403-7)
- Eperon LP, Graham IR, Griffiths AD, Eperon IC. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* 1988; 54:393-401; PMID:2840206; [http://dx.doi.org/10.1016/0092-8674\(88\)90202-4](http://dx.doi.org/10.1016/0092-8674(88)90202-4)
- Kornbliht AR. Chromatin, transcript elongation and alternative splicing. *Nat Struct Mol Biol* 2006; 13:5-7; PMID:16395314; <http://dx.doi.org/10.1038/nsmb0106-5>
- Kornbliht AR, Schor IE, Allo M, Blencowe BJ. When chromatin meets splicing. *Nat Struct Mol Biol* 2009; 16:902-3; PMID:19739285; <http://dx.doi.org/10.1038/nsmb0909-902>
- Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 2009; 16:990-5; PMID:19684600; <http://dx.doi.org/10.1038/nsmb.1659>
- Alexander RD, Innocente SA, Barras JD, Beggs JD. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* 2010; 40:582-93; PMID:21095588; <http://dx.doi.org/10.1016/j.molcel.2010.11.005>
- Gupta S, Dennis J, Thurman RE, Kingston R, Stamatoyannopoulos JA, Noble WS. Predicting human nucleosome occupancy from primary sequence. *PLoS Comput Biol* 2008; 4:e1000134; PMID:18725940; <http://dx.doi.org/10.1371/journal.pcbi.1000134>
- Levitsky VG, Podkolodnaya OA, Kolchanov NA, Podkolodny NL. Nucleosome formation potential of exons, introns, and Alu repeats. *Bioinformatics* 2001; 17:1062-4; PMID:11724736; <http://dx.doi.org/10.1093/bioinformatics/17.11.1062>
- Englander EW, Howard BH. Nucleosome positioning by human Alu elements in chromatin. *J Biol Chem* 1995; 270:10091-6; PMID:7730313; <http://dx.doi.org/10.1074/jbc.270.17.10091>
- Tanaka Y, Yamashita R, Suzuki Y, Nakai K. Effects of Alu elements on global nucleosome positioning in the human genome. *BMC Genomics* 2010; 11:309; PMID:20478020; <http://dx.doi.org/10.1186/1471-2164-11-309>
- Thebault P, Boutin G, Bhat W, Rufiange A, Martens J, Nourani A. Transcription regulation by the non-coding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. *Mol Cell Biol* 2011; 31:1288-300; PMID:21220514; <http://dx.doi.org/10.1128/MCB.01083-10>