# When chromatin meets splicing

Alberto R Kornblihtt, Ignacio E Schor, Mariano Allo & Benjamin J Blencowe

**Using bioinformatics analysis of previously published global genome deep-sequencing data, two papers now show that DNA sequences associated with nucleosomes are preferentially located in exons. The correlation between nucleosome distribution and the exon-intron organization of genes may have a key role in exon recognition at the pre-mRNA level during co-transcriptional splicing, consistent with previous findings indicating chromatin-mediated regulation of alternative splicing.**



Chromatin and mRNA splicing are fields of modern molecular biology that, at first sight, have been following independent paths for many years. However, a deeper look shows that the first link between them was proposed almost 17 years ago. With early sequence data from human and mouse genes, Trifonov and colleagues[1] suggested a connection between exon-intron architecture and nucleosome positioning, on the basis of the periodic distribution of splice sites. However, because at that time the splicing process was thought to be independent of transcription, no connection between chromatin structure and the actual process of splicing was made. Two reports[2,3] on pages 990 and 996 of this issue from the Ast and Guigó laboratories now independently reveal evidence for cross-talk between chromatin structure and exon-intron architecture. These studies build on growing evidence for extensive coupling between transcription and pre-mRNA processing and suggest that the recognition of exons by the splicing machinery might need 'a little help from a chromatin friend' that, surprisingly, tends to mark exons specifically at the DNA level.

There is a common misconception that exons almost exclusively form the protein-coding regions of eukaryotic genes. In fact, many protein-coding genes contain exons whose sequences are never translated into protein, such as those located in 5′ and 3′ untranslated regions. Furthermore, noncoding RNA genes, including some ribosomal genes, contain exons. Therefore, an exon can be defined more generally as a gene segment that remains represented in the mature RNA after the process of splicing, independently of its protein-coding capacity. It is therefore not surprising that protein-coding capacity does not itself have a substantial role as an evolutionary driving force in shaping exon recognition. A striking observation that does, however, seem to reflect a strong evolutionary driving force underlying exon recognition in higher eukaryotic protein-coding genes is that they are usually short and of relatively uniform length (~140–150 nt on average). This is in contrast to introns, which are generally much longer (thousands of nucleotides) and of more variable length. What are the selective forces that influence exon length bias, and what drives exon recognition?

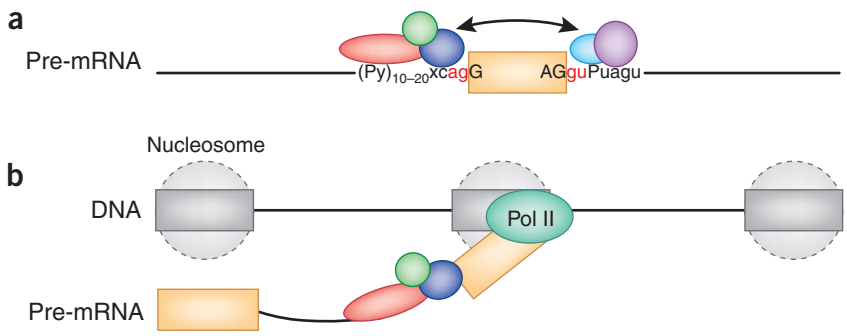


**Figure 1** Possible mechanisms for exon definition. (**a**) Exon definition is achieved at the pre-mRNA level. Spliceosomal and auxiliary factors are recruited to the splice sites flanking an exon on the mRNA precursor. Direct and indirect interactions between the 3′ and 5′ complexes favor exon recognition and splicing and exert selective pressure for a conserved exon length of 140–150 nt. Pu, purine; Py, pyrimidine. (**b**) Nucleosomes (broken circles) are preferentially bound to exons, whereas introns are mostly devoid of nucleosomes. Exons are therefore marked at the DNA level by nucleosome positioning, which may act as 'speed bumps' for RNA polymerase II, helping in the co-transcriptional recruitment of splicing factors to the nascent pre-mRNA and improving exon definition. As nucleosomes accommodate DNA stretches of approximately 147 nt, their preferential location on exons may act as the selective pressure factor for the conservation in exon length.

The boundaries of each internal exon are the 3′ and 5′ splice sites. These consensus sequences (**Fig. 1**) are the target sites for dozens of spliceosomal components and auxiliary factors that at the pre-mRNA level promote exon recognition and ligation, with the concomitant precise excision of flanking introns. A mechanism that acts at the pre-mRNA level, known as 'exon definition', was proposed by Berget and colleagues[4] to explain how short exons are accurately recognized in a 'sea of introns' by the splicing machinery. The 140–150-nt exon length might represent an optimal distance for protein-protein interactions between the factors assembling at the 5′ and 3′ splice sites (**Fig. 1a**). However, because the complete set of *cis*-acting sequence elements and *trans*-acting factors that facilitate the precise recognition of exon 5′ and 3′ splice sites during splicing has remained poorly understood, other explanations for the length bias of exons in addition to a more complete picture of the code(s) that govern exon recognition have been highly sought after.

Schwartz *et al.*[2] and Tilgner *et al.*[3] made use of published experimental data on nucleosome positioning within the human genome derived from deep sequencing of DNA fragments attached to mononucleosomes[5], obtained from chromatin digestion with micrococcal nuclease. They then performed a thorough bioinformatics analysis to find out whether there was differential distribution of nucleosome-associated sequences across exons and introns. Both groups

*Alberto R. Kornblihtt, Ignacio E. Schor and Mariano Allo are at the Departamento de Fisiología y Biología Molecular, LFBM and IFIByNE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina. Benjamin J. Blencowe is at the Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada. e-mail: ark@fbmc.fcen.uba.ar or b.blencowe@utoronto.ca*

found a distinct peak of nucleosome occupancy within exons that is paralleled by nucleosome depletion from introns. This difference in distribution seems to be independent of the transcriptional activity of the genes analyzed. Moreover, Schwartz *et al.*[2] could essentially recapitulate their observations by using, instead of deep-sequencing data, computationally based predictions of nucleosome positioning that rely entirely on sequence features reported to favor or disfavor nucleosome assembly[6]. The favoring sequences were concentrated on exons, whereas the disfavoring sequences were depleted in exons but enriched in the 50-nt intronic regions immediately preceding and succeeding exons, as if these nucleosome-free regions were marking the beginnings and the ends of exons at the DNA level. This pattern is suspiciously coincident with the 'splicing code', previously thought to be an exclusive property of pre-mRNA. In fact, as the authors note, consistent with Trifonov and co-worker's formative proposal, the mean length of exons is strikingly similar to the 147-nt length of DNA that is required to wrap entirely around a mononucleosome. These findings provide an alternative explanation for the average size of exons, because a length of 140–150 nt would accommodate exon wrapping around a nucleosome, perhaps facilitating exon recognition (**Fig. 1b**).

Additionally, Tilgner *et al.*[2] showed that, unlike canonical exons, nucleosomes are depleted from pseudoexons, that is, intronic regions that are flanked by strong splice sites but are not included in mature mRNA. More excitingly, there seems to be an inverse correlation between nucleosome occupancy and splice site strength of exons: the weaker the splice sites of an exon, the higher the nucleosome occupancy, as if the latter were contributing to promote exon inclusion despite the presence of weak splice sites. This situation is mirrored in pseudoexons, where, despite having strong splice sites, inclusion is not favored, perhaps in part because of low nucleosome occupancy. However, Schwartz *et al.* do not entirely agree with these observations, because their data reveal a positive correlation between nucleosome occupancy and the degree of exon inclusion when analyzing three categories: alternative exons with less than 50% inclusion, those with more than 50% inclusion and constitutive exons.

The pioneering ideas of Trifonov and colleagues were that nucleosomes were associated with intron-exon junctions to protect coding sequences and splices sites from mutation[1]. However, the availability of large experimental data provided by high-throughput sequencing allows deeper insight into the profound biological implications that may ultimately change our conception of the earliest steps responsible for exon recognition during the course of gene expression in eukaryotes. Highly relevant to the results of Schwartz *et al.*[2] and Tilgner *et al.*[3], recent work performed in *Caenorhabditis elegans*[7] demonstrated that the chromatin modification trimethylated histone H3 lysine 36 (H3K36me3), which is associated with RNA polymerase II (Pol II) elongation, was preferentially enriched on exons. Upon further analysis of this and other epigenetic marks (H3K79me1, H4K20me1 and H2BK5me1) in humans, both Schwartz *et al.*[2] and Tilgner *et al.*[3] conclude that the H3K36me3 peak, rather than being unique to exons, actually reflects underlying nucleosome occupancy. These findings further reveal that the preferential occupancy of nucleosomes on exons relative to introns is a conserved feature from worms to humans.

Exons are known to have a slightly higher GC content than introns[8]. In parallel, nucleosomes are preferentially assembled on GC-rich sequences. This raises the question of whether the nucleosome 'marking' of exons is just the mere circumstantial reflection of two independent events. Schwartz *et al.*[2] essentially rule out this possibility by showing that low nucleosome occupancy prevails in intronic regions with GC contents that are much higher than those of exons displaying high nucleosome occupancy. Similar conclusions were reported by Tilgner *et al.*[3], who showed that true exons have a substantially higher nucleosome occupancy than pseudoexons with the same GC content, pointing to splicing as a contributing force in determining nucleosome occupancy over exons. Consequently, the higher GC content of exons might be the result of not only selective pressure exerted by the GC bias in codon usage during translation, but also of that exerted by nucleosome occupancy. What is the advantage of the preferential association of nucleosomes with exons relative to introns? Answers to this question may be best addressed in light of what is known about co-transcriptional splicing. Many different approaches, both *in vivo* and *in vitro*, have revealed mechanistic coupling between transcription and pre-mRNA processing and have illustrated how the efficiency and fidelity of the splicing process can be facilitated by associations with the Pol II transcriptional machinery[9]. Moreover, the mechanistic connection between transcription and splicing has been shown to influence alternative splicing choices as well[10], both by affecting the recruitment of splicing factors and by a kinetic effect whereby decreased elongation rate or pausing of Pol II promotes the recognition of weak splice sites.

The influence of transcription on pre-mRNA processing has prompted many groups to search for possible connections between chromatin structure and splicing. Indeed, chromatin structure and histone modifications were shown to affect both constitutive and alternative splicing, whether by recruitment of splicing factors[11] or by modulating Pol II elongation[12–14]. Recent evidence demonstrated that reduction of elongation by chromatin-remodeling factors such as SWI/SNF[12], or by small interfering RNAs (siRNAs) that create local heterochromatin marks[13], affects alternative splicing in an opposite manner to histone hyperacetylation agents that promote elongation through chromatin opening[14]. In this context, nucleosome positioning on exons may function to create roadblocks or 'speed bumps' for Pol II elongation that increase transit time, allowing co-transcriptional recognition of newly synthesized splice signals by splicing factors in the nascent pre-mRNA. In agreement with these ideas, a recent study using optical tweezers to follow individual Pol II complexes as they transcribe nucleosomal DNA strongly supports the 'bumpy' nature of Pol II elongation, where the nucleosome behaves as a fluctuating barrier that locally increases pause density, slows pause recovery and reduces the apparent pause-free velocity of Pol II[15]. This may represent a basal level of control upon which specific covalent histone modifications could add an additional level of regulation. Finally, the possibility of dynamic chromatin remodeling *in vivo* implies a likely role for nucleosome positioning in the regulation of alternative splicing, an interesting possibility that has yet to be experimentally tested.


1. Beckmann, J.S. & Trifonov, E.N. *Proc. Natl. Acad. Sci. USA* **88**, 2380–2383 (1991).
2. Schwartz, S., Meshorer, E. & Ast, G. *Nat. Struct. Mol. Biol.* **16**, 990–995 (2009).
3. Tligner, H. *et al. Nat. Struct. Mol. Biol.* **16**, 996–1001 (2009).
4. Berget, S.M. *J. Biol. Chem.* **270**, 2411–2414 (1995).
5. Schones, D.E. *et al. Cell* **132**, 887–898 (2008).
6. Kaplan, N. *et al. Nature* **458**, 362–366 (2009).
7. Kolasinska-Zwierz, P. *et al. Nat. Genet.* **41**, 376–381 (2009).
8. Zhu, L. *et al. BMC Genomics* **10**, 47 (2009).
9. Bentley, D.L. *Curr. Opin. Cell Biol.* **17**, 251–256 (2005).
10. Kornblihtt, A.R. *Adv. Exp. Med. Biol.* **623**, 175–189 (2007).
11. Sims, R.J. III *et al. Mol. Cell* **28**, 665–676 (2007).
12. Batsché, E., Yaniv, M. & Muchardt, C. *Nat. Struct. Mol. Biol.* **13**, 22–29 (2006).
13. Alló, M. *et al. Nat. Struct. Mol. Biol.* **16**, 717–724 (2009).
14. Schor, I.E., Rascovan, N., Pelisch, F., Allo, M. & Kornblihtt, A.R. *Proc. Natl. Acad. Sci. USA* **106**, 4325–4330 (2009).
15. Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M. & Bustamante, C. *Science* **325**, 626–628 (2009).