The processing of pre-mRNAs, including the selection of polyadenylation sites, is influenced by the surrounding chromatin context. We review here recent studies in Arabidopsis thaliana highlighting the intricate and reciprocal interplay between chromatin state and RNA processing. The studies have revealed that transcription can be influenced by the presence, in gene introns, of combination of epigenetic marks typical of heterochromatin. New factors binding to these marks have been identified and shown to play key roles in controlling the use of polyadenylation sites and processing of functional mRNAs. Concomitantly, several proteins of both the splicing and the polyadenylation machineries are also emerging as regulators of DNA methylation patterns and chromatin silencing.

Chromatin marks modulate the use of polyadenylation signals

In plants, the Jumonji C (JmjC) domain-containing protein INCREASE IN BONSAI METHYLATION1 (IBM1) is essential to prevent ectopic epigenetic modifications at protein-coding genes [1–3]. IBM1 is a histone demethylase that specifically removes mono-methylation and di-methylation of the lysine 9 of histone H3 tails (H3K9me1,2). In ibm1 mutants, genes not only accumulate H3K9me2 but also cytosine methylation at CHG (where H = A, T or C) sequence contexts. This occurs as the DNA methyltransferase CHROMOMETHYLASE3 (CMT3) that methylates CHG positions is recruited to chromatin through its binding to H3K9me2 marks. In combination with others, both H3K9me2 and CHG methylation are two general hallmarks of highly silenced regions such as transposons. The function of IBM1 in maintaining genes free of H3K9me2 and CHG methylation is essential for normal development and ibm1 mutants show severe and pleiotropic phenotypes after only a few generations. How the switch in epigenetic marks occurring in ibm1 mutants impacts gene expression is unclear. Heterochromatin-associated epigenetic marks such as CHG DNA methylation and H3K9me2 can potentially interfere with transcription and/or disturb RNA processing; however, their presence is now shown to be also required for proper RNA processing at loci including the IBM1 gene itself.

The IBM1 gene encodes two different transcripts (Figure 1). Only the longest one is functional and can complement an ibm1 mutant [4]. The production of full-length IBM1 mRNAs requires splicing of an unusually large intron of about 2 kb located in the middle of the gene, and which harbors an at least 700-bp long heterochromatin domain associated with dense CG and CHG methylation and H3K9me2. Mutants that impact CG (i.e. met1) or CHG methylation (i.e. cmt3 or kyp) in this intronic region do not accumulate normal levels of long IBM1 transcripts, implying that both types of DNA methylation are necessary to correctly process the functional transcript. Consequently, downregulation of the functional IBM1 mRNA in met1 mutants [4*] results in the ectopic appearance of both H3K9me2 and CHG methylation at several thousand genes [5,6].
We [7] and others [8**, 9] found that mutants of the IBM2/ ASII/SG1 gene, hereafter called IBM2, are morphologically very similar to *ibm1* mutants, and the two mutants share the same molecular phenotype with thousands of genomic regions invaded by heterochromatic silencing marks. However, IBM1 and IBM2 have no particular impact on transposable elements located in constitutive heterochromatin. IBM2 encodes a protein of unknown function containing a Bromo-Adjacent Homology (BAH) domain shared by several chromatin regulators and an RNA-Recognition Motif (RRM). In *ibm2* mutants, as in *met1*, the long intron of *IBM1* is not correctly processed, and consequently accumulation of full-length functional *IBM1* mRNAs is drastically compromised. This accounts almost exclusively for the observed *ibm2* phenotype as it can be complemented by the introduction of a transgene producing the long version of *IBM1* transcripts. Complementation only occurs when the long intron was omitted from the *IBM1* transgene, implying that IBM2 functions by promoting transcription of *IBM1* over the intronic heterochromatin domain in the wild-type (Figure 1). IBM2 targets additional genes that contain long introns associated with DNA methylation and chromatin immunoprecipitation assays confirmed that IBM2 physically interacts with these regions. In the absence of IBM2, mRNAs corresponding to these genes are shorter, precociously ending in their large introns because proximal polyadenylation sites seem to be favored. This is raising the possibility that IBM2 might function by promoting the use of distal polyadenylation sites over proximal ones located in large introns. Using a new method allowing sequencing of single mRNA molecules and direct identification of mRNA cleavage and polyadenylation sites, a recent study has demonstrated that use of alternative polyadenylation sites within introns of pre-mRNAs is rare, FCA and FPA (see below) being two exceptions [10]. Combined with the fact that large introns (>2 kb) containing transposon-like sequences are also uncommon, at least in *Arabidopsis* [8**], the role of IBM2 in controlling polyadenylation is likely limited to few, very specific loci. Accordingly, only four IBM2 target genes have been characterized so far. First, *IBM1* that controls chromatin structure to ensure that highly transcribed genes remain free of silencing marks. Second, *AT1G11270* that shows heat responsiveness due to the insertion of a retrotransposon of the ONSEN family in its second intron [11]. Third, a gene encoding a photosystem II reaction center PsbP family protein. And finally, a disease-resistance gene, *RPP7*, which contains a COPIA retrotransposon in its first large intron. Although all these genes contain intrinsic heterochromatin domains, their nature is different, including a LINE element, two COPIA retrotransposons, and in the case of *IBM1*, a sequence homologous to both chloroplast and mitochondrial genes. Therefore, IBM2 might recognize its target genes through a specific chromatin signature rather than a precise DNA sequence.

Interestingly, polyadenylation of *RPP7* is also controlled by another protein, ENHANCED DOWNY MILDEW2
(EDM2) that contains several zinc-finger domains and a region similar to the active domains of methyltransferases [12**]. edm2 mutants were isolated from a genetic screen to identify genes that are essential for RPP7 function [13]. Similar to ibm2, edm2 mutants show impaired accumulation of functional RPP7 transcripts. Levels of H3K9me2 at the intronic COPIA retrotransposon are reduced in edm2 and this directly impacts the transcription of the full-length RPP7 transcript [14**]. In the wild-type, H3K9me2 EDM2-dependent marks promote the use of a distal polyadenylation site, thereby enhancing production of the long RPP7 functional transcript; while in edm2 mutants, a proximal polyadenylation site localized in the retrotransposon is used, leading to a shorter transcript (named ECL). In natural Arabidopsis accessions containing an RPP7 gene with no COPIA elements, the alternative polyadenylation of the transcript becomes EDM2-independent. The EDM2 protein can physically bind H3K9me2 EDM2-dependent marks, and unlike IBM2, EDM2 controls levels of H3K9me2 at certain transposons [12**]. The precise function of EDM2 remains to be determined. Not surprisingly, EDM2 also binds heterochromatic marks of IBM1 and controls the expression of the functional IBM1 long transcript by promoting the distal polyadenylation site [15]. Consequently, the similar phenotypic alterations observed in edm2, ibm1 and ibm2 mutants all result from the misregulation of IBM1.

A role for epigenetic marks in regulating alternative polyadenylation has also been revealed in the unique case of two mouse imprinted genes: H13 and Herc3 that host in their introns two genes (called retrogene) Mts2 and Nap15, respectively, sharing characteristics with retrotransposons [16,17]. When the retrogene are silenced by DNA methylation, a distal polyadenylation site is used to process the host genes. In contrast, the transcription of the unsilenced retrogenes favors a proximal polyadenylation site to the detriment of the host gene. Experimental evidence points toward a mechanism of direct competition for the transcription of the host and the retrogene transcripts, the latter one being transcribed and polyadenylated first when silencing marks are absent. In the case of both IBM2 and EDM2, the mechanism involved is co-transcriptional or post-transcriptional since the Pol II occupancy of their target genes remains similar between the wild-type and the corresponding ibm2 or edm2 mutants [5**,14**,15].

RNA 3’-processing factors are required for chromatin silencing

Plants control their flowering to ensure that they reproduce under favorable conditions. This complex control is extensively studied in Arabidopsis, and studies of the flowering repressor gene FLC have certainly been pioneer in revealing a link between RNA processing activities and chromatin regulation in gene silencing. Two RNA-binding proteins, FCA and FPA, are implicated in the regulation of alternative polyadenylation, thereby playing important roles in limiting intergenic transcription in the Arabidopsis genome [18]. FCA not only regulates its own expression by enhancing proximal polyadenylation but also controls expression of the major floral regulator gene FLC [19**]. Together with FPA, FCA represses FLC expression through alternative 3’-processing of the FLC antisense transcript termed COOLAIR [20,21]. FCA and FPA act to favor proximal polyadenylation of COOLAIR. The resulting short FLC antisense RNA prompts transcriptional downregulation of FLC through stimulating the activity of the FLD histone H3K4me1/2 demethylase in an unknown process, leading to a repressed chromatin state in the body of the FLC gene [19**]. Such a role for FCA and FPA in chromatin silencing is not restricted to the FLC gene and appears to be more widespread than may have been originally anticipated. Indeed, mutant alleles of these two genes were recovered in a genetic screen aimed at identifying further components required for RNA-mediated chromatin silencing [22]. FCA and FPA were revealed to be required for efficient trans-gene-induced chromatin silencing of the endogenous PDS gene, as well as silencing of several, mainly single-copy, endogenous sequences including some transposable elements. At certain targets, transcriptional reactivation in fca fpa double mutants was coupled to reduced DNA methylation, mainly in non-CG sequence contexts, suggesting for the first time a possible link between 3’-RNA processing and DNA methylation.

RNA-directed DNA methylation and splicing pathways are interconnected

The presence of non-CG methylation at a given DNA sequence, and of CHH methylation in particular, is often a signature of RNA-directed DNA methylation (RdDM) [23]. The efficient maintenance of CHH methylation requires the continuous production of 24-nt small interfering RNAs (siRNAs) that act as targeting molecules, although a siRNA-independent CHH methylation pathway involving CMT2 has recently been reported in Arabidopsis [24]. There is ample evidence that RdDM contributes to gene silencing in plants, in particular at transposable elements and other types of repeated DNA sequences [25,26]. This complex pathway is likely primed by the production of double-stranded RNAs through the action of RNA polymerase IV and the RNA-dependent RNA polymerase RDR2 [27]. The resulting dsRNAs are then cleaved by DICER-LIKE3 (DCL3) into 24-nt siRNAs that are subsequently loaded onto an ARGONAUTE protein (AGO4 or AGO6). siRNAs loaded onto AGO proteins then serve as sequence specific guides for the DNA methyltransferase DRM2 by pairing with nascent scaffold RNAs produced from target loci by the RNA polymerase Pol V or Pol II [28,29]. Some of the factors involved in RdDM, including RDR2, DCL3 and AGO4, as well as siRNAs were shown to colocalize with nucleolus-associated organelles named Cajal bodies
These bodies, described in 1903 by Ramón y Cajal, appear to be universal nuclear organelles that have key roles in nuclear physiology. Among those are processing of ribosomal RNA and 3'-end histone mRNA, and the modification and assembly of U small nuclear ribonucleoproteins (snRNPs), some of which eventually form the RNA splicing machinery also known as the spliceosome. Of the seven U snRNPs, only five are involved in the spliceosome (U1, U2, U5 and the U4/U6 complex) that is additionally composed of numerous non-snRNP splicing factors [31,32]. The simultaneous presence of RDR2, DCL3, AGO4 and siRNAs in Cajal bodies suggests that these nuclear compartments also function as centers for siRNA processing and AGO4 effector complex assembly [30,33].

Therefore, given this spatial proximity, it may not sound inappropriate to expect some type of functional relationship between RdDM and RNA splicing components. A series of recent reports provide experimental support for a role of RNA splicing factors in RdDM. The DRM2 DNA methyltransferase involved in the RdDM pathway also performs all de novo DNA methylation that occurs when some naive/unmethylated DNA sequences are stably introduced in the Arabidopsis genome [34]. Transformation of Arabidopsis with a transgene containing the FWA gene has been extensively used as an assay for de novo DNA methylation. Wild-type plants transformed with FWA efficiently methylate repeats encompassing the promoter of the FWA transgene leading to silencing, whereas dnm2 mutants do not initiate DNA methylation allowing FWA expression, which induces a late flowering phenotype. In an effort to identify additional factors required for the establishment of DNA methylation, Ausin and colleagues screened a collection of pre-selected T-DNA insertion lines using the FWA transgene silencing reporter system [35**]. A late flowering phenotype was observed when plants mutant for ARGININE/SERINE-RICH45 (SR45) were transformed with the FWA transgene. SR45 contains an RRM and was previously shown to be a bona fide plant-specific splicing factor [36]. The sr45 mutant not only shows impaired de novo DNA methylation capacity at the FWA transgene, but also reduced non-CG methylation at several endogenous RdDM targets, as well as impaired accumulation of siRNAs associated with RdDM. Therefore, the SR45 splicing factor appears to be required for RdDM and is probably acting in this pathway before the production of siRNAs. Resembling FCA and FPA, SR45 was interestingly additionally shown to be required for proper silencing of FLC [35**,36], suggesting that SR45 role in RNA-mediated silencing extends beyond the RdDM pathway.

Last year, additional support for a role of the whole mRNA splicing machinery in RdDM was provided by studies from two other groups. Three mutants for different splicing factors were recovered from forward genetic screens designed to identify suppressors of RdDM-dependent silencing of a luciferase transgene. Mutants were recovered in genes encoding ZINC-FINGER AND OCRE DOMAIN-CONTAINING PROTEIN1 (ZOP1), a novel pre-mRNA splicing factor conserved in green algae and various angiosperms [37*]; STABILIZED1 (STA1), a U5 snRNP-associated protein required for both splicing and mRNA stability [38*,39]; and RNA-DIRECTED DNA METHYLATION16 (RDM16), a component of the U4/U6 snRNP protein complex conserved in euakaryotes [40*]. Similar to sr45, all three mutants show reduced non-CG methylation at several endogenous RdDM targets, including AxSN1. However, both the impact of these mutants on siRNA accumulation and the nuclear localization of the corresponding wild-type proteins suggest that these splicing factors act at different steps of the RdDM pathway. Indeed, mutants for SR45, ZOP1 and STA1, but not RDM16, affect siRNA abundance; whereas sta1 and rdm16, but not zop1, reduces Pol V transcripts accumulation suggesting that STA1 and RDM16 likely function in a later step of RdDM. ZOP1 can specifically associates with STA1 [37*], and the two factors colocalize with AGO4 at the Cajal body [37*,38*], while RDM16 protein is dispersed throughout the nucleoplasm [40*]. In support to a role of the whole splicing machinery in RdDM rather than anodic contribution of particular components, Zhang and colleagues showed that mutants of four other spliceosome factors are also defective to various extents for siRNA accumulation, DNA methylation and transcriptional silencing [37*]. Importantly, all these mutations do not significantly alter transcript accumulation and pre-mRNA splicing of known RdDM factors, indicating that their contribution to RdDM and RNA-mediated silencing is likely to be direct [37*,38*,40*]. However, the functional contribution of the spliceosome components into RdDM remains unclear. To this point, no interactions between splicing factors and RdDM components have been detected. Besides, many RdDM targets with altered DNA methylation or transcriptional silencing in mutants of these splicing factors are intronless sequences. Thus, it is plausible that it is more the RNA processing activities of these factors than the pre-mRNA splicing function itself that plays a role in RdDM. Somehow similar to the function of FCA/FPA in processing non-coding FLC transcripts, the splicing factors may interact with non-coding transcripts generated by either Pol V or Pol II and process these co-transcriptionally so they are able to act as scaffold RNAs in the later steps or RdDM. Some of these spliceosome components such as SR45 may also be required to process Pol IV transcripts and allow them to be recognized as a template by RDR2 at early steps of RdDM. Alternatively, as sr45 shares a very similar phenotype as dcl3 [35**], and making the parallel with the role of the core splicing factor SmD1 in RNAi in Drosophila [41], SR45 may interact with DCL3 and/or dsRNAs to allow optimal siRNA biogenesis.
Conclusions
Chromatin state and RNA processing meet at different levels from which certainly only a few aspects are currently known. The recent discovery of IBM2 and EDM2 raises interesting questions related to mechanisms of interplay between RNA polyadenylation and chromatin marks. The mode of action of proteins linking pre-RNAs and chromatin is little understood and the discovery of protein partners and complexes will certainly help resolve that issue. An evolutionary common scheme in which pre-mRNA splicing factors modulate RNA-mediated silencing processes, including RdDM is emerging from studies in fission yeast, Caenorhabditis elegans, Drosophila melanogaster and Arabidopsis [35*,37*,38*,40*,41–43]. Although it is likely that these factors act in these processes independently of their pre-mRNA splicing activities, the underlying molecular mechanisms remain elusive. Isolation of additional mutants through forward genetics and combination of genomic and biochemical approaches will undoubtedly soon provide a deeper understanding of the mode of action of the splicing machinery in the RdDM pathway and in RNA-mediated silencing in general. pre-mRNAs are processed while Pol II is still elongating transcripts, meaning that splicing and alternative polyadenylation are potentially competing. The relationship between the rate of elongation by the Pol II and pre-mRNA processing will also need to be addressed in plants. In animals, the control of alternative polyadenylation is emerging as an important contributor to the complexity of transcriptomes [44]. The discovery of underlying mechanisms of regulation in both plants and animals is thus becoming an important research priority.

Acknowledgements
We would like to thank Isabelle Gyi for critical reading of the manuscript and helpful comments. NB is supported by the ANR (Project 11-JS7V-0013). Work in OM’s laboratory is supported by the regional council of Auvergne and by the European Community’s Seventh Framework Programme (FP7, 2007/2013) through a Starting Independent Researcher grant from the ERC (257°=260742). OM is an EMBO Young Investigator.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest
•• of outstanding interest


This paper brings the first experimental evidence showing that heterochromatin marks located in the large intron of IBM1 play a crucial role in regulating the IBM1 transcription.


A landmark paper describing for the first time IBM2 and revealing a functional link with IBM1 in addition to three other IBM2 targets, one of them being the resistant gene RPP7.


This paper shows how EDM2 affects epigenetic marks at certain transposon locus.


A paper presenting the functional role of EDM2 (originally identified in 2007 by the same group; see [13]) in the alternative polyadenylation of RPP7). EDM2-dependent repressive marks of an intronic RPP7 retrotransposon are essential to promote the distal polyadenylation site and the functional transcript.


This paper shows that FCA functions with 3′-end processing factors to favor proximal FLC antisense transcript polyadenylation, which leads to
FLC silencing through promoting the activity of the FLD H3K4 histone
demethylase and associated changes in FLC chromatin structure.

20. Hornýk C, Terzi LC, Simpson GG: The open family protein FPA
controls alternative cleavage and polyadenylation of RNA. Dev

21. Ietswaart R, Wu Z, Dean C: Flowering time control: another
window to the connection between antisense RNA and

22. Baurle I, Smith L, Baulcombe DC, Dean C: Widespread role for
the flowering-time regulators FCA and FPA in RNA-mediated

23. Law JA, Jacobsen SE: Establishing, maintaining and modifying
DNA methylation patterns in plants and animals. Nat Rev Genet
2010, 11:204-220.

Thao K, Harmer SL, Zilberman D: The Arabidopsis nucleosome
remodeler DDM1 allows DNA methyltransferases to access


26. Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ: RNA-
mediated chromatin-based silencing in plants. Curr Opin Cell

27. Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasat-
olic L, Pikaard CS: In vitro transcription activities of Pol IV, Pol
V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA

V transcription guides ARGONAUTE4 to chromatin. Nat Genet
2009, 41:630-634.

transcription by RNA polymerase II coordinates Pol IV and Pol
V in siRNA-directed transcriptional gene silencing in

30. Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV,
Chan SW, Lagrange T, Pikaard CS, Jacobsen SE: An
ARGONAUTE4-containing nuclear processing center
colocalized with Cajal bodies in Arabidopsis thaliana. Cell

31. Wahl MC, Will CL, Lührmann R: The spliceosome: design

32. Morris GE: The Cajal body. Biochim Biophys Acta 2008,
1783:2108-2115.

33. Pontes O, Li CF, Costa Nunes P, Haag J, Ream T, Vittas A,
Jacobsen SE, Pikaard CS: The Arabidopsis chromatin-
modifying nuclear siRNA pathway involves a nucleolar RNA

34. Cao X, Jacobsen SE: Role of the Arabidopsis DRM
methyltransferases in de novo DNA methylation and gene

35. Ausin I, Greenberg MV, Li CF, Jacobsen SE: The splicing factor
SR45 affects the RNA-directed DNA methylation pathway in

First report showing the involvement of a spliceosome component in
siRNA accumulation and RdDM in Arabidopsis.

Regulation of plant developmental processes by a novel

37. Zhang CJ, Zhou JX, Liu J, Ma ZY, Zhang SW, Dou K, Huang HW,
RNA-directed DNA methylation and transcriptional silencing in
Arabidopsis. EMBO J 2013, 32:1128-1140.

See comments of [40*].

38. Dou K, Huang CF, Ma ZY, Zhang CJ, Zhou JX, Huang HW, Cai T,
Tang K, Zhu JK, He XJ: The PRP6-like splicing factor STA1 is
involved in RNA-directed DNA methylation by facilitating the
production of Pol V-dependent scaffold RNAs. Nucleic Acids
Res 2013, 41:8489-8502.

See comments of [40*].

39. Lee BH, Kapoor A, Zhu J, Zhu JK: STABILIZED1, a stress-
upregulated nuclear protein, is required for pre-mRNA
splicing, mRNA turnover, and stress tolerance in Arabidopsis.

40. Huang CF, Miki D, Tang K, Zhou HR, Zheng Z, Chen W, Ma ZY,
Yang L, Zhang H, Liu R et al.: A pre-mRNA-splicing factor is
required for RNA-directed DNA methylation in Arabidopsis.

Using forward genetics, this series of three papers [37*,38*,40*] identifies
three distinct pre-mRNA splicing factors as being required for proper
transcriptional silencing and RdDM.

41. Xiong XP, Kurthkoti K, Chang KY, Lichinchi G, De N,
Schneemann A, Macrae UJ, Rana TM, Perrimon N, Zhou R: Core
small nuclear ribonucleoprotein particle splicing factor SmD1
modulates RNA interference in Drosophila. Proc Natl Acad Sci

42. Bayne EH, Portoso M, Kagansky A, Kos-Braun IC, Urano T,
Ekwall K, Alves F, Rappaport J, Allshire RC: Splicing factors
facilitate RNAi-directed silencing in fission yeast. Science

43. Tabach Y, Billi AC, Hayes GD, Newman MA, Zuk O, Gabel H,
Kamath R, Yacoby K, Chapman B, Garcia SM et al.: Identification
of small RNA pathway genes using patterns of phylogenetic

44. Elkon R, Ugalde AP, Agami R: Alternative cleavage and
polyadenylation: extent, regulation and function. Nat Rev
Genet 2013, 14:496-506.