



INSTITUT DE FRANCE  
Académie des sciences

# *Comptes Rendus*

---

## *Biologies*

Martin Crespi

**Long non-coding RNAs reveal new regulatory mechanisms controlling gene expression**

Volume 345, issue 4 (2022), p. 15-39

Published online: 10 January 2023

Issue date: 16 February 2023

<https://doi.org/10.5802/crbio.106>

**Part of Special Issue:** Insights onto Plant Biology

**Guest editors:** Michel Delseny (Laboratoire Génome et Développement des plantes, UMR5096 CNRS-Université de Perpignan Via Domitia, Université de Perpignan 66860 Perpignan, France) and Georges Pelletier (Directeur de recherche émérite à l'Institut national de la recherche agronomique, Versailles, France)



This article is licensed under the  
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.  
<http://creativecommons.org/licenses/by/4.0/>



*Les Comptes Rendus. Biologies* sont membres du  
Centre Mersenne pour l'édition scientifique ouverte

[www.centre-mersenne.org](http://www.centre-mersenne.org)

e-ISSN : 1768-3238



---

Insights onto Plant Biology / *Regards sur la biologie végétale*

# Long non-coding RNAs reveal new regulatory mechanisms controlling gene expression

*Les longs ARN non codants révèlent de nouveaux mécanismes de régulation de l'expression des gènes*

Martin Crespi<sup>® a, b</sup>

<sup>a</sup> Institute of Plant Sciences Paris-Saclay IPS2, University of Paris Saclay, CNRS, INRAE, Bat. 630, 91192 Gif sur Yvette, France

<sup>b</sup> Institute of Plant Sciences Paris-Saclay IPS2, University of Paris Cité, CNRS, INRAE, Bat. 630, 91192 Gif sur Yvette, France

*E-mail:* martin.crespi@universite-paris-saclay.fr

**Abstract.** A plethora of non-coding RNAs have been found in eukaryotes, notably with the advent of modern sequencing technologies to analyze the transcriptome. Apart from the well-known house-keeping RNA genes (such as the ribosomal RNA or the transfer RNA), many thousands of transcripts detected are not evidently linked to a protein-coding gene. These, so called non-coding RNAs, may code for crucial regulators of gene expression, the small si/miRNAs, for small peptides (translated under specific conditions) or may act as long RNA molecules (antisense, intronic or intergenic long non-coding RNAs or lncRNAs). The lncRNAs interact with members of multiple machineries involved in gene regulation. In this review, we discussed about how plant lncRNAs permitted to discover new regulatory mechanisms acting in epigenetic control, chromatin 3D structure and alternative splicing. These novel regulations diversified the expression patterns and protein variants of target protein-coding genes and are an important element of the response of plants to environmental stresses and their adaptation to changing conditions.

**Résumé.** Une pléthore d'ARN non codants a été découverte chez les eucaryotes, notamment avec l'avènement des technologies modernes de séquençage pour analyser le transcriptome. En dehors des gènes ARN domestiques bien connus (tels que l'ARN ribosomal ou l'ARN de transfert), plusieurs milliers de transcrits détectés ne sont pas manifestement liés à un gène codant pour une protéine. Ces ARN, appelés ARN non codants, peuvent coder pour des régulateurs cruciaux de l'expression des gènes, les petits si/miARN, pour de petits peptides (traduits dans des conditions spécifiques) ou peuvent agir comme de longues molécules d'ARN (ARN non codants antisens, introniques ou intergénomiques ou lncRNA). Les lncRNAs interagissent avec les membres des multiples machineries impliquées dans la régulation des gènes. Dans cette revue, nous avons discuté de la façon dont les lncRNAs des plantes ont permis de découvrir de nouveaux mécanismes de régulation agissant dans le contrôle épigénétique, la structure 3D de la chromatine et l'épissage alternatif. Ces nouvelles régulations ont diversifié les profils d'expression des gènes codants pour les protéines cibles et constituent un élément important de la réponse des plantes aux stress environnementaux et de leur adaptation aux conditions changeantes.

**Keywords.** Long non-coding RNAs, Oligopeptides and mi/siRNAs, Chromatin regulation, Alternative splicing, Messenger RNA.

**Mots-clés.** Longs ARN non codants, Oligopeptides et mi/siRNAs, Régulation de la chromatine, Épissage alternatif, ARN messenger.

*Published online: 10 January 2023, Issue date: 16 February 2023*

## La version française de l'article est disponible à la suite de la version anglaise

### 1. Introduction: diverse non-coding RNAs exist in cells

The availability of genome sequences of numerous living organisms markedly changed our vision of the genome. Genomes are widely transcribed and produce thousands of long non-coding RNAs (lncRNAs), which are an abundant class of transcripts longer than 200 nucleotides with low protein coding capacities [1]. lncRNAs are generally transcribed by RNA polymerase II (Pol II), are poly-adenylated and may contain introns as a coding mRNA. It is considered that a long non-coding RNA has a function independent of its protein coding potential. Initially, genetic approaches in bacteria mainly revealed the role of protein coding genes in gene expression (e.g. transcription factors) and similarly, in eukaryotes, identification of mutations controlling various phenotypes emerging from genetic screenings also pointed mainly to protein coding genes. Indeed, until the 1940s, it was proposed that genome length could positively correlate with the organism's complexity [2]. Nevertheless, the strong differences of genome length observed between closely related organisms and the existence of very long genomes within simple organisms, noted in the early 1950s, give rise to a convolitional quandary, the so-called "C-value paradox" [3]. In fact, evolutionary complexity of eukaryotic organisms was supposed to deal with rare protein-coding genes immersed in the so-called "junk" DNA likely containing no function. However, the advent of genome sequencing revealed that in yeast the number of protein-coding genes was three times less than those detected in humans despite the large diversity of cell types and specialized functions of human cells in the organisms as compared to the 2 main cells (meiotic or not) linked to yeast biology. This led to the hypothesis that genome complexity maybe linked to the non-coding portion of the genome more than to the protein cod-

ing genome and that this junk DNA, from which functions could not be revealed using classical genetic approaches maybe hidden. Uncovering of large non-protein-coding DNA regions within the genome raised multiple questions concerning the relevance of these non-protein-coding sequences: How these sequences appear and are they functional? The extent advances of RNA sequencing methods revealed that more than 90% of eukaryotic genomes are transcribed to RNA with only 2–10% of individual transcripts translated into protein [4]. As expected, this tremendous quantity of transcripts is significantly different from each other, some are constitutively activated and participate in the basal function of cell life, such as gene transcription or translation whereas others are specifically activated upon particular stimuli or tissues, suggesting potential regulatory functions. In fact, many ncRNAs were linked to mRNA splicing and translation in all organisms. A coding gene is transcribed into an RNA molecule which is getting mature through the addition of a poly(A) tail and a N7-methylated guanosine linked to the first nucleotide in 5', modifications that increased mRNA stability and translation efficiency. In the nuclei, after transcription, the linear mRNA sequence can also be internally modified, or spliced, by the spliceosome, leading to the eviction of certain portions of the transcript, the so-called introns [5]. In humans, 73 mutations impairing the splicing process have been linked to cancer predisposition [6], strengthening the importance of this mechanism for cell life. Most of the time, during splicing all intronic regions are removed from the transcript. However, certain introns can be retained and certain exons can be skipped, leading to different mature mRNAs from one locus through alternative splicing (AS), permitting in certain cases to expand protein variants from the same gene. The spliceosome is constituted of small nuclear non-coding RNAs (snRNAs) that bind to proteins which together form a

ribonucleoprotein complex and catalyze the splicing reactions [7]. Notably, more than 90% of pre-mRNA transcripts are spliced in humans, many in a tissue-dependent or developmental-related manner [8]. Similarly, in higher plants, more than 60% of transcripts containing intron undergo AS [9]. Once mature, the coding transcripts exit the nucleus and are bound and decoded by the ribosomal complex, again a large complex formed with multiple proteins and the ribosomal non-coding RNAs (rRNAs) [10]. This complex is able to recognize specific sequences from the coding transcripts and produce the corresponding protein with the help of another type of non-coding transcripts: the transfer RNA (tRNA), linked to amino-acids [11]. Hence, both splicing and translation are realized through the cooperation of proteins and non-coding transcripts.

## 2. The small RNAs, new regulators of gene expression

In addition to these well characterized house-keeping non-coding RNAs, new regulatory RNAs emerged in recent years. Small regulatory non-coding RNAs or small RNAs (sRNAs) are defined as transcripts with a length in-between 19–25 nt and not translated into protein. They are issued from double stranded RNA (dsRNA). Depending on their biogenesis and their mechanism of action they can be classified in different major types: small interfering RNAs (siRNAs), microRNAs (miRNAs) and tasi/phasiRNAs [12], all of which, fine-tune the transcriptional or post-transcriptional gene activity of their target mRNAs through DNA methylation, cleavage or translation regulation. Several excellent reviews dedicated to small RNA biogenesis and function have been published and we will not describe them here [12, 13]. These small RNAs are encoded in the genome in large precursors that are being processed into small RNAs through the RDR/DCL machinery and their biogenesis allowed the classification of these different types of siRNAs [12, 14]. Hence, many long non-coding RNAs encode those small non-coding RNAs that are major regulators of gene expression as their action has been involved in many aspects of eukaryotic development. In plants, they were linked to developmental processes as well as to the response to the environment. Interestingly, certain lncRNAs, such as TAS3, which are not precursors

of miRNAs, are targeted by a miRNA, miR390, to induce a cleavage reaction in the lncRNA. This process triggered the formation of double-stranded RNA and the processing by DCL4 to form the so-called tasiRNAs. These secondary tasiRNAs now target mRNAs of the ARF family (auxin-related transcription factors) to post-transcriptionally control both spatial or temporal ARF expression. Interestingly many miRNA pathways function in a wide number of processes, for example, the *miR390-TAS3-ARF* pathway is implicated in lateral root growth, leaf formation, embryogenesis and is conserved in all land plants [15]. Also, a wheat lncRNA *WSGAR* is targeted by *miR9678* to generate phasiRNAs (processed similarly as the tasiRNAs) that are involved in seed development and germination [16]. Altogether, miRNAs, siRNAs and tasi/phasiRNAs generated after miRNA-mediated cleavage revealed already a portion of lncRNAs contributing the complexity of gene regulation through the formation of small RNAs.

In addition, many lncRNAs act as such to modulate the function of miRNAs *in vivo*. For example, the gene *PHOSPHATE 2 (PHO2)*, a key regulator of phosphate (Pi) homeostasis, is targeted by *miR399* and the *AtIPS1* lncRNA shares a highly similar 23nt sequence with *PHO2*, corresponding to the recognition sequence of *miR399*. Interestingly, in *IPS1*, there are 3nt sequence mismatches in the middle of the *miR399* binding site leading to the formation of a bulge avoiding *IPS1* cleavage by miR399. Consequently, *IPS1* is acting like a “suicide” substrate of *miR399*-mediated cleavage and sequesters miR399 to consequently inhibit its action on *PHO2* regulation. This mechanism called target mimicry [17] is a nice example of new regulatory mechanisms linked to lncRNAs and was initially discovered in plants. These miRNA “sponges” can blocked miRNA-mediated post-transcriptional regulation of specific mRNA targets. Interestingly, the *IPS1*-mediated *miR399* inhibition is also conserved in maize, tomato and *Medicago truncatula* [18, 19] and bioinformatic screenings in *Arabidopsis* revealed 407 putative lncRNA miRNA-sponge acting in response to blue-light; one of which directly implicated in the blue-light-mediated photomorphogenesis and mannitol stress response through the sequestration of *miR167*, involved in hypocotyl elongation [20]. This type of regulation in which a lncRNA trap miRNA was named target mimicry and was also

found in animals, where two thirds of the human transcriptome is likely to be regulated by miRNA. These miRNA mimics or “sponges” open wide perspectives to the role of these class of lncRNAs in post-transcriptional regulation [21]. Interestingly, Circular RNAs (CircRNAs) are also a recent class of non-coding RNAs formed through reverse splicing of an exon or cyclization of an intron within a gene, generally coding but also non-coding in specific cases [22]. These CircRNAs are short, present a covalent 5′–3′ ends linkage [23], are non-polyadenylated and significantly more stable than linear lncRNA. Abundant in the human genome, and expressed in a tissue-specific manner [24], many CircRNAs have been shown to act mainly as miRNA sponges as IPS1. Thousands of CircRNAs have been identified in humans, and also in plants in nearly 30 species, among which *Arabidopsis* and many crops [25]. Despite their responsiveness to abiotic and biotic stresses, their biological functions in plants remain to be elucidated. Hence, small RNAs either through their biogenesis or interaction with lncRNAs revealed specific mechanisms of gene regulation where lncRNAs play a significant role in plant development and stress responses.

### 3. To code or not to code: that is the question; Peptides or RNAs?

The increased sensitivity of next-generation sequencing methods allowed to detect large amounts of long non-coding RNAs in eukaryotes [14]. lncRNAs are transcripts superior to 200 nt in length, presenting a low ability to encode for proteins and are classified according to their genomic position in relation to other genes. First, the intronic lncRNAs are lncRNAs transcribed within the intron of another gene that does not overlap with any exons, independently of its orientation [1], while they are classified as exonic or sense lncRNAs if they overlap with the exon of a gene and on the same strand. In contrast, lncRNAs overlapping with exonic genomic regions but on the other strand of the DNA are classified as natural antisense lncRNA (NAT-lncRNA). lncRNAs transcribed from the promoter of a gene may constitute a promoter-lncRNA and the intergenic lncRNAs (lincRNAs) are lncRNAs transcriptional units localized in-between two genes (including their promoter region (e.g. 1 kb before the TSS)

although this is not an absolute definition [1]. The number of discovered lncRNAs increased together with the development of sequencing and bioinformatic prediction tools. For example, in 2017 the FANTOM5 project identified around 27,919 lncRNAs in humans, whereas in 2018 it has been estimated that the human genome contains around 270,044 lncRNAs (<https://fantom.gsc.riken.jp/5/>). Similarly in plants, the number of annotated lncRNAs significantly increased over time. For example, the study of Ben Amor (2009) identified only 76 lncRNAs in *Arabidopsis* from detailed expertise of full length cDNA databases [26], whereas RNAseq identified 6510 lncRNAs, among which 4050 NAT-lncRNAs and 2460 lincRNAs [27]. Now, thousands of lncRNAs have been identified within more than 40 plant species. For example the *Green Non-Coding Database* (GreeNC) annotated more than 120,000 lncRNAs within 37 plant species and six *algae* [28].

There is an increasing number of pipelines able to decipher if a transcript is coding or non-coding based on multiple parameters including, the transcript sequence, secondary structure and RNA conservation that allowed to significantly improve genomic annotations. For example, the CPC program uses both sequence alignment and Open Reading Frame (ORF) length and coverage whereas the phyloCSF uses known protein databases together with sequence alignment to classify a transcript as coding or non-coding [29]. CNCI and PLEK use nucleotide composition such as GC content and *k-mer* occurrence [30,31], whereas COME uses structural features of the transcript and epigenetic information to decipher its coding ability [32]. These tools are based on a machine learning approach, where the software is trained on a set of known non-coding and coding transcripts. More recently, newly developed tools use a whole set of alignment and non-alignment methods to analyze multiple transcripts features, such as the ORF, RNA secondary structure, encoded peptides isoelectric point and Ficket score. Among these tools, CPC2, emerged as a widely used coding prediction tool for a large range of species [33].

Even though lncRNAs share common features with coding RNAs (polyA, spliced transcripts) they also have distinct characteristics such as generally be shorter in length, contain less exons and produce a lower number of isoforms as compared to coding RNAs [34, 35]. Moreover, they are often less abundant

than coding RNAs and frequently retained in the nucleus whereas the coding RNAs are rapidly translocated to the cytoplasm to be translated. In mammals, promoters of lncRNAs tend to have less TF binding sites and fewer histone modifications that could explain their lower abundance as compared to coding transcripts. Indeed, a massively parallel reporter assay, which compared the activity of 2078 coding and lncRNAs gene promoters, shows that coding gene promoters have higher activity than lncRNA promoters [36]. The low stability of lncRNA transcript could also participate in their poor abundance.

As lncRNA molecules are generally more than 200 nt in length, it is likely that they contain small ORF that could be recognized and decoded by ribosomal units [37], even though classification generally states that lncRNA have no discernable coding potential. More importantly, a long transcript is considered as lncRNA if it is biologically functional in its RNA state, even if it may produce certain peptides [1]. Alternatively, some transcripts can have a function as RNA and as protein, suggesting that they can be “dual” RNAs [38]. For example, the *ENOD40* transcript is involved in nucleocytoplasmic trafficking of MtRBP1, but can also generate small peptides participating in root symbiotic nodule organogenesis. Similarly, *Steroid receptor RNA activator (SRA)*, regulates steroid receptor-dependent gene expression in its RNA state but can also produce a peptide that modulates the transcriptional activity of *SRA1* gene [39]. Interestingly, the SRA gene produces a lncRNA or peptide depending on the isoform produced, strengthening the relevance of AS for the production of bifunctional RNA molecules [40]. Conversely, the *TAD* gene in *Drosophila* implicated in embryonic development is classified as lncRNA but a 7aa motif present in an ORF is critical for its function. Furthermore, even if the RNA molecule contains a sORF, the RNA moiety may modulate its ability to interact with the ribosomes and yield translation products [41]. Indeed, RNA secondary structure can influence the splicing process leading to alternatively spliced mRNA that will have less affinity with ribosomes or lacked a start/stop codon, avoiding their translation into protein [42]. More generally, increasing evidence from bioinformatic analyses and ribosome profiling, shows that certain lncRNA associates with ribosomes in plants and animals [43–45] although out of the thousands of sORF discovered

in lncRNAs, very few seem to produce detectable peptides such as *MLN* [46]. As the proportion of coding lncRNA varies greatly between ribosomal-profiling studies [47, 48], the Mass spectrometry (MS) emerged as a complementary method to decipher the coding ability of lncRNAs and more recently peptidomics approaches were combined together with enrichment protocols to overcome the weak sensitivity of MS when compared to transcriptomics [49]. In plants, several sORFs present in lncRNA such as miRNA or siRNA precursors were shown to be associated with ribosomes by RiboSeq approaches, however these sORF were not evolutionary conserved, in contrast to the encoded regulatory small RNA. For the TAS3 lncRNA, the translation process per se may be required to stabilize the lncRNA molecular to be able to generate dsRNA and tasiRNAs under particular conditions [44]. Even though we would like to classify lncRNAs as exclusively coding or non-coding, it seems that the frontier between coding and non-coding is subtle and likely RNA sequences and sORF-encoded peptides may develop functions in different cellular contexts across evolution.

The increasing availability of whole genome sequences from different species revealed highly conserved genomic regions between all living organisms, such as the ribosomal related genes and the homeobox-genes. On the other hand, DNA from non-coding genes and/or intergenic regions are lowly conserved between species and even specific to some clades or individuals. Indeed, structural genetic variants are less frequent in Coding DNA Sequences (CDSs) and introns as compared to other DNA regions such as non-coding genes and intergenic regions in humans [50]. Similarly, in *Arabidopsis thaliana*, the 1001 genome project shows that coding regions accumulate less Single Nucleotide Polymorphism (SNP) than non-coding or intergenic regions [51]. Likewise, in 66 rice accessions, less than 5% of SNPs and Insertion/deletions (indels) were located in coding regions while representing around 10% of the genome [52]. Only less than 2% of the *Arabidopsis thaliana* lncRNAs are conserved at the sequence level across the plant kingdom. Concomitantly, it has been observed that highly expressed genes tend to be more conserved than lowly or specifically expressed genes, such as lncRNAs [53]. Logically, the conserved *Brassicaceae* lncRNA (18% of sequence-level) present a higher level of expres-

sion as compared to the non-conserved ones [54]. Genome analysis of five monocotyledon and five dicotyledon species shows that lncRNA conservation remains high within the same species but strongly decreases at the inter-species level [55]. Even between ecotypes the landscape of non-coding RNAs showed variability at expression level independently of minor SNPs between accessions [56]. Thus, it is tempting to assume that the non-conserved genome reflects the specific adaptation of an accession or an individual to its environment [56, 57]. It is noteworthy that the sequence of the lncRNA gene may not be conserved although their relative position (synteny) within the genome may be [58] suggesting that the interaction of the lncRNA with its syntenic neighboring coding genes may have biological relevance for their co-regulation.

#### 4. Expression and subcellular localization of lncRNAs

Rather than possessing recognizable specific domains across evolution like functional proteins, lncRNAs harbor specific short sequences and secondary structures which can shape their 3D structure and affect their interaction with other molecules. Interactor elements (IEs) serve for physical interactions with various partners through base complementarity (with other nucleic acids) and sequence-specific recognition by RNA-binding proteins (RBPs). On the other hand, structural elements (SEs) allow the formation of secondary and/or 3D lncRNA structures, directing their functional interactions with other cellular partners. Structural domains contain both IEs and SEs in various combinations and permit interactions with RBP complexes [59]. The association of these different elements could represent one of the languages that serve for directing lncRNAs interactions. As an example, the lncRNA Xist harbors 33 regions that form well-defined secondary structures linked by structurally variable regions, including a conserved A-repeat element. This region forms an inter-repeat structure which is essential for its control over X chromosome inactivation [60].

The understanding of the RNA repertoire still requires further efforts for seizing the underlying signals that allow lncRNAs to exert their functions. Nevertheless, a few studies have tried to associate

well known protein-domains with their RNA counterparts [61]. One of these signals consists in the information guiding their subcellular localization. Indeed, lncRNAs can either be exported to the cytosol or reside in the nucleus, contributing to various cell processes in both cases [62]. lncRNAs are generally more enriched in the nucleus compared to mRNAs, possibly due to their less efficient splicing and their interactions with nuclear RBPs. An RNA motif that recognizes the U1 small nuclear ribonucleoprotein (snRNP) and is essential for mobilizing lncRNAs to chromatin was discovered, therefore retaining them into the nucleus [63]. Similarly, the lncRNA BORG exhibits a pentamer RNA motif which is essential for nuclear retention. The mutation of this motif to a scrambled sequence resulted in the loss of nuclear localization. Conversely, the addition of a single copy of the motif in a cytoplasmic RNA was sufficient to induce its retention in the nucleus [64]. These specific RNA domains could therefore represent the RNA counterparts of the well described nuclear localization signal (NLS) present in proteins. The characterization of these RNA signals will be of major relevance to develop new tools for molecular biology and the study of the mechanisms of action of lncRNAs.

Besides the regulation of their subcellular localization, lncRNAs expression at the tissue level is also tightly regulated. Interestingly, lncRNAs are generally more expressed in a tissue-specific manner than coding-genes in both plants and animals [65, 66]. For instance, more than 30% of *Arabidopsis* lncRNAs display an organ- or developmental-specific expression pattern [65]. The growing number of discovered lncRNAs, their diverse sequences and their high tissue-specificity suggest that lncRNAs could serve as markers of tissues and developmental stages, and that they may be at the basis of the large diversity of regulatory patterns of expression across evolution to generate the wide variability of forms and functions throughout development and stress responses in eukaryotes.

#### 5. lncRNA in the regulation of gene expression

lncRNAs have emerged as important regulators of gene expression both at transcriptional and post-transcriptional levels. We will now focus the next part on the mechanisms involving plant long non-coding RNAs although we will cite few papers from animal studies in relation to these mechanisms

as several concepts are common and give a major relevance to the generality of these mechanisms in evolution. Nevertheless, there are excellent recent reviews to describe the plethora of information about long non-coding RNAs in animal cells [67] and briefly refer to plant lncRNAs.

Transcription involves the cooperation between the RNA Polymerase machinery, TFs and other complexes. Among them, the MEDIATOR complex mediates the communication between TFs and Pol II, influencing the transcriptional rate. Interestingly, in *Arabidopsis*, the lncRNA *ELF18-INDUCED LONG-NONCODING RNA1 (ELENA1)* is able to bind to *MED19a*, a subunit of the *MEDIATOR* complex, and modulate its recruitment to specific promoters [68]. A pathogen attack triggers the accumulation of *ELENA1* which binds to the promoter region of *PATHOGENESIS-RELATED1 (PR1)* gene allowing to enhance the recruitment of *MED19a* in this region activating PR1 expression. Likewise, the *PANDA* and *DHFR* human lncRNAs were also shown to directly interact with TFs in order to modulate their binding to gene promoters, directly influencing the transcription of these lncRNA-recognized targets [69]. lncRNAs can also act negatively on the regulation of target gene transcription. In *Arabidopsis*, the *HIDDEN TREASURE1 (HID1)* lncRNA decreases *PHYTOCHROME-INTERACTING FACTOR 3 (PIF3)* transcriptional activity by binding to its first intron. Downregulation of *HID1* expression increases *PIF3* gene activity and the subsequent elongation of the hypocotyl. The mechanism by which *HID1* decreases gene expression is still unclear, however *HID1* forms a ribonucleoprotein complex interacting with the *PIF3* genomic region suggesting that this recruitment may impair Pol II progression along the *PIF3* gene [70].

When two genes oriented in the opposite direction overlap are generally considered antisense RNAs and, in many cases, one of the members of this couple is a lncRNA. However, two genes cannot be transcribed at the same moment, otherwise the two Pol II complexes will encounter each other and their respective progression on the DNA will be blocked. In *Arabidopsis*, the antisense lncRNA *SVALKA* is a nice example where this process controls its neighboring protein-coding gene *CBF1* involved in cold tolerance. Prolonged cold exposure triggers a transcriptional read-through of the *SVALKA* genomic region

leading to Pol II collision between *SVALKA* and *CBF1* transcriptions. This down-regulates *CBF1* and fine tunes the response of the plants to this stress [71]. In agreement, in yeast, Pol II collision of natural antisense RNAs lead to the removal of both Pol II polymerases from the DNA through proteolysis [72]. An alternative explanation for the co-existence of transcription in both strands on the same region can be that each member of the couple is expressed in different cells as shown for specific antisense RNAs [73]. In this case, the role of the lncRNA is to spatially control gene expression. Taken together, lncRNAs can change the transcriptional efficiency of a gene through interaction with transcription-related proteins or through transcriptional readthrough leading to different transcriptional outputs.

### 5.1. lncRNAs in chromatin regulation

There are a large number of evidences that lncRNAs can physically interact with proteins able to modify the epigenetic landscape in cells. One of the best studied examples is the epigenetic regulation by lncRNAs of the *FLC* flower regulator [74]. The *FLOWERING LOCUS C (FLC)* gene encodes a MADS TF involved in flowering transition through a complex epigenetic switch likely involving the action of three lncRNAs. First, *COOLAIR*, an antisense transcript of the *FLC* gene, physically interacts with the 5' region of *FLC* and mediates, with *FLOWERING LOCUS D (FLD)*, a histone demethylase the demethylation of H3K4me2 in the locus resulting in *FLC* transcriptional repression. Additionally, the *COLD ASSISTED INTRONIC NONCODING RNA (COLDAIR)* *FLC* intronic lncRNA interacts with the CLF PRC2 sub-unit to enhance the epigenetic silencing of *FLC* gene through H3K27me3 deposition [75]. Finally, another lncRNA, *COLDWRAP* arising from the promoter of *FLC* is also able to interact with the PRC2 complex and modulates *FLC* silencing in different ecotypes [76]. Another example in *Arabidopsis* for such a mechanism was linked to the *AGAMOUS INTRONIC RNA 4 (AG-incRNA4)*. This lncRNA binds to the CLF sub-unit from the PRC2 complex and is encoded in the first intron of the *AGAMOUS (AG)* gene which codes for another MADS TF involved in flower development. In PRC2 mutants, a reduction in the repressive mark H3K27me3 on *AG* chromatin was detected with a concomitant induction of *AG*. Interestingly,



the intronic *AG-incRNA4* lncRNA seems inked to the recruitment of PRC2 complexes to this locus leading to its own silencing and establishing a feedback loop as production of the lncRNA requires AG expression resulting in gene repression [77]. Related mechanisms were analyzed also in other plants such as rice. The expression of the *LRKs* gene cluster involved in rice grain yield is modulated by an antisense lncRNA, named *LRK ANTISENSE INTERGENIC RNA (LAIR)*, transcribed from a region inside the *LRK1* gene. This lncRNA is able to recruit the *OsWDR5 (WD REPEAT DOMAIN5)*, involved in the H4K16 acetylation, of the *LRK* genomic region, increasing their expression. Conversely, the overexpression of *LAIR* lncRNA drastically increased grain yields in rice [78] and reduces the presence of this histone modification in this region.

lncRNAs can also modulate the chromatin state through DNA methylation and the production of siRNAs. In this case, the lncRNA region can be processed into specific siRNAs able to trigger gene silencing through DNA methylation via the RdDM mechanism [14]. For example, in *Arabidopsis*, an auxin stimulus triggers a strong upregulation of both the PIN-related protein kinase PID, regulating root gravitropism, and its neighboring lncRNA *AUXIN-REGULATED PROMOTER LOOP (APOLO)*. This auxin induction is mediated by the disruption of a chromatin loop encompassing the *APOLO* region and the PID promoter likely via an active DNA demethylation process. The demethylation of the PID-*APOLO* dual promoter by auxin opens this chromatin loop and permits the expression of both genes. Gradually, Pol II-dependent *APOLO* transcripts recruit LHP1 protein and the RdDM machinery to reform the chromatin loop and strengthen the compaction of the chromatin loop [79]. Then, both repressive marks and DNA methylation of this region maintains the silencing of the *APOLO-PID* regions. As the epigenetic status and chromatin condensation influences genome topology, it was proposed that chromatin-related lncRNAs are actors in the 3D spatial configuration of the genome. Indeed, in the last decade, many evidences linked lncRNAs to the modulation of chromatin conformation in different species [80]. This was correlated with classical experiments showing that RNase A treatment of nuclei resulted in a global decondensation of chromatin [81]. In plants, the previously mentioned *COLDWRAP* lncRNA tran-

scribed within the *FLC* flowering repressor gene mediates the formation of a repressive intragenic chromatin loop which blocks Pol II transcription, inhibiting *FLC* expression allowing plant to flower [76]. On the other hand, the *APOLO lncRNA* was shown to also act in trans, through the formation of RNA:DNA duplexes named R-loops. R-loops containing the *APOLO RNA* are able to displace LHP1 at distant loci and modulate their chromatin conformation [82]. Interestingly, a significant proportion of genomic distant loci recognized by the *APOLO RNA* are auxin-responsive genes. Among them, the *LEUCINE RICH EXTENSIN2 (LRX2)*, involved in cell wall remodeling upon lateral root emergence, and *ROOT HAIR DEFECTIVE 6 (RHD6)*, a key regulator of root hair initiation [83, 84].

## 5.2. *lncRNAs mediating post-transcriptional regulation of gene expression*

In addition to the regulatory mechanisms involving small RNAs (e.g. target mimicry or siRNA production as mentioned above), recently alternative mechanisms for the post-transcriptional control of gene expression by lncRNAs emerged. These lncRNAs target the splicing machinery and this interaction highlights the potential of lncRNAs to contribute to the generation of protein diversity through the regulation of alternative splicing (AS). In plants, the *Arabidopsis* lncRNA *Alternative Splicing Competitor (ASCO)* was shown to physically interact with the nuclear speckle RNA-binding proteins (NSRs), which are alternative splicing regulators during developmental processes. By overexpressing the *ASCO* lncRNA in plants, many NSR-alternatively spliced targets were modified showing that *ASCO*-NSRs interactions affect AS of multiple auxin related genes and influence auxin-driven lateral root formation in *Arabidopsis* [85, 86]. More recently, the *ASCO* lncRNA have been found to interact also with other core components of the spliceosome (SmD1b and PRP8a, components directly participating in splicing reactions) suggesting that dynamic and complex interaction of lncRNAs and spliceosomal components, such as the U snRNAs, may intervene in the regulation of specific targets [87]. There are several studies providing clues about the potential role of lncRNAs in the control of AS in animal cells and three types of regulation seem to emerge: (a) chromatin remodelling

and RNA-DNA hybrids, (b) formation of lncRNA-RNA hybrids and (c) alteration of splicing factors (SFs) proteins or AS mRNA targets interaction with the spliceosome to affect directly the splicing reactions.

### 5.2.1. *LncRNAs affecting chromatin remodeling in the context of splicing*

In addition to the previously mentioned regulation of chromatin conformation and epigenetic status by lncRNA, the chromatin context of a locus, notably in the gene body, was also found to affect AS. Since splicing mainly occurs co-transcriptionally, the modulation of Pol II elongation rate by the chromatin context fine-tunes the choice of alternative splice sites [88]. Hence, as lncRNA can affect chromatin remodeling, it seems plausible that certain lncRNAs interacting with chromatin, notably in gene bodies, may indirectly affect alternative splicing. In plants, this mechanism was recently analyzed by uncovering the role of the non-coding circRNA *SEP3* in the AS modulation of its own gene [89]. *SEP3* is a member of the MADS (MCM1-AGAMOUS-DEFICIENS-SRF)-box superfamily and was shown to participate in flower development. Defects in *SEP3* splicing provoke floral homeotic phenotypes, underlying its importance in this developmental process. Surprisingly, the overexpression of a circRNA containing the entire exon 6 of the *SEP3* gene led to the accumulation of the AS variant *SEP3.3* lacking this exon 6. It was further shown that *SEP3* exon 6 circRNA can directly interact with its cognate DNA locus through the formation of R-loops (RNA-DNA hybrids). This R-loop promotes transcriptional pausing on its own gene and affects the recruitment of splicing factors [89]. This specific mechanism suggests that circRNAs may participate in the AS regulation of their cognate exon-skipped messenger RNAs. However, it also opens the possibility that certain lncRNAs only encode circRNAs acting in splicing *in trans*. Interestingly, in animals many circRNAs are encoded in intergenic genome regions [23]. Other examples in animals suggest that lncRNA modulate AS through chromatin interactions. For example, the lncRNA *as-FGFR2* (for antisense *FGFR2*) is transcribed from the human *FGFR2* locus and induces epithelial-specific AS of *FGFR2* [90]. This AS is permitted by the recruitment of chromatin modifiers specifically to its own locus, leading to changes in the DNA methylation status of this region. The presence of *asF-*

*GRF2* ensures the deposition of H3K27me3 mark and the decrease of H3K36me2/3 marks, resulting in impaired recruitment of the chromatin-binding protein MRG15 and the negative splicing regulator PTBP1. The MRG15-PTBP1 complex cannot any longer inhibit the inclusion of the exon IIIb in *FGFR2*, leading to the epithelial-specific AS of this gene. Hence, lncRNAs can modulate chromatin conformation either by direct binding to DNA (through R-loops) or by recruiting specific epigenetic regulators to genomic loci that may show changing epigenetic and AS patterns in different cells or in response to environmental stresses.

### 5.2.2. *LncRNA-RNA hybrids as AS regulators*

Regardless of their genomic origin, natural antisense transcripts or NATs, can hybridize with pre-mRNAs and form RNA-RNA duplexes. These complexes were shown to impact AS during various biological processes, notably in mammalian cells [85]. One major apoptotic pathway in animals is activated through the interaction between the Fas receptor (Fas) and the Fas ligand (FasL) [91]. At the *FAS* locus, the antisense lncRNA *SAF* is transcribed in reverse orientation and from the opposite strand of the first intron of *FAS*. *SAF* localizes in the nucleus where it binds to the Fas receptor pre-mRNA and the human SPLICING FACTOR 45 (SPF45). This interaction facilitates the AS and exclusion of the exon 6, leading to the production of a soluble Fas protein that protects cells against FasL-induced apoptosis [91]. Thus, NAT transcripts can facilitate interactions between pre-mRNAs and specific SFs both in cis (on its own transcript) or in trans whereas they can also mask specific splice sites to prevent their processing by the spliceosome. In *A. thaliana*, the increased frequency of alternatively spliced and variably polyadenylated transcripts when an intron overlaps with a NAT suggests that the formation of NAT lncRNA-RNA pairs may regulate the AS of protein-coding genes [92]. Consistently, a genome-wide screen of trans-NATs in *A. thaliana* allowed the identification of 1320 putative trans-NAT pairs [93]. Even though certain trans-NATs were predicted to produce double stranded RNAs and siRNAs (so-called natsiRNAs) leading to RNA silencing of their sense target, the study showed that many trans-NAT pairs have a much higher proportion of AS events compared to all transcription units in the genome. Thus, lncRNAs can take advantage of

sequence similarity to bind specific transcripts and modulate their splicing.

### 5.2.3. *lncRNAs altering splicing factors activity and interactions with the spliceosome*

The spliceosome is a giant ribonucleoprotein complex where splicing factors, small nuclear RNAs and target mRNAs meet to define the splicing outcome of a gene. The modulation of SF proteins activity can occur through many ways: by promoting post-translational modification of the protein, changing its subcellular localization, or impacting protein binding to other partners or transcript targets. One of the most deeply characterized lncRNAs associated with AS regulation are the *NUCLEAR PARASPECKLE ASSEMBLY TRANSCRIPT 1 (NEAT1)* and *METASTASIS ASSOCIATED LUNG ADENOCARCINOMA TRANSCRIPT 1 (MALAT1)/NUCLEAR PARASPECKLE ASSEMBLY TRANSCRIPT 2 (NEAT2)*. Both of these lncRNAs were shown to modulate the localization and phosphorylation status of specific SFs, and to exhibit differential expression in a wide range of human and murine tissues [94]. *NEAT1* is a highly abundant lncRNA found in paraspeckles, nuclear domains controlling the sequestration of splicing-related proteins. During adipocyte differentiation, the abundance of *NEAT1* is dynamically regulated to modulate the relative levels of PPAR $\gamma$  mRNA isoforms, the major TF driving adipogenesis. In short, *NEAT1* was shown to interact with the SR protein SRp40 (SFRS5), leading to SRp40 retention in paranuclear bodies. The *NEAT1*-SRp40 interaction enhances SRp40 phosphorylation by CDC2-LIKE KINASE 1 (CLK1), a kinase specifically targeting SFs. This change in SRp40 phosphorylation promotes PPAR $\gamma$  AS, therefore fine-tuning the adipogenesis process [95].

The lncRNA *MALAT1/NEAT2* localizes in nuclear speckles and exerts oncogenic roles within the cell. Like *NEAT1*, *MALAT1/NEAT2* can effectively modulate SF distribution and phosphorylation, leading to AS changes in their target pre-mRNAs [96]. In plants, the characterization of lncRNAs directly involved in AS is still at its infancy. Nevertheless, a few examples showing lncRNAs interactions with SFs are slowly emerging. In the legume plant *Medicago truncatula*, the lncRNA *ENOD40* is rapidly induced upon interaction with symbiotic rhizobial bacteria. It is expressed in the root pericycle and in the differentiating cells of the nodule primordia [97]. The over-

expression of *ENOD40* leads to accelerated nodulation, mainly caused by increased initiation of primordia and an enhanced sensitivity to nodulation signals. *ENOD40* was found to be highly structured and did not associate with polysomes. Yeast three-hybrid assays revealed direct interaction between *ENOD40* and the constitutive RNA Binding Protein 1 (RBP1), which localizes into nuclear speckles where the splicing machinery is also hosted [98]. During nodulation, RBP1 is re-localized to cytoplasmic granules through its association with *ENOD40*. Therefore, the highly structured *ENOD40* lncRNA contributes to nucleocytoplasmic trafficking of RBP1, suggesting that RBP1 role in the nucleus may be perturbed during nodule development. MtRBP1 is a close homolog of AtNSRs splicing factors; known to regulate AS and it was demonstrated major changes in AS during nodule differentiation concomitant to *ENOD40* expression [99]. Although a direct effect of *ENOD40* on AS has not yet been shown in *Medicago truncatula*, the relocalisation of an SF into the cytoplasm may affect AS patterns.

As mentioned before the overexpression of the lncRNA named *ASCO* was shown to modulate AS of NSR mRNA targets during lateral root development in *A. thaliana* [86]. Identified first by Ben Amor *et al.* [26] using a genome-wide bioinformatics analysis of *Arabidopsis* full-length cDNA databases, *ASCO* is encoded within an intergenic region of chromosome 1 (AT1G67105). It has a size of 786 nt and no obvious protein-coding capacity and accumulates in the nucleus. Consistently, *ASCO* was not found to be associated with ribosomes [44]. A ClustalW and MUSCLE alignment identified at least 4 *ASCO* paralogs in *A. thaliana*, and suggested a wide conservation of *ASCO* and its paralogs in *Arabidopsis* ecotypes and even between different *Brassicaceae* species. Interestingly, detailed analysis of AS in the *nsra/b* double mutant uncovered an important number of AS events, notably in response to auxin [100] that were similarly observed in the *ASCO* overexpressing lines. Moreover, RNA immunoprecipitation assays using NSR-GFP lines established that NSRs bind *in vivo* not only to their alternatively spliced mRNA targets, but also to the *ASCO* lncRNA and many other lncRNAs [101]. Using *in vitro* experiments to assess the binding of NSR-containing complexes to its mRNA substrates, Bardou *et al.* showed that the *ASCO* lncRNA was able to compete out of the complex the endogenous tar-

get mRNA in a very efficient sequence-specific manner. This suggested that *ASCO* can displace AS targets from an NSR-containing complex to modify their AS pattern [26] and proposed a model where the lncRNA *ASCO* hijacks the NSR splicing factors to prevent their interaction with its own targets. It remains to be seen whether this is a more general mechanism for other SF-lncRNA interactions but it is tempting to speculate that these “hijacker competitor lncRNAs” may be acting as intron mimics inside the spliceosome to regulate their function in a cell-specific manner.

Taken together, these studies suggest that lncRNAs integrate a dynamic splicing network including many SFs, chromatin states and their associated pre-mRNA targets. Through these varied range of interactions with different partners, lncRNAs exert control over transcriptome reprogramming through AS in eukaryotes.

## 6. Concluding remarks

Altogether, lncRNA molecules are emerging as new elements in regulatory networks controlling plant growth and development. Physically interacting with epigenetic-related complexes, TFs or the DNA itself, they fine-tune transcriptional activity of target protein-coding genes. Variation in lncRNA expression in response to external abiotic or biotic

stimuli, which show major variability between cell layers inside tissues, may lead to diverse expression

### *Version française*

#### 1. Introduction : Divers ARN non codants existent dans les cellules

La disponibilité des séquences génomiques de nombreux organismes vivants a considérablement modifié notre vision du génome. Les génomes sont largement transcrits et produisent des milliers de longs ARN non codants (lncRNAs), qui constituent une classe abondante de transcrits de plus de 200 nucléotides ayant de faibles capacités de codage des protéines [1]. Les lncRNAs sont généralement transcrits par l'ARN polymérase II (Pol II), sont polyadénylés et peuvent contenir des introns comme un

patterns of their targets at cellular level, enhancing specific differentiation processes or responses to external stimuli. Additionally lncRNAs interacting with key modulators of the mRNA splicing and stability, may also influence mRNA abundance of target genes at post-transcriptional level and may diversify protein outputs from a same gene. Globally, the discovery of these lncRNA-dependent regulatory mechanisms opens wide perspectives to understand the diversity of morphologies and adaptation capacities of plants.

#### Conflicts of interest

The author has no conflict of interest to declare.

#### Acknowledgements

I would like to thank Michel Caboche for interesting discussion about non-coding RNA (which were cautiously perceived by geneticists) and his support of non-coding RNA research in Genoplante programs. I also thanks Adam Kondorosi for his encouragement and mentoring to follow on non-coding RNA biology when this was just emerging. I would also like to deeply thank people of my group and all PhD students along these years for their enthusiasm and passion on this quest for understanding non-coding RNAs, notably Federico Ariel who was for many years a close collaborator. I would also like to thank Hervé Vaucheret for fruitful and inspiring discussions about “strange” cellular RNAs as well as for long-term collaborations with his team.

ARNm codant. On considère qu'un long ARN non codant a une fonction indépendante de son potentiel de codage des protéines. Initialement, les approches génétiques chez les bactéries ont principalement révélé le rôle des gènes codants pour des protéines dans l'expression des gènes (par exemple, les facteurs de transcription) et de même, chez les eucaryotes, l'identification des mutations contrôlant divers phénotypes émergeant des criblages génétiques a pointé principalement vers les gènes codants pour des protéines. En effet, jusque dans les années 1940, il a été proposé que la longueur du génome soit en corrélation positive avec la complexité de l'orga-

nisme [2]. Néanmoins, les fortes différences de longueur du génome observées entre des organismes étroitement apparentés et l'existence de très longs génomes au sein d'organismes simples, constatée au début des années 1950, donnent lieu à un dilemme circonstanciel, le « paradoxe de la valeur C » [3]. En fait, la complexité évolutive des organismes eucaryotes était censée concerner de rares gènes codants pour des protéines, immergés dans l'ADN dit « poubelle », susceptible de ne contenir aucune fonction. Or, l'avènement du séquençage du génome a révélé que chez la levure, le nombre de gènes codants pour des protéines était trois fois inférieur à celui détecté chez l'homme, malgré la grande diversité des types cellulaires et des fonctions spécialisées des cellules humaines dans les organismes par rapport aux 2 principales cellules (méiotiques ou non) liées à la biologie de la levure. Cela a conduit à l'hypothèse que la complexité du génome était peut-être liée à la partie non codante du génome plus qu'au génome codant pour les protéines et que cet ADN poubelle, dont les fonctions ne pouvaient être révélées par les approches génétiques classiques, était peut-être caché. La découverte de grandes régions d'ADN non codant pour les protéines au sein du génome a soulevé de multiples questions concernant leur pertinence : comment ces séquences apparaissent-elles et sont-elles fonctionnelles ? Les progrès considérables des méthodes de séquençage de l'ARN ont révélé que plus de 90 % des génomes eucaryotes sont transcrits en ARN et que seulement 2 à 10 % des transcrits individuels sont traduits en protéines [4]. Comme on pouvait s'y attendre, cette énorme quantité de transcrits recouvre une diversité très significative, puisque certains transcrits sont constitutivement activés et participent à la fonction basale de la vie cellulaire, comme la transcription ou la traduction des gènes, tandis que d'autres sont spécifiquement activés lors de stimuli ou dans des tissus particuliers, suggérant des fonctions régulatrices potentielles. En fait, de nombreux ARNnc ont été liés à l'épissage et à la traduction des ARNm dans tous les organismes. Un gène codant est transcrit en une molécule d'ARN qui devient mature par l'ajout d'une queue poly(A) et d'une guanosine N7-méthylée liée au premier nucléotide en 5', modifications qui augmentent la stabilité de l'ARNm et l'efficacité de la traduction. Dans les noyaux, après la transcription, la séquence linéaire de l'ARNm peut également être modifiée

de façon interne, ou épissée, par le spliceosome, ce qui conduit à l'éviction d'une partie du transcrit, les fameux introns [5]. Chez l'homme, 73 mutations altérant le processus d'épissage ont été liées à une prédisposition au cancer [6], ce qui confirme l'importance de ce mécanisme pour la vie cellulaire. La plupart du temps, lors de l'épissage, toutes les régions introniques sont éliminées du transcrit. Cependant, certains introns peuvent être conservés et certains exons peuvent être sautés, ce qui conduit à des ARNm matures différents à partir d'un locus par épissage alternatif (EA), permettant dans certains cas de développer des variantes de protéines à partir du même gène. Le spliceosome est constitué de petits ARN nucléaires non codants (snRNA) qui se lient à des protéines qui forment ensemble un complexe ribonucléoprotéique et catalysent les réactions d'épissage [7]. Notamment, plus de 90 % des transcriptions de pré-ARNm sont épissées chez l'homme, souvent d'une manière qui dépend du tissu ou du développement [8]. De même, chez les plantes supérieures, plus de 60 % des transcrits contenant un intron subissent une AS [9]. Une fois matures, les transcrits codants sortent du noyau et sont liés et décodés par le complexe ribosomal, là encore un grand complexe formé de multiples protéines et des ARN non codants ribosomiaux (ARNr) [10]. Ce complexe est capable de reconnaître des séquences spécifiques des transcrits codants et de produire la protéine correspondante à l'aide d'un autre type de transcrits non codants : l'ARN de transfert (ARNt), lié aux acides aminés [11]. Ainsi, l'épissage et la traduction sont tous deux réalisés par la coopération de protéines et de transcrits non codants.

## 2. Les petits ARN, nouveaux régulateurs de l'expression des gènes

En plus de ces ARN non codants domestiques bien caractérisés, de nouveaux ARN régulateurs sont apparus ces dernières années. Les petits ARN non codants régulateurs ou petits ARN (ARNs) sont définis comme des transcrits d'une longueur comprise entre 19 et 25 nt et non traduits en protéines. Ils sont issus de doubles brins (ds) d'ARN. En fonction de leur biogenèse et de leur mécanisme d'action, ils peuvent être classés en différents types principaux : les petits ARN interférents (siRNA), les microARN (miRNA) et les phasiARN [12], qui régulent

tous finement l'activité génique transcriptionnelle ou post-transcriptionnelle de leurs ARNm cibles par clivage ou régulation de la traduction. Plusieurs excellentes revues consacrées à la biogenèse et à la fonction des petits ARN ont été publiées et nous ne les décrirons pas ici en détail [12, 13]. Ces petits ARN sont codés dans le génome sous forme de grands pré-curseurs qui sont transformés en petits ARN par la machinerie RDR/DCL et leur biogenèse a permis la classification de ces différents types de siRNA [12, 14]. Ainsi, de nombreux longs ARN non codants codent pour ces petits ARN non codants et sont des régulateurs majeurs de l'expression des gènes car leur action est impliquée dans de nombreux aspects du développement des eucaryotes. Chez les plantes, ils ont été liés aux processus de développement ainsi qu'à la réponse à l'environnement. De manière intéressante, certains lncRNAs, tels que TAS3, qui ne sont pas des pré-curseurs de miRNAs, sont ciblés par un miRNA, miR390, pour induire une réaction de clivage dans le lncRNA. Ce processus déclenche la formation d'ARN double brin et la transformation par DCL4 pour former ce que l'on appelle les tasiRNAs. Ces tasiRNA secondaires ciblent maintenant les mRNAs de la famille ARF (facteurs de transcription liés à l'auxine) pour contrôler de manière post-transcriptionnelle l'expression des ARF dans l'espace ou dans le temps. Il est intéressant de noter que de nombreuses voies de miRNA fonctionnent dans un grand nombre de processus, par exemple, la voie miR390-TAS3-ARF est impliquée dans la croissance latérale des racines, la formation des feuilles, l'embryogenèse et est conservée dans toutes les plantes terrestres [15]. De même, un lncRNA du blé, WSGAR, est ciblé par miR9678 pour générer des phasiRNAs qui sont impliqués dans le développement et la germination des graines [16]. Dans l'ensemble, les miRNA, les siRNA et les phasiRNA générés après le clivage médié par les miRNA ont déjà révélé une partie des lncRNA contribuant à la complexité de la régulation des gènes par la formation de petits ARN.

De plus, de nombreux lncRNAs agissent en tant que tels pour moduler la fonction des miRNAs *in vivo*. Par exemple, le gène PHOSPHATE 2 (PHO2), un régulateur clé de l'homéostasie du phosphate (Pi), est ciblé par miR399 et le lncRNA AtIPS1 partage une séquence de 23nt très similaire avec PHO2, correspondant à la séquence de reconnaissance de

miR399. De manière intéressante, dans IPS1, il y a un mésappariement de séquence de 3 nt au milieu du site de liaison de miR399 conduisant à la formation d'un renflement évitant le clivage de IPS1. Par conséquent, IPS1 agit comme un substrat « suicide » du clivage médié par miR399 et séquestre miR399 et inhibe son action sur la régulation de PHO2. Ce mécanisme appelé mimétisme de cible [17] est un bel exemple de nouveaux mécanismes de régulation liés aux lncRNAs et a été initialement découvert chez les plantes. Ces « éponges » à miARN peuvent bloquer la régulation post-transcriptionnelle, médiée par les miARN, de cibles spécifiques d'ARNm. Il est intéressant de noter que l'inhibition du miR399 médiée par IPS1 est également conservée chez le maïs, la tomate et *Medicago truncatula* [18, 19] et que des criblages bio-informatiques chez *Arabidopsis* ont révélé 407 lncRNA miRNA-sponge putatifs agissant en réponse à la lumière bleue; l'un d'entre eux est directement impliqué dans la photomorphogenèse médiée par la lumière bleue et la réponse au stress du mannitol par la séquestration du miR167, impliqué dans l'élongation de l'hypocotyle [20]. Ce type de régulation dans lequel un lncRNA piège un miRNA est nommé mimétisme de cible. Ce mécanisme a également été trouvé chez les animaux, où deux tiers du transcriptome humain sont susceptibles d'être régulés par des miRNA, ouvrant de larges perspectives au rôle de cette classe de lncRNAs dans la régulation post-transcriptionnelle [21]. Il est intéressant de noter que les ARN circulaires (ARNc) constituent une classe récente d'ARN non codants formés par épissage inverse d'un exon ou par cyclisation d'un intron au sein d'un gène, généralement codant mais aussi dans certains cas non codant [22]. Ces ARNc sont courts, présentent une liaison covalente aux extrémités 5'-3' [23], sont non polyadénylés et significativement plus stables que les lncRNA linéaires. Ils sont abondants dans le génome humain, et exprimés de manière spécifique aux tissus [24], et il a été démontré que beaucoup d'entre eux agissent principalement comme des éponges à miARN comme IPS1. Des milliers d'ARN circulaires ont été identifiés chez l'homme, et également chez les plantes dans près de 30 espèces, parmi lesquelles *Arabidopsis* et de nombreuses cultures [25]. Malgré leur réactivité aux stress abiotiques et biotiques, leurs fonctions biologiques chez les plantes restent à élucider. Ainsi, les petits ARN, soit par leur biogenèse, soit par leur interac-

tion avec les lncRNA, ont révélé des mécanismes spécifiques de régulation des gènes où les lncRNA jouent un rôle important dans le développement des plantes et les réponses aux stress.

### 3. Coder ou ne pas coder : telle est la question. Peptides ou ARN ?

La sensibilité accrue des méthodes de séquençage de nouvelle génération a permis de détecter de grandes quantités de longs ARN non codants chez les eucaryotes [14]. Les lncRNAs sont des transcrits d'une longueur supérieure à 200 nt, présentant une faible capacité à coder pour des protéines et sont classés en fonction de leur position génomique par rapport aux autres gènes. Tout d'abord, les lncRNA introniques sont des lncRNA transcrits dans l'intron d'un autre gène qui ne chevauche aucun exon, indépendamment de son orientation [1], tandis qu'ils sont classés comme lncRNA exoniques ou sens naturel s'ils chevauchent l'exon d'un gène sur le même brin. En revanche, les lncRNAs chevauchant des régions génomiques exoniques mais sur l'autre brin de l'ADN sont classés comme lncRNAs antisens naturels (NAT-lncRNA). Les lncRNAs transcrits à partir du promoteur d'un gène peuvent constituer un promoteur-lncRNA et les lncRNAs intergéniques (lincRNAs) sont des unités transcriptionnelles lncRNAs localisées entre deux gènes (y compris leur région promotrice, par exemple 1 kb avant le TSS) bien que cela ne soit pas absolu [1]. Le nombre de lncRNAs découverts a augmenté en même temps que le développement du séquençage et des outils de prédiction bio-informatique. Par exemple, en 2017, le projet FANTOM5 a identifié environ 27 919 lncRNA chez l'homme, alors qu'en 2018, il a été estimé que le génome humain contient environ 270 044 lncRNA (<https://fantom.gsc.riken.jp/5/>). De même chez les plantes, le nombre de lncRNA annotés a significativement augmenté au fil du temps. Par exemple, l'étude de Ben Amor (2009) n'a identifié que 76 lincRNAs dans *Arabidopsis* à partir d'une expertise détaillée des bases de données d'ADNc complets [26], alors qu'une étude plus récente utilisant RNAseq a identifié 6510 lncRNAs, dont 4050 NAT-lncRNAs et 2460 lincRNAs [27]. Aujourd'hui, plus de milliers de lncRNA ont été identifiés dans plus de 40 espèces végétales. Par exemple, la *Green Non-Coding Database*

(GreeNC) a annoté plus de 120 000 lncRNA dans 37 espèces végétales et six algues [28].

Il existe un nombre croissant de pipelines capables de déchiffrer si un transcrit est codant ou non codant sur la base de multiples paramètres, notamment la séquence du transcrit, la structure secondaire et la conservation de l'ARN, qui ont permis d'améliorer considérablement les annotations génomiques. Par exemple, le programme CPC utilise à la fois l'alignement des séquences et la longueur et la couverture des cadres de lecture ouverts (ORF), tandis que le programme phyloCSF utilise des bases de données de protéines connues ainsi que l'alignement des séquences pour classer un transcrit comme codant ou non codant [29]. CNCI et PLEK utilisent la composition nucléotidique telle que la teneur en GC et l'occurrence des *k*-mer [30, 31], tandis que COME utilise les caractéristiques structurales du transcrit et les informations épigénétiques pour déchiffrer sa capacité de codage [32]. Ces outils sont basés sur une approche d'apprentissage automatique, où le logiciel est entraîné sur un ensemble de transcriptions codantes et non codantes connues. Plus récemment, des outils récemment développés utilisent un ensemble de méthodes d'alignement et de non-alignement pour analyser les caractéristiques de plusieurs transcrits, comme l'ORF, la structure secondaire de l'ARN, le point isoélectrique des peptides codés et le score de Fickett. Parmi ces outils, CPC2 est apparu comme un outil de prédiction de codage largement utilisé pour un large éventail d'espèces [33].

Bien que les lncRNA partagent des caractéristiques communes avec les ARN codants (polyA, transcrits épissés), ils ont également des caractéristiques distinctes telles que généralement une longueur plus courte, moins d'exons et un nombre plus faible d'isoformes par rapport aux ARN codants [34, 35]. De plus, ils sont souvent moins abondants que les ARNs codants et fréquemment retenus dans le noyau alors que les ARNs codants sont rapidement transloqués vers le cytoplasme pour être traduits. Chez les mammifères, les promoteurs des lncRNA ont tendance à avoir moins de sites de liaison TF et moins de modifications des histones, ce qui pourrait expliquer leur plus faible abondance par rapport aux transcrits codants. En effet, un test de rapporteur massivement parallèle, qui a comparé l'activité du promoteur de 2078 gènes codants et lnc

cRNAs, montre que les promoteurs de gènes codants ont une activité plus élevée que les promoteurs de lncRNAs [36]. La faible stabilité de la transcription des lncRNA pourrait également participer à leur faible abondance.

Comme les molécules de lncRNA ont une longueur de plus de 200 nt, il est probable qu'elles contiennent de petits ORF qui pourraient être reconnus et décodés par les unités ribosomiques [37], même si la classification stipule généralement que les lncRNA n'ont pas de potentiel de codage discernable. Plus important encore, un long transcrit est considéré comme un lncRNA s'il est biologiquement fonctionnel à l'état d'ARN, même s'il peut produire certains peptides [1]. Par ailleurs, certains transcrits peuvent avoir une fonction d'ARN et de protéine, ce qui suggère qu'ils peuvent être des ARN « doubles » [38]. Par exemple, le transcrit ENOD40 est impliqué dans le trafic nucléocytoplasmique de MtRBP1, mais peut également générer de petits peptides participant à l'organogenèse du nodule symbiotique racinaire. De même, l'activateur ARN du récepteur stéroïdien (SRA) régule l'expression des gènes dépendants des récepteurs stéroïdiens à l'état d'ARN mais peut également produire un peptide qui module l'activité transcriptionnelle du gène SRA1 [39]. Il est intéressant de noter que le gène SRA produit un lncRNA ou un peptide en fonction de l'isoforme produite, ce qui renforce la pertinence du SA pour la production de molécules d'ARN bifonctionnelles [40]. À l'inverse, le gène TAD de la drosophile impliqué dans le développement embryonnaire est classé comme lncRNA mais un motif 7aa présent dans un ORF est critique pour sa fonction. De plus, même si la molécule d'ARN contient un sORF, la partie de l'ARN peut moduler sa capacité à interagir avec les ribosomes et à donner des produits de traduction [41]. En effet, la structure secondaire de l'ARN peut influencer le processus d'épissage conduisant à des ARNm épissés alternativement qui auront moins d'affinité avec les ribosomes ou seront dépourvus d'un codon start/stop, évitant ainsi leur traduction en protéine [42]. Plus généralement, de plus en plus de preuves provenant d'analyses bio-informatiques et de profilage de ribosomes montrent que certains lncRNA s'associent aux ribosomes chez les plantes et les animaux [43–45] bien que parmi les milliers de sORF découverts dans les lncRNA, très peu semblent produire des peptides détectables tels que la MLN [46]. Comme la pro-

portion de lncRNA codants varie grandement entre les études de profilage ribosomal [47, 48], la spectrométrie de masse (MS) est apparue comme une méthode complémentaire pour déchiffrer la capacité de codage des lncRNAs et plus récemment des approches peptidomiques ont été combinées avec des protocoles d'enrichissement pour surmonter la faible sensibilité de la MS par rapport à la transcriptomique (‘ç). Chez les plantes, plusieurs sORFs présents dans les lncRNA tels que les précurseurs de miRNA ou de siRNA se sont avérés être associés aux ribosomes par des approches RiboSeq, cependant ces sORFs n'étaient pas conservés au cours de l'évolution, contrairement aux petits ARN régulateurs codés. Pour le lncRNA TAS3, le processus de traduction en soi peut être nécessaire pour stabiliser le lncRNA moléculaire afin de pouvoir générer des dsRNA et des tasiRNA dans des conditions particulières [44]. Même si nous aimerions classer les lncRNA comme étant exclusivement codants ou non codants, il semble que la frontière entre codant et non codant soit subtile et que les séquences d'ARN et les peptides codés par le sORF puissent développer des fonctions dans différents contextes cellulaires au cours de l'évolution.

La disponibilité croissante des séquences de génomes entiers de différentes espèces a révélé des régions génomiques hautement conservées entre tous les organismes vivants, comme les gènes liés aux ribosomes et les gènes homéobox. En revanche, les gènes non codants et/ou les régions intergéniques sont peu conservés entre les espèces, voire spécifiques à certains clades ou individus. En effet, les variants génétiques structurels sont moins fréquents dans les séquences d'ADN codantes (CDS) et les introns par rapport à d'autres régions d'ADN telles que les gènes non codants et les régions intergéniques chez l'homme [50]. De même, chez *Arabidopsis thaliana*, le projet 1001 Génomes montre que les régions codantes accumulent moins de *Single Nucleotide Polymorphism* (SNP) que les régions non codantes ou intergéniques [51]. De même, dans 66 accessions de riz, moins de 5 % des SNP et des insertions/délétions (indels) étaient situés dans les régions codantes alors qu'ils représentaient environ 10 % du génome [52]. Seulement moins de 2 % des lncRNA d'*Arabidopsis thaliana* sont conservés au niveau de la séquence dans le règne végétal. De manière concomitante, il a été observé que les gènes for-



tement exprimés ont tendance à être plus conservés que les gènes faiblement ou spécifiquement exprimés, tels que les lncRNAs [53]. Logiquement, les 18 % de lncRNA des *Brassicaceae* conservés au niveau de la séquence présentent un niveau d'expression plus élevé par rapport aux lncRNA non conservés [54]. L'analyse du génome de cinq espèces de monocotylédones et de cinq espèces de dicotylédones montre que la conservation des lncRNA reste élevée au sein d'une même espèce mais diminue fortement au niveau inter-espèces [55]. Même entre écotypes, le paysage des ARN non codants a montré une variabilité au niveau de l'expression indépendamment des SNP mineurs entre les accessions [56]. Ainsi, il est tentant de supposer que le génome non conservé reflète l'adaptation spécifique d'une accession ou d'un individu à son environnement [56, 57]. Il convient de noter que la séquence du gène de l'ARNnc peut ne pas être conservée bien que sa position relative (synténie) au sein du génome puisse l'être [58], ce qui suggère que l'interaction de l'ARNnc avec ses gènes codants synténiques voisins peut avoir une pertinence biologique pour leur corégulation.

#### 4. Expression et localisation subcellulaire des lncRNAs

Plutôt que de posséder des domaines spécifiques reconnaissables au cours de l'évolution comme les protéines fonctionnelles, les lncRNAs hébergent des séquences spécifiques et des structures secondaires qui peuvent façonner leur structure 3D et affecter leur interaction avec d'autres molécules. Les éléments interacteurs (IE) servent aux interactions physiques avec divers partenaires par la complémentarité des bases (avec d'autres acides nucléiques) et la reconnaissance spécifique de la séquence par les protéines de liaison à l'ARN (RBP). D'autre part, les éléments structuraux (SEs) permettent la formation de structures secondaires et/ou 3D des lncRNA, dirigeant leurs interactions fonctionnelles avec d'autres partenaires cellulaires. Les domaines structuraux contiennent à la fois des IEs et des SEs dans diverses combinaisons et permettent des interactions avec les complexes RBP [59]. L'association de ces différents éléments pourrait représenter l'un des langages qui servent à diriger les interactions des lncRNA. À titre d'exemple, le lncRNA Xist possède 33 régions qui forment des structures secondaires bien

définies reliées par des régions structurellement variables, dont un élément A-répétitif conservé. Cette région forme une structure inter-répétitive qui est essentielle pour son contrôle de l'inactivation du chromosome X [60].

La compréhension du répertoire des ARN nécessite encore des efforts supplémentaires pour saisir les signaux sous-jacents qui permettent aux lncRNAs d'exercer leurs fonctions. Néanmoins, quelques études ont tenté d'associer des domaines protéiques bien connus à leurs homologues ARN [61]. L'un de ces signaux consiste en l'information guidant leur localisation subcellulaire. En effet, les lncRNAs peuvent soit être exportés vers le cytosol, soit résider dans le noyau, contribuant dans les deux cas à divers processus cellulaires [62]. Les lncRNA sont généralement plus enrichis dans le noyau que les ARNm, probablement en raison de leur épissage moins efficace et de leurs interactions avec les RBP nucléaires. On a découvert un motif d'ARN qui reconnaît la petite ribonucléoprotéine nucléaire (snRNP) U1 et qui est essentiel pour mobiliser les lncRNAs vers la chromatine, les retenant ainsi dans le noyau [63]. De même, le lncRNA BORG présente un motif d'ARN pentamère qui est essentiel pour la rétention nucléaire. La mutation de ce motif en une séquence brouillée a entraîné la perte de la localisation nucléaire. Inversement, l'ajout d'une seule copie du motif dans un ARN cytoplasmique était suffisant pour induire sa rétention dans le noyau [64]. Ces domaines spécifiques de l'ARN pourraient donc représenter les homologues ARN du signal de localisation nucléaire (NLS) bien décrit présent dans les protéines. La caractérisation de ces signaux ARN sera d'une importance majeure pour le développement de nouveaux outils de biologie moléculaire et l'étude des mécanismes d'action des lncRNAs.

Outre la régulation de leur localisation subcellulaire, l'expression des lncRNAs au niveau tissulaire est également étroitement régulée. Il est intéressant de noter que les lncRNA sont généralement plus exprimés de manière spécifique à un tissu que les gènes codants, tant chez les plantes que chez les animaux [65, 66]. Par exemple, plus de 30 % des lncRNA d'*Arabidopsis* présentent un modèle d'expression spécifique à un organe ou à un développement [65]. Le nombre croissant de lncRNAs découverts, leurs séquences diverses et leur haute spécificité tissulaire suggèrent que les lncRNAs pourraient

servir de marqueurs de tissus et de stades de développement, et qu'ils pourraient être à la base de la grande diversité des modèles d'expression régulateurs au cours de l'évolution pour générer la grande variabilité des formes et des fonctions tout au long du développement et des réponses au stress chez les eucaryotes.

## 5. Les lncRNA dans la régulation de l'expression des gènes

Les lncRNAs sont apparus comme d'importants régulateurs de l'expression des gènes tant au niveau transcriptionnel que post-transcriptionnel. Nous allons maintenant concentrer la partie suivante sur les mécanismes impliquant les longs ARN non codants végétaux, même si nous citerons quelques articles sur les cellules animales en relation avec ces mécanismes, car plusieurs concepts sont communs et donnent une importance majeure à la généralité de ces mécanismes dans l'évolution. Néanmoins, il existe d'excellentes revues récentes pour décrire la pléthore d'informations sur les longs ARN non codants dans les cellules animales [67]. Dans la vaste bibliographie, elles se concentrent principalement sur les lncRNAs animaux et font brièvement référence aux lncRNAs végétaux.

La transcription implique la coopération entre la machinerie de l'ARN polymérase, les TF et d'autres complexes. Parmi ces derniers, le complexe MEDIATOR est le médiateur de la communication entre les TF et la Pol II, influençant le taux de transcription (Allen et Taatjes, 2015). De manière intéressante, chez *Arabidopsis*, le lncRNA ELF18-INDUCED LONG-NONCODING RNA1 (ELENA1) est capable de se lier à MED19a, une sous-unité du complexe MEDIATOR, et de moduler son recrutement vers des promoteurs spécifiques [68]. Une attaque pathogène déclenche l'accumulation d'ELENA1 qui se lie à la région promotrice du gène PATHOGENESIS-RELATED1 (PR1) permettant de renforcer le recrutement de MED19a dans cette région activant l'expression de PR1. De même, il a été démontré que les lncRNA humains PANDA et DHFR interagissent directement avec les TF afin de moduler leur liaison aux promoteurs des gènes, influençant directement la transcription de ces cibles reconnues par les lncRNA [69]. Les lncRNA peuvent également agir négativement sur la régulation de la transcription des

gènes cibles. Chez *Arabidopsis*, le lncRNA HIDDEN TREASURE1 (HID1) diminue l'activité transcriptionnelle du PHYTOCHROME-INTERACTING FACTOR 3 (PIF3) en se liant à son premier intron. La régulation négative de l'expression de HID1 augmente l'activité du gène PIF3 et l'élongation subséquente de l'hypocotyle. Le mécanisme par lequel HID3 diminue l'expression du gène n'est pas encore clair, cependant HID1 forme un complexe ribonucléoprotéique interagissant avec la région génomique de PIF3, suggérant que ce recrutement pourrait entraver la progression de Pol II le long du gène PIF3 [70].

Lorsque deux gènes orientés dans la direction opposée se chevauchent, ils sont généralement considérés comme des ARN antisens et, dans de nombreux cas, l'un des membres de ce couple est un lncRNA. Cependant, deux gènes ne peuvent pas être transcrits au même moment, sinon les deux complexes Pol II se rencontreront et leurs progressions respectives sur l'ADN seront bloquées. Chez *Arabidopsis*, le lncRNA antisens SVALKa est un bel exemple où ce processus contrôle son gène codant pour une protéine voisine, CBF1, impliqué dans la tolérance au froid. Une exposition prolongée au froid déclenche une lecture transcriptionnelle de la région génomique de SVALKa, ce qui entraîne une collision Pol II entre les transcriptions de SVALKa et de CBF1. Ceci a pour effet de réguler à la baisse CBF1 et d'affiner la réponse des plantes à ce stress [71]. En accord, chez la levure, la collision Pol II d'ARN antisens naturels conduit à l'élimination des deux polymérases Pol II de l'ADN par protéolyse [72]. Une explication alternative à la coexistence de la transcription dans les deux brins sur la même région peut être que chaque membre du couple est exprimé dans des cellules différentes comme cela a été montré pour les ARN antisens spécifiques [73]. Dans ce cas, le rôle du lncRNA est de contrôler spatialement l'expression des gènes. Dans l'ensemble, les lncRNA peuvent modifier l'efficacité transcriptionnelle d'un gène par le biais d'une interaction avec des protéines liées à la transcription ou par le biais d'une lecture transcriptionnelle conduisant à des sorties transcriptionnelles différentes.

### 5.1. Les lncRNA dans la régulation de la chromatine

Il existe un grand nombre de preuves que les lncRNA peuvent interagir physiquement avec des protéines

capables de modifier le paysage épigénétique des cellules. L'un des exemples les mieux étudiés est la régulation épigénétique par les lncRNA du régulateur floral FLC [74]. Le gène FLOWERING LOCUS C (FLC) est un MADS TF impliqué dans la transition vers la floraison par un commutateur épigénétique complexe impliquant probablement l'action de trois lncRNAs. Tout d'abord, COOLAIR, un transcrit antisens du gène FLC, interagit physiquement avec la région 5' de FLC et est médiateur, avec FLOWERING LOCUS D (FLD, une histone déméthylase), de la déméthylation de H3K4me2 dans le locus résultant en la répression transcriptionnelle de FLC. De plus, le lncRNA intronique FLC COLD ASSISTED INTRONIC NONCODING RNA (COLDAIR) interagit avec la sous-unité PRC2 de CLF pour renforcer la répression épigénétique du gène FLC par le dépôt de H3K27me3 [75]. Enfin, un autre lncRNA, COLDWRAP provenant du promoteur de FLC est également capable d'interagir avec le complexe PRC2 et module l'extinction de FLC dans différents écotypes [76]. Un autre exemple chez *Arabidopsis* d'un tel mécanisme a été lié à l'ARN INTRONIQUE AGAMIQUE 4 (AG-incRNA4). Ce lncRNA se lie à la sous-unité CLF du complexe PRC2 et est codé dans la première intro du gène AGAMOUS (AG) qui code pour un MADS TF impliqué dans le développement des fleurs. Chez les mutants PRC2, une réduction de la marque répressive H3K27me3 sur la chromatine d'AG a été détectée avec une induction concomitante d'AG. Il est intéressant de noter que le lncRNA intronique AG-incRNA4 semble lié au recrutement des complexes PRC2 à ce locus, ce qui conduit à sa propre extinction et établit une boucle de rétroaction, car la production du lncRNA nécessite l'expression de AG, ce qui entraîne la répression du gène [77]. Des mécanismes similaires ont également été analysés chez d'autres plantes comme le riz. L'expression du groupe de gènes LRKs impliqués dans le rendement en grains du riz est modulée par un lncRNA antisens, appelé LRK ANTISENSE INTERGENIC RNA (LAIR), transcrit à partir d'une région à l'intérieur du gène LRK1. Ce lncRNA est capable de recruter les OsWDR5 (WD REPEAT DOMAIN5), impliqués dans l'acétylation de H4K16, de la région génomique LRK, augmentant ainsi leur expression. À l'inverse, la surexpression du lncRNA LAIR augmente drastiquement le rendement en grains chez le riz [78] et réduit la présence de cette modification d'histone dans cette région.

Les lncRNAs peuvent également moduler l'état de la chromatine par la méthylation de l'ADN et la production de siRNAs. Dans ce cas, la région lncRNA peut être transformée en siRNAs spécifiques capables de déclencher l'extinction des gènes par méthylation de l'ADN via le mécanisme RdDM [14]. Par exemple, chez *Arabidopsis*, un stimulus auxin déclenche une forte régulation à la hausse de la protéine kinase PID liée au PIN, qui régule le gravitropisme racinaire, et de son lncRNA voisin AUXIN-REGULATED PROMOTER LOOP (APOLO). Cette induction par l'auxine est médiée par la perturbation d'une boucle chromatinienne englobant la région APOLO et le promoteur PID, probablement via un processus actif de déméthylation de l'ADN. La déméthylation du double promoteur PID-APOLO par l'auxine ouvre cette boucle chromatinienne et permet l'expression des deux gènes. Progressivement, les transcrits APOLO dépendants de Pol II recrutent la protéine LHP1 et la machinerie RdDM pour reformer la boucle chromatinienne et renforcer la compaction de la boucle chromatinienne [79]. Ensuite, les marques répressives et la méthylation de l'ADN de cette région maintiennent le silençage des régions APOLO-PID. Comme le statut épigénétique et la condensation de la chromatine influencent la topologie du génome, il a été proposé que les lncRNAs liés à la chromatine soient des acteurs de la configuration spatiale 3D du génome. En effet, au cours de la dernière décennie, de nombreuses preuves ont lié les lncRNA à la modulation de la conformation de la chromatine chez différentes espèces [80]. Ceci a été corrélé avec des expériences classiques montrant que le traitement des noyaux par la RNase A entraîne une décondensation globale de la chromatine [81]. Chez les plantes, le lncRNA COLDWRAP mentionné précédemment, transcrit dans le gène répresseur de la floraison FLC, agit comme médiateur de la formation d'une boucle chromatinienne intragénique répressive qui bloque la transcription de Pol II, inhibant l'expression de FLC et permettant à la plante de fleurir [76]. D'autre part, il a été démontré que le lncRNA APOLO agit également en trans, par la formation de duplex ARN : ADN appelés R-loops. Les boucles R contenant l'ARN APOLO sont capables de déplacer LHP1 à des loci distants et de moduler leur conformation chromatinienne [82]. Il est intéressant de noter qu'une proportion significative de loci génomiques distants reconnus par l'ARN APOLO

sont des gènes sensibles à l'auxine. Parmi eux, la LEUCINE RICH EXTENSIN2 (LRX2), impliquée dans le remodelage de la paroi cellulaire lors de l'émergence des racines latérales, et ROOT HAIR DEFECTIVE 6 (RHD6), un régulateur clé de l'initiation du cheveu racinaire [83, 84].

## 5.2. Les lncRNAs médiateurs de la régulation post-transcriptionnelle de l'expression génétique

En plus des mécanismes de régulation impliquant les petits ARN (par exemple, le mimétisme de cible ou la production de siRNA), des mécanismes alternatifs pour le contrôle post-transcriptionnel de l'expression des gènes par les lncRNA ont récemment émergé. Ces lncRNAs ciblent la machinerie d'épissage et cette interaction souligne le potentiel des lncRNAs à contribuer à la génération de la diversité protéique par la régulation de l'épissage alternatif (AS). Chez les plantes, il a été démontré que le lncRNA *Alternative Splicing COmpetitor* (ASCO) d'*Arabidopsis* interagit physiquement avec les protéines de liaison à l'ARN du spectre nucléaire (NSRs), qui sont des régulateurs de l'épissage alternatif au cours des processus de développement. En surexprimant uniquement le lncRNA ASCO dans les plantes, de nombreuses cibles épissées alternativement par les NSR ont été modifiées, montrant que les interactions ASCO-NSRs affectent l'AS de plusieurs gènes liés à l'auxine et influencent la formation de racines latérales pilotée par l'auxine chez *Arabidopsis* [85, 86]. Plus récemment, on a découvert que le lncRNA ASCO interagissait également avec d'autres composants centraux du spliceosome (SmD1b et PRP8a, composants participant directement aux réactions d'épissage), ce qui suggère que l'interaction dynamique et complexe des lncRNA et des composants du spliceosome, tels que les UsnRNA, peut intervenir dans la régulation de cibles spécifiques [87]. Plusieurs études fournissent des indices sur le rôle potentiel des lncRNA dans le contrôle de l'AS dans les cellules animales et trois types de régulation semblent émerger : (a) remodelage de la chromatine et hybrides ARN-ADN, (b) formation d'hybrides lncRNA-ARN et (c) altération des protéines des facteurs d'épissage (SFs) ou interaction des cibles de l'ARNm de l'AS avec le spliceosome pour affecter directement les réactions d'épissage.

### 5.2.1. LncRNAs affectant le remodelage de la chromatine dans le contexte de l'épissage

En plus de la régulation de la conformation de la chromatine et du statut épigénétique par les lncRNA mentionnée précédemment, on a découvert que le contexte chromatinien d'un locus, notamment dans le corps du gène, affecte également le SA. Comme l'épissage se produit principalement de manière co-transcriptionnelle, la modulation de la vitesse d'élongation de Pol II par le contexte chromatinien permet d'affiner le choix des sites d'épissage alternatifs [88]. Ainsi, comme les lncRNA peuvent affecter le remodelage de la chromatine, il semble plausible que certains lncRNA interagissant avec la chromatine, notamment dans les corps des gènes, puissent indirectement affecter l'épissage alternatif. Chez les plantes, ce mécanisme a été récemment analysé en découvrant le rôle du circRNA non codant SEP3 dans la modulation de l'AS de son propre gène [89]. SEP3 est un membre de la superfamille des boîtes MADS (MCM1-AGAMOUS-DEFICIENS-SRF) et il a été démontré qu'il participe au développement des fleurs. Les défauts d'épissage de SEP3 provoquent des phénotypes homéotiques floraux, ce qui souligne son importance dans ce processus de développement. De manière surprenante, la surexpression d'un circRNA contenant l'exon 6 entier du gène SEP3 a conduit à l'accumulation du variant AS SEP3.3 dépourvu de cet exon 6. Il a également été démontré que le circRNA de l'exon 6 du gène SEP3 peut interagir directement avec son locus d'ADN correspondant par la formation de boucles R (hybrides ARN-ADN). Cette boucle R favorise la pause transcriptionnelle sur son propre gène et affecte le recrutement des facteurs d'épissage [89]. Ce mécanisme spécifique suggère que les circRNAs peuvent participer à la régulation AS de leurs ARN messagers cognés à exon sauté. Cependant, il ouvre également la possibilité que certains lncRNAs ne codent que des circRNAs agissant dans l'épissage en trans. Il est intéressant de noter que chez les animaux, de nombreux circRNA sont codés dans des régions génomiques intergéniques [23]. D'autres exemples chez les animaux suggèrent que les lncRNA modulent le SA par le biais d'interactions chromatinienne. Par exemple, le lncRNA asFGFR2 (pour antisense) est transcrit à partir du locus FGFR2 humain et induit un SA spécifique de l'épithélium de FGFR2 [90]. Cette AS est permise

par le recrutement de modificateurs de chromatine spécifiquement pour son propre locus, conduisant à des changements dans le statut de méthylation de l'ADN de cette région. La présence de asFGFR2 assure le dépôt de la marque H3K27me3 et la diminution des marques H3K36me2/3, ce qui entraîne une altération du recrutement de la protéine de liaison à la chromatine MRG15 et du régulateur d'épissage négatif PTBP1. Le complexe MRG15-PTBP1 ne peut plus inhiber l'inclusion de l'exon IIIb dans le FGFR2, ce qui conduit à l'AS spécifique de l'épithélium de ce gène. Par conséquent, les lncRNA peuvent moduler la conformation de la chromatine soit en se liant directement à l'ADN (par le biais de boucles R), soit en recrutant des régulateurs épigénétiques spécifiques à des loci génomiques qui peuvent présenter des schémas de SA changeants dans différentes cellules ou en réponse à des stress environnementaux.

### 5.2.2. *Hybrides LncRNA-ARN en tant que régulateurs AS*

Indépendamment de leur origine génomique, les transcrits antisens naturels ou NATs, peuvent s'hybrider avec des pré-ARNm et former des duplex ARN-ARN. Il a été démontré que ces complexes ont un impact sur la SA au cours de divers processus biologiques, notamment dans les cellules de mammifères [85]. Une des principales voies apoptotiques chez les animaux est activée par l'interaction entre le récepteur Fas (Fas) et le ligand Fas (FasL) [91]. Au locus FAS, le lncRNA antisens SAF est transcrit dans une orientation inverse et à partir du brin opposé du premier intron de FAS. SAF se localise dans le noyau où il se lie au pré-mRNA du récepteur Fas et au SPF45 (SPLICING FACTOR 45) humain. Cette interaction facilite l'AS et l'exclusion de l'exon 6, conduisant à la production d'une protéine Fas soluble qui protège les cellules contre l'apoptose induite par FasL [91]. Ainsi, les transcriptions NAT peuvent faciliter les interactions entre les pré-ARNm et les FS spécifiques à la fois en cis (sur sa propre transcription) ou en trans alors qu'elles peuvent également masquer des sites d'épissage spécifiques pour empêcher leur traitement par le spliceosome. Chez *A. thaliana*, la fréquence accrue de transcrits alternativement épissés et variablement polyadénylés lorsqu'un intron chevauche un NAT suggère que la formation de paires lncRNA-ARN NAT peut réguler l'AS des gènes codant pour des protéines [92]. De manière

cohérente, un criblage à l'échelle du génome des trans-NAT chez *A. thaliana* a permis d'identifier 1320 paires trans-NAT putatives [93]. Bien que certains trans-NAT aient été prédits pour produire des ARN double brin et des siRNA (appelés natsiRNA) conduisant à l'ARN silencieux de leur cible sens, l'étude a montré que de nombreuses paires trans-NAT ont une proportion beaucoup plus élevée d'événements AS par rapport à toutes les unités de transcription du génome. Ainsi, les lncRNA peuvent profiter de la similarité de séquence pour se lier à des transcriptions spécifiques et moduler leur épissage.

### 5.2.3. *LncRNAs modifiant l'activité des facteurs d'épissage et les interactions avec le spliceosome*

Le spliceosome est un complexe ribonucléoprotéique géant où les facteurs d'épissage, les petits ARN nucléaires et les ARNm cibles se rencontrent pour définir le résultat de l'épissage d'un gène. La modulation de l'activité des protéines SF peut se faire de plusieurs manières : en favorisant la modification post-traductionnelle de la protéine, en changeant sa localisation subcellulaire, ou en ayant un impact sur la liaison de la protéine à d'autres partenaires ou cibles de transcription. Parmi les lncRNA les mieux caractérisés et associés à la régulation des SA, on compte le NUCLEAR PARASPECKLE ASSEMBLY TRANSCRIPT 1 (NEAT1) et le METASTASIS ASSOCIATED LUNG ADENOCARCINOMA TRANSCRIPT 1 (MALAT1) / NUCLEAR PARASPECKLE ASSEMBLY TRANSCRIPT 2 (NEAT2). Il a été démontré que ces deux lncRNA modulent la localisation et l'état de phosphorylation de FS spécifiques, et qu'ils présentent une expression différentielle dans un large éventail de tissus humains et murins [94]. NEAT1 est un lncRNA très abondant que l'on trouve dans les paraspeckles, des domaines nucléaires contrôlant la séquestration des protéines liées à l'épissage. Au cours de la différenciation des adipocytes, l'abondance de NEAT1 est régulée dynamiquement pour moduler les niveaux relatifs des isoformes de l'ARNm PPAR $\gamma$ , le principal TF dirigeant l'adipogenèse. En bref, il a été démontré que NEAT1 interagit avec la protéine SRp40 (SFRS5), entraînant la rétention de SRp40 dans les corps paranucléaires. L'interaction NEAT1-SRp40 augmente la phosphorylation de SRp40 par la CDC2-LIKE KINASE 1 (CLK1), une kinase ciblant spécifiquement les SFs. Cette modification de la phosphorylation de SRp40

favorise l'AS PPAR $\gamma$ , et donc le réglage fin du processus d'adipogenèse [95].

Le lncRNA MALAT1/NEAT2 se localise dans les mouchetures nucléaires et exerce des rôles oncogènes dans la cellule. Comme NEAT1, MALAT1/NEAT2 peut moduler efficacement la distribution et la phosphorylation de SF, conduisant à des changements de SA dans leurs pré-ARNm cibles [96]. Chez les plantes, la caractérisation des lncRNAs directement impliqués dans la SA en est encore à ses débuts. Néanmoins, quelques exemples montrant les interactions des lncRNA avec les AS émergent lentement. Chez la plante légumineuse *Medicago truncatula*, le lncRNA ENOD40 est rapidement induit lors de l'interaction avec les bactéries rhizobiennes symbiotiques. Il est exprimé dans le péricycle de la racine et dans les cellules de différenciation des primordia des nodules [97]. La surexpression de ENOD40 conduit à une nodulation accélérée, principalement causée par une initiation accrue des primordia et une sensibilité accrue aux signaux de nodulation. ENOD40 s'est avéré être hautement structuré et ne s'est pas associé aux polysomes. Des essais de levure à trois hybrides ont révélé une interaction directe entre ENOD40 et la protéine constitutive de liaison à l'ARN 1 (RBP1), qui se localise dans les mouchetures nucléaires où la machinerie d'épissage est également hébergée [98]. Pendant la nodulation, RBP1 est relocalisée dans les granules cytoplasmiques par son association avec ENOD40. Par conséquent, le lncRNA ENOD40 hautement structuré contribue au trafic nucléocytoplasmique de RBP1, ce qui suggère que le rôle de RBP1 dans le noyau peut être perturbé pendant le développement du nodule. MtRBP1 est un homologue proche des facteurs d'épissage AtNSRs; il est connu pour réguler la SA et il a été démontré que des changements majeurs dans la SA pendant la différenciation des nodules sont concomitants à l'expression de ENOD40 [99]. Bien qu'un effet direct d'ENOD40 sur les SA n'ait pas encore été démontré chez *Medicago truncatula*, la relocalisation d'un FS dans le cytoplasme peut affecter les modèles de SA.

Comme mentionné précédemment, il a été démontré que la surexpression du lncRNA nommé ASCO modulait l'AS des cibles de l'ARNm NSR pendant le développement des racines latérales chez *A. thaliana* [86]. Identifié pour la première fois par Ben Amor *et al.* [26] en utilisant une analyse bio-informatique à l'échelle du génome des bases de

données d'ADNc complets d'*Arabidopsis*, ASCO est codé dans une région intergénique du chromosome 1 (AT1G67105). Il a une taille de 786 nt et aucune capacité évidente de codage de protéines et s'accumule dans le noyau. De manière cohérente, on n'a pas trouvé que l'ASCO était associé aux ribosomes [44]. Un alignement ClustalW et MUSCLE a identifié au moins 4 paralogues d'ASCO dans *A. thaliana*, et a suggéré une large conservation d'ASCO et de ses paralogues dans les écotypes d'*Arabidopsis* et même entre différentes espèces de *Brassicaceae*. Il est intéressant de noter que l'analyse détaillée de l'AS dans le double mutant *nsra/b* a mis en évidence un nombre important d'événements AS, notamment en réponse à l'auxine [100], qui a été observé de manière similaire dans les lignées surexprimant l'ASCO. De plus, des essais d'immunoprécipitation d'ARN utilisant des lignées NSR-GFP ont établi que les NSR se lient *in vivo* non seulement à leurs cibles ARNm alternativement épissées, mais aussi au lncRNA ASCO et à de nombreux autres lncRNA [101]. En utilisant des expériences *in vitro* pour évaluer la liaison des complexes contenant des NSR à leurs substrats d'ARNm, Bardou *et al.* ont montré que le lncRNA ASCO était capable de concurrencer hors du complexe l'ARNm cible endogène d'une manière très efficace et spécifique à la séquence. Cela a suggéré que l'ASCO peut déplacer les cibles d'AS d'un complexe contenant des NSR pour modifier leur modèle d'AS [26] et a proposé un modèle où le lncRNA ASCO détourne les facteurs d'épissage des NSR pour empêcher leur interaction avec ses propres cibles. Il reste à voir s'il s'agit d'un mécanisme plus général pour d'autres interactions SF-lncRNA, mais il est tentant de spéculer que ces « lncRNA concurrents pirates » peuvent agir comme des imitateurs d'intron à l'intérieur du spliceosome pour réguler leur fonction d'une manière spécifique à la cellule.

Dans l'ensemble, ces études suggèrent que les lncRNA intègrent un réseau d'épissage dynamique comprenant de nombreux SF, des états chromatiniens et leurs cibles pré-mRNA associées. Grâce à cette gamme variée d'interactions avec différents partenaires, les lncRNA exercent un contrôle sur la reprogrammation du transcriptome par le biais de l'AS chez les eucaryotes.

## 6. Remarques finales

Dans l'ensemble, les molécules lncRNA apparaissent comme de nouveaux éléments des réseaux de régulation contrôlant la croissance et le développement des plantes. En interagissant physiquement avec des complexes épigénétiques, des TF ou l'ADN lui-même, ils règlent avec précision l'activité transcriptionnelle de gènes codant pour des protéines cibles. Les variations de l'expression des lncRNA en réponse à des stimuli abiotiques ou biotiques externes, qui présentent une grande variabilité entre les couches cellulaires à l'intérieur des tissus, peuvent conduire à des schémas d'expression différents de leurs cibles au niveau cellulaire, améliorant ainsi des processus de différenciation spécifiques ou des réponses à des stimuli externes. En outre, les lncRNA interagissent avec des modulateurs clés de l'épissage et de la stabilité de l'ARNm et peuvent également influencer l'abondance de l'ARNm des gènes cibles au niveau post-transcriptionnel et diversifier les sorties protéiques d'un même gène. Globalement, la découverte de ces mécanismes de régulation dépendants des lncRNA ouvre de larges perspectives pour comprendre la diversité des morphologies et des capacités d'adaptation des plantes.

## Conflit d'intérêt

L'auteur n'a aucun conflit d'intérêt à déclarer.

## Remerciements

Je tiens à remercier Michel Caboche pour nos discussions intéressantes sur les ARN non codants (qui étaient alors perçus avec prudence par les généticiens) et son soutien à la recherche sur les ARN non codants dans les programmes de Génoplante. Je remercie également Adam Kondorosi pour ses encouragements et son mentorat dans le domaine de la biologie des ARN non codants à l'époque de son émergence. Je tiens également à remercier profondément les membres de mon groupe de recherche et tous mes doctorants au cours de ces années pour leur enthousiasme et leur passion dans cette quête de compréhension des ARN non codants, notamment Federico Ariel qui a été pendant de nombreuses

années un proche collaborateur. Je voudrais également remercier Hervé Vaucheret pour nos discussions fructueuses et inspirantes sur les ARN cellulaires « étranges », ainsi que pour les collaborations à long terme avec son équipe.

## References

- [1] F. Ariel, N. Romero-Barrios, T. Jégu, M. Benhamed, M. Crespi, "Battles and hijacks: Noncoding transcription in plants", *Trends Plant Sci.* **20** (2015), p. 362-371.
- [2] T. A. Elliott, T. R. Gregory, "What's in a genome? The C-value enigma and the evolution of eukaryotic genome content", *Philos. Trans. R. Soc. B* **370** (2015), article no. 20140331.
- [3] C. A. Thomas, "The genetic organization of chromosomes", *Annu. Rev. Genet.* **5** (1971), p. 237-256.
- [4] N. Gil, I. Ulitsky, "Regulation of gene expression by cis-acting long non-coding RNAs", *Nat. Rev. Genet.* **21** (2020), p. 102-117.
- [5] S. Filichkin, H. D. Priest, M. Megraw, T. C. Mockler, "Alternative splicing in plants: Directing traffic at the crossroads of adaptation and environmental stress", *Curr. Opin. Plant Biol.* **24** (2015), p. 125-135.
- [6] J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson *et al.*, "COSMIC: the catalogue of somatic mutations in cancer", *Nucleic Acids Res.* **47** (2019), p. D941-D947.
- [7] M. C. Wahl, C. L. Will, R. Lührmann, "The spliceosome: design principles of a dynamic RNP machine", *Cell* **136** (2009), p. 701-718.
- [8] E. T. Wang, R. Sandberg, S. Luo, I. Khrebukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, C. B. Burge, "Alternative isoform regulation in human tissue transcriptomes", *Nature* **456** (2008), p. 470-476.
- [9] S. Chaudhary, W. Khokhar, I. Jabre, A. S. N. Reddy, L. J. Byrne, C. M. Wilson, N. H. Syed, "Alternative splicing and protein diversity: Plants versus animals", *Front. Plant Sci.* **10** (2019), p. 1-14.
- [10] C. Merchante, A. N. Stepanova, J. M. Alonso, "Translation regulation in plants: an interesting past, an exciting present and a promising future", *Plant J.* **90** (2017), p. 628-653.
- [11] Sanchita, P. K. Trivedi, M. H. Asif, "Updates on plant long non-coding RNAs (lncRNAs): the regulatory components", *Plant Cell. Tissue Organ Cult.* **140** (2020), p. 259-269.
- [12] M. J. Axtell, "Classification and comparison of small RNAs from plants", *Annu. Rev. Plant Biol.* **64** (2013), p. 137-159.
- [13] X. Chen, "Small RNAs and their roles in plant development", *Annu. Rev. Cell Dev. Biol.* **25** (2009), p. 21-44.
- [14] A. T. Wierzbicki, T. Blevins, S. Swiezewski, "Long noncoding RNAs in plants", *Annu. Rev. Plant Biol.* **72** (2021), p. 245-271.
- [15] T. Yifhar, I. Pekker, D. Peled, G. Friedlander, A. Pistunov, M. Sabban, G. Wachsman, J. P. Alvarez, Z. Amselem, Y. Eshehed, "Failure of the tomato trans-acting short interfering RNA program to regulate AUXIN response FACTOR3 and ARF4 underlies the wiry leaf syndrome", *Plant Cell* **24** (2012), p. 3575-3589.
- [16] G. Guo, X. Liu, F. Sun, J. Cao, N. Huo, B. Wuda, M. Xin, Z. Hu, J. Du, R. Xia, V. Rossi, H. Peng, Z. Ni, Q. Sun, Y. Yao, "Wheat

- miR9678 affects seed germination by generating phased siRNAs and modulating abscisic acid/gibberellin signaling”, *Plant Cell* **30** (2018), p. 796-814.
- [17] J. M. Franco-Zorrilla, A. Valli, M. Todesco, I. Mateos, M. I. Puga, I. Rubio-Somoza, A. Leyva, D. Weigel, J. A. García, J. Paz-Ares, “Target mimicry provides a new mechanism for regulation of microRNA activity”, *Nat. Genet.* **39** (2007), p. 1033-1037.
- [18] Q. Du, K. Wang, C. Zou, C. Xu, W. X. Li, “The PILNCR1-miR399 regulatory module is important for low phosphate tolerance in maize”, *Plant Physiol.* **177** (2018), p. 1743-1753.
- [19] T. Wang, M. Zhao, X. Zhang, M. Liu, C. Yang, Y. Chen, R. Chen, J. Wen, K. S. Mysore, W. H. Zhang, “Novel phosphate deficiency-responsive long non-coding RNAs in the legume model plant *Medicago truncatula*”, *J. Exp. Bot.* **68** (2017), p. 5937-5948.
- [20] Z. Sun, K. Huang, Z. Han, P. Wang, Y. Fang, “Genome-wide identification of Arabidopsis long noncoding RNAs in response to the blue light”, *Sci. Rep.* **10** (2020), p. 1-10.
- [21] P. Borah, A. Das, M. J. Milner, A. Ali, A. R. Bentley, R. Pandey, “Long non-coding rnas as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants”, *Genes (Basel)* **9** (2018), p. 1-17.
- [22] Y. Tian, Y. Xing, Z. Zhang, R. Peng, L. Zhang, Y. Sun, “Bioinformatics analysis of key genes and circRNA-miRNA-mRNA regulatory network in gastric cancer”, *Biomed. Res. Int.* **2020** (2020), article no. 2862701.
- [23] L. Santer, C. Bär, T. Thum, “Circular RNAs: a novel class of functional RNA molecules with a therapeutic perspective”, *Mol. Ther.* **27** (2019), p. 1350-1363.
- [24] Y. Shao, J. Li, R. Lu, T. Li, Y. Yang, B. Xiao, J. Guo, “Global circular RNA expression profile of human gastric cancer and its clinical significance”, *Cancer Med.* **6** (2017), p. 1173-1180.
- [25] C. Y. Ye, L. Chen, C. Liu, Q. H. Zhu, L. Fan, “Widespread noncoding circular RNAs in plants”, *New Phytol.* **208** (2015), p. 88-95.
- [26] B. Ben Amor, S. Wirth, F. Merchan, P. Laporte, Y. d’Aubenton Carafa, J. Hirsch, A. Maizel, A. Mallory, A. Lucas, J. M. Deragon, H. Vaucheret, C. Thermes, M. Crespi, “Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses”, *Genome Res.* **19** (2009), no. 1, p. 57-69.
- [27] X. Zhao, J. Li, B. Lian, H. Gu, Y. Li, Y. Qi, “Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA”, *Nat. Commun.* **9** (2018), p. 1-12.
- [28] A. P. Gallart, A. H. Pulido, I. A. M. De Lagrán, W. Sanseverino, R. A. Cigliano, “GREENC: a wiki-based database of plant lncRNAs”, *Nucleic Acids Res.* **44** (2016), p. D1161-D1166.
- [29] M. F. Lin, I. Jungreis, M. Kellis, “PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions”, *Bioinformatics* **27** (2011), p. 275-282.
- [30] A. Li, J. Zhang, Z. Zhou, “PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k*-mer scheme”, *BMC Bioinform.* **15** (2014), p. 1-10.
- [31] L. Sun, H. Luo, D. Bu, G. Zhao, K. Yu, C. Zhang, Y. Liu, R. Chen, Y. Zhao, “Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts”, *Nucleic Acids Res.* **41** (2013), article no. e166.
- [32] L. Hu, Z. Xu, B. Hu, Z. J. Lu, “COME: A robust coding potential calculation tool for lncRNA identification and characterization based on multiple features”, *Nucleic Acids Res.* **45** (2017), p. 1-13.
- [33] Y. J. Kang, D. C. Yang, L. Kong, M. Hou, Y. Q. Meng, L. Wei, G. Gao, “CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features”, *Nucleic Acids Res.* **45** (2017), p. W12-W16.
- [34] A. A. Golicz, M. B. Singh, P. L. Bhalla, “The long intergenic noncoding RNA (lincRNA) Landscape of the soybean genome”, *Plant Physiol.* **176** (2018), p. 2133-2147.
- [35] I. Sarropoulos, R. Marin, M. Cardoso-Moreira, H. Kaessmann, “Developmental dynamics of lncRNAs across mammalian organs and species”, *Nature* **571** (2019), p. 510-514.
- [36] K. Mattioli, P. J. Volders, C. Gerhardinger, J. C. Lee, P. G. Maass, M. Melé, J. L. Rinn, “High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity”, *Genome Res.* **29** (2019), p. 344-355.
- [37] J. Li, C. Liu, “Coding or noncoding, the converging concepts of RNAs”, *Front. Genet.* **10** (2019), p. 1-10.
- [38] F. Bardou, F. Merchan, F. Ariel, M. Crespi, “Dual RNAs in plants”, *Biochimie* **93** (2011), p. 1950-1954.
- [39] F. Hubé, G. Velasco, J. Rollin, D. Furling, C. Francastel, “Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation”, *Nucleic Acids Res.* **39** (2011), p. 513-525.
- [40] S. M. Colley, P. J. Leedman, “Steroid receptor RNA activator—A nuclear receptor coregulator with multiple partners: Insights and challenges”, *Biochimie* **93** (2011), p. 1966-1972.
- [41] P. Xie, H. Chen, “Mechanism of ribosome translation through mRNA secondary structures”, *Int. J. Biol. Sci.* **13** (2017), p. 712-722.
- [42] S. W. Foley, M. C. Kramer, B. D. Gregory, “RNA structure, binding, and coordination in Arabidopsis”, *Wiley Interdiscip. Rev. RNA* **8** (2017), article no. e1426.
- [43] J. W. Nam, S. W. Choi, B. H. You, “Incredible RNA: Dual functions of coding and noncoding”, *Mol. Cells* **39** (2016), p. 367-374.
- [44] J. Bazin, K. Baerenfaller, S. J. Gosai, B. D. Gregory, M. Crespi, J. Bailey-Serres, “Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation”, *Proc. Natl. Acad. Sci. USA* **114** (2017), p. E10018-E10027.
- [45] F. Yeasmin, T. Yada, N. Akimitsu, “Micropeptides encoded in transcripts previously identified as long noncoding RNAs: A new chapter in transcriptomics and proteomics”, *Front. Genet.* **9** (2018), p. 1-10.
- [46] D. M. Anderson, K. M. Anderson, C. L. Chang, C. A. Makarewich, B. R. Nelson, J. R. McAnally, P. Kasaragod, J. M. Shelton, J. Liou, R. Bassel-Duby, E. N. Olson, “A micropeptide encoded by a putative long noncoding RNA regulates muscle performance”, *Cell* **160** (2015), p. 595-606.
- [47] M. Guttman, P. Russell, N. T. Ingolia, J. S. Weissman, E. S. Lander, “Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins”, *Cell* **154** (2013), p. 240-251.
- [48] N. T. Ingolia, G. A. Brar, N. Stern-Ginossar, M. S. Harris, G. J. S. Talhouarne, S. E. Jackson, M. R. Wills, J. S. Weissman, “Ribosome profiling reveals pervasive translation outside of



- annotated protein-coding genes”, *Cell Rep.* **8** (2014), p. 1365-1379.
- [49] G. M. Mustafa, D. Larry, J. R. Petersen, C. J. Elferink, “Targeted proteomics for biomarker discovery and validation of hepatocellular carcinoma in hepatitis C infected patients”, *World J. Hepatol.* **7** (2015), p. 1312-1324.
- [50] P. H. Sudmant, T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. H. Y. Fritz *et al.*, “An integrated map of structural variation in 2,504 human genomes”, *Nature* **526** (2015), p. 75-81.
- [51] The 1001 Genomes Consortium, “1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*”, *Cell* **166** (2016), p. 481-491.
- [52] Q. Zhao, Q. Feng, H. Lu, Y. Li, A. Wang, Q. Tian, Q. Zhan, Y. Lu, L. Zhang, T. Huang *et al.*, “Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice”, *Nat. Genet.* **50** (2018), p. 278-284.
- [53] B. Contreras-Moreira, C. P. Cantalapiedra, M. J. García-Pereira, S. P. Gordon, J. P. Vogel, E. Igartua, A. M. Casas, P. Vinuesa, “Analysis of plant pan-genomes and transcriptomes with GET\_HOMOLOGUES-EST, a clustering solution for sequences of the same species”, *Front. Plant Sci.* **8** (2017), p. 1-16.
- [54] A. D. L. Nelson, E. S. Forsythe, U. K. Devisetty, D. S. Clausen, A. K. Haug-Batzell, A. M. R. Meldrum, M. R. Frank, E. Lyons, M. A. Beilstein, “A genomic analysis of factors driving lincRNA diversification: Lessons from plants”, *G3 Genes Genomes Genet.* **6** (2016), p. 2881-2891.
- [55] P. Deng, S. Liu, X. Nie, S. Weining, L. Wu, “Conservation analysis of long non-coding RNAs in plants”, *Sci. China Life Sci.* **61** (2018), p. 190-198.
- [56] T. Blein, C. Balzergue, T. Roulé, M. Gabriel, L. Scalis, T. François, C. Sorin, A. Christ, C. Godon, E. Delannoy *et al.*, “Landscape of the non-coding transcriptome response of two *Arabidopsis* ecotypes to phosphate starvation”, *Plant Physiol.* **183** (2020), p. 1058-1072.
- [57] G. Vernikos, D. Medini, D. R. Riley, H. Tettelin, “Ten years of pan-genome analyses”, *Curr. Opin. Microbiol.* **23** (2015), p. 148-154.
- [58] S. Mohammadin, P. P. Edger, J. C. Pires, M. E. Schranz, “Positionally-conserved but sequence-diverged: Identification of long non-coding RNAs in the Brassicaceae and Cleomeaceae”, *BMC Plant Biol.* **15** (2015), p. 1-12.
- [59] M. Fabbri, L. Girmita, G. Varani, G. A. Calin, “Decrypting noncoding RNA interactions, structures, and functional networks”, *Genome Res.* **29** (2019), p. 1377-1388.
- [60] G. Pintacuda, A. N. Young, A. Cerase, “Function by structure: Spotlights on xist long non-coding RNA”, *Front. Mol. Biosci.* **4** (2017), article no. 90.
- [61] L. Goff, J. L. Rinn, “Linking RNA biology to lncRNAs”, *Genome Res.* **25** (2015), p. 1456-1465.
- [62] J. Carlevaro-Fita, R. Johnson, “Global positioning system: understanding long noncoding RNAs through subcellular localization”, *Mol. Cell* **73** (2019), p. 869-883.
- [63] Y. Yin, J. Y. Lu, X. Zhang, W. Shao, Y. Xu, P. Li, Y. Hong, L. Cui, G. Shan, B. Tian *et al.*, “U1 snRNP regulates chromatin retention of noncoding RNAs”, *Nature* **580** (2020), p. 147-150.
- [64] B. Zhang, L. Gunawardane, F. Niazi, F. Jahanbani, X. Chen, S. Valadkhan, “A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA”, *Mol. Cell. Biol.* **34** (2014), p. 2318-2329.
- [65] J. Liu, C. Jung, J. Xu, H. Wang, S. Deng, L. Bernad, C. Arenas-Huerta, N.-H. Chua, “Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*”, *Plant Cell* **24** (2012), p. 4333-4345.
- [66] L. C. Tsoi, M. K. Iyer, P. E. Stuart, W. R. Swindell, J. E. Gudjonsson, T. Tejasvi, M. K. Sarkar, B. Li, J. Ding, J. J. Voorhees *et al.*, “Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin”, *Genome Biol.* **16** (2015), article no. 24.
- [67] L. Statello, C. J. Guo, L. L. Chen, M. Huarte, “Gene regulation by long non-coding RNAs and its biological functions”, *Nat. Rev. Mol. Cell Biol.* **22** (2021), p. 96-118.
- [68] J. S. Seo, H. X. Sun, B. S. Park, C. H. Huang, S. D. Yeh, C. Jung, N. H. Chua, “ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in *Arabidopsis*”, *Plant Cell* **29** (2017), p. 1024-1038.
- [69] P. K. Puvvula, R. D. Desetty, P. Pineau, A. Marchio, A. Moon, A. Dejean, O. Bischof, “Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit”, *Nat. Commun.* **5** (2014), article no. 5323.
- [70] Y. Wang, X. Fan, F. Lin, G. He, W. Terzaghi, D. Zhu, X. W. Deng, “*Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light”, *Proc. Natl. Acad. Sci. USA* **111** (2014), p. 10359-10364.
- [71] P. Kindgren, R. Ard, M. Ivanov, S. Marquardt, “Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation”, *Nat. Commun.* **9** (2018), article no. 4561.
- [72] D. J. Hobson, W. Wei, L. M. Steinmetz, J. Q. Svejstrup, “RNA polymerase II collision interrupts convergent transcription”, *Mol. Cell* **48** (2012), p. 365-374.
- [73] E. Zubko, P. Meyer, “A natural antisense transcript of the *Petunia hybrida* Sho gene suggests a role for an antisense mechanism in cytokinin regulation”, *Plant J.* **52** (2007), p. 1131-1139.
- [74] F. Liu, S. Marquardt, C. Lister, S. Swiezewski, C. Dean, “Targeted 3' processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing”, *Science* **327** (2010), p. 94-97.
- [75] J. B. Heo, S. Sung, “Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA”, *Science* **331** (2011), p. 76-79.
- [76] D.-H. Kim, S. Sung, “Vernalization-triggered intragenic chromatin-loop formation by long noncoding RNAs”, *Dev. Cell* **176** (2017), p. 100-106.
- [77] H. W. Wu, S. Deng, H. Xu, H. Z. Mao, J. Liu, Q. W. Niu, H. Wang, N. H. Chua, “A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves”, *New Phytol.* **219** (2018), p. 1480-1491.
- [78] Y. Wang, X. Luo, F. Sun, J. Hu, X. Zha, W. Su, J. Yang, “Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice”, *Nat. Commun.* **9** (2018), p. 1-9.
- [79] F. Ariel, T. Jegu, D. Latrasse, N. Romero-Barrios, A. Christ,

- M. Benhamed, M. Crespi, "Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop", *Mol. Cell* **55** (2014), p. 383-396.
- [80] S. Quinodoz, M. Guttman, "Long non-coding RNAs: An emerging link between gene regulation and nuclear organization", *Trends Cell Biol.* **24** (2014), p. 651-663.
- [81] M. Caudron-Herger, K. Rippe, "Nuclear architecture by RNA", *Curr. Opin. Genet. Dev.* **22** (2012), p. 179-187.
- [82] F. Ariel, L. Lucero, A. Christ, M. F. Mammarella, T. Jegu, A. Veluchamy, K. Mariappan, D. Latrasse, T. Blein, C. Liu *et al.*, "R-Loop mediated trans action of the APOLO long noncoding RNA", *Mol. Cell* **77** (2020), p. 1-11.
- [83] Q. Lin, Y. Ohashi, M. Kato, T. Tsuge, H. Gu, L. J. Qu, T. Aoyama, "GLABRA2 directly suppresses basic helix-loop-helix transcription factor genes with diverse functions in root hair development", *Plant Cell* **27** (2015), p. 2894-2906.
- [84] M. Moison, J. M. Pacheco, L. Lucero, C. Fonouni-Farde, J. Rodríguez-Melo, N. Mansilla, A. Christ, J. Bazin, M. Benhamed, F. Ibañez *et al.*, "The lncRNA APOLO interacts with the transcription factor WRKY42 to trigger root hair cell expansion in response to cold", *Mol. Plant* **14** (2021), p. 937-948.
- [85] N. Romero-Barrios, M. F. Legascue, M. Benhamed, F. Ariel, M. Crespi, "Splicing regulation by long noncoding RNAs", *Nucleic Acids Res.* **46** (2018), p. 2169-2184.
- [86] F. Bardou, F. Ariel, C. G. Simpson, N. Romero-Barrios, P. Laporte, S. Balzergue, J. W. S. Brown, M. Crespi, "Long non-coding RNA modulates alternative splicing regulators in Arabidopsis", *Dev. Cell* **30** (2014), p. 166-176.
- [87] R. Rigo, J. Bazin, N. Romero-Barrios, M. Moison, L. Lucero, A. Christ, M. Benhamed, T. Blein, S. Huguet, C. Charon *et al.*, "The Arabidopsis lnc RNA ASCO modulates the transcriptome through interaction with splicing factors", *EMBO Rep.* **21** (2020), p. 1-19.
- [88] R. F. Luco, M. Allo, I. E. Schor, A. R. Kornblihtt, T. Misteli, "Epigenetics in alternative pre-mRNA splicing", *Cell* **144** (2011), p. 16-26.
- [89] V. M. Conn, V. Hugouvieux, A. Nayak, S. A. Conos, G. Capovilla, G. Cildir, A. Jourdain, V. Tergaonkar, M. Schmid, C. Zubieta *et al.*, "A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation", *Nat. Plants* **3** (2017), p. 4-8.
- [90] I. Gonzalez, R. Munita, E. Agirre, T. A. Dittmer, K. Gysling, T. Misteli, R. F. Luco, "A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature", *Nat. Struct. Mol. Biol.* **22** (2015), p. 370-376.
- [91] O. Villamizar, C. B. Chambers, J. M. Riberdy, D. A. Persons, A. Wilber, "Long noncoding RNA Saf and splicing factor 45 increase soluble Fas and resistance to apoptosis", *Oncotarget* **7** (2016), p. 13810-13826.
- [92] C. H. Jen, I. Michalopoulos, D. R. Westhead, P. Meyer, "Natural antisense transcripts with coding capacity in Arabidopsis may have a regulatory role that is not linked to double-stranded RNA degradation", *Genome Biol.* **6** (2005), article no. R51.
- [93] H. Wang, N. H. Chua, X. J. Wang, "Prediction of trans-antisense transcripts in Arabidopsis thaliana", *Genome Biol.* **10** (2006), article no. 7.
- [94] J. L. Rinn, H. Y. Chang, "Long noncoding RNAs: molecular modalities to organismal functions", *Annu. Rev. Biochem.* **89** (2020), p. 283-308.
- [95] D. R. Cooper, G. Carter, P. Li, R. Patel, J. E. Watson, N. A. Patel, "Long non-coding RNA NEAT1 associates with SRp40 to temporally regulate PPAR $\gamma$ 2 splicing during adipogenesis in 3T3-L1 cells", *Genes (Basel)* **5** (2014), p. 1050-1063.
- [96] V. Tripathi, J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya *et al.*, "The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation", *Mol. Cell* **39** (2010), p. 925-938.
- [97] M. D. Crespi, E. Jurkevitch, M. Poirer, Y. d'Aubenton Carafa, G. Petrovics, E. Kondorosi, A. Kondorosi, "Enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth", *EMBO J.* **13** (1994), p. 5099-5112.
- [98] A. Campalans, A. Kondorosi, M. Crespi, "Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*", *Plant Cell* **16** (2004), p. 1047-1059.
- [99] L. Lucero, J. Bazin, J. Rodriguez Melo, F. Ibañez, M. D. Crespi, F. Ariel, "Evolution of the small family of alternative splicing modulators nuclear speckle RNA-binding proteins in plants", *Genes (Basel)* **11** (2020), article no. 207.
- [100] V. D. T. Tran, O. Souiai, N. Romero-Barrios, M. Crespi, D. Gautheret, "Detection of generic differential RNA processing events from RNA-seq data", *RNA Biol.* **13** (2016), p. 59-67.
- [101] J. Bazin, N. Romero, R. Rigo, C. Charon, T. Blein, F. Ariel, M. Crespi, "Nuclear speckle RNA binding proteins remodel alternative splicing and the non-coding arabidopsis transcriptome to regulate a cross-talk between auxin and immune responses", *Front. Plant Sci.* **9** (2018), p. 1-13.