



HAL
open science

Long Noncoding RNAs in Plants

Andrzej T Wierzbicki, Todd Blevins, Szymon Swiezewski

► **To cite this version:**

Andrzej T Wierzbicki, Todd Blevins, Szymon Swiezewski. Long Noncoding RNAs in Plants. Annual Review of Plant Biology, 2021, 72 (1), pp.245-271. 10.1146/annurev-arplant-093020-035446 . hal-03456471

HAL Id: hal-03456471

<https://hal.science/hal-03456471>

Submitted on 10 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author manuscript file - for proofreading

Annu. Rev. Plant Biol (2021) 72:X–X

Long Noncoding RNAs in Plants

Andrzej T. Wierzbicki,¹ Todd Blevins,² and Szymon Swiezewski³

¹Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109, USA; email: wierzbic@umich.edu

²Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, F-67084 Strasbourg, France; email: todd.blevins@ibmp-cnrs.unistra.fr

³Institute of Biochemistry and Biophysics, Polish Academy of Science, 02-106 Warsaw, Poland; email: sswiesz@ibb.waw.pl

Keywords

lncRNA, RNA polymerases, RNA-directed DNA methylation, gene regulation

Abstract

Plants have an extraordinary diversity of transcription machineries, including five nuclear DNA-dependent RNA polymerases. Four of these enzymes are dedicated to the production of long noncoding RNAs (lncRNAs), which are ribonucleic acids with functions independent of their protein-coding potential. lncRNAs display a broad range of lengths and structures, but they are distinct from the small RNA guides of RNA interference (RNAi) pathways. lncRNAs frequently serve as structural, catalytic, or regulatory molecules for gene expression. They can affect all elements of genes, including promoters, untranslated regions, exons, introns, and terminators, controlling gene expression at various levels including modifying chromatin accessibility, transcription, splicing, and translation. Certain lncRNAs protect genome integrity, while others respond to environmental cues like temperature, drought, nutrients, and pathogens. In this review, we explain the challenge of defining lncRNAs, introduce the machineries responsible for their production, and organize this knowledge by viewing the functions of lncRNAs throughout the structure of a typical plant gene.

1. INTRODUCTION

1.1. Discovery, Middle Ages, and Renaissance of Long Noncoding RNA

Ribonucleic acids (RNAs) have a complex relationship with genetic information. The discovery of protein biosynthesis mechanisms revealed that RNA relates to genetic information in at least three distinct ways ([104](#)). It can directly carry genetic information as messenger RNA (mRNA). It can perform structural and enzymatic roles as ribosomal RNA (rRNA). Or it can serve at the interface between genetic information and protein sequence as transfer RNA (tRNA). Although mRNA was not the first RNA discovered ([57](#)), it serves as the reference point from which all other RNA types are termed noncoding RNA (ncRNA), an ambiguous and negatively inflected definition. As their catalog expanded, ncRNAs were not merely seen as molecular fossils but rather as the new frontier for discovery. In this modern RNA world, numerous essential processes were understood to be ncRNA driven ([34](#)).

The eighteenth-century physicist Felice Fontana was the first to study a cellular structure made of ncRNA. While observing “vesicles” (cells) derived from eel epidermis, he discerned “an oval body, with a spot in its center” ([42](#)). Thanks to Emil Heitz and Barbara McClintock in the 1930s, Oscar Miller and Barbara Beatty in the 1960s, and contemporary cell and molecular biologists, we now understand Fontana’s oval to be the cell nucleus and the central spot to be the nucleolus. This nuclear body is the dynamic nexus of rRNA synthesis, rRNA processing, and preribosome assembly tethered to the nucleolus organizer regions in eukaryotic chromosomes ([96](#)).

Although the widespread presence of ncRNA has been known for a long time ([43a](#)), our knowledge about ncRNA exploded with the advent of technologies for genome-wide RNA detection. Microarrays and high-throughput sequencing in animals, fungi, and plants revealed that, beyond conventional gene loci, large portions of eukaryotic genomes are transcribed ([30](#), [33](#), [82](#), [143](#)). Large genomes appeared to be full of countless, previously unknown ncRNAs. However, the lack of evidence for functionality soon challenged the significance of this genome-wide transcription ([45](#)). Most transcribed loci that do not encode proteins lack sequence conservation and are thus unlikely to be under selection. Moreover, disrupting their expression or sequence often yields no evident phenotype. Even their existence is being questioned, as some of these transcripts may be technical artifacts or products of transcriptional noise ([73](#)).

Systematic functional studies of the last decade put a spotlight on the mechanisms and functions of ncRNA, especially on the longer subset termed long noncoding RNA (lncRNA). This

work revealed conclusive biological functions for a limited number of plant lncRNAs and explained certain aspects of their molecular mechanisms. These lncRNAs are the focus of this review. Before we proceed to describing their biosynthesis, mechanisms, and functions, we must clarify what we include in our definition of lncRNA.

1.2. Definition of Long Noncoding RNA

We propose the following definition: a lncRNA is an RNA that has a function independent of its protein-coding potential and that is produced by a mechanism other than molecular ruler-based dicing or trimming. Below we explain the three essential criteria of this revised definition and how it mitigates the shortcomings of past definitions used widely in the literature.

1.2.1. Functional RNA.

The first criterion we apply to define lncRNAs is the test of functionality. The mere presence of an RNA does not imply function because transcription and other mechanisms of RNA biosynthesis are inherently noisy (117). RNA transcripts that frequently occur at a locus may ultimately become functionalized, but the stochastic products of transcriptional noise are excluded from our definition of lncRNA. Admittedly, the act of transcription may sometimes be functional and the RNA just a by-product (17). Such transcripts are included in our definition because it is technically hard to distinguish the functional role of transcription from its transcript products. Finally, a large pool of RNA is detectable only in mutants defective in the RNA degradation machinery (66, 120). These RNAs may reflect RNA processing noise and are excluded from lncRNA, unless clear evidence of function exists. Exclusion from our definition does not imply that transcriptional noise and RNA processing noise are unworthy research subjects. Rather, we argue that these stochastic RNAs are quantitative evidence about the fidelity of the RNA biogenesis machinery, not units of cellular function.

Defining functionality is not trivial. In a strict definition, function is a product of selection (45). Sequence conservation is the easiest test of selection. However, conservation is also relatively rare: Only 3.8% of the *Arabidopsis* genome that does not encode proteins shows evidence of selection (50). An alternative test relies on genetic evidence and observing phenotypic consequences after altering the candidate RNA. To assign functionality, we require an RNA to meet one of these two tests. A lower bar for functionality is the causal role test, which relies on the RNA's involvement in a molecular process (45). RNAs that have been shown to be involved in a genetic or biochemical

pathway, but lack evidence of selection or have not been tested for phenotypic effect upon disruption, are termed putative lncRNAs. In many cases, altered phenotypes are seen only after lncRNA overexpression, rather than deficiency. While overexpression remains an important tool, it is vulnerable to technical artifacts. lncRNAs only validated using this latter method are also included in the putative lncRNA category.

1.2.2. Noncoding.

The second criterion we apply to define lncRNAs is that they have a function independent of any protein-coding potential. This criterion mitigates a shortcoming of the literal view of noncoding because a lack of encoded polypeptides is impossible to prove conclusively (60). We recognize here that an RNA molecule may combine both coding and noncoding functions (9, 76). When these functions are unrelated, the noncoding role clearly fits the definition of lncRNA. However, mRNAs also use a variety of mechanisms to affect their own expression. Examples of such mechanisms are sequence elements affecting RNA processing, signals that control subcellular localization, and riboswitches. Such autoregulatory features within mRNAs are directly related to their coding potential and do not fit the definition of lncRNA. Finally, our definition includes structural RNAs, in which the functional structure is encoded in the nucleotide sequence but is independent of the genetic code.

1.2.3. Long.

The third criterion we apply to define lncRNAs is their biogenesis mechanism, which is clearly distinct from that of small RNA biogenesis. This criterion is consistent with early uses of the term lncRNA (113) and avoids the need for an arbitrary size cutoff (traditionally at 200 nt), which has no justification in the origin or function of various lncRNAs. Small RNAs, in contrast, have a well-defined size range and conserved biogenesis mechanisms, which make them distinct from lncRNAs (8). Therefore, we define lncRNAs as products of biogenesis mechanisms other than the activity of Dicer endonucleases or exonuclease protection within Argonaute proteins. Our definition includes rRNA, tRNA, small nuclear RNA (snRNA), and various outlier classes of RNA (Table 1), which were previously categorized with little consistency (80, 98).

<COMP: PLEASE INSERT TABLE 1 HERE>

1.3. Unique Properties of lncRNA

Successful efforts have been made to systematically classify lncRNAs by their distinct properties (21, 106). Many of the proposed common features of lncRNAs that determine biological function closely parallel well-established attributes of proteins. In fact, lncRNAs resemble proteins in being structurally and functionally heterogeneous. At the same time, lncRNAs mostly lack common evolutionary origins and do not have universally conserved mechanisms of action. The category of lncRNA is thus extremely broad. Although our understanding of plant lncRNAs is still limited, the lack of universal lncRNA properties may be intrinsic to this class of RNAs.

The ability to combine genetic and structural capacities in one molecule distinguishes lncRNAs from proteins. A good example of this combination is tRNA, which has genetic information in the anticodon and adopts a conserved structure, which facilitates reversible binding of amino acids and association with the ribosome. Further examples of this combination are instructive because they reveal conceptually novel mechanisms of lncRNA action.

An important implication of these distinct lncRNA properties is that they expand our understanding of evolutionary conservation. In a typical scenario, conservation of sequence determines the conservation of structure and function, as in the case of the *Telomerase RNA subunit* (*AtTR*), which provides the RNA template for telomere DNA synthesis and is essential for catalysis in the ribonucleoprotein telomerase enzyme (37). However, structure and function may remain conserved even in the absence of strong sequence conservation, as suggested for the lncRNA *COOLAIR* (51). Moreover, functional conservation may occur even without sequence or structural similarity by virtue of a consistent location relative to a gene (39, 44, 92).

Plants have evolved the most diverse transcriptional machinery yet described for any taxonomic group of eukaryotes (123). This enzymatic machinery includes two novel nuclear DNA-dependent RNA polymerases that are dedicated to lncRNA synthesis. Given these specialized enzymes, the lncRNA production machinery is particularly complex in plants and thus deserves an in-depth introduction.

2. THE TRANSCRIPTIONAL MACHINERY RESPONSIBLE FOR lncRNA PRODUCTION

Eukaryotes deploy three conserved multisubunit RNA polymerases (Pol I, II, and III) to transcribe their nuclear genomes into various coding and noncoding transcripts (27). All three of these RNA

polymerases produce lncRNA: Pol I and Pol III are dedicated to lncRNA production, and Pol II, despite its primary function in mRNA biogenesis, also produces a range of lncRNAs. Remarkably, plants use two additional enzymes, Pol IV and Pol V, to produce lncRNAs that are essential for recognizing and silencing transposable elements (TEs).

Nuclear RNA polymerases assemble from 12 to 17 subunits, including unique combinations of the largest two subunits, which enclose the catalytic center ([27](#), [102](#), [103](#)). The plant nomenclature assigns equivalent numbers to the first 12 subunits, which are structurally and functionally homologous in Pol I, II, III, IV, and V (**Figure 1a**). Each transcription machinery in plants generates distinct lncRNAs (**Table 1**) that are governed by specialized transcription factors, the core enzyme's sequence specificity, its other biochemical properties, and associated RNA processing activities. Armed with deep sequencing, proteomics, and structure determination techniques, scientists are probing ever deeper into the enzyme complexes that synthesize lncRNA.

<COMP: PLEASE INSERT FIGURE 1 HERE>

Figure 1 Nuclear RNA polymerases and their diverse lncRNA products. (a) The subunit architectures of the five nuclear RNA polymerases in plants (Pol I, Pol II, Pol III, Pol IV, and Pol V) are shown on the basis of comparison to the structures of homologous yeast enzymes. Equivalent numbers are assigned to the first 12 subunits, which are structurally and functionally homologous. Common and enzyme-specific subunits are colored according to the published proteomics analyses of the affinity-purified *Arabidopsis thaliana* complexes ([102](#), [103](#)). Enzyme type is indicated by the last letter of each subunit name: NRPA for Pol I subunits, NRPB for Pol II subunits, NRPC for Pol III subunits, NRPD for Pol IV subunits, and NRPE for Pol V subunits. The Pol I subunit NRPA4 is white because the corresponding protein has not yet been identified in *A. thaliana*. (b) Typical RNA products of each plant RNA polymerase machinery are shown in red, with schematics that briefly summarize the features of their biogenesis and/or the mature RNA structure. lncRNAs are depicted in light blue boxes to the right of each enzyme. Pol II produces both protein-coding RNAs (*light brown*) and lncRNAs (*light blue*), which is a dual function unique to this transcriptional machinery. Abbreviations: AGO, ARGONAUTE; lncRNA, long noncoding RNA; RDR2, RNA-DEPENDENT RNA POLYMERASE 2; TE, transposable element.

2.1. Pol I

The most abundant lncRNA class in the cell is rRNA that catalyzes protein synthesis. To produce this rRNA, Pol I transcribes tandemly repeated genes into long rRNA precursors, which are processed into 18S, 5.8S, and 25S rRNAs (**Figure 1b**). These mature rRNAs, in addition to 5S rRNAs (see below), associate with ribosomal proteins to assemble into the small and large ribosomal subunits. Pol I transcription is thus coupled to ribosome biogenesis, which requires

further lncRNA-containing ribonucleoproteins and is beyond the scope of our review (110). A distinct property of Pol I is the tight coordination of its activity with cell growth and organismal development. Structurally, Pol I is distinguished by its catalytic subunits NRPA1 and NRPA2 and comprises several other fully or partially specialized proteins for a total of 14 subunits (102) (**Figure 1a**). The primary rRNA transcripts, their processed forms, and the assembled catalytic ribozyme are all lncRNA, according to our definition.

2.2. Pol II

While Pol II transcripts represent only 3% of the total steady-state RNA in a cell (129), Pol II transcription accounts for protein-coding transcripts (pre-mRNAs) and a significant subset of lncRNA (**Figure 1b**). A distinct feature of Pol II transcripts is the precise regulation of their production and activity accomplished at both the transcriptional and posttranscriptional levels. These regulatory mechanisms include general transcription factors (GTFs, or TFII complexes), promoter-specific factors, the Mediator complex, and RNA processing steps that impact transcription initiation, elongation, splicing, and termination (27).

Pol II assembles from largest and second-largest catalytic subunits (NRPB1 and NRPB2), as well as smaller polypeptides that are, in part, common to other RNA polymerases (**Figure 1a**). These plant proteins form a 12-subunit Pol II complex, homologous to the archetypal yeast Pol II (27, 103). Pol II is unique among RNA polymerases in that its largest catalytic subunit (NRPB1) contains a long array of heptapeptide repeats in the C-terminal domain (CTD). Like other Pol II enzymes, the plant Pol II CTD is extensively posttranslationally modified, providing a platform to integrate the state of transcription with downstream RNA processing (74).

In addition to the 12 core subunits, Pol II relies on specialized factors to escort it through the steps of transcription. These proteins facilitate both coding and noncoding transcription by Pol II. Promoters contain a range of specific transcription factor binding sites, providing an opportunity to respond and integrate multiple environmental and internal signals that regulate Pol II initiation. Pol II transcribing through a gene body is assisted by a range of transcription elongation factors (4) that support Pol II processivity during elongation. Once Pol II reaches the end of a gene, transcription must reliably terminate to avoid read-through transcription of neighboring genes. In plants, this termination process has not been extensively studied, but the torpedo model adapted from work in other systems is gaining support (79).

Both mRNA and lncRNA products of Pol II undergo extensive cotranscriptional and

posttranscriptional processing (**Figure 1b**). These processes often include RNA capping, splicing, polyadenylation, and nuclear export, depending on the transcript. In the case of mRNAs, processing ensures subsequent translation on ribosomes, while in the case of lncRNAs, it assures proper structure, localization, and accumulation (88). In addition, Pol II products are subject to nonsense-mediated decay, RNA uridylation, and additional quality control steps implicating the RNA exosome and RNAi pathways (22, 31).

In plants, the distinction between coding and noncoding Pol II transcription has not yet been attributed to a specialized Pol II complex or to specialized GTFs, as was reported in yeast and *Drosophila* (2, 41). However, the genes encoding subunits of the GTFs TFIIA, TFIIB, TFIID, and TFIIE are duplicated in plant evolution (72), which may indicate that the plant Pol II transcription machinery is more complex than currently appreciated.

2.3. Pol III

Pol III synthesizes hundreds of relatively short (<500-nt) lncRNAs with diverse but essential functions. Most notably, Pol III produces the tRNA adaptor molecules that transform mRNA codon information into amino acids during translation (**Figure 1b**). In addition, Pol III transcribes 5S rRNA, an essential part of ribosomes, and U6 snRNA, a crucial component of the spliceosome catalytic core (24, 126). Plant cells have diverged evolutionarily from yeast and vertebrates to use Pol III, instead of Pol II, for U3 small nucleolar RNA and telomerase RNA production (20a, 37, 64). In terms of enzyme composition, Pol III is distinguished by the catalytic subunits NRPC1 and NRPC2, as well as by several specialized polypeptides, for a total of 17 subunits (102) (**Figure 1a**). Due to Pol III being mostly coordinated with cell growth and division, the regulation of Pol III transcription is less complex than that of Pol II.

2.4. Pol IV

Studies of gene silencing and RNA polymerase subunits in *Arabidopsis* revealed two novel transcription machineries in plants; both are dedicated to lncRNA biogenesis (53, 61, 94, 97). The first enzyme, Pol IV, produces precursors for small interfering RNAs (siRNAs) that silence TEs via RNA-directed DNA methylation (RdDM). Pol IV has been implicated in flowering time control in *Arabidopsis* and epigenetic interactions between chromosomes, such as paramutation in maize (1, 20, 36). Work in *Brassicaceae* species, tomato, and other crops has revealed roles for Pol IV in sexual reproduction and seed development (23).

Pol IV, like its molecular cousin Pol II, is composed of 12 subunits (103) (**Figure 1a**). The 3D structure of Pol IV has not been determined, but comparing Pol IV molecular data to the yeast Pol II structure is informative (40, 47, 48, 103). Pol IV's first and second largest subunits (NRPD1 and NRPD2) assemble to form the catalytic core. Pol IV copurifies with SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) and CLASSY family (CLSY1/2/3/4) proteins, which facilitate the recruitment, initiation, and/or elongation of Pol IV transcription in silent chromatin (48, 71) (**Figure 2**). SHH1 specifically binds dimethylated histone H3 lysine 9 (H3K9me2) and unmethylated histone H3 lysine 4 (H3K4) and is involved in recruiting Pol IV to its target loci (70). CLSY proteins are SWI/SNF-like factors critical for locus-specific Pol IV transcription, with the SHH1 transcription factor selectively partnered to CLSY1/2 (148).

<COMP: PLEASE INSERT FIGURE 2 HERE>

Figure 2 lncRNAs in RdDM. Plants use two multisubunit RNA polymerases, Pol IV and Pol V, to transcribe chromosomal DNA into lncRNAs that recognize and silence TEs. This RdDM pathway involves CLSY and SHH1 proteins, which help recruit Pol IV to TEs in chromatin that contains repressive histone marks (H3K9me2 and unmethylated H3K4). Pol IV transcribes TEs into ~30-nt lncRNAs, directly feeding its transcripts to a partner enzyme, RDR2. RDR2 uses Pol IV transcripts to synthesize dsRNAs, which are processed by DCL3 into 24-nt siRNAs. The 5' termini of the siRNAs are 2'-O-methylated (m). An siRNA strand then guides the effector protein AGO4 to sites of Pol V transcription. The binding of an AGO4-siRNA complex to a nascent Pol V transcript is thought to recruit DRM2, leading to de novo cytosine methylation, repressive chromatin marks, and TE silencing that protects genome integrity. The bottom part of the figure is faded to distinguish the siRNAs (orange) from lncRNAs (red). Abbreviations: AGO4, ARGONAUTE 4; CLSY, CLASSY; CTD, C-terminal domain; DCL3, DICER-LIKE 3; DRM2, DOMAINS REARRANGED METHYLTRANSFERASE; dsRNA, double-stranded RNA; H3K4, histone 3 lysine 4; H3K9me2, dimethylated histone 3 lysine 9; lncRNA, long noncoding RNA; RdDM, RNA-directed DNA methylation; RDR2, RNA-DEPENDENT RNA POLYMERASE 2; SHH1, SAWADEE HOMEODOMAIN HOMOLOG 1; siRNA, small interfering RNA; TE, transposable element.

Unlike Pol II, which typically transcribes DNA downstream of gene promoters to produce transcripts thousands of nucleotides long, Pol IV does not rely on sequence-encoded promoters and transcribes mostly TEs and intergenic regions. Although Pol IV transcripts are ~26–45 nt in length and are templates for RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), both primary Pol IV transcripts and secondary RDR2 products fit our definition of lncRNA until they are

processed by a Dicer (**Figure 2**). These transcripts have an A or G as their first nucleotide, which is also observed in products of other RNA polymerases (13, 142). Pol IV transcription is error prone, likely because it lacks proofreading mediated by the canonical trigger loop and bridge helix in the active site of Pol II (87).

Two models for Pol IV termination have been proposed to explain the short Pol IV transcription units. DNA methylation at positions of the Pol IV transcript 3' termini in vivo led the authors of one study to argue that 5-methylcytosine induces Pol IV termination (142). However, Pol IV transcribes methylated DNA without difficulty in vitro. Transcription assays using a series of DNA templates showed that an annealed DNA-DNA hybrid downstream of Pol IV is sufficient to terminate RNA extension (115). These results suggest a different model, in which the Pol IV transcription bubble constrains RNA product lengths, because Pol IV cannot displace a nontemplate DNA strand.

The distinct transcription activities of Pol IV and RDR2 can be coupled in vitro, showing that Pol IV synthesizes primary transcripts and feeds these lncRNAs directly to RDR2 as templates for dsRNA synthesis (48, 115) (**Figure 2**). Pol IV-RDR2 physical coupling is conserved in *Arabidopsis* and maize and so is likely common to flowering plants (47, 48). The resulting dsRNA precursors are then processed by DICER-LIKE 3 (DCL3) into 24-nt siRNAs, which are loaded onto ARGONAUTE 4 (AGO4) and help specify sites of de novo DNA methylation in concert with Pol V.

2.5. Pol V

Pol V is another plant-specific RNA polymerase specialized in lncRNA production and TE silencing (90). In contrast to Pol IV, Pol V does not directly contribute to siRNA production but is needed for siRNAs to recognize their complementary target loci in the genome. The main distinguishing feature of Pol V is its largest subunit, NRPE1 (**Figure 1a**), which contains a unique CTD that is essential for its specific properties (35, 61, 97, 131). Another unique subunit of Pol V is NRPE5 (68, 103). The other subunits of Pol V are shared either with Pol IV or with Pol II and Pol IV (103), reflecting that Pol IV and Pol V originated from Pol II during the evolution of terrestrial plants (86, 123).

Pol V, like Pol IV, has the remarkable property of initiating transcription in a sequence-independent manner. There is no evidence of Pol V transcription initiation being mediated by GTFs binding promoter sequences (16, 58, 132). Instead, Pol V has been proposed to transcribe

pervasively, possibly even transcribing the majority of the genome (120b). The presence of DNA methylation is correlated with higher rates of Pol V transcription, which facilitates efficient maintenance of TE methylation (58, 85, 120b). This mark is recognized by the SET and RING-associated domain proteins SUVH2 and SUVH9 (59), which recruit Pol V and/or facilitate transcript elongation (58, 85). Pol V transcription also requires the DDR complex, composed of the factors DRD1, DMS3, and RDM1; this complex has been proposed to allow the active transcription of chromatin containing repressive marks (69, 133, 134, 147). Pol V transcription is highly accurate but substantially slower than other RNA polymerases (87). Little is known about the mechanisms of Pol V transcription termination and further processing or degradation of Pol V transcripts, beyond the observation that Pol V interacts with RRP6L1, a nuclear exosome subunit with 3'-to-5' exonuclease activity (131).

Transcription by Pol V allows for the recruitment of AGO4-siRNA complexes to their complementary targets in the genome (120b, 134) (**Figure 2**). This recruitment is facilitated by the interaction between the CTD of NRPE1 and AGO4 (35). One model assumes that sequence specificity of AGO4-siRNA recruitment is achieved by base-pairing between siRNA and DNA (67), which leads to a commonly asked question: What is more important, noncoding transcripts or the act of noncoding transcription? Several lines of evidence support the importance of lncRNA itself and suggest base-pairing between siRNA and nascent Pol V-produced lncRNA (84, 101, 134). Binding of AGO4 leads to the recruitment of other RNA-binding proteins and eventually de novo DNA methyltransferase DRM2 and other enzymes responsible for repressive chromatin modifications (15, 146, 149).

3. FUNCTIONS OF lncRNA THROUGHOUT THE GENE

Most lncRNAs are studied in the context of protein-coding gene regulation and as a result can be functionally or mechanistically connected to mRNA expression. In certain cases, lncRNAs target elements in these genes to control their expression. But this connection also means that certain lncRNAs are produced within coding genes or are processed from mRNA transcripts during their maturation. Below we summarize our current understanding of plant lncRNAs by discussing the best characterized examples organized along functionally relevant elements of protein-coding genes (**Figure 3**).

<COMP: PLEASE INSERT FIGURE 3 HERE>

Figure 3 Functions of lncRNA throughout the gene. lncRNAs are best understood in the context of gene regulation. This figure provides a graphical overview of plant lncRNA studies, organizing the information along the gene axis. The lncRNA products of Pol IV and Pol V are mostly synthesized from TEs. When TEs are near gene promoters, RdDM can silence both the TE and the flanking gene. Certain genes display alternative TSS selection, such as at the *Arabidopsis DCL4* gene promoter, where TSS choice is controlled by RdDM. Other promoter-targeting lncRNAs include *ELENA*, *COLDWRAP*, and *APOLO*. Several classes of lncRNA catalyze or regulate splicing: snRNAs in the spliceosome remove introns from pre-mRNAs, whereas *SEPALLATA3* circRNA and *ASCO* help control splicing at particular targets. In addition, *AG-lncRNA* and *COLDAIR* are examples of lncRNAs encoded within intronic regions. *P5SM* is a lncRNA produced as the result of alternate splicing of an otherwise typical mRNA. Downstream of the gene terminator, antisense transcripts can arise: *IGOD* (*DOG1* antisense), *COOLAIR* (*FLC* antisense), and *SVALKA* are examples of antisense lncRNAs. Finally, lncRNAs also originate from intergenic regions. For example, the lncRNA *SUF* is transcribed from an intergenic promoter but extends to overlap the protein-coding gene *FGMYB* in the antisense orientation, controlling female identity in *Marchantia*. Abbreviations: *AG*, *AGAMOUS*; circRNA, circular RNA; *DCL4*, *DICER-LIKE 4*; *DOG1*, *DELAY OF GERMINATION 1*; lncRNA, long noncoding RNA; LTR, long terminal repeat; *P5SM*, 5S rRNA structural mimic; RdDM, RNA-directed DNA methylation; snRNA, small nuclear RNA; TE, transposable element; TSS, transcription start site.

3.1. Gene Promoters

Promoters of protein-coding genes are sequence elements located upstream of transcription start sites (TSSs) and are among the primary determinants of gene expression. Their main role is to efficiently recruit Pol II and regulate the timing and context of transcription initiation. Several studies have observed that promoters are bidirectional and can thus also be a source of promoter upstream transcripts (63a, 120). Conversely, Pol II transcription into the gene body often terminates prematurely, resulting in the generation of short promoter-proximal RNAs (120a). Promoters are therefore a rich source of Pol II transcripts with little apparent protein-coding potential. Following our definition, we refer to these cases as putative lncRNAs, because in the majority of cases they have not been assigned a function.

Not only are promoters a source of lncRNAs, but their activity can also be directly or indirectly controlled by lncRNA. A prime example, discussed below, is the lncRNA that facilitates Mediator complex assembly, a key step in transcription initiation. Additional cases include lncRNAs that control DNA topology and the deposition of specific histone marks that control promoter activity.

3.1.1. Core transcriptional machinery.

A well-studied example of lncRNA acting at a gene promoter is *ELENAI*, which is transcribed by Pol II within the promoter region of the pathogen response gene *PRI* (112). *ELENAI* acts by controlling Mediator occupancy at the *PRI* promoter (Figure 3). Mediator is a large multisubunit protein complex that integrates transcription factor effects and transmits these regulatory signals to Pol II. *ELENAI* is transcribed from and acts on the *PRI* promoter but is also active when ectopically transcribed from a plasmid. It acts by binding the Mediator subunit MED19. *ELENAI* transcription is induced by flg22, a bacterial peptide elicitor of pathogen response in plants. Supporting the hypothesized role of *ELENAI* in pathogen response, artificial miRNA knockdown of *ELENAI* causes a compromised pathogen response, whereas *ELENAI* overexpression lines show enhanced pathogen response (112). The modulation of lncRNA-Mediator interactions is thus one mechanism by which lncRNAs can control promoter activity and regulate plant responses to environmental challenges like bacterial infection.

3.1.2. Chromosome looping.

Promoters are also known to engage in interactions with other chromosomal regions. These interactions span a range of distances and are thought to be responsible for connecting distant regulatory elements with proximal promoters. At least two *Arabidopsis* lncRNAs have been shown to affect the formation of chromosome loops involving promoters of their target loci. The *APOLO* lncRNA is transcribed 5 kb upstream of the gene *PID*, which encodes a regulator of auxin transport (6). *APOLO* is transcribed by both Pol II and Pol V and affects root development by regulating *PID* expression. Experiments based on RNAi knockdown of *APOLO* lncRNA showed that *APOLO* controls *PID* expression by contributing to the formation of a chromosome loop between the *APOLO* locus and the *PID* promoter. Changes in chromosomal interactions associated with *APOLO* expression are dependent on DNA methylation and histone H3 lysine 27 trimethylation (H3K27me3) at the *PID* promoter. *APOLO* also controls local chromatin looping and engages in R-loop formation at a set of its *trans* targets (7).

COLDWRAP is derived from the promoter of *FLOWERING LOCUS C (FLC)*, a key regulator of flowering time; *COLDWRAP* controls the intragenic gene loop between the *FLC* gene promoter and its first intron (62) (Figure 4). *COLDWRAP* is required for efficient silencing of *FLC* transcription in response to prolonged cold exposure, a process termed vernalization, which accelerates flowering. This function was discovered using RNAi knockdown, as well as the

expression of *FLC* from a transgene with the *COLDWRAP* region deleted. The molecular mechanisms of *COLDWRAP* action include establishment of a repressive intragenic loop that forms in response to cold exposure and is dependent on *COLDWRAP*. In addition, *COLDWRAP* binds the Polycomb repressive complex 2 (PRC2) subunit CLF, contributing to its localization to the *FLC* gene, as well as facilitating H3K27me3 of *FLC* chromatin in response to vernalization. Together, these examples demonstrate that the modulation of chromosome looping is another mechanism used by lncRNA to affect the function of gene promoters.

<COMP: PLEASE INSERT FIGURE 4 HERE>

Figure 4 *FLC* gene regulation by lncRNAs. The *FLC* gene (*black region*) codes for the *FLC* mRNA (*red rectangles*), which is translated into a MADS-box transcription factor that regulates flowering time. *FLC* expression is regulated by three distinct classes of lncRNA (*rectangles with white arrows*). *COLDWRAP* is transcribed from within the *FLC* promoter and has been reported to control *FLC* promoter looping during cold-mediated *FLC* silencing, a process known as vernalization. *COLDAIR* is a different lncRNA transcribed from the genomic region of *FLC* intron 1, which has been reported to recruit PRC2 during vernalization. *COOLAIR* is a collection of antisense lncRNAs transcribed from a promoter overlapping the *FLC* terminator. *COOLAIR* contributes to the vernalization process but also has a function in *FLC* regulation in moderate temperatures. *COOLAIR* expression is negatively regulated by an R-loop formed at its promoter and stabilized by the AtNDX protein. *COOLAIR* is alternatively spliced. The PRP8 factor, and a splice site where PRP8 acts, are shown; this lncRNA splicing event is required for *COOLAIR* to control *FLC* expression. In addition, the *COOLAIR* transcript is alternatively polyadenylated (denoted by polyA). The FPA and FCA proteins localize to the *FLC* gene and promote selection of the proximal *COOLAIR* polyA site. *COOLAIR* structure has also been shown to be evolutionarily conserved. While the precise mode of *COOLAIR* action is not fully understood, it is thought to form an RNA cloud at the *FLC* locus and to mediate H3K36me3 and H3K4me1 removal. Abbreviations: *FLC*, *FLOWERING LOCUS C*; H3K4me1, histone 3 lysine 4 monomethylation; H3K36me3, histone 3 lysine 36 trimethylation; lncRNA, long noncoding RNA; PRC2, Polycomb repressive complex 2; PRP8, PRE-MRNA PROCESSING 8.

3.1.3. Chromatin modifications and transcription.

Another mechanism used by lncRNA to control promoter function is chromatin modifications, which include posttranslational histone modifications and DNA methylation. A typical promoter of an actively transcribed plant protein-coding gene is associated with high levels of histone H3 lysine 4 trimethylation (H3K4me3) and histone acetylation. These modifications peak near the 5' end of the transcribed region, corresponding to the early elongation phase of Pol II transcription. Such promoters are typically devoid of the chromatin marks present further downstream in the

transcribed gene body, such as histone 2B ubiquitination (H2Bub), histone H3 lysine 4 monomethylation (H3K4me1), and histone H3 lysine 36 dimethylation (H3K36me2) (74). lncRNA mediated regulation of gene promoters includes changes in histone modifications, such as deposition of repressive histone H3 lysine 27 trimethylation (H3K27me3) marks, a mechanism associated with functions of the previously discussed lncRNAs *APOLO* and *COLDWRAP*. Importantly, posttranslational histone modifications are often established cotranscriptionally by histone-modifying enzymes associated with the elongating Pol II holoenzyme (4). It is therefore often difficult to disentangle the effects of the act of transcription and changes in cotranscriptional histone modifications.

An example of lncRNAs that act on gene promoters via transcription are Pol II transcripts derived from improper transcription termination. Such transcripts are observed in *Arabidopsis* mutants defective in BORDER proteins, which are transcription elongation factors containing TFIIS-like domains (139). While abundant read-through transcripts are observed in *border* mutants, most of these RNAs lack any assigned function and cannot be termed lncRNA. However, some such transcripts appear to silence downstream genes, presumably because of read-through into the target gene promoters and TSSs. On the basis of the deficiency phenotype of *border* triple mutants, this mechanism was implicated in photomorphogenesis (139). This is an example of how the phenomenon of transcriptional interference may repress transcription units (44a, 113a). A similar process was reported for transgene-derived putative lncRNAs that target the *QUAI* gene promoter (93).

A common source of noncoding transcripts in plant gene promoters are TEs, which not only are genomic parasites but also contribute to gene regulation (54). Approximately 36% of *Arabidopsis* genes and 85% of maize genes are present within 1 kb of a TE (54, 78). Noncoding transcription by Pol II, as well as by the specialized RNA polymerases Pol IV and Pol V, provides a surveillance system that recognizes TE sequences, allowing for the establishment and maintenance of DNA methylation and repressive histone modifications (95). Consequently, the transcriptional silencing machinery, including Pol IV and Pol V, is also commonly found to target gene promoters (145, 147).

RNA silencing of gene promoter-localized TEs affects the expression of limited subsets of protein-coding genes in flowering plants (3, 20, 80a, 128, 145). Effects of TE silencing on gene expression were initially proposed to rely on heterochromatin spreading from the TEs into gene

promoters, which turned out not to be the case (78). Instead, RdDM at TE edges is responsible for the maintenance of heterochromatic marks within TEs (78, 119a), and the mechanism that allows RdDM to affect gene promoters is likely related to the silencing of TEs (54). Moreover, TEs not only serve as a source of lncRNAs that regulate neighboring genes, but these lncRNAs also affect distant genes by controlling long-range chromosomal interactions (109). Despite having little sequence conservation, TE edges have features similar to those of promoters, especially the frequent initiation of Pol IV and Pol V transcription at the edges of long TEs (16, 141a, 142).

A specialized instance of lncRNA targeting promotor regions is nucleolar dominance, which is the selective silencing of rRNA genes in certain hybrid species (96). For example, in allotetraploid *Arabidopsis suecica* (*Arabidopsis arenosa* × *A. thaliana*), only rRNA genes derived from *A. arenosa* are expressed. By contrast, rRNA genes derived from *A. thaliana* are silenced by DNA methylation and repressive histone modifications at their Pol I promoters (122) (**Figure 1**). lncRNAs synthesized from the intergenic spacer (IGS) separating adjacent rRNA genes have been implicated in this dominance of *A. arenosa* over *A. thaliana* rRNA genes in the hybrid (99). Pol I transcription traversing the *A. arenosa* IGS is thought to read-through the downstream Pol I promoter, generating 24-nt siRNAs capable of triggering RdDM in *trans* to silence *A. thaliana* genes residing on different chromosomes. Other RNA polymerases that transcribe the IGS and rRNA genes (e.g., Pol II or Pol IV), may also generate primary precursors of siRNAs (99). Together, these examples demonstrate the multifaceted roles of lncRNA in controlling promoters by establishing chromatin modifications (**Figure 3**).

3.2. Transcription Start Site and 5' Untranslated Region

The 5' untranslated region (5'UTR) is the gene interval between a Pol II TSS and the corresponding mRNA's AUG start codon. 5'UTRs are noncoding features that control mRNA stability and translation (141), but they do not qualify as lncRNAs by default. Identifying bona fide lncRNAs that arise at the TSS or within the 5'UTR is complicated because of the many mRNA variants and small upstream open reading frames (ORFs) that overlap the 5' regions of plant genes (116, 124). Putative lncRNAs detected in genic regions must thus be carefully validated using molecular genetics to test for mechanisms of lncRNA biogenesis and function.

Pol II selection of a TSS delimits the beginning of the 5'UTR and can govern the 5' extremity of the mRNA-coding region. The *DICER-LIKE 4* (*DCL4*) gene uses lncRNA to control TSS selection and protein sequence in *Arabidopsis* (100). In most tissues, the *DCL4* promoter is

targeted by RdDM, forcing Pol II to initiate at a position downstream in the gene to yield the predominant, cytoplasmic DCL4 isoform. In mutants deficient for lncRNA transcription by Pol IV, *DCL4* promoter methylation is lost, allowing Pol II to initiate at the alternative, upstream TSS. This longer mRNA includes a 5' nuclear localization signal and encodes a nuclear DCL4 isoform accountable for a class of 21-nt siRNAs specific to wild-type siliques and seeds. Nuclear DCL4 likely accumulates in those particular tissues because they have reduced levels of *DCL4* promoter methylation (100).

Properties of 5'UTRs that can convey lncRNA function include RNA secondary structures and conserved sequence elements (75, 76). Such properties were observed for the untranslated 5' leader region of the cauliflower mosaic virus (CaMV) 35S RNA, which plant Pol II synthesizes by transcribing viral DNA in the nucleus. The highly structured 35S leader and related viral 8S lncRNA are processed into massive quantities of siRNAs, overwhelming the host antiviral defense machinery. This lncRNA decoy mechanism is hypothesized to protect CaMV-coding regions from antiviral cleavage (14). No endogenous lncRNAs have yet been reported to directly regulate a plant 5'UTR. However, in other systems, lncRNAs do act at 5'UTRs (19), leaving open the prospect that lncRNA-5'UTR interactions could be discovered for plant mRNAs. Together, these two examples demonstrate that TSS regions and 5'UTRs may be targeted by and give rise to lncRNA.

3.3. Exons and Introns

While diverse RNA polymerases produce lncRNA (Figure 1), the majority of these lncRNAs are involved in processes that influence how Pol II transcripts are expressed, processed, or translated as mRNAs (Figure 3). An essential lncRNA-controlled process is splicing, which removes introns from mRNA sequences. Introns are a fertile landscape for lncRNA evolution because of their abundance and complex relationship with protein production (56). The spliceosome consists of protein factors combined with snRNAs, which perform structural and catalytic functions in the spliceosome (135). Although snRNAs have the term small in their name (Table 1), they share no commonalities with Dicer-produced small RNAs and fit our definition of lncRNAs. Acting at various levels in splicing, these lncRNAs determine the ultimate protein-coding sequence in most, if not all, plant mRNAs (Figure 3). Recent studies in plants have begun to show how the intron-exon architecture of Pol II transcripts also facilitates lncRNA production.

3.3.1. Regulation of splicing by lncRNA.

Beyond having a direct enzymatic function in splicing, plant lncRNAs also regulate alternative splicing, which can be responsive to developmental or environmental cues (18, 107). An intriguing example is the 60-nt circular RNA (circRNA) that forms from exon 6 backsplicing in the *SEPALLATA3* (*SEP3*) gene of *Arabidopsis* (26). This circRNA is hypothesized to form an R-loop (RNA:DNA hybrid) at the *SEP3* gene locus to regulate *SEP3* pre-mRNA splicing, resulting in defects in flower development (26). Another example of a lncRNA controlling splicing is *ASCO*, which associates with the core splicing component PRE-MRNA PROCESSING 8 (PRP8) as well as with SmD1 in *Arabidopsis*. RNAi-mediated knockdown of the *ASCO* lncRNA led to splicing defects at a subset of genes and reduced PRP8 binding at a subset of its pre-mRNA targets (105). These examples demonstrate that lncRNA may regulate alternative splicing in plants (Figure 3).

3.3.2. lncRNA production from introns.

Although introns are typically degraded quickly, they may also give rise to lncRNAs. One example is *FLC*, for which a capped, nonpolyadenylated sense RNA is produced from intron 1 (52). This lncRNA is known as *COLDAIR* and is clearly distinct from *COOLAIR*, which is a collection of capped, spliced, and polyadenylated antisense lncRNAs (119). The *COLDAIR* promoter resides within *FLC* intron 1 and is sufficient for cold-responsive expression. Whereas *COOLAIR* expression peaks after 10 days of cold treatment, *COLDAIR* expression peaks after 20 days. Because *COLDAIR* associates with the CLF-containing PRC2, it has been proposed to recruit PRC2 to *FLC* in response to cold. PRC2, in turn, methylates histone H3 lysine 27 (H3K27me3) and causes the repression of *FLC* (52). This work shows that intronic transcripts, like *COLDAIR*, can affect gene expression in *cis*.

Another example of introns contributing to lncRNA production comes from studies of floral development. *AGAMOUS* (*AG*) encodes a MADS-box protein that specifies stamen and carpel fates in *Arabidopsis* flowers. The second intron of *AG* gives rise to a lncRNA termed *AG-incRNA4*, which interacts with CLF. Consistent with *AG-incRNA4* acting to control *AG* expression, CLF is enriched in chromatin corresponding to the *AG-incRNA4* genomic segment. Furthermore, RNAi knock down of *AG-incRNA4* causes decreased H3K27me3 at *AG* and increased accumulation of *AG* mRNA in leaves, along with developmental phenotypes typical of *AG* overexpression (136). This work suggests that *AG-incRNA4* may recruit PRC2 and repress *AG*, therefore contributing to tissue-specific expression of *AG*. Although the connection between lncRNAs and PRC2 remains debated in various organisms (91), this example further shows that introns may give rise to

lncRNAs that regulate gene expression (**Figure 3**).

3.3.3. Alternative splicing of lncRNAs.

In addition to lncRNAs produced from introns or lncRNA function in the regulation of splicing, lncRNA can be subject to alternative splicing. The *COOLAIR* lncRNAs that silence *FLC* during vernalization are capped, spliced, and polyadenylated (119). Cold treatment induces *COOLAIR* antisense transcription, which extends to cover nearly the entire *FLC* gene. Because of splice sites oriented antisense to *FLC*, *COOLAIR* occurs as three alternatively spliced isoforms (119) (**Figure 4**). The longest *COOLAIR* isoform includes three noncoding exons: one downstream of the *FLC* transcript and the other two spanning *FLC* exon 7 and exon 1. Proper splicing of *COOLAIR* is essential for its function in *FLC* repression, as demonstrated in splicing factor mutants and *COOLAIR* mutants with disrupted splice sites (88) (**Figure 4**). Moreover, natural variations in *COOLAIR* splicing have been linked to its function in *FLC* repression (77).

Alternative splicing also regulates the plant Pol III machinery via expression of an exonic lncRNA. In *Arabidopsis*, the Pol III GTF TFIIIA is encoded by two alternatively spliced mRNA isoforms. The first isoform encodes the functional TFIIIA protein and is thus essential for viability. The second isoform expressed from the *TFIIIA* gene is a lncRNA, which is a noncoding 5S rRNA structural mimic (*P5SM*) and serves a purely regulatory function (49). Because *P5SM* binds the L5 ribosomal protein (mimicking the 5S rRNA) and promotes Pol III synthesis of actual 5S rRNA, *P5SM* is thought to sense free L5 protein and adjust transcription to maintain equivalent amounts of L5 protein and 5S rRNA, whose transcripts are products of distinct RNA polymerases. The exon-skipping architecture and *P5SM* form of *TFIIIA* RNA are conserved throughout terrestrial plants, likely making this an ancient type of gene regulation by a plant lncRNA (49). Together, these examples demonstrate the importance of alternative splicing in lncRNA biogenesis and function.

3.3.4. microRNAs and their relationship with lncRNAs.

Not all Pol II–transcribed genes code for proteins, as indicated above (**Figure 1b**). One such class comprises the genes for primary microRNAs (pri-miRNAs), which are capped and polyadenylated transcripts that fold into stem-loop hairpin structures. The processing of pri-miRNAs into miRNAs by DICER-LIKE 1 (DCL1) is a complex pathway that has been extensively studied in plants. miRNA biogenesis, miRNA targeting of mRNAs, and related gene regulatory processes have been

reviewed elsewhere and are beyond the scope of our review (8, 140). While mature miRNAs do not fit our definition of lncRNAs, pri-miRNAs may have Dicer-independent functions and are, intrinsically, a form of lncRNA that controls miRNA accumulation. Recent evidence indicates that pri-miRNAs are methylated by the adenosine methylase MTA (12). The presence of *N*⁶-methyladenosine is needed for proper pri-miRNA structure, which in turn facilitates the activity of the DCL1-containing microprocessor complex and miRNA production (12).

The activity of miRNAs can also be controlled by lncRNAs that mimic miRNA targets, also known as miRNA sponges (43). Such lncRNAs contain miRNA binding sites and are properly recognized by a miRNA. However, this binding has no consequence for lncRNAs. Instead, it sequesters the specific miRNA and prevents it from functionally repressing its target mRNAs (43, 143). Together, these cases show that lncRNAs control miRNA biogenesis and can also modify the biological activity of miRNAs in gene regulation.

3.4. 3'UTR

The 3'UTR is defined by the interval between the translation stop codon and the mRNA polyadenylation site. Compared to mammalian 3'UTRs, which have a median length of 700 nt, plant 3'UTRs are shorter, ranging from 240 nt in *Arabidopsis* to 470 nt in rice (116). In analogy to the case of alternative TSSs, described above, alternative polyadenylation sites can generate RNA isoforms with different primary sequences, secondary structures, 3'UTR lengths, and regulatory potential (116).

Alternative polyadenylation may affect the length and function of lncRNAs, as reported for *COOLAIR* (81), in which the selection of a proximal polyadenylated site leads to suppression of *FLC* expression (Figure 4). This case demonstrates that changes in the 3'UTR linked to alternative polyadenylation may disrupt lncRNA function. Other forms of polyadenylation site selection are probably widespread in plants, but the extent to which they affect lncRNAs is not yet known.

3.5. Terminator

Transcription terminators are DNA segments with sequences that signal for RNA polymerase to terminate transcription, marking the end of the gene. lncRNAs can control a terminator's ability to stop Pol II transcription. A failure to terminate results in transcriptional read-through, as observed in *border* mutants (139), in plants expressing a mutant Pol II enzyme with accelerated transcription (73a), as well as in mutants for XRN 5'-to-3' exoribonucleases (66). In most cases,

these read-through transcripts have not been assigned a function and do not meet the definition of lncRNA. One example in which failure to terminate transcription generates a functional lncRNA is the lncRNA *SVALKKA*, which is antisense to *CBFI* and initiates from the adjacent downstream gene (63). *SVALKKA* modulates cold response by controlling expression of the full-length *CBFI* mRNA.

Conversely, terminators are also a source of antisense transcripts, a class of lncRNAs that are widespread in the *Arabidopsis* genome (144). More than 1,800 *Arabidopsis* genes have promoter sequence elements near their terminators and a promoter-associated SWI-SNF complex bound at their 3' ends (5). The ability of these terminators to serve as promoters may be controlled by other lncRNAs, as well as by the act of transcription termination itself. Below we describe this complicated interplay between lncRNA and terminators.

3.5.1. Terminators as lncRNA promoters.

Extensive evidence supports the view that terminators can serve as promoters to drive lncRNA transcription in the antisense orientation relative to protein-coding genes (127, 143). Such antisense transcripts may work not only in *cis* but potentially also in *trans*, affecting other genomic regions with sequence similarity (alleles, homologs, and homeologs). Significantly, not all antisense transcripts are initiated from within terminators. Read-through transcription of adjacent genes, TEs, introns, and other sources also contributes to antisense transcription.

A well-known example of an antisense transcript starting within a terminator is *COOLAIR* (119). It originates from an autonomous promoter that overlaps with the terminator of *FLC* (Figure 4). *COOLAIR* is strongly induced by cold and contributes to the initial repression of *FLC*; such repression is then reinforced by Polycomb complexes in concert with two additional lncRNAs: *COLDAIR* transcribed from the first *FLC* intron (52) and *COLDWRAP* transcribed from the *FLC* promoter (62). While the mechanism of *COOLAIR* action is unclear, single-molecule RNA FISH revealed a cloud of *COOLAIR* surrounding the *FLC* locus, suggesting that antisense lncRNA accumulates there (108). This process leads to removal of the active chromatin marks H3K36me3 and H3K4me1, which contributes to *FLC* repression (28, 38) (Figure 4).

Another case of an antisense transcript that originates from a gene terminator occurs on the *DELAY OF GERMINATION 1* (*DOG1*) locus, which controls the natural variability of seed dormancy among *Arabidopsis* accessions (11, 39). *DOG1* has two polyadenylation sites, and the proximal one overlaps the promoter of an antisense transcript known as *IGOD* (29). How *IGOD*

causes repression of *DOG1* is currently unknown, but *DOG1* expression is extensively regulated, allowing *DOG1* to respond to abscisic acid and revealing a role of *DOG1* in drought regulation (138).

Both *COOLAIR* and *IGOD* act in *cis*, meaning that they affect the alleles from which they are produced. However, other antisense transcripts act in *trans*. This finding is mostly based on studies that rely on antisense lncRNA overexpression from transgenes (32). As such, the function and mechanisms of these antisense transcripts are supported by varying degrees of experimental evidence. Together, these examples illustrate the importance of antisense lncRNAs transcribed from promoters within terminators of protein-coding genes.

3.5.2. Interplay between transcription termination and lncRNA production.

Because transcription terminators often serve as lncRNA start sites, an interesting consequence is that transcription termination itself may control the activity of antisense promoters. In both case studies of antisense lncRNAs in *Arabidopsis* (*COOLAIR* from *FLC* and *IGOD* from *DOG1*), the antisense promoter can autonomously drive expression of a transgenic reporter. However, in the latter case, changing from the proximal to the distal *DOG1* polyadenylation site on the sense strand (the protein-coding strand) suppresses *IGOD* lncRNA expression (65).

Transcription termination may control lncRNA production in one additional manner. Failure to efficiently terminate transcription leads to read-through transcription, which produces RNA that spans fragments of neighboring genes. Such RNA has the potential to control expression of those neighboring genes, which has been proposed in mutants defective in the transcription termination machinery or RNA exonucleases (66).

lncRNAs may also control the initiation of antisense transcription from transcription terminators, which has been shown for putative lncRNAs forming R-loops (137). Within the *FLC* terminator, R-loop formation stabilized by the R-loop binding protein AtNDX was reported to suppress *COOLAIR* transcription initiation (118) (Figure 4). Together, these examples demonstrate the interplay between transcription termination and lncRNA production.

3.5.3. Antisense lncRNAs derived from intergenic regions.

Antisense transcripts may also originate from genomic regions stretching beyond the transcription terminator. A good example of this class of antisense lncRNA is the *SUF* transcript in the liverwort *Marchantia polymorpha* (55). *SUF* is transcribed from an intergenic promoter and overlaps the

second half of the *MpFGMYB* gene, a key MYB-type transcription factor contributing to female identity. *SUF* acts in *cis* and represses *MpFGMYB* in male plants. These actions that antisense lncRNAs may control gene expression even if they are transcribed from promoters beyond transcription terminators.

3.6. Pervasive Transcription

Pervasive transcription generates a unique category of lncRNAs produced across the genome with limited sequence specificity (30, 33, 82, 143). Pervasive transcription has been reported in many species, but it is essential to distinguish between pervasive lncRNAs of defined biological function and the products of transcriptional noise or sequencing artifacts.

One example of pervasive lncRNA production is Pol V's function in the maintenance of genome integrity. Pol V transcription has been proposed to provide the lncRNA scaffolds needed to silence newly integrated or activated TEs. To this end, Pol V transcribes a large portion, or potentially even the entirety, of the *Arabidopsis* genome (120b). Pervasive Pol V transcription may be essential because new TE insertions lack universal DNA motifs or chromatin marks that may recruit Pol V for TE silencing. In this model, silencing specificity is determined primarily by TE-derived siRNAs (28a) that require Pol V pervasive transcription for target recognition.

The role of Pol V in pervasive transcription is further supported by its involvement in DNA repair (130). DNA double-strand breaks are repaired by a mechanism that shares elements with RNA silencing, including the requirement for Pol V (83, 130). Because DNA double-strand breaks form at random genomic positions, Pol V may transcribe chromosomes without much sequence specificity. These examples indicate that widespread noncoding transcription may be functional even when it is not constrained by the gene structure or other features of the genome.

4. CONCLUSIONS

As explained above, plant lncRNAs have remarkably diverse biogenesis mechanisms, as highlighted by the four distinct RNA polymerases (Pol I, III, IV, and V) dedicated to lncRNA production. The products of these polymerases and noncoding products of Pol II display a broad range of sizes (Table 1). Therefore, no arbitrary size cutoff adequately splits ncRNAs into small RNA and lncRNA categories. Instead, we offer an inclusive definition based on RNA biogenesis mechanisms. The diversity of plant lncRNAs is also reflected in their complex relationship with

protein-coding RNAs. For this reason, the lack of protein annotations or predicted ORFs is of little use in defining lncRNAs. Our expanded lncRNA definition thus relies on the presence of a demonstrated biological function that is independent of the protein-coding potential of an RNA.

The mechanisms of plant lncRNA function are still poorly understood, with many fundamental questions open. One big question concerns the commonality of transcriptional regulation by lncRNA. Are most or even all plant genes regulated by lncRNA mechanisms similar to those attributed to the genes *FLC*, *DOG1*, and *PID*? This scenario is quite possible, but the impact of lncRNA relative to protein factors at each gene likely varies and should be the subject of future research efforts. Another open question concerns the most common mechanisms of lncRNA function. While some overarching themes have been proposed ([21](#), [106](#)), our knowledge about the generality of these mechanisms is limited. More plant lncRNAs have to be studied in depth to determine which aspects of lncRNA biology are universal and which are locus specific. Another unresolved question concerns the limits of lncRNA functionality. Current evidence supports the involvement of lncRNA in all steps of gene expression that engage nucleic acids, from chromatin modifications to translation. How common are the purely structural roles of lncRNA, unrelated to gene expression and protein translation?

Finally, lncRNA evolution is another exciting frontier. The tremendous variety of plant lncRNAs helps explain the lack of any narrowly conserved mechanism of action. Through their diversity, lncRNAs have endless adaptive potential, which resembles the functional variation of proteins in some respects. In extreme cases, convergent evolution in plants has independently produced different protein and lncRNA enzymes with the same catalytic activity and function ([46](#)). While some lncRNAs are highly conserved and can be expected to fit into paradigms established for the evolution of proteins, many lncRNAs lack sequence conservation, meaning that only a blurry line distinguishes them from noisy or pervasive transcription.

Two hundred and forty years after the first observation of a lncRNA-containing structure ([42](#)), our knowledge of this fascinating group of nucleic acids has exploded and is still quickly expanding. Many exciting new discoveries certainly await.

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Sebastian Marquardt and Michał Krzysztoń for critical reading of the article. The research of A.T.W. is funded by the National Institute of General Medical Sciences of the National Institutes of Health under award R01GM108722 and by National Science Foundation grant MCB1934703. T.B. is supported by the LabEx consortia ANR-10-LABX-0036_NETRINA and ANR-17-EURE-0023 from French Programme d'Investissements d'Avenir (PIA) funds and by French Agence Nationale de la Recherche grant ANR-17-CE20-0004-01. S.S. is funded by the Foundation for Polish Science (TEAM POIR.04.04.00-00-3C97/16) and by Polish National Science Centre grants OPUS UMO-2016/23/B/NZ1/02989 and SONATA BIS UMO-2018/30/E/NZ1/00354.

LITERATURE CITED

1. Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, et al. 2006. An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* 442(7100):295–98
2. Andersen PR, Tirian L, Vunjak M, Brennecke J. 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* 549(7670):54–59
3. Anderson SN, Zynda GJ, Song J, Han Z, Vaughn MW, et al. 2018. Subtle perturbations of the maize methylome reveal genes and transposons silenced by chromomethylase or RNA-directed DNA methylation pathways. *G3 Genes Genomes Genet.* 8(6):1921–32
4. Antosz W, Pfab A, Ehrnsberger HF, Holzinger P, Köllen K, et al. 2017. The composition of the *Arabidopsis* RNA polymerase II transcript elongation complex reveals the interplay between elongation and mRNA processing factors. *Plant Cell* 29(4):854–70
5. Archacki R, Yatusevich R, Buszewicz D, Krzyczmonik K, Patryn J, et al. 2017. *Arabidopsis* SWI/SNF chromatin remodeling complex binds both promoters and terminators to regulate gene expression. *Nucleic Acids Res.* 45(6):3116–29
6. Ariel F, Jegu T, Latrassé D, Romero-Barrios N, Christ A, et al. 2014. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* 55(3):383–96
7. Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, et al. 2020. R-loop mediated *trans* action of the APOLO long noncoding RNA. *Mol. Cell* 77(5):1055–65

8. Axtell MJ. 2013. Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* 64:137–59
9. Bardou F, Merchan F, Ariel F, Crespi M. 2011. Dual RNAs in plants. *Biochimie* 93(11):1950–54
10. Ben Amor B, Wirth S, Merchan F, Laporte P, d’Aubenton-Carafa Y, et al. 2009. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res.* 19(1):57–69
11. Bentsink L, Jowett J, Hanhart CJ, Koornneef M. 2006. Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *PNAS* 103(45):17042–47
12. Bhat SS, Bielewicz D, Gulanicz T, Bodi Z, Yu X, et al. 2020. mRNA adenosine methylase (MTA) deposits m⁶A on pri-miRNAs to modulate miRNA biogenesis in *Arabidopsis thaliana*. *PNAS* 117(35):21785–95
13. Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, et al. 2015. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *eLife* 4:e09591
14. Blevins T, Rajeswaran R, Aregger M, Borah BK, Schepetilnikov M, et al. 2011. Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Res.* 39(12):5003–14
15. Böhmendorfer G, Rowley MJ, Kuciński J, Zhu Y, Amies I, Wierzbicki AT. 2014. RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA. *Plant J.* 79(2):181–91
16. Böhmendorfer G, Sethuraman S, Rowley MJ, Krzyszton M, Rothi MH, et al. 2016. Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *eLife* 5:e19092
17. Brosius J, Raabe CA. 2016. What is an RNA? A top layer for RNA classification. *RNA Biol.* 13(2):140–44
18. Capovilla G, Pajoro A, Immink RGH, Schmid M. 2015. Role of alternative pre-mRNA splicing in temperature signaling. *Curr. Opin. Plant Biol.* 27:97–103
19. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, et al. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491(7424):454–57

20. Chan SW-L, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. 2004. RNA silencing genes control de novo DNA methylation. *Science* 303(5662):1336
- 20a. Chapon C, Cech TR, Zaug AJ. 1997. Polyadenylation of telomerase RNA in budding yeast. *RNA* 3(11):1337-1351
21. Chekanova JA. 2015. Long non-coding RNAs and their functions in plants. *Curr. Opin. Plant Biol.* 27:207–16
22. Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, et al. 2007. Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell* 131(7):1340–53
23. Chow HT, Chakraborty T, Mosher RA. 2020. RNA-directed DNA methylation and sexual reproduction: expanding beyond the seed. *Curr. Opin. Plant Biol.* 54:11–17
24. Cloix C, Yukawa Y, Tutois S, Sugiura M, Tourmente S. 2003. In vitro analysis of the sequences required for transcription of the *Arabidopsis thaliana* 5S rRNA genes. *Plant J.* 35(2):251–61
25. Cognat V, Pawlak G, Duchêne A-M, Daujat M, Gigant A, et al. 2013. PlantRNA, a database for tRNAs of photosynthetic eukaryotes. *Nucleic Acids Res.* 41(Database Issue):D273–79
26. Conn VM, Hugouvieux V, Nayak A, Conos SA, Capovilla G, et al. 2017. A circRNA from *SEPALLATA3* regulates splicing of its cognate mRNA through R-loop formation. *Nat. Plants* 3:17053
27. Cramer P. 2019. Organization and regulation of gene transcription. *Nature* 573(7772):45–54
28. Csorba T, Questa JI, Sun Q, Dean C. 2014. Antisense *COOLAIR* mediates the coordinated switching of chromatin states at *FLC* during vernalization. *PNAS* 111(45):16160–65
- 28a. Cuerda-Gil D, Slotkin RK. 2016. Non-canonical RNA-directed DNA methylation. *Nat. Plants* 2:16163
29. Cyrek M, Fedak H, Ciesielski A, Guo Y, Sliwa A, et al. 2016. Seed dormancy in *Arabidopsis* is controlled by alternative polyadenylation of *DOG1*. *Plant Physiol.* 170(2):947–55
30. David L, Huber W, Granovskaia M, Toedling J, Palm CJ, et al. 2006. A high-resolution map of transcription in the yeast genome. *PNAS* 103(14):5320–25
31. De Almeida C, Scheer H, Zuber H, Gagliardi D. 2018. RNA uridylation: a key posttranscriptional modification shaping the coding and noncoding transcriptome. *Wiley Interdiscip Rev RNA* 9(1):e1440
32. Deforges J, Reis RS, Jacquet P, Sheppard S, Gadekar VP, et al. 2019. Control of cognate sense

- mRNA translation by *cis*-natural antisense RNAs. *Plant Physiol.* 180(1):305–22
33. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, et al. 2012. Landscape of transcription in human cells. *Nature* 489(7414):101–8
 34. Eddy SR. 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2(12):919–29
 35. El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, et al. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* 21(20):2539–44
 36. Erhard KFJ, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB. 2009. RNA polymerase IV functions in paramutation in *Zea mays*. *Science* 323(5918):1201–5
 37. Fajkus P, Peska V, Zavodnik M, Fojtova M, Fulneckova J, et al. 2019. Telomerase RNAs in land plants. *Nucleic Acids Res.* 47(18):9842–56
 38. Fang X, Wu Z, Raitskin O, Webb K, Voigt P, et al. 2020. The 3' processing of antisense RNAs physically links to chromatin-based transcriptional control. *PNAS* 117(26):15316–21
 39. Fedak H, Palusinska M, Krzyczmonik K, Brzezniak L, Yatusевич R, et al. 2016. Control of seed dormancy in *Arabidopsis* by a *cis*-acting noncoding antisense transcript. *PNAS* 113(48):E7846–55
 40. Ferrafiat L, Pflieger D, Singh J, Thieme M, Böhrer M, et al. 2019. The NRPD1 N-terminus contains a Pol IV–specific motif that is critical for genome surveillance in *Arabidopsis*. *Nucleic Acids Res.* 47(17):9037–52
 41. Fischl H, Howe FS, Furger A, Mellor J. 2017. Paf1 has distinct roles in transcription elongation and differential transcript fate. *Mol. Cell* 65(4):685–698.e8
 42. Fontana F. 1781. *Traité sur le vénin de la vipere sur les poisons américains sur le laurier-cerise et sur quelques autres poisons végétaux*. Florence: Nyon l'Ainé
 43. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, et al. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39(8):1033–37
 - 43a. Goldberg RB, Hoschek G, Kamalay JC, Timberlake WE. 1978. Sequence complexity of nuclear and polysomal RNA in leaves of the tobacco plant. *Cell* 14(1):123–131
 44. Golicz AA, Singh MB, Bhalla PL. 2018. The long intergenic noncoding RNA (lincRNA) landscape of the soybean genome. *Plant Physiol.* 176(3):2133–47
 - 44a. Gowthaman U, García-Pichardo D, Jin Y, Schwarz I, Marquardt S. 2020. DNA processing in

- the context of noncoding transcription. *Trends Biochem Sci.* 45(12):1009-1021
45. Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. 2013. On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* 5(3):578–90
 46. Gutmann B, Gobert A, Giegé P. 2012. PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*. *Genes Dev.* 26(10):1022–27
 47. Haag JR, Brower-Toland B, Krieger EK, Sidorenko L, Nicora CD, et al. 2014. Functional diversification of maize RNA polymerase IV and V subtypes via alternative catalytic subunits. *Cell Rep.* 9(1):378–90
 48. Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, et al. 2012. In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol. Cell* 48(5):811–18
 49. Hammond MC, Wachter A, Breaker RR. 2009. A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. *Nat. Struct. Mol. Biol.* 16(5):541–49
 50. Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, et al. 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet.* 45(8):891–98
 51. Hawkes EJ, Hennelly SP, Novikova IV, Irwin JA, Dean C, Sanbonmatsu KY. 2016. *COOLAIR* antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell Rep.* 16(12):3087–96
 52. Heo JB, Sung S. 2011. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331(6013):76–79
 53. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* 308(5718):118–20
 54. Hirsch CD, Springer NM. 2017. Transposable element influences on gene expression in plants. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860(1):157–65
 55. Hisanaga T, Okahashi K, Yamaoka S, Kajiwara T, Nishihama R, et al. 2019. A *cis*-acting bidirectional transcription switch controls sexual dimorphism in the liverwort. *EMBO J.* 38(6):e100240
 56. Hong X, Scofield DG, Lynch M. 2006. Intron size, abundance, and distribution within

- untranslated regions of genes. *Mol. Biol. Evol.* 23(12):2392–404
57. Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3(3):318–56
58. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, et al. 2014. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507(7490):124–28
59. Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE. 2008. SRA-domain proteins required for DRM2-mediated de novo DNA methylation. *PLOS Genet.* 4(11):e1000280
60. Juntawong P, Girke T, Bazin J, Bailey-Serres J. 2014. Translational dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*. *PNAS* 111(1):E203–12
61. Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, et al. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* 37(7):761–65
62. Kim D-H, Sung S. 2017. Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Dev. Cell* 40(3):302–12.e4
63. Kindgren P, Ard R, Ivanov M, Marquardt S. 2018. Transcriptional read-through of the long non-coding RNA *SVALK*A governs plant cold acclimation. *Nat. Commun.* 9(1):4561
- 63a. Kindgren P, Ivanov M, Marquardt S. 2020. Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in *Arabidopsis*. *Nucleic Acids Res.* 48(5):2332–47
64. Kiss T, Marshallsay C, Filipowicz W. 1991. Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and in vitro. *Cell* 65(3):517–26
65. Kowalczyk J, Palusinska M, Wroblewska-Swiniarska A, Pietras Z, Szewc L, et al. 2017. Alternative polyadenylation of the sense transcript controls antisense transcription of *DELAY OF GERMINATION 1* in *Arabidopsis*. *Mol. Plant* 10(10):1349–52
66. Krzyszton M, Zakrzewska-Placzek M, Kwasnik A, Dojer N, Karlowski W, Kufel J. 2018. Defective XRN3-mediated transcription termination in *Arabidopsis* affects the expression of protein-coding genes. *Plant J.* 93(6):1017–31
67. Lahmy S, Pontier D, Bies-Etheve N, Laudié M, Feng S, et al. 2016. Evidence for ARGONAUTE4-DNA interactions in RNA-directed DNA methylation in plants. *Genes Dev.* 30(23):2565–70
68. Lahmy S, Pontier D, Cavel E, Vega D, El-Shami M, et al. 2009. PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-

- specific subunit. *PNAS* 106(3):941–46
69. Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, et al. 2010. A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in *Arabidopsis*. *Curr. Biol.* 20(10):951–56
70. Law JA, Du J, Hale CJ, Feng S, Krajewski K, et al. 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498(7454):385–89
71. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE. 2011. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLOS Genet.* 7(7):e1002195
72. Lawit SJ, O’Grady K, Gurley WB, Czarnecka-Verner E. 2007. Yeast two-hybrid map of *Arabidopsis* TFIID. *Plant Mol. Biol.* 64(1–2):73–87
73. Lee H, Zhang Z, Krause HM. 2019. Long noncoding RNAs and repetitive elements: junk or intimate evolutionary partners? *Trends Genet.* 35(12):892–902
- 73a. Leng X, Ivanov M, Kindgren P, Malik I, Thieffry A, et al. 2020. Organismal benefits of transcription speed control at gene boundaries. *EMBO Rep.* 21(4):e49315
74. Leng X, Thomas Q, Rasmussen SH, Marquardt S. 2020. A G(enomic)P(ositioning)S(ystem) for plant RNAPII transcription. *Trends Plant Sci.* 25(8):744–64
75. Leppek K, Das R, Barna M. 2018. Functional 5’ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* 19(3):158–74
76. Li J, Liu C. 2019. Coding or noncoding, the converging concepts of RNAs. *Front. Genet.* 10:496
77. Li P, Tao Z, Dean C. 2015. Phenotypic evolution through variation in splicing of the noncoding RNA *COOLAIR*. *Genes Dev.* 29(7):696–701
78. Li Q, Gent JJ, Zynda G, Song J, Makarevitch I, et al. 2015. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *PNAS* 112(47):14728–33
79. Libri D. 2015. Endless quarrels at the end of genes. *Mol. Cell* 60(2):192–94
80. Lipovich L, Johnson R, Lin C-Y. 2010. MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. *Biochim. Biophys. Acta Gene Regul. Mech.* 1799(9):597–615
- 80a. Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, et al. 2004. Role of transposable

- elements in heterochromatin and epigenetic control. *Nature*. 430(6998):471–76
81. Liu F, Marquardt S, Lister C, Swiezewski S, Dean C. 2010. Targeted 3' processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. *Science* 327(5961):94–97
 82. Liu J, Jung C, Xu J, Wang H, Deng S, et al. 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24(11):4333–45
 83. Liu M, Ba Z, Costa-Nunes P, Wei W, Li L, et al. 2017. IDN2 interacts with RPA and facilitates DNA double-strand break repair by homologous recombination in *Arabidopsis*. *Plant Cell* 29(3):589–99
 84. Liu W, Duttke SH, Hetzel J, Groth M, Feng S, et al. 2018. RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in *Arabidopsis*. *Nat. Plants* 4(3):181–88
 85. Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, et al. 2014. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLOS Genet.* 10(1):e1003948
 86. Luo J, Hall BD. 2007. A multistep process gave rise to RNA polymerase IV of land plants. *J. Mol. Evol.* 64(1):101–12
 87. Marasco M, Li W, Lynch M, Pikaard CS. 2017. Catalytic properties of RNA polymerases IV and V: accuracy, nucleotide incorporation and rNTP/dNTP discrimination. *Nucleic Acids Res.* 45(19):11315–26
 88. Marquardt S, Raitskin O, Wu Z, Liu F, Sun Q, Dean C. 2014. Functional consequences of splicing of the antisense transcript *COOLAIR* on *FLC* transcription. *Mol. Cell* 54(1):156–65
 89. Marz M, Stadler PF. 2009. Comparative analysis of eukaryotic U3 snoRNA. *RNA Biol.* 6(5):503–7
 90. Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15(6):394–408
 91. Mocavini I, Di Croce L. 2020. RNA closing the Polycomb circle. *Nat. Genet.* 52(9):866–67
 92. Mohammadin S, Edger PP, Pires JC, Schranz ME. 2015. Positionally-conserved but sequence-diverged: identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. *BMC Plant Biol.* 15(1):217
 93. Nielsen M, Ard R, Leng X, Ivanov M, Kindgren P, et al. 2019. Transcription-driven chromatin repression of intragenic transcription start sites. *PLOS Genet.* 15(2):e1007969

94. Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation–dependent heterochromatin formation. *Cell* 120(5):613–22
95. Panda K, Ji L, Neumann DA, Daron J, Schmitz RJ, Slotkin RK. 2016. Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. *Genome Biol.* 17(1):170
96. Pikaard CS. 2002. Transcription and tyranny in the nucleolus: the organization, activation, dominance and repression of ribosomal RNA genes. *Arabidopsis Book* 1:e0083
97. Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, et al. 2005. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* 19(17):2030–40
98. Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell* 136(4):629–41
99. Preuss SB, Costa-Nunes P, Tucker S, Pontes O, Lawrence RJ, et al. 2008. Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins. *Mol. Cell* 32(5):673–84
100. Pumplun N, Sarazin A, Jullien PE, Bologna NG, Oberlin S, Voinnet O. 2016. DNA methylation influences the expression of *DICER-LIKE4* isoforms, which encode proteins of alternative localization and function. *Plant Cell* 28(11):2786–804
101. Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 443(7114):1008–12
102. Ream TS, Haag JR, Pontvianne F, Nicora CD, Norbeck AD, et al. 2015. Subunit compositions of *Arabidopsis* RNA polymerases I and III reveal Pol I– and Pol III–specific forms of the AC40 subunit and alternative forms of the C53 subunit. *Nucleic Acids Res.* 43(8):4163–78
103. Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, et al. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol. Cell* 33(2):192–203
104. Rheinberger H-J. 2006. A history of protein biosynthesis and ribosome research. In *Protein Synthesis and Ribosome Structure: Translating the Genome*, ed. KH Nierhaus, DN Wilson, pp. 1–51. Weinheim, Ger.: Wiley

105. Rigo R, Bazin J, Romero-Barrios N, Moison M, Lucero L, et al. 2020. The *Arabidopsis* lncRNA *ASCO* modulates the transcriptome through interaction with splicing factors. *EMBO Rep.* 21(5):e48977
106. Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81:145–66
107. Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M. 2018. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* 46(5):2169–84
108. Rosa S, Duncan S, Dean C. 2016. Mutually exclusive sense-antisense transcription at *FLC* facilitates environmentally induced gene repression. *Nat. Commun.* 7:13031
109. Rowley MJ, Rothi MH, Böhmendorfer G, Kuciński J, Wierzbicki AT. 2017. Long-range control of gene expression via RNA-directed DNA methylation. *PLOS Genet.* 13(5):e1006749
110. Saez-Vasquez J, Delseny M. 2019. Ribosome biogenesis in plants: from functional 45S ribosomal DNA organization to ribosome assembly factors. *Plant Cell* 31(9):1945–67
111. Seo JS, Diloknawarit P, Park BS, Chua N-H. 2019. ELF18-INDUCED LONG NONCODING RNA 1 evicts fibrillarlin from mediator subunit to enhance *PATHOGENESIS-RELATED GENE 1 (PRI)* expression. *New Phytol.* 221(4):2067–79
112. Seo JS, Sun H-X, Park BS, Huang C-H, Yeh S-D, et al. 2017. ELF18-INDUCED LONG-NONCODING RNA associates with Mediator to enhance expression of innate immune response genes in *Arabidopsis*. *Plant Cell* 29(5):1024–38
113. Shamovsky I, Nudler E. 2006. Gene control by large noncoding RNAs. *Sci. STKE* 2006(355):pe40
- 113a. Shearwin KE, Callen BP, Egan JB. 2005. Transcriptional interference—a crash course. *Trends Genet.* 21(6):339–45
114. Deleted in proof
115. Singh J, Mishra V, Wang F, Huang H-Y, Pikaard CS. 2019. Reaction mechanisms of Pol IV, RDR2, and DCL3 drive RNA channeling in the siRNA-directed DNA methylation pathway. *Mol. Cell* 75(3):576–89.e5
116. Srivastava AK, Lu Y, Zinta G, Lang Z, Zhu J-K. 2018. UTR-dependent control of gene expression in plants. *Trends Plant Sci.* 23(3):248–59
117. Struhl K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* 14(2):103–5

118. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. 2013. R-loop stabilization represses antisense transcription at the *Arabidopsis FLC* locus. *Science* 340(6132):619–21
119. Swiezewski S, Liu F, Magusin A, Dean C. 2009. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* 462(7274):799–802
- 119a. Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccara M, et al. 2009. A role for RNAi in the selective correction of DNA methylation defects. *Science*. 323(5921):1600–1604
120. Thieffry A, Vigh ML, Bornholdt J, Ivanov M, Brodersen P, Sandelin A. 2020. Characterization of *Arabidopsis thaliana* promoter bidirectionality and antisense RNAs by inactivation of nuclear RNA decay pathways. *Plant Cell* 32(6):1845–67
- 120a. Thomas QA, Ard R, Liu J, Li B, Wang J, et al. 2020. Transcript isoform sequencing reveals widespread promoter-proximal transcriptional termination in *Arabidopsis*. *Nat Commun*. 11(1):2589
- 120b. Tsuzuki M, Sethuraman S, Coke AN, Rothi MH, Boyle AP, Wierzbicki AT. 2020. Broad noncoding transcription suggests genome surveillance by RNA polymerase V. *PNAS* 117(48):30799–804
122. Tucker S, Vitins A, Pikaard CS. 2010. Nucleolar dominance and ribosomal RNA gene silencing. *Curr. Opin. Cell Biol.* 22(3):351–56
123. Tucker SL, Reece J, Ream TS, Pikaard CS. 2010. Evolutionary history of plant multisubunit RNA polymerases IV and V: subunit origins via genome-wide and segmental gene duplications, retrotransposition, and lineage-specific subfunctionalization. *Cold Spring Harb. Symp. Quant. Biol.* 75:285–97
124. van der Horst S, Snel B, Hanson J, Smeekens S. 2019. Novel pipeline identifies new upstream ORFs and non-AUG initiating main ORFs with conserved amino acid sequences in the 5' leader of mRNAs in *Arabidopsis thaliana*. *RNA* 25(3):292–304
126. Waibel F, Filipowicz W. 1990. U6 snRNA genes of *Arabidopsis* are transcribed by RNA polymerase III but contain the same two upstream promoter elements as RNA polymerase II-transcribed U-snRNA genes. *Nucleic Acids Res.* 18(12):3451–58
127. Wang H, Chung PJ, Liu J, Jang I-C, Kean MJ, et al. 2014. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* 24(3):444–53
128. Wang Z, Butel N, Santos-González J, Borges F, Yi J, et al. 2020. Polymerase IV plays a

- crucial role in pollen development in *Capsella*. *Plant Cell* 32(4):950–66
129. Warner JR. 1999. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* 24(11):437–40
130. Wei W, Ba Z, Gao M, Wu Y, Ma Y, et al. 2012. A role for small RNAs in DNA double-strand break repair. *Cell* 149(1):101–12
131. Wendte JM, Haag JR, Singh J, McKinlay A, Pontes OM, Pikaard CS. 2017. Functional dissection of the Pol V largest subunit CTD in RNA-directed DNA methylation. *Cell Rep.* 19(13):2796–808
132. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, et al. 2012. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the *Arabidopsis* epigenome. *Genes Dev.* 26(16):1825–36
133. Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135(4):635–48
134. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* 41(5):630–34
135. Wilkinson ME, Charenton C, Nagai K. 2020. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* 89:359–88
136. Wu H-W, Deng S, Xu H, Mao H-Z, Liu J, et al. 2018. A noncoding RNA transcribed from the *AGAMOUS* (*AG*) second intron binds to CURLY LEAF and represses *AG* expression in leaves. *New Phytol.* 219(4):1480–91
137. Xu W, Xu H, Li K, Fan Y, Liu Y, et al. 2017. The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat. Plants* 3(9):704–14
138. Yatushevich R, Fedak H, Ciesielski A, Krzyczmonik K, Kulik A, et al. 2017. Antisense transcription represses *Arabidopsis* seed dormancy QTL *DOG1* to regulate drought tolerance. *EMBO Rep.* 18(12):2186–96
139. Yu X, Martin PGP, Michaels SD. 2019. BORDER proteins protect expression of neighboring genes by promoting 3' Pol II pausing in plants. *Nat. Commun.* 10(1):4359
140. Yu Y, Jia T, Chen X. 2017. The “how” and “where” of plant microRNAs. *New Phytol.* 216(4):1002–17
141. Zanetti ME, Blanco F, Reynoso M, Crespi M. 2020. To keep or not to keep: mRNA stability

- and translatability in root nodule symbiosis. *Curr. Opin. Plant Biol.* 56:109–17
- 141a. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, et al. 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*. 153(1):193–205
142. Zhai J, Bischof S, Wang H, Feng S, Lee T-F, et al. 2015. A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. *Cell* 163(2):445–55
143. Zhang Y-C, Liao J-Y, Li Z-Y, Yu Y, Zhang J-P, et al. 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 15(12):512
144. Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. 2018. Global identification of *Arabidopsis* lncRNAs reveals the regulation of *MAF4* by a natural antisense RNA. *Nat. Commun.* 9(1):5056
145. Zheng Q, Rowley MJ, Böhmendorfer G, Sandhu D, Gregory BD, Wierzbicki AT. 2013. RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. *Plant J.* 73(2):179–89
146. Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, et al. 2014. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* 157(5):1050–60
147. Zhong X, Hale CJ, Law JA, Johnson LM, Feng S, et al. 2012. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* 19(9):870–75
148. Zhou M, Palanca AMS, Law JA. 2018. Locus-specific control of the de novo DNA methylation pathway in *Arabidopsis* by the CLASSY family. *Nat. Genet.* 50(6):865–73
149. Zhu Y, Rowley MJ, Böhmendorfer G, Wierzbicki AT. 2013. A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. *Mol. Cell* 49(2):298–309

Table 1 Plant lncRNAs discussed in this article. The lncRNAs are listed with their biological functions and observed lengths, and grouped by the RNA polymerase enzyme(s) required for their biogenesis.

lncRNA	Function	Enzyme	Length	Reference(s)
45S rRNA	rRNA precursor	Pol I	~8,400 nt	96
• 18S rRNA	Ribosomal SSU		1,800 nt	110
• 5.8S rRNA	Ribosomal LSU		161 nt	110
• 25S rRNA	Ribosomal LSU		3,380 nt	110
<i>AG-incRNA4</i>	Regulates <i>AG</i> gene expression	Pol II	1,000 nt	136
<i>APOLO</i>	Regulates <i>PID</i> gene expression	Pol II/Pol V	910 nt	6, 10
<i>ASCO</i>	Regulates alternative splicing	Pol II	786 nt	105
<i>COLDAIR</i>	<i>FLC</i> gene silencing during vernalization	Pol II	1,100 nt	52
<i>COLDWRAP</i>			~316 nt	62
<i>COOLAIR</i>			Unspliced: 6,000 nt Spliced: 558 nt, 318 nt	119
<i>DOG1 antisense (IGOD)</i>	Seed dormancy	Pol II	~300 nt	40
<i>ELENAI</i>	Pathogen response	Pol II	589 nt	111
<i>IPSI</i> (miRNA sponge)	Phosphate homeostasis	Pol II	542 nt	43
<i>SEP3</i> circRNA	R-loops, alternative splicing	Pol II	60 nt	26
<i>SUF</i> (intergenic and antisense to <i>FGMYB</i>)	Female identity in <i>Marchantia</i>	Pol II	Unspliced: 6,093 nt	55
<i>SVALKA</i>	Freezing response	Pol II	587 nt	63
<i>P5SM</i>	Pol III, TFIIA regulation	Pol II	~145 nt	49
Telomere RNA	Telomere maintenance in yeast (<i>TLC1</i>)	Pol II	1,157 nt	20a
	Telomere maintenance in plants (<i>AtTR</i>)	Pol III	268 nt	37
U6 snRNA	mRNA splicing	Pol III	102 nt	126
U3 snoRNA	Ribosome biogenesis in most other eukaryotes	Pol II	143–442 nt	89
	Ribosome biogenesis in plants	Pol III	~220 nt	89
tRNAs	Translation	Pol III	73–91 nt	25
5S rRNA	Ribosomal LSU	Pol III	120 nt	24
Pol IV transcripts	TE silencing, RdDM	Pol IV	~26–45 nt	13, 142
Pol V transcripts	TE silencing, RdDM	Pol V	~200 nt	133

Abbreviations: *AG*, *AGAMOUS*; *AtTR*, *Telomerase RNA subunit*; circRNA, circular RNA; *FLC*,

FLOWERING LOCUS C; *IPSI, INDUCED BY PHOSPHATE STARVATION1*; lncRNA, long noncoding RNA; LSU, large ribosomal subunit; *P5SM, 5S rRNA structural mimic*; RdDM, RNA-directed DNA methylation; *SEP3, SEPALLATA3*; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; SSU, small ribosomal subunit; *TLC1, TELOMERASE COMPONENT 1*; TE, transposable element.