

# Linking discoveries, mechanisms, and technologies to develop a clearer perspective on plant long noncoding RNAs

Kyle Palos <sup>1,\*</sup> Li'ang Yu <sup>1,†</sup> Caylyn E. Railey <sup>1,2,†</sup> Anna C. Nelson Dittrich <sup>1</sup> and Andrew D.L. Nelson <sup>1,\*</sup>

<sup>1</sup> Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA

<sup>2</sup> Plant Biology Graduate Field, Cornell University, Ithaca, NY 14853, USA

\*Author for correspondence: [krp75@cornell.edu](mailto:krp75@cornell.edu) (K.P.), [an425@cornell.edu](mailto:an425@cornell.edu) (A.D.L.N.)

†These authors contributed equally to this paper.

The author responsible for the distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plcell>) is Andrew D.L. Nelson ([an425@cornell.edu](mailto:an425@cornell.edu)).

## Abstract

Long noncoding RNAs (lncRNAs) are a large and diverse class of genes in eukaryotic genomes that contribute to a variety of regulatory processes. Functionally characterized lncRNAs play critical roles in plants, ranging from regulating flowering to controlling lateral root formation. However, findings from the past decade have revealed that thousands of lncRNAs are present in plant transcriptomes, and characterization has lagged far behind identification. In this setting, distinguishing function from noise is challenging. However, the plant community has been at the forefront of discovery in lncRNA biology, providing many functional and mechanistic insights that have increased our understanding of this gene class. In this review, we examine the key discoveries and insights made in plant lncRNA biology over the past two and a half decades. We describe how discoveries made in the pregenomics era have informed efforts to identify and functionally characterize lncRNAs in the subsequent decades. We provide an overview of the functional archetypes into which characterized plant lncRNAs fit and speculate on new avenues of research that may uncover yet more archetypes. Finally, this review discusses the challenges facing the field and some exciting new molecular and computational approaches that may help inform lncRNA comparative and functional analyses.

## Introduction

The basic definition of a long noncoding RNA (lncRNA) dictates that the RNA transcript must be 200 nucleotides or longer and not be translated into a protein. Traditional lncRNA definitions also exclude housekeeping RNAs, such as ribosomal (rRNA), transfer (tRNA), and small nuclear or nucleolar (sn/snoRNA; [Amaral et al. 2011](#)). This definition is problematic when considering RNAs that share similarity to portions of transposable or repetitive elements ([Cho 2018](#)) and those that give rise to small RNAs (sRNA) such as microRNAs (miRNAs), small interfering RNAs (siRNAs),

and phasiRNAs ([Wierzbicki et al. 2021](#)). While the latter precursor RNAs can be considered bona fide lncRNAs, here we will not heavily focus on these RNAs, as there are many other excellent reviews written by experts on these RNA classes and pathways ([Matzke and Mosher 2014](#); [Wang et al. 2019a](#); [Erdmann and Picard 2020](#); [Liu et al. 2020](#)). lncRNAs are also commonly defined based on the genomic context from which they are transcribed. For instance, lncRNAs are commonly separated into those that do not overlap other genes (intergenic lncRNAs or lincRNAs) and those that do. lncRNAs overlapping protein-coding genes are then separated

based on the strand of overlap (antisense vs. sense) and context of overlap (intronic and exonic; Rinn and Chang 2012; Ma et al. 2013). Finally, a sometimes contentious point of the lncRNA definition concerns their noncoding nature. Current identification efforts rely on ORF length, protein similarity, and machine learning (ML) algorithms to distinguish between coding and noncoding RNAs. However, as discussed below, there are a number of described proteins and lncRNAs that defy these coding/noncoding definitions. Thus, determining whether a transcript is a lncRNA is a nontrivial task requiring both computational and molecular approaches, but one with important implications for plant biology.

In this review, we highlight the major contributions that plant researchers have made to lncRNA biology. We describe how discoveries around plant lncRNAs lit the path towards our functional understanding of these enigmatic transcripts, and how technological and algorithmic improvements have increased the number of identified plant lncRNAs from hundreds to thousands. We introduce the predominant computational algorithms and pipelines used to identify lncRNAs, and discuss where there are still challenges in lncRNA identification and analysis. We examine how plant lncRNAs fit into the functional paradigms developed for eukaryotic lncRNAs, with a particular focus on transcriptional regulation, as this is the predominant functional archetype seen to date for plant lncRNAs. We then discuss what is known about how lncRNAs themselves are regulated, and end with what we believe are the exciting new areas in lncRNA research in which we believe plants will continue to make major contributions.

## Historical introduction to lncRNAs

### lncRNAs in the pregenomics era

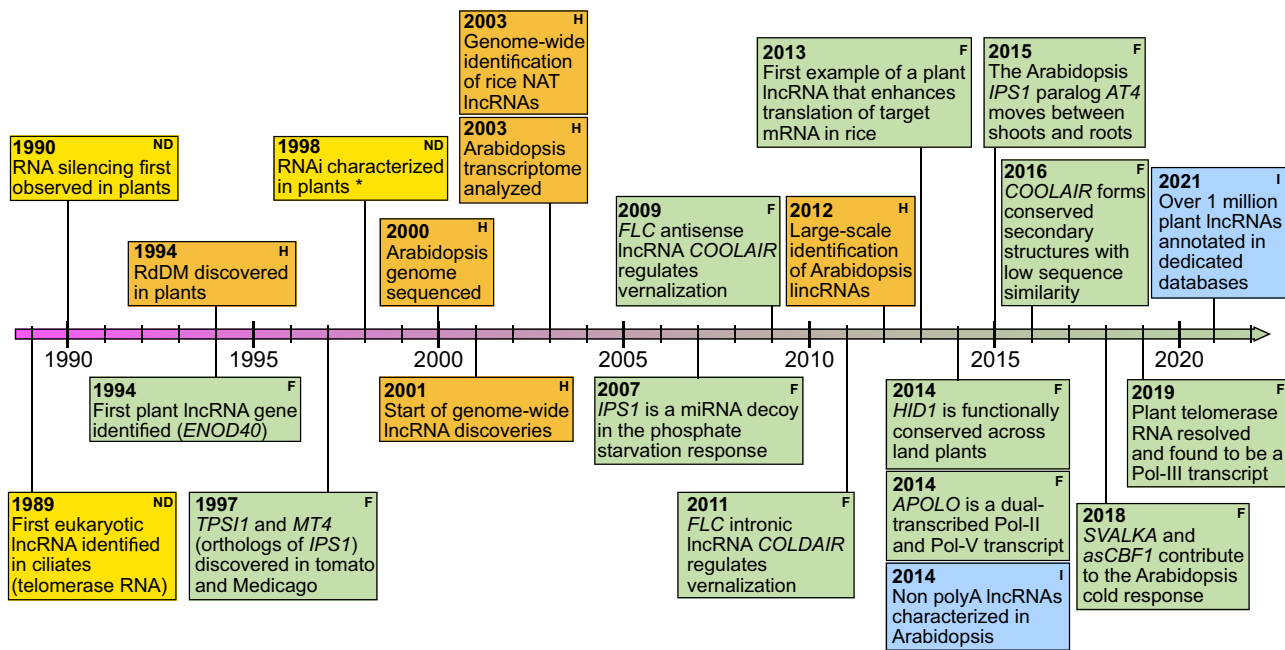
Plant biology has a rich history of supplying key discoveries in eukaryotic RNA biology (Fig. 1), including the initial observations of posttranscriptional gene silencing and stress-induced RNA-protein cytoplasmic aggregates, now referred to as stress granules (SGs; Nover et al. 1983; Matzke et al. 1989; Nover et al. 1989; Napoli et al. 1990). Plant-specific evolutionary innovations have also helped us understand how RNAs contribute to DNA methylation through RNA-dependent DNA methylation (RdDM; Wassenaar et al. 1994; Matzke and Mosher 2014). lncRNAs represent another class of RNA biology in which plant biologists have made seminal contributions to a field where foundational findings are generally attributed to nonplant model systems. Notably, many of these early lncRNAs were identified in agriculturally relevant species due to their involvement in physiologically important traits, highlighting the contributions even these nonmodel crop species have made to eukaryotic RNA biology.

Even prior to the sequencing of the *Arabidopsis* (*Arabidopsis thaliana*) genome in 2000 (Arabidopsis Genome Initiative 2000) and the beginning of the genomics era, a small number of lncRNAs were already emerging as

functional players in a wide range of cellular activities. Due to the molecular and genetic technologies available, these first lncRNAs were identified based on their biological, rather than mechanistic, functions. By 2000, at least five lncRNAs had been functionally described: *ENOD40* (EARLY NODULIN 40) in *Medicago truncatula*, *CR20* (CYTOKININ REPRESSED 20) in cucumber (*Cucumis sativus*), *GUT15* (GENE WITH UNSTABLE TRANSCRIPT 15) in tobacco (*Nicotiana tabacum*) and *Arabidopsis*, *MT4* in *M. truncatula*, and *TPS11* (TOMATO PHOSPHATE STARVATION INDUCED 1) in tomato (*Solanum lycopersicum*). *ENOD40*, independently discovered in both *M. truncatula* and soybean (*Glycine max*), was the first lncRNA discovered in plants and plays a role in root nodulation in legumes (Crespi et al. 1994). Meanwhile, *MT4* and *TPS11* are part of a dicot-conserved family of lncRNAs that respond to phosphate stress and, as we discuss below, contribute to appropriate responses to phosphate starvation (Burleigh and Harrison 1997; Burleigh and Harrison 1998; Burleigh and Harrison 1999; Liu et al. 1997; Bari et al. 2006; Franco-Zorrilla et al. 2007). *CR20* and *GUT15* represent a family of lncRNAs that are hormonally regulated, alternatively spliced, and conserved across angiosperms, yet there is limited functional data on these lncRNAs (Teramoto et al. 1996; MacIntosh et al. 2001; Plewka et al. 2018).

At the time of their discovery, it was unclear to the community if there was an RNA-specific molecular mechanism for these noncoding functional RNAs, as this ran counter to the predominant protein-centric views of molecular biology. However, many important observations were made to suggest a functional role for these early lncRNAs that was independent of any potential coding sequence. For instance, Crespi and co-authors determined that *ENOD40* likely performed its role as an RNA rather than a protein based on its free energy of folding being more similar to other noncoding RNAs compared to coding RNAs. And, despite transient overexpression of the soybean *ENOD40* resulting in the translation of a small signaling peptide, no peptide was observed from in vitro translation experiments or under native conditions in vivo (Crespi et al. 1994; van de Sande et al. 1996). These data, paired with the observation that *ENOD40* is most conserved outside its ORF, continue to support the model of *ENOD40* acting as a lncRNA. Thus, based on the definitions of noncoding RNAs of the era, these five genes represented puzzling, but exciting examples of functional lncRNAs.

Many experiments and observations from these early studies in plants laid the foundation for our current knowledge and definition of lncRNAs across all eukaryotes. While it was unclear to what degree these noncoding RNAs were present in plant transcriptomes, commonalities among them served as the basis for future identification efforts. For instance, these lncRNAs were typically expressed under very specific cellular or environmental conditions, a characteristic that holds true for many lncRNAs identified since. It is unclear if early researchers knew that lncRNAs, in general, were lowly expressed, as many of the early identified



**Figure 1.** Important plant lncRNA discoveries over the past three decades. A timeline of seminal RNA and lncRNA discoveries with an emphasis on plant-specific pathways and mechanisms. \*RNA interference and its characterization in plants were not covered in this review (Voinnet et al. 1998; Waterhouse et al. 1998). Green boxes with the letter 'F' in the top corner denote discoveries described in the plant lncRNA functional mechanisms section. Blue boxes with the letter 'I' in the top corner denote findings discussed in the section on major issues associated with lncRNA identification. Orange boxes with the letter 'H' in the top corner are discussed in the historical introduction, and yellow boxes with the letters 'ND' in the top corner are not explicitly discussed in the review, but place plant lncRNA discoveries in a greater context.

transcripts were likely the most abundant of the total lncRNA pool. At a molecular level, these early lncRNAs displayed mRNA-like attributes such as 5' caps and polyadenosine tails and were considered to be mRNA-like ncRNAs (Rymarquis et al. 2008). However, this is likely biased by how the original lncRNAs were identified and is not representative of all lncRNAs (discussed below). While the essential definition has not changed since these initial discoveries, we have more clarity on what it means to be a plant lncRNA, a more standardized definition, and much more functional data to guide mechanistic experiments.

### lncRNAs in the genomics era

In the decade that followed the sequencing of the Arabidopsis genome (Arabidopsis Genome Initiative 2000), there were numerous studies that noted widespread transcription in unannotated regions of plant genomes (Yamada et al. 2003; Meyers et al. 2004; Li et al. 2006; Chekanova et al. 2007; Li et al. 2007; Matsui et al. 2008; Okamoto et al. 2010). These observations aligned with reports from transcriptomic studies in animals, suggesting that pervasive transcription is a common feature in eukaryotic genomes (ENCODE Project Consortium et al. 2007; Kapranov et al. 2007). The first direct attempt at genome-wide plant lncRNA identification took place soon after the Arabidopsis genome sequence was published, when MacIntosh and co-authors identify 39 noncoding RNA candidates from two Arabidopsis expressed sequence tag (EST)

collections representing ~20,000 polyadenylated and size-selected transcripts (MacIntosh et al. 2001). Other analyses followed, primarily using ORF length (<100 amino acids (AA)) and similarity to known proteins, to identify a small number of mRNA-like noncoding RNAs in Arabidopsis and *M. truncatula* (Riaño-Pachón et al. 2005; Wen et al. 2007).

After these initial studies and during the following decade, our view of the lncRNA portion of the Arabidopsis transcriptome became clearer. Numerous groups, using varying scopes, technologies, and computational methodologies, identified suites of Arabidopsis lncRNAs (Marker et al. 2002; Osato et al. 2003; Wang et al. 2005; Hirsch et al. 2006; Wang et al. 2006; Amor et al. 2009; Song et al. 2009; Swiezewski et al. 2009). Most of these studies still relied on the ever-expanding databases of mRNA-focused ESTs and full-length cDNA sequences generated by a variety of groups and consortiums (Pontius et al. 2003; Castelli et al. 2004; Sakurai et al. 2005), however, their scope and methodology distinguished these studies. Some, such as Wang et al. (2005) and Wang et al. (2006), used full-length cDNA sequences from UniGene and RIKEN databases to identify ~3,000 natural antisense transcripts (NATs). These NATs might share sequence complementarity in *cis* (directly overlapping a gene) or in *trans* (complementary sequences at separate loci). Other groups, such as Hirsch et al. (2006) and Amor et al. (2009), used a variety of strict criteria to identify high-confidence nonprotein-coding RNAs, including high GC content, other nucleotide biases, stable RNA structures, and

features that would preclude successful translation. These findings uncovered a number of important functional lncRNAs. For example, the *APOLO* and *ALTERNATIVE SPLICING COMPETITOR* (*ASCO*) lncRNAs (discussed below) were both annotated as lncRNAs by Amor et al. (2009) prior to their biological function being described. In addition, algorithmic advances made it easier to annotate protein-coding genes accurately by incorporating comparative and transcriptomic information, e.g. *MAKER* (Holt and Yandell 2011), thereby making it easier to distinguish between unannotated protein-coding genes and lncRNAs. Thus, even before the large-scale adoption of RNA-sequencing in the 2010s, it was clear that thousands of transcriptionally active putative lncRNA loci existed in plant genomes.

### **lncRNAs in the next generation RNA-sequencing era**

As transcriptomic technologies improved in the 2010s (e.g. higher density tiling arrays and next generation RNA-sequencing), it became cheaper and easier to perform more comprehensive lncRNA identification efforts that spanned numerous tissues or conditions. One of the first of these studies came from Liu et al. (2012) in which they used 200 publicly available and custom-made tiling arrays, as well as their own RNA-seq data from four tissues to identify nearly 7,000 lincRNAs and a similar number of NAT-lncRNAs. These public arrays targeted poly-A RNA from 14 *Arabidopsis* mutants, 18 heat treatments, and 6 different plant tissues. Importantly, reproducibility across multiple tiling arrays was used as a criterion to generate high-confidence lncRNAs. Finally, Liu et al.'s study used both tiling array technology and RNA-seq, allowing for a near-direct comparison of technologies for lncRNA discovery. While ~40% of their 6,480 lincRNAs had some measure of RNA-seq-based transcriptional support, fewer than 300 of them were fully supported by their relatively deep sequencing (~250 million reads/tissue). This updated annotation was instrumental in the discovery of many now functionally characterized lncRNAs, such as *lncCOBRA1*, *FLORE*, *DRIR*, and *AGAMOUS* (*AG*)-*lncRNA4* (Henriques et al. 2017; Qin et al. 2017; Wu et al. 2018; Kramer et al. 2022). In addition, these data highlighted differences between tiling arrays and RNA-seq, and also pointed to the necessity for sequencing breadth being just as, or more, important as depth when attempting to comprehensively identify lncRNAs.

A number of other lncRNA identification efforts closely followed the initial work by Liu et al. (2012). Importantly, these included further efforts in *Arabidopsis* and many agriculturally important species (Boerner and McGinnis 2012; Moghe et al. 2013; Li et al. 2014a; Zhang et al. 2014; Shumayla et al. 2017). Similar findings to those in *Arabidopsis* were found in maize (*Zea mays*) by Boerner and McGinnis where they identified ~2,500 lncRNAs using a public dataset of full-length cDNA sequences (Boerner and McGinnis 2012). About half of these novel lncRNAs were categorized as siRNA precursors. This finding differs from *Arabidopsis* studies, where Liu et al. found that ~2.5% of their identified

lincRNAs were sRNA associated. In support of Boerner and McGinnis' study, a follow-up lncRNA identification study in Maize using ESTs and RNA-seq from diverse tissues found evidence for over 20,000 lncRNAs of which more than 90% were sRNA precursors (Li et al. 2014b). In the past decade since Liu et al. RNA-seq-based lncRNA identification efforts have expanded to almost every model or agronomically relevant plant species, with remarkably consistent findings, lncRNAs are abundant but lowly expressed, making their identification difficult but worth the effort.

## **Predominant methods of identifying lncRNAs in plants**

The flood of transcriptional evidence supplied by next generation RNA-sequencing data necessitated improvements in how lncRNAs are identified from these data. Traditionally, lncRNA identification pipelines discard transcripts based on size (<200 nts), abundance (varies, e.g., less than 1 transcript per million), similarity to known genes (e.g., using Rfam (an RNA family database, Griffiths-Jones et al. 2003) and Pfam (The protein families database, Mistry et al. 2021), and ORF length (>100 AA). While imperfect, this approach helped to identify a number of functionally important lncRNAs in eukaryotes (Cabili et al. 2011; Liu et al. 2012). Over time, additional approaches have been developed that build on characteristics of previously identified lncRNAs to better determine what is coding and noncoding. In this section, we summarize some of the computational resources, algorithms, and strategies that have been developed over the past decade to aid in lncRNA annotation and functional prediction.

### **Algorithms**

To date, there are more than 30 distinct algorithms and pipelines developed for lncRNA identification. In general, many of these algorithms (Table 1) rely on sequence intrinsic features, such as ORF length and composition, coding potential, and sequence decomposition (i.e. k-mers). Additionally, some of these bioinformatic packages require reference genome and annotation files (referred to as *alignment-based*), whereas others do not (*alignment-free*; Table 1). Alignment-based algorithms such as CPC2 and PlncPRO tend to be faster and more accurate, as the input transcripts typically lack sequencing errors and have a more accurate gene structure than those used in alignment-free methods (i.e. de novo assembly; Kang et al. 2017; Singh et al. 2017). However, in nonmodel systems where genomes or genome annotations are lacking, alignment-free approaches such as PLEK (predictor of long non-coding RNAs and messenger RNAs based on an improved k-mer scheme) and CNCI (Coding-Non-Coding Index) are quite useful (Sun et al. 2013; Li et al. 2014a; Schneider et al. 2017; Guo et al. 2019). Of note: for all algorithms, the input information is a set of transcript sequences—thus alignments are not necessary, but alignment-free methods have been explicitly



**Table 1.** Algorithms and pipelines used to identify lncRNAs from RNA-sequencing data

Name	Method (specific algorithm)	Features for classification	Organism	Manner	References
Annocript	Genomic feature-based	Homology, ORF lengths, other	Animals	Alignment-based	<a href="#">Musacchia et al. (2015)</a>
BASINET	DL (decision tree on complex network)	Topological measures of sequence networks	Vertebrates	Alignment-free	<a href="#">Ito et al. (2018)</a>
CNCI	ML (support vector machine (SVM))	Adjoining nucleotide triplets, ORF structure, other	Vertebrates	Alignment-free	<a href="#">Sun et al. (2013)</a>
CNIT	ML (XGBoost)	66 features, including most-like CDS, adjoining nucleotide triplets, etc.	Animals and plants	Alignment-based	<a href="#">Guo et al. (2019)</a>
COME	ML (random forest)	Nine features, including sequence-derived, expression, and histone features	Human	Alignment-based	<a href="#">Hu et al. (2007)</a>
CPAT	Logistic regression	ORF size and coverage, Fickett score, hexamer usage	Animals	Alignment-free	<a href="#">Wang et al. (2013)</a>
CPC2	ML (random forest)	ORF length, integrity, Fickett score, isoelectric point	Animals	Alignment-free	<a href="#">Kang et al. (2017)</a>
CREMA	ML (multiple)	mRNA length, ORF length, GC content, conservation, other	Animals and plants	Alignment-based	<a href="#">Simopoulos et al. (2019)</a>
DeepLnc	DL (deep neural networks)	K-mer	Human	Alignment-free	<a href="#">Tripathi et al. (2016)</a>
Evolinc	Pipeline	Utilizes CPC2 and common lncRNA heuristics	Animals and plants	Alignment-based	<a href="#">Nelson et al. (2017)</a>
FEElnc	ML (random forest)	Relation to known transcripts, ORF characteristics, other	Mammals	Alignment-free	<a href="#">Wucher et al. (2017)</a>
iSeeRNA	ML (SVM)	Conservation, ORF length, ORF proportion	Animals	Alignment-based	<a href="#">Sun et al. (2013)</a>
LGC	ML	ORF length and GC content	Animals and plants	Alignment-free	<a href="#">Wang et al. (2019a)</a>
LncADeep	DL (deep belief network)	Sequences intrinsic and homology feature in a deep belief network	Human	Alignment-free	<a href="#">Yang et al. (2018)</a>
LncFinder	ML (multiple)	ORF, structure, physiochemical property, other	Animals and plants	Alignment-free	<a href="#">Han et al. (2019)</a>
lncRNA-ID	ML (random forest)	Ribosome interaction, protein conservation features, other	Animals	Alignment-based	<a href="#">Achawanantakun et al. (2015)</a>
lncRNA-MFDL	DL (deep stacking networks)	ORF, k-mers, structure, other	Human	Alignment-free	<a href="#">Fan and Zhang (2015)</a>
lncRNA-screen	Genomic feature-based	Relation to known transcripts, no small RNA overlap, other	Animals	Alignment-based	<a href="#">Gong et al. (2017)</a>
lncRNA-net	DL (recurrent neural network)	Intrinsic features extracted by recurrent neural networks	Human	Alignment-free	<a href="#">Baek et al. (2018)</a>
lncRScan-SVM	ML (SVM)	Transcript length, stop codon presence, conservation, other	Animals	Alignment-based	<a href="#">Sun et al. (2015)</a>
LncScore	Logistic regression	Coding potential, ORF characteristics, exon hexamers, and GC content	Animals	Alignment-free	<a href="#">Zhao et al. (2016)</a>
longdist	ML (SVM)	Nucleotide pattern frequencies, ORF characteristics	Animals	Alignment-based	<a href="#">Schneider et al. (2017)</a>
PLEK	ML (SVM)	K-mer	Vertebrates	Alignment-free	<a href="#">Li et al. (2014a)</a>
PLIT	ML (iterative random forests)	ORF characteristics, codon-bias, other	Plants	Alignment-free	<a href="#">Deshpande et al. (2019)</a>
PLncPRO	ML (random forest)	Homology to known proteins, codon-bias, length, other	Plants	Alignment-based	<a href="#">Singh et al. (2017)</a>
PlncRNA-HDeep	DL	Sequence composition encoded into vectors	Plants	Alignment-free	<a href="#">Meng et al. (2021)</a>
PORTRAIT	ML (SVM)	Nucleotide composition, translated ORF characteristics, other	Fungi	Alignment-based	<a href="#">Arrial et al. (2009)</a>
RNAplonc	ML (multiple)	16 features including K-mer, sequence length, GC content, coding potential	Plants	Alignment-free	<a href="#">Da Negri et al. (2019)</a>

Abbreviations: BASINET, BiologicAI Sequences NETWORK; CDS, coding sequence; CNIT, Coding-Non-Coding Identifying Tool; COME, a coding potential calculation tool based on multiple features; CPAT, Coding-Potential Assessment Tool; CREMA, Classifying RNA by Ensemble Machine learning Algorithm; DL, Deep learning; FEElnc, FEXible Extraction of lncRNAs; lncRNA-ID, Long non-coding RNA IDentification; LGC, ORF Length and GC content.

designed to overlook sequencing or transcript assembly errors that might influence feature comparisons.

The algorithms incorporated into many researcher's lncRNA identification workflows, such as the commonly

used CPC2 (coding potential calculator 2) ([Kang et al. 2017](#)), were developed and trained predominantly on vertebrate lncRNAs. Given lineage-specific genomic differences (e.g. GC content), models trained on vertebrate lncRNAs

may incorrectly assign plant lncRNAs. Recognizing this potential issue, a few tools (Table 1) have been developed and optimized specifically for plant lncRNA identification, including PlncRNA-HDeep (Plant lncRNA hybrid deep learning model) and RNAplonc, which both employ deep learning approaches (Da Negri et al. 2019; Meng et al. 2021), as well as PlncPRO (Plant Long Non-Coding Rna Prediction by Random Forests) and PLIT (Plant lncRNA Identification Tool) which utilize random forest models (Singh et al. 2017; Deshpande et al. 2019). While each of these algorithms was more accurate in predicting plant lncRNAs, it is likely how the models were trained (on plant lncRNAs), rather than the machine learning approach itself that lends them this higher accuracy. A comparison of each of these tools using the same well-curated set of plant lncRNAs would be useful in determining which was the most appropriate for a given set of input transcripts.

### Integrative pipelines

In efforts to streamline lncRNA discovery and evolutionary analysis across large datasets, pipelines, such as Evolinc (Nelson et al. 2017), have been developed. Evolinc utilizes multiple out of the box ML algorithms (e.g. CPC2) as well as traditional heuristics (e.g. transcript length and similarity to known proteins) for lncRNA identification and is simple to use, but generally is not amenable to changes in the underlying filtering mechanisms and reliant on predefined lncRNA features. Another useful aspect of Evolinc is the evolutionary portion of the pipeline, which searches for sequence homologs in a user-defined set of related species (Nelson Ditttrich and Nelson 2022). Another useful pipeline is lncRNA-screen (Gong et al. 2017), which is designed to incorporate additional genomic features, such as histone marks, HiC data, and transcript abundance, to better functionally annotate already identified lncRNAs. Similarly, lncADeep is a deep learning approach to identify and functionally annotate lncRNAs, and furthermore, infer putative protein interaction partners based on numerous sequence characteristics (Yang et al. 2018). As the community settles on a few primary lncRNA identification algorithms, these simplified pipelines will likely become automated aspects of genome annotation projects.

The increasing number of novel algorithms and computational pipelines for annotating lncRNAs has invited challenging discourse around “best” identification practices. Some pressing questions at the moment are: How do users determine which method is best suited to their data and how does the community compare lncRNAs identified using different methods? One potential solution would be for each research group annotating a novel lncRNA, or group of RNAs, to utilize multiple approaches and take the intersection of those approaches (i.e. predictions shared between those approaches). However, selecting seemingly disparate algorithms that utilize different lncRNA features for identification raises additional concerns about which is most appropriate. In addition, there are potentially species-specific tradeoffs to

consider, particularly when annotating lncRNAs in nonmodel plant systems lacking reference genomes. Possible solutions would be for the plant community to (1) develop a common test dataset for benchmarking purposes and (2) have multiple groups collaborate to perform this benchmarking on extant algorithms to compare performance and accuracy.

### Major issues associated with lncRNA identification in the genomics era

The many attempts at annotating lncRNAs in plant transcriptomes have revealed a number of key properties that have made their identification problematic. One of these issues pertains to sequencing resolution and appropriate supporting data at cellular, tissue, and organismal scales. For example, even though most model plant species have sufficient transcriptomic data across all major plant organs and developmental stages, very few concerted efforts have been undertaken to utilize these data to annotate lncRNAs. In addition, companion data that would provide greater support for lncRNA annotation, such as information about transcription start sites (cap analysis of gene expression (CAGE); Kawaji et al. 2014), structure and protein interaction (protein interaction profiling sequencing; Silverman et al. 2014), and degradation (e.g., parallel analysis of RNA ends or genome-wide mapping of uncapped and cleaved transcripts; Gregory et al. 2008; German et al. 2009; Willmann et al. 2014), are often limiting. This restricts the confident development of lncRNA repertoires for many species to only a few commonly sampled tissues, often leaves or seedlings. Because of the importance of comparative approaches to characterize lncRNAs, expanding the tissues and environmental conditions used to identify and profile their expression will help to propel their functional characterization. This should be a priority for plant lncRNA biology as much of these data already exist and are publicly available.

### Sampling bias

lncRNAs have classically been thought of as mRNA-like transcripts. As a result, most identification efforts have sampled polyadenylated (polyA) pools of RNA. However, polyA-focused studies may overlook a significant fraction of the ncRNA transcriptome and thus ignore biologically significant lncRNAs. In animals, there are thousands of non-polyA noncoding transcripts which do not fall into the classical definition of housekeeping RNAs (rRNA, tRNA, sn/snoRNA; Livyatan et al. 2013). The initial characterization efforts of non-polyA plant ncRNAs came in 2013 and 2014 (Liu et al. 2013; Di et al. 2014; Wang et al. 2014b), with Di and co-authors focusing specifically on non-polyA lncRNAs in Arabidopsis. Di et al. used rRNA depletion and polyA minus RNA-seq in four stresses to identify lncRNAs lacking polyA tails. More recently, a combined transcriptomic approach that sampled both polyA and non-polyA lncRNAs

and their localization within the cell uncovered a suite of stress-responsive lncRNAs, including MAS (discussed below; Zhao et al. 2018). The biological significance of these non-polyA and likely nonpolymerase II (non-Pol-II) transcribed ncRNAs is hinted at by the discovery of the highly conserved lncRNA telomerase RNA (TR), a Pol-III transcript whose discovery relied on surveying ribosomal depleted RNA-seq libraries (Fajkus et al. 2019; Song et al. 2019; Dew-Budd et al. 2020). Separately, some of these non-polyA transcripts may be associated with or are produced by the RdDM pathway and therefore have implications in epigenetic silencing (Tsuzuki et al. 2020). Indeed, the RdDM pathway, and plant sRNA pathways in general, may be substantial contributors to the overall lncRNA pool in plants. Thus, we believe that it may be necessary to shift RNA-sequencing efforts to ribodepleted libraries in order to capture the noncoding RNA portions of plant transcriptomes in a more complete manner.

### Low expression

Perhaps the greatest difficulty in annotating lncRNAs comes from their intrinsic characteristics. Unannotated transcripts that are fed into lncRNA identification pipelines are typically mono-exonic transcripts with low expression, and may have insufficient data to infer gene structure. Determining which of these lowly expressed, mono-exonic transcripts fit into the definition of a lncRNA is further complicated by the relatively broad, and nonstandardized, definition of a lncRNA. For example, a typical deep RNA-sequencing experiment identifies thousands of unannotated and lowly expressed transcripts which fit the criteria of a lncRNA (Liao et al. 2017; Wang et al. 2020). However, many of these transcripts have inconsistent expression patterns and may not appear in independent RNA-sequencing data from the same tissues (Palos et al. 2022). In addition, their low expression makes it difficult to distinguish them algorithmically from transcriptional noise derived from deep sequencing. It is important to note here that low expression does not equate to lack of function, as a lncRNA may have rapid turnover or be functional at very low stoichiometries (Unfried and Ulitsky 2022). Emerging work in human cells suggests that many lncRNAs may function by initiating liquid–liquid phase separation, a phenomenon whereby distinct membrane-less compartments form within the cell and contribute to cellular stability. These compartments would also explain the substoichiometric nature of lncRNA copy-number and function within the cell (Guo et al. 2021; Wu et al. 2021). To navigate issues assembling lowly expressed transcripts, some groups have suggested filtering out single exon transcripts, as well as those that are within 500 base pairs of another gene (Cabili et al. 2011; Cemel et al. 2017). These are rational and conservative decisions but do not work well with smaller genomes, such as Arabidopsis, and fail to consider the large number of functionally described mono-exonic lncRNAs in plants (Franco-Zorrilla et al. 2007; Wang et al. 2014a; Fajkus et al. 2019). In short, there always seem to be exceptions to the most carefully thought-out lncRNA definition. Thus,

until high-throughput genetic screens such as Perturb-seq (Adamson et al. 2016) are translated to plant systems, we would argue that replication and variability in expression in response to stimuli are key. To sum up, a high-confidence lncRNA must be identified in multiple experimental and biological replicates, and the expression of the lncRNA should be induced under specific circumstances. As mentioned below, the induction pattern can assist in functional prediction.

### Poor sequence conservation

Beyond Arabidopsis and other model plants, the increased number of sequenced genomes and transcriptomes largely enabled the comparative and evolutionary studies of plant lncRNAs (Mohammadin et al. 2015; Nelson et al. 2016; Simopoulos et al. 2019; Corona-Gomez et al. 2020; Fesenko et al. 2021; Zhu et al. 2022). These comparative analyses suggest that plant lncRNAs are more evolutionarily labile, with much shorter apparent evolutionary half-lives than those seen in vertebrates (Cabili et al. 2011; Necșulea et al. 2014; Washietl et al. 2014; Hezroni et al. 2015). Sequence homologs are difficult to find in even closely related species for most plant lncRNAs. In addition, homology does not necessarily coincide with collinearity (synteny). Interestingly, sequence-divergent lncRNAs, transcribed adjacent to orthologous protein-coding genes, have been observed in the mustards, suggesting that transcriptional conservation may be more important than sequence for certain lncRNAs (Walden et al. 2020; Palos et al. 2022). For the subset of lncRNAs with identifiable homologs, conserved domains, structures, splice sites, and interaction partners have been discerned. Interestingly, even these conserved lncRNAs are rarely identified through sequence-based homology searches outside of a plant family, requiring more detailed co-variation and structure-based models for homology inference (Hawkes et al. 2016; Fajkus et al. 2019, 2021). Thus, plant lncRNAs may be divided into at least three evolutionary classes: (1) the species-specific lncRNAs that appear to make up the majority of lncRNA populations, (2) the lncRNAs with potential transcriptional and positional conservation, but little sequence conservation, and (3) the much smaller number of lncRNAs with easily observable modes of conservation more typical of protein-coding genes. These three evolutionary classes are likely linked to functional mechanism and biological significance, and will help guide future exploration into plant lncRNAs. Thus, developing high-throughput comparative tools to evolutionarily classify lncRNAs will be critical for future functional work.

### Lack of a common lncRNA definition

Difficulties in coalescing around a common definition for lncRNAs are even visible in the repositories meant to serve the plant lncRNA community. There are three comprehensive and plant-specific lncRNA databases, with over one million lncRNAs across close to 100 species, that have been developed in the last 5 yr (Szczeniuk et al. 2019; Jin et al. 2021a; Di Marsico et al. 2022). GreeNC (<http://greenc>.

sequentiabiotech.com/wiki2/; Di Marsico et al. 2002) is unique among these databases in that it utilized a purely in silico approach to identify lncRNAs from previously annotated transcripts. They identified ~500,000 putative lncRNAs from 94 plant and algae species. A unique and helpful resource that GreeNC provides is their description of lncRNA orthogroups which facilitates accessible comparative analyses and prediction for function (Di Marsico et al. 2022). A notable feature missing from GreeNC is the information pertaining to lncRNA expression. In contrast, both CANTATadb (<http://cantata.amu.edu.pl/>; Szcześniak et al. 2019) and PLncDB ([www.tobacodb.org/plncdb/](http://www.tobacodb.org/plncdb/); Jin et al. 2021a) have curated publicly available expression data to predict lncRNAs. To date, over 1.2 million lncRNAs across 80 plant species and over 200,000 lncRNAs across nearly 40 plant species are represented in PLncDB and CANTATadb, respectively. The data curated by PLncDB are especially relevant for hypothesis generation, as it includes experimental, expression, and intermolecular regulatory network information. While the developers of PLncDB annotated an expansive repertoire of lncRNAs utilizing diverse sets of sequencing data, those behind CANTATadb arguably took a more conservative approach in data utilization and lncRNA identification. lncRNAs in the GreeNC and PLncDB databases were defined with an ORF cutoff of 120 AA, whereas those in CANTATadb were defined by an ORF cutoff of 100 AA. Additionally, only paired-end sequencing data were used to identify lncRNAs for the CANTATadb, which improves read mapping quality and transcript assembly. While the utility of these resources is apparent, the lack of harmonization across them, as well as the different ways in which they define lncRNAs, makes it difficult to easily compare across them. Both to facilitate discovery and reduce upkeep costs, we would propose a unified database that was linked to species-specific genomic resources (e.g. the Arabidopsis information resource or MaizeGDB; Swarbreck et al. 2008; Portwood et al. 2019).

### Plant lncRNA functional mechanisms

Many of the functionally characterized plant lncRNAs are predominantly nuclear-localized and exhibit some role in transcriptional or posttranscriptional gene regulation (Fig. 2). For the sake of clarity, we have separated these mechanisms below, with an additional section for lncRNAs whose function falls outside of regulating gene expression. Each of the described lncRNAs, their functional archetype, and when and where they were identified, can be found in Table 2.

## Pretranscriptional regulation of gene expression by lncRNAs

### Modulating expression through the formation of chromatin loops

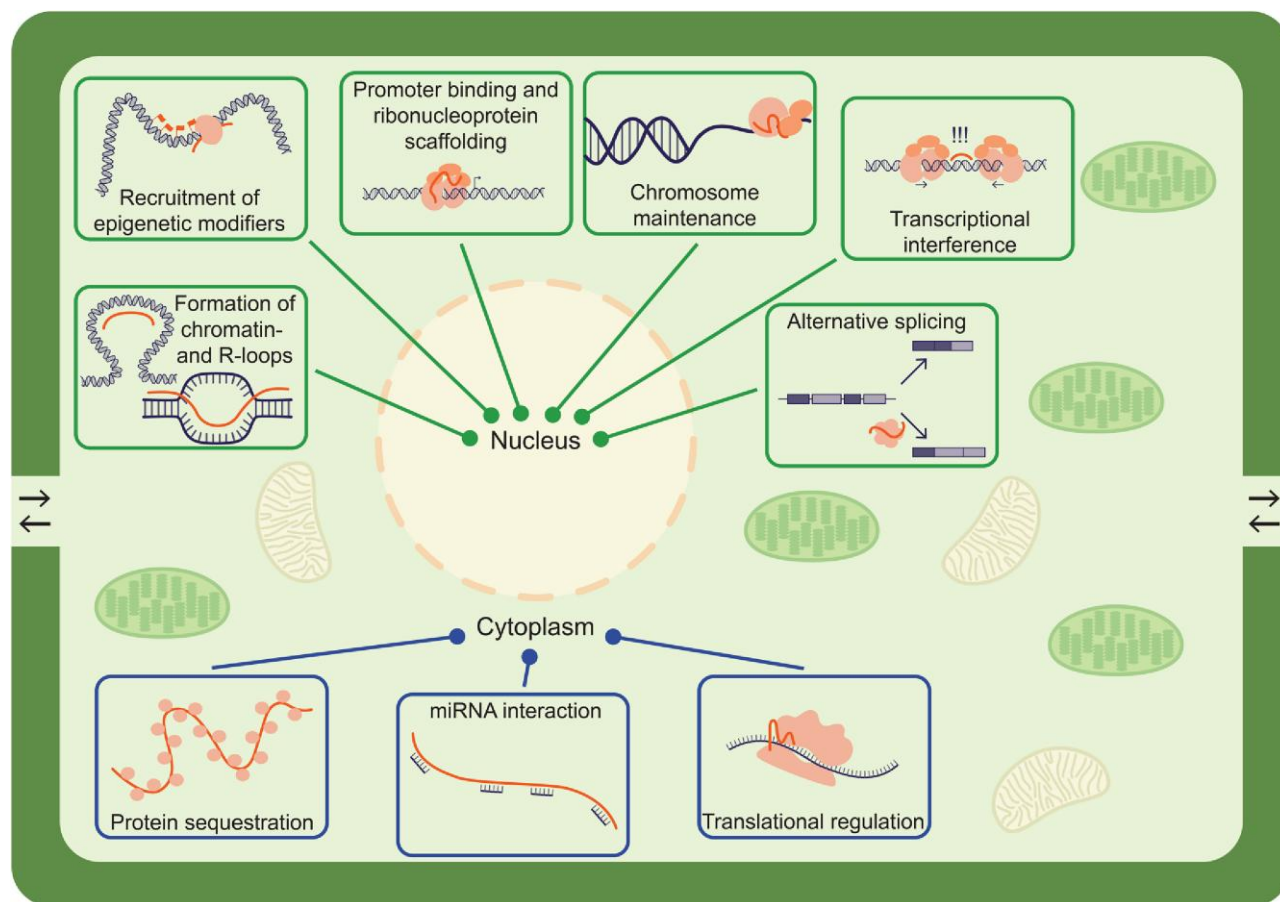
lncRNAs can mediate changes in gene expression through alterations to chromatin topology. One prominent and well-

studied example of this functional class is the lncRNA APOLO (AUXIN REGULATED PROMOTER LOOP) and its role in regulating the expression patterns of auxin-responsive genes (Ariel et al. 2014, 2020; Mas and Huarte 2020; Moison et al. 2021). Upon auxin treatment, the APOLO locus undergoes chromatin relaxation. This results in RNA Pol-II expression of both APOLO and upstream neighbor, PINOID (PID), a protein kinase associated with polar auxin transport (Friml et al. 2004). In a self-regulating cycle, Pol-II APOLO accumulation encourages the recruitment of RNA polymerase V (Pol-V) and the production of Pol-V APOLO variants. Accumulation of the Pol-II APOLO variant leads to recruitment of the polycomb repressive complex 1 and 2 (PRC1/2) which maintain and deposit the repressive H3K27me3 (histone3, lysine 27 trimethylation) mark, respectively. This subsequently leads to repression at the locus (partially through the RdDM pathway), a return of the repressive chromatin loop, and loss of PID expression. Interestingly, APOLO is also able to navigate to target sites *in trans* via sequence complementarity, without a requirement for topological association (Ariel et al. 2020). To our knowledge, this appears to be one of the first examples of a eukaryotic lncRNA impacting chromatin topology in *cis* and in *trans* merely through sequence complementarity. In sum, dual transcription of the APOLO locus facilitates tight transcriptional control of auxin-responsive genes in a highly dynamic fashion (Ariel et al. 2020).

A new and compelling avenue that lncRNAs have been shown to act in is the regulation of biosynthetic gene clusters. In the last two decades, numerous examples of genomic co-localized biosynthetic gene clusters have been identified and are often co-regulated (Bharadwaj et al. 2021; Polturak and Osbourn 2021). lncRNAs derived from biosynthetic gene clusters present a unique mechanism for the *cis*-regulation of gene clusters and associated genetic elements. In a mechanism similar to APOLO, MARS (MARneral Silencing), a lncRNA transcribed from within the Arabidopsis marneral cluster, also functions as a chromatin topology modulator (Roulé et al. 2022). In response to the hormone abscisic acid (ABA), elevated MARS expression likely decoys PRC1 away from the marneral cluster and facilitates chromatin loop formation; this seems to enhance the expression of the MARNERAL SYNTHASE1 gene. This chromatin loop brings the marneral cluster closer to distal ABA-responsive elements, effectively modulating the expression of the marneral gene cluster during seed germination and osmotic stress (Roulé et al. 2022).

In plants, as in many organisms, developmental transitions need to be connected to environmental cues, and thus are often under tight and tunable regulation. In Arabidopsis and throughout the Brassicaceae, the FLOWERING LOCUS C (FLC) gene is a critical regulator of vernalization, the process that initiates flowering after prolonged cold exposure (Whittaker and Dean 2017). FLC is a MADS-box transcription factor that represses a suite of floral activation genes. Therefore, FLC expression is repressed upon cold exposure to allow for the appropriate timing of flowering to facilitate successful reproduction. Nested within the FLC locus is a





**Figure 2.** Functional lncRNA archetypes. A simplified cellular diagram displaying various nuclear or cytoplasmic mechanisms reported in plants. lncRNAs are shown as thick orange lines in each diagram. Nuclear activities are shown in green boxes with lines pointing towards the nucleus, whereas cytoplasmic activities are shown in blue boxes with lines pointing toward ‘Cytoplasm’.

group of lncRNAs that play critical roles in the vernalization response through their repression of *FLC*. These lncRNAs have been reviewed extensively elsewhere (Wang and Chekanova 2017; Whittaker and Dean 2017; Lucero et al. 2021), so we will only touch on them here in the context of how they relate to archetypal lncRNA functions. *COLDWRAP* is a lncRNA transcribed from the repressed promoter of *FLC* and contributes to the establishment of H3K27me3 at the *FLC* promoter through interactions with PRC2. In addition, *COLDWRAP* is necessary to form a repressive intragenic chromatin loop at the 5' end of the *FLC* gene; this loop appears necessary for spreading of H3K27me3 (Kim and Sung 2017). This PRC2-*COLDWRAP* interaction plays a part in preventing precocious flowering. As we discuss below, the regulation of *FLC* is complex and provides examples of multiple lncRNA archetypes.

The mechanisms of *APOLLO*, *MARS*, and *COLDWRAP* collectively represent a common archetype of regulating the formation and regulation of chromatin loops, paired with interactions with repressive histone complexes (Fig. 2). As more plant lncRNAs are characterized over the next decade, we anticipate a large number of lncRNAs to fall into this functional archetype.

### Modulating expression in a chromatin loop-independent manner

There are an accumulating number of plant lncRNAs that can regulate gene expression by recruiting histone modifiers independent of any reported chromatin looping. Along with *COLDWRAP* as mentioned above, the lncRNAs *COOLAIR* and *Antisense Long* (*ASL*) are both necessary for the proper transcriptional repression of *FLC*. *COOLAIR* is a capped, polyadenylated, and alternatively spliced lncRNA that is transcribed from the antisense strand of *FLC*, initiating near the terminator of *FLC*. *COOLAIR* is a component of the vernalization-controlled flowering pathway, in that it is upregulated during cold conditions and contributes to the initial repression of *FLC* by promoting a repressed chromatin state through removal of active epigenetic marks, namely H3K36me3 and H3K4me1 (Liu et al. 2010; Csorba et al. 2014; Tian et al. 2019; Fang et al. 2020). This removal of epigenetic marks seems to occur through the H3K4 demethylase, *FLOWERING LOCUS D* (Fang et al. 2020). *ASL* is transcribed from the same promoter as *COOLAIR*, is also alternatively spliced, but is not polyadenylated (Shin and Chekanova 2014). While *COOLAIR* and *COLDWRAP* are expressed in response to cold, *ASL* biogenesis is dependent upon

**Table 2.** Functionally and mechanistically annotated plant lncRNAs

lncRNA name	Subcellular localization	Species of initial identification	Initial publication date	Functional archetype (Fig. 2)
<i>AG-incRNA4</i>	N	Arabidopsis	Wu et al. (2018)	Recruitment of epigenetic modifiers
<i>APOLO</i>	N	Arabidopsis	Ariel et al. (2014)	Formation of chromatin- and r-loops + recruitment of epigenetic modifiers
<i>ASCO</i>	N	Arabidopsis	Bardou et al. (2014)	Alternative splicing
<i>COLD AIR</i>	N	Arabidopsis	Heo and Sung (2011)	Recruitment of epigenetic modifiers
<i>COLDWRAP</i>	N	Arabidopsis	Kim and Sung (2017)	Formation of chromatin loop + recruitment of epigenetic modifiers
<i>COOLAIR</i>	N	Arabidopsis	Liu et al. (2010)	Recruitment of epigenetic modifiers
<i>DRIR</i>	N	Arabidopsis	Qin et al. (2017)	ND
<i>Ef-cd</i>	N	Rice	Fang et al. (2019)	Recruitment of epigenetic modifiers
<i>ELENA1</i>	N	Arabidopsis	Seo et al. (2017)	Promoter binding and ribonucleoprotein scaffolding
<i>ENOD40</i>	N, C	Medicago	Crespi et al. (1994)	Alternative splicing + protein sequestration
<i>FLORE</i>	N	Arabidopsis	Henriques et al. (2017)	ND
<i>FLAIL</i>	N	Arabidopsis	Jin et al. (2021b)	ND
<i>HID1</i>	N	Arabidopsis	Wang et al. (2014b)	Promoter binding and ribonucleoprotein scaffolding
<i>IPS1</i>	C	Tomato and Medicago	Burleigh and Harrison (1997) and Liu et al. (1997)	miRNA interaction
<i>LAIR</i>	N	Rice	Wang et al. (2018)	Recruitment of epigenetic modifiers
<i>IncCOBRA1</i>	N	Arabidopsis	Kramer et al. (2022)	Ribonucleoprotein scaffolding
<i>lncRNA39896</i>	C	Tomato	Hong et al. (2022)	miRNA interaction
<i>MARS</i>	N	Arabidopsis	Roulé et al. (2022)	Formation of chromatin loop + recruitment of epigenetic modifiers
<i>MAS</i>	N	Arabidopsis	Zhao et al. (2018)	Recruitment of epigenetic modifiers
<i>MISSEN</i>	C	Rice	Zhou et al. (2021)	Protein sequestration
<i>PHO1;2-NAT</i>	C	Rice	Jabnoun et al. (2013)	Translational regulation
<i>SEP3 circRNA</i>	N	Arabidopsis	Conn et al. (2017)	Formation of R-loop + alternative splicing
<i>SVALKA/asCBF1</i>	N	Arabidopsis	Kindgren et al. (2018)	Transcriptional interference
<i>TR</i>	N	Onion and other land plants	Fajkus et al. (2019) and Song et al. (2019)	Chromosome maintenance
<i>TWISTED LEAF</i>	N	Rice	Liu et al. (2018)	Recruitment of epigenetic modifiers

components of the exosome and thereby likely regulates *FLC* in a temperature-independent (autonomous) manner. ASL RNA immunoprecipitation experiments and epigenetic analyses revealed that ASL physically associates with both exosomal proteins and H3K27me3 regions within the *FLC* locus (Shin and Chekanova 2014). Thus, in the characterization of ASL, Shin and Chekanova established a mechanism for how *FLC* could be regulated by both the vernalization and autonomous flowering pathways in Arabidopsis.

Another repressive regulatory lncRNA arising from the *FLC* locus is *COLD AIR*, transcribed from the sense strand of the first intron of *FLC*. Working in tandem with *COOLAIR* to epigenetically silence *FLC*, *COLD AIR* associates with the *FLC* locus, recruiting the PRC2 complex to deposit H3K27me3 (Kim and Sung 2017). H3K27me3 further reinforces transcriptional repression of the *FLC* locus during vernalization. In addition to the *FLC*-derived lncRNAs, *MAS* (antisense to the *MADS AFFECTING FLOWERING4-MAF4*) is also essential to maintaining proper flowering control in Arabidopsis (Zhao et al. 2018). *MAF4* encodes for another MADS-box transcription factor that is paralogous to *FLC* and also acts as a vernalization-regulated floral repressor (Gu et al. 2013). During prolonged cold temperatures, *MAS* and *MAF4* expression are consistently low until a burst of expression at 20 d

post cold exposure when expression peaks for both transcripts. This burst of *MAF4* expression is hypothesized to prevent premature flowering during the early vernalization response (Kim and Sung 2013). *MAS* was shown to directly influence the expression of *MAF4* through recruitment of the COMPASS-like complex to the sense strand of *MAF4* which deposits H3K4me4 to promote the transcriptional burst of *MAF4* (Zhao et al. 2018). The transcription of *MAF4* enhances the suppression of early flowering (Zhao et al. 2018). In addition to governing floral transitions and morphogenesis, lncRNAs have the capacity to epigenetically drive tissue-specific expression patterns. In Arabidopsis, an intronic lncRNA of *AG*, *AG-incRNA4* recruits PRC2 components to facilitate *AG* repression. *AG*, encoding for yet another MADS-box transcription factor, is a key regulator of stamen and carpel identity in Arabidopsis flowers that must be repressed in vegetative tissues (Pelayo et al. 2021). *AG-incRNA4* interacts with PRC2 to promote the deposition of H3K27me3 at *AG* in vegetative tissues, in a mechanism remarkably similar to *COLD AIR*-mediated repression of *FLC* (Wu et al. 2018). Both *AG-incRNA4* and *COLD AIR* directly bind the PRC2 component CURLY LEAF to recruit the complex to DNA (Wu et al. 2018; Tian et al. 2019). Thus, cis-regulation of MADS-box transcription factors through overlapping lncRNAs might be a widespread mechanism to

ensure specificity of expression in this family of transcription factors.

In rice, *LAIR* (*LRK antisense intergenic RNA*), a nuclear-localized lncRNA found within the *leucine-rich kinase* (*LRK*) gene cluster functions as a positive regulator of the *LRK* cluster in a PRC2-dependent manner. *LAIR* directly interacts with the histone modifiers OsMOF and OsWDR5, components of an H4K16 acetyltransferase and associated protein complex (Taipale et al. 2005; Yang et al. 2014; Wang et al. 2018). This interaction positively regulates the expression of *LRK* genes through the deposition of H4K16ac and H3K4me3, and overexpression of *LAIR* ultimately leads to an increase in grain yield (Wang et al. 2018). *Ef-cd* (*early flowering-completely dominant*), an antisense lncRNA transcribed from the *OsSOC1* locus, represents another example by which yield is manipulated by lncRNAs. *OsSOC1* is the putative ortholog of the Arabidopsis *SUPPRESSOR OF OVEREXPRESSION OF CO1*, a floral integrator that is regulated by *FLC* and other floral regulators (Lee and Lee 2010). Although it is mechanistically unclear how *Ef-cd* functions, it is suggested that *Ef-cd* interacts with the histone modifier SDG724 to deposit H3K36me3 promoting transcription at *OsSOC1* (Fang et al. 2019). *Ef-cd* was identified as a major quantitative trait locus (QTL) for early flowering, where it was determined that disrupting *Ef-cd*, but not *OsSOC1*, led to delayed maturity. Thus, genetic and molecular data suggest that both *LAIR* and *Ef-cd* act as positive epigenetic regulators and reflect the potential for lncRNAs to impact agronomic traits.

### Regulating expression through promoter binding and ribonucleoprotein scaffolding

lncRNAs can also directly regulate transcription through targeted association with regulatory elements. One prominent example of this can be found in *HIDDEN TREASURE 1* (*HID1*). In response to constant red light, *HID1*, along with unknown protein partners, targets the promoter of *PHYTOCHROME INTERACTING FACTOR3* (*PIF3*), resulting in transcriptional inhibition of *PIF3* (Wang et al. 2014a). *PIF3* is a transcription factor that inhibits developmental responses to red light and experiences rapid degradation by the light-activated phytochrome A and B photoreceptors (Al-Sady et al. 2006; Ni et al. 2013). In one of the most striking demonstrations of plant lncRNA conservation to date, the Arabidopsis *hid1* mutant phenotype was rescued using the *HID1* homolog from Rice, an evolutionary distance of ~160 million years (Doyle 2012; Magallón et al. 2013). Indeed, sequence homologs of *HID1* have been identified in the moss, *Physcomitrium patens*. Interestingly, this sequence and functional conservation appear to be largely driven by two highly structured snoRNA-like domains, further blending the boundaries between snoRNAs and lncRNAs (Wang et al. 2014a).

Another example of lncRNAs targeting regulatory elements comes from the lncRNA *ELENA1* (*ELF18-INDUCED LONG-NONCODING RNA1*), which positively regulates plant pathogen defense. Specifically, *ELENA1* expression is induced

by the bacterial pathogen-associated molecular patterns, elf18 and flg22 (Seo et al. 2017). *ELENA1* then directly interacts with and enriches the mediator subunit 19A (*MED19A*) at the distal promoter region of the *PATHOGENESIS-RELATED GENE1* (*PR1*) gene. *PR1* expression is induced and subsequently upregulates a suite of genes involved in the biotic stress response (Seo et al. 2017). In a follow-up study, Seo et al. (2019) further dissect this mechanism by showing that *ELENA1* removes FIBRILLARIN 2 (*FIB2*), a direct interactor of *MED19A* and transcriptional repressor, from the *PR1* promoter. Thus, pretranscriptional regulation of gene expression by lncRNAs can occur through a number of mechanisms, enhancing or repressing depending on context.

## Co- and posttranscriptional regulation of gene expression, mRNA abundance, and translation

### Influencing expression through transcriptional interference

lncRNAs can also regulate mRNA abundance and turnover after transcription has initiated, either co- or posttranscriptionally. Co-transcriptional regulation has been observed through the physical interference of Pol-II complexes when transcribing antisense or adjacent gene pairs (i.e. polymerase collision; Hobson et al. 2012). *SVALKA* and antisense-*CBF1* transcripts (*asCBF1*), lncRNAs located in the *C-repeat/dehydration-responsive element binding factors* (*CBFs*) gene cluster, function in this fashion (Kindgren et al. 2018). During the early cold stress response in Arabidopsis, rapid upregulation of the *CBF* gene cluster occurs (Medina et al. 1999, 2011). The *CBFs* are transcription factors that induce expression of the *COLD REGULATED* (*COR*) gene family (Fowler and Thomashow 2002). Activation of the *COR* genes results in swift biochemical and physiological changes allowing for freezing tolerance (Fowler and Thomashow 2002; Zhao et al. 2016). After several hours (>4) of cold exposure, transcription of an adjacent and antisense lncRNA to *CBF1*, *SVALKA*, begins. Transcription of *SVALKA* continues into the 3' portion of the *CBF1* gene, generating unstable *asCBF1* through a phenomenon known as read-through transcription. This transcriptional read-through leads to stalling of the Pol-II complex transcribing the sense *CBF1*, resulting in attenuated expression of *CBF1* and tight control of the acclimation response. While we are only aware of one plant lncRNA currently described with this mechanism, the compact intergenic space of the Arabidopsis genome and the large number of lncRNAs residing in this space (Liu et al. 2012; Palos et al. 2022) suggest this mechanism may be more widespread.

### Influencing mRNA fate through alternative splicing and isoform selection

Alternative splicing has long been recognized as a critical mechanism impacting mRNA fate and function that can be dictated by lncRNAs. While much of our understanding of



this lncRNA-guided mechanism comes from humans, there are now prominent examples of this in plants as well (Hutchinson et al. 2007; Clemson et al. 2009). The Arabidopsis lncRNA, ASCO, localizes to nuclear speckles (nuclear domains enriched in pre-mRNAs and splicing factors) and mediates alternative splicing of numerous target precursor mRNAs (Bardou et al. 2014). ASCO directly interacts with the spliceosomal subunits PRP8a (PRE-MRNA PROCESSING 8a) and SmD1b (Sm ring D1b) and mediates spliceosomal interactions with target mRNAs (Rigo et al. 2020). Loss or overexpression of ASCO leads to significant alternative splicing events across hundreds of transcripts. Initial studies showed that ASCO can mediate alternative splicing in roots, and assist in shifting isoform abundances to regulate cell fate during auxin-induced organogenesis (Bardou et al. 2014). However, more recent studies have shown that ASCO can also mediate the splicing of genes involved in plant biotic responses (Rigo et al. 2020). The role of ASCO in alternative splicing indirectly affects the function of hundreds of target transcripts, and thus the full influence, and other lncRNAs like it, remain unknown.

Alternative isoforms can be enriched through lncRNA-mediated co-transcriptional mechanisms. For example, an exon-derived long circular RNA (long-circRNA) transcribed from *SEPALLATA3* (*Sep3*) can impact the abundance of *SEP3* isoforms. *SEP3* encodes for a MADS-box transcription factor that is important for flowering time and floral organ identity (Liu et al. 2009). By forming a DNA–RNA hybrid, or R-loop, with the cognate locus, the *SEP3* long-circRNA encourages transcriptional pausing and effectively forces the locus to favor the production of alternative isoforms (Conn et al. 2017). Furthermore, the overexpression of the *SEP3* long-circRNA results in increased petal numbers and reduced stamen numbers through the promotion of alternative *SEP3* mRNA isoforms, suggesting that the *SEP3* long-circRNA is associated with proper floral organogenesis. Thus, the *SEP3* long-circRNA serves as the first example of a plant lncRNA mediating homeotic phenotypes via both R-looping and alternative splicing. In addition, *SEP3* represents the first mechanistic characterization of a circRNA in plants.

### **lncRNAs that influence mRNA stability by sequestering miRNAs**

At the posttranscriptional level, cytoplasmic-localized lncRNAs can influence gene expression by interacting with miRNA-mediated pathways and directly interacting with cytoplasmic proteins. *IPS1* (induced by phosphate starvation 1) and other noncoding transcripts within the *TPS11/MT4* family described above (Burleigh and Harrison 1997; Liu et al. 1997; Chiou 2007) present another functional archetype of plant lncRNAs in modulating gene expression. *IPS1* was the first lncRNA, across all biology, to exhibit the capacity to sequester miRNA (Franco-Zorrilla et al. 2007). Under ambient conditions, *miRNA-399* acts to repress translation of its target mRNA—*PHO2*, a negative regulator of inorganic phosphate

(Pi) accumulation (Lin et al. 2008). However, under phosphate starvation, *IPS1* inhibits *PHO2* degradation by sequestering *miRNA-399* away from *PHO2* mRNAs. Interestingly, *IPS1* evades miRNA-mediated degradation through incomplete sequence complementarity, creating a robust regulatory mechanism (Franco-Zorrilla et al. 2007). In this context, *IPS1* target mimicry allows for the fine-tuning of plant responses to Pi starvation by modulating the regulatory capacity of *PHO2*.

The functional characterization of *IPS1* and discovery of miRNA target mimicry often go understated. This discovery fueled new avenues of research and presented an entirely novel layer of gene regulation. One month after the *IPS1* result was published, a group working on mammalian miRNAs published a method to suppress miRNA action through transient expression of a RNA with miRNA binding sites (Ebert et al. 2007). This method was adapted in Arabidopsis to generate a collection of miRNA knockdown lines and examine their phenotypes (Todesco et al. 2010). During the preparation of this review, a group reported the mechanism of *lncRNA39896* which is involved in the tomato response to *Phytophthora infestans* infection (Hong et al. 2022). This lncRNA was previously predicted to be a target mimic for *miRNA-166b* based on sequence complementarity, expression characteristics, and degradome sequencing analysis (Cui et al. 2020). The *miRNA-166b* targets two mRNAs encoding homeodomain leucine zipper transcription factors (HD-ZIP III), *SIHDZ34* and *SIHDZ45*, which are responsible for attenuating jasmonic acid and ethylene responses during biotic infection. The authors showed that *lncRNA39896* effectively decoys *miRNA-166b*, allowing translation of the target mRNAs and regulation of the hormone responses. Mutating the miRNA binding site of *lncRNA39896* led to increased miRNA targeting of the *SIHDZ34* and *SIHDZ45* mRNAs, a higher level of jasmonic acid and ethylene abundance, and increased resistance to *Phytophthora* infection. Thus, the discovery of endogenous target mimicry has contributed to our understanding of a gene regulatory mechanism with clear biotechnological applications.

### **Impacting the last step of the central dogma: lncRNAs regulating translation**

There are very few known instances of lncRNAs directly acting as regulators of mRNA translation. In rice, a cis-NAT of *PHOSPHATE1; 2* (*PHO1; 2*) is strongly upregulated during phosphate starvation and leads to the increase of *PHO1; 2* protein levels (Jabnour et al. 2013). Importantly, *PHO1; 2* mRNA levels remain stable across phosphate levels, as well as its isoform abundance and nuclear export patterns. The *PHO1; 2* cis-NAT achieves this increase in *PHO1; 2* protein abundance through increased occupancy of both the cis-NAT and *PHO1; 2* mRNA occupancy at polysomes. Overexpression of the *PHO1; 2* NAT in trans leads to increased *PHO1; 2* protein levels even in phosphate-replete conditions. Findings such as this have led to the development



of technologies to activate translation through the expression of natural or synthetic long antisense RNAs (Zucchelli et al. 2015). In a separate global analysis of phosphate starvation stress, Bazin and colleagues performed ribosome footprinting to analyze noncoding RNA occupancy in roots shifted from phosphate replete to limited conditions (Bazin et al. 2017). They identify over 1,000 annotated lncRNAs with ribosome footprint signatures and nearly half of these lncRNAs are cis-NATs. These data suggest a potentially widespread and largely unexplored mechanism of translational regulation akin to the *PHO1*; 2 cis-NAT discussed above. As the number of transcriptome datasets increase (e.g. ribo-seq, polysome profiling, etc.) in plants, they will allow researchers to distinguish between novel small ORF containing transcripts and those that regulate translation in a manner similar to the cis-NAT of *PHO1*; 2, and perhaps identify novel cytoplasmic lncRNA functions outside of translational regulation away from the ribosome.

## Plant lncRNAs outside of regulating gene expression

While a majority of the mechanistically described plant lncRNAs govern every aspect of the progression from DNA to protein, there are several lncRNAs that function outside of this role. As large, multidomain molecules with complex structures, lncRNAs are ideal platforms on which other molecules (proteins or RNA) can dock. This innate ability to interact with other molecules has been alluded to above in the context of gene expression, but also occurs in other contexts in both the nucleus and the cytoplasm.

Perhaps the most famous of these lncRNAs is the *TR*, the RNA component of a ribonucleoprotein complex called telomerase which is essential for maintaining chromosome ends. Despite its characterization in ciliates, yeast, and vertebrates in the 20th century (Greider and Blackburn 1989; Singer and Gottschling 1994; Feng et al. 1995), *TR* was only recently characterized in plants (Fajkus et al. 2019; Song et al. 2019; Dew-Budd et al. 2020). In each of these lineages, *TR* serves as a scaffold for the binding of the reverse transcriptase, TERT, as well as a number of accessory proteins critical for complex maturation and function. This scaffolding function is dependent upon conserved structures within *TR*, however, *TRs* are highly divergent in both their sequences and biogenesis pathways, suggesting the acquisition of novel *TRs* into the telomerase complex. Excitingly, plant genomes have a preponderance of *TR* paralogs and additional TERT interacting RNAs (Nelson and Shippen 2015), suggesting that plants may have more lessons to teach in terms of lncRNA-mediated regulation and the relaxed parameters under which even functionally conserved lncRNAs evolve.

Another example of a scaffolding RNA with a putative role in ribosome assembly and biogenesis is *lncCOBRA1* (CONSERVED IN BRASSICA RAPA 1; Kramer et al. 2022). As the name suggests, this lncRNA was identified through an

examination of Arabidopsis nuclear lncRNAs which were protein bound and conserved across Brassicaceae (Gosai et al. 2015). Mutant *lncCOBRA1* plants show delayed germination and generally grow slower than wild-type. *lncCOBRA1* seems to act through scaffolding of various protein partners, particularly RACK1A, which is important for ribosome assembly and biogenesis (Guo et al. 2011). Finally, it is notable that *lncCOBRA1*, similar to *HID1*, is a polycistronic transcript containing two highly conserved snoRNAs. Although these snoRNA domains do not appear to be further processed, they are highly structured and display high sequence conservation rates across Brassicaceae relative to the rest of the *lncCOBRA1* locus. In addition, these two domains were reported as the targets for protein binding (Gosai et al. 2015; Qu et al. 2015). Thus, *lncCOBRA1* and *HID1* represent scaffolding lncRNAs that impact Arabidopsis development and were identified through comparative and molecular signatures.

Another important RNA–protein interaction can be seen in the first identified plant lncRNA, *ENOD40*. In *M. truncatula*, the *ENOD40* RNA interacts with a pre-mRNA splicing factor RBP1 (RNA BINDING PROTEIN1), a nuclear speckle RNA-binding protein (Campalans et al. 2004). In non-nodulating plant cells, RBP1 is nuclear localized. However, during root nodulation, *ENOD40* and RBP1 translocate to the cytoplasm and this translocation is dependent on *ENOD40* expression. Thus, *ENOD40* appears to act in a similar manner to *ASCO* through the regulation of nuclear speckle splicing factors. However, in this case, *ENOD40* is acting as both an environmental sensor and a guide, facilitating the re-localization of its target protein.

In addition to aiding in protein localization, lncRNAs can also compete for protein interaction partners, with profound impacts on plant development. One of the early lncRNA identification efforts in rice (Zhang et al. 2014) led to the identification of *MISSEN*, a cytoplasmically localized lncRNA that regulates endosperm development (Zhou et al. 2021). *MISSEN* was identified based on its high expression levels in young flowers and pistils. Disruption of the *MISSEN* locus by T-DNA insertional mutagenesis led to decreased seed set, misshapen seeds, and reduced endosperm size. Mechanistically, *MISSEN* acts by competitively inhibiting a helicase family protein from interacting with tubulin, subsequently impacting endosperm development (Zhou et al. 2021). Finally, *MISSEN* was shown to be specifically expressed from the maternal allele during endosperm development, a feature common during endosperm and seed development (Huh et al. 2007). Given the contexts of expression, *MISSEN* provides a model for how regulatory properties of lncRNAs can affect fundamental stages of plant development.

## Novel approaches for inferring function

De novo structural identification and classification of lncRNAs in a transcriptomic and genomic context provides

the foundation to explore the functional roles of plant lncRNAs. However, prioritizing lncRNAs for functional analysis is not as straightforward as for protein-coding genes. Functional and comparative resources (e.g. RNA functional domain databases) are lacking, restricting the development of functional clues when a lncRNA is identified in a researcher's RNA-seq data. While we expect functional annotation to become easier as more lncRNAs are described, there are currently only a few bioinformatic methods for assigning putative functions to lncRNAs, including integrating lncRNAs into co-expression and gene regulatory networks as well as motif enrichment analyses that we will describe below.

### Guilt-by-association

To better understand the function of lncRNAs, clustering lncRNAs with highly correlated expression to protein-coding genes using weighted gene co-expression networks, often referred to as “Guilt-by-Association” (GBA), is gaining traction in both plants and animals (Langfelder and Horvath 2008; Liu et al. 2022; Nolte et al. 2022; Waseem et al. 2022; Zhang et al. 2022a). Given similarities in expression profiles across a complex dataset, genes will group together into modules. Gene ontology term and other similar pathway analysis databases (e.g., Kyoto Encyclopedia of Genes and Geomes or KEGG) can identify the protein-associated biological processes enriched within that module. The GBA comes from the assumption that lncRNAs with statistically similar expression patterns will likely be involved in similar processes as proteins within a module. Overlaying additional information, such as transcription factor binding data (e.g. coming from DNA affinity purification sequencing or DAP-seq; O'Malley et al. 2016; Bartlett et al. 2017), can further constrain these modules and assist in inferring the directionality of interactions. Ultimately, these types of analyses help researchers build testable functional hypotheses.

When paired with additional types of data, GBA approaches can be quite informative and help pinpoint functional lncRNAs. For instance, a functional annotation of chromatin-enriched ncRNAs in rice incorporated expression, chromatin-interaction data, and phenotypic information to identify lncRNAs associated with yield (Zhang et al. 2022b). In addition, Palos et al. (2022) identified germination-associated lncRNAs using a GBA approach through the incorporation of expression data from the Klepikova tissue atlas (Klepikova et al. 2016). The authors identified a module of co-expressed genes with peak expression in embryogenesis and germination. A further phenotypic screen of mutants in lncRNAs found within this module revealed reduced germination rates. These approaches benefit from experimental complexity, be it temporal, tissue, or treatment based, as well as pairing with complementary epigenetic, structural, or interaction datasets. Under these circumstances, GBA approaches can be incredibly useful in helping researchers identify when and where to look for function.

### Identification of functional domains

lncRNA functional domains, as with protein domains, are often the regions of the RNA through which intermolecular interactions (RNA–protein, RNA–RNA, and RNA–metabolite) occur. These domains, or motifs, are often short (6 to 12 nt) and may occur multiple times within the same RNA to increase binding affinity or number of interactions (Ulitsky et al. 2011; Bitetti et al. 2018). As a result, those motifs, and their enrichment, may serve as signatures of function and can even help uncover patterns of conservation. Sequence homology-based approaches of Brassicaceae lncRNAs revealed higher levels of conservation within regions found to be structured and protein bound (Nelson et al. 2016; Palos et al. 2022). More elegant comparative analyses of vertebrate lncRNAs revealed a higher degree of conservation within functional motifs as compared to nonmotif fragments within the same RNA, indicating some level of selection was occurring (Hezroni et al. 2015; Ross and Ulitsky 2022). These approaches can further help identify orthologous lncRNAs in the absence of overall sequence similarity. One example of this comes from the lncLOOM framework (lncrna Linear Order cOnserved Motifs) which searches for retention in the order of multiple short motifs within the same lncRNA to determine homology and ultimately functional conservation (Ross et al. 2021). lncLOOM was used in vertebrates to identify orthologs of the vertebrate-conserved lncRNA, CYRANO (Ulitsky et al. 2011). The *Cyrano* gene family was originally identified due to transcription arising from a syntenic locus. However, despite the identification of homologs across vertebrates, *Cyrano* is highly sequence and length divergent and it was difficult to infer functional orthology. Upon deeper inspection of the *Cyrano* gene family with lncLOOM, seven functional motifs, always arranged in the same order, were uncovered within this lncRNA across vertebrates. Syntenic, but sequence-divergent lncRNAs have also been observed in plants (Palos et al. 2022), and lncLOOM, or similar approaches, may help uncover conserved functional domains and guide in their evolutionary analyses.

Innovations in natural language processing and machine learning methods have made it possible to use functional motifs from characterized lncRNAs to infer function of unknown lncRNAs. In this context, algorithms such as SEquence Evaluation from K-mer Representation (SEEKR) search for the enrichment of particular motifs, or k-mers, within a lncRNA (Kirk et al. 2018). K-mer signatures are developed for a particular query lncRNA and then pairwise Pearson-correlation is used to search for other lncRNAs with similar k-mer signatures. SEEKR was successfully used to identify two human lncRNAs (*KCNQ1OT1* and *AIRN*) that exhibited similar k-mer profiles to the *XIST* lncRNA, a well-known lncRNA involved in the epigenetic silencing of the x chromosome through interactions with the polycomb repressor complex (Zhao et al. 2008). Further analysis revealed that the observed correlation between k-mer signatures was driven

by PRC interaction domains found in each of the lncRNAs. Indeed, through a clustering analysis, SEEKR identified hundreds of lncRNAs with similar k-mer signatures as *XIST*, suggesting a large set of PRC-interacting epigenetic regulators may exist in mammals, an observation further bolstered by an abundance of genome-wide PRC-RNA interaction data. Thus, k-mer or motif-based lncRNA analyses may help to predict lncRNA function. One caveat to the SEEKR and LncLOOM approaches outlined above is that they were both developed predominantly in mammalian systems and thus may require retooling for plant lncRNA functional motifs (SEEKR) and plant genome evolution (LncLOOM). However, in silico functional predictions and inferences of orthology are likely to only get stronger as more interaction partners and functional domains are identified for plant lncRNAs.

### High-throughput genetic screens for perturbing lncRNA expression

In addition to bioinformatic approaches, an improved molecular toolkit for plant systems would greatly facilitate the identification of functional lncRNAs. CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats) gene editing approaches have already demonstrated their utility in plants for assessing gene function in a highly targeted fashion (Zhou et al. 2021; Kramer et al. 2022). However, lncRNAs lack many of the fundamental features typically used as targets by CRISPR approaches (i.e. disrupting ORFs through frame-shift mutations). CRISPR-mediated gene or promoter deletions are options, and are certainly ideal over standard insertional mutagenesis approaches long used in Arabidopsis (e.g. T-DNA insertional mutagenesis). Indeed, CRISPR-mediated lncRNA deletion was used to ascertain the function of at least two lncRNAs, including *lncCOBRA1* and *FLAIL* (Jin et al. 2021b; Kramer et al. 2022). However, targeted deletion of both lncRNA loci or their promoters is largely guesswork, particularly in the absence of well-annotated transcription start site information, and is not particularly high-throughput as a functional screen. A promising CRISPR-based alternative for screening for functional lncRNAs can be found in CRISPR activation/inactivation systems (CRISPRi/a) that were recently developed for mammalian systems (Jensen et al. 2021) and are starting to gain traction in plants (Gilbert et al. 2013, 2014; Liu et al. 2017). The CRISPRi/a systems work by fusing either repressive or activating effector domains to Cas9 and then using guide RNAs designed to the approximate transcriptional start site of a lncRNA of interest to effectively modulate expression. This type of system can be multiplexed with the addition of 100s–1000s of guide RNAs to screen through a suite of target lncRNAs (Liu et al. 2017; Covarrubias et al. 2020). This technology is in the early stages in plants, but has successfully been applied in a few protein-centric manners (Lowder et al. 2017; Ochoa-Fernandez et al. 2020; Kar et al. 2022) and will undoubtedly have an outsized impact on lncRNA biology.

## Future functional forays: understudied aspects of plant lncRNAs

### Localization

A lncRNA's subcellular localization can inform in which functional archetype the gene may be involved (Fig. 2). Nuclear lncRNAs may perform a variety of pre or cotranscriptional modes of gene regulation, genome stability, or mediating chromosomal interactions (Guh et al. 2020; Li et al. 2021; Xiao et al. 2022). In contrast, cytoplasmic lncRNAs might scaffold proteins, decoy miRNAs, regulate translation, or a variety of other less understood lncRNA actions (Noh et al. 2018). Multiple studies have investigated lncRNAs from nuclear and/or cytoplasmic compartments (Zhao et al. 2018; Do et al. 2019), and there are emerging computational approaches to predict lncRNA subcellular localization (Su et al. 2018; Ahmad et al. 2020). These datasets represent valuable resources to the plant lncRNA community as they present the foundation for follow-up molecular experiments; integration of such datasets to a central repository or database of plant lncRNAs (e.g. STRING for lncRNAs; Jensen et al. 2008) would bolster and accelerate all aspects of plant RNA biology.

One avenue of lncRNA subcellular localization that remains untouched in plants is their role in the formation of SGs. SGs are cytoplasmic protein, RNA, and metabolite aggregates that form in response to stress conditions, and are thought to serve as crucial reservoirs of ribosomal machinery, chaperones, and untranslated mRNAs (Kearly et al. 2022). SGs are a conserved phenomenon across all eukaryotes, and in mammals, lncRNAs play some role in their formation, potentially through disordered domains or through their scaffolding abilities (Khong et al. 2017; Campos-Melo et al. 2021; Maruri-López et al. 2021). This likely holds true in plants and would highlight another important cytoplasmic function for lncRNAs.

### Mobilization

In addition to subcellular localization, the role of cell–cell mobile lncRNAs is another exciting new avenue for plant lncRNA research. There is ample evidence that the plant vasculature serves as a long-distance communication system for transporting proteins, metabolites, hormones, as well as RNAs (Morris 2018; Thomas and Frank 2019). Indeed, one group has analyzed the lncRNA repertoires that show movement through the phloem during phosphate deficiency in cucumber (Zhang et al. 2019). They identified hundreds of mobile lncRNAs that are responsive to phosphate deficiency, including the cucumber homolog to *IPS1* and other putative miRNA mimic lncRNAs. In support of this finding, Thieme et al. (2015) identified an Arabidopsis paralog of *IPS1* (*AT4*) in a pool of mobile mRNAs. Furthermore, these mobile lncRNAs may be important for interspecies communication. A recent study demonstrated that a green peach aphid (*Myzus persicae*) lncRNA (*Ya1*) is part of a pool of molecules injected into the aphid's plant host (Arabidopsis) and that



this particular lncRNA promotes aphid feeding (Chen et al. 2020). Interestingly, *Ya1* was observed to migrate from the aphid feeding site to other leaves, perhaps interacting with a target molecule in those leaves to reduce plant defenses. Thus, lncRNA mobility, within plants, and between organisms, is an important aspect to consider when predicting targets and molecular mechanisms.

### The role of alternative splicing in regulating lncRNA function

Alternative splicing is an important mechanism for delivering regulatory plasticity in response to changing cellular or organismal environments by allowing for multiple RNAs from the same locus (Syed et al. 2012; Reddy et al. 2013). The scale and regulation of alternative splicing in plant protein-coding pre-messenger RNAs is well documented; over 60% of multi-exonic genes undergo alternative splicing (AS) in plants (Chaudhary et al. 2019). While most plant lncRNAs appear to be mono-exonic (~90% in Arabidopsis; Nelson et al. 2016; Palos et al. 2022), a large number of the multiexonic Arabidopsis lncRNAs display splice site conservation across Brassicaceae (Corona-Gomez et al. 2020). In addition, AS has been well-described for certain lncRNAs, such as *COOLAIR* and *FLORE* (Csorba et al. 2014; Henriques et al. 2017). The biological role of AS in plant transcripts is unclear, as many mRNA isoforms are rapidly degraded by the nonsense-mediated decay machinery (Filichkin et al. 2015). Noncoding, alternative isoforms of protein-coding genes may represent a novel form of posttranscriptional gene regulation, as they have been found to influence mRNA levels arising from the same locus (Reddy 2007). While the mechanistic reason behind lncRNA AS is not as immediately clear as it is for mRNAs, AS may allow for alternative targeting (for cis or trans-regulatory lncRNAs), or for alternative protein–RNA interactions that then facilitate unique functions. Thus, AS is an exciting aspect of plant lncRNAs that requires further exploration.

### Regulating the regulator

While lncRNAs have broadly become established as regulators of gene expression, there is substantially less known regarding how lncRNAs themselves are regulated and how this feeds back into their function. Genome-wide assessment of DNA methylation and histone modification profiles in multiple plant species has demonstrated that lincRNA loci closely resemble protein-coding loci. In particular, lincRNAs in Arabidopsis and *Eutrema salsugineum* displayed enrichment of H3K4me3 (trimethylation of lysine 4 on histone 3) near the promoter and 5' start site along with H3K36me3 (trimethylation of lysine 36 on histone 3) across the gene body and flanking both ends of the transcript start and end sites (Zhang et al. 2009; Heo et al. 2013; Palos et al. 2022). This local connection of H3K4me3 and H3K36me3, also referred to as the K4-K36 domain, is associated with actively transcribed protein-coding genes and was used to

identify numerous deeply conserved, highly expressed, and multiexonic lincRNAs in humans and mouse (Guttman et al. 2009; Khalil et al. 2009). Less is known about the relationship between DNA methylation and lncRNA expression, but from the limited studies to date, the relationship between expression and DNA methylation is generally negative, particularly in the gene body (Zhou et al. 2021; Palos et al. 2022; Yu et al. 2022). Much more work is needed to better understand how the epigenome modulates lncRNA expression.

Once the chromatin is relaxed, transcriptional machinery needs to be recruited in order to initiate lncRNA transcription. Interestingly, distinct differences have been observed in this regard in Arabidopsis. Tokizawa et al. (2017) mapped transcription start sites and characterized the promoters and capping characteristics of coding and noncoding genes. They observed that antisense lncRNA and lincRNA promoters have lower ratios of TATA boxes and Y patches (pyrimidine patches) that are generally necessary for transcription. This relative depletion of TATA boxes has been noted in animals for antisense lncRNAs (Lin et al. 2015). These findings suggest a potential explanation for the low abundance of lncRNAs as a class: if lncRNAs contain nonoptimal promoters, Pol-II may be less efficiently recruited in the absence of additional factors. Another intriguing finding was that only ~42% of antisense lncRNAs and ~74% of lincRNAs were associated with CAGE-generated capped transcription start sites, potentially suggesting rapid turnover or alternative biogenesis pathways (Tokizawa et al. 2017).

RNA biogenesis pathways are another underexplored aspect of lncRNA regulation and turnover. Work from Zhi John Lu's group (as mentioned above in the sampling bias section) has shed light upon lncRNAs that do not contain polyA tails, a facet of lncRNA research that is underexplored. For instance, Yuan et al. (2018) showed that while hundreds of Rice lncRNAs are downregulated during abiotic stress, isoforms of these lncRNAs that do not contain polyA tails actually increase in abundance (Yuan et al. 2018). One possibility is that the abundance of these lncRNAs are modulated by the exosome in a manner similar to ASL in Arabidopsis (Shin and Chekanova 2014). The biological implications and reasoning for this phenomenon are not understood, yet it presents exciting opportunities for uncovering even more complex modes of lncRNA-involved transcriptional regulation, as these RNAs may be overlooked in standard sequencing approaches.

### Understanding the impact of the epitranscriptome on lncRNAs

Finally, every aspect of an RNA's lifecycle, including its shape, interacting partners, and ultimately function, can be influenced by chemical base modifications often referred to as the epitranscriptome (Bhatia et al. 2022). Aside from the 5' trimethylguanosine cap and recently identified alternatives (Wang et al. 2019b; Yu et al. 2021), there are a number of RNA base modifications known to impact mRNA structure



and function in plants (Shen et al. 2019). While a more exhaustive review of the epitranscriptome can be found elsewhere in this issue, we believe that these modifications are likely also widespread on lncRNAs and have just been overlooked due to the general low abundance and poor annotation of lncRNAs. However, targeted studies have demonstrated that the epitranscriptome is just as critical for lncRNAs as it is for mRNAs and sRNAs. For instance, N6-methyladenosine was recently shown to be important for COOLAIR's role in regulating the *FLC* locus by modulating R-loop stability (Xu et al. 2021). Given COOLAIR's ability to adopt multiple structural conformations that impact flowering (Yang et al. 2022), considering both RNA modifications and structure in tandem will be critical for dissecting mechanism. Global analyses of RNA modifications in Arabidopsis using the bioinformatic tool HAMR found distinct patterns between stable lncRNAs and those targeted for degradation, suggesting modification state may be indicative of the functional state (Ryvkin et al. 2013; Vandivier et al. 2015). Single-molecule techniques capable of directly identifying modifications on RNAs, such as Oxford Nanopore's direct RNA-sequencing approach (Kirov et al. 2020), are poised to dramatically alter our understanding of the epitranscriptome in plants. In sum, the impact of the epitranscriptome on lncRNA biology in plants is an emerging field with the potential to explain how lncRNAs function at substoichiometric levels, how they recruit or interact with binding partners, and ultimately how they impact plant biology.

## Conclusions with an eye towards the future of plant lncRNA biology

Tremendous achievements have been made over the last 25 yr in plant lncRNA biology. Characteristics and mechanisms of lncRNAs first discovered in nonmodel species, then elaborated on in Arabidopsis, are now being used to identify functionally important lncRNAs across the plant lineage. However, at present, the function of the vast majority of annotated plant lncRNAs remains unknown, and traditional molecular investigations into each one of these transcripts would be cost-prohibitive. We anticipate modern-omics technology, in combination with rapid CRISPR-based genetic screening and ML approaches, will close the gap between putative and actual functions for plant lncRNAs.

Discoveries in a few key plant species will undoubtedly translate to more and more distant relatives as we begin to understand the rules that govern lncRNA evolution. Moving away from a protein-centric perspective will be critical. Selective pressures at the nucleotide, or even structural level, are likely less important for some lncRNA functions than others. In addition, expression abundance—typically thought of as a sign of functionality in protein-coding genes—may be less important than expression at the right or optimal time or arising from the correct portion of a

genome. Developing comparative frameworks that incorporate these types of conservation will be critical in transferring functional information from one species to another.

New and exciting tools and techniques are being brought to bear on the identification, functional dissection, and comparative analyses of lncRNAs. As computational algorithms and approaches are specifically developed for lncRNAs, robust sets of functional lncRNAs will be needed for training. Critical thought will need to be placed into which lncRNAs are selected as positive controls to avoid training around non-lncRNA features. To aid in algorithm development, the current plant lncRNA databases need to be harmonized in terms of the criteria each used to identify their lncRNAs.

In all of this, the variation in functional archetypes is important to keep in mind. While many lncRNAs likely influence the genome and transcriptome in the nucleus, there is growing evidence that cytoplasmic functions are just as varied. The ability of lncRNAs, alongside other molecules, to translocate between cells or organisms further deepens the mystery of what role these enigmatic transcripts play in plant biology. The functional possibilities seem endless, and the chance for mistaken identities is high, but the long and storied history of lncRNAs in plants makes it clear that their study is worth the effort.

## Acknowledgments

The authors would like to thank other members of the Nelson lab as well as the Gregory lab at the University of Pennsylvania for helpful discussion.

## Author contributions

All authors contributed to the writing of the review. K.R.P., A.C.N.D., and C.E.R. contributed to the figures, and L.Y. and C.E.R. contributed to the tables.

## Funding

The authors would like to acknowledge NSF-IOS 2023310 (awarded to A.D.L.N.) and NSF-IOS 2120131 (awarded to A.D.L.N.). We would also like to acknowledge the NSF Research Traineeship #1922551 for supporting CER.

*Conflict of interest statement.* None declared.

## Data availability

No data were generated in the preparation of this review.

## References

- Achawanantakun R, Chen J, Sun Y, Zhang Y. lncRNA-ID: Long non-coding RNA IDentification using balanced random forests. *Bioinformatics*. 2015;31(24):3897–3905. <https://doi.org/10.1093/bioinformatics/btv480>

- Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*. 2016;**167**(7):1867–1882.e21. <https://doi.org/10.1016/j.cell.2016.11.048>
- Ahmad A, Lin H, Shatabda S. Locate-R: subcellular localization of long non-coding RNAs using nucleotide compositions. *Genomics*. 2020;**112**(3):2583–2589. <https://doi.org/10.1016/j.ygeno.2020.02.011>
- Al-Sady B, Ni W, Kircher S, Schäfer E, Quail PH. Photoactivated phytochrome induces rapid PIF3 phosphorylation prior to proteasome-mediated degradation. *Mol Cell*. 2006;**23**(3):439–446. <https://doi.org/10.1016/j.molcel.2006.06.011>
- Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res*. 2011;**39**(suppl 1):D146–D151. <https://doi.org/10.1093/nar/gkq1138>
- Amor BB, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, et al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res*. 2009;**19**(1):57–69. <https://doi.org/10.1101/gr.080275.108>
- Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000;**408**(6814):796–815. <https://doi.org/10.1038/35048692>
- Ariel F, Jegu T, Latrasse D, Romero-Barrios N, Christ A, Benhamed M, Crespi M. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell*. 2014;**55**(3):383–396. <https://doi.org/10.1016/j.molcel.2014.06.011>
- Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, Mariappan K, Latrasse D, Blein T, Liu C, et al. R-loop mediated trans action of the APOLO long noncoding RNA. *Mol Cell*. 2020;**77**(5):1055–1065.e4. <https://doi.org/10.1016/j.molcel.2019.12.015>
- Arriall RT, Togawa RC, Brigido MDM. Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus *Paracoccidioides brasiliensis*. *BMC Bioinformatics*. 2009;**10**(1):239. <https://doi.org/10.1186/1471-2105-10-239>
- Baek J, Lee B, Kwon S, Yoon S. LncRNA-net: long non-coding RNA identification using deep learning. *Bioinformatics*. 2018;**34**(22):3889–3897. <https://doi.org/10.1093/bioinformatics/bty418>
- Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, Balzergue S, Brown JWS, Crespi M. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Dev Cell*. 2014;**30**(2):166–176. <https://doi.org/10.1016/j.devcel.2014.06.017>
- Bari R, Datt Pant B, Stitt M, Scheible W-R. PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiol*. 2006;**141**(3):988–999. <https://doi.org/10.1104/pp.106.079707>
- Bartlett A, O'Malley RC, Huang S-SC, Galli M, Nery JR, Gallavotti A, Ecker JR. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc*. 2017;**12**(8):1659–1672. <https://doi.org/10.1038/nprot.2017.055>
- Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, Bailey-Serres J. Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proc Natl Acad Sci USA*. 2017;**114**(46):E10018–E10027. <https://doi.org/10.1073/pnas.1708433114>
- Bharadwaj R, Kumar SR, Sharma A, Sathishkumar R. Plant metabolic gene clusters: evolution, organization, and their applications in synthetic biology. *Front Plant Sci*. 2021;**12**:697318. <https://doi.org/10.3389/fpls.2021.697318>
- Bhatia G, Prall W, Sharma B, Gregory BD. Covalent RNA modifications and their budding crosstalk with plant epigenetic processes. *Curr Opin Plant Biol*. 2022;**69**:102287. <https://doi.org/10.1016/j.pbi.2022.102287>
- ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;**447**(7146):799–816. <https://doi.org/10.1038/nature05874>
- Bitetti A, Mallory AC, Golini E, Carrieri C, Carreño Gutiérrez H, Perlas E, Pérez-Rico YA, Tocchini-Valentini GP, Enright AJ, Norton WHJ, et al. MicroRNA degradation by a conserved target RNA regulates animal behavior. *Nat Struct Mol Biol*. 2018;**25**(3):244–251. <https://doi.org/10.1038/s41594-018-0032-x>
- Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One*. 2012;**7**(8):e43047. <https://doi.org/10.1371/journal.pone.0043047>
- Burleigh SH, Harrison MJ. A novel gene whose expression in *Medicago truncatula* roots is suppressed in response to colonization by vesicular-arbuscular mycorrhizal (VAM) fungi and to phosphate nutrition. *Plant Mol Biol*. 1997;**34**(2):199–208. <https://doi.org/10.1023/A:1005841119665>
- Burleigh SM, Harrison MJ. Characterization of the Mt4 gene from *Medicago truncatula*. *Gene*. 1998;**216**(1):47–53. [https://doi.org/10.1016/S0378-1119\(98\)00326-6](https://doi.org/10.1016/S0378-1119(98)00326-6)
- Burleigh SH, Harrison MJ. The down-regulation of Mt4-like genes by phosphate fertilization occurs systemically and involves phosphate translocation to the Shoots1. *Plant Physiol*. 1999;**119**(1):241–248. <https://doi.org/10.1104/pp.119.1.241>
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;**25**(18):1915–1927. <https://doi.org/10.1101/gad.17446611>
- Campalans A, Kondorosi A, Crespi M. Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell*. 2004;**16**(4):1047–1059. <https://doi.org/10.1105/tpc.019406>
- Campos-Melo D, Hawley ZCE, Droppelmann CA, Strong MJ. The integral role of RNA in stress granule formation and function. *Front Cell Dev Biol*. 2021;**9**:621779. <https://doi.org/10.3389/fcell.2021.621779>
- Castelli V, Aury J-M, Jaillon O, Wincker P, Clepet C, Menard M, Cruaud C, Quétiér F, Scarpelli C, Schächter V, et al. Whole genome sequence comparisons and “full-length” cDNA sequences: a combined approach to evaluate and improve Arabidopsis genome annotation. *Genome Res*. 2004;**14**(3):406–413. <https://doi.org/10.1101/gr.1515604>
- Cemel IA, Ha N, Schermann G, Yonekawa S, Brunner M. The coding and noncoding transcriptome of *Neurospora crassa*. *BMC Genomics*. 2017;**18**(1):978. <https://doi.org/10.1186/s12864-017-4360-8>
- Chaudhary S, Khokhar W, Jabre I, Reddy ASN, Byrne LJ, Wilson CM, Syed NH. Alternative splicing and protein diversity: plants versus animals. *Front Plant Sci*. 2019;**10**:708. <https://doi.org/10.3389/fpls.2019.00708>
- Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, Hooker T, Yazaki J, Li P, Skiba N, Peng Q, et al. Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell*. 2007;**131**(7):1340–1353. <https://doi.org/10.1016/j.cell.2007.10.056>
- Chen Y, Singh A, Kaithakottil GG, Mathers TC, Gravino M, Mugford ST, van Oosterhout C, Swarbrick D, Hogenhout SA. An aphid RNA transcript migrates systemically within plants and is a virulence factor. *Proc Natl Acad Sci USA*. 2020;**117**(23):12763–12771. <https://doi.org/10.1073/pnas.1918410117>
- Chiou T-J. The role of microRNAs in sensing nutrient stress. *Plant Cell Environ*. 2007;**30**(3):323–332. <https://doi.org/10.1111/j.1365-3040.2007.01643.x>
- Cho J. Transposon-derived non-coding RNAs and their function in plants. *Front Plant Sci*. 2018;**9**:600. <https://doi.org/10.3389/fpls.2018.00600>
- Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. An architectural role for a nuclear noncoding RNA: nEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell*. 2009;**33**(6):717–726. <https://doi.org/10.1016/j.molcel.2009.01.026>

- Conn VM, Hugouvieux V, Nayak A, Conos SA, Capovilla G, Cildir G, Jourdain A, Tergaonkar V, Schmid M, Zubieta C, et al. A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat Plants*. 2017;3(5):17053. <https://doi.org/10.1038/nplants.2017.53>
- Corona-Gomez JA, Garcia-Lopez IJ, Stadler PF, Fernandez-Valverde SL. Splicing conservation signals in plant long noncoding RNAs. *RNA*. 2020;26(7):784–793. <https://doi.org/10.1261/rna.074393.119>
- Covarrubias S, Vollmers AC, Capili A, Boettcher M, Shulkin A, Correa MR, Halasz H, Robinson EK, O'Briain L, Vollmers C, et al. High-throughput CRISPR screening identifies genes involved in macrophage viability and inflammatory pathways. *Cell Rep*. 2020;33(13):108541. <https://doi.org/10.1016/j.celrep.2020.108541>
- Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A. Enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J*. 1994;13(21):5099–5112. <https://doi.org/10.1002/j.1460-2075.1994.tb06839.x>
- Csorba T, Questa JI, Sun Q, Dean C. Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc Natl Acad Sci USA*. 2014;111(45):16160–16165. <https://doi.org/10.1073/pnas.1419030111>
- Cui J, Jiang N, Hou X, Wu S, Zhang Q, Meng J, Luan Y. Genome-wide identification of lncRNAs and analysis of ceRNA networks during tomato resistance to *Phytophthora infestans*. *Phytopathology*. 2020;110(2):456–464. <https://doi.org/10.1094/PHYTO-04-19-0137-R>
- Da Negri TC, Alves WAL, Bugatti PH, Saito PTM, Domingues DS, Paschoal AR. Pattern recognition analysis on long noncoding RNAs: a tool for prediction in plants. *Brief Bioinform*. 2019;20(2):682–689. <https://doi.org/10.1093/bib/bby034>
- Deshpande S, Shuttlesworth J, Yang J, Taramonli S, England M. PLIT: an alignment-free computational tool for identification of long non-coding RNAs in plant transcriptomic datasets. *Comput Biol Med*. 2019;105:169–181. <https://doi.org/10.1016/j.combiomed.2018.12.014>
- Dew-Budd K, Cheung J, Palos K, Forsythe ES, Beilstein MA. Evolutionary and biochemical analyses reveal conservation of the Brassicaceae telomerase ribonucleoprotein complex. *PLoS One*. 2020;15(4):e0222687. <https://doi.org/10.1371/journal.pone.0222687>
- Di C, Yuan J, Wu Y, Li J, Lin H, Hu L, Zhang T, Qi Y, Gerstein MB, Guo Y, et al. Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J*. 2014;80(5):848–861. <https://doi.org/10.1111/tpj.12679>
- Di Marsico M, Paytavi Gallart A, Sanseverino W, Aiese Cigliano R. GreenC 2.0: a comprehensive database of plant long non-coding RNAs. *Nucleic Acids Res*. 2022;50(D1):D1442–D1447. <https://doi.org/10.1093/nar/gkab1014>
- Do T, Qu Z, Searle I. Purification and functional analysis of plant long noncoding RNAs (lncRNA). *Methods Mol Biol*. 2019;1933:131–147. [https://doi.org/10.1007/978-1-4939-9045-0\\_7](https://doi.org/10.1007/978-1-4939-9045-0_7)
- Doyle JA. Molecular and fossil evidence on the origin of angiosperms. *Annu Rev Earth Planet Sci*. 2012;40(1):301–326. <https://doi.org/10.1146/annurev-earth-042711-105313>
- Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods*. 2007;4(9):721–726. <https://doi.org/10.1038/nmeth1079>
- Erdmann RM, Picard CL. RNA-directed DNA methylation. *PLoS Genet*. 2020;16(10):e1009034. <https://doi.org/10.1371/journal.pgen.1009034>
- Fajkus P, Kilar A, Nelson ADL, Holá M, Peška V, Goffová I, Fojtová M, Zachová D, Fulnečková J, Fajkus J. Evolution of plant telomerase RNAs: farther to the past, deeper to the roots. *Nucleic Acids Res*. 2021;49(13):7680–7694. <https://doi.org/10.1093/nar/gkab545>
- Fajkus P, Peška V, Závodník M, Fojtová M, Fulnečková J, Dobias Š, Kilar A, Dvořáčková M, Zachová D, Nečasová I, et al. Telomerase RNAs in land plants. *Nucleic Acids Res*. 2019;47(18):9842–9856. <https://doi.org/10.1093/nar/gkz695>
- Fan XN, Zhang SW. lncRNA-MFDL: identification of human long non-coding RNAs by fusing multiple features and using deep learning. *Mol Biosyst*. 2015;11(3):892–897. <https://doi.org/10.1039/c4mb00650j>
- Fang X, Wu Z, Raitskin O, Webb K, Voigt P, Lu T, Howard M, Dean C. The 3' processing of antisense RNAs physically links to chromatin-based transcriptional control. *Proc Natl Acad Sci USA*. 2020;117(26):15316–15321. <https://doi.org/10.1073/pnas.2007268117>
- Fang J, Zhang F, Wang H, Wang W, Zhao F, Li Z, Sun C, Chen F, Xu F, Chang S, et al. Locus shortens rice maturity duration without yield penalty. *Proc Natl Acad Sci USA*. 2019;116(37):18717–18722. <https://doi.org/10.1073/pnas.1815030116>
- Feng J, Funk WD, Wang SS, Weinrich SL, Avilion AA, Chiu CP, Adams RR, Chang E, Allsopp RC, Yu J. The RNA component of human telomerase. *Science*. 1995;269(5228):1236–1241. <https://doi.org/10.1126/science.7544491>
- Fesenko I, Shabalina SA, Mamaeva A, Knyazev A, Glushkevich A, Lyapina I, Ziganshin R, Kovalchuk S, Kharlampieva D, Lazarev V, et al. A vast pool of lineage-specific microproteins encoded by long non-coding RNAs in plants. *Nucleic Acids Res*. 2021;49(18):10328–10346. <https://doi.org/10.1093/nar/gkab816>
- Filichkin S, Priest HD, Megraw M, Mockler TC. Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr Opin Plant Biol*. 2015;24:125–135. <https://doi.org/10.1016/j.pbi.2015.02.008>
- Fowler S, Thomashow MF. Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell*. 2002;14(8):1675–1690. <https://doi.org/10.1105/tpc.003483>
- Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet*. 2007;39(8):1033–1037. <https://doi.org/10.1038/ng2079>
- Friml J, Yang X, Michniewicz M, Weijers D, Quint A, Tietz O, Benjamins R, Ouwerkerk PBF, Ljung K, Sandberg G, et al. A PINOID-dependent binary switch in apical-basal PIN polar targeting directs auxin efflux. *Science*. 2004;306(5697):862–865. <https://doi.org/10.1126/science.1100618>
- German MA, Luo S, Schroth G, Meyers BC, Green PJ. Construction of parallel analysis of RNA ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat Protoc*. 2009;4(3):356–362. <https://doi.org/10.1038/nprot.2009.8>
- Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*. 2014;159(3):647–661. <https://doi.org/10.1016/j.cell.2014.09.029>
- Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013;154(2):442–451. <https://doi.org/10.1016/j.cell.2013.06.044>
- Gong Y, Huang H-T, Liang Y, Trimarchi T, Aifantis I, Tsiganos A. lncRNA-screen: an interactive platform for computationally screening long non-coding RNAs in large genomics datasets. *BMC Genomics*. 2017;18(1):434. <https://doi.org/10.1186/s12864-017-3817-0>
- Gosai SJ, Foley SW, Wang D, Silverman IM, Selamoglu N, Nelson ADL, Beilstein MA, Daldal F, Deal RB, Gregory BD. Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the Arabidopsis nucleus. *Mol Cell*. 2015;57(2):376–388. <https://doi.org/10.1016/j.molcel.2014.12.004>
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res*. 2003;31(1):439–441. <https://doi.org/10.1093/nar/gkg006>
- Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH, Ecker JR. A link between RNA metabolism and



- silencing affecting Arabidopsis development. *Dev Cell*. 2008;**14**(6): 854–866. <https://doi.org/10.1016/j.devcel.2008.04.005>
- Greider CW, Blackburn EH.** A telomeric sequence in the RNA of tetrahymena telomerase required for telomere repeat synthesis. *Nature*. 1989;**337**(6205):331–337. <https://doi.org/10.1038/337331a0>
- Gu X, Le C, Wang Y, Li Z, Jiang D, Wang Y, He Y.** Arabidopsis FLC clade members form flowering-repressor complexes coordinating responses to endogenous and environmental cues. *Nat Commun*. 2013;**4**(1):1947. <https://doi.org/10.1038/ncomms2947>
- Guh C-Y, Hsieh Y-H, Chu H-P.** Functions and properties of nuclear lncRNAs-from systematically mapping the interactomes of lncRNAs. *J Biomed Sci*. 2020;**27**(1):44. <https://doi.org/10.1186/s12929-020-00640-3>
- Guo J-C, Fang S-S, Wu Y, Zhang J-H, Chen Y, Liu J, Wu B, Wu J-R, Li E-M, Xu L-Y, et al.** CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition. *Nucleic Acids Res*. 2019;**47**(W1): W516–W522. <https://doi.org/10.1093/nar/gkz400>
- Guo Q, Shi X, Wang X.** RNA and liquid–liquid phase separation. *Noncoding RNA Res*. 2021;**6**(2):92–99. <https://doi.org/10.1016/j.ncrna.2021.04.003>
- Guo J, Wang S, Valerius O, Hall H, Zeng Q, Li J-F, Weston DJ, Ellis BE, Chen J-G.** Involvement of Arabidopsis RACK1 in protein translation and its regulation by abscisic acid. *Plant Physiol*. 2011;**155**(1): 370–383. <https://doi.org/10.1104/pp.110.160663>
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al.** Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;**458**(7235):223–227. <https://doi.org/10.1038/nature07672>
- Han S, Liang Y, Ma Q, Xu Y, Zhang Y, Du W, Wang C, Li Y.** Lncfinder: an integrated platform for long non-coding RNA identification utilizing sequence intrinsic composition, structural information and physicochemical property. *Brief Bioinformatics*. 2019;**20**(6): 2009–2027. <https://doi.org/10.1093/bib/bby065>
- Hawkes EJ, Hennelly SP, Novikova IV, Irwin JA, Dean C, Sanbonmatsu KY.** COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell Rep*. 2016;**16**(12): 3087–3096. <https://doi.org/10.1016/j.celrep.2016.08.045>
- Henriques R, Wang H, Liu J, Boix M, Huang L-F, Chua N-H.** The anti-phasic regulatory module comprising CDF5 and its antisense RNA FLORE links the circadian clock to photoperiodic flowering. *New Phytol*. 2017;**216**(3):854–867. <https://doi.org/10.1111/nph.14703>
- Heo JB, Lee Y-S, Sung S.** Epigenetic regulation by long noncoding RNAs in plants. *Chromosome Res*. 2013;**21**(6–7):685–693. <https://doi.org/10.1007/s10577-013-9392-6>
- Heo JB, Sung S.** Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*. 2011;**331**(6013):76–79. <https://doi.org/10.1126/science.1197349>
- Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I.** Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep*. 2015;**11**(7): 1110–1122. <https://doi.org/10.1016/j.celrep.2015.04.023>
- Hirsch J, Lefort V, Vankerschaver M, Boualem A, Lucas A, Thermes C, d'Aubenton-Carafa Y, Crespi M.** Characterization of 43 non-protein-coding mRNA genes in Arabidopsis, including the MIR162a-derived transcripts. *Plant Physiol*. 2006;**140**(4):1192–1204. <https://doi.org/10.1104/pp.105.073817>
- Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ.** RNA polymerase II collision interrupts convergent transcription. *Mol Cell*. 2012;**48**(3): 365–374. <https://doi.org/10.1016/j.molcel.2012.08.027>
- Holt C, Yandell M.** MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 2011;**12**(1):491. <https://doi.org/10.1186/1471-2105-12-491>
- Hong Y, Zhang Y, Cui J, Meng J, Chen Y, Zhang C, Yang J, Luan Y.** The lncRNA39896-miR166b-HDZs module affects tomato resistance to *Phytophthora infestans*. *J Integr Plant Biol*. 2022;**64**(10):1979–1993. <https://doi.org/10.1111/jipb.13339>
- Huh JH, Bauer MJ, Hsieh T-F, Fischer R.** Endosperm gene imprinting and seed development. *Curr Opin Genet Dev*. 2007;**17**(6):480–485. <https://doi.org/10.1016/j.gde.2007.08.011>
- Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A.** A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*. 2007;**8**(1):39. <https://doi.org/10.1186/1471-2164-8-39>
- Ito EA, Katahira I, Vicente FFDR, Pereira LFP, Lopes FM.** BASiNET—BiologicAI Sequences NETwork: a case study on coding and non-coding RNAs identification. *Nucleic Acids Res*. 2018;**46**(16):e96. <https://doi.org/10.1093/nar/gky462>
- Jabnourne M, Secco D, Lecampion C, Robaglia C, Shu Q, Poirier Y.** A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *Plant Cell*. 2013;**25**(10):4166–4182. <https://doi.org/10.1105/tpc.113.116251>
- Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, et al.** STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 2008;**37**(Database):D412–D416. <https://doi.org/10.1093/nar/gkn760>
- Jensen TI, Mikkelsen NS, Gao Z, Foßelteder J, Pabst G, Axelgaard E, Laustsen A, König S, Reinisch A, Bak RO.** Targeted regulation of transcription in primary cells using CRISPRa and CRISPRi. *Genome Res*. 2021;**31**(11):2120–2130. <https://doi.org/10.1101/gr.275607.121>
- Jin Y, Ivanov M, Dittrich AN, Nelson ADL, Marquardt S.** A trans-acting long non-coding RNA represses flowering in Arabidopsis; 2021b. [bioRxiv:2021.11.15.468639](https://doi.org/10.1101/2021.11.15.468639)
- Jin J, Lu P, Xu Y, Li Z, Yu S, Liu J, Wang H, Chua N-H, Cao P.** PLncDB V2.0: a comprehensive encyclopedia of plant long noncoding RNAs. *Nucleic Acids Res*. 2021a;**49**(D1):D1489–D1495. <https://doi.org/10.1093/nar/gkaa910>
- Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, Gao G.** CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res*. 2017;**45**(W1):W12–W16. <https://doi.org/10.1093/nar/gkx428>
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, et al.** RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007;**316**(5830):1484–1488. <https://doi.org/10.1126/science.1138341>
- Kar S, Bordiya Y, Rodriguez N, Kim J, Gardner EC, Gollihar JD, Sung S, Ellington AD.** Orthogonal control of gene expression in plants using synthetic promoters and CRISPR-based transcription factors. *Plant Methods*. 2022;**18**(1):42. <https://doi.org/10.1186/s13007-022-00867-1>
- Kawaji H, Lizio M, Itoh M, Kanamori-Katayama M, Kaiho A, Nishiyori-Sueki H, Shin JW, Kojima-Ishiyama M, Kawano M, Murata M, et al.** Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Res*. 2014;**24**(4):708–717. <https://doi.org/10.1101/gr.156232.113>
- Kearly A, Nelson ADL, Skirycz A, Chodasiewicz M.** Composition and function of stress granules and P-bodies in plants. In: *Seminars in cell & developmental biology*. Academic Press; 2022.
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al.** Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA*. 2009;**106**(28):11667–11672. <https://doi.org/10.1073/pnas.0904715106>
- Khong A, Matheny T, Jain S, Mitchell SF, Wheeler JR, Parker R.** The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol Cell*. 2017;**68**(4):808–820.e5. <https://doi.org/10.1016/j.molcel.2017.10.015>
- Kim D-H, Sung S.** Coordination of the vernalization response through a VIN3 and FLC gene family regulatory network in Arabidopsis.



- Plant Cell. 2013;**25**(2):454–469. <https://doi.org/10.1105/tpc.112.104760>
- Kim D-H, Sung S.** Vernalization-triggered intragenic chromatin loop formation by long noncoding RNAs. *Dev Cell*. 2017;**40**(3):302–312.e4. <https://doi.org/10.1016/j.devcel.2016.12.021>
- Kindgren P, Ard R, Ivanov M, Marquardt S.** Transcriptional read-through of the long non-coding RNA SVALKa governs plant cold acclimation. *Nat Commun*. 2018;**9**(1):4561. <https://doi.org/10.1038/s41467-018-07010-6>
- Kirk JM, Kim SO, Inoue K, Smola MJ, Lee DM, Schertzer MD, Wooten JS, Baker AR, Sprague D, Collins DW, et al.** Functional classification of long non-coding RNAs by k-mer content. *Nat Genet*. 2018;**50**(10):1474–1482. <https://doi.org/10.1038/s41588-018-0207-8>
- Kirov I, Dudnikov M, Merkulov P, Shingaliev A, Omarov M, Kolganova E, Sigaeva A, Karlov G, Soloviev A.** Nanopore RNA sequencing revealed long non-coding and LTR retrotransposon-related RNAs expressed at early stages of triticale SEED development. *Plants*. 2020;**9**(12):1794. <https://doi.org/10.3390/plants9121794>
- Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA.** A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J*. 2016;**88**(6):1058–1070. <https://doi.org/10.1111/tpj.13312>
- Kramer MC, Kim HJ, Palos KR, Garcia BA, Lyons E, Beilstein MA, Nelson ADL, Gregory BD.** A conserved long intergenic non-coding RNA containing snoRNA sequences, IncCOBRA1, affects *Arabidopsis* germination and development. *Front Plant Sci*. 2022;**13**:906603. <https://doi.org/10.3389/fpls.2022.906603>
- Langfelder P, Horvath S.** WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;**9**(1):559. <https://doi.org/10.1186/1471-2105-9-559>
- Lee J, Lee I.** Regulation and function of SOC1, a flowering pathway integrator. *J Exp Bot*. 2010;**61**(9):2247–2254. <https://doi.org/10.1093/jxb/erq098>
- Li L, Eichten SR, Shimizu R, Petsch K, Yeh C-T, Wu W, Chetoor AM, Givan SA, Cole RA, Fowler JE, et al.** Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*. 2014b;**15**(2):R40. <https://doi.org/10.1186/gb-2014-15-2-r40>
- Li L, Luo H, Lim D-H, Han L, Li Y, Fu X-D, Qi Y.** Global profiling of RNA-chromatin interactions reveals co-regulatory gene expression networks in *Arabidopsis*. *Nat Plants*. 2021;**7**(10):1364–1378. <https://doi.org/10.1038/s41477-021-01004-x>
- Li L, Wang X, Sasidharan R, Stolc V, Deng W, He H, Korbel J, Chen X, Tongprasit W, Ronald P, et al.** Global identification and characterization of transcriptionally active regions in the rice genome. *PLoS One*. 2007;**2**(3):e294. <https://doi.org/10.1371/journal.pone.0000294>
- Li L, Wang X, Stolc V, Li X, Zhang D, Su N, Tongprasit W, Li S, Cheng Z, Wang J, et al.** Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet*. 2006;**38**(1):124–129. <https://doi.org/10.1038/ng1704>
- Li A, Zhang J, Zhou Z.** PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics*. 2014a;**15**(1):311. <https://doi.org/10.1186/1471-2105-15-311>
- Liao J-Y, Wu J, Wang Y-J, He J-H, Deng W-X, Hu KS, Zhang Y-C, Zhang Y, Yan H, Wang D-L, et al.** Deep sequencing reveals a global reprogramming of lncRNA transcriptome during EMT. *Biochim Biophys Acta Mol Cell Res*. 2017;**1864**(10):1703–1713. <https://doi.org/10.1016/j.bbamcr.2017.06.003>
- Lin S-I, Chiang S-F, Lin W-Y, Chen J-W, Tseng C-Y, Wu P-C, Chiou T-J.** Regulatory network of microRNA399 and PHO2 by systemic signaling. *Plant Physiol*. 2008;**147**(2):732–746. <https://doi.org/10.1104/pp.108.116269>
- Lin S, Zhang L, Luo W, Zhang X.** Characteristics of antisense transcript promoters and the regulation of their activity. *Int J Mol Sci*. 2015;**17**(1):9. <https://doi.org/10.3390/ijms17010009>
- Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, He D, Attenello FJ, Villalta JE, Cho MY, Chen Y, et al.** CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*. 2017;**355**(6320):aah7111. <https://doi.org/10.1126/science.aah7111>
- Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H.** Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell*. 2012;**24**(11):4333–4345. <https://doi.org/10.1105/tpc.112.102855>
- Liu X, Li D, Zhang D, Yin D, Zhao Y, Ji C, Zhao X, Li X, He Q, Chen R, et al.** A novel antisense long noncoding RNA, *TWISTED LEAF*, maintains leaf blade flattening by regulating its associated sense R2R3-MYB gene in rice. *New Phytologist*. 2018;**218**(2):774–788. <https://doi.org/10.1111/nph.15023>
- Liu F, Marquardt S, Lister C, Swiezewski S, Dean C.** Targeted 3' processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing. *Science*. 2010;**327**(5961):94–97. <https://doi.org/10.1126/science.1180278>
- Liu C, Muchhal US, Raghothama KG.** Differential expression of TPS11, a phosphate starvation-induced gene in tomato. *Plant Mol Biol*. 1997;**33**(5):867–874. <https://doi.org/10.1023/A:1005729309569>
- Liu Y, Teng C, Xia R, Meyers BC.** PhasiRNAs in plants: their biogenesis, genic sources, and roles in stress responses, development, and reproduction. *Plant Cell*. 2020;**32**(10):3059–3080. <https://doi.org/10.1105/tpc.20.00335>
- Liu C, Xi W, Shen L, Tan C, Yu H.** Regulation of floral patterning by flowering time genes. *Dev Cell*. 2009;**16**(5):711–722. <https://doi.org/10.1016/j.devcel.2009.03.011>
- Liu P, Zhang Y, Zou C, Yang C, Pan G, Ma L, Shen Y.** Integrated analysis of long non-coding RNAs and mRNAs reveals the regulatory network of maize seedling root responding to salt stress. *BMC Genomics*. 2022;**23**(1):50. <https://doi.org/10.1186/s12864-021-08286-7>
- Liu T-T, Zhu D, Chen W, Deng W, He H, He G, Bai B, Qi Y, Chen R, Deng XW.** A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in *Oryza sativa*. *Mol Plant*. 2013;**6**(3):830–846. <https://doi.org/10.1093/mp/sss087>
- Liyatan I, Harikumar A, Nissim-Rafinia M, Duttagupta R, Gingeras TR, Meshorer E.** Non-polyadenylated transcription in embryonic stem cells reveals novel non-coding RNA related to pluripotency and differentiation. *Nucleic Acids Res*. 2013;**41**(12):6300–6315. <https://doi.org/10.1093/nar/gkt316>
- Lowder LG, Paul JW, Qi Y.** Multiplexed transcriptional activation or repression in plants using CRISPR-dCas9-based systems. In: **Kaufmann K, Mueller-Roeber B**, editors. *Plant gene regulatory networks: methods and protocols*. New York (NY): Springer; 2017. p. 167–184.
- Lucero L, Ferrero L, Fonouni-Farde C, Ariel F.** Functional classification of plant long noncoding RNAs: a transcript is known by the company it keeps. *New Phytol*. 2021;**229**(3):1251–1260. <https://doi.org/10.1111/nph.16903>
- Ma L, Bajic VB, Zhang Z.** On the classification of long non-coding RNAs. *RNA Biol*. 2013;**10**(6):925–933. <https://doi.org/10.4161/rna.24604>
- MacIntosh GC, Wilkerson C, Green PJ.** Identification and analysis of *Arabidopsis* expressed sequence tags characteristic of non-coding RNAs. *Plant Physiol*. 2001;**127**(3):765–776. <https://doi.org/10.1104/pp.010501>
- Magallón S, Hilu KW, Quandt D.** Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am J Bot*. 2013;**100**(3):556–573. <https://doi.org/10.3732/ajb.1200416>
- Marker C, Zemann A, Terhöst T, Kiefmann M, Kastenmayer JP, Green P, Bachelier J-P, Brosius J, Hüttenhofer A.** Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr Biol*. 2002;**12**(23):2002–2013. [https://doi.org/10.1016/S0960-9822\(02\)01304-0](https://doi.org/10.1016/S0960-9822(02)01304-0)

- Maruri-López I, Figueroa NE, Hernández-Sánchez IE, Chodasiewicz M. Plant stress granules: trends and beyond. *Front Plant Sci*. 2021;12:722643. <https://doi.org/10.3389/fpls.2021.722643>
- Mas AM, Huarte M. lncRNA-DNA hybrids regulate distant genes. *EMBO Rep*. 2020;21(3):e50107. <https://doi.org/10.15252/embr.202050107>
- Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M, Nambara E, Nakajima M, Kawashima M, et al. Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol*. 2008;49(8):1135–1149. <https://doi.org/10.1093/pcp/pcn101>
- Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*. 2014;15(6):394–408. <https://doi.org/10.1038/nrg3683>
- Matzke MA, Primig M, Trnovsky J, Matzke AJ. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J*. 1989;8(3):643–649. <https://doi.org/10.1002/j.1460-2075.1989.tb03421.x>
- Medina J, Bagues M, Terol J, Pérez-Alonso M, Salinas J. The Arabidopsis CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression is regulated by low temperature but not by abscisic acid or Dehydration1. *Plant Physiol*. 1999;119(2):463–470. <https://doi.org/10.1104/pp.119.2.463>
- Medina J, Catalá R, Salinas J. The CBFs: three Arabidopsis transcription factors to cold acclimate. *Plant Sci*. 2011;180(1):3–11. <https://doi.org/10.1016/j.plantsci.2010.06.019>
- Meng J, Kang Q, Chang Z, Luan Y. PlncRNA-HDeep: plant long non-coding RNA prediction using hybrid deep learning based on two encoding styles. *BMC Bioinformatics*. 2021;22(S3):242. <https://doi.org/10.1186/s12859-020-03870-2>
- Meyers BC, Tej SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S. The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. *Genome Res*. 2004;14(8):1641–1653. <https://doi.org/10.1101/gr.2275604>
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res*. 2021;49(D1):D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- Moghe GD, Lehti-Shiu MD, Seddon AE, Yin S, Chen Y, Juntawong P, Brandizzi F, Bailey-Serres J, Shiu S-H. Characteristics and significance of intergenic polyadenylated RNA transcription in Arabidopsis. *Plant Physiol*. 2013;161(1):210–224. <https://doi.org/10.1104/pp.112.205245>
- Mohammad S, Edgar PP, Pires JC, Schranz ME. Positionally-conserved but sequence-diverged: identification of long non-coding RNAs in the Brassicaceae and Cleomaceae. *BMC Plant Biol*. 2015;15(1):217. <https://doi.org/10.1186/s12870-015-0603-5>
- Moison M, Pacheco JM, Lucero L, Fonouni-Farde C, Rodríguez-Melo J, Mansilla N, Christ A, Bazin J, Benhamed M, Ibañez F, et al. The lncRNA APOLO interacts with the transcription factor WRKY42 to trigger root hair cell expansion in response to cold. *Mol Plant*. 2021;14(6):937–948. <https://doi.org/10.1016/j.molp.2021.03.008>
- Morris RJ. On the selectivity, specificity and signalling potential of the long-distance movement of messenger RNA. *Curr Opin Plant Biol*. 2018;43:1–7. <https://doi.org/10.1016/j.pbi.2017.11.001>
- Musacchia F, Basu S, Petrosino G, Salvemini M, Sanges R. Annocript: a flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. *Bioinformatics*. 2015;31(13):2199–2201. <https://doi.org/10.1093/bioinformatics/btv106>
- Napoli C, Lemieux C, Jorgensen R. Introduction of a chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell*. 1990;2(4):279–289. <https://doi.org/10.2307/3869076>
- Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grützner F, Kaessmann H. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*. 2014;505(7485):635–640. <https://doi.org/10.1038/nature12943>
- Nelson ADL, Devisetty UK, Palos K, Haug-Baltzell AK, Lyons E, Beilstein MA. Evolinc: a tool for the identification and evolutionary comparison of long intergenic non-coding RNAs. *Front Genet*. 2017;8:52. <https://doi.org/10.3389/fgene.2017.00052>
- Nelson ADL, Forsythe ES, Devisetty UK, Clausen DS, Haug-Baltzell AK, Meldrum AMR, Frank MR, Lyons E, Beilstein MA. A genomic analysis of factors driving lincRNA diversification: lessons from plants. *G3*. 2016;6(9):2881–2891. <https://doi.org/10.1534/g3.116.030338>
- Nelson ADL, Shippen DE. Evolution of TERT-interacting lncRNAs: expanding the regulatory landscape of telomerase. *Front Genet*. 2015;6:277. <https://doi.org/10.3389/fgene.2015.00277>
- Nelson Ditttrich AC, Nelson ADL. High-throughput evolutionary comparative analysis of long intergenic noncoding RNAs in multiple organisms. *Methods Mol Biol*. 2022;2512:45–60. [https://doi.org/10.1007/978-1-0716-2429-6\\_3](https://doi.org/10.1007/978-1-0716-2429-6_3)
- Ni W, Xu S-L, Chalkley RJ, Pham TND, Guan S, Maltby DA, Burlingame AL, Wang Z-Y, Quail PH. Multisite light-induced phosphorylation of the transcription factor PIF3 is necessary for both its rapid degradation and concomitant negative feedback modulation of photoreceptor phyB levels in Arabidopsis. *Plant Cell*. 2013;25(7):2679–2698. <https://doi.org/10.1105/tpc.113.112342>
- Noh JH, Kim KM, McClusky WG, Abdelmohsen K, Gorospe M. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip Rev RNA*. 2018;9(3):e1471. <https://doi.org/10.1002/wrna.1471>
- Nolte W, Weikard R, Albrecht E, Hammon HM, Kühn C. Metabogenomic analysis to functionally annotate the regulatory role of long non-coding RNAs in the liver of cows with different nutrient partitioning phenotype. *Genomics*. 2022;114(1):202–214. <https://doi.org/10.1016/j.ygeno.2021.12.004>
- Nover L, Scharf KD, Neumann D. Formation of cytoplasmic heat shock granules in tomato cell cultures and leaves. *Mol Cell Biol*. 1983;3:1648–1655. <https://doi.org/10.1128/mcb.3.9.1648-1655.1983>
- Nover L, Scharf KD, Neumann D. Cytoplasmic heat shock granules are formed from precursor particles and are associated with a specific set of mRNAs. *Mol Cell Biol*. 1989;9:1298–1308. <https://doi.org/10.1128/mcb.9.3.1298-1308.1989>
- Ochoa-Fernandez R, Abel NB, Wieland F-G, Schlegel J, Koch L-A, Miller JB, Engesser R, Giuriani G, Brandl SM, Timmer J, et al. Optogenetic control of gene expression in plants in the presence of ambient white light. *Nat Methods*. 2020;17(7):717–725. <https://doi.org/10.1038/s41592-020-0868-y>
- Okamoto M, Tatematsu K, Matsui A, Morosawa T, Ishida J, Tanaka M, Endo TA, Mochizuki Y, Toyoda T, Kamiya Y, et al. Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of Arabidopsis using tiling arrays. *Plant J*. 2010;62(1):39–51. <https://doi.org/10.1111/j.1365-313X.2010.04135.x>
- O'Malley RC, Huang S-SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. Cistrome and episcistrome features shape the regulatory DNA landscape. *Cell*. 2016;165(5):1280–1292. <https://doi.org/10.1016/j.cell.2016.04.038>
- Osato N, Yamada H, Satoh K, Ooka H, Yamamoto M, Suzuki K, Kawai J, Carninci P, Ohtomo Y, Murakami K, et al. Antisense transcripts with rice full-length cDNAs. *Genome Biol*. 2003;5(1):R5. <https://doi.org/10.1186/gb-2003-5-1-r5>
- Palos K, Nelson Ditttrich AC, Yu L, Brock JR, Railey CE, Wu H-YL, Sokolowska E, Skirycz A, Hsu PY, Gregory BD, et al. Identification and functional annotation of long intergenic non-coding RNAs in Brassicaceae. *Plant Cell*. 2022;34(9):3233–3260. <https://doi.org/10.1093/plcell/koac166>
- Pelayo MA, Yamaguchi N, Ito T. One factor, many systems: the floral homeotic protein AGAMOUS and its epigenetic regulatory mechanisms. *Curr Opin Plant Biol*. 2021;61:102009. <https://doi.org/10.1016/j.pbi.2021.102009>
- Plewka P, Thompson A, Szymanski M, Nuc P, Knop K, Rasinska A, Bialkowska A, Szwejkowska-Kulinska Z, Karłowski WM, Jarmolowski A. A stable tRNA-like molecule is generated from the

- long noncoding RNA GUT15 in *Arabidopsis*. *RNA Biol.* 2018;**15**:726–738. <https://doi.org/10.1080/15476286.2018.1445404>
- Polturak G, Osbourn A.** The emerging role of biosynthetic gene clusters in plant defense and plant interactions. *PLoS Pathog.* 2021;**17**(7): e1009698. <https://doi.org/10.1371/journal.ppat.1009698>
- Pontius JU, Wagner L, Schuler GD.** UniGene: a unified view of the transcriptome. *The NCBI handbook.* 2003;1:1–2.
- Portwood JL, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Sen TZ, Cho KT, Schott DA, et al.** MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.* 2019;**47**(D1):D1146–D1154. <https://doi.org/10.1093/nar/gky1046>
- Qin T, Zhao H, Cui P, Albeshier N, Xiong L.** A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. *Plant Physiol.* 2017;**175**(3):1321–1336. <https://doi.org/10.1104/pp.17.00574>
- Qu G, Kruska K, Plewka P, Yang S-Y, Chiou T-J, Jarmolowski A, Szwedkowska-Kulinska Z, Echeverria M, Karlowski WM.** Promoter-based identification of novel non-coding RNAs reveals the presence of dicistronic snoRNA-miRNA genes in *Arabidopsis thaliana*. *BMC Genomics.* 2015;**16**(1):1009. <https://doi.org/10.1186/s12864-015-2221-x>
- Reddy ASN.** Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol.* 2007;**58**(1):267–294. <https://doi.org/10.1146/annurev.plant.58.032806.103754>
- Reddy ASN, Marquez Y, Kalyana M, Barta A.** Complexity of the alternative splicing landscape in plants. *Plant Cell.* 2013;**25**(10):3657–3683. <https://doi.org/10.1105/tpc.113.117523>
- Riaño-Pachón DM, Dreyer I, Mueller-Roeber B.** Orphan transcripts in *Arabidopsis thaliana*: identification of several hundred previously unrecognized genes. *Plant J.* 2005;**43**(2):205–212. <https://doi.org/10.1111/j.1365-3113.2005.02438.x>
- Rigo R, Bazin J, Romero-Barrios N, Moison M, Lucero L, Christ A, Benhamed M, Blein T, Hugué S, Charon C, et al.** The *Arabidopsis* lncRNA ASCO modulates the transcriptome through interaction with splicing factors. *EMBO Rep.* 2020;**21**(5):e48977. <https://doi.org/10.15252/embr.201948977>
- Rinn JL, Chang HY.** Genome regulation by long noncoding RNAs. *Annu Rev Biochem.* 2012;**81**(1):145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902>
- Ross CJ, Rom A, Spinrad A, Gelbard-Solodkin D, Degani N, Ulitsky I.** Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. *Genome Biol.* 2021;**22**(1):29. <https://doi.org/10.1186/s13059-020-02247-1>
- Ross CJ, Ulitsky I.** Discovering functional motifs in long noncoding RNAs. *Wiley Interdiscipl Rev RNA.* 2022;**13**(4):e1708. <https://doi.org/10.1002/wrna.1708>
- Roulé T, Christ A, Hussain N, Huang Y, Hartmann C, Benhamed M, Gutierrez-Marcos J, Ariel F, Crespi M, Blein T.** The lncRNA MARS modulates the epigenetic reprogramming of the maternal cluster in response to ABA. *Mol Plant.* 2022;**15**(5):840–856. <https://doi.org/10.1016/j.molp.2022.02.007>
- Rymarquis LA, Kastenmayer JP, Hüttenhofer AG, Green PJ.** Diamonds in the rough: mRNA-like non-coding RNAs. *Trends Plant Sci.* 2008;**13**(7):329–334. <https://doi.org/10.1016/j.tplants.2008.02.009>
- Ryvkin P, Leung YY, Silverman IM, Childress M, Valladares O, Dragomir I, Gregory BD, Wang L-S.** HAMR: high-throughput annotation of modified ribonucleotides. *RNA.* 2013;**19**(12):1684–1692. <https://doi.org/10.1261/rna.036806.112>
- Sakurai T, Satou M, Akiyama K, Iida K, Seki M, Kuromori T, Ito T, Konagaya A, Toyoda T, Shinozaki K.** RARGE: a large-scale database of RIKEN *Arabidopsis* resources ranging from transcriptome to phenotype. *Nucleic Acids Res.* 2005;**33**(Database issue):D647–D650. <https://doi.org/10.1093/nar/gki014>
- Schneider HW, Raiol T, Brigido MM, Walter MEMT, Stadler PF.** A support vector machine based method to distinguish long non-coding RNAs from protein coding transcripts. *BMC Genomics.* 2017;**18**(1):804. <https://doi.org/10.1186/s12864-017-4178-4>
- Seo JS, Diloknawarit P, Park BS, Chua N-H.** ELF18-INDUCED LONG NONCODING RNA 1 evicts fibrillarin from mediator subunit to enhance PATHOGENESIS-RELATED GENE 1 (PR1) expression. *New Phytol.* 2019;**221**(4):2067–2079. <https://doi.org/10.1111/nph.15530>
- Seo JS, Sun H-X, Park BS, Huang C-H, Yeh S-D, Jung C, Chua N-H.** ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in *Arabidopsis*. *Plant Cell.* 2017;**29**(5):1024–1038. <https://doi.org/10.1105/tpc.16.00886>
- Shen L, Liang Z, Wong CE, Yu H.** Messenger RNA modifications in plants. *Trends Plant Sci.* 2019;**24**(4):328–341. <https://doi.org/10.1016/j.tplants.2019.01.005>
- Shin J-H, Chekanova JA.** *Arabidopsis* RRP6L1 and RRP6L2 function in FLOWERING LOCUS C silencing via regulation of antisense RNA synthesis. *PLoS Genet.* 2014;**10**(9):e1004612. <https://doi.org/10.1371/journal.pgen.1004612>
- Shumayla, Sharma S, Taneja M, Tyagi S, Singh K, Upadhyay SK.** Survey of high throughput RNA-seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. *Front Plant Sci.* 2017;**8**:1019. <https://doi.org/10.3389/fpls.2017.01019>
- Silverman IM, Li F, Alexander A, Goff L, Trapnell C, Rinn JL, Gregory BD.** RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol.* 2014;**15**(1):R3. <https://doi.org/10.1186/gb-2014-15-1-r3>
- Simopoulos CMA, Weretilnyk EA, Golding GB.** Molecular traits of long non-protein coding RNAs from diverse plant species show little evidence of phylogenetic relationships. *G3.* 2019;**9**(8):2511–2520. <https://doi.org/10.1534/g3.119.400201>
- Singer MS, Gottschling DE.** TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. *Science.* 1994;**266**(5184):404–409. <https://doi.org/10.1126/science.7545955>
- Singh U, Khemka N, Rajkumar MS, Garg R, Jain M.** PLncPRO for prediction of long non-coding RNAs (lncRNAs) in plants and its application for discovery of abiotic stress-responsive lncRNAs in rice and chickpea. *Nucleic Acids Res.* 2017;**45**(22):e183. <https://doi.org/10.1093/nar/gkx866>
- Song J, Logeswaran D, Castillo-González C, Li Y, Bose S, Aklilu BB, Ma Z, Polkhovskiy A, Chen JJ-L, Shippen DE.** The conserved structure of plant telomerase RNA provides the missing link for an evolutionary pathway from ciliates to humans. *Proc Natl Acad Sci USA.* 2019;**116**(49):24542–24550. <https://doi.org/10.1073/pnas.1915312116>
- Song D, Yang Y, Yu B, Zheng B, Deng Z, Lu B-L, Chen X, Jiang T.** Computational prediction of novel non-coding RNAs in *Arabidopsis thaliana*. *BMC Bioinformatics.* 2009;**10**(S1):S36. <https://doi.org/10.1186/1471-2105-10-S1-S36>
- Su Z-D, Huang Y, Zhang Z-Y, Zhao Y-W, Wang D, Chen W, Chou K-C, Lin H.** iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics.* 2018;**34**(24):4196–4204. <https://doi.org/10.1093/bioinformatics/bty508>
- Sun L, Liu H, Zhang L, Meng J.** lncRScan-SVM: a tool for predicting long non-coding RNAs using support vector machine. *PLoS One.* 2015;**10**(10):e0139654. <https://doi.org/10.1371/journal.pone.0139654>
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y.** Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 2013;**41**(17):e166. <https://doi.org/10.1093/nar/gkt646>
- Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, et al.** The *Arabidopsis* information resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 2008;**36**(Database):D1009–D1014. <https://doi.org/10.1093/nar/gkm965>
- Swiezewski S, Liu F, Magusin A, Dean C.** Cold-induced silencing by long antisense transcripts of an *Arabidopsis* polycomb target. *Nature.* 2009;**462**(7274):799–802. <https://doi.org/10.1038/nature08618>



- Syed NH, Kalyna M, Marquez Y, Barta A, Brown JWS. Alternative splicing in plants—coming of age. *Trends Plant Sci.* 2012;**17**(10): 616–623. <https://doi.org/10.1016/j.tplants.2012.06.001>
- Szcześniak MW, Bryzghalov O, Ciomborowska-Basheer J, Makalowska I. CANTATAdb 2.0: expanding the collection of plant long noncoding RNAs. *Methods Mol Biol.* 2019;**1933**:415–429. [https://doi.org/10.1007/978-1-4939-9045-0\\_26](https://doi.org/10.1007/978-1-4939-9045-0_26)
- Taipale M, Rea S, Richter K, Vilar A, Lichter P, Imhof A, Akhtar A. hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol Cell Biol.* 2005;**25**(15):6798–6810. <https://doi.org/10.1128/MCB.25.15.6798-6810.2005>
- Teramoto H, Toyama T, Takeba G, Tsuji H. Noncoding RNA for CR20, a cytokinin-repressed gene of cucumber. *Plant Mol Biol.* 1996;**32**(5): 797–808. <https://doi.org/10.1007/BF00020478>
- Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, Miñambres M, Walther D, Schulze WX, Paz-Ares J, et al. Endogenous Arabidopsis messenger RNAs transported to distant tissues. *Nat Plants.* 2015;**1**(4):15025. <https://doi.org/10.1038/nplants.2015.25>
- Thomas HR, Frank MH. Connecting the pieces: uncovering the molecular basis for long-distance communication through plant grafting. *New Phytol.* 2019;**223**(2):582–589. <https://doi.org/10.1111/nph.15772>
- Tian Y, Zheng H, Zhang F, Wang S, Ji X, Xu C, He Y, Ding Y. PRC2 recruitment and H3K27me3 deposition at FLC require FCA binding of COOLAIR. *Sci Adv.* 2019;**5**(4):eaau7246. <https://doi.org/10.1126/sciadv.aau7246>
- Todesco M, Rubio-Somoza I, Paz-Ares J, Weigel D. A collection of target mimics for comprehensive analysis of microRNA function in *Arabidopsis thaliana*. *PLoS Genet.* 2010;**6**(7):e1001031. <https://doi.org/10.1371/journal.pgen.1001031>
- Tokizawa M, Kusunoki K, Koyama H, Kurotani A, Sakurai T, Suzuki Y, Sakamoto T, Kurata T, Yamamoto YY. Identification of Arabidopsis genic and non-genic promoters by paired-end sequencing of TSS tags. *Plant J.* 2017;**90**(3):587–605. <https://doi.org/10.1111/tpj.13511>
- Tripathi R, Patel S, Kumari V, Chakraborty P, Varadwaj PK. DeepLNC, a long non-coding RNA prediction tool using deep neural network. *Netw Model Anal Health Inform Bioinform.* 2016;**5**(1):21. <https://doi.org/10.1007/s13721-016-0129-2>
- Tsuzuki M, Sethuraman S, Coke AN, Rothi MH, Boyle AP, Wierzbicki AT. Broad noncoding transcription suggests genome surveillance by RNA polymerase V. *Proc Natl Acad Sci USA.* 2020;**117**(48):30799–30804. <https://doi.org/10.1073/pnas.2014419117>
- Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell.* 2011;**147**(7):1537–1550. <https://doi.org/10.1016/j.cell.2011.11.055>
- Unfried JP, Ulitsky I. Substoichiometric action of long noncoding RNAs. *Nat Cell Biol.* 2022;**24**(5):608–615. <https://doi.org/10.1038/s41556-022-00911-1>
- van de Sande K, Pawlowski K, Czaja I, Wieneke U, Schell J, Schmidt J, Walden R, Matvienko M, Wellink J, van Kammen A, et al. Modification of phytohormone response by a peptide encoded by ENOD40 of legumes and a nonlegume. *Science.* 1996;**273**(5273): 370–373. <https://doi.org/10.1126/science.273.5273.370>
- Vandivier LE, Campos R, Kuksa PP, Silverman IM, Wang L-S, Gregory BD. Chemical modifications mark alternatively spliced and uncapped messenger RNAs in Arabidopsis. *Plant Cell.* 2015;**27**(11):3024–3037. <https://doi.org/10.1105/tpc.15.00591>
- Voinnet O, Vain P, Angell S, Baulcombe DC. Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell.* 1998;**95**(2):177–187. [https://doi.org/10.1016/S0092-8674\(00\)81749-3](https://doi.org/10.1016/S0092-8674(00)81749-3)
- Walden N, Nguyen T-P, Mandáková T, Lysak MA, Schranz ME. Genomic blocks in *Aethionema arabicum* support Arabideae as next diverging clade in Brassicaceae. *Front Plant Sci.* 2020;**11**:719. <https://doi.org/10.3389/fpls.2020.00719>
- Wang H-LV, Chekanova JA. Long noncoding RNAs in plants. In: Rao MRS, editor. Long non coding RNA biology. Singapore: Springer; 2017. p. 133–154.
- Wang H, Chua N-H, Wang X-J. Prediction of trans-antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* 2006;**7**(10):R92. <https://doi.org/10.1186/gb-2006-7-10-r92>
- Wang Y, Fan X, Lin F, He G, Terzaghi W, Zhu D, Deng XW. Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light. *Proc Natl Acad Sci USA.* 2014a;**111**(28):10359–10364. <https://doi.org/10.1073/pnas.1409457111>
- Wang X-J, Gaasterland T, Chua N-H. Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* 2005;**6**(4):R30. <https://doi.org/10.1186/gb-2005-6-4-r30>
- Wang Y, Li S, Zhao Y, You C, Le B, Gong Z, Mo B, Xia Y, Chen X. NAD<sup>+</sup>-capped RNAs are widespread in the *Arabidopsis* transcriptome and can probably be translated. *Proc Natl Acad Sci USA.* 2019b;**116**(24):12094–12102. <https://doi.org/10.1073/pnas.1903682116>
- Wang Y, Luo X, Sun F, Hu J, Zha X, Su W, Yang J. Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat Commun.* 2018;**9**(1):3516. <https://doi.org/10.1038/s41467-018-05829-7>
- Wang J, Mei J, Ren G. Plant microRNAs: biogenesis, homeostasis, and degradation. *Front Plant Sci.* 2019a;**10**:360. <https://doi.org/10.3389/fpls.2019.00360>
- Wang T, Song H, Wei Y, Li P, Hu N, Liu J, Zhang B, Peng R. High throughput deep sequencing elucidates the important role of lncRNAs in foxtail millet response to herbicides. *Genomics.* 2020;**112**(6):4463–4473. <https://doi.org/10.1016/j.ygeno.2020.07.045>
- Wang Y, Wang X, Deng W, Fan X, Liu T-T, He G, Chen R, Terzaghi W, Zhu D, Deng XW. Genomic features and regulatory roles of intermediate-sized non-coding RNAs in Arabidopsis. *Mol Plant.* 2014b;**7**(3):514–527. <https://doi.org/10.1093/mp/ss177>
- Waseem M, Yang X, Aslam MM, Li M, Zhu L, Chen S, Zhu G, Li Y, Liu P. Genome-wide identification of long non-coding RNAs in two contrasting rapeseed (*Brassica napus* L.) genotypes subjected to cold stress. *Environ Exp Bot.* 2022;**201**:104969. <https://doi.org/10.1016/j.envexpbot.2022.104969>
- Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* 2014;**24**(4):616–628. <https://doi.org/10.1101/gr.165035.113>
- Wassenegger M, Heimes S, Riedel L, Sönger HL. RNA-directed de novo methylation of genomic sequences in plants. *Cell.* 1994;**76**(3): 567–576. [https://doi.org/10.1016/0092-8674\(94\)90119-8](https://doi.org/10.1016/0092-8674(94)90119-8)
- Waterhouse PM, Graham MW, Wang MB. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci USA.* 1998;**95**(23): 13959–13964. <https://doi.org/10.1073/pnas.95.23.13959>
- Whittaker C, Dean C. The FLC locus: a platform for discoveries in epigenetics and adaptation. *Annu Rev Cell Dev Biol.* 2017;**33**(1): 555–575. <https://doi.org/10.1146/annurev-cellbio-100616-060546>
- Wen J, Parker BJ, Weiller GF. In Silico identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula*. In *Silico Biol.* 2007;**7**(4-5):485–505
- Wierzbicki AT, Blevins T, Swiezewski S. Long noncoding RNAs in plants. *Annu Rev Plant Biol.* 2021;**72**(1):245–271. <https://doi.org/10.1146/annurev-arplant-093020-035446>
- Willmann MR, Berkowitz ND, Gregory BD. Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes—gMUCT 2.0. *Methods.* 2014;**67**(1):64–73. <https://doi.org/10.1016/j.ymeth.2013.07.003>
- Wu H-W, Deng S, Xu H, Mao H-Z, Liu J, Niu Q-W, Wang H, Chua N-H. A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves.



- New Phytol. 2018;219(4):1480–1491. <https://doi.org/10.1111/nph.15231>
- Wu M, Yang L-Z, Chen L-L. Long noncoding RNA and protein abundance in lncRNPs. RNA. 2021;27(12):1427–1440. <https://doi.org/10.1261/rna.078971.121>
- Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H, et al. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Res. 2017;45(8):e57. <https://doi.org/10.1093/nar/gkw1306>
- Xiao Q, Huang X, Zhang Y, Xu W, Yang Y, Zhang Q, Hu Z, Xing F, Sun Q, Li G, et al. The landscape of promoter-centred RNA-DNA interactions in rice. Nat Plants. 2022;8(2):157–170. <https://doi.org/10.1038/s41477-021-01089-4>
- Xu C, Wu Z, Duan H-C, Fang X, Jia G, Dean C. R-loop resolution promotes co-transcriptional chromatin silencing. Nat Commun. 2021;12(1):1790. <https://doi.org/10.1038/s41467-021-22083-6>
- Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, et al. Empirical analysis of transcriptional activity in the Arabidopsis genome. Science. 2003;302(5646):842–846. <https://doi.org/10.1126/science.1088305>
- Yang YW, Flynn RA, Chen Y, Qu K, Wan B, Wang KC, Lei M, Chang HY. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. Elife. 2014;3:e02046. <https://doi.org/10.7554/eLife.02046>
- Yang C, Yang L, Zhou M, Xie H, Zhang C, Wang MD, Zhu H. LncADeep: an ab initio lncRNA identification and functional annotation tool based on deep learning. Bioinformatics. 2018;34(22):3825–3834. <https://doi.org/10.1093/bioinformatics/bty428>
- Yang M, Zhu P, Cheema J, Bloomer R, Mikulski P, Liu Q, Zhang Y, Dean C, Ding Y. In vivo single-molecule analysis reveals COOLAIR RNA structural diversity. Nature. 2022;609(7926):394–399. <https://doi.org/10.1038/s41586-022-05135-9>
- Yu X, Willmann MR, Vandivier LE, Trefely S, Kramer MC, Shapiro J, Guo R, Lyons E, Snyder NW, Gregory BD. Messenger RNA 5' NAD<sup>+</sup> capping is a dynamic regulatory epitranscriptome mark that is required for proper response to abscisic acid in Arabidopsis. Dev Cell. 2021;56(1):125–140.e6. <https://doi.org/10.1016/j.devcel.2020.11.009>
- Yu S, Zhang Z, Li J, Zhu Y, Yin Y, Zhang X, Dai Y, Zhang A, Li C, Zhu Y, et al. Genome-wide identification and characterization of lncRNAs in sunflower endosperm. BMC Plant Biol. 2022;22(1):494. <https://doi.org/10.1186/s12870-022-03882-5>
- Yuan J, Li J, Yang Y, Tan C, Zhu Y, Hu L, Qi Y, Lu ZJ. Stress-responsive regulation of long non-coding RNA polyadenylation in *Oryza sativa*. Plant J. 2018;93(5):814–827. <https://doi.org/10.1111/tpj.13804>
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. Genome Biol. 2009;10(6):R62. <https://doi.org/10.1186/gb-2009-10-6-r62>
- Zhang Y-C, Liao J-Y, Li Z-Y, Yu Y, Zhang J-P, Li Q-F, Qu L-H, Shu W-S, Chen Y-Q. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biol. 2014;15(12):512. <https://doi.org/10.1186/s13059-014-0512-1>
- Zhang Z, Zheng Y, Ham B-K, Zhang S, Fei Z, Lucas WJ. Plant lncRNAs are enriched in and move systemically through the phloem in response to phosphate deficiency. J Integr Plant Biol. 2019;61(4):492–508. <https://doi.org/10.1111/jipb.12715>
- Zhang Z, Zhong H, Nan B, Xiao B. Global identification and integrated analysis of heat-responsive long non-coding RNAs in contrasting rice cultivars. Theor Appl Genet. 2022b;135(3):833–852. <https://doi.org/10.1007/s00122-021-04001-y>
- Zhang Y-C, Zhou Y-F, Cheng Y, Huang J-H, Lian J-P, Yang L, He R-R, Lei M-Q, Liu Y-W, Yuan C, et al. Genome-wide analysis and functional annotation of chromatin-enriched noncoding RNAs in rice during somatic cell regeneration. Genome Biol. 2022a;23(1):28. <https://doi.org/10.1186/s13059-022-02608-y>
- Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nat Commun. 2018;9(1):5056. <https://doi.org/10.1038/s41467-018-07500-7>
- Zhao J, Sun BK, Erwin JA, Song J-J, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322(5902):750–756. <https://doi.org/10.1126/science.1163045>
- Zhao C, Zhang Z, Xie S, Si T, Li Y, Zhu J-K. Mutational evidence for the critical role of CBF transcription factors in cold acclimation in Arabidopsis. Plant Physiol. 2016;171(4):2744–2759. <https://doi.org/10.1104/pp.16.00533>
- Zhou Y-F, Zhang Y-C, Sun Y-M, Yu Y, Lei M-Q, Yang Y-W, Lian J-P, Feng Y-Z, Zhang Z, Yang L, et al. The parent-of-origin lncRNA MISSEN regulates rice endosperm development. Nat Commun. 2021;12(1):6525. <https://doi.org/10.1038/s41467-021-26795-7>
- Zhu Y, Chen L, Hong X, Shi H, Li X. Revealing the novel complexity of plant long non-coding RNA by strand-specific and whole transcriptome sequencing for evolutionarily representative plant species. BMC Genomics. 2022;23(S4):381. <https://doi.org/10.1186/s12864-022-08602-9>
- Zucchelli S, Cotella D, Takahashi H, Carrieri C, Cimatti L, Fasolo F, Jones MH, Sblattero D, Sanges R, Santoro C, et al. SINEUPs: a new class of natural and synthetic antisense long non-coding RNAs that activate translation. RNA Biol. 2015;12(8):771–779. <https://doi.org/10.1080/15476286.2015.1060395>