Published in final edited form as:

Methods Mol Biol. 2019 ; 1933: 1–30. doi:10.1007/978-1-4939-9045-0_1.

An Overview of Methodologies in Studying lncRNAs in the High-Throughput Era: When Acronyms ATTACK!

Hsiao-Lin V. Wang[1,2], **Julia A. Chekanova**[3]

[1]Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning, Guangxi, China.

[2]Present address: Department of Biology, Emory University, Atlanta, GA, USA.

[3]Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning, Guangxi, China.

Abstract

The discovery of pervasive transcription in eukaryotic genomes provided one of many surprising (and perhaps most surprising) findings of the genomic era and led to the uncovering of a large number of previously unstudied transcriptional events. This pervasive transcription leads to the production of large numbers of noncoding RNAs (ncRNAs) and thus opened the window to study these diverse, abundant transcripts of unclear relevance and unknown function. Since that discovery, recent advances in high-throughput sequencing technologies have identified a large collection of ncRNAs, from microRNAs to long noncoding RNAs (lncRNAs). Subsequent discoveries have shown that many lncRNAs play important roles in various eukaryotic processes; these discoveries have profoundly altered our understanding of the regulation of eukaryotic gene expression. Although the identification of ncRNAs has become a standard experimental approach, the functional characterization of these diverse ncRNAs remains a major challenge. In this chapter, we highlight recent progress in the methods to identify lncRNAs and the techniques to study the molecular function of these lncRNAs and the application of these techniques to the study of plant lncRNAs.

Keywords

High-throughput methods; RNA methods; Noncoding RNAs; lncRNAs; Plant lncRNAs; RNA secondary structures; RNA interactions

1 Introduction

Recent studies using high-throughput technologies have identified increasing numbers of lncRNAs in various eukaryotic transcriptomes. Functional studies have shown that some of these lncRNAs have diverse and important functions [1–3] in gene silencing and imprinting, transcription, mRNA splicing, translation, trafficking of nuclear factors, genome rearrangements, and regulation of chromatin modifications. In plants, lncRNAs are involved in the regulation of flowering, root development, plant immunity, responses to biotic and abiotic stresses, and many other important biological processes [1,2].

The detection of lncRNA transcripts has become easier, particularly due to the recent development and improvement of high-throughput sequencing technologies, but the

functions of most lncRNA remain largely unknown. Therefore, how to decipher the functions of lncRNAs has become an important topic in genome research. In this chapter, we provide an overview of methods for the identification and functional characterization of lncRNAs and focus on how these techniques could be adapted to study plant lncRNAs. Following this Introduction (Subheading 1), this chapter has two parts: one focuses on the identification of lncRNAs (Subheading 2), and the other focuses on analysis of the biological and molecular functions of lncRNAs (Subheading 3). In Subheading 2, we focus on methods to identify different populations of lncRNAs, revolving around different adaptations of RNA-seq methodology. We include some tag-based methods that were developed before the next-generation sequencing (NGS) era and present them through a historical lens. The details of how RNA-seq and high-throughput sequencing can be applied in a wide range of applications, including lncRNAs, have been extensively summarized by Wang and Snyder [4] and Reuter et al. [5]. As the functions of most lncRNAs largely remain to be elucidated, in Subheading 3 we present several selected methods to address various aspects of lncRNA functions in a high-throughput manner. The functional aspects addressed include tissue or cell type-specific analysis and examination of RNA-protein interactions, RNA-DNA/chromatin interactions, RNA-RNA interactions, RNA secondary structures, as well as RNA modifications. In addition to providing an overview of selected methods that are available to study lncRNAs, we detail the topics covered in this volume of *Methods in Molecular Biology, Plant Long Non-Coding RNAs: Methods and Protocols.* Another aim of this chapter is to give the reader links to information on specific technologies that were beyond the scope of this book or that have not yet been used in plants.

With the massive amount of data generated in each high-throughput sequencing experiment, data analysis has become a crucially important subject. Therefore, several chapters in this book provide step-by-step protocols for analyzing the large-scale sequencing data produced in the high-throughput experiments. However, a general summary of data analysis is a massive topic that requires separate, dedicated reviews and is thus outside of the scope of this chapter; therefore, we direct the reader to additional reviews and resources.

RNA biology is a vast, fast-moving field with myriad methods to study RNAs that are being improved and spawning variants as fast as discoveries can be disseminated to the research community; therefore this chapter only scratches the surface of the methodologies available to study lncRNAs.

2 Identification of lncRNAs Using High-Throughput Methodologies

The identification of lncRNAs is the first step in elucidating the role of lncRNAs in plants, and in recent years, most studies on plant ncRNAs have focused on identification of plant lncRNAs [1, 2]. Many of these lncRNAs are curated in various lncRNA databases, which we summarize in Table 1. The easy-to-follow protocols for how to use the three plant lncRNA databases, GreeNC [9], CANTATAdb 2.0 [11], and EVLncRNAs [13], are described in Part VI of this book, Chapters 25–27, respectively. lncRNAs are largely tissue specific and typically have a relatively low expression level; therefore, choosing the appropriate experimental techniques to identify and study lncRNAs is extremely important. In addition

to the identification of novel lncRNAs, the methods described below can be used to examine the expression levels of known lncRNAs.

Before the high-throughput era, RNAs were traditionally detected using Northern blotting analysis, nuclease protection assays (NPA), in situ hybridization, reverse transcription-polymerase chain reaction (RT-PCR), etc. Although most of these methods are not used in the initial steps of identifying lncRNAs genome-wide, they are often employed to validate the expression of lncRNAs and to examine them in the context of specific molecular functions. In this section of the chapter, we first provide an overview of selected high-throughput methods for identifying lncRNAs; these are summarized in Table 2 and described in Subheadings 2.1–2.6. We include widely used high-throughput sequencing techniques (Subheadings 2.3–2.6), hybridization-based approaches, and tag-based methods, which were developed before the high- throughput era (Subheadings 2.1 and 2.2, respectively).

2.1 Hybridization-Based Approaches

Before the emergence of high-throughput sequencing technologies, hybridization-based methods, including custom-designed and high-density oligo microarrays and genomic tiling microarrays, have been developed to analyze the transcriptome quantitatively [15–20]. In these approaches, cDNAs produced from a population of RNAs are hybridized to microarrays of tiled oligonucleotides that cover the non-repetitive sequences of the target genome at a very high resolution. Since the cDNAs and tiled oligonucleotides are labeled with different fluorophores, the relative abundance of RNAs can be inferred from the differences in fluorescent signal produced upon hybridization. For example, in *Arabidopsis thaliana,* the Affymetrix ATH 1.0F arrays and 100 ATH 1.0R arrays have been used to determine the transcriptional activity of the *Arabidopsis* genome and identify ncRNAs [20–25, 70].

Although hybridization-based approaches are relatively inexpensive and high-throughput, they have several limitations [4]. These limitations include reliance on the coverage and density of probes, sufficient knowledge of genome sequence and gene annotations, high background noise due to cross-hybridization, etc. Many of these limitations have made microarray analysis unsuitable for non-model plant species. However, despite these limitations, microarrays with probes representing already identified lncRNAs are now widely used to detect lncRNA expression with high sensitivity in many organisms, including plants. For example, Liu et al. used a custom microarray with 60-mer oligonucleotide probes for *Arabidopsis thaliana* long intergenic ncRNAs (lincR-NAs; ATH lincRNAvl array) to verify the expression of identified lincRNAs and to facilitate detection of lncRNA in different tissues, in response to biotic stresses, and in various mutants [70].

2.2 Sequence Tag-Based Approaches

Other large-scale methodologies for quantitatively analyzing expression of RNAs involve the production of very short sequence tags from the cDNAs derived from a given RNA sample. These short sequence tags are then sequenced using various platforms. The abundance of individual sequence tags corresponding to specific transcripts determines the relative

abundance of each transcript. Unlike microarray probes, which must be preselected from known sequences, sequence tags are discovered by random sequencing; therefore, this approach allows researchers to find novel RNA sequences. For example, expressed sequence tags (ESTs) are a collection of short subsequences derived from pools of cDNAs. ESTs can be used to examine gene expression [71], but EST-based approaches are low throughput, costly, and nonquantitative.

Other tag-based methods have overcome these limitations [72]. These new methods include serial analysis of gene expression (SAGE) [26], cap analysis of gene expression (CAGE) [36, 37,73], and massively parallel signature sequencing (MPSS) [31, 74, 75], which are described in Subheadings 2.2.1–2.2.3. These high-throughput tag-based approaches provide precise transcript levels. However, many of the short tags do not map uniquely to the reference genome. Moreover, these methods analyze only a small segment of each transcript and cannot distinguish transcript isoforms, which limits their use in studying the dynamic structures of many transcripts.

2.2.1 Serial Analysis of Gene Expression (SAGE)—SAGE was one of the first tag-based methods for high-throughput analysis of transcriptomes [26]. SAGE uses short sequence tags of cDNAs made from all the polyadenylated RNAs in a given sample. Each RNA is first converted into biotinylated cDNAs, which are captured on streptavidin beads. A few rounds of restriction enzyme digestions, ligation, and PCR result in a collection of short sequence tags representing each of the RNAs in the sample. The tag length must allow the tags to be mapped to the genes that they represent in the reference genome. Although, in theory, a short sequence tag of 9–10 nucleotides could be enough to identify individual transcripts, there is still the possibility that multiple genes could have the same tags. In practice, SAGE generally uses tags of 14–20 bp; the superSAGE variant uses tags of about 26 bp.

After the digested cDNAs are released from the beads, the tags are concatenated so that they can be cloned and sequenced in large groups. Counting the occurrences of each tag in the sequence data will give relative RNA expression levels. Because the SAGE technique maps the tags to a reference genome to identify genes, it works best in organisms that have a complete genome sequence. SAGE and superSAGE have been used in different plant species, including *Arabidopsis*, wheat, and chickpea, to analyze and detect existing transcripts and novel ncRNAs [27–30]. However, SAGE has been largely replaced by NGS technologies, which can examine more transcripts in greater depth. In addition, NGS methods generally skip the concatenation of tags, which SAGE uses to improve yields in Sanger sequencing.

2.2.2 Massively Parallel Signature Sequencing (MPSS)—Another sequence tag-based expression technique, massively parallel signature sequencing (MPSS), was developed to quantitatively analyze gene expression [31]. MPSS involves the acquisition of 17–20-nt tags (signatures) from cDNAs cloned on beads, using an unconventional, massively parallel sequencing method. MPSS uses a unique cloning strategy where every mRNA (and the corresponding cDNA) in a sample is represented by a single microbead; these microbeads are analyzed in a flow cell setup in an array format containing thousands of beads. The bases

of mRNAs are systematically removed after the sequencer reads the mRNA bases by hybridization to a labeled coder. This produces a collection of 17–20 bp signature tags representing each of the mRNAs in the sample. Work in multiple plants, including *Arabidopsis,* has used MPSS for analysis of the transcriptome [32–35], but MPSS has also been largely replaced by NGS.

2.2.3 Cap Analysis of Gene Expression (CAGE)—In contrast to other tag-based sequencing methodologies, like SAGE and MPSS, which largely depend on 3′ end of the RNA transcript, cap analysis of gene expression (CAGE or 5′-SAGE) is designed to capture the expression of 5′-capped RNAs quantitatively by using sequence tags from the 5′ ends of cDNAs [36, 37, 76]. The original CAGE method used the biotinylated CAP-trapper method [76], in which the cap structure of capped and poly-adenylated RNAs was chemically biotinylated. CAGE has been also used to analyze full-length cDNAs in *Arabidopsis* [38]. One of the advantages of CAGE is that it allows effective detection of the transcriptional activity around the promoter regions and RNA polymerase II-driven transcription start sites. However, the major limitation of CAGE is that non-capped RNAs are not detected.

In addition to the original tag-based CAGE, the same methodology now can be coupled with high-throughput sequencing (CAGE-seq) to examine the 5′-capped RNAs in a high-throughput manner not limiting to mRNAs [41,42]. Several plant studies have used the high-throughput sequencing-based CAGE or nano-CAGE to analyze transcriptional activity around transcription start sites [39, 40]. Additionally, paired-end analysis of transcription start sites (PEAT) is another approach that has been developed to capture 5′-capped RNAs [77]. Morton et al. successfully analyzed the transcriptional activity around transcription start sites using PEAT in wild-type Columbia-0 *Arabidopsis thaliana* whole root tissues [78].

2.3 RNA Sequencing (RNA-seq)

RNA-seq is currently widely used for the detection of RNA expression and for the discovery of novel lncRNAs [4]. In addition, RNA-seq can be used to find alternatively spliced mRNAs and splice junctions [79], as well as different isoforms. For RNA-seq, transcripts are reverse-transcribed into a pool of cDNAs that are cloned into a library for sequencing. The first such libraries were reverse-transcribed with oligo (dT) primers, thus capturing mostly polyadenylated RNAs and only a few non-polyadenylated RNAs, particularly rRNA. Oligo (dT) priming also excludes non-polyadenylated transcripts and many transcripts from the degradome. Therefore, most RNA-seq libraries are now reverse-transcribed with random primers, using a pool of RNAs that has been depleted of rRNA.

RNA-seq typically produces 30–400 bp reads, depending on the platform and methods used. The high-throughput sequencing platforms include Illumina, ABI’s SOLiD, Life Technologies/ThermoFisher/Ion Torrent, Oxford Nanopore Technologies, and Pacific Biosciences, as well as the recently retired Roche 454 sequencing platform. These various high-throughput sequencing platforms are discussed in detail by Reuter et al. [5]. Most platforms fragment RNA molecules (which generates a population of short sequences) and use short read-based technique, but Pacific Biosciences’s SMRT sequencing and Oxford Nanopore sequencing use single-molecule-based sequencing technology, with an average

read length of >14 kb and individual reads as long as 60 kb for Pacific Biosciences's SMRT sequencing and a medium read length of 6 kb and maximum of >60 kb for Oxford Nanopore sequencing. Although both of these single-molecule-based sequencing technologies can provide longer reads compared to other sequencing plat-forms, their high error rates are one of the most cumbersome technical problems.

A large number of studies have conducted RNA-seq to identify and categorize lncRNAs in plants and metazoans. Several of the chapters in this book provide detailed protocols for analyzing lncRNAs using RNA-seq followed by extensive bioinformatic and functional analysis (Chapters 11–16, all chapters in Part IV of the book: identification and functional analysis of lncRNAs). Specifically, Chapter 2 from the Pikaard lab provides a comprehensive protocol for how to analyze the ncRNAs that are produced by RNA polymerase IV; these lncRNAs serve as precursors for small interfering RNAs (siRNAs) in the RNA-directed DNA methylation pathway. Additionally, Chapters 3 and 4 provide protocols for how to use RNA-seq to identify differentially expressed lncRNAs during development. Chapters 9 and 10 provide protocols for how to use RNA-seq to identify lncRNAs produced in response to biotic and abiotic stresses, including drought and salt tolerances *(see* Chapter 9) and virus infection (*see* Chapter 10). Additionally, Chapter 8, by Matsui and Seki, provides a comprehensive review on the subject of lncRNAs and stress responses in plants.

2.4 Parallel Analysis of RNA Ends (PARE)/Genome-Wide Mapping of Uncapped and Cleaved Transcripts (GMUCT)/Degradome-Seq

Like all RNAs, lncRNAs are eventually degraded; moreover, some lncRNAs serve as targets for miRNAs, and degradation of some lncRNAs yields miRNAs [80]. The intense interest in the RNA interference pathway (RNAi) in the plant field has led to the development of techniques to examine transcripts that are in the process of being degraded or could be miRNA targets and precursors. These techniques include parallel analysis of RNA ends (PARE) [44] (protocol, [45]), degradome sequencing [48] (protocol, [49]), and genome-wide mapping of uncapped and cleaved transcripts (GMUCT) [46] (protocol, [47]). These three nearly identical techniques generate equivalent data and were developed using *Arabidopsis thaliana* and used in plants. These approaches all target RNA degradation products that have been uncapped at their 5′ end; these RNAs are ligated to an RNA adapter that allows them to be converted to cDNAs, which are then amplified by PCR and sequenced.

2.5 Transcript Isoform Sequencing (TIF-seq)

The techniques described above examine the 3′ or 5′ ends of transcripts. By contrast, TIF-seq examines both ends of transcripts [50, 51], thus enabling genome-wide assessment of transcripts based on the precise positions of their 5′ and 3′ ends. TIF-seq was originally designed to study transcriptional heterogeneity and unique transcript isoforms in *Saccharomyces cerevisiae* and relies on the usage of oligo-capping, which identifies the 5′ cap structure to allow for ligation of an oligo tag at the 5′ end. After oligo-capping, the resulting RNA molecules containing both 5′ cap structure and poly(A) tail undergo reverse transcription, generating full-length cDNAs with barcodes at both 5′ and 3′ ends. The barcoded cDNAs undergo intramolecular circularization, which allows sequencing of the

junction of the 5′ and 3′ ends of the transcript. In contrast to approaches that target only the 3′ or 5′ ends of transcripts, pinpointing the 3′ and 5′ ends of transcripts by TIF-seq allows the researcher to distinguish full-length and truncated transcripts, transcription through multiple open reading frames (bicistronic messages), and transcripts that originate from different start sites or terminate at different end sites. However, TIF-seq has not yet been used in plants nor in eukaryotes other than *S. cerevisiae.*

2.6 Nascent RNA Sequencing

In many cases, it is important to detect nascent transcription and nascent transcripts to capture the RNAs that are in the process of being transcribed. However, nascent transcripts can be unstable and difficult to distinguish from degraded or complete transcripts. The abundance of RNA polymerase II (Pol II) is often utilized to determine the level of nascent transcription at particular genomic locus. For example, chromatin immunoprecipitation microarray (ChIP-ChIP) or ChIP sequencing (ChIP-seq) methods are typically used to immunoprecipitate Pol II and associated chromatin. However, IP of Pol II collects paused Pol II and active Pol II-RNA complexes [81]. On the other hand, simply sequencing total RNA by RNA-seq or CAGE-seq detects the pool of steady-state RNAs and is also inefficient for detecting unstable nascent RNAs.

Several methods were designed to capture nascent RNAs that are associated with Pol II; rather than immunoprecipitation, these methods rely on “run-on” extension of nascent transcripts. In nuclear run-on experiments, cells are treated to halt transcription in vivo; reinitiation of transcription in isolated nuclei supplied with labeled RNA precursors (often 5 ′-bromo-uridine, BrU) labels only the nascent RNAs. These nuclear run-on assays include generic run-on assays and global run-on sequencing assay (GRO-seq) [52, 82], precision nuclear run-on sequencing (PRO-seq) [57], and native elongating transcript sequencing (NET-seq) [59–61]. Additional methods, like BRIC-Seq/BrU-Seq/BrU-Chase-Seq, were also developed to capture nascent transcripts [65, 68]. Although each method was designed with the similar goal of capturing actively transcribed RNA Pol II transcripts and nascent RNAs, they differ in technical details and have specific limitations, as described in the subsections below. Moreover, only GRO-seq has been used in plants [53, 54].

2.6.1 Global Run-On Sequencing (GRO-Seq)—Nuclear run-on assays and global run-on sequencing (GRO-seq) were developed to capture nascent RNAs and to measure RNA half-life [52]. GRO-seq reveals Pol Il-engaged transcripts genome-wide, with high resolution and specific information on the orientation and exact 5′ end of the transcript. In GRO-seq, nuclear run-on assays use BrU as the label and release paused Pol II with sarkosyl, to label only transcripts from engaged polymerases. The BrU-labeled transcripts are purified with anti-Br-UTP antibodies and deep sequenced. This very sensitive and specific method gives high-throughput, genome-wide data on nascent transcripts. However, purification of nuclei, reinitiation of transcription under non-physiological conditions, and precipitation of labeled RNAs have proven difficult. GRO-seq also has a limited resolution of 30–50 bases due to the necessity to allow polymerase to run on and incorporate labeled BrU into RNAs.

Recently, GRO-seq and 5´GRO-seq (also called GRO-cap, see description below), which use a 7-methylguanylate (m7G) cap, have been used to capture the characteristics of the nascent transcriptome in *Arabidopsis thaliana* seedlings [53] and in maize [54]. Moreover, protocols for GRO-seq have been described in multiple publications (e.g., *see* [55, 56]), and this technique will likely see a wider application in the plant studies in the future.

2.6.2 Precision Nuclear Run-On Sequencing (PRO-Seq)—Similar to GRO-seq, precision nuclear run-on sequencing (PRO-seq) was developed to examine Pol II that is actively engaged in transcription at high resolution and on a genome-wide scale [57] (protocol, [58]). However, in contrast to GRO-seq, PRO-seq can reveal the mapping and distribution ofPol II pausing at single-base resolution. Similar to traditional Sanger sequencing, PRO-seq uses chain-terminating ribonucleotide triphosphate analogs labeled with biotin (biotin-NTPs, either all four, or one with additional unlabeled NTPs) for run-on assays. The nascent RNAs can be purified using the biotin label and used for high-throughput sequencing.

PRO-seq can be modified to capture 5´ capped RNAs, a method termed PRO-cap [57]. In PRO-cap, uncapped RNAs are first removed, leaving only the pool of capped RNAs. The cap of each RNA is then modified to allow the ligation of adapters to the 5’ end. Therefore, PRO-cap allows identification of the transcription start sites at the RNA synthesis level. PRO-seq and PRO-cap can be coupled to compare the differences in the Pol II initiation and pause sites. However, PRO-seq has not been used in plants yet. As a nuclear run-on-based methodology, PRO-seq has the same technical difficulties as GRO-seq. Moreover, PRO-seq only identifies Pol II complexes that are competent to elongate nascent transcripts; it cannot map complexes that are backtracking or arrested.

2.6.3 Native Elongating Transcript Sequencing (NET-Seq)—NET-seq or mammalian NET-seq (mNET-seq) detects nascent, actively transcribed Pol II RNAs, through the capture of 3´ RNAs [59–61] (protocol, [62–64]). In this method, the affinity-tagged Pol II elongation complex is immunoprecipitated, and then coprecipitated RNA is extracted and reverse-transcribed into cDNA. Deep sequencing of the cDNAs produces 3´-end sequences of nascent RNA, providing nucleotide resolution mapping of transcripts. This immunoprecipitation-based method captures elongating complexes and complexes that are backtracked or arrested, an advantage, depending on the goal of the experiment, compared to GRO-seq and PRO-seq. However, immunoprecipitation requires that Pol II complexes be solubilized, which can be challenging in metazoan cells where they are typically insoluble and strongly associated with chromatin under native conditions. NET-seq and mNET-seq have been used in *Saccharomyces cerevisiae* and HeLa cells, but the NET-seq protocol has not been used in plants yet.

2.6.4 BRIC-Seq/BrU-Seq/BrUChase-Seq—In addition to examining nascent RNAs, other methods can measure the half-lives of mRNAs or lncRNAs, which can inform analysis of their physiological functions and regulation. In organisms with established cell cultures, endogenous transcripts can be pulse-labeled by adding BrU to the culture media. In different variants of the classic pulse-chase method, label can be added for different times and removed from the media; labeled RNA can be immuno-precipitated and sequenced. For

example, to establish the half-lives of RNAs or lncRNAs, in 5′-BrU immunoprecipitation chase-deep sequencing analysis (BRIC-seq) [65] (protocol, [66, 67]), total RNAs containing BrU-labeled RNAs (BrU-RNAs) are isolated at sequential time intervals after removal of BrU from the culture medium. BrU-RNAs are then recovered by immunopurification, which is followed by RT-qPCR or deep sequencing.

BrU labeling and sequencing (BrU-seq) and BrU pulse-chase sequencing (BrUChase-Seq) also involve BrU pulse-labeling that is chased with uridine, giving pools of RNA of different ages [68] (protocol, [69]). Following immunocapture, the BrU-labeled RNA is deep sequenced. However, none of the BRIC-Seq/BrU-Seq/BrUChase-Seq protocols have been used in plants yet.

3 Analyzing the Biological and Molecular Functions of lncRNAs

After the identification of lncRNAs using high-throughput approaches, one next step would be determining if these lncRNAs have biological functions, followed by identification of these functions. However, these experimental approaches have proven challenging, particularly for high-throughput studies, because of the diverse functions of lncRNAs, their potential tissue and stage specificity, and the varied mechanisms by which lncRNAs achieve these functions. For example, as one approach, overexpression or knock-down of the target lncRNAs can be used to study the functions of lncRNAs; however, such approaches are difficult to conduct in a high-throughput fashion in multicellular organisms.

Despite these challenges, new methods are emerging that can examine lncRNA function in a high-throughput manner. For example, in addition to modulating gene expression, lncRNAs have been implicated in genome architecture. The local interaction and looping events of the genomic regions where ncRNAs originated from can be captured by chromosome conformation capture followed by massively parallel sequencing [83, 84]. In this book, Chapter 28 by Padmarasu et al. describes an improved method to detect long-range chromatin interactions using in situ Hi-C for plants. Below, we present several selected methods that can be used to analyze the functions of lncRNAs in a high-throughput manner; these techniques are summarized in Table 3 (Subheadings 3.1–3.6).

3.1 Tissue or Cell Type-Specific Analysis

In multicellular organisms, specialized cell types each have a specific phenotype, function, and transcriptional program. However, our knowledge of how cells implement these programs during differentiation, particularly the effects of lncRNAs, remains limited, in part because purifying individual cell types for transcriptional and epigenomic profiling remains challenging. However, ongoing research has developed multiple methods to study lncRNAs in specific plant cell types by purifying individual cell types for analysis. These methods include laser microdissection (LM or laser capture microdissection, LCM) of fixed tissue sections [85], fluorescence-activated cell sorting (FACS) of fluorescently labeled cells or nuclei [89], and isolation of nuclei tagged in specific cell types (INTACT) using affinity-based isolation [92]. All three of these methods are commonly used in different plant species, and additional information is provided below (Subheadings 3.1.1–3.1.3). Three

chapters in the Part II of this book (Chapters 5–7) also provide different protocols in studying the tissue and cell type-specific lncRNAs.

In addition to these methods, cryosectioning and cryostat sectioning can be used to study specific cell types [143, 144]; this is often less invasive compared to other techniques. Cryosectioning is also the first step of LM (or LCM) for sample preparation and can be coupled with other cell type-specific or tissue-specific techniques to isolate and study plant lncRNAs [145, 146]. Chapter 3 in this book by Kim et al. describes a protocol that uses cryostat sectioning to isolate distinct tissue types in the developing endosperm in maize followed by transcriptome and epigenome analysis to identify lncRNAs. Other methods like ex vivo differentiation from progenitor cells and the use of cultured cell lines are commonly used in metazoan studies; however, cultured cell lines are not commonly used in plant studies.

3.1.1 Laser Microdissection (LM)—Laser microdissection (LM; laser-captured microdissection, LCM; laser-assisted microdissection, LMD or LAM) uses a laser beam and direct microscopic visualization to isolate specific cells from heterogeneous tissues [85–87]. When coupled with high-throughput sequencing or microarray analysis, LM allows genome-wide analysis of gene expression in specific cell types. The detailed methodology and technological requirements of LM are comprehensively discussed in the review by Bevilacqua and Ducos [147]. LM has been used in separation of specific cell types in *Arabidopsis* [148], maize [149, 150], and other plants. The recent advances and applications of LM in the context of plant biology and transcrip-tome studies were comprehensively reviewed by Gautam and Sarkar [88]. Chapter 5 in this book by Gautam et al. describes a protocol that adapts the LM to obtain high-quality RNA of low abundance from specific tissues, followed by RT-PCR or stem-loop RT-PCR.

3.1.2 Fluorescence-Activated Cell Sorting (FACS)—Fluorescence-activated cell sorting (FACS) to separate cells or nuclei into different populations is based on the green fluorescent protein (GFP) labeling of specific cells, which are then separated from unlabeled cells using flow cytometry [89, 90]. FACS is followed by RNA extraction from each subpopulation of cells and high-throughput sequencing or microarray analysis. Based on the use of enhancer trap lines or promoter-GFP fusions that express GFP in specific tissues, these techniques have been widely used in plants. This has allowed deep analysis of RNA expression and the transcriptome in distinct cell types, cells in different developmental stages, and cells in response to biotic or abiotic stresses (reviewed by Carter et al. [91]). For example, recently cell type expression analyses in *Arabidopsis* roots were used to characterize intergenic lncRNAs [151].

3.1.3 Isolation of Nuclei Tagged in Specific Cell Types (INTACT)—LM and FACS require specialized equipment and the manipulation of whole cells; by contrast, the isolation of transgenically tagged nuclei in specific cell types (INTACT) uses affinity-based methods to isolate tagged nuclei from total nuclei. INTACT does not require the dissociation and manipulation of whole cells [92] (protocol, [93, 94]). The INTACT method was initially developed to study cell types in the *Arabidopsis thaliana* root epidermis with high yield and efficiency. Although it has its advantages, in order to successfully obtain transgenically

tagged nuclei, it requires a promoter or enhancer trap line that is expressed in the specific cell type to be examined. Chapter 7 in this book by Do et al. describes a protocol that adapts the INTACT methodology to isolate specific cell types with tagged nuclei, followed by RNA-seq and bioinformatic analysis to identify nuclear lncRNAs in *Arabidopsis*.

3.2 In Situ Hybridization (ISH) and Fluorescence In Situ Hybridization (FISH)

FISH can be used to visualize the subcellular localization of lncRNAs and possibly provide information on their potential functions [95]. DNA and RNA can be visualized in situ using DNA FISH and RNA FISH, respectively, and multiplex FISH can simultaneously assay multiple targets within the same specimen. FISH techniques have been used for decades; however, emerging work has brought FISH into the genomics era. For example, the fluorescent in situ RNA sequencing (FISSEQ) amplifies cDNAs in cells and tissues [96]; compared with RNA-FISH, FISSEQ gives a higher resolution and can identify more targets. FISSEQ produces fewer reads than regular RNA-seq, but could provide cell-specific spatial information on lncRNAs. Chapter 7 in this book by Francoz et al. describes a protocol that integrates ISH and transcriptomics resulting in a medium-throughput RNA in situ hybridization methodology.

3.3 RNA-Protein Interactions

Many lncRNAs function in complexes with proteins, but very few of the proteins that interact with lncRNAs have been identified. Identification of the protein interactors of lncRNAs will shed substantial light on the mechanisms of lncRNA function. Below, in Subheadings 3.3.1–3.3.3, we describe selected methods to analyze RNA-protein interactions. Also, three chapters in this book, from the Chua lab, provide three distinct and comprehensive protocols for identification and analysis of lncRNAs and protein interactions *(see* Chapters 17–19 in Part V of this book). Additional techniques that are not presented here are comprehensively summarized in the review by Ferre et al. [152].

3.3.1 RNA Immunoprecipitation Followed by Sequencing (RIP-Seq)—The versatile technique of RNA immunoprecipitation (RIP)-seq can examine multiple aspects of RNA-protein interactions, from either the RNA or the protein side. If the interacting protein is known, then antibodies against the target (or the affinity-tagged targets) can be used for RIP of RNAs that interact with the protein of interest [153]. RIP can be coupled with microarray (RIP-Chip) or high-throughput sequencing (RIP-seq) to identify the RNAs that interact with proteins genome-wide [97, 100]. Additionally, RIP can be used to identify the binding sites for specific proteins. RIP and RIP-seq have been widely used in metazoans and plants. For example, in *Arabidopsis,* RIP-seq was used to identify the transcriptome-wide RNA targets of SR34, a serine/arginine-rich (SR)-like RNA-binding protein that functions in constitutive and alternative splicing [98]. RIP was also used to identify Argonaute (AGO)-associated smRNAs (RIP smRNA-seq) [99] or RNAs (AGO RIP-seq) [101] in the RNA interference (RNAi) pathway.

RIP can also be used to identify regions in the RNA molecule that interact with proteins. Indeed, the first studies used RIP to find proteins that interact with the lncRNA Xist, which functions in X chromosome inactivation in mammals [154]. Usually, chemical agents are

used to cross-link RNAs and proteins, but this can introduce artifacts. RIP can also be conducted without cross-linking, thus reducing the potential generation of artifacts. Moreover, various nuclease treatments can provide additional information on protein-nucleic acid interactions. For example, RNase H digests the RNA in RNA-DNA hybrids, while DNase I digests DNA, and the combination of the treatments with these nucleases can help distinguish the indirect binding of protein of interest to neighboring DNA from direct binding between protein of interest and RNAs. Different nuclease treatments can also distinguish protein-RNA interactions that involve single-stranded or stem- loop RNAs.

The modifications of RNAs play an important role in regulating the functions of RNA molecules, and RIP-seq can be adopted to map RNA modifications genome-wide, such as mapping 5-methyl-cytosine (m^5C) or *N*6-methyladenosine (m^5A) of RNAs (m^5C-RIP-seq, m^5A-seq, respectively) [102, 155, 156]. Chapter 24 in this book by Liang and Gu describes a protocol for m^5C-RIP-seq to map m^5C RNA modifications in plants genome-wide. There are additional techniques for identifying RNA modification sites of both mRNAs and lncRNAs transcriptome-wide, such as coupling RNA bisulfite conversion with sequencing (bsRNA-seq) [157,158] and 5-azacytidine-mediated RNA immunoprecipitation (Aza-IP) [159].

3.3.2 High-Throughput Sequencing Cross-Linking Immunoprecipitation (HITSCLIP or CLIP-seq)—

CLIP (cross-linking immunoprecipitation) examines RNA-protein interactions by UV cross-linking cells before immunoprecipitation [160] (protocol, [161]). CLIP-seq, also known as HITS-CLIP, is a method for genome-wide mapping of RNA-protein binding sites, by CLIP, followed by high-throughput sequencing of the RNA [103, 104]. In contrast to chromatin immunoprecipitation sequencing (ChIP-seq), which uses formaldehyde cross-linking, in CLIP-Seq UV cross-linking covalently links the RNA and protein. The pools of cross-linked and immunoprecipitated RNA molecules are first fragmented with RNase followed by proteinase digestion and purification. One of the main advantages of this method is that it identifies the essential protein-binding sites on the RNA molecule. However, the UV light can cause mutations and CLIP-Seq does not give full-length sequence of the immunoprecipitated RNA. These disadvantages can be particularly problematic for systems that lack collections of full-length, annotated lncRNA sequences and for lncRNAs that are present at very low levels.

CLIP-Seq/HITS-CLIP has been widely used in mammalian studies, including identifying genome-wide interactions of RNAs and the neuron-specific splicing factor Nova in mouse brains [103]; however, no study has used CLIP-Seq in plants so far. Several other methods have been developed to improve the efficiency of CLIP-Seq, including enhanced CLIP (eCLIP) and infrared-CLIP [105–107]. Protocols for CLIP-seq are described in these references [108, 109].

3.3.3 Photoactivatable Ribonucleotide-Enhanced Cross-Linking and Immunoprecipitation (PAR-CLIP)—

PAR-CLIP, a variant of CLIP-Seq, has better cross-linking efficiency and resolution, as well as a higher signal-to-noise ration compared with other methods [110, 112]. PAR-CLIP uses the ribonucleoside analogs, 4-thiouridine (4SU) and 6-thioguanosine (6SG). Photoactivation of 4SU and 6SG by UV light produces strong

cross-links and specific mutations of the nucleic acid sequence: 4SU produces T to C changes, and 6SG produces G to A. PAR-CLIP can therefore be used to identify the binding sites of RNA-binding proteins. Moreover, PAR-CLIP can be used to identify miRNA targets [113]. PAR-CLIP has been widely implemented in metazoan studies [111]; however, no study has used PAR-CLIP in plants so far. Protocols for PAR-CLIP are described in these references [112, 113].

3.4 RNA-DNA/Chromatin Interactions

lncRNAs can physically associate with chromatin, indirectly through an RNA-protein interaction [162] or directly through RNA-DNA hybridization in a triple helix [163, 164]. DNA-RNA FISH can show this association only at low resolution; new technologies can examine these lncRNA-DNA interactions at higher resolution. In Subheadings 3.4.1–3.4.3, we describe three recently developed methodologies for mapping the interactions of RNAs with chromatin in a high-throughput manner; however, none of these methods have been used in plants so far. These three techniques are very similar and detect the lncRNA by probing, differing only in some specifics.

3.4.1 Chromatin Isolation by RNA Purification Sequencing (ChIRP or ChIRP-Seq)—The techniques described above in Subheading 3.3 examine RNA-protein interactions to identify the RNAs that bind to a known protein. ChIRP can be used to identify the proteins and chromatin regions that are bound by a known RNA [114]. After cross-linking and sonication of chromatin, ChIRP uses tiled biotinylated oligonucleotides (20-mers) to affinity purify a known lncRNA in complex with its associated chromatin and proteins. The DNA genomic regions associated with the RNA of interest can be identified by sequencing (ChIRP-Seq), and the RNA can be quantified by qPCR. Chu et al. used ChIRP to identify DNA regions associated with the lncRNA HOTAIR. In addition to identifying associated DNA regions, RNA-associated proteins can be purified from ChIRP reactions and examined by mass spectrometry (ChIRP-MS) [115]. For example, ChIRP-MS identified 81 proteins associated with the Xist lncRNA, which plays key roles in X chromosome silencing in mammals. A comprehensive protocol, with video of ChIRP-seq, can be found in this article [116].

3.4.2 RNA Antisense Purification (RAP)—RNA antisense purification (RAP) or RNA antisense purification followed by DNA sequencing (RAP-DNA) can be used to find the chromatin regions that associate with a specific RNA [118]. In contrast to ChIRP, which uses the tiled biotinylated 20-nt oligo-nucleotides, RAP uses 120-nt-long antisense RNA probes, thus improving target lncRNA binding and increasing the signal-to-noise ratio. Like ChIRP, RAP uses tiled overlapping probes that cover target transcripts without considering whether the regions are accessible for hybridization. DNase I digestion then produces genomic DNA fragments of <300 bp; these fragments are then sequenced. RAP was used to find the exact X chromosome binding sites of Xist in mouse ES cells [118]. A detailed protocol for RAP-DNA can be found on the Guttman lab's website (http://www.lncrna-test.caltech.edu/protocols.php) and [119].

3.4.3 Capture Hybridization Analysis of RNA Targets (CHART)—Like ChIRP and RAP, CHART involves cross-linking and purification of RNA-DNA complexes [120] (protocol, [122]). However, CHART uses capture oligonucleotides (C-oligos) that carry a tag such as biotin for affinity purification. CHART probe design involves the detection of open binding sites by hybridization of oligonucleotides and RNase H mapping; then this information is used to design biotinylated 24-nt-long C-oligos that target the open sites. Similar to other methods, CHART can recover the chromatin and proteins associated with the target lncRNA. CHART has been used to study the function of roX2 lncRNA in *Drosophila* (lncRNA involved in dosage compensation) [120] and Xist lncRNA [121].

3.5 RNA-RNA Interactions

Interactions between RNAs involve direct base-pairing interactions (including miRNA-mRNA and mRNA-lncRNA sense-antisense interactions) and indirect interactions mediated by protein intermediates. Some lncRNAs associate with RNA-processing proteins, indicating that these lncRNAs may target other RNAs via proteins [165,166]. Below, in Subheadings 3.5.1 and 3.5.2, we present two recently developed methods for mapping RNA-RNA interactions for a RNA molecule of interest genome-wide. However, none of these methods have been used in plants so far.

3.5.1 RNA Antisense Purification Followed by RNA Sequencing (RAP-RNA)—RNA antisense purification (RAP) and variants thereof detect RNA-RNA interactions. For example, to identify the intermolecular contacts of lncRNA-RNA interactions, RAP-RNA, a method based on RAP, was developed to systematically map RNA-RNA interactions [123]. RAP-RNA cross-links RNAs in vivo and then uses antisense oligonucleotide to purify the RNA, followed by high-throughput RNA sequencing. Cross-linking reagents that differ in their specificity for proteins and nucleic acids can give additional information on the interactions. Additional variants, RAP-RNA$^{[AMT]}$, RAP-RNA$^{[FA]}$, and RAP-RNA$^{[FA\text{-}DSG]}$, can identify direct and indirect RNA-RNA interactions [123]. Each variant uses a different cross-linking agent, 4′-aminomethyltrioxsalen (AMT), formaldehyde (FA), and both FA and disuccinimidyl glutarate (DSG), respectively. These methodologies were applied to investigate two ncRNAs implicated in RNA processing: U1 small nuclear RNA, a component of the spliceosome, and Malat1, a large lncRNA that localizes to nuclear speckles [123]. This study revealed that U1 hybridizes to the 5′ splice sites of RNAs and Malat1 indirectly binds pre-mRNAs by interacting with proteins. A detailed protocol for RAP-RNA can be found on the Guttman lab's website (http://www.lncrna-test.caltech.edu/protocols.php).

3.5.2 Cross-Linking, Ligation, and Sequencing of Hybrids (CLASH)—CLASH uses UV cross-linking to capture direct RNA-RNA hybridization in RNA-protein complexes [124] (protocol, [126, 127]). UV light cross-links interacting nucleic acids and interacting RNA-protein complexes, and UV cross-linking offers advantages compared to chemical cross-linking. For examples, chemical cross-linking can also cross-link between interacting protein-protein complexes, making it hard to differentiate direct and indirect (i.e., protein-mediated) interactions. UV cross-linked RNA-protein complexes are affinity-purified for the protein of interest. RNA-RNA hybrids are ligated together to generate chimeric RNAs and

isolated, thus giving high-throughput data on RNA-RNA interactions. Despite the advantages of high-throughput data, ligation of the RNAs remains a challenging step. Work in yeast used CLASH to find novel snoRNA-rRNA interactions [124], and work in humans found miRNA-mRNA interactions in complexes with Argonaute [125]. This technique has a potential to be used to capture lncRNA-RNA interactions.

3.6 RNA Secondary Structures

Functional RNAs act via sequence complementarity with RNAs or DNA and via their structures, which involve intricate base pairing with multiple stems and loops. These RNA structures can give RNAs the ability to catalyze reactions, scaffold macromolecular complexes, and bind ligands. RNAs can thus affect epigenetic regulation; mRNA splicing, stability, and translation; and signal transduction. Therefore, understanding the structure of lncRNAs provides key information for understanding their functions [167–169].

Computational prediction of RNA secondary structure works well for smaller RNAs, but for lncRNAs, the number of possible structures increases exponentially. For experimental methods, recent studies have developed genome-wide techniques to examine RNA secondary structures [169]. Different methodologies have been extensively summarized in the review articles by Wan et al. [169], Vandivier et al. [167], and Bevilacqua et al. [168]. Below, in Subheadings 3.6.1–3.6.4, we present five high-throughput methods that have been (or are being) used to solve RNA secondary structures in plants or have the potential to be adapted in the plant fields. Some of these methods are also provided in Part V of this book in Chapters 20–22. Recently, the secondary structure of COOLAIR, an antisense RNAs involved in the regulation of *Flowering Locus C (FLC)* and vernalization, was determined using chemical probing across the Brassicaceae [170]; the methodology is also described in Chapter 23 of this book.

3.6.1 Selective 2´-Hydroxyl Acylation by Primer Extension Sequencing (SHAPE-Seq)—Selective 2´-hydroxyl acylation by primer extension (SHAPE) provides large-scale data on RNA secondary structure and could be used to determine the secondary structure of lncRNAs [169, 171]. SHAPE uses the electrophile N-methylisatoic anhydride (NMIA), which can attack the 2ÓHs of bases in flexible regions of the RNA, forming 2-O adducts, which can be detected because they terminate reverse transcription. SHAPE reverse transcription followed by capillary sequencing has been used to examine the structures of the 16S rRNA and the 9-kb HIV RNA genome [169, 172]. Recently, SHAPE has also been coupled to high-throughput sequencing (SHAPE-seq) to provide the secondary and tertiary structural information of seven RNAs in vitro [128] (protocol, [129, 130]).

3.6.2 Structure-Seq and Structure-Seq2—RNA structures, which are influenced by the cellular environment, often differ in vitro and in vivo [173]. Structure-seq2 and its predecessor, Structure-seq, are high-throughput methods that provide an efficient way to study RNA structures in vivo [131, 132]. They do this by using DMS (dimethyl suberimidate) to react with RNA in vivo, similar to DMS-seq [174] and Mod-seq [175], and in contrast to SHAPE. The Galaxy environment (https://usegalaxy.org/) provides open access to the user-friendly Structure-seq computational pipeline [133]. Structure-seq2 offers

improvements compared to its predecessor and can be used in additional plant species [132]. A comprehensive computational pipeline, StructureFold2, accompanies this new and improved method [134].

Comprehensive protocol for Structure-seq is described in this article [135]. Also, Chapter 20 in this book, by Ritchey et al. (the authors are the developers of this method), describes two versions of the Structure-seq2 protocol and provides detailed step-by-step instructions on how to produce and analyze Structure-seq2 data. This protocol has been used in plants and can be easily applied to other organisms.

3.6.3 Protein Interaction Profile Sequencing (PIP-Seq)—To simultaneously examine protein-RNA binding sites and RNA secondary structure globally and in an unbiased manner, protein interaction profile sequencing (PIP-seq) uses RNase-based foot-printing assays for RNA-associated proteins, which are cross-linked on the RNA [136,137]. RNase digestion then removes the regions of the RNA that are not bound by the protein, leaving the bound regions as the protein footprint. After reversal of the cross-linking, the bound regions are recovered by ligation to linkers, PCR amplification, and cloning as strand-specific libraries for sequencing. To identify RNase-resistant regions, which would appear as protein-bound regions, control reactions are conducted in which the proteins are denatured and partially digested before RNase treatment. Moreover, to examine RNA structure, PIP-seq uses RNases specific for double-and single-stranded RNA; comparison of the results reveals the secondary structure of the RNAs. Comprehensive protocols for PIP-seq can be found in the following articles [138, 176]. Also, two chapters in this book, Chapters 21 and 22, from the Gregory lab (the developer of this method) provide detailed step-by-step instructions on how to perform PIP-seq experiments and analyze the data in the context of plant transcriptomes.

3.6.4 Parallel Analysis of RNA Structure (PARS) and Fragmentation Sequencing (Frag-Seq)—RNA structures can be determined in vitro by parallel analysis of RNA structure (PARS) and fragmentation sequencing (Frag-seq) [139, 141]. PARS uses RNase V1 and S1 nucleases which are specific to double- or single-stranded regions of RNAs, respectively, and compares the results [139]. PARS uses genome-wide RNA structure probing and was validated on the known structure of the HOTAIR lncRNA. Frag-seq uses RNase P1, which is specific to single-stranded nucleic acids, to sequence RNA [141]. PARS and Frag-seq both use nucleases with specific activities on different structures, but differ in how they map RNA structure and thus provide complementary information. Structure prediction programs can be used to examine Frag-seq and PARS data to improve the prediction of RNA secondary structure. However, PARS and Frag-seq have not been used to examine RNA secondary structure in plants. Comprehensive protocols can be found here [140] for PARS and here for Fraq-seq [142].

4 Conclusions and Future Outlook

lncRNAs function at multiple levels in gene regulatory networks, where they modulate complex biological processes in eukaryotes. A large amount of effort has been spent studying lncRNAs in the metazoan field, effort that has translated into a better understanding

of lncRNA functions in animals. However, unlike protein-coding genes, functional RNA coding genes (i.e., tRNAs, rRNAs, etc.), and microRNAs, lncRNAs show little conservation between species, making it difficult to identify functional lncRNAs and use the findings from one species to inform the study of another. This is why it is so important to study plant lncRNAs; however, very few plant lncRNAs have been examined. It will be crucial to identify factors controlling the expression and biogenesis ofplants lncRNAs, as well as the protein and RNA components functioning with these lncRNAs. Together, this information will contribute to the much-needed understanding of plant lncRNA functions. Furthermore, these studies should be extended to a wide range of plant species.

Perturbation of lncRNA expression, in situ hybridization, structural characterization, and genomics are among the major tools available to dissect the molecular mechanisms of lncRNA functions. Revealing the regulatory roles of lncRNAs may require improvements in various techniques, as well as adoption of technologies from other fields to plants. These technologies include in vivo imaging of RNAs through single-molecule techniques, identifying the binding partners of lncRNAs by high-throughput methods, as well as using single-cell methods to decipher the heterogeneity of the transcriptome in a cellular population. We hope that the continuing interest in the biology of lncRNAs will bring new insights and discoveries to the functional mechanisms of lncRNAs in plants.

References

References

1. Chekanova JA (2015) Long non-coding RNAs and their functions in plants. Curr Opin Plant Biol 27:207–216 [PubMed: 26342908]
2. Wang H-LV, Chekanova JA (2017) Long Noncoding RNAs in Plants. Adv Exp Med Biol 1008:133–154 [PubMed: 28815539]
3. Quinn JJ, Chang HY (2016) Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17:47–62 [PubMed: 26666209]
4. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63 [PubMed: 19015660]
5. Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol Cell 58:586–597 [PubMed: 26000844]
6. Lamesch P, Berardini TZ, Li D et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40:D1202–D1210 [PubMed: 22140109]
7. Cheng C-Y, Krishnakumar V, Chan AP et al. (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J 89:789–804 [PubMed: 27862469]
8. Jin J, Liu J, Wang H et al. (2013) PLncDB: plant long non-coding RNA database. Bioin-formatics 29:1068–1071
9. Paytuví Gallart A, Hermoso Pulido A, Anzar Martínez de Lagran I et al. (2015) GREENC: a Wiki-based database of plant lncRNAs. Nucleic Acids Res 44(Database issue): D1161–D1166. 10.1093/nar/gkv1215 [PubMed: 26578586]
10. Zhao Y, Li H, Fang S et al. (2016) NON-CODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic Acids Res 44:D203–D208 [PubMed: 26586799]
11. Szcześniak MW, Rosikiewicz W, Makalowska I (2016) CANTATAdb: a collection of plant long non-coding RNAs. Plant Cell Physiol 57:e8–e8 [PubMed: 26657895]
12. Yi X, Zhang Z, Ling Y et al. (2015) PNRD: a plant non-coding RNA database. Nucleic Acids Res 43:D982–D989 [PubMed: 25398903]

13. Zhou B, Zhao H, Yu J et al. (2018) EVLncR- NAs: a manually curated database for long non-coding RNAs validated by low-throughput experiments. Nucleic Acids Res 46(D1):D100–D105 [PubMed: 28985416]
14. Chen D, Yuan C, Zhang J et al. (2012) Plant-NATsDB: a comprehensive database of plant natural antisense transcripts. Nucleic Acids Res 40:D1187–D1193 [PubMed: 22058132]
15. Kapranov P, Cawley SE, Drenkow J et al. (2002) Large-scale transcriptional activity in chromosomes 21 and 22. Science 296:916–919 [PubMed: 11988577]
16. Shoemaker DD, Schadt EE, Armour CD et al. (2001) Experimental annotation of the human genome using microarray technology. Nature 409:922–927 [PubMed: 11237012]
17. Yazaki J, Gregory BD, Ecker JR (2007) Mapping the genome landscape using tiling array technology. Curr Opin Plant Biol 10:534–542 [PubMed: 17703988]
18. Gregory BD, Yazaki J, Ecker JR (2008) Utilizing tiling microarrays for whole-genome analysis in plants. Plant J 53:636–644 [PubMed: 18269573]
19. Yamada K, Lim J, Dale JM et al. (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. Science 302:842–846 [PubMed: 14593172]
20. Chekanova JA, Gregory BD, Reverdatto SV et al. (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. Cell 131:1340–1353 [PubMed: 18160042]
21. Macintosh GC, Wilkerson C, Green PJ (2001) Identification and analysis of Arabidopsis expressed sequence tags characteristic of non-coding RNAs. Plant Physiol 127:765–776 [PubMed: 11706161]
22. Marker C, Zemann A, Terhorst T et al. (2002) Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana. Curr Biol 12:2002–2013 [PubMed: 12477388]
23. Rymarquis LA, Kastenmayer JP, Hüttenhofer AG et al. (2008) Diamonds in the rough: mRNA-like non-coding RNAs. Trends Plant Sci 13:329–334 [PubMed: 18448381]
24. Song D, Yang Y, Yu B et al. (2009) Computational prediction of novel non-coding RNAs in Arabidopsis thaliana. BMC Bioinformatics 10(Suppl 1):S36 [PubMed: 19208137]
25. Jouannet V, Crespi M (2011) Long nonprotein-coding RNAs in plants. Prog Mol Subcell Biol 51:179–200 [PubMed: 21287139]
26. Velculescu VE, Zhang L, Vogelstein B et al. (1995) Serial analysis of gene expression. Science 270:484–487 [PubMed: 7570003]
27. Robinson SJ, Cram DJ, Lewis CT et al. (2004) Maximizing the efficacy of SAGE analysis identifies novel transcripts in Arabidopsis. Plant Physiol 136:3223–3233 [PubMed: 15489285]
28. Robinson SJ, Parkin IAP (2008) Differential SAGE analysis in Arabidopsis uncovers increased transcriptome complexity in response to low temperature. BMC Genomics 9:434 [PubMed: 18808718]
29. Poole RL, Barker GLA, Werner K et al. (2008) Analysis of wheat SAGE tags reveals evidence for widespread antisense transcription. BMC Genomics 9:475 [PubMed: 18847483]
30. Molina C, Rotter B, Horres R et al. (2008) SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics 9:553 [PubMed: 19025623]
31. Brenner S, Johnson M, Bridgham J et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630–634 [PubMed: 10835600]
32. Meyers BC, Vu TH, Tej SS et al. (2004) Analysis of the transcriptional complexity of Arabi-dopsis thaliana by massively parallel signature sequencing. Nat Biotechnol 22:1006–1011 [PubMed: 15247925]
33. Meyers BC, Tej SS, Vu TH et al. (2004) The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. Genome Res 14:1641–c1653 [PubMed: 15289482]
34. Meyers BC, Lee DK, Vu TH et al. (2004) Arabidopsis MPSS. An online resource for quantitative expression analysis. Plant Physiol 135:801–813 [PubMed: 15173564]
35. Quattro CD, Enrico Pè M, Bertolini E (2017) Long noncoding RNAs in the model species Brachypodium distachyon. Sci Rep 7:11252 [PubMed: 28900227]

36. Kodzius R, Kojima M, Nishiyori H et al. (2006) (2006)CAGE: cap analysis of gene expression. Nat Methods 3:211–222 [PubMed: 16489339]
37. Shiraki T, Kondo S, Katayama S et al. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A 100:15776–15781 [PubMed: 14663149]
38. Seki M, Carninci P, Nishiyama Y et al. (1998) High-efficiency cloning of Arabidopsis full-length cDNA by biotinylated CAP trapper. Plant J 15:707–720 [PubMed: 9778851]
39. Mejia-Guerra MK, Li W, Galeano NF et al. (2015) Core promoter plasticity between maize tissues and genotypes contrasts with predominance of sharp transcription initiation sites. Plant Cell 27:3309–3320 [PubMed: 26628745]
40. Cumbie JS, Ivanchenko MG, Megraw M (2015) NanoCAGE-XL and CapFilter: an approach to genome wide identification of high confidence transcription start sites. BMC Genomics 16:597 [PubMed: 26268438]
41. Takahashi H, Lassmann T, Murata M et al. (2012) 5′ end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. Nat Protoc 7:542–561 [PubMed: 22362160]
42. Takahashi H, Kato S, Murata M et al. (2012) CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks. Methods Mol Biol 786:181–200 [PubMed: 21938627]
43. Nagalakshmi U, Wang Z, Waern K et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349 [PubMed: 18451266]
44. German MA, Pillay M, Jeong D-H et al. (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol 26:941–946 [PubMed: 18542052]
45. Zhai J, Arikit S, Simon SA et al. (2013) Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. Methods 67(1):84–90 [PubMed: 23810899]
46. Gregory BD, O'Malley RC, Lister R et al. (2008) A link between RNA metabolism and silencing affecting Arabidopsis development. Dev Cell 14:854–866 [PubMed: 18486559]
47. Willmann MR, Berkowitz ND, Gregory BD (2014) Improved genome-wide mapping of uncapped and cleaved transcripts in eukar-yotes--GMUCT 2.0. Methods 67:64–73 [PubMed: 23867340]
48. Addo-Quaye C, Eshoo TW, Bartel DP et al. (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol 18:758–762 [PubMed: 18472421]
49. Addo-Quaye C, Miller W, Axtell MJ (2009) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioin-formatics 25:130–131
50. Pelechano V, Wei W, Steinmetz LM (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. Nature 497:127–131 [PubMed: 23615609]
51. Pelechano V, Wei W, Jakob P et al. (2014) Genome-wide identification of transcript start and end sites by transcript isoform sequencing. Nat Protoc 9:1740–1759 [PubMed: 24967623]
52. Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848 [PubMed: 19056941]
53. Hetzel J, Duttke SH, Benner C et al. (2016) Nascent RNA sequencing reveals distinct features in plant transcription. Proc Natl Acad Sci 113:12316–12321 [PubMed: 27729530]
54. Erhard KF, Talbot J-ERB, Deans NC et al. (2015) Nascent transcription affected by RNA polymerase IV in Zea mays. Genetics 199:1107–1125 [PubMed: 25653306]
55. Gardini A (2017) Global Run-On Sequencing (GRO-Seq). Methods Mol Biol 1468:111–120 [PubMed: 27662873]
56. Lopes R, Agami R, Korkmaz G (2017) GRO-seq, A tool for identification of transcripts regulating gene expression. Methods Mol Biol 1543:45–55 [PubMed: 28349421]
57. Kwak H, Fuda NJ, Core LJ et al. (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 339:950–953 [PubMed: 23430654]
58. Mahat DB, Kwak H, Booth GT et al. (2016) Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc 11:1455–1476 [PubMed: 27442863]

59. Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469:368–373 [PubMed: 21248844]
60. Larson MH, Mooney RA, Peters JM et al. (2014) A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science 344:1042–1047 [PubMed: 24789973]
61. Nojima T, Gomes T, Grosso ARF et al. (2015) Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell 161:526–540 [PubMed: 25910207]
62. Nojima T, Gomes T, Carmo-Fonseca M et al. (2016) Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nat Protoc 11:413–428 [PubMed: 26844429]
63. Churchman LS, Weissman JS (2012) Native elongating transcript sequencing (NET-seq) Curr Protoc Mol Biol edited by Ausubel Frederick M.… [et al.] Chapter 4:Unit 4.14.1–Unit 4.1417
64. Mayer A, Churchman LS (2016) Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. Nat Protoc 11:813–833 [PubMed: 27010758]
65. Tani H, Mizutani R, Salam KA et al. (2012) Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res 22:947–956 [PubMed: 22369889]
66. Yamada T, Imamachi N, Onoguchi-Mizutani R et al. (2018) 5'-Bromouridine IP Chase (BRIC)-Seq to determine RNA half-lives. Methods Mol Biol 1720:1–13 [PubMed: 29236247]
67. Imamachi N, Tani H, Mizutani R et al. (2014) BRIC-seq: a genome-wide approach for determining RNA stability in mammalian cells. Methods 67:55–63 [PubMed: 23872059]
68. Paulsen MT, Veloso A, Prasad J et al. (2013) Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. Proc Natl Acad Sci 110:2240–2245 [PubMed: 23345452]
69. Paulsen MT, Veloso A, Prasad J et al. (2014) Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. Methods 67(1):45–54 [PubMed: 23973811]
70. Liu J, Jung C, Xu J et al. (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. Plant Cell 24:4333–4345 [PubMed: 23136377]
71. Adams MD, Kerlavage AR, Fleischmann RD et al. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377:3–174 [PubMed: 7566098]
72. Harbers M, Carninci P (2005) Tag-based approaches for transcriptome research and genome annotation. Nat Methods 2:495–502 [PubMed: 15973418]
73. Nakamura M, Carninci P (2004) [Cap analysis gene expression: CAGE], Tanpakushitsu kakusan koso. Protein, nucleic acid, enzyme 49,2688–2693
74. Peiffer JA, Kaushik S, Sakai H et al. (2008) A spatial dissection of the Arabidopsis floral transcriptome by MPSS. BMC Plant Biol 8:43 [PubMed: 18426585]
75. Reinartz J, Bruyns E, Lin J-Z et al. (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Brief Funct Genomic Proteomic 1:95–104 [PubMed: 15251069]
76. Carninci P, Kvam C, Kitamura A et al. (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. Genomics 37:327–336 [PubMed: 8938445]
77. Ni T, Corcoran DL, Rach EA et al. (2010) A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat Methods 7:521–527 [PubMed: 20495556]
78. Morton T, Petricka J, Corcoran DL et al. (2014) Paired-end analysis of transcription start sites in Arabidopsis reveals plant-specific promoter signatures. Plant Cell 26:2746–2760 [PubMed: 25035402]
79. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111 [PubMed: 19289445]
80. Wang H-LV, Chekanova JA (2016) Small RNAs: essential regulators of gene expression and defenses against environmental stresses in plants. Wiley Interdiscip Rev RNA 7:356–381 [PubMed: 26924473]

81. Nechaev S, Adelman K (2011) Pol II waiting in the starting gates: regulating the transition from transcription initiation into productive elongation. Biochim Biophys Acta 1809:34–45 [PubMed: 21081187]
82. Min IM, Waterfall JJ, Core LJ et al. (2011) Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes Dev 25:742–754 [PubMed: 21460038]
83. Lieberman-Aiden E, van Berkum NL, Williams L et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293 [PubMed: 19815776]
84. Dekker J, Rippe K, Dekker M et al. (2002) Capturing chromosome conformation. Science 295:1306–1311 [PubMed: 11847345]
85. Emmert-Buck MR, Bonner RF, Smith PD et al. (1996) Laser capture microdissection. Science 274:998–1001 [PubMed: 8875945]
86. Kerk NM, Ceserani T, Tausta SL et al. (2003) Laser capture microdissection of cells from plant tissues. Plant Physiol 132:27–35 [PubMed: 12746508]
87. Ohtsu K, Takahashi H, Schnable PS et al. (2007) Cell type-specific gene expression profiling in plants by using a combination of laser microdissection and high-throughput technologies. Plant Cell Physiol 48:3–7 [PubMed: 17148694]
88. Gautam V, Sarkar AK (2014) Laser assisted microdissection, an efficient technique to understand tissue specific gene expression patterns and functional genomics in plants. Mol Biotechnol 57:299–308
89. Birnbaum K, Jung JW, Wang JY et al. (2005) Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. Nat Methods 2:615–619 [PubMed: 16170893]
90. Birnbaum K, Shasha DE, Wang JY et al. (2003) A gene expression map of the Arabi- dopsis root. Science 302:1956–1960 [PubMed: 14671301]
91. Carter AD, Bonyadi R, Gifford ML (2013) The use of fluorescence-activated cell sorting in studying plant development and environmental responses. Int J Dev Biol 57:545–552 [PubMed: 24166437]
92. Deal RB, Henikoff S (2010) A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell 18:1030–1040 [PubMed: 20627084]
93. Deal RB, Henikoff S (2011) The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat Protoc 6:56–68 [PubMed: 21212783]
94. Wang D, Deal RB (2015) Epigenome profiling of specific plant cell types using a streamlined INTACT protocol and ChIP-seq. Methods Mol Biol 1284:3–25 [PubMed: 25757765]
95. Speicher MR, Carter NP (2005) The new cytogenetics: blurring the boundaries with molecular biology. Nat Rev Genet 6:782–792 [PubMed: 16145555]
96. Lee JH, Daugharthy ER, Scheiman J et al. (2015) Fluorescent in situ sequencing (FIS-SEQ) of RNA for gene expression profiling in intact cells and tissues. Nat Protoc 10:442–458 [PubMed: 25675209]
97. Sephton CF, Cenik C, Kucukural A et al. (2011) Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. J Biol Chem 286:1204–1215 [PubMed: 21051541]
98. Xing D, Wang Y, Hamilton M et al. (2015) Transcriptome-wide identification of RNA targets of Arabidopsis SERINE/ARGI- NINE-RICH45 uncovers the unexpected roles of this RNA binding protein in RNA processing. Plant Cell 27:3294–3308 [PubMed: 26603559]
99. Mi S, Cai T, Hu Y et al. (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. Cell 133:116–127 [PubMed: 18342361]
100. Keene JD, Komisarow JM, Friedersdorf MB (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nat Protoc 1:302–307 [PubMed: 17406249]
101. Carbonell A (2017) Immunoprecipitation and high-throughput sequencing of ARGONAUTE-bound target RNAs from plants. Methods Mol Biol 1640:93–112 [PubMed: 28608336]
102. Cui X, Liang Z, Shen L et al. (2017) 5-methylcytosine RNA methylation in Arabi-dopsis Thaliana. Mol Plant 10:1387–1399 [PubMed: 28965832]

103. Licatalosi DD, Mele A, Fak JJ et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469 [PubMed: 18978773]
104. Wang T, Xiao G, Chu Y et al. (2015) Design and bioinformatics analysis of genome-wide CLIP experiments. Nucleic Acids Res 43:5263–5274 [PubMed: 25958398]
105. Haque N, Hogg JR (2016) Easier, better, faster, stronger: improved methods for RNA-protein interaction studies. Mol Cell 62:650–651 [PubMed: 27259196]
106. Van Nostrand EL, Pratt GA, Shishkin AA et al. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods 13:508–514 [PubMed: 27018577]
107. Zarnegar BJ, Flynn RA, Shen Y et al. (2016) irCLIP platform for efficient characterization of protein-RNA interactions. Nat Methods 13:489–492 [PubMed: 27111506]
108. Moore MJ, Zhang C, Gantman EC et al. (2014) Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat Protoc 9:263–f293 [PubMed: 24407355]
109. König J, Zarnack K, Rot G et al. (2011) iCLIP--transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. J Vis Exp 50:e2638
110. Hafner M, Landthaler M, Burger L et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141 [PubMed: 20371350]
111. Ascano M, Hafner M, Cekan P et al. (2012) Identification of RNA-protein interaction networks using PAR-CLIP. Wiley Interdiscip Rev RNA 3:159–177 [PubMed: 22213601]
112. Garzia A, Meyer C, Morozov P et al. (2017) Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. Methods 118-119:24–40
113. Hafner M, Lianoglou S, Tuschl T et al. (2012) Genome-wide identification of miRNA targets by PAR-CLIP. Methods 58:94–105 [PubMed: 22926237]
114. Chu C, Qu K, Zhong FL et al. (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44:667–678 [PubMed: 21963238]
115. Chu C, Zhang QC, da Rocha ST et al. (2015) Systematic discovery of Xist RNA binding proteins. Cell 161:404–416 [PubMed: 25843628]
116. Chu C, Quinn J, Chang HY (2012) Chromatin isolation by RNA purification (ChIRP). J Vis Exp 61:e3912
117. Chu C, Chang HY (2016) Understanding RNA-Chromatin Interactions Using Chromatin Isolation by RNA Purification (ChIRP). Methods Mol Biol 1480:115–123 [PubMed: 27659979]
118. Engreitz JM, Pandya-Jones A, McDonel P et al. (2013) The Xist lncRNA exploits three dimensional genome architecture to spread across the X chromosome. Science 341:1237973
119. Engreitz J, Lander ES, Guttman M (2015) RNA antisense purification (RAP) for mapping RNA interactions with chromatin. Methods Mol Biol 1262:183–197 [PubMed: 25555582]
120. Simon MD, Wang CI, Kharchenko PVet al. The genomic binding sites of a non-coding RNA. Proc Natl Acad Sci 108:20497–20502
121. Simon MD, Pinter SF, Fang R et al. (2013) High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature 504:465–469 [PubMed: 24162848]
122. Simon MD (2013) Capture hybridization analysis of RNA targets (CHART) Curr Protoc Mol Biol edited by Ausubel Frederick M.… [et al.] Chapter 21:Unit 21.25
123. Engreitz JM, Sirokman K, McDonel P et al. (2014) RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell 159:188–199 [PubMed: 25259926]
124. Kudla G, Granneman S, Hahn D et al. (2011) Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. Proc Natl Acad Sci 108:10010–10015 [PubMed: 21610164]
125. Helwak A, Kudla G, Dudnakova T et al. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665 [PubMed: 23622248]
126. Helwak A, Tollervey D (2014) Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). Nat Protoc 9:711–728 [PubMed: 24577361]

127. Helwak A, Tollervey D (2016) Identification of miRNA-target RNA interactions using CLASH. Methods Mol Biol 1358:229–251 [PubMed: 26463387]
128. Lucks JB, Mortimer SA, Trapnell C et al. (2011) Multiplexed RNA structure characterization with selective 2^0-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A 108:11063–11068 [PubMed: 21642531]
129. Watters KE, Yu AM, Strobel EJ et al. (2016) Characterizing RNA structures in vitro and in vivo with selective 2´-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Methods 103:34–48 [PubMed: 27064082]
130. Mortimer SA, Trapnell C, Aviran S et al. (2012) SHAPE-seq: high-throughput RNA structure analysis. Curr Protoc Chem Biol 4:275–297 [PubMed: 23788555]
131. Ding Y, Tang Y, Kwok CK et al. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505:696–700 [PubMed: 24270811]
132. Ritchey LE, Su Z, Tang Y et al. (2017) Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. Nucleic Acids Res 45:e135–e135 [PubMed: 28637286]
133. Tang Y, Bouvier E, Kwok CK et al. (2015) StructureFold: genome-wide RNA secondary structure mapping and reconstruction in vivo. Bioinformatics 31:2668–2675 [PubMed: 25886980]
134. Tack DC, Tang Y, Ritchey LE et al. (2018) StructureFold2: bringing chemical probing data into the computational fold of RNA structural analysis. Methods 143:12–15 [PubMed: 29410279]
135. Ding Y, Kwok CK, Tang Y et al. (2015) Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. Nat Protoc 10:1050–1066 [PubMed: 26086407]
136. Silverman IM, Li F, Alexander A et al. (2014) RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. Genome Biol 15: R3 [PubMed: 24393486]
137. Gosai SJ, Foley SW, Wang D et al. (2015) Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the Arabidopsis nucleus. Mol Cell 57:376–388 [PubMed: 25557549]
138. Foley SW, Gregory BD (2016) Protein Interaction Profile Sequencing (PIP-seq) Curr Protoc Mol Biol. / edited by Ausubel Frederick M… [et al.] 116:27.5.1–27.5.15
139. Kertesz M, Wan Y, Mazor E et al. (2010) Genome-wide measurement of RNA secondary structure in yeast. Nature 467:103–107 [PubMed: 20811459]
140. Wan Y, Qu K, Ouyang Z et al. (2013) Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. Nat Protoc 8:849–869 [PubMed: 23558785]
141. Underwood JG, Uzilov AV, Katzman S et al. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods 7:995–1001 [PubMed: 21057495]
142. Uzilov AV, Underwood JG (2016) High-throughput nuclease probing of RNA structures using FragSeq. Methods Mol Biol 1490:105–134 [PubMed: 27665596]
143. Lazof DB, GOLDSMITH JKG, RUFTY TW et al. (2011) The preparation of cryosections from plant tissue: an alternative method appropriate for secondary ion mass spectrometry studies of nutrient tracers and trace metals. J Microsc 176:99–109
144. Kim E-D, Xiong Y, Pyo Y et al. (2017) Spatio-temporal analysis of coding and long noncoding transcripts during maize endosperm development. Sci Rep 7:3838 [PubMed: 28630499]
145. Barcala M, Fenoll C, Escobar C (2012) Laser microdissection of cells and isolation of high-quality RNA after cryosectioning. Methods Mol Biol 883:87–95 [PubMed: 22589126]
146. Blokhina O, Valerio C, Sokolowska K et al. (2016) Laser capture microdissection protocol for xylem tissues of woody plants. Front Plant Sci 7:1965 [PubMed: 28101088]
147. Bevilacqua C, Ducos B (2018) Laser micro-dissection: a powerful tool for genomics at cell level. Mol Aspects Med 59:5–27 [PubMed: 28927943]
148. Gautam V, Singh A, Singh S et al. (2016) An efficient LCM-based method for tissue specific expression analysis of genes and miRNAs. Sci Rep 6:21577 [PubMed: 26861910]
149. Nakazono M, Qiu F, Borsuk LA et al. (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 15:583–596 [PubMed: 12615934]

150. Ohtsu K, Smith MB, Emrich SJ et al. (2007) Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). Plant J 52:391–404 [PubMed: 17764504]
151. Li S, Yamada M, Han X et al. (2016) High-resolution expression map of the Arabidopsis root reveals alternative splicing and lincRNA regulation. Dev Cell 39:508–522 [PubMed: 27840108]
152. Ferré F, Colantoni A, Helmer-Citterich M (2016) Revealing protein-lncRNA interaction. Brief Bioinform 17:106–116 [PubMed: 26041786]
153. Wheeler EC, Van Nostrand EL, Yeo GW (2018) Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. Wiley Interdiscip Rev RNA 9: e1436
154. Augui S, Nora EP, Heard E (2011) Regulation of X-chromosome inactivation by the X-inactivation centre. Nat Rev Genet 12:429–442 [PubMed: 21587299]
155. Dominissini D, Moshitch-Moshkovitz S, Salmon-Divon M et al. (2013) Transcriptome-wide mapping of N6-methyladenosine by m6A-seq based on immunocapturing and massively parallel sequencing. Nat Protoc 8:176–189 [PubMed: 23288318]
156. Dominissini D, Moshitch-Moshkovitz S, Schwartz S et al. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485:201–206 [PubMed: 22575960]
157. Squires JE, Patel HR, Nousch M et al. (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res 40:5023–5033 [PubMed: 22344696]
158. David R, Burgess A, Parker B et al. (2017) Transcriptome-wide mapping of RNA 5-methylcytosine in Arabidopsis mRNAs and noncoding RNAs. Plant Cell 29:445–460 [PubMed: 28062751]
159. Khoddami V, Cairns BR(2013) Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat Biotechnol 31:458–464 [PubMed: 23604283]
160. Ule J, Jensen KB, Ruggiu M et al. (2003) CLIP identifies Nova-regulated RNA net-works in the brain. Science 302:1212–1215 [PubMed: 14615540]
161. Ule J, Jensen K, Mele A et al. (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods 37:376–386 [PubMed: 16314267]
162. Jeon Y, Lee JT (2011) YY1 tethers Xist RNA to the inactive X nucleation center. Cell 146:119–133 [PubMed: 21729784]
163. Martianov I, Ramadass A, Serra Barros A et al. (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature 445:666–670 [PubMed: 17237763]
164. Schmitz K-M, Mayer C, Postepska A et al. (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 24:2264–2269 [PubMed: 20952535]
165. Tollervey JR, Curk T, Rogelj B et al. (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci 14:452–458 [PubMed: 21358640]
166. Wang G, Chen H-W, Oktay Y et al. (2010) PNPASE regulates RNA import into mito-chondria. Cell 142:456–467 [PubMed: 20691904]
167. Vandivier LE, Anderson SJ, Foley SW et al. (2016) The conservation and function of RNA secondary structure in plants. Annu Rev Plant Biol 67:463–488 [PubMed: 26865341]
168. Bevilacqua PC, Ritchey LE, Su Z et al. (2016) Genome-wide analysis of RNA secondary structure. Annu Rev Genet 50:235–266 [PubMed: 27648642]
169. Wan Y, Kertesz M, Spitale RC et al. (2011) Understanding the transcriptome through RNA structure. Nat Rev Genet 12:641–655 [PubMed: 21850044]
170. Hawkes EJ, Hennelly SP, Novikova IV et al. (2016) COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. Cell Rep 16:3087–3096 [PubMed: 27653675]
171. Merino EJ, Wilkinson KA, Coughlan JL et al. (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 127:4223–4231 [PubMed: 15783204]
172. Watts JM, Dang KK, Gorelick RJ et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460:711–716 [PubMed: 19661910]

173. Zaug AJ, Cech TR (1995) Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. RNA 1:363–374 [PubMed: 7493315]

174. Rouskin S, Zubradt M, Washietl S et al. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505:701–705 [PubMed: 24336214]

175. Talkish J, May G, Lin Y et al. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. RNA 20:713–720 [PubMed: 24664469]

176. Silverman IM, Gregory BD (2015) Transcriptome-wide ribonuclease-mediated protein footprinting to identify RNA-protein interaction sites. Methods 72:76–85 [PubMed: 25448484]

Table 1

A selected list of plant lncRNA databases

Name	Descriptions/features	Link	Refs
The Arabidopsis Information Resource (TAIR)	TAIR and its latest version (TAIR10) offer a comprehensive database of the *Arabidopsis thaliana* genome. In addition to gene structures and transcriptome data for coding and nonprotein-coding loci, it offers many analytical tools	https://www.arabidopsis.org/	[6]
Araport11	Similar to TAIR, Araport11 curates comprehensive genomic information of *Arabidopsis thaliana* using the ecotype Col-0 version 11 genome	https://www.araport.org/	[7]
Plant long noncoding RNA database (PLncDB)	This is a plant-specific database that has >13,000 lncRNAs. The organ-specific expression and the differential expression in RNA-directed DNA methylation mutants of the curated lncRNAs are provided. A genome browser allows the user to examine the association between lincRNAs and epigenetic markers	http://chualab.rockefeller.edu/gbrowse2/homepage.html	[8]
Green Non-coding Database (GreeNC)[a]	Including lncRNAs from 37 plant species and algae, this database curates a total of >120,000 annotated lncRNAs and offers the coding potential and folding energy for each curated lncRNA	http://greenc.sciencedesigners.com/wiki/Main_Page	[9]
NONCODE v4.0	This database mainly curated noncoding RNAs from metazoans. Although *Arabidopsis* is the only plant species considered in this database, it has >500,000 lncRNAs from all 17 species and > 3500 *Arabidopsis* lncRNAs	http://www.noncode.org/index.php	[10]
CANTATAdb[a]	This database has >45,000 plant lncRNAs from 10 plant species. Each lncRNA is also evaluated based on potential roles in splicing regulation and miRNA modulations, as well as their tissue-specific expressions and coding potential	http://cantata.amu.edu.pl/	[11]
Plant ncRNA database (PNRD)	Includes lncRNAs from 150 plant species, this database curated a total of >25,000 ncRNAs of 11 types. In addition, this database offers several analytical tools and includes a customized genome browser, a coding potential calculator, and miRNA predictor	http://structuralbiology.cau.edu.cn/PNRD/	[12]
EVLncRNAs[a]	This database has lncRNAs that are validated by experiments and also integrated information from other databases. The database currently has >1500 lncRNAs from >75 species, including animals, plants, and microbes	http://biophy.dzu.edu.cn/EVLncRNAs/	[13]
Plant Natural Antisense Transcripts DataBase (PlantNATsDB)	This database has the natural antisense transcripts (NATs) of 70 plant species. Other information is housed in the database, including associated gene information, small RNA expression, and GO annotation	http://bis.zju.edu.cn/pnatdb/	[14]

[a]The descriptions and protocols for how to use them are included in the same book

Table 2

Selected high-throughput methods for identifying lncRNAs

Method	Purpose	Reference
Tiling microarray	Transcripts and transcriptome analysis	[15, 16]; plant references [17–20]; plant lncRNAs references [20–25]
Serial analysis of gene expression (SAGE)	Transcripts and transcriptome analysis	[26]; plant references [27–30]
Massively parallel signature sequencing (MPSS)	Transcripts and transcriptome analysis	[31]; plant references [32–35])
Cap analysis of gene expression (CAGE, CAGE-seq)	Identify transcripts with 5′ caps	[36, 37]; plant references [38–40]; protocol [41, 42]
RNA-seq	Transcripts and transcriptome analysis	First report [43] and comprehensive review by Wang et al. [4]; in this book: RNA-seq protocols (Part IV, Chapters 11–16), biotic and abiotic stress-related protocol (Part III, Chapters 8–10)
Parallel analysis of RNA ends (PARE)/genome-wide mapping of uncapped and cleaved transcripts (GMUCT)/degradome-seq	Identify transcripts that are being degraded and/or microRNA targets	PARE: [44], protocol [45]; GMUCT: [46], protocol [47]; degradome sequencing [48], protocol [49]
Transcript isoform sequencing (TIF-seq)	High-throughput identification of transcript isoforms with 5′ caps and poly(A) tails	[50]; protocol [51]
Global run-on sequencing (GRO-seq)	Identify the binding sites of transcriptionally active RNA Pol II and nascent RNAs	[52]; plant references [53, 54]; protocol [55, 56]
Precision nuclear run-on sequencing (PRO-seq)	Identify the binding sites of transcriptionally active RNA Pol II and nascent RNAs	[57]; protocol [58]
Native elongating transcript sequencing (NET-seq)	Identify the binding sites of elongating RNA Pol II and the associated nascent RNAs	[59–61]; protocols [62–64]
BRIC-Seq/BrU-Seq/BrUChase-Seq	Identify nascent RNAs and RNA half-lives, and measure RNA decay	BRIC-seq:[65], protocol [66, 67]; BrU-seq/BrUChase-seq: [68], protocol [69]

Table 3

Selected methods for analyzing the functions of lncRNAs

Method	Purpose	Reference
Laser microdissection (LM)	Tissue or cell-type specific analysis	[85–88]; protocol can be found in Chapter 5 of this book
Fluorescence-activated cell sorting (FACS)	Tissue or cell-type specific analysis	[89–91]
Isolation of nuclei tagged in specific cell types (INTACT)	Tissue or cell-type specific analysis	[92]; protocol can be found in Chapter 7 of this book and [93, 94]
Cryosectioning	Tissue or cell-type specific analysis	Protocol can be found in Chapter 3 of this book
In situ hybridization (ISH) and fluorescence in situ hybridization (FISH)	Visualization of RNA and DNA; tissue or cell-type specific analysis	[95]; protocol can be found in Chapter 6 of this book and [96]
RNA immunoprecipitation followed by microarray or sequencing (RIP-Chip or RIP-seq)	High-throughput identification of RNA-protein interactions	[97–99]; protocol can be found in Chapters 17 and 18 of this book and [100, 101]
m^5CRNA immunoprecipitation followed by sequencing (m^5C-RIP-seq)	High-throughput identification of RNA modifications	[102]; protocol can be found in Chapter 24 of this book
High-throughput sequencing cross-linking immunoprecipitation (HITSCLIP, CLIP-seq, eCLIP, irCLIP)	High-throughput identification of RNA-protein interactions	[103–107]; protocol [108, 109]
Photoactivatable ribonucleotide-enhanced cross-linking and immunoprecipitation (PAR-CLIP)	High-throughput identification of RNA-protein interactions	[110, 111]; protocol [112, 113]
Chromatin isolation by RNA purification sequencing (ChIRP, ChIRP-Seq)	Genome-wide identification of RNA-DNA/chromatin interactions	[114]; ChIRP-MS [115]; protocol [116, 117]
RNA antisense purification (RAP)	Genome-wide identification of RNA-DNA/chromatin interactions	[118]; protocol can be found on the Guttman lab's website at http://www.lncrna-test. caltech.edu/protocols.php and [119]
Capture hybridization analysis of RNA targets (CHART, CHART-seq)	Genome-wide identification of RNA-DNA interactions	[120, 121]; protocol [122]
RNA antisense purification followed by RNA sequencing (RAP-RNA)	High-throughput mapping of RNA-RNA interactions	[123]; protocol can be found on the Guttman lab's website at http://www.lncrna-test. caltech.edu/protocols.php
Cross-linking, ligation, and sequencing of hybrids (CLASH)	High-throughput mapping of RNA-RNA interactions	[124, 125]; protocol [126, 127]
Selective 2′-hydroxyl acylation by primer extension sequencing (SHAPE-seq)	In vitro high-throughput profiling of RNA secondary structure	[128]; protocol [129, 130]
Structure-seq/Structure-seq2	In vivo high-throughput profiling of RNA secondary structure; method developed using *Arabidopsis thaliana*	[131–134]; protocol can be found in Chapter 20 of this book and [135]
Protein interaction profile sequencing (PIP-seq)	High-throughput profiling of RNA-protein interactions and RNA secondary structure; method developed using HeLa cells and *Arabidopsis thaliana*	[136, 137]; protocol can be found in Chapters 21 and 22 of this book and [138]
Parallel analysis of RNA structure (PARS)	High-throughput profiling of RNA secondary structure	[139]; protocol [140]
Fragmentation sequencing (frag-seq)	High-throughput profiling of RNA secondary structure	[141]; protocol [142]