Briefings in Bioinformatics, 20(5), 2019, 1781–1794

doi: 10.1093/bib/bby050
Advance Access Publication Date: 4 June 2019
Review article

Versatile interactions and bioinformatics analysis of noncoding RNAs

Qi Chen, Xianwen Meng, Qi Liao and Ming Chen

Corresponding author: Ming Chen, Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, P. R. China. Tel.: +86 (0)571-88206612; Fax: +86 (0)571-88206612; E-mail: mchen@zju.edu.cn

Abstract

Advances in RNA sequencing technologies and computational methodologies have provided a huge impetus to noncoding RNA (ncRNA) study. Once regarded as inconsequential results of transcriptional promiscuity, ncRNAs were later found to exert great roles in various aspects of biological functions. They are emerging as key players in gene regulatory networks by interacting with other biomolecules (DNA, RNA or protein). Here, we provide an overview of ncRNA repertoire and highlight recent discoveries of their versatile interactions. To better investigate the ncRNA-mediated regulation, it is necessary to make full use of innovative sequencing techniques and computational tools. We further describe a comprehensive workflow for *in silico* ncRNA analysis, providing up-to-date platforms, databases and tools dedicated to ncRNA identification and functional annotation.

Key words: noncoding RNAs; ncRNA transcription; ncRNA–RNA interaction; ncRNA–DNA interaction; ncRNA–protein interaction; bioinformatics resources

Introduction

The central dogma of molecular biology raised in the mid-20th century has largely confined the role of RNA as the simple template for protein synthesis. Messenger RNA (mRNA) has long been the major research focus, while noncoding RNA (ncRNA) was considered as a by-product of massive transcription with less biological meaning. Notably, the past few decades have witnessed the emergence of the previously unsuspected noncoding world (Figure 1). Since the initial discovery of transfer RNA (tRNA) and ribosome RNA (rRNA) in the late 1950s, ncRNAs have gradually surfaced, which turned out to encompass a huge variety of RNA species. The joint analysis of large-scale sequencing data with computational tools represents a powerful approach for the ncRNA exploration. At the beginning of the 21st century, initial sequencing and analysis of human [1] and mouse genome [2] have first revealed a large number of ncRNAs in animals unexpectedly [3]. Soon afterwards, the Human Genome Project (HGP) was achieved in 2005 [4] with abundant lncRNAs detected in mammals [5, 6]. Later, the wide application of next-generation sequencing has further allowed a more accurate profiling of ncRNAs [7, 8]. The ENCODE (Encyclopedia of DNA Elements) project launched in 2005 has revealed that up to 80% of our genome is transcribed into ncRNAs in its recent reports [9, 10]. The large ncRNA data sets generated from sequencing projects have promoted the establishment of many public databases such as Rfam [11], NONCODE [12], miRBase [13] and circBase [14].

Generally, ncRNAs are found to regulate many physiological, developmental and disease processes. They have been identified as oncogenic drivers and tumor suppressors in various cancer types [15]. In addition, accumulating evidences indicate that ncRNAs function as key regulatory molecules in plant stress

Qi Chen is a Master student in Ming Chen's laboratory in Zhejiang University. Her research focuses on the analysis of circular RNAs in plants.
Xianwen Meng is a Ph.D. student in Ming Chen's laboratory in Zhejiang University. His research focuses on biogenesis, properties and functions of circular RNAs in eukaryotic cells.
Qi Liao is a assistant professor at the Department of Preventative Medicine, School of Medicine, Ningbo University, Ningbo, Zhejiang, China. Her current research involves the identification and functional annotation of lncRNAs in bioinformatics way.
Ming Chen is a full professor in the Department of Bioinformatics, College of Life Sciences, Zhejiang University. His current research focuses on the construction of noncoding RNA-mediated regulatory networks and establishing useful web servers and platforms to help biologists browse and analyze massive biological data sets.
Submitted: 28 March 2018; **Received (in revised form):** 2 May 2018

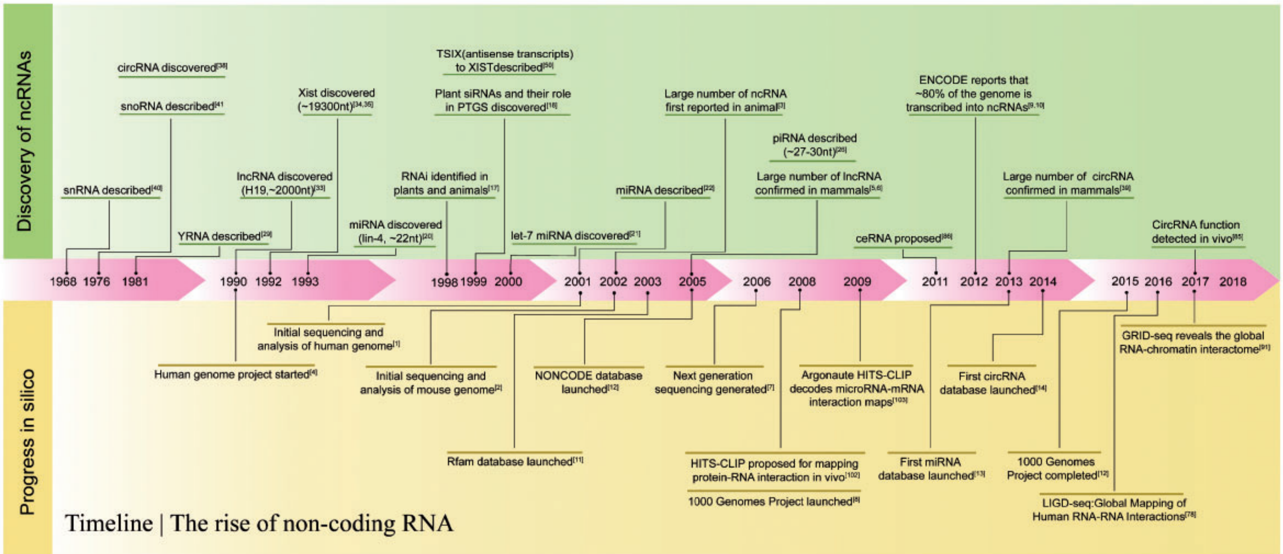

Figure 1. Timeline of ncRNA research in the past half-century. The upper panel lists the important events of ncRNA discoveries from 1968 to 2017 and the lower panel lists the bioinformatics progresses in ncRNA analysis from 1990 to 2017.

responses [16]. It is clear that ncRNA analysis has become a cutting-edge trend, and current progresses have already yielded novel insights into their functions. However, the annotation and interpretation of ncRNAs are still a challenging task because of the huge volumes of data and the diversity of ncRNAs, which has made it necessary to dispose of bioinformatics methods to store, analyze and visualize ncRNA information. Considering the rapid growth of the field, a comprehensive review of the latest development of ncRNA transcriptomes that involves multiple ncRNA species and their interrelationship is scientifically appealing. In this manuscript, we start with a brief overview of noncoding members and ncRNA transcription. We then summarize three layers of ncRNA-related interactions to see the panorama of ncRNA function. At last, bioinformatics resources dedicated to ncRNA studies are provided with up-to-date platforms, databases and tools.

ncRNA repertoire in eukaryotes

Eukaryotic transcription from different genomic regions and RNA processing yield a diverse catalog of ncRNA species. Based on their functional features, ncRNAs are roughly divided into two large parts: the housekeeping RNA and the regulatory RNA. The former one mainly involves in basic cell maintenance, while the regulatory RNA participates in various biological processes.

Regulatory ncRNAs

Small interfering RNAs (siRNAs) are defined as a class of double-stranded RNA molecules (~20–25 nt) that are derived from the fold-back structure with nearly perfect complementarity. It can be divided into two large categories based on their origination. Exogeneous siRNA (exo-siRNA) is a kind of exogenous RNA because of artificial insertion or virus infections. Endogenous siRNAs (endo-siRNA) have been detected within cells that play an essential role in transposon control and DNA rearrangement via RNA interference [17, 18]. The endo-siRNAs are mostly transcribed from transposon elements (TEs) and can be further divided into several subcategories: repeat-associated siRNAs, heterochromatic siRNAs, cis-acting RNAs, trans-acting RNAs and natural antisense siRNAs [19].

MicroRNAs (miRNAs) are small endogenous ncRNAs (~21–23 nt) that are processed from transcribed hairpin loop structures [20–22]. They are abundant in the cytoplasm that regulate gene expression at the levels of mRNA stability and translation [23]. MiRNA has gained great attention in past few years as a key player in the intricate interplay among diverse RNA species. Many other ncRNA species such as small nucleolus RNAs (snoRNAs), rRNAs, tRNAs, piRNAs and lncRNAs can be processed into miRNAs through miRNA machinery [24]. Additionally, both mRNAs and some ncRNAs can communicate with and co-regulate each other by competing for binding to miRNAs [25].

Piwi-interacting RNAs (piRNAs) are named after piwi proteins, as many piwi-like proteins could process precursor piRNAs into mature piRNAs through ping-pong pathway [26, 27]. Unlike miRNAs and siRNAs, piRNAs (~24–32 nt) are processed from single-stranded RNA precursors in a Dicer-independent manner. This animal-specific ncRNA species usually forms a piRNA-induced silencing complex to target transposons in the germ line of many animal species [28], which acts in a similar manner to endo-siRNAs in plants to silence TEs.

Y RNAs were first identified in the early 1980s during investigations of autoimmune proteins and associated RNAs in systemic lupus erythematosus patients [29, 30]. The size of Y RNAs varies from 70 to 115 nucleotides, and they are able to produce smaller RNA fragments during apoptosis. Y RNA-derived small RNAs (~22–36 nt) and full-length Y RNAs are highly abundant in various cell types [31]. Intriguingly, some of Y RNA fragments (~24 nt) were initially mis-annotated as a novel type of miRNAs. Recent reports have speculated that some cleavage products of Y RNAs are likely to enter the miRNA pathway [32].

Long ncRNAs (lncRNAs) are defined as nonprotein coding transcripts longer than 200 nt, which were first discovered in the early 1990s [33–35]. According to the genomic location, they can be divided into three subcategories. Long intergenic ncRNAs (lincRNAs) are located and transcribed within the intergenic region. Long intronic ncRNAs lie in the intronic region of protein-coding genes. Other lncRNAs overlap with or intersperse between multiple coding and noncoding transcripts. Some lncRNAs can be processed into small ncRNAs (such as miRNAs, piRNAs and snoRNAs) to perform distinct functions [36]. Therefore, lncRNAs are characterized to have weak functional constraint and rapid turnover [37].

Circular RNAs (circRNAs) are emgerging as a unique group of ncRNAs that forms a covalently closed loop structure from exon circularization. Although circRNAs were discovered decades ago [38], little attention was paid to this nonlinear ncRNAs until recent years with the advance of sequencing techniques. Different from small ncRNAs, circRNAs have a wide range of molecular size from 100 nt to 4 kb, while commonly these are a few hundred nucleotides in human cells [39]. They can be classified into exonic circRNAs, intronic circRNAs, untranslated regions (UTRs) circRNAs, intergenic circRNAs and other circRNAs on the basis of the genome region from which circRNAs arise.

Housekeeping ncRNAs

Housekeeping RNAs comprise rRNAs, tRNAs, small nuclear RNAs (snRNAs) and small nucleolus RNAs (snoRNAs), as early discovered ncRNA species [40, 41]. They are uniformly expressed with little variance in all cells to maintain the basic cellular function. Housekeeping ncRNAs have a wide length scale, ranging from 50 nt to 500 nt. Advances in ncRNA research have revealed some housekeeping RNAs that are cleaved to perform regulatory roles. For example, tRNA-derived RNA fragments (tRFs) and translation interfering tRNAs are two new classes of regulatory ncRNAs that are derived from the cleavage of tRNAs [42]. Studies have revealed that translation interfering tRNAs could inhibit translation through recruitment of innovative packed aggregates of proteins and RNAs under stress situation [43, 44]. In addition, deep sequencing together with bioinformatic analyses has also discovered some short RNAs derived from snoRNAs: sno-miRNAs [45] and sno-piRNAs [46].

ncRNA transcription from different genomic regions

Eukaryotic genomes have a much lower gene density than prokaryotic genomes, which are considered as efficient evolutionary consequences to meet their high biological requirements. It has been found that ncRNA transcripts cover >98% of all genomic output in humans [47]. Genomic regions that produce ncRNA transcripts are important for gene expression regulation; thus, it is necessary to study them for a better understanding of ncRNA biogenesis and biological functions. Additionally, it is worth mentioning the RNA amplification mechanism in which small ncRNAs are used as templates for the synthesis of secondary small RNAs, usually termed 22G RNAs [48]. In this section, however, we mainly focus on new findings of ncRNA transcription from three different genomic regions: protein-coding genes, enhancer elements and TEs.

ncRNAs derived from protein-coding genes

Exon sequences in mRNAs are the main source for protein synthesis, while many ncRNAs also contain exons and overlap with protein-coding genes (Figure 2A). For example, processed pseudogenes as a result of retrotransposition only contain exons and have introns discarded. Some unprocessed pseudogenes derived from duplication of functional genes maintain the genomic features of intron-connected exons [49]. Some lncRNAs are transcribed from the antisense strand of protein-coding genes or overlap with them [50]. A majority of circRNAs identified by sequencing rRNA-depleted, RNase treatment RNAs are found to contain non-colinear exons [51]. And one host gene can generate multiple circRNAs that comprise different numbers of exons via alternative splicing.

In protein-coding genes, introns are usually degraded via splicing and used to be deemed as junk sequences. However, the advent of ncRNAs in higher organisms, especially large amount of intronic-derived ncRNAs such as miRNAs, snoRNAs, lncRNAs and circRNAs (Figure 2A) suggests otherwise. Notably, nearly half of human miRNAs and a majority of snoRNAs are derived from introns. The ENCODE project has reported that around 20% of lncRNAs are sense intronic with no intersection with exons [52]. Recent deep sequencing of non-polyadenylated transcriptomes of human cells has identified a unique type of intron-derived lncRNAs: sno-lncRNAs [53]. They appear to derive from the intron that imbeds two snoRNA genes, and the internal sequences between two snoRNAs are not degraded, leading to the accumulation of lncRNAs flanked by snoRNA sequences but lacking 5′caps and 3′poly(A) tails. Some intronic RNAs resistant to RNase R treatment are likely to be circRNAs. Zhang *et al.* (2013) [54] identified 103 RNase R-enriched intronic circRNAs in H9 cells, and 485 intronic circRNAs (4.0%) were identified in *Oryza sativa* [55].

ncRNAs derived from enhancers

Mammalian genomes are populated with thousands of enhancers [56], commonly defined as a type of cis-regulatory sequences (∼50–1500 bp) that are positioned far from the target genes (∼20 kb–2 Mb). Their transcriptional regulation is achieved via promoter–enhancer loops that form higher-order chromatin structures [57]. Intriguingly, studies have observed that RNA polymerase II can aggregate at enhancer elements and respond dynamically to signal transduction [58]. Later, with total RNA deep sequencing techniques, enhancer-derived RNAs were found in nerve cells [59] and T cells [58, 60].

Enhancer RNAs (eRNAs) share similar transcriptional features with lncRNAs and mRNAs, but they are generally less stable and easily degraded by the exosome complex [61]. Bidirectional transcription results in largely non-polyadenylated eRNAs [62], while unidirectional transcribed eRNAs (>∼3–4 kb) are polyadenylated at 3′sites (Figure 2B) [63]. Several studies have demonstrated an ambiguous overlap between polyadenylated unidirectional transcripts produced at enhancer regions and lincRNAs [64], as they share many similar features. For enhancers in the intragenic region, alternative transcription start sites are able to generate both polyA$^-$ and polyA$^+$ eRNAs (Figure 2B) [65]. The latter one is known as multi-exonic eRNA: an alternative isoform of its host gene with low-coding potential [64]. Cell-based experiments have validated functional requirements for eRNAs in the enhancer-mediated gene regulation [66]. Differential expression signatures of eRNAs across cell types and tissues correlate with specific enhancer activities [67]. They are actively involved in the regulation of chromatin remodeling [68] and gene expression as effective indicators of enhancer activity [67].

ncRNAs derived from TEs

Protein-coding genes are mainly derived from single-copy sequences, while the rest of the background regions are filled

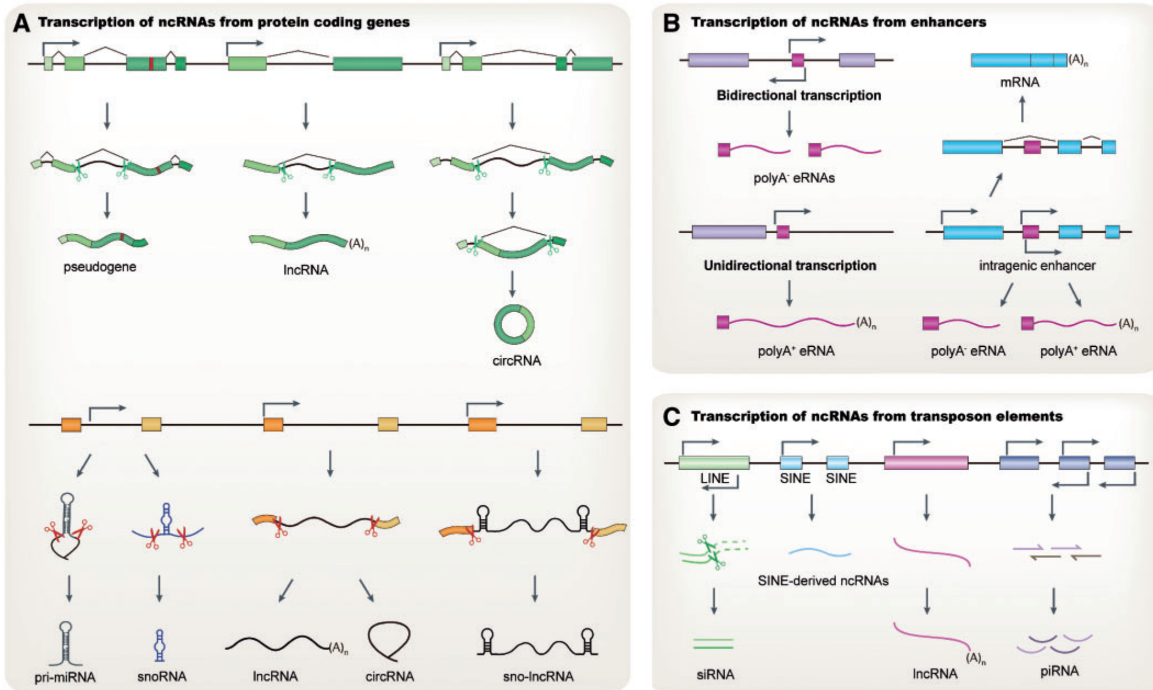

Figure 2. Transcription of ncRNAs from different DNA elements. (**A**) Exons in the protein-coding genes can be transcribed into pseudogene, lncRNA and circRNA, most of which have introns discarded. Introns in the protein-coding genes can be transcribed into miRNA, snoRNA, lncRNA, circRNA and sno-lncRNA via special processing events. (**B**) Enhancer regions can be transcribed into polyA$^-$ or polyA$^+$ eRNAs. (**C**) TEs can be transcribed into siRNA, SNIE-derived ncRNA, lncRNA and piRNA.

with a myriad of repetitive elements. Over two-thirds of the human genome consists of repetitive elements [69]. They are either tandemly repeated sequences or dispersed throughout the genome as TEs. TEs are the major contributors to the evolutionary origination and biogenesis of regulatory RNAs. Both long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) can generate RNA intermediates during transposition. LINEs (~7000 bp) can be transcribed and processed into endo-siRNAs that target genome regions where LINEs reside (Figure 2C). SINEs are transcribed by RNA polymerase III into ncRNAs, typically shorter than 600 bp (Figure 2C). SINE loci are normally silent [70] but can be robustly activated on DNA virus infection [71]. Several studies have shown that SINE-derived ncRNAs not only regulate gene expression of RNA polymerase II inside the nucleus [72, 73] but also involve in the stimulation of cytoplasmic immune signaling [74].

In human, mouse and zebrafish, up to two-thirds of lncRNAs were found to contain exonic TE sequences [75]. Upregulated expression of some lncRNAs was found to be closely related with highly enriched TEs in the upstream regions of these lncRNA genes [75]. It is highly possible to link the less tractable lncRNA evolution with the dynamic transposition activity. While some lncRNAs are merely results of pervasive transcriptional activity of TEs with little biological function, TEs-derived piRNAs are well known to silence transposable elements in the animal germline [19]. Obviously, accumulated TE insertions are pernicious to host genes, thereby the dicer-independent piRNA pathway serves as a beneficial mechanism to target transposon clusters and maintain the healthy transgenerational inheritance [76].

Versatile roles of ncRNA-associated interaction

The far-reaching regulatory effects of ncRNAs are largely benefited from the originally single-strand form, which avails its bases for hydrogen bonding with other complementary molecules or conjugates intramolecularly to form secondary structures [77]. Recent high-throughput techniques have produced remarkable evidences for ncRNA-associated interactions in different kinds of cellular functions. Based on the current studies, we delineate a panorama of ncRNA functions in Figure 3. Of note, it does not include all modes of ncRNA action and will be updated with future advances in ncRNA biology. Specifically, nuclear ncRNAs are found to have roles in chromosome replication, chromosomal modification and organization, transcriptional regulation, alternative splicing and telomere elongation. Cytoplasmic ncRNAs actively participate in 22G-RNA production, mRNA stability and translation, protein degradation and translocation (Figure 3).

ncRNAs interact with RNA

RNA–RNA interactions between different RNA species well show the nature of the interlaced noncoding world. In the context of their diversity and abundance, systematic mapping of RNA–RNA interactions is of great significance for the comprehensive charting of ncRNAs. Recently, a newly developed sequencing technique: LIGation of interacting RNA followed by high-throughput sequencing (LIGR-seq) [78], has enabled global-scale detection of possible RNA duplexes *in vivo* with no need of prior knowledge. LIGR-seq analysis of human transcriptome has revealed significant interactions among housekeeping RNAs. Thereinto, intermolecular rRNA–rRNA interactions provide the structural framework for ribosomal proteins during the generation of ribosomes, and snRNA–snRNA contacts function in a similar manner in the formation of spliceosomes. Besides, snoRNAs-associated interactions with rRNAs and snRNAs suggest the snoRNA-guided modification in the maturation of housekeeping RNAs. LIGR-seq also detected significant

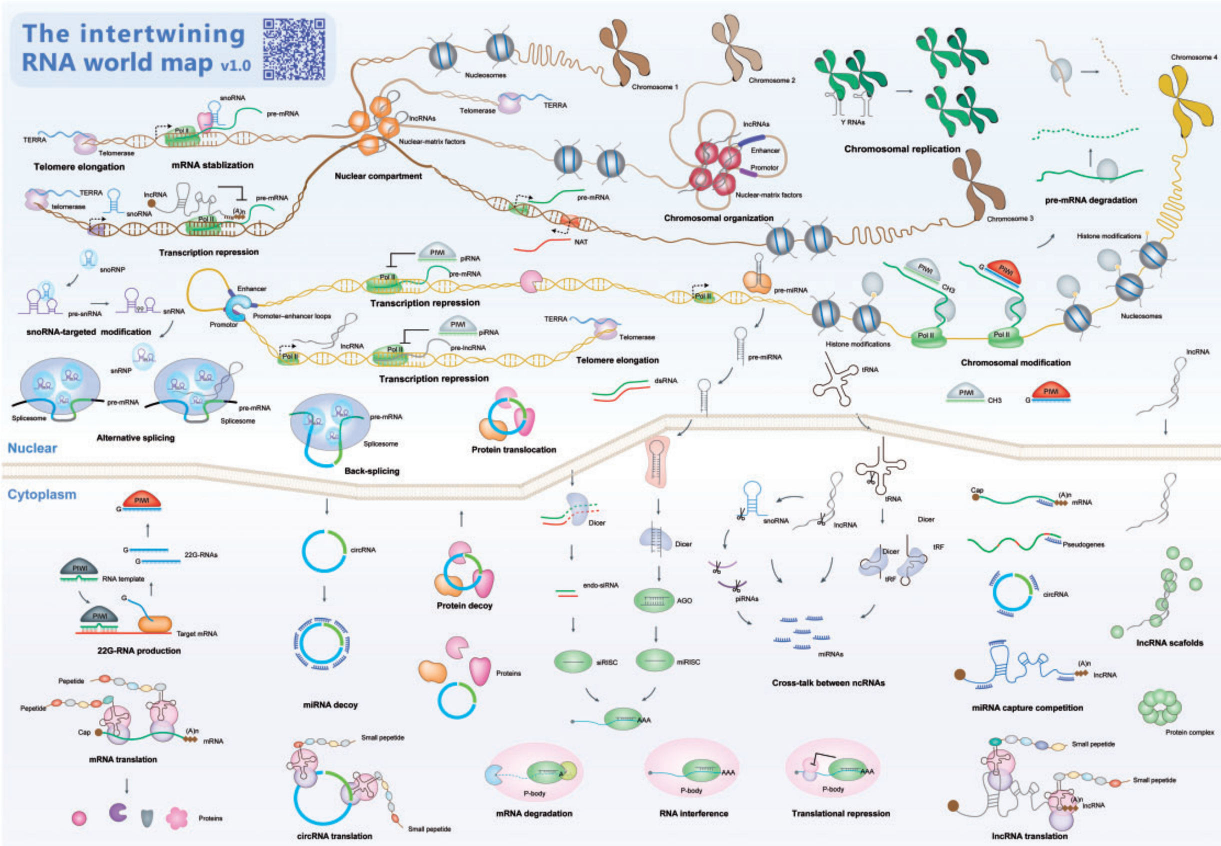

Figure 3. The functional roles of ncRNA in nuclear and cytoplasm. The upper panel presents the ncRNA activities in nuclear. Thereinto, telomeric repeat-containing RNAs together with telomerase maintain the genome integrity via telomere elongation. SnoRNAs play a part in epigenetic modification and may also facilitate the mRNA stabilization during pre-mRNA transcription. SnRNPs are important components of the splicesome and lncRNAs can interfere splicing. lncRNAs inside the nuclear actively participate in nuclear compartment, chromosomal organization and transcription repression. PiRNAs regulate RNA silencing via transcription repression. Y RNAs are involved in the chromosomal DNA replication. Nuclear-acting small RNAs can induce chromatin modifications, such as H3K9 methylation. The lower panel presents the ncRNA activities in the cytoplasm. CircRNAs can function as a decoy of proteins and achieve protein translocation. Small ncRNAs (miRNAs and siRNAs) share a common set of cellular proteins (Dicer and the RNA-induced silencing complex) in their biogenesis and RNA interference. Some ncRNAs are processed into miRNAs that target multiple transcripts and form ceRNA pairs. Small ncRNAs are used as templates for the synthesis of secondary small RNAs (22 G RNAs). LncRNAs possibly act as scaffolding molecules for protein complexes. Both lncRNAs and circRNAs can be translated into small peptides.

snoRNA–mRNA interactions, which provides evidence for the previous implication of snoRNA-mediated regulation [79]. However, the method is unable to capture short ncRNA-related interactions [78], thus missing small ncRNA targets.

For small ncRNAs (miRNAs, siRNAs and piRNAs), target binding usually involves partial pairing within ‘seed regions’: short stretches of nucleotides (~6–10 nt) [80]. The seemingly primary interactions are based on the secondary and tertiary structures [81] so as to recruit necessary factors and expose specific sequences to bind functional targets. Among small RNAs (sRNAs), miRNAs are particularly fascinating in terms of their highly active relationship with target genes. Increasing evidence suggests miRNAs as guide strands for mRNA degradation and translational inhibition to a large extent through imperfect base pairing at 3′UTRs of target mRNAs [25]. Other ncRNA species with miRNA-response elements such as pseudogenes [82], lncRNAs [83] and circRNAs [84, 85] are also targeted by miRNAs, thus acting as competitive endogenous RNAs (ceRNAs). Those miRNA sponges can inhibit specific miRNA activity and relieve the repression to the originally targeted mRNAs. The ceRNA hypothesis was raised in 2011 [86], and its derivative ceRNA network has largely enriched our vision of how cellular networks may operate. Because of the diversity and abundance of ceRNA pairs, computational methods have risen as effective tools to predict miRNA-associated interactions on a whole genome-wide.

ncRNAs interact with DNA

Deep sequencing of small RNAs in the nuclear has discovered the abundant existence of chromatin-associated ncRNAs: promotor-associated small RNA (PASR) [87], transcription initiation small RNA [88], transcription start site-associated RNA [89] and splice site-associated RNA [90]. They are possibly involved in nucleosome positioning, chromatin marking and transcriptional regulation. A new RNA sequencing technique: global RNA interactions with DNA by deep sequencing [91] has recently been developed to detect the whole genomic chromatin-interacting RNAs comprising both mRNAs and ncRNAs. It revealed a large number of snoRNAs enriched near the active gene loci. Some were previously reported to interact with nascent pre-mRNAs during transcription to protect their integrity [92], therefore proximately locating to chromatin at active genes. The method also detected various trans-acting lncRNAs, including two well-characterized mammalian lncRNAs: MALAT1 and NEAT1 [93].

Those nuclear-enriched lncRNAs are able to bring widely separated functional elements (within a chromosome or between chromosomes) into close spatial proximity, thus compartmentalizing the nucleus [94]. Their involvement in the organization of multi-chromosomal regions largely relies on nuclear-matrix factors [95, 96] through which they attain the affinity with chromosomes. The arrangement of chromosomal 3D conformation is important in the precise execution of nuclear functions. It also provides favorable conditions for lncRNA-associated interactions with transcriptional regulators on functional DNA elements. On the one hand, lncRNAs can interfere with the expression of a protein-coding gene that is in close proximity to their transcriptional sites [94], such as some antisense lncRNAs. On the other hand, highly abundant and stable lncRNAs can diffuse throughout the nucleus to spatially search for affinity sites [93, 97] and broadly modulate gene expression on various chromosomes. Besides, Y RNAs in the nuclear also have intimate communication with chromosomes. They were found to be key factors in a Ro ribonucleoproteins (RoRNPs) independent manner during the initiation of DNA replication [98, 99]. In the start of DNA replication, Y RNAs associate preferentially with replicated chromatin. Four kinds of Y RNAs (Y1, Y3, Y4 and Y5) were detected in the chromatin-associated fraction with similar abundance in both caner and non-cancer cell lines [100, 101].

ncRNAs interact with protein

Development of deep-sequencing approaches coupled with immunoprecipitation of RNA binding proteins (RBPs) has revealed a wide range of ncRNA-associated proteins [102, 103]. Housekeeping ncRNAs interacting with proteins can form various ribonucleoprotein (RNP) complexes that perform diverse functions. For instance, snRNAs along with multiple proteins make up spliceosomes (snRNPs) that take part in both canonical splicing and alternative splicing. Many nucleotides in the pre-rRNAs, pre-snRNA and pre-tRNA undergo post-transcriptional modifications via small nucleolar RNP particles [104]. Y RNA was first discovered as components of Ro60 RNP particle, and Ro60 proteins were found critical in the stabilization of Y RNAs. Vault RNPs known as vaults also contain a small portion of ncRNAs (vault RNA) that bind with several vault proteins.

Interactions between regulatory RNAs and proteins are essential in mediating fundamental cellular processes. Small ncRNAs (miRNAs, siRNAs and piRNAs) are well known to interact with the Argonaute family proteins during RNA interference pathway that affects RNA stability and translation. Comparatively, lncRNAs and circRNAs seem to be wonderful berths for multiple proteins. Binding with scaffold attachment factor A (SAFA) tethers Xist to chromatin, while other Xist-associated proteins target for transcriptional silencing [96, 105]. As flexible scaffolds, lncRNAs are able to recruit protein modules with distinct functions but share same compartmentalization to ensure biological efficiency. Some lncRNAs with multicopy repeating RNA domain achieve continuous combination with SAFAs, thus bringing widely separated functional elements into close spatial proximity through the 3D organization of chromosomal architecture [106]. CircRNAs can act as protein sponges that translocate proteins to the specific subcellular compartment. In tumor apoptosis studies, circ-Foxo3 was actively expressed and binded with p53 and MDM2, which promoted MDM2-induced p53 ubiquitination and subsequent degradation [107]. It is clear that protein recruitment by lncRNAs and circRNAs alters the cellular concentration and localization of captured proteins. With accumulated evidence of ncRNAs as agencies of proteins communication, their potential involvement in protein co-localization and protein–protein interaction (PPI) network warrants more attention.

Bioinformatics resources for ncRNA analysis

Reviews on ncRNA-related computational methods are emerging, while most of them are focused on one particular category or already out of style because of the boom of newly constructed tools and data sets. Here, we summarize a comprehensive workflow for ncRNA analysis (Figure 4), providing up-to-date platforms, databases and tools dedicated to ncRNAs identification and functional annotation. The first step is the preprocessing of RNA-seq data that involves quality control and adapter removal. The trimmed data can be mapped to the reference sequences via Tophat [108], STAR [109] and HISAT2 [110] accordingly. The aligned reads are associated with the annotation information of ncRNAs from public databases to be classified into different ncRNA categories, while unclassified sequences are used to predict novel ncRNAs. The ncRNA data sets are then applied for differential expression analysis. In the annotation part, ncRNA-associated interaction is highlighted as accumulated ncRNAs were reported to function by interfacing with diverse classes of biomolecules. The wide application of high-throughput sequencing has enabled global-scale mapping of RNA-associated interactions *in vivo*. Bioinformatics resources, on the other hand, make computationally feasible and biologically relevant prediction to complement experimentally identified interactions. Thereinto, a large number of miRNA target prediction tools have been developed, such as TargetScan [111] and PicTar [112] in animals, targetFinder [113], psRNATarget [114] and TAPIR [115] in plants. Next, results from both experiments and prediction can be applied to construct an *in silico* network for further understanding of ncRNA functions. By network visualization and pattern recognition, key ncRNAs and corresponding targets can be identified based on hierarchical and topologic characteristics of the network.

ncRNA analysis platform

Data processing from RNA-seq for ncRNA analysis takes several steps with a series of tools. There is a substantial need for systematic interpretation platforms that allow a convenient analysis of RNA-seq data sets. Here, we summarize seven fully automated and easy to use web services suitable for ncRNA detection, profiling and functional annotation based on high-throughput sequencing.

DARIO [116], launched in 2011, is the first web service for small ncRNAs analysis from high-throughput sequencing data. It provides comprehensive quantification of different ncRNA classes (miRNAs, C/D snoRNAs, H/ACA snoRNAs, tRNAs, scRNAs and rRNAs) based on the annotation in the public ncRNA databases. Besides classification of known ncRNAs by a RandomForest classifer, DARIO is capable of non-annotated ncRNA prediction. The final results include ncRNA loci, expression pattern and genome browser. While DARIO mainly works for animals, PlantDARIO [117], an extension of DARIO was later developed for plant-specific small ncRNA analysis.

MirTools 2.0 [118] detects and profiles various types of ncRNAs, such as miRNAs, tRNAs, snRNAs, snoRNAs, rRNAs and piRNAs. Users can input either raw sequences or alignment data to analyze any sequenced genomes for either single case

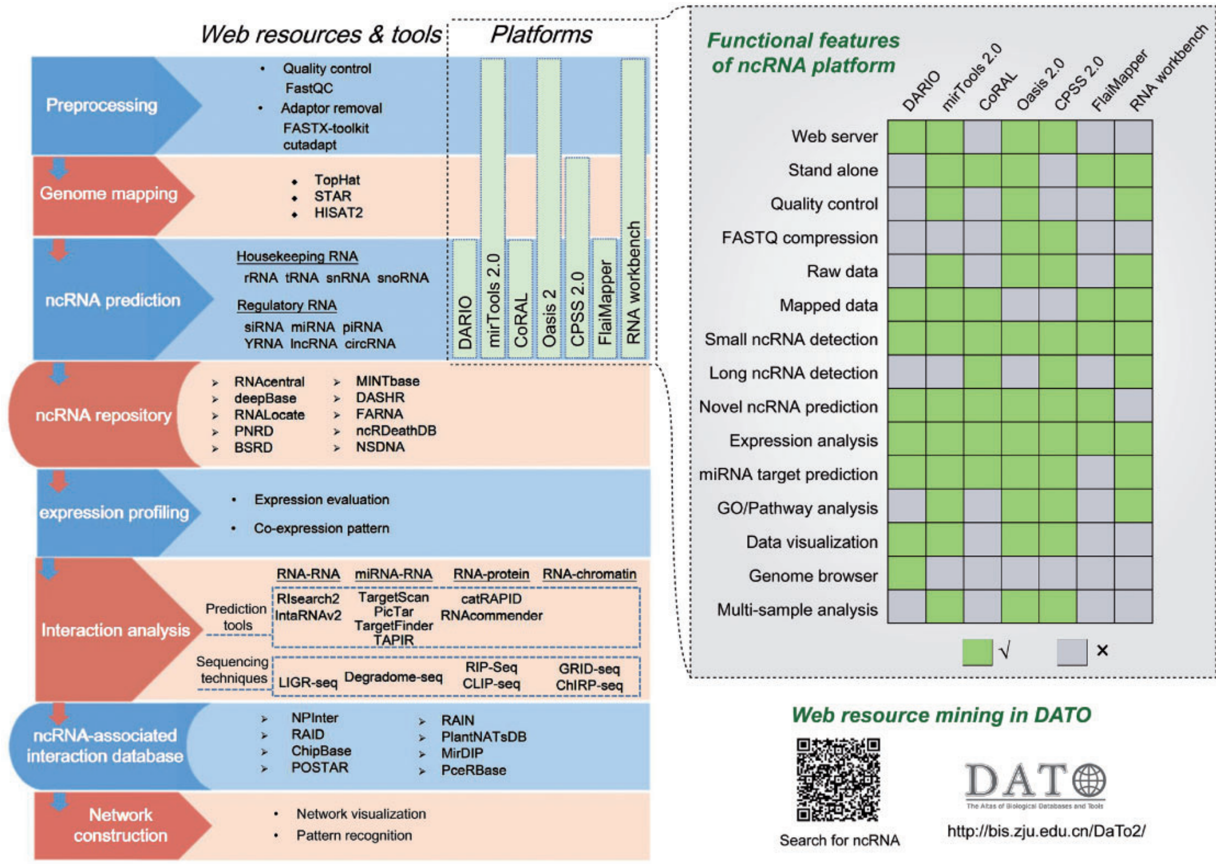

Figure 4. Bioinformatics workflow for ncRNA analysis. The left panel shows the main steps of the ncRNA analysis from RNA-seq data that involves data preprocessing, genome mapping, ncRNA prediction, expression profiling, interaction analysis and network construction. The ncRNA analysis platforms are appended beside according to their processing abilities. The right panel provides a comparison of seven ncRNA analysis platforms according to their processing abilities. Green bars refer to the feasible features of the platforms, while gray bars denote the opposite.

or multiple cases, which enables comparison of differentially expressed ncRNAs between experimental groups. The pipeline also predicts novel miRNAs, piRNAs and identifies miRNA targets with detailed function annotation.

Differentially, CoRAL [119] applies an updated multi-class classification algorithm that integrates most informative features (fragment length, cleavage specificity and antisense transcription) of each RNA class and cross-validation from ncRNA databases. Based on a machine learning-based approach, CoRAL is able to distinguish different ncRNA populations and allows detection of both small and lncRNAs. However, it is not easy to run the pipeline, as CoRAL depends on an array of bioinformatics tools that need to be installed accordingly.

Oasis 2 [120] is an improved major release of the Oasis web application for small RNA deep sequencing data analysis. Before data uploading, users can compress their FASTQ files by the platform-independent application in the web. The most selling point of Oasis is the downloadable, visually appealing and interactive output of several separated modules: quality assessment, small RNA detection and disease biomarker identification. Intriguingly, for frequent users, Oasis supports the automated submission via an Oasis' advanced programming interface (API) that can be achieved by a few lines of python scripts. Oasis 2 has optimized the speed and accuracy of sRNA detection module and expanded its analysis support of previous 14 animal species to all organisms.

CPSS 2.0 is a ready-to-use computational platform for ncRNA analysis that assembles the latest version of its dependent databases and software [121]. It currently serves as the most comprehensive and effective web server that supports genome reference of totally 48 species (vertebrates, insects, deuterostomes, nematodes and plants) and nearly all types of ncRNAs (miRNAs, known piRNAs, repeat-associated RNAs, circRNAs, lncRNAs, sRNAs, tRNAs, snRNAs and snoRNAs). It provides target prediction for miRNAs and their functional enrichment analysis. The platform allows users to modify default parameters like P-value and enrichment fold if special cases are required and multiple samples can be run easily in one go.

Notably, some small ncRNA-derived RNAs (sncdRNAs) are not merely degradation byproducts but are indeed functional and have specific maturation mechanisms. However, these sncdRNAs are often ignored in sequencing data and partially understood. To explore their quantities and characteristics, FlaiMapper [122] was presented to extract and annotate the locations of sncdRNAs based on the start and end position densities of mapped reads. The expression level of annotated fragments is also estimated.

RNA workbench [123] is an all-around platform for the analysis of RNA-based regulation. The versatile tool allows processing of multiple sequencing data such as RNA-seq, CLIP-seq and Ribo-seq. The ncRNA work-flow is incorporated in RNA workbench to detect functional structures in ncRNAs and their coding potential for small peptides. Equipped with several relevant

Table 1. List of ncRNA repositories and ncRNA interaction repositories

Databases	Data source	Stored ncRNAs/ncRNA-associated interaction	Link	Reference
ncRNA repository				
RNAcentral	Database integration	10.2 million distinct ncRNA sequences from over 720 000 organisms	http://rnacentral.org/	[124]
deepBase	SmRNA-seq and RNA-seq data sets from GEO and SRA	Diverse ncRNAs (small RNA, lncRNA and circRNA) across 19 species	http://rna.sysu.edu.cn/deepBase/	[125]
RNALocate	Database integration	High-quality RNA subcellular localization resource with 42 subcellular localizations in 65 species	http://www.rna-society.org/rnalocate/	[126]
PNRD	Database integration and text mining	11 different types of ncRNAs from 150 plant species.	http://structuralbiology.cau.edu.cn/PNRD	[127]
BSRD	RNA-seq data sets, database integration and text mining	964 experimentally validated sRNAs, 8248 sRNA homologs and 507 candidate sRNAs	http://kwanlab.bio.cuhk.edu.hk/BSRD	[128]
MINTbase	RNA-seq data sets from TCGA	Human nuclear and mitochondrial tRNA-derived fragments	http://cm.jefferson.edu/MINTbase/	[129]
DASHR	SmRNA-seq data sets from GEO and SRA	Human sncRNA genes and mature sncRNA products across 42 normal tissues and cell types	http://lisanwanglab.org/DASHR	[130]
FARNA	CAGE data sets from FANTOM5	Human ncRNA transcripts (2734 miRNA and 7555 lncRNA) in 119 tissues and 177 primary cells of human	http://cbrc.kaust.edu.sa/farna	[131]
ncRDeathDB	Database integration and text mining	Diverse ncRNAs involved in cell death system across 12 species	http://www.rna-society.org/ncrdeathdb/	[132]
NSDNA	Text mining	Experimentally supported ncRNAs associated with nervous system diseases for 11 speices	http://www.bio-bigdata.net/nsdna/	[133]
ncRNA interaction repository				
NPInter	CLIP-seq data sets, text mining, computational prediction	491 416 ncRNA–protein, miRNA–lncRNA pairs in 22 species	http://www.bioinfo.org/NPInter/	[134]
RAID	Database integration, text mining	4 million RNA–RNA interactions and > 1.2 million RNA–protein interactions across 60 species	www.rna-society.org/raid/	[135]
ChIPBase	ChIP-seq data sets	TF-miRNA, TF-lncRNA and TF-PCG interactions across 10 species	http://deepbase.sysu.edu.cn/chipbase/	[136]
POSTAR	CLIP-seq data sets, computational prediction	Experimentally probed (∼23 million) and computationally predicted (∼117 million) RBP binding sites in the human and mouse transcriptomes	http://POSTAR.ncrnalab.org	[137]
RAIN	Database integration, text mining, computational prediction	270 242 ncRNA–RNA/protein interactions in human, mouse, rat and yeast	http://rth.dk/resources/rain	[138]
PlantNATsDB	Computational prediction	2 million natural antisense transcript (NAT) pairs in 70 plant species	http://bis.zju.edu.cn/pnatdb/	[139]
MirDIP	Computational prediction	Unique miRNA-target interactions (∼48 million), comprising 2586 unique miRNAs and 27 667 unique gene	http://ophid.utoronto.ca/mirDIP/	[140]
PceRBase	Computational prediction	Potential ceRNA target-target and ceRNA target-mimic pairs from 26 plant species	http://bis.zju.edu.cn/pcernadb/index.jsp	[141]

tools (RNAcode, MAFFT, locARNA and RNAz), RNA workbench provides an easy way for users to accomplish functional analysis of any interested ncRNAs.

To have a fast grasp of different platforms, we compared 15 essential features among seven ncRNA analysis platforms (Figure 4) mentioned earlier. They include platform package, the applicability of data types and functional modules. Comparatively, Mirtools 2.0, Oasis 2 and CPSS 2.0 have a more outstanding performance that integrate currently prevalent technologies and toolkits in their web servers and provide more functional modules for ncRNA analysis.

ncRNA repository

The intense scientific interest in ncRNAs has resulted in a large number of ncRNA databases, while most of them are limited to one particular ncRNA type. The scattered data makes ncRNA search and comparison challenging and incompatible because

Table 2. List of Web resources for prediction of ncRNA-associated interactions

Web resources	Type	Description	Link	Reference
RIsearch2	Software	Suffix array-based prediction of RNA–RNA interactions and siRNA off-targets	http://rth.dk/resources/risearch	[144]
IntaRNAv2	Web server	RNA–RNA interaction prediction based on up-to-date benchmark data	http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp	[145]
catRAPID	Web server	Prediction of protein–RNA interactions and detection of RNA motifs and protein RNA-binding domains	http://s.tartaglialab.com/catrapid/omics	[146]
RNAcommender	Software	Genome-wide recommendation of RNA–protein interactions	http://rnacommender.disi.unitn.it	[147]
RPI-Pred	Web server	Prediction of protein–RNA interaction based on both the sequences and structures	http://ctsb.is.wfubmc.edu/projects/rpi-pred	[148]
Circinteractome	Web server	Prediction of RBP- and miRNA-binding sites on human circRNAs	http://circinteractome.nia.nih.gov	[149]

of the lack of a uniform way to access ncRNA information. This section summarizes several high-caliber databases that integrate multiple resources for a comprehensive ncRNA mining. Data source and stored ncRNA information are listed in Table 1.

RNAcentral [124] provides an easy access to retrieve high-quality ncRNA sequences that covers over 720 000 organisms. Modification information such as pseudouridine and methyluridine is also available for part of ncRNA sequences. Besides three search manners: ncRNA ID, ncRNA type and species, RNAcentral allows similarity search for a query sequence against its comprehensive ncRNAs sequence repository.

While RNAcentral is based on various ncRNA databases, deepBase v2.0 [125] combines both small RNA-Seq and RNA-Seq data sets and extends ncRNA analysis to expression evaluation, accurate annotation and evolutional conservation analysis, particularly for lncRNAs and circRNAs. The database covers 19 animal species currently.

RNALocate [126] is a comprehensive RNA repository organized by subcellular localization for mRNA and eight kinds of ncRNAs. It benefits ncRNA analysis, as biological function largely relies on their location at different compartments in cells. The current release covers 42 subcellular localizations in 65 species.

Plant ncRNA databases (PNRD) [127] is the extension of plant miRNA database [142]. PNRD is a plant-specific platform for multiple ncRNA-related searchings: ncRNA ID, ncRNA literature, miRNA targets and miRNA-related epigenetic modifications. Users can upload their interested sequences for novel miRNA prediction, coding potential assessment and blast against known plant ncRNAs.

BSRD [128] is a comprehensive bacterial sRNAs repository that aggregates data from both databases and literature. For each sRNAs, basic profile, secondary structure, expression pattern and their possible targets can be accessed. Additionally, sRNADeep, an RNA-Seq analysis platform, is built and implemented in the database that allows users to submit their own data sets for bacterial sRNAs characterization.

MINTbase v2.0 [129] collects both nuclear and mitochondrial tRNA-derived fragments in multiple human tissues. Based on MINITmap2 [143], a fast and exhaustive tRF mining pipeline, a total of 23 413 tRFs are identified from transcriptomic data sets in The Cancer Genome Atlas (TCGA) data sets.

Some ncRNA databases are only focused on human transcripts and expert at deep mining of specific tissues and cells, which benefits from large data sets of the human body system. Database of small human ncRNAs (DASHR) [130] offers various types of small ncRNAs across 42 normal tissues and cell types with processing information. Recently, Function Annotation of human ncRNA transcripts (FARNA) [131] provides function annotations of two key ncRNAs: miRNAs and lncRNAs. It covers 119 tissues and 177 primary cells based on the co-expression pattern of transcripts with transcription factors (TFs) and TF co-factors (TcoFs).

With the advance of ncRNA research in diseases, databases that specialized in disease-related ncRNAs are gradually emerging. ncRDeathDB [132] collects three ncRNAs (miRNAs, lncRNAs and snoRNAs) in the programmed cell death that links to many diseases. It allows users to search ncRNAs and associated interaction by three cell death pathway: apoptosis, autophagy and programmed necrosis for 12 species. NSDNA [133] comprises experimentally supported ncRNAs (miRNAs, lncRNAs, siRNAs, snoRNAs and piRNAs) in nervous system diseases for 11 species. It also constructs miRNA-NSD bipartite network based on experimentally supported relationships between miRNAs and NSDs.

ncRNA interaction resources

A complete spectrum of ncRNAs interacting partners is significant to deepen our understanding of how ncRNAs modulate biological processes. This section collects web resources, including databases that integrate diverse RNA-associated interaction data sets and bioinformatics tools for interaction prediction. Features of databases and predictors are listed in Tables 1 and 2, respectively.

NPInter [134] has updated to the third version that contains experimentally verified ncRNA-associated interactions, especially lncRNA–miRNA pairs. The interactions stored in NPInter v3.0 are curated from various sources: CLIP-seq data sets, literature mining and computational prediction. It provides ncRNA–protein pairs, miRNA–lncRNA pairs with detailed notes: co-expression values, cellular sub-location and interactive sites of pairing molecules. Moreover, functions of lncRNAs in human and mouse are predicted through lnc-GFP, which benefits from their extensive interactions.

RAID v2.0 [135] collects RNA-associated interactions across 60 species. While NPInter emphasizes on miRNA and lncRNA, RAID covers a wide range of RNA classes for RNA–RNA (protein) interactions. Data sets from nearly 20 ncRNA-related databases are curated in RAID 2.0, and confidence score of each paring is calculated. However, the current version is still lacking circRNA-associated interactions.

ChIPBase2.0 [136] takes great advantage of ChIP-seq data that expands our understanding of ncRNA–TF interactions. It integrates over 10 000 samples across 10 species for millions of TFs. Around 5 million transcriptional regulatory relationships of TF-ncRNA are extracted and annotated, mainly for miRNA and lncRNA.

CLIPdb 2 (POSTAR) [137] collects large data sets of RBP binding sites in the human and mouse transcriptomes from experiments and computational prediction. Besides protein-coding genes, the platform annotates a majority of ncRNA types, such as canonical ncRNAs (snoRNA, snRNA and tRNA), pseudogenes and lncRNAs. To provide a comprehensive post-transcriptional regulatory map, POSTAR integrates multiple regulatory events: miRNA binding, RNA modification, RNA editing and gene mutations.

RAIN [138] comprises 270 242 ncRNA–protein and ncRNA–RNA interactions. They are further associated with the PPIs in the STRING database. RAIN integrates experimentally supported interactions from several publicly available resources and interactions identified by text mining. Predicted interactions from five respective miRNA predictors are summarized and filtered by specific score thresholds. The curated knowledge is collected for nine classes of ncRNAs, namely, miRNA, lncRNA, snoRNA, snRNA, rRNA, tRNA, Y RNA, vRNA, telomerase RNA and signal recognition particle RNA. RAIN benchmarks the overall resources to assess the reliability of each interaction. Currently, the database is largely constituted of human interactions, as well as that of rat, mouse and yeast.

PlantNATsDB [139] is a comprehensive resource of natural antisense transcripts (NATs) in the plant kingdom. NAT pairs are acquired by computational prediction and divided into two groups: *cis*-NATs and *trans*-NATs. The database provides easy access to the functional investigation of each NAT pairs with rich annotation, including NAT summarization, gene information, GO annotation and sRNA expression. Notably, a graphical browser is incorporated to display the network formed by different NAT pairs. The database currently stores approximately 2 million NAT pairs in 70 plant species.

Nowadays, verified interacting partner(s) for ncRNA are far sparse, as experimental methods are expensive and labor-intensive. In this case, computational approach serves as a good alternative to elucidate their potential functional relationships. Identification of ceRNA pairs is a major part of the ncRNA-related interaction. Tools for miRNA target prediction have emerged as the boom of miRNA research [150]. Not limited to miRNA-associated interaction, RIsearch2 [144] is a novel RNA–RNA interaction predictor that enables fast discovery of interactions between two RNA molecules. It bases on a single integrated seed-and-extend framework and allows prediction on a transcriptome-wide scale. IntaRNAv2 [145] performs high predictive quality of RNA–RNA hybrids by incorporating restrictive seed constraints and interaction site accessibility. Moreover, the web server provides visualization of minimal energy profiles for interacting RNAs, which benefits further study of experimental validation.

catRAPID [146] omics is a specialized server for large-scale prediction of protein–RNA pairs based on secondary structure, hydrogen bonding and van der Waals contributions. The web server contains six tools to comprehensively estimate binding propensity of protein–RNA pairs at a genome-wide scale. While catRAPID online server allows limited sequences on one run, RNAcommender [147] is capable of suggesting RNA targets for large data sets. It basically works as a recommender system that outputs a ranking of candidate RNA targets. Sequence information of proteins (transcripts) is calculated into appropriate features to measure the similarity of binding capabilities between proteins (transcripts), which makes it possible to predict little known RBPs or transcripts. As the high-order structures of proteins and RNAs are critical to their functions, RPI-Pred [148] predicts ncRNA–protein interactions considering both sequence and structural information. Circinteractome [149] is a web tool for exploring the possible role of circRNA in sequestering RBPs and miRNAs, largely based on the prediction.

There are also several databases of predicted ncRNA-associated interactions. For example, MirDIP [140] comprises 30 miRNA targets prediction resources with reduced bias, and finally gets almost 152 million human miRNA-target results, each assigned with an integrative score. PceRBase [141], a collection of predicted plant ceRNA pairs of 26 plant species, mainly focuses on the miRNA-associated interplay among diverse RNA transcripts.

Discussion

We are now in a golden era of ncRNA transcriptome. As our understanding of the ncRNA transcriptome improves, novel species continue to emerge, which poses great challenges for precise sorting of ncRNAs. Current ncRNA catalogs are far from exhaustive and certainly contain false positives. To create accurate and comprehensive catalogs of ncRNAs, computational tools become effective alternatives that can integrate multiple features of ncRNAs. Although ncRNA repertoire has rapidly expanded, their biological function and regulation remain largely elusive. Considering the large number and functional diversity of ncRNAs, one important challenge will be the great demand of systematic and integrative annotation tools.

As ncRNA-associated interactions participate extensively in diverse physiological programs, ncRNA interactome becomes increasingly important for functional investigation. However, the lack of experimentally validated ncRNA interaction has largely hindered the development of *in silico* predictors, as they rely on the collected data to construct a well-defined training model. Given this realization, much effort is now required to develop approaches for systematically characterizing the ncRNA interactome with the aid of high-throughput sequencing. Based on the available resources, databases that integrate experimental data, literature mining and computational predictions are emerging. And the research field is expecting more high-caliber web servers with rich annotation and user-friendly interface to provide clues for follow-up experimental studies, as well as inspire novel hypotheses.

Key Points

- The big noncoding family comprises a diverse catalog of ncRNA species that exhibit a surprising range of sizes and structure.
- ncRNA transcripts account for a majority in eukaryotic RNA transcription that are shown to be derived from multiple genomic regions.
- ncRNAs participate actively in diverse biological processes via versatile interactions with other molecules.
- A comprehensive workflow for *in silico* ncRNA analysis is provided to make full use of the existing wealth of information about ncRNAs.

Funding

This work was supported by the National Key Research and Development Program of China (grant number 2016YFA0501704), National Natural Sciences Foundation of China (grant numbers 31571366 and 31771477), Jiangsu Collaborative Innovation Center for Modern Crop Production and the Fundamental Research Funds for the Central Universities.

References

1. Consortium IHGS. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860.
2. Consortium MGS. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;**420**:520.
3. Okazaki Y, Furuno M, Kasukawa T, *et al*. Analysis of the mouse transcriptome based on functional annotation of 60, 770 full-length cDNAs. *Nature* 2002;**420**:563.
4. Watson JD. The human genome project: past, present, and future. *Science* 1990;**248**:44–9.
5. Carninci P, Kasukawa T, Katayama S, *et al*. The transcriptional landscape of the mammalian genome. *Science* 2005; **309**:1559–63.
6. Consortium IH. A haplotype map of the human genome. *Nature* 2005;**437**:1299.
7. Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010;**11**:31.
8. Kaiser J. DNA sequencing. A plan to capture human diversity in 1000 genomes. *Science* 2008;**319**:395.
9. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57.
10. Djebali S, Davis CA, Merkel A, *et al*. Landscape of transcription in human cells. *Nature* 2012;**489**:101.
11. Griffiths-Jones S, Bateman A, Marshall M, *et al*. Rfam: an RNA family database. *Nucleic Acids Res* 2003;**31**:439–41.
12. Liu C, Bai B, Skogerbø G, *et al*. NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res* 2005;**33**:D112–15.
13. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2013;**42**:D68–73.
14. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA* 2014;**20**:1666–70.
15. Pavet V, Portal M, Moulin J, *et al*. Towards novel paradigms for cancer therapy. *Oncogene* 2011;**30**:1.
16. Wang J, Meng X, Dobrovolskaya OB, *et al*. Non-coding RNAs and their roles in stress response in plants. *Genomics Proteomics Bioinformatics* 2017;**15**:301–12.
17. Waterhouse PM, Graham MW, Wang M-B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci USA* 1998;**95**:13959–64.
18. Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 1999;**286**:950–2.
19. McCue AD, Slotkin RK. Transposable element small RNAs as regulators of gene expression. *Trends Genet* 2012;**28**:616–23.
20. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 1993;**75**:843–54.
21. Reinhart BJ, Slack FJ, Basson M, *et al*. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 2000;**403**:901.
22. Lagos-Quintana M, Rauhut R, Lendeckel W, *et al*. Identification of novel genes coding for small expressed RNAs. *Science* 2001;**294**:853–8.
23. Pillai RS, Bhattacharyya SN, Artus CG, *et al*. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* 2005;**309**:1573–6.
24. Yamamura S, Imai-Sumida M, Tanaka Y, *et al*. Interaction and cross-talk between non-coding RNAs. *Cell Mol Life Sci* 2017;**75**:467–84.
25. Subramanian S. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Front Genet* 2014;**5**:8.
26. Kim JK, Gabel HW, Kamath RS, *et al*. Functional genomic analysis of RNA interference in *C. elegans*. *Science* 2005;**308**: 1164–7.
27. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 2007;**318**:761–4.
28. Siomi MC, Sato K, Pezic D, *et al*. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 2011; **12**:246.
29. Lerner MR, Boyle JA, Hardin JA, *et al*. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science* 1981;**211**:400–2.
30. Wolin SL, Steitz JA. The Ro small cytoplasmic ribonucleoproteins: identification of the antigenic protein and its binding site on the Ro RNAs. *Proc Natl Acad Sci USA* 1984;**81**: 1996–2000.
31. Meiri E, Levy A, Benjamin H, *et al*. Discovery of microRNAs and other small RNAs in solid tumors. *Nucleic Acids Res* 2010; **38**:6234–46.
32. Nicolas FE, Hall AE, Csorba T, *et al*. Biogenesis of Y RNA—derived small RNAs is independent of the microRNA pathway. *FEBS Lett* 2012;**586**:1226–30.
33. Brannan CI, Dees EC, Ingram RS, *et al*. The product of the H19 gene may function as an RNA. *Mol Cell Biol* 1990;**10**: 28–36.
34. Brockdorff N, Ashworth A, Kay GF, *et al*. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 1992;**71**:515–26.
35. Brown CJ, Hendrich BD, Rupert JL, *et al*. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 1992;**71**:527–42.
36. Röther S, Meister G. Small RNAs derived from longer non-coding RNAs. *Biochimie* 2011;**93**:1905–15.
37. Kapusta A, Feschotte C. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet* 2014;**30**:439–52.
38. Sanger HL, Klotz G, Riesner D, *et al*. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc Natl Acad Sci USA* 1976;**73**:3852–6.
39. Jeck WR, Sorrentino JA, Wang K, *et al*. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013;**19**:141–57.
40. Cerutti P, Holt JW, Miller N. Detection and determination of 5, 6-dihydrouridine and 4-thiouridine in transfer ribonucleic acid from different sources. *J Mol Biol* 1968;**34**: 505–18.
41. Zieve G, Penman S. Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell* 1976;**8**:19–31.

42. Schimmel P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol* 2018;**19**:45.
43. Venkatesh T, Suresh PS, Tsutsumi R. tRFs: miRNAs in disguise. *Gene* 2016;**579**:133–8.
44. Schaefer M, Pollex T, Hanna K, *et al*. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev* 2010;**24**:1590–5.
45. Ono M, Scott MS, Yamada K, *et al*. Identification of human miRNA precursors that resemble box C/D snoRNAs. *Nucleic Acids Res* 2011;**39**:3879–91.
46. He X, Chen X, Zhang X, *et al*. An Lnc RNA (GAS5)/SnoRNA-derived piRNA induces activation of TRAIL gene by site-specifically recruiting MLL/COMPASS-like complexes. *Nucleic Acids Res* 2015;**43**:3712–25.
47. Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* 2001;**2**:986–91.
48. Ashe A, Sapetschnig A, Weick E-M, *et al*. piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* 2012;**150**:88–99.
49. Zhang ZD, Frankish A, Hunt T, *et al*. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol* 2010; **11**:R26.
50. Lee J, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* 1999;**21**:400.
51. Chen G, Cui J, Wang L, *et al*. Genome-wide identification of circular RNAs in *Arabidopsis thaliana*. *Front Plant Sci* 2017;**8**:1678.
52. Rosenbloom KR, Dreszer TR, Long JC, *et al*. ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res* 2011;**40**:D912–17.
53. Yin Q-F, Yang L, Zhang Y, *et al*. Long noncoding RNAs with snoRNA ends. *Mol Cell* 2012;**48**:219–30.
54. Zhang Y, Zhang X-O, Chen T, *et al*. Circular intronic long noncoding RNAs. *Mol Cell* 2013;**51**:792–806.
55. Ye CY, Chen L, Liu C, *et al*. Widespread noncoding circular RNAs in plants. *New Phytol* 2015;**208**:88–95.
56. Andersson R, Gebhard C, Miguel-Escalada I, *et al*. An atlas of active enhancers across human cell types and tissues. *Nature* 2014;**507**:455.
57. Rao SS, Huntley MH, Durand NC, *et al*. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;**159**:1665–80.
58. De Santa F, Barozzi I, Mietton F, *et al*. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* 2010;**8**:e1000384.
59. Kim T-K, Hemberg M, Gray JM, *et al*. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010; **465**:182.
60. Abarrategui I, Krangel MS. Noncoding transcription controls downstream promoters to regulate T-cell receptor α recombination. *EMBO J* 2007;**26**:4380–90.
61. Wyers F, Rougemaille M, Badis G, *et al*. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly (A) polymerase. *Cell* 2005;**121**:725–37.
62. Melgar MF, Collins FS, Sethupathy P. Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol* 2011;**12**:R113.
63. Koch F, Fenouil R, Gut M, *et al*. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* 2011;**18**:956.
64. Natoli G, Andrau J-C. Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* 2012;**46**:1–19.
65. Kowalczyk MS, Hughes JR, Garrick D, *et al*. Intragenic enhancers act as alternative promoters. *Mol Cell* 2012;**45**: 447–58.
66. Struhl K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* 2007;**14**:103.
67. Wu H, Nord AS, Akiyama JA, *et al*. Tissue-specific RNA expression marks distant-acting developmental enhancers. *PLoS Genetics* 2014;**10**:e1004610.
68. Lai F, Orom UA, Cesaroni M, *et al*. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 2013;**494**:497.
69. Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. *Science* 2001;**291**:1304–51.
70. Varshney D, Vavrova-Anderson J, Oler AJ, *et al*. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nat Commun* 2015; **6**:6569.
71. Tucker JM, Glaunsinger BA. Host noncoding retrotransposons induced by DNA viruses: a SINE of infection? *J Virol* 2017;**91**:e00982–17.
72. Allen TA, Von Kaenel S, Goodrich JA, *et al*. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* 2004;**11**:816.
73. Mariner PD, Walters RD, Espinoza CA, *et al*. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 2008;**29**:499–509.
74. Karijolich J, Abernathy E, Glaunsinger BA. Infection-induced retrotransposon-derived noncoding RNAs enhance herpesviral gene expression via the NF-κB pathway. *PLoS Pathog* 2015;**11**:e1005260.
75. Kapusta A, Kronenberg Z, Lynch VJ, *et al*. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* 2013;**9**:e1003470.
76. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 2017;**18**:71.
77. Gorski SA, Vogel J, Doudna JA. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol* 2017;**18**:215.
78. Sharma E, Sterne-Weiler T, O'Hanlon D, *et al*. Global mapping of human RNA-RNA interactions. *Mol Cell* 2016;**62**: 618–26.
79. Gong J, Li Y, Liu C-j, *et al*. A pan-cancer analysis of the expression and clinical relevance of small nucleolar RNAs in human cancer. *Cell Rep* 2017;**21**:1968–81.
80. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009;**136**:215–33.
81. Schirle NT, Sheu-Gruttadauria J, MacRae IJ. Structural basis for microRNA targeting. *Science* 2014;**346**:608–13.
82. Karreth FA, Reschke M, Ruocco A, *et al*. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell* 2015;**161**:319–32.
83. Cesana M, Cacchiarelli D, Legnini I, *et al*. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 2011;**147**:358–69.
84. Hansen TB, Jensen TI, Clausen BH, *et al*. Natural RNA circles function as efficient microRNA sponges. *Nature* 2013;**495**:384.
85. Piwecka M, Glažar P, Hernandez-Miranda LR, *et al*. Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science* 2017;**357**:eaam8526.
86. Salmena L, Poliseno L, Tay Y, *et al*. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;**146**: 353–8.

87. Fejes-Toth K, Sotirova V, Sachidanandam R, *et al*. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs: affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project. *Nature* 2009;**457**: 1028.
88. Taft RJ, Glazov EA, Cloonan N, *et al*. Tiny RNAs associated with transcription start sites in animals. *Nat Genet* 2009;**41**:572.
89. Seila AC, Calabrese JM, Levine SS, *et al*. Divergent transcription from active promoters. *Science* 2008;**322**:1849–51.
90. Taft RJ, Simons C, Nahkuri S, *et al*. Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. *Nat Struct Mol Biol* 2010;**17**:1030.
91. Li X, Zhou B, Chen L, *et al*. GRID-seq reveals the global RNA–chromatin interactome. *Nat Biotechnol* 2017;**35**:940.
92. Engreitz JM, Sirokman K, McDonel P, *et al*. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* 2014;**159**:188–99.
93. West JA, Davis CP, Sunwoo H, *et al*. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell* 2014;**55**:791–802.
94. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* 2016;**17**:756.
95. Hasegawa Y, Brockdorff N, Kawano S, *et al*. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* 2010;**19**:469–76.
96. Chu C, Zhang QC, Da Rocha ST, *et al*. Systematic discovery of Xist RNA binding proteins. *Cell* 2015;**161**:404–16.
97. Engreitz JM, Pandya-Jones A, McDonel P, *et al*. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 2013;**341**:1237973.
98. Kowalski MP, Krude T. Functional roles of non-coding Y RNAs. *Int J Biochem Cell Biol* 2015;**66**:20–9.
99. Zhang AT, Langley AR, Christov CP, *et al*. Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. *J Cell Sci* 2011;**124**: 2058–69.
100. Collart C, Christov CP, Smith JC, *et al*. The midblastula transition defines the onset of Y RNA-dependent DNA replication in *Xenopus laevis*. *Mol Cell Biol* 2011;**31**:3857–70.
101. Christov C, Trivier E, Krude T. Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *Br J Cancer* 2008;**98**:981.
102. Licatalosi DD, Mele A, Fak JJ, *et al*. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 2008;**456**:464.
103. Chi SW, Zang JB, Mele A, *et al*. Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* 2009;**460**:479.
104. Lafontaine DL. Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nat Struct Mol Biol* 2015;**22**:11.
105. McHugh CA, Chen C-K, Chow A, *et al*. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 2015;**521**:232.
106. Hacisuleyman E, Goff LA, Trapnell C, *et al*. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* 2014;**21**:198.
107. Du WW, Fang L, Yang W, *et al*. Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell Death Differ* 2017;**24**:357.
108. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;**25**:1105–11.
109. Dobin A, Davis CA, Schlesinger F, *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.
110. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015; **12**:357.
111. Lewis BP, Shih I-H, Jones-Rhoades MW, *et al*. Prediction of mammalian microRNA targets. *Cell* 2003;**115**:787–98.
112. Krek A, Grün D, Poy MN, *et al*. Combinatorial microRNA target predictions. *Nat Genet* 2005;**37**:495.
113. Fahlgren N, Carrington JC. miRNA target prediction in plants. In: *Plant MicroRNAs*. Berlin: Springer, 2010, 51–57.
114. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 2011;**39**:W155–9.
115. Bonnet E, He Y, Billiau K, *et al*. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 2010;**26**:1566–8.
116. Fasold M, Langenberger D, Binder H, *et al*. DARIO: a ncRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res* 2011;**39**:W112–17.
117. Patra D, Fasold M, Langenberger D, *et al*. plantDARIO: web based quantitative and qualitative analysis of small RNA-seq data in plants. *Front Plant Sci* 2014;**5**:708.
118. Wu J, Liu Q, Wang X, *et al*. mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. *RNA Biol* 2013;**10**:1087–92.
119. Leung YY, Ryvkin P, Ungar LH, *et al*. CoRAL: predicting non-coding RNAs from small RNA-sequencing data. *Nucleic Acids Res* 2013;**41**:e137.
120. Rahman R-U, Gautam A, Bethune J, *et al*. Oasis 2: improved online analysis of small RNA-seq data. *BMC Bioinformatics* 2018;**19**:54.
121. Wan C, Gao J, Zhang H, *et al*. CPSS 2.0: a computational platform update for the analysis of small RNA sequencing data. *Bioinformatics* 2017;**33**:3289–91.
122. Hoogstrate Y, Jenster G, Martens-Uzunova ES. FlaiMapper: computational annotation of small ncRNA-derived fragments using RNA-seq high-throughput data. *Bioinformatics* 2014;**31**:665–73.
123. Grüning BA, Fallmann J, Yusuf D, *et al*. The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy. *Nucleic Acids Res* 2017;**45**:W560–6.
124. RNAcentral Consortium. RNAcentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res* 2017;**45**:D128–34.
125. Zheng L-L, Li J-H, Wu J, *et al*. deepBase v2. 0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res* 2015;**44**:D196–202.
126. Zhang T, Tan P, Wang L, *et al*. RNALocate: a resource for RNA subcellular localizations. *Nucleic Acids Res* 2016;**45**:D135–8.
127. Yi X, Zhang Z, Ling Y, *et al*. PNRD: a plant non-coding RNA database. *Nucleic Acids Res* 2014;**43**:D982–9.
128. Li L, Huang D, Cheung MK, *et al*. BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Res* 2012;**41**: D233–8.
129. Pliatsika V, Loher P, Magee R, *et al*. MINTbase v2. 0: a comprehensive database for tRNA-derived fragments that includes nuclear and mitochondrial fragments from all The Cancer Genome Atlas projects. *Nucleic Acids Res* 2017;**46**:D152–9.
130. Leung YY, Kuksa PP, Amlie-Wolf A, *et al*. DASHR: database of small human noncoding RNAs. *Nucleic Acids Res* 2015;**44**: D216–22.
131. Alam T, Uludag M, Essack M, *et al*. FARNA: knowledgebase of inferred functions of non-coding RNA transcripts. *Nucleic Acids Res* 2016;**45**:2838–48.

132. Wu D, Huang Y, Kang J, *et al*. ncRDeathDB: a comprehensive bioinformatics resource for deciphering network organization of the ncRNA-mediated cell death system. *Autophagy* 2015;**11**:1917–26.
133. Wang J, Cao Y, Zhang H, *et al*. NSDNA: a manually curated database of experimentally supported ncRNAs associated with nervous system diseases. *Nucleic Acids Res* 2016;**45**: D902–7.
134. Hao Y, Wu W, Li H, *et al*. NPInter v3. 0: an upgraded database of noncoding RNA-associated interactions. *Database* 2016; **2016**:baw157.
135. Yi Y, Zhao Y, Li C, *et al*. RAID v2. 0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res* 2016;**45**:D115–18.
136. Zhou K-R, Liu S, Sun W-J, *et al*. ChIPBase v2. 0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res* 2017;**45**:D43–50.
137. Hu B, Yang Y-CT, Huang Y, *et al*. POSTAR: a platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res* 2016;**45**:D104–14.
138. Junge A, Refsgaard JC, Garde C, *et al*. RAIN: rNA–protein association and interaction networks. *Database* 2017;**2017**:baw167.
139. Chen D, Yuan C, Zhang J, *et al*. PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic Acids Res* 2011;**40**:D1187–93.
140. Tokar T, Pastrello C, Rossos AE, *et al*. mirDIP 4.1—integrative database of human microRNA target predictions. *Nucleic Acids Res* 2017;**46**:D360–70.
141. Yuan C, Meng X, Li X, *et al*. PceRBase: a database of plant competing endogenous RNA. *Nucleic Acids Res* 2016;**45**: D1009–14.
142. Zhang Z, Yu J, Li D, *et al*. PMRD: plant microRNA database. *Nucleic Acids Res* 2009;**38**:D806–13.
143. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**1**:7.
144. Alkan F, Wenzel A, Palasca O, *et al*. RIsearch2: suffix array-based large-scale prediction of RNA–RNA interactions and siRNA off-targets. *Nucleic Acids Res* 2017;**45**:e60.
145. Mann M, Wright PR, Backofen R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res* 2017;**45**:W435–9.
146. Agostini F, Zanzoni A, Klus P, *et al*. cat RAPID omics: a web server for large-scale prediction of protein–RNA interactions. *Bioinformatics* 2013;**29**:2928–30.
147. Corrado G, Tebaldi T, Costa F, *et al*. RNAcommender: genome-wide recommendation of RNA–protein interactions. *Bioinformatics* 2016;**32**:3627–34.
148. Suresh V, Liu L, Adjeroh D, *et al*. RPI-Pred: predicting ncRNA-protein interaction using sequence and structural information. *Nucleic Acids Res* 2015;**43**:1370–9.
149. Dudekula DB, Panda AC, Grammatikakis I, *et al*. CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* 2016;**13**: 34–42.
150. Fan X, Kurgan L. Comprehensive overview and assessment of computational prediction of microRNA targets in animals. *Brief Bioinform* 2014;**16**:780–94.