

REVIEW

Open Access



Principles and innovative technologies for decrypting noncoding RNAs: from discovery and functional prediction to clinical application

Yu-Meng Sun and Yue-Qin Chen* 

Abstract

Noncoding RNAs (ncRNAs) are a large segment of the transcriptome that do not have apparent protein-coding roles, but they have been verified to play important roles in diverse biological processes, including disease pathogenesis. With the development of innovative technologies, an increasing number of novel ncRNAs have been uncovered; information about their prominent tissue-specific expression patterns, various interaction networks, and subcellular locations will undoubtedly enhance our understanding of their potential functions. Here, we summarized the principles and innovative methods for identifications of novel ncRNAs that have potential functional roles in cancer biology. Moreover, this review also provides alternative ncRNA databases based on high-throughput sequencing or experimental validation, and it briefly describes the current strategy for the clinical translation of cancer-associated ncRNAs to be used in diagnosis.

Keywords: Novel ncRNAs, Sequencing technologies, Functional ncRNA discovery, Subcellular localization, ncRNA database, Diagnostic kits

Background

More than half a century after being considered as the central component in the central dogma of biology, RNA has been accepted to play various essential roles in different biological processes [1–4]. With recent developments in sequencing methods and information analysis, an increasing number of novel ncRNAs have been identified, including long noncoding RNAs (lncRNAs) [5, 6], circular RNAs (circRNAs) [7, 8], and novel small ncRNAs [9–11]. Growing studies have uncovered the characteristics of these ncRNAs, including their origins, mechanisms of generation, structures, and potential functions [6, 8, 12], which can be summarized into a

principle for the identification of known species of ncRNAs or even novel ncRNA discovery. As many ncRNAs exhibit highly tissue-specific expression patterns and important roles in biological processes related to cancer [13–19], ncRNAs have been considered as ideal therapeutic targets for cancer diagnosis and treatment [20–22]. Due to the enormous transcription potential of mammalian genomes and multiple mechanisms of ncRNA generation [8, 9, 23, 24], the ncRNA world is still full of infinite mysteries, in which unknown species of RNAs could play important roles. Technological innovation makes it possible to discover more novel functional ncRNAs.

This review focuses on the principles and innovative technologies currently available for the discovery of novel ncRNAs or functional ncRNAs within specific subcellular compartments. The particular classes of ncRNAs

* Correspondence: lscyq@mail.sysu.edu.cn

MOE Key Laboratory of Gene Function and Regulation, State Key Laboratory for Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, People's Republic of China



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

that are either novel transcripts or “old dogs” performing “new tricks” are especially emphasized. Moreover, this review also provides an overview of ncRNA-associated databases and applications of cancer-related ncRNA identification for therapeutic strategies.

Principle for novel ncRNA discovery

Early sequencing data revealed that the mammalian genome encodes many thousands of noncoding transcripts, especially those that resemble message RNAs (mRNAs) in length and splicing structure but cannot code for proteins, revealing that the world of RNA genes is far more complex than originally imagined [25]. Here, we summarized the features into a principle that could be used for the identification of known species of ncRNAs or even for novel ncRNA discovery.

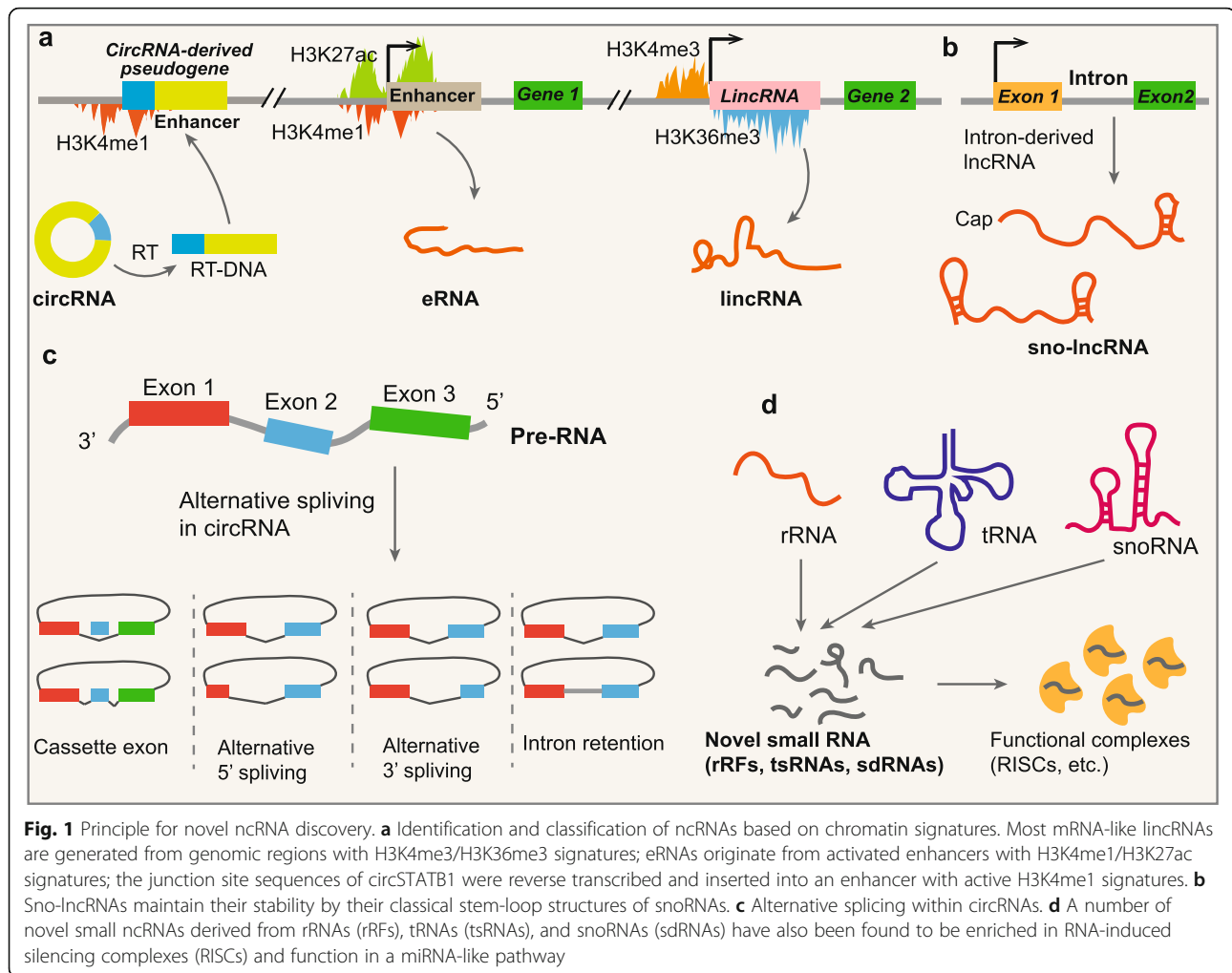
Chromatin signatures for novel ncRNA discovery

The definition of genes has become a major hurdle following the sequencing of the human genome. As histones can be modified in different ways that are indicative of the underlying DNA functional state [26–29], chromatin modifications of the corresponding genomic region could represent important biological information for the identification and classification of noncoding transcripts. The increased occurrence of trimethylation of lysine 4 of histone 3 (H3K4me3) at the promoter regions of transcripts and trimethylation of lysine 36 of histone 3 (H3K36me3) along the entire transcribed region is a signature for active transcription; these occurrences are always found at active sites of mRNA transcription [27, 28]. By searching for H3K4me3/H3K36me3 signatures that failed to overlap with known genes, there was the identification of approximately 2500 regions in the human genome and approximately 1600 regions in the mouse genome that were actively transcribed [30, 31]. However, the vast majority of these intergenic regions with H3K4me3/H3K36me3 signatures produced multi-exonic RNAs that had a little capability to encode a conserved protein; they were termed as long intergenic ncRNAs (lincRNAs) (Fig. 1a) [30, 32]. A fraction of genes encoding ncRNAs display monomethylation of lysine 4 of histone 3 (H3K4m1) and histone H3 acetylation at lysine 27 (H3K27ac), which cover their initiation sites, indicating that they are transcribed from activated enhancers as enhancer-derived RNAs (eRNAs) (Fig. 1a) [29, 33]. Although both lincRNAs and eRNAs are categorized as lncRNAs because of their lengths, distinguishing different classes of ncRNAs based on distinct chromatin modifications is necessary because specific ncRNAs generated from given gene regulatory elements could function in classic modes [34, 35]. For example, eRNAs are thought to play an important role

in regulating the 3D architecture of chromosomes near their site of transcription [34].

With developments in sequencing technologies and bioinformatics analysis, novel ncRNAs generated from alternative splicing processing or degradation of their parent RNAs have been discovered [8, 9, 36]. This kind of ncRNA does not have independent genomic regions or transcriptional regulatory elements and can be produced following parent gene transcription or degradation. Therefore, it is unable to accurately identify and describe the characteristics of these kinds of ncRNAs at the level of chromatin modification. As a typical example, circRNAs are mainly generated from alternative splicing of precursor RNA (pre-RNA), and then, they form covalently closed loop structures [8, 37]. Exonic circRNAs are produced from back-spliced exons of precursor linear RNAs, including mRNAs and lncRNAs, and they account for a major portion of the circRNA family. In addition, the intron lariats escaping from degradation can also form intronic circRNAs. Although there are some other variant forms of circRNAs, such as circular formats of small nucleolar RNAs (snoRNAs) and P RNA [38], the majority of circRNAs in humans are mainly produced from actively transcribed mRNA and lncRNA genes with H3K4me3-H3K36me3 signatures [39, 40]. Interestingly, the junction site sequences of circRNAs, such as circSTATB1 in mice, have been discovered to be inserted into an enhancer with active H3K4me1 signatures (Fig. 1a) [41]. The H3K4me1 modifications suggest that the functions of circRNAs in the regulation of enhancer and genome structure by forming pseudogenes, which may provide evidence for further classification of circSTATB1 as a retrotransposed circRNA (Fig. 1a) [41]. Although chromatin modifications cannot be used in the discovery of circRNAs, the modification signatures may be useful for more detailed classification of circRNAs.

In addition to circRNAs, there are many other novel ncRNAs that are generated from the degradation of typical transcripts from well-known genomic regions [9, 11, 42]. The excised intron-derived lncRNAs with snoRNA-like ends (sno-lncRNAs) are formed when one intron contains two snoRNA genes [42]. After splicing, the sequences between two snoRNAs escape degradation, resulting in the accumulation of certain lncRNAs. Another example is novel functional small ncRNAs, such as small ribosomal RNA-derived fragments (rRFs) [11], tRNA-derived small RNAs (tsRNAs) [9], and snoRNA-derived RNAs (sdrRNAs) [10], which are derived from “old dogs” including ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and snoRNAs. An increasing number of discoveries of novel ncRNAs have indicated the limitation of chromatin modification signatures in novel ncRNA identification. However,



chromatin signatures are still an available tool of ncRNA classification for efficient investigation of their functions.

Principles for evaluating coding potential

As ncRNAs, especially lincRNAs and circRNAs, are likely to contain open reading frames (ORFs) purely by chance, it has been a challenge to determine whether a transcript is noncoding [43]. As a growing number of studies have shown that several lincRNAs and circRNAs can produce functional micropeptides [44–47], it is necessary to evaluate the RNA coding potential of novel ncRNAs.

The lack of evolutionary conservation in identified ORFs is evidence for the absence of coding potential of ncRNAs [48, 49]. Novikova et al. reported that a human lincRNA, SRA, has different isoforms that either function at the ncRNA level or produce proteins, and there is higher evolutionary stabilization of the RNA structural core than that of the translational product under evolutionary pressure [50]. Another example is Xist, a lincRNA involved in X chromosome inactivation in

mammals that originates from the protein-coding gene *Lnx3* [51]. Interestingly, the *Lnx3* gene is still a protein-coding gene in opossum; however, it has been transformed into a noncoding transcript with frame-shifting mutations in later vertebrates [51]. In addition, the lack of homology to known protein domains and the inability to template significant protein production are the other important factors that are needed to be considered [48, 49]. These principles have been generalized to classify ncRNA coding potential by scoring conserved ORFs across diverse species with computational methods [52, 53], by searching for homology using protein-domain databases [54], and by sequencing ncRNAs associated with polyribosomes [55].

However, the coding potential of some novel ncRNAs, especially circRNAs, could fail to be determined with the principle mentioned above. Most circRNAs derived from mRNA back-splicing lose translational capacity because of the lack of effective ORFs or ribosome entry approaches, while a few circRNAs from coding or

Table 1 Characteristics of diverse sequencing methods

Classification	Techniques	Short description	Strengths of the approach	Weakness	Ref
Microarrays	Tiling arrays	A method based on probes for discovering transcripts from specific genomic regions.	This approach can provide in-depth analysis of transcripts from target regions of genome.	Suffer from potential noise as a result of weak binding or cross-hybridization of transcripts to probes.	[56]
	Microarrays	A method based on a large number of oligonucleotide probes for performing quick global or parallel expression analysis of transcriptome.	Small size and high-throughput capabilities.	This method is not able to discover novel transcripts.	[57]
RNA-seq	RNA-seq	A technique that is currently the most widespread sequencing technology for both detecting RNA expression and discovering novel RNAs.	The method provides a global high-throughput detection and identification of RNAs greater than 200 nt.	Its standard procedure is not suitable for detection of RNAs less than 200 nt. It also suffers from sequence errors at the reverse-transcription step or primer bias.	[58]
	RNA capture sequencing	A derivative technology combining RNA-seq with tiling arrays.	The method can specifically elevate the sequencing depth of target regions.	Suffer from disadvantages of both tiling arrays and RNA-seq.	[59]
scRNA-seq	Smart-seq	A scRNA-seq method based on a full-length cDNA amplification strategy.	Provide a full-length cDNA amplification of polyadenylated RNAs.	The limitations are lack of strand-specific identification, inability to read transcripts longer than 4 kb and only for polyadenylated RNAs.	[60]
	DP-seq	A scRNA-seq method using heptamer primers.	Suitable for smaller size samples or transcripts longer than 4 kb. This approach also suppresses highly expressed rRNAs in the cDNA library.	Captured RNAs are limited to polyadenylated RNAs.	[61]
	Quartz-seq	A scRNA-seq method which reduces background noise.	Reduce background noise by using specially suppression PCR primers to reduce side products.	The method is limited to detecting polyadenylated RNAs.	[62]
	SUPeR-seq	A single-cell universal polyadenylated tail-independent RNA sequencing.	Detect polyadenylated and nonpolyadenylated RNAs. Minimal rRNAs contamination.	Relatively low sensitivity for nonpolyadenylated RNAs.	[63]
	RamDA-seq	A full-length total RNA-sequencing method for analyzing single cells.	High sensitivity for nonpolyadenylated RNAs. It can also uncover the dynamics of recursive splicing.	Unknown	[64]
	Small RNA-seq	Small RNA-seq	A type of RNA-seq that discriminates small RNA from larger RNA to better evaluate and discover novel small RNAs.	Specifically detect and discover small or intermediate-sized RNAs with target sizes.	Adapter ligation bias leads to reverse transcription bias or amplification bias.
Single-cell small-RNA sequencing	Small-seq	A method which detects small RNAs in a single cell.	The method can detect small RNAs in a single cell.	The limitation may be similar to small RNA-seq.	[66]
Nascent RNA-seq	GRO-seq	A method labeling nascent RNAs with 5Br-UTP and immunoprecipitating RNAs for sequencing.	Detect nascent RNAs and provide a genome-wide view of the location, orientation, and density of Pol II-engaged transcripts.	The method is confounded by contamination due to nonspecific binding, which could possibly result in experimental bias.	[67]
	SLAM-seq	A method distinguishing nascent RNA from total RNA via s^4U -to-C conversion induced by nucleophilic substitution chemistry.	It is an enrichment-free method which can avoid contamination induced by affinity purification.	The oxidation condition caused certain oxidative damage to guanine, which may impact the accuracy of sequencing.	[68]
	TimeLapse-seq	A method distinguishing nascent RNA from total RNA via s^4U -to-C conversion induced by an oxidative nucleophilic aromatic substitution reaction.	It is an enrichment-free method which can avoid contamination induced by affinity purification.	The oxidation condition caused certain oxidative damage to guanine, which may impact the accuracy of sequencing.	[69]
	AMUC-seq	A method distinguishing nascent RNA from total RNA via transforming s^4U into a cytidine derivative using acrylonitrile.	More efficient and reliable because it has a minimal influence on the base-pairing manner of other nucleosides.	Unknown	[70]
Identification of RNA-	GRID-seq	A method that aims to comprehensively detect and	Use a bivalent linker to ligate RNA to DNA in situ and provide exact	Usable sequence length for mapping RNA is 18–23 bp. However, short	[71]

Table 1 Characteristics of diverse sequencing methods (*Continued*)

Classification	Techniques	Short description	Strengths of the approach	Weakness	Ref
chromatin interaction		determine the localization of all potential chromatin-interacting RNAs.	profiles of RNA-chromatin interactome.	sequence length can result in ambiguity in mapping.	
	iMARGI	A method providing a in situ mapping of RNA-genome interactome.	iMARGI needs less number of input cells and is suitable for paired-end sequencing.	Unknown	[72]
	ChAR-seq	A chromatin-associated RNA sequencing that maps genome-wide RNA-to-DNA contacts.	Uncover chromosome-specific dosage compensation ncRNAs, and genome-wide trans-associated RNAs.	The method needs more than 100 million input cells.	[73]
Identification of RNA-RNA interaction	CLASH	A relatively early method that uses UV cross-linking to capture direct RNA-RNA hybridization.	Avoid noise from protein intermediate-mediated interactions.	This method only detects the RNA-RNA interactions base on proteins.	[74]
	RIPPLIT	A transcriptome-wide method for probing the 3D conformations of RNAs stably associated with defined proteins.	The method can capture 3D RNP structural information independent of base pairing.	This method only detects the RNA-RNA interactions base on proteins.	[75]
	MARIO	A method identifying RNA-RNA interactions in the vicinity of all RNA-binding proteins using a biotin-linked reagent.	This method can identify RNA-RNA interactions in the vicinity of all RNA-binding proteins.	The method only detects the RNA-RNA interactions base on proteins.	[76]
	PARIS	Psoralen analysis of RNA interactions and structures with high throughput and resolution.	Directly measure RNA-RNA interactions independent of proteins in living cells.	Unknown	[77]
	LIGR-seq	A method for the global-scale mapping RNA-RNA interactions in vivo.	Provide global-scale mapping RNA-RNA interactions independent of proteins in vivo	Unknown	[78]
	SPLASH	A method providing pairwise RNA-RNA partnering information genome-wide.	Map pairwise RNA interactions in vivo with high sensitivity and specificity, genome-wide.	Unknown	[79]
	RIC-seq	RNA in situ conformation sequencing technology for the global mapping of intra- and intermolecular RNA-RNA interactions.	The method performs RNA proximity ligation in situ and can facilitate the generation of 3D RNA interaction maps.	Unknown	[80]
	RNA proximity sequencing	A method based on massive-throughput RNA barcoding of particles in water-in-oil emulsion droplets.	This method can detect multiple RNAs in proximity to each other without ligation and is fit for studying the spatial organization of RNAs in the nucleus.	Unknown	[81]
RNAs in protein complexes or subcellular structures	FISSEQ	A method that offers in situ information of RNAs at high-throughput levels.	Provide information of RNAs at high-throughput levels. Visualization.	Unknown	[82]
	CeFra-seq	A method that physically isolates subcellular compartments and identifies their RNAs.	The methods have high sensitivity for low-abundance transcripts.	The method is limited to isolation protocols and the purity of resulting isolates.	[83]
	APEX-RIP	A method can map organelle-associated RNAs in living cells via proximity biotinylation combined with protein-RNA crosslinking.	The technique can offer high specificity and sensitivity in targeting the transcriptome of membrane-bound organelles.	Unknown	[84]

noncoding transcripts could also obtain novel ORFs and may be translated into new proteins [47, 85]. The deficiency of coding-potential evaluation could be due to the incomplete circRNAs databases across diverse species, the complex mechanism of ribosome entry and translational initiation of circRNAs [86], and the lack of databases that document the information of new peptides or proteins transcribed from novel templates

containing the sequences of circRNA junction sites. Ribosome profiling has provided a strategy to identify ribosome occupancy on RNA, which has been proposed to be an available method for distinguishing noncoding transcripts from coding ones [55]. Nevertheless, some transcripts playing clear roles as ncRNAs have been detected in ribosomes, indicating that an association of RNA with a ribosome alone cannot be taken as evidence

of protein-coding potential [87, 88]. These ribosome-associated ncRNAs may serve as translational regulators or may produce nonfunctional translation noise [89, 90]. Thus, experimental technologies such as mass spectrometry proteomics have been used to improve the accuracy of noncoding transcript definition [91].

Characteristics of known ncRNAs

With the development of sequencing methods and information analysis, a vast number of diverse types of ncRNAs have been identified, such as microRNAs (miRNAs), lncRNAs, circRNAs, and novel small ncRNAs derived from well-known RNAs. Understanding the characteristics of the known ncRNAs would be helpful for novel ncRNA discovery.

NcRNAs are very heterogeneous in terms of their length and conformation [92]. They can be separated into 3 categories: (1) small ncRNAs (< 50 nt), including miRNAs (19–25 nt) [93], small interfering RNAs (siRNAs, 19–29 nt) [94], piwi-interacting RNAs (piRNAs, 25–31 nt) [95], and other functional small RNAs such as transcription initiation RNAs (tiRNAs, 17–18 nt) [96], tsRNAs (14–36 nt) [9], sdRNAs (17–24 nt or > 27 nt) [10], and sectional rRFs (15–81 nt) [11]; (2) intermediate-sized ncRNAs (50–500 nt), including 5S rRNAs (~120 nt) [97], 5.8S rRNA (~150 nt) [98], tRNAs (76–90 nt) [99], snoRNAs (60–300 nt) [100], and small nuclear RNAs (snRNAs, ~150 nt) [101]; (3) long noncoding transcripts greater than 500 nt, including linear lncRNAs [30] and circular circRNAs [40].

Most large ncRNAs, including lncRNAs and circRNAs, have been reported to be tissue-specific and expressed at relatively low levels [24, 102–104]. Different types of ncRNAs have distinct structures that maintain their stability. The most abundant lncRNAs are transcribed by RNA polymerase II (Pol II), and then, they undergo mRNA-like posttranscriptional processes, leading to 5'-caps and polyadenylated tails at their 3' ends [30]. However, studies of novel ncRNA identification that were not based on polyadenylated tails have shown the existence of nonpolyadenylated ncRNAs such as sno-lncRNAs with snoRNA-like ends and circRNAs (Fig. 1b, c) [42]. Several sno-lncRNAs have been reported to stabilize their structures by interacting with classical snoRNA binding proteins (snoRBPs) via the classical stem-loop structures of snoRNAs (Fig. 1b) [105]. In addition, circRNAs are processed to form covalently closed loop structures without open terminals, which makes them resistant to degradation by exonucleases, causing them to have relatively high stability (Fig. 1c) [8]. In contrast, most eRNAs are nonpolyadenylated transcripts that have shorter half-lives than polyadenylated lncRNAs and are difficult to discover according to their even lower levels in organisms [24, 106].

Intermediate-sized and small ncRNAs possess specifically structural features as well, such as the conversed stem-box structures of snoRNAs (C/D box or H/ACA box) [100], unique 5'-caps of snRNAs (5'-trimethylguanosine caps or 5'-monomethylphosphate caps) [101, 107], the cloverleaf-like secondary structure of tRNA [99], and hairpin loop of miRNA precursor. Most types of intermediate-sized and small ncRNAs do not have specific modification at the 5' or 3' ends, and they maintain their stabilities via binding specific proteins to form complexes. For example, snoRNAs stabilize their structures by interacting with classical snoRBPs via the classical stem-loop structures [108]. Another example is miRNA, whose precursor yields a miRNA:miRNA duplex with Dicer processing [109]. In most cases, only one strand of the duplex is usually incorporated into the RNA-induced silencing complex (RISC) to exist and function, and the other free strand is normally degraded. Together, RNA structures could affect their expression levels in cells, which always influences the discovery of potential novel ncRNAs.

Principle and strategy for identification of novel ncRNAs

Nowadays, increased types of ncRNAs have been detected and identified by the development of next-generation sequencing (NGS) [58], which can be roughly divided into the process sections of sample preprocessing, library preparation, sequencing, and bioinformatics. Importantly, it should be noted that the ways of RNA isolation and library preparation greatly affect the detection of target species of ncRNAs.

Organic reagent method using isothiocyanate/phenol/chloroform or Trizol (Invitrogen) is an universal RNA extraction way to obtain total RNA containing small and intermediate-sized RNA. However, it has been reported that phenol contamination has influences on RNA yields and subsequent sequencing [110]. Spin column chromatography using commercial kits without phenol can avoid this contamination and obtain relatively high-quality RNA from the same samples. However, silica-based spin column chromatography fails to efficiently capture RNA shorter than 200 nt, which leads to massive loss of small and intermediate-sized ncRNAs and makes the way unsuitable for small RNA-seq [111, 112]. In contrast, the ways using spin column that can capture all RNA greater than 10 nt can be selected when we aim to obtain total ncRNAs or specifically enrich small ncRNAs. Choosing appropriate ways of RNA extraction is important for identification of novel ncRNAs with a specific size.

Library with appropriate RNA selection/depletion is also pivotal in the detection of specific types of ncRNAs. In library preparation for mRNA sequencing, RNAs with polyadenylated tails are specifically isolated by

hybridization with poly(dT) oligomers from nonpolyadenylated RNAs which include a vast number of rRNAs. As a part of lncRNAs do not have polyadenylated tails, polyadenylated tail selection can only capture mRNA-like lncRNAs [113]. As for total lncRNA sequencing, library preparation is generally dependent on rRNA depletion methods. Next, the filtered RNAs are fragmented, reverse transcribed into cDNA by random primers, and undergo end repair, sequencing adaptor ligation, and size selection for subsequent sequencing. In this way, not only lncRNA but also mRNA, circRNA, and a part of intermediate-sized ncRNAs can be detected. However, reverse transcription (RT) by random priming and size selection leads to the deficiency of small ncRNAs such as miRNAs [114]. Depletion of linear RNAs by Rnase R treatment for circRNA sequencing and separation of RNAs with specific size by gel electrophoresis can specifically enrich target types of ncRNAs for RNA-seq, which are as far as possible to reduce interference signal from other transcripts. In addition, due to the shortened size, small RNA is hard to be successfully acquired through cDNA synthesis (first or second cDNA synthesis) by random priming and be always removed by size selection after sequencing adaptor ligation [114]. Thus, in small RNA-seq, both ends of the RNA fragments are firstly ligated to the adapters and followed by the cDNA synthesis and library construction. We also need to pay attention to the effects of RNA modifications on library preparation, which usually influence adapter ligation. For example, 5' caps of snRNAs should be removed before adapter ligation. Selecting appropriate methods of library preparation is also important for identification of novel ncRNAs [101, 107].

It is worth noting that alternative splicing processes enable great complexity in transcripts from the same genomic regions [115]. For linear ncRNAs, various isoforms can be relatively easy to identify by RNA-seq. Nevertheless, despite the identification of circRNAs based on the junction site, extra sequence identification is still needed to determine the actual sequences of circRNAs because of potential circRNA variants being generated from a single gene locus [116]. This issue results from alternative splicing that occurs within circRNAs with multiple exons (Fig. 1c) [116]. All four basic types of canonical alternative splicing were found to occur in circRNAs as well: cassette exon, intron retention, alternative 5' splicing and alternative 3' splicing (Fig. 1c) [116]. For example, the human *XPO1* gene locus has been demonstrated to contain a circRNA-predominant cassette exon, the *CAMSAP1* gene locus generates two circRNA isoforms with or without a retained intron, and the human *EIF3J* and *PAIP2* gene loci can also produce circRNAs containing both exon

and intron sequences [104, 117, 118]. Other factors, such as read-through transcription and the fusion of genes derived from chromatin rearrangement, also generate read-through circRNAs and fusion circRNAs, respectively, which increase the diversity of ncRNAs [119, 120].

Traditionally well-known small noncoding RNAs, including miRNAs, siRNAs, and piRNAs, function in concert with the Argonaute (Ago) family of proteins to regulate gene expression at the level of transcription, mRNA stability, or translation [121, 122]. Interestingly, sdRNAs were initially discovered from an analysis of small RNAs associated with human Ago1 and Ago2 revealed by immunoprecipitation and RNA-seq (Fig. 1d) [10]. In addition, a number of novel small ncRNAs derived from both rRNAs (rRFs) and tRNAs (tsRNAs) have also been found to be enriched in RNA-induced silencing complexes (RISCs), and they function in a miRNA-like pathway (Fig. 1d) [9, 11, 36]. Immunoprecipitation of members of the Ago family proteins followed by small RNA-seq has revealed a series of novel small ncRNAs that play roles in RNA-induced target gene silencing. These data suggested that functional ncRNAs in well-known complexes should have more extensive sources and that transcripts derived from canonical DNA regions could have functions in addition to their classical ones by interacting with nonclassical RNA binding proteins (RBPs) or being located in novel complexes. This method of identifying RNA found in specific complexes or associating with subcellular components followed by RNA-seq represents an ideal way to discover new species of functional small ncRNAs. For example, the Vault complex, a novel ribonucleoprotein that probably functions in the nuclear export of large molecules, was isolated and characterized in 1986 [123]. By analyzing the components of Vaults, researchers discovered a novel and single species of small ncRNAs that is 86-141 nt in length, which was termed Vault RNAs (vRNAs) [124]. vRNAs that are derived from *VTRNA* genes by RNA polymerase III (Pol III) have been reported to be associated with multidrug resistance and, interestingly, also be the origin of miRNA-like small ncRNAs processed by Dicer [125]. Another example of identification of RNAs in complexes is snoRNAs, whose canonical functions are generally considered to guide the pseudouridylation and 2'-O-methylation of rRNA in the nucleolus [126]. However, in situ global RNA interactions with DNA identified by immunoprecipitation and RNA-seq showed that snoRNAs represent a vast population and a high enrichment in the chromatin-bound fractions, suggesting the other potential functions of these well-known small ncRNAs located in the nonclassical complexes [71, 127, 128].

Lack of sequence conservation, low level or high tissue-specific expression pattern, or derivation from

canonical DNA sequences are potential factors that make the discovery and identification of novel ncRNAs difficult. We provided the identification principle of recently discovered functional ncRNAs, which would be a referential principle for novel ncRNA discovery. Importantly, recent technological developments, especially specific sequencing technological developments, have provided multiple approaches for the discovery and study novel ncRNAs.

Approaches for discovering ncRNAs

Most ncRNAs, such as lncRNAs and circRNAs, have the characteristics of spatiotemporal specificity and low expression levels, which make it difficult to identify them [24, 102–104]. Therefore, it is necessary for us to purposefully choose the appropriate methods in sample preparation and sequencing techniques. Here, we will review innovative and novel sequencing methods that significantly improve the process of RNA identification and investigation, placing special emphasis on their advantages and limitations (Table 1).

Tiling arrays and microarrays

Tiling array is an alternative and classic method for discovering RNA [56]. This approach hybridizes complementary DNAs (cDNAs) to microarray slides containing tiled oligonucleotide probes that are designed to hybridize with nonrepetitive sequences of specific genomic regions or the entire genome [56]. For example, tiling arrays were used to specifically identify the potential transcripts from four human *HOX* gene clusters with 400,000 probes, leading to the discovery of intergenic ncRNAs, including the well-known lncRNA *HOX* antisense intergenic RNA (*HOTAIR*) [129]. Tiling arrays can also provide in-depth analysis of alternative splicing, polymorphism, and novel transcription site identification by elevating the resolutions of designed probes [56, 130]. Nevertheless, because microarrays suffer from potential noise as a result of weak binding or cross-hybridization of transcripts to probes, tiling arrays have been replaced by NGS technologies and now preferably serve as a supplemental step for RNA-seq to increase the sequencing depth of target regions.

Microarray is an important method for performing quick global or parallel expression analysis of the transcriptome in different cell/tissue types, experimental systems, developmental stages, or pathological conditions [57]. This classic method consists of a large number of oligonucleotide probes spotted on a solid surface that are then allowed to hybridize to target sequences from samples, which are further detected by fluorescently labeled target sequences. The intensity of fluorescence is used to quantify target sequences. Their small size and high-throughput capabilities have brought

microarrays to the forefront of RNomic research. However, this approach can only detect RNAs whose sequences are known and have specific hybridization probes; this method is not able to discover novel transcripts.

RNA-seq

RNA-seq is currently the most widespread sequencing technology for both detecting RNA expression and discovering novel species of ncRNAs (Fig. 2a) [24, 58]. In addition, this approach can also be used to identify single nucleotide polymorphisms, alternative splicing isoforms, gene fusion events, and novel splice junctions [131–134]. RNA-seq is based on the conversion of RNA into a pool of cDNA with either oligo (dT) primers or random primers, depending on the purpose of the sequencing. However, because cDNA libraries prepared with oligo (dT) selectively enrich for polyadenylated RNA and simultaneously deplete nonpolyadenylated and partially degraded transcripts, RNA-seq with random primers for cDNA synthesis on rRNA-depleted transcripts is currently a more widely used approach. Analysis of human or mouse cell types using RNA-seq revealed the presence of more than 8000 human and over 1000 mouse long intergenic ncRNAs (lincRNAs), the majority of which had not been previously identified [32, 135]. Interestingly, in one study using RNA-seq for the specific identification of nonpolyadenylated RNA, a novel species of lncRNAs with snoRNA-like ends was discovered to be produced from excised introns [42, 105]. Moreover, the first identification of large numbers of circRNAs in humans and mice occurred following the combination of RNA-seq and RNase R treatment, which uncovered the effective presence of 1950 human and 1903 mouse circRNAs in human cell lines (HEK293 and leukocytes) and mouse tissues such as the brains and fetal heads [7]. In addition, RNA-seq with specific preparation for small RNA identification is also the primary approach for discovering and detecting miRNAs, snoRNAs, piRNAs, and other novel small ncRNAs, including IRFs, tsRNA, and sdrRNAs [10, 126, 136–139].

There is a derivative technology based on RNA-seq, RNA capture sequencing, which is combined with tiling arrays to elevate the sequencing depth of target regions [59]. In brief, tiling arrays are performed first with specific oligonucleotide probes to enrich cDNAs from specific genomic regions. Second, the hybridized cDNAs are eluted and then sequenced by RNA-seq. RNA capture sequencing increases the sequencing depth in specific genomic regions compared to RNA-seq and has uncovered multiple unannotated isoforms of both mRNAs and ncRNAs, including a novel alternative splicing transcript of *HOTAIR* that lacks the binding domain for the polycomb repressive complex (PRC2) [59].

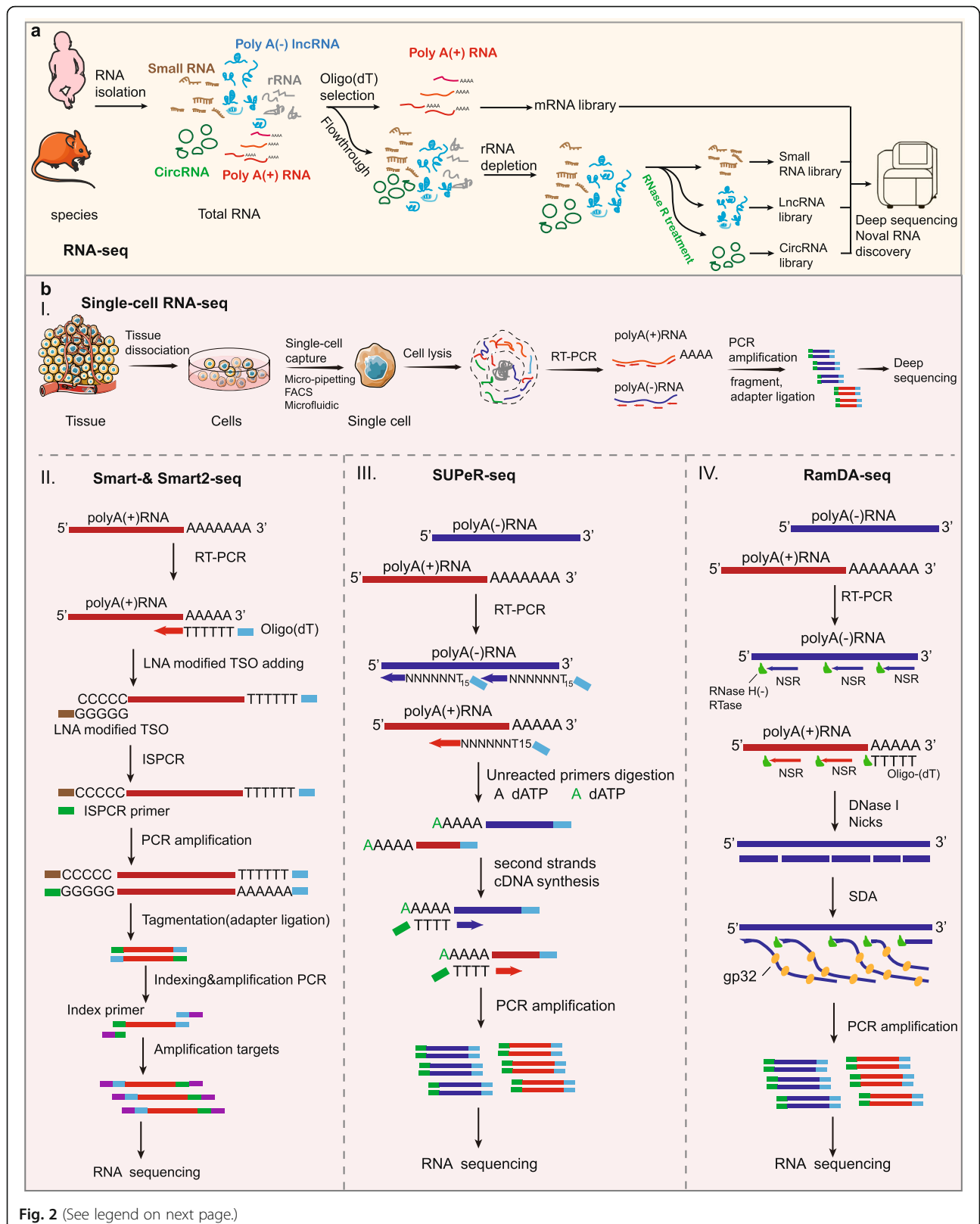


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Technologies for novel ncRNA discovery. **a** Process diagrams of RNA-seq. RNA-seq with purposeful experimental treatments can be used to detect diverse species of ncRNAs, including lncRNAs, circRNAs, and small ncRNAs. **b** Process diagrams of scRNA-seq. (I) The schematic of single-cell RNA-seq. Single cells are isolated and lysed to release total RNAs. RNAs are then reverse transcribed into first-strand cDNAs using designed primers followed by amplification for RNA-seq. (II–IV) The detailed schematic of innovative and novel methods such as Smart-seq (II), SUPeR-seq (III), and RamDA-seq (IV). In Smart-seq, polyadenylated RNAs are reverse transcribed into a pool of cDNAs by oligo (dT) primers followed by adding nontemplate C nucleotide tails to the 3' ends (II); however, SUPeR-seq uses random primers with fixed anchor sequences for cDNA synthesis, followed by adding poly(A) tails to the 3' ends (III). (IV) RamDA-seq uses both oligo (dT) and random primers for cDNA synthesis. cDNA is synthesized by the RNA-dependent DNA polymerase activity of RNase H minus reverse transcriptase (RTase). DNase I selectively nicks the cDNA of the RNA:cDNA hybrid strand. The 3' cDNA strand is displaced by the strand displacement activity of RTase mediated by the T4 gene 32 protein (gp32), starting from the nick randomly introduced by DNase I. cDNA is amplified and protected by gp32 from DNase I. NSR: not-so-random primer

Over the years, many technologies based on basic RNA-seq have been developed to identify RNAs at the transcriptome scale, some of which will be discussed in the following sections. It is inferred that advanced algorithms for analysis of sequencing data are also likely to promote transcriptome analysis. Nevertheless, RNA-seq may suffer from disadvantages such as the introduction of sequence errors at the reverse-transcription step or primer bias, which require further optimization [140].

Small RNA-seq and single-cell small-RNA sequencing

Because sample preparation for RNA-seq is not suitable for small RNAs, such as reverse transcription with random priming (short RNA species yield even shorter cDNAs that are not long enough for efficient alignment), small RNA-seq with modified library preparation, such as miRNA-seq, was developed [65, 114, 141]. Small RNA-seq is a type of RNA-seq that discriminate small RNAs from larger RNAs to better evaluate and discover novel small RNAs [65]. In this method, small RNAs are fractionated by gel electrophoresis, and then, universal adapters are ligated to the both ends of RNA fragments, which are acted as primer binding sites during reverse transcription and PCR amplification. Previous studies using small RNA-seq detect specific expression profiles of miRNAs in various biological processes and cancer; reveal asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNA, and tRNAs; and even provide evidence for human miRNA-offset RNAs [65, 142, 143]. Although adapter ligation bias which lead to reverse transcription bias or amplification bias still need to be optimized [144, 145], small RNA-seq currently remains a high-efficiency way to detect and discover novel small ncRNAs.

A recent study provided a method to detect small ncRNAs in a single cell and the method was named as Small-seq [66]. In brief, single cell is lysed, and 5.8S rRNA is masked with a complementary oligonucleotide during adapter ligation. Then 3' adapters are ligated to small RNAs, and unligated adapters are subsequently digested. The 5' adapters containing a unique molecular

identifier (UMI) are ligated, and reverse transcription is carried out. In the original article, the method captured a complex set of small RNAs, including miRNAs, fragments of tRNAs, and snoRNAs [66].

Single-cell RNA sequencing (scRNA-seq)

The fundamental unit of an organism is a single cell. Along with in-depth studies on development and disease occurrence, there is a growing sense that some single cells possess nonnegligible abilities that can affect organic growth or lead to the downfall of the entire organism [146]. It is helpful for researchers to further understand the mechanisms of growth or disease progression by revealing the gene expression pattern of specific single cells. However, the sample sizes from a single cell are insufficient for general RNA-seq, which has led to the development of scRNA-seq methods (Fig. 2b(I)). In addition, scRNA-seq techniques are also appropriate for other small samples, such as limited clinical patient samples or cells sorted with fluorescence-activated cell sorting (FACS) [61, 147].

Previous scRNA-seq techniques include Smart-seq [60, 148], designed primer-based sequencing (DP-seq) [61], and Quartz-seq [62], and each of them exhibits prominent advantages and disadvantages. Smart-seq is a method based on a full-length cDNA amplification strategy (Fig. 2b(II)) [60]. In this approach, polyadenylated RNAs are reverse transcribed into a pool of cDNAs by oligo (dT) primers and Moloney murine leukemia virus reverse transcriptase (MMLV RT). As a result, the terminal transferase activity of MMLV can add several nontemplate C nucleotides to the 3' ends of the reverse transcribed products when the reverse transcription reaction reaches the 5' end of a template transcript during first-strand cDNA generation (Fig. 2b(II)). Then, the poly-cytidine overhangs are used to complete the double-strand cDNA generation, which ensures that the prepared library for scRNA-seq only contains full-length cDNAs. However, the lack of strand-specific identification and inability to read transcripts longer than 4 kb partly limit the application of this method [149].

Compared to Smart-seq, DP-seq shows the advantage of being to amplify RNAs from smaller size samples, as low as 50 pg, or from transcripts longer than 4 kb [61]. DP-seq uses a defined set of heptamer primers, which target regions less likely to form secondary structures and reside upstream of the unique regions on certain transcriptomes, and they amplify the majority of expressed transcripts from a limited number of RNAs [61]. In the original study, preparation of a DP-seq library successfully amplified over 80% of the mouse transcriptome with 44 heptamer primers. Moreover, this approach can also suppress highly expressed rRNAs in the cDNA library and is able to detect transcripts at relatively low levels [61]. In addition, Quartz-seq is an alternative scRNA-seq method with reduced background noise that utilizes specially designed suppression polymerase chain reaction (PCR) primers to reduce the generation of unwanted side products [62].

Recent studies on scRNA-seq methods preferably focused on total RNA sequencing, which provided rich information on biological systems in addition to the abundance of mRNAs. Thus far, much efforts have been made to develop scRNA-seq techniques with full-length coverage or sensitivity to nonpolyadenylated RNAs. There are several scRNA-seq methods, such as Smart-seq, that can provide full-length coverage of transcripts [60]. Nevertheless, these methods fail to measure nonpolyadenylated transcripts due to oligo (dT) priming [60]. Single-cell universal poly(A)-independent sequencing (SUPeR-seq), which uses random primers with fixed anchor sequences to replace oligo (dT) primers for cDNA synthesis, has been reported for the detection of nonpolyadenylated RNAs, especially circRNAs, in a single cell with robust precision and accuracy (Fig. 2b(III)) [63]. In the original study, researchers discovered 2891 circRNAs and 913 novel linear RNAs in mouse preimplantation embryos using SUPeR-seq and deciphered regulation mechanism of circRNA during early embryonic development [63]. However, SUPeR-seq also exhibits relatively low sensitivity for nonpolyadenylated RNAs [64].

Random displacement amplification sequencing (RamDA-seq) is a full-length total RNA-sequencing method for analyzing single cells, but it has a high sensitivity for nonpolyadenylated RNAs [64]. This approach can measure not only polyadenylated but also nonpolyadenylated RNAs, including nascent RNAs, lncRNAs, circRNAs, and eRNAs, and it can uncover the dynamics of recursive splicing [64]. Furthermore, it can provide full-length coverage for extremely long transcripts (more than 10 kb). RamDA-seq simplifies the experimental procedure to amplify cDNA as early as possible by using a novel RT technology, RT with random displacement amplification (RT-RamDA), which aims to obtain higher capture efficiency of RNAs and global cDNA

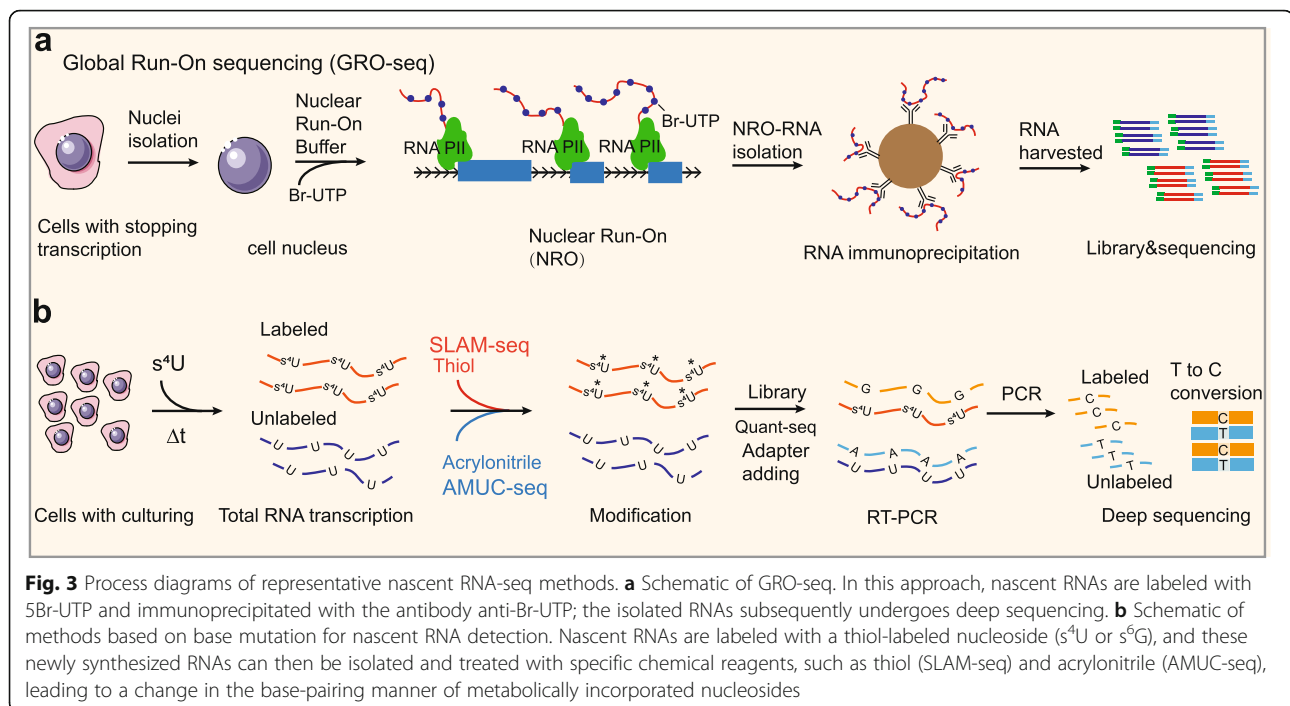
amplification for further sequencing (Fig. 2b(IV)). Moreover, not-so-random primers (NSRs) are used to enable random priming while preventing the synthesis of cDNA from rRNAs [64]. Analysis of mouse embryonic stem cells undergoing differentiation using RamDA-seq revealed the cell state-dependent expression of known and novel nonpolyadenylated RNAs, including nonpolyadenylated isoforms of the lncRNA Neat1, and the specific genome-wide eRNA expression patterns in single cells [64].

Nascent RNA-seq

RNA-seq is a revolutionary tool for transcriptome profiling that provides information on the dynamic changes of gene expression against different conditions or after exposure to different stimuli [58]. However, the traditional RNA-seq technique is generally performed to determine steady-state RNA levels, and changes in RNA transcription and decay rates cannot be easily distinguished [150]. Moreover, common RNA-seq also fails to provide efficient temporal information on RNA kinetics [150]. To address these issues, new sequencing methods for measuring nascent transcripts, as opposed to total RNAs, have been developed [151].

Nascent RNA-seq can reveal the temporal information of gene expression changes. Metabolic labeling and affinity purification of labeled nascent RNAs followed by RNA-seq is a well-known approach for analyzing nascent RNAs [151]. For example, global run-on sequencing (GRO-seq) labels nascent RNAs with 5Br-UTP, enabling labeled nascent RNAs to be immunoprecipitated with the antibody anti-Br-UTP; the isolated RNAs subsequently undergoes deep sequencing (Fig. 3a) [67]. By sequencing nascent RNAs, GRO-seq can also provide a genome-wide view of the location, orientation, and density of Pol II-engaged transcripts, revealing divergent transcription at active promoters that yield antisense ncRNAs [152]. In recent studies, labeling/purifying RNA analysis has also been used to detect nascent ncRNAs, including nascent circRNAs. Nevertheless, the conventional purification assay in GRO-seq is confounded by contamination due to nonspecific binding, which could possibly result in experimental bias [70].

Recently, innovative enrichment-free methods for nascent RNA detection have been developed, which avoid contamination induced by affinity purification [153]. These methods directly distinguish nascent RNA from total RNA in single-base resolution by marking the mapping reads of nascent RNAs with introduced base mutations. In brief, nascent transcripts are labeled by adding a thiol-labeled nucleoside (s^4U or s^6G) to cell culture media, and these newly synthesized RNAs can then be isolated and treated with specific chemical reagents, leading to a change in the base-pairing manner of



metabolically incorporated nucleosides (Fig. 3b) [153]. For example, SLAM-seq uses nucleophilic substitution chemistry to induce s⁴U-to-C conversion in an RT-dependent manner [68], and TimeLapse-seq employs s⁴U-to-C conversion via an oxidative nucleophilic aromatic substitution reaction (Fig. 3b) [69]; however, this oxidation condition caused certain oxidative damage to guanine [69]. A recent study reported an improved method, AMUC-seq, which transformed s⁴U into a cytidine derivative using acrylonitrile (Fig. 3b) [70]. Compared to other enrichment-free methods for nascent RNA detection, AMUC-seq has been reported to be more efficient and reliable because it has a minimal influence on the base-pairing manner of other nucleosides and can quantitatively analyze RNA at the transcriptome scale [70].

Innovative techniques based on RNA location and interactome for functional ncRNA discovery

As discussed above, the vast majority of the human genome can be transcribed into ncRNAs; thus, it is important to reveal potentially functional ncRNAs that may play a role in certain biological processes, especially in cancer occurrence and development. It has been shown that ncRNAs are commonly folded into highly ordered structures that play a role within their interactome [154, 155]. Therefore, in this section, we will discuss the discovery and identification of functional ncRNAs based on their interaction networks and subcellular location levels, and we will provide some novel

techniques that can be used to screen purposefully for functional ncRNAs.

RNA-chromatin interaction

An increasing number of studies have reported that diverse species of ncRNAs show regulatory functions in different layers of and gene expression. Many ncRNAs perform direct actions on chromatin, some of which may mediate genomic interactions predominantly in *cis*, whereas others are capable of acting extensively in *trans* [156–158]. These findings suggest a common role of specific RNA-chromatin interactions in modulating gene expression. Global RNA interactions with DNA by deep sequencing (GRID-seq) is a method that aims to comprehensively detect and determine the localization of all potential chromatin-interacting RNAs [71]. This approach uses a bivalent linker to ligate RNA to DNA *in situ* in fixed nuclei (Fig. 4a). Briefly, cells are fixed with disuccinimidyl glutarate (DSG) and formaldehyde first to stabilize RNAs on chromatin. Then, nuclei are extracted, and DNA is digested *in situ* by the frequent 4-base cutter AluI. A specifically designed bivalent linker labeled by biotin that consists of single-stranded RNA (ssRNA) portions, to ligate RNA, and a double-stranded DNA (dsDNA) portion, to ligate DNA, is used to link RNAs to AluI-digested genomic DNAs. Then, the DNA-RNA complexes are purified, filtered, and sequenced. In the original article, GRID-seq performed in human, mouse, and *Drosophila* cells revealed a large set of tissue-specific coding and noncoding RNAs that bind to active

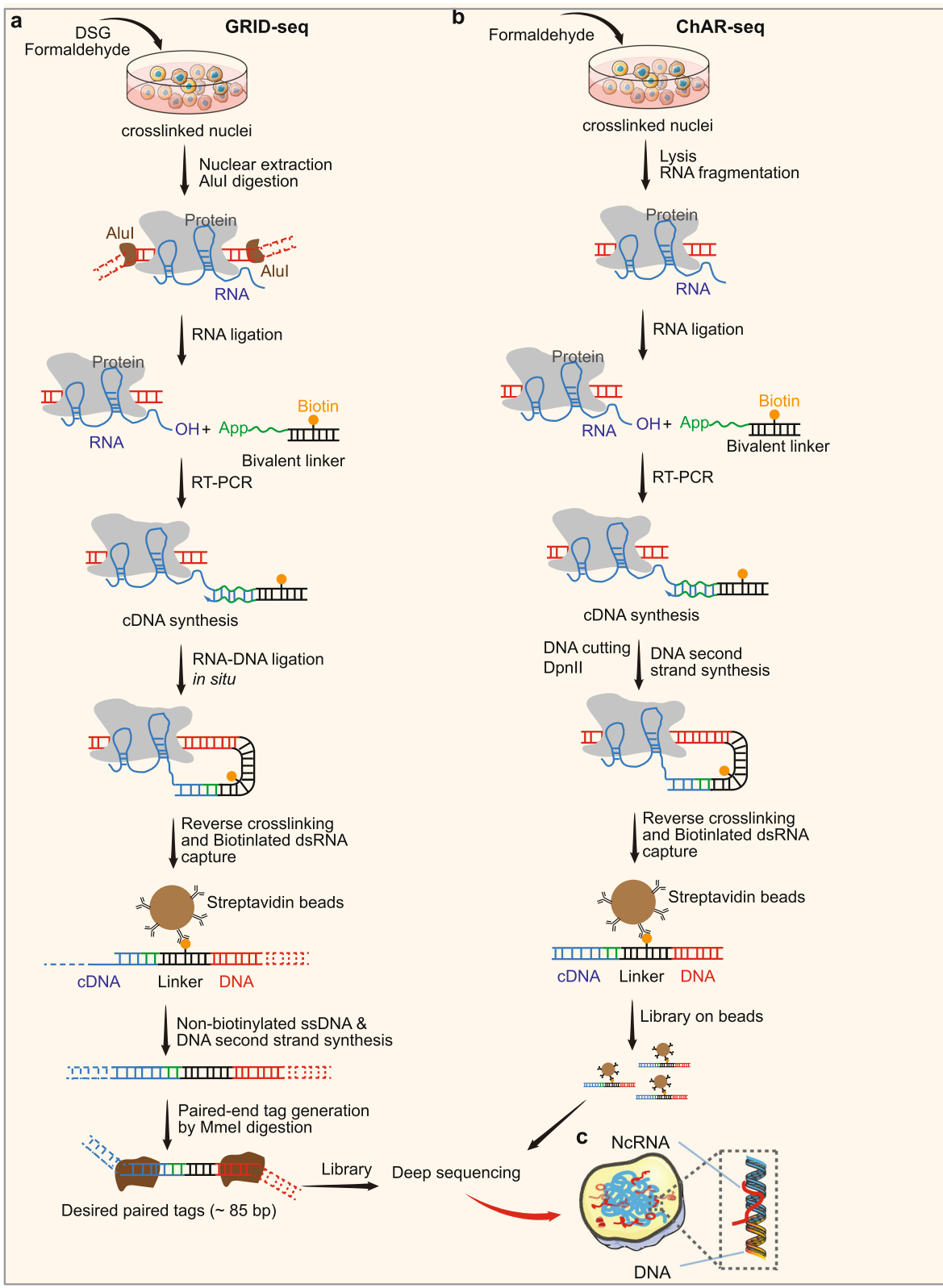


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Technologies for discovery of RNA-chromatin interaction. **a** Process diagrams of GRID-seq. Cells are fixed with disuccinimidyl glutarate (DSG) and formaldehyde. Then, nuclei are extracted, and DNA is digested in situ by the frequent 4-base cutter AluI. A specifically designed bivalent linker labeled by biotin that consists of single-stranded RNA (ssRNA) portions, to ligate RNA, and a double-stranded DNA (dsDNA) portion, to ligate DNA, is used to link RNAs to AluI-digested genomic DNAs. DNA ligation to AluI-digested genomic DNA are performed in situ followed by affinity purification on streptavidin beads. Then, ssDNA are released from the beads, generated into dsDNA, cleaved by a type II restriction enzyme MmeI and sequenced. **b** Overview of the ChAR-seq method. RNA-DNA contacts are preserved by crosslinking, followed by in situ ligation of the 3' end of RNAs to the 5' end of the ssDNA tail of a bivalent linker containing biotin and a DpnII-complementary overhang on the opposite end. After generating a strand of cDNA complementary to the RNA, the genomic DNA is then digested with DpnII and then re-ligated, capturing proximally associated bridge molecules and RNA. The chimeric molecules are reverse transcribed, purified, and sequenced

promoters and enhancers, especially super-enhancers [71]. Interestingly, the study also exhibited a large number of snoRNAs interacting with chromatin, suggesting possibly important roles of snoRNAs at the chromatin level [71].

Other alternative techniques based on the ligation of RNA to DNA have been reported for detecting genome-wide RNA-chromatin interactions, including MARGI and its improved version iMARGI [72, 159], and chromatin-associated RNA sequencing (ChAR-seq) [73]. Analysis of chromatin-associated RNA (caRNA) sequencing by MARGI and iMARGI revealed that caRNAs not only are associated with genomic regions where they are generated (proximal interactions) but also are attached to distal genomic regions (distal interactions) on the same chromosomes or on other chromosomes (inter-chromosomal interactions) [72, 159]. Interestingly, transcription start sites (TSSs) were identified as the preferred genomic regions targeted by chromatin-associated ncRNAs through distal or interchromosomal interactions. ChAR-seq also uncovered a range of chromatin-associated RNAs, especially chromosome-specific dosage compensation ncRNAs, and genome-wide trans-associated RNAs, which are involved in cotranscriptional RNA processing (Fig. 4b) [73].

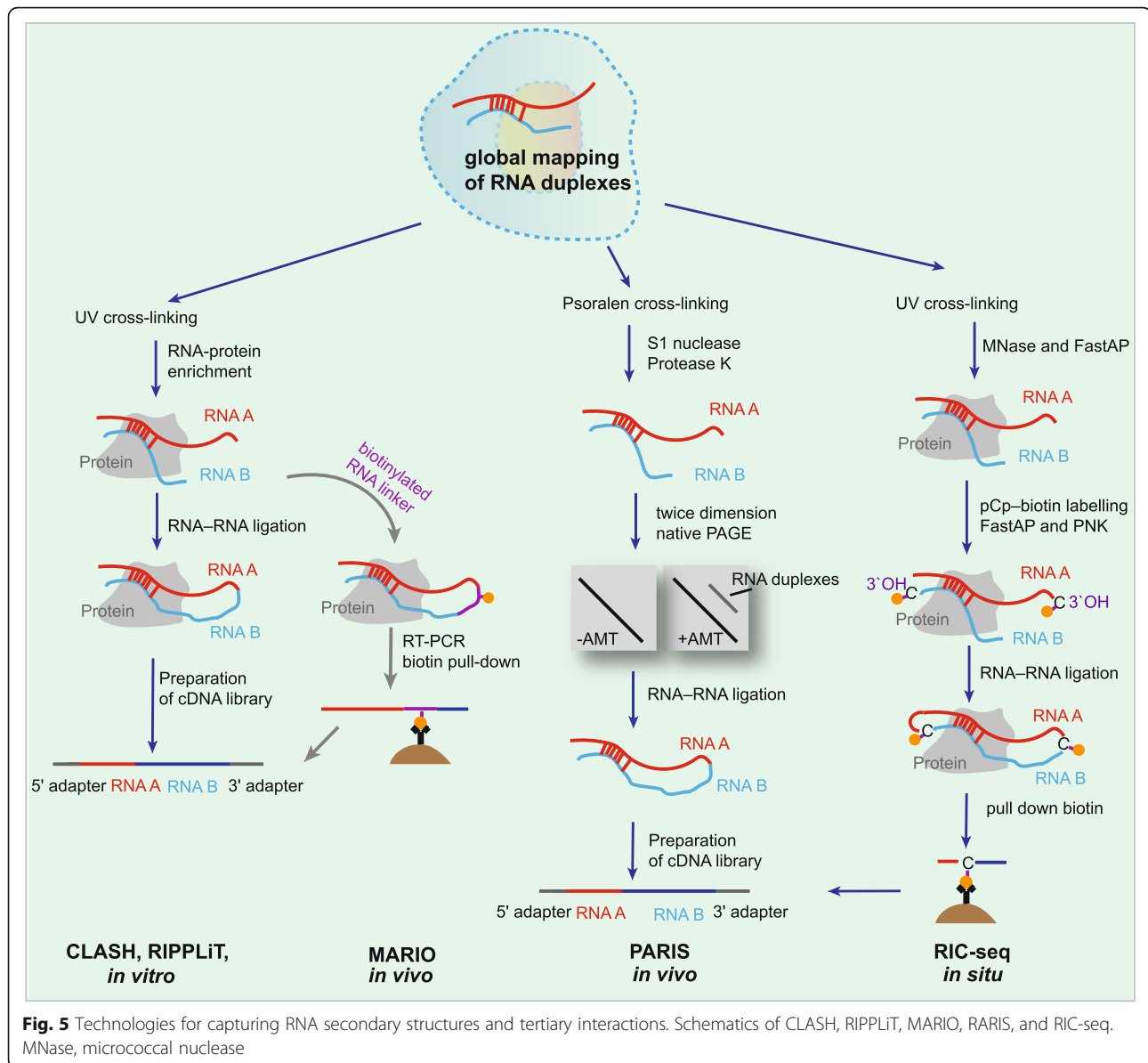
In addition to the sequencing methods for identification of global RNA-chromatin interactomes mentioned above, various techniques were developed to detect specific localization on chromatin of target RNAs [160–162]. These techniques use hybridization of complementary oligonucleotides to pull down a single target RNA, and then NGS or mass spectrometry is performed to identify its DNA- or protein-binding partners.

RNA-RNA spatial interactions

Structured RNAs such as duplexes represent a feature that is critical for most steps in the gene expression pathway. Numerous characterized ncRNAs function via base pairing with target RNAs to control their biological activities, such as dynamic interactions involving snRNA-snRNA and snRNA-pre-mRNA during the assembly and disassembly of spliceosomes, interactions between snoRNAs and their target RNAs to guide RNA

modification, and interactions between ncRNAs and mRNAs that regulate transcript turnover and translation. Thus far, an increasing number of sequencing techniques have been developed for global mapping of RNA-RNA interactions (Fig. 5).

RNA proximity ligation is a set of molecular biological techniques that can be used to analyze the conformation and spatial proximity of RNAs in cells [74]. The typical first steps in these approaches involves cross-linking biological samples with UV light or psoralen, which is followed by partial fragmentation of RNA, RNA-RNA ligation, library preparation, and high-throughput sequencing. UV light and psoralen are two widely used methods for sample preparation prior to proximity ligation: UV light treatment stabilizes and enriches the RNA duplexes that are bound to a specific protein or protein complex; however, psoralen is used to stabilize and enrich RNA-RNA interactions. Studies on RNA conformation have shown different emphases, as some approaches identified pairs of RNAs that are in direct contact or in close proximity with each other, while others recovered pairs of RNAs that are part of the same protein complex or subcellular compartment [163]. Alternative cross-linking methods provide alternative treatments for diverse purposes (Fig. 5). Cross-linking ligation and sequencing of hybrids (CLASH) is a relatively early method that uses UV cross-linking to capture direct RNA-RNA hybridization [74]. Compared to chemically cross-linking methods, which also induce extra protein-protein cross-linking, CLASH has the advantage of avoiding noise from protein intermediate-mediated interactions, and has been used to identify novel snoRNA-rRNA interactions in yeast [74], miRNA-mRNA interactions in human HEK293 cells [164], and piRNAs interactomes [164]. In another method, RNA immunoprecipitation and proximity ligation in tandem (RIPPLiT), sequential pull-down of components of exon junction complexes showed a mapping of mRNA conformations when bound to this complex [75]. Moreover, another approach, mapping the RNA interactome in vivo (MARIO), has identified RNA-RNA interactions in the vicinity of all RNA-binding proteins using a biotin-linked reagent [76].



Methods for identifying RNA-RNA interactions at the transcriptome scale by cross-linking with psoralen have been reported since 2016 [77, 79, 165]. Unlike CLASH, psoralen-based approaches do not depend on the pull-down of RNA-RNA interactions with a specific protein, and in principle, they can yield transcriptome-wide RNA interactomes. The methods using this principle of cross-linking RNAs are combined with different means to enrich cross-linked fragments, such as two-dimensional gel electrophoresis in PARIS (psoralen analysis of RNA interactions and structures) [77], digestion by RNase R in LIGR-seq (ligation of interacting RNA followed by high-throughput sequencing) [78], and biotin-streptavidin enrichment in SPLASH (sequencing of psoralen cross-linked, ligated, and selected hybrids) [79]. The

psoralen cross-linking methods uncovered general properties of RNA-RNA interactomes in mammalian cells (Fig. 5). For example, PARIS uncovered alternative base pairing in intramolecular interactions, which suggests substantial structural heterogeneity in cells, and it also elucidated the structure produced by a repeat of adenosines in Xist *in vivo* [77]. LIGR-seq in HEK293 cells detected novel snRNA-snRNA and snoRNA-rRNA interactions [78]. More importantly, this approach also revealed that SNORD83B can regulate gene expression by binding to target mRNAs, revealing an unexpected function of these snoRNAs [78]. Psoralen cross-linking methods such as PARIS and SPLASH were also applied to detect dense networks of RNA-RNA interactions within viral genomes inside infected cells [166, 167].

A recent study reported a novel method, RNA in situ conformation sequencing (RIC) technology, for the global mapping of intra- and intermolecular RNA-RNA interactions (Fig. 5) [80]. Compared to the RNA ligation induced in vitro in previous methods, RIC-seq performs RNA proximity ligation in situ, and it enriches chimeric reads using a biotinylated cytidine phosphate (pCp-biotin) [80]. Briefly, the cells are cross-linked by formaldehyde, and then, RNA is randomly cut with micrococcal nuclease and dephosphorylated at 3' overhangs. The 3' ends are labeled with pCp-biotin and ligated to proximal 5' overhangs under in situ and non-denaturing conditions. Total RNAs are fragmented in vitro, and RNAs containing C-biotin are enriched followed by conversion into cDNA libraries for sequencing. In the original article, RIC-seq was used to facilitate the generation of 3D RNA interaction maps in human cells, and it revealed global noncoding RNA targets, RNA topological domains, and trans-interacting hubs [80].

In addition to the sequencing methods using RNA proximity ligation, there are some other approaches without ligation that have been developed because of the possible limitation in efficiency of enzymatic ligations affected by short-range distances between RNA ends [81]. RNA proximity sequencing is a method based on massive-throughput RNA barcoding of particles in water-in-oil emulsion droplets [81]. In brief, this approach uniquely barcodes RNA in millions of subnuclear particles in parallel by a rapid vortexing step that combines fragmented nuclear particles with barcoded beads in a water-in-oil emulsion; then, the cDNA is sequenced. The detection of multiple RNAs in proximity to each other by RNA proximity sequencing distinguished RNA-dense and RNA-sparse compartments, and this technique is an alternative approach for studying the spatial organization of transcripts in the nucleus, including ncRNAs and their functional relevance.

RNAs in protein complexes or subcellular structures

The location of ncRNAs in cells is the primary determinant of their molecular functions. ncRNAs, especially lncRNAs, are often considered as chromatin-restricted modulators of gene transcription and chromatin structure [157, 158]. In addition, a rich population of cytoplasmic ncRNAs, such as extra lncRNAs and exonic circRNAs, have been reported to play roles in diverse biological processes, including translational regulation and signal transduction [8, 168]. Elution-based methods promise to detect RNAs at the transcriptome scale associated with all organelles of mammalian cells, and RNA maps of increasing resolution reveal a subcellular world of highly specific localization patterns.

In situ hybridization (ISH) is the most widely used method of RNA localization using labeled complementary

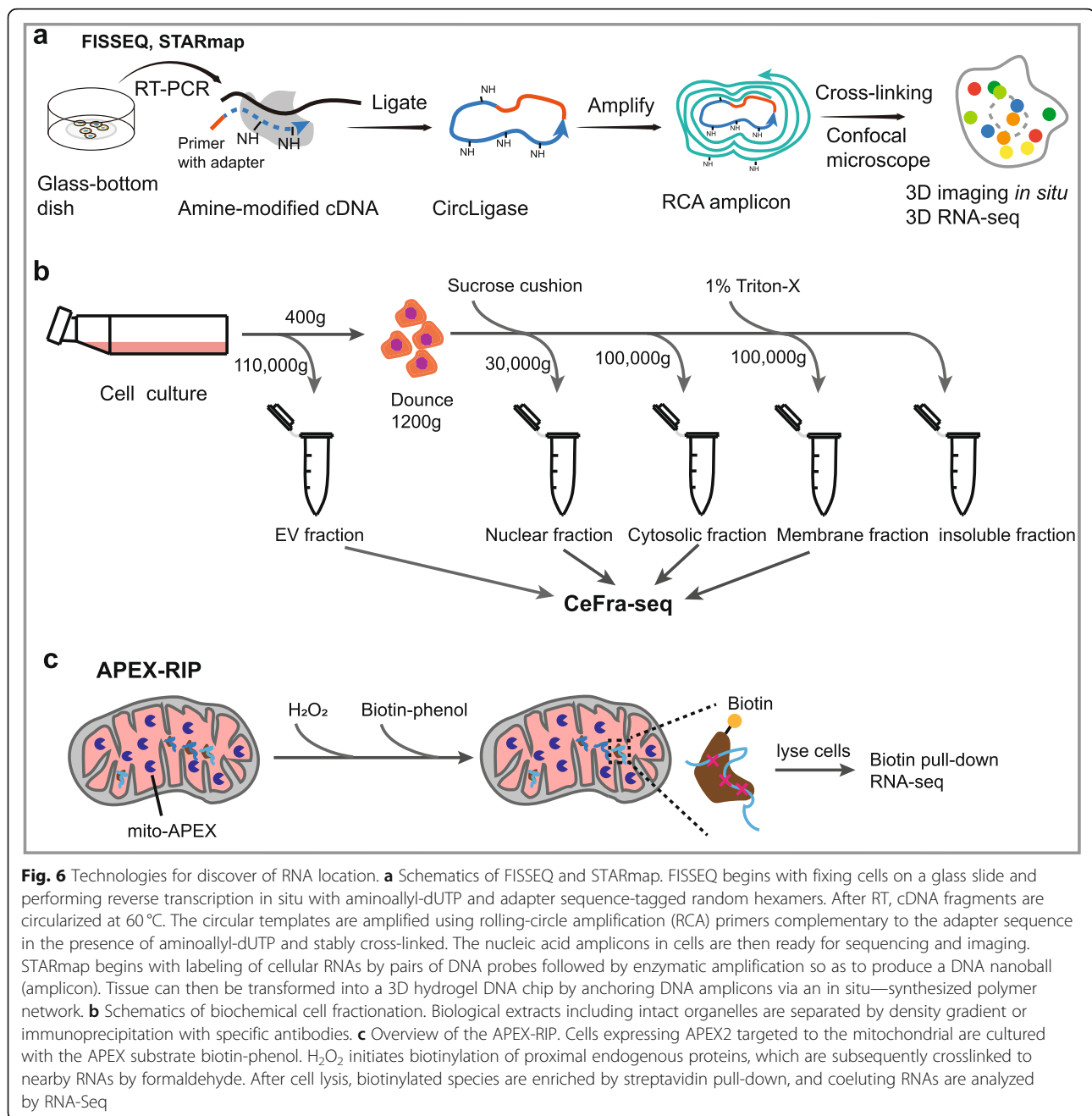
oligonucleotide probes to visualize target RNAs [169, 170]. Single-molecule fluorescence ISH (smFISH) uses multiple probes to amplify the fluorescent signal for the detection of target RNAs at low levels, and it is thought of as the gold-standard technique for single-gene studies [169, 171]. In contrast to RNA smFISH, fluorescent in situ RNA sequencing (FISSEQ) offers in situ information at high-throughput levels [82]. In this approach, RNA is reverse transcribed in situ into cDNA in cross-linked cells and tissue samples, which is then analyzed by sequencing (Fig. 6a). However, compared to standard RNA-seq, FISSEQ also comes at the expense of lower read coverage, which reduces sensitivity for lowly expressed RNAs, especially ncRNAs [82]. Another alternative related technique, spatially resolved transcript amplicon readout mapping (STARmap), provided 3D locational information of RNA expression in intact tissue samples [172].

Biochemical cell fractionation is a fractionation-based method that physically isolates subcellular compartments and identifies their RNAs (Fig. 6b). These types of methods can be based on protein immunoprecipitation, intact organelle purification, or partitioning through sucrose gradients [173]. Then, RNA-seq (biochemical cell fractionation combined with RNA-seq, CeFra-seq) was performed to detect specific RNAs at the transcriptome scale [83]. Fractionation-based methods have high sensitivity for low-abundance transcripts due to aggregation across many cells; however, they are restricted by isolation protocols and the purity of resulting isolates, which possibly induce technical noise by contamination across fractions [24, 174, 175].

Recently, innovative techniques have been developed to overcome the deficiencies of conventional methods. A new fractionation-based method, APEX-RIP [84], was developed, and it combines APEX (engineered ascorbate peroxidase)-catalyzed proximity biotinylation [176] and RNA immunoprecipitation (RIP) [177] to map RNAs at vastly improved spatial resolution (Fig. 6c). In brief, APEX-catalyzed proximity biotinylation is targeted by genetic fusion to proteins from various subcellular compartments of interest. This is followed by protein-RNA crosslinking and RIP to pull down the biotinylated subcellular fraction for further high-throughput sequencing. Using this method, thousands of ncRNAs have been mapped to specific compartments without the need for purification of specific organelles, and it offers high specificity and sensitivity in targeting the transcriptome of membrane-bound organelles [84]. Moreover, in a recent study, a transcriptome-wide subcellular RNA atlas was generated by APEX-RIP [178].

ncRNA database

Various sequencing methods have provided systematic expression profiling of ncRNAs in diverse cells, tissues,



and organisms, and they have mapped the interaction networks or subcellular localization of ncRNAs, which inform their potential biological functions. Databases provide important references based on theoretical analysis, sequencing data, and even experimental verification, which play a guiding role in the identification and functional investigation of ncRNAs. Here, we will introduce a series of ncRNA databases that emphasize basic ncRNA information, cancer-associated ncRNA expression patterns, or specific ncRNA interaction

networks based on experimental techniques followed by high-throughput sequencing.

The correlations between ncRNA expression and cancer progression provide important hints whether a ncRNA could play a role in certain cancers. There are an increasing number of databases providing comprehensive associations between ncRNAs and human cancers, which are supported by sequencing data or even experiments, such as TANRIC [179], Lnc2Cancer 2.0 [180], InCaNet [181], and LncRNADisease [182] for

lncRNAs, CSCD [183], Circ2Traits [184], CircR2Disease [185], and MiOncoCirc [119] for circRNAs, miRCancer [186], SomamiR 2.0 [187], OncomiR [188], miRCancerdb [189], and dbDEMC 2.0 [190] for miRNAs and YM500v3 [191], tRF2Cancer [192], and MINbase v2.0 [193] for other small ncRNAs, as summarized in Table 2. A recently reported MiOncoCirc is the first database that mainly consists of circRNAs directly detected in tumor tissues [119]. It was established by detecting and characterizing circRNAs across more than 2000 cancer samples with an exome capture RNA sequencing protocol. In the article that originally described the process, candidate circRNAs identified from MiOncoCirc were determined to be useful as biomarkers for prostate cancer and were found to be detected in urine, suggesting that MiOncoCirc could be an alternative tool to uncover novel diagnostic biomarkers for clinical translational strategies [119]. Another interesting, recently reported database is SELER, which collects specific super-enhancer-associated lncRNA profiles from different cancers [195]. In addition, some databases document the basic annotation and functional information on ncRNAs, including lncRNA-associated resources LNCipedia [199], LNCediting [200], lncRNadb v2.0 [201], circRNA-associated ones circAtlas [206], circBase [207], CIRCpedia v 2 [208], TSCD [209], miRNA-associated ones starBase v2.0 [210], miRTarBase [211], miRmine [212], EVmiRNA [213], miRGate [214], miRBase [215], and even other small ncRNA-associated ones DASHR 2.0 [217]. A growing number of databases have undoubtedly played important roles in the discovery and investigation of novel functional ncRNAs.

Several specific RNA-seq datasets have revealed the subcellular locations and potential interactomes of ncRNAs, which provide more real information than what is learned from bioinformatics prediction. There are some databases that provide high-quality RNA subcellular location resources in accordance with the results of subcellular compartment sequencing, such as RNALocate [218] and LncATLAS [219]. RNALocate documents more than 37,700 manually curated RNA subcellular location entries with experimental evidence, and it has data on 65 organisms, 42 subcellular locations (such as cytoplasm, nucleus, endoplasmic reticulum), and 9 RNA categories, such as lncRNAs [218]. However, thus far, few interactome database of ncRNAs except miRNA [210, 211, 216], has been established based on experimental techniques and sequencing. NPInter v3.0 is a database of ncRNA-associated interactions based on experimental techniques followed by high-throughput sequencing, such as crosslinking and immunoprecipitation followed by deep sequencing (CLIP-seq) [220], and chromatin isolation by RNA purification followed by high-throughput sequencing (ChIRP-seq) [161, 221]. NPInter v3.0 documented approximately 500,000

interactions in 188 tissues (or cell lines) from 68 kinds of experiments and predicted the functions of lncRNAs in humans on the basis of their interactions in the database [221]. Furthermore, a database of RNA interactomes identified by sequencing at the transcriptome scale is lacking, and it is needed for identification of novel functional ncRNAs.

Application of cancer-related ncRNA identification for diagnosis

Due to their highly tissue-specific expression patterns identified by various sequencing techniques and their key roles in regulating biological activity in cancer, ncRNAs, including miRNAs, lncRNAs, and circRNAs, are generally considered to have potential as novel biomarkers for cancer diagnosis [20, 222, 223]. This section aims to present new developments in diagnostic kits for cancer diagnosis by the analysis of cancer-related ncRNAs.

Cancer seriously threatens the human life and gives rise to an enormous burden on society. However, the incidence and mortality of cancer could be decreased effectively by preventative measures, including early detection tests and monitoring of cancer prognosis. Therefore, searching for novel biomarkers that are easy to use, are not invasive, and exhibit high sensitivity, specificity, and stability for cancer diagnosis and prognosis has been a key clinical translational strategy. In addition to the features of specific expression patterns, some types of ncRNAs, such as miRNAs, lncRNAs, and circRNAs, have also been shown to be relatively stable in serum, plasma, saliva, or urine, which can be easier to collect and is less harmful or invasive for patients than other collection methods. In the past few years, seeking novel biomarkers in cancer diagnosis has mainly focused on miRNAs [224]. Recently, growing research has shown that other ncRNAs, especially lncRNAs and circRNAs, could also serve as a hallmark of carcinomas.

MiRNAs, lncRNAs, and circRNAs have been observed to have highly specific expression patterns in diverse types of cancers, and this aberrant expression usually occurs in certain tumor cells or cancer tissues at a specific stage of disease progression [13, 14, 225]. According to patent searches in resources such as the EPO (<https://worldwide.espacenet.com>), there are growing uses of these three types of ncRNAs in the preparation of diagnostic kits for various cancers, including hepatocellular, cervical, stomach, liver, breast, prostatic, and bladder cancers (Table 3). Generally, detecting cancer-related nucleic acids in patient samples using qRT-PCR with specific primers or probes is the main method for diagnosis based on ncRNAs, which is also the primary approach for diagnosing disease in the recent COVID-19 (CoronaVirusDisease2019) pandemic [226]. For example, a recent patent provided a circRNA *hsacirc_0028185*

Table 2 Database of ncRNAs

Cancer or basis	Database	Species	Website	Short description	Ref
Cancer	Lnc2Cancer v2.0	lncRNA	http://www.bio-bigdata.net/lnc2cancer	An updated database that provides comprehensive experimentally supported associations between lncRNAs and human cancers.	[180]
	TANRIC	lncRNA	http://bioinformatics.mdanderson.org/main/TANRIC:Overview	This database characterizes the expression profiles of lncRNAs in large patient cohorts of 20 cancer types, including TCGA and independent datasets (> 8000 samples overall).	[179]
	InCaNet	lncRNA	http://lncanet.bioinformatics.zhao.org/	This database provides a comprehensive co-expression data resource which reveals the interactions between lncRNA and non-neighbouring cancer genes.	[181]
	LncRNADisease 2.0	lncRNA	http://www.rnanut.net/lncrnadisease/	A database integrating comprehensive experimentally supported and predicted lncRNA-disease associations.	[182]
	The Cancer LncRNome Atlas	lncRNA	http://tcla.fcgportal.org/	An academic research database to explore the lncRNA alternations across multiple human cancer types.	[194]
	SELER	lncRNA	http://www.seler.cn/download.php	A database of super-enhancer-associated lncRNA-directed transcriptional regulation in human cancers.	[195]
	CSCD	circRNA	http://gb.whu.edu.cn/CSCD	A database that focuses on distinguishing cancer-specific circRNAs from noncancerous circRNAs, and reports predicted cellular location, RBP sites, and ORFs.	[183]
	Circ2Traits	circRNA	http://gyanxet-beta.com/circdb/	Provide circRNA-disease association based on the interaction of circRNAs with disease-related miRNAs and SNP mapped on circRNA loci.	[184]
	CircR2Disease	circRNA	http://bioinfo.snnu.edu.cn/CircR2Disease/	Provide a comprehensive resource for circRNA deregulation in various diseases, containing 725 associations between 661 circRNAs and 100 diseases.	[185]
	CircRNA disease	circRNA	http://cgga.org.cn:9091/circRNADisease/	A manually curated database of experimentally supported circRNA-disease associations.	[196]
	MiOncoCirc	circRNA	https://nguyenjoshvo.github.io/	circRNA detection in 2093 clinical human cancer samples using exome capture sequencing.	[119]
	CircRiC	circRNA	https://hanlab.uth.edu/cRic	A database focusing on lineage-specific circRNAs in 935 cancer cell lines including drug response.	[197]
	miRCancer	miRNA	http://mircancer.ecu.edu/	A database currently documents more than 9000 relationships between 57,984 miRNAs and 196 human cancers.	[186]
	SomamiR 2.0	miRNA	http://compbio.uthsc.edu/SomamiR/	A database of cancer somatic mutations in microRNAs (miRNA) and their target sites that potentially alter the interactions between miRNAs and competing endogenous RNAs (ceRNA).	[187]
	OncomiR	miRNA	http://www.oncomir.org/	An online resource for exploring miRNA dysregulation in cancer.	[188]
miRCancerdb	miRNA	https://mahshaaban.shinyapps.io/miRCancerdb/	An easy-to-use database to investigate the microRNAs-dependent regulation of target genes involved in development of cancer.	[189]	
miR2Disease	miRNA	http://www.miR2Disease.org	A database aiming at providing a comprehensive resource of microRNA deregulation in various human diseases.	[198]	
YM500v3	small ncRNA	http://ngs.ym.edu.tw/ym500/	A database which contains more than 8000 small RNA-seq datasets and focuses on piRNAs, tRFs, snRNAs, snoRNAs, and miRNAs.	[191]	
tRF2Cancer	small ncRNA	http://rna.sysu.edu.cn/tRFfinder/	A web server to detect tRFs and their expression in multiple cancers.	[192]	
MINTbase v2.0	Small ncRNA	https://cm.jefferson.edu/MINTbase/	A framework for the interactive exploration of mitochondrial and nuclear tRNA fragments.	[193]	
Basis	LNCipedia	lncRNA	https://lncipedia.org	A public database for lncRNA sequence and annotation.	[199]
	LNCediting	lncRNA	http://bioinfo.life.hust.edu.cn/LNCediting/	This database provides a comprehensive resource for the functional prediction of RNA editing in lncRNAs.	[200]
	lncRNADB v2.0	lncRNA	http://lncrnadb.com/	This database provides comprehensive annotations of eukaryotic lncRNAs.	[201]
	LncRNAWiki	lncRNA	http://lncrna.big.ac.cn	This database is a publicly editable and open-content platform for community curation of human lncRNAs.	[202]
	LncBook	lncRNA	http://bigd.big.ac.cn/lncbook	This database is a curated knowledgebase of human lncRNAs.	[203]

Table 2 Database of ncRNAs (Continued)

Cancer or basis	Database	Species	Website	Short description	Ref
	MONOCLdb	lncRNA	https://www.monocldb.org/	20,728 mouse lncRNA genes.	[204]
	NONCODE	lncRNA	http://www.bioinfo.org/noncode/	An interactive database that aims to present the most complete collection and annotation of ncRNAs especially lncRNAs from 17 species.	[205]
	CircAtlas	circRNA	http://circatlas.biols.ac.cn/	An integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes.	[206]
	circBase	circRNA	http://www.circbase.org/	A database containing thousands of recently identified circRNAs in eukaryotic cells.	[207]
	CIRCpedia v2	circRNA	http://www.picb.ac.cn/momics/circpedia	A database for comprehensive circRNA annotation from over 180 RNA-seq datasets across six different species.	[208]
	TSCD	circRNA	http://gb.whu.edu.cn/TSCD	A tissue-specific circRNA database from RNA-seq datasets and characterized the features of circRNAs in human and mouse.	[209]
	starBase v2.0	miRNA	http://starbase.sysu.edu.cn/	A database decoding miRNA-ceRNA, miRNA-ncRNA, and protein-RNA interaction networks from large-scale CLIP-Seq data.	[210]
	miRTarBase	miRNA	http://mirtarbase.cuhk.edu.cn/php/index.php	A resource for experimentally validated microRNA-target interactions.	[211]
	miRmine	miRNA	http://guanlab.ccmb.med.umich.edu/mirmine	A database of human miRNA expression profiles.	[212]
	EVmiRNA	miRNA	http://bioinfo.life.hust.edu.cn/EVmiRNA#!	A database focusing on miRNA expression profiles in extracellular vesicles.	[213]
	miRGate	miRNA	http://mirgate.bioinfo.cnio.es/miRGate/	A curated database of human, mouse, and rat miRNA-mRNA targets.	[214]
	miRBase	miRNA	http://www.mirbase.org/	A database containing microRNA sequences from 271 organisms: 38,589 hairpin precursors and 48,860 mature microRNAs.	[215]
	DIANA-TarBase v8	miRNA	http://www.microrna.gr/tarbase	A reference database devoted to the indexing of experimentally supported miRNA targets.	[216]
	DASHR 2.0	small ncRNA	http://lisanwanglab.org/DASHR	A database that integrates human small ncRNA gene and mature products derived from all major RNA classes.	[217]

qPCR assay kit for the diagnosis of hepatocellular carcinoma. By detecting expression changes of serum hscirc_0028185, it is possible to assess the occurrence and development of hepatocellular carcinoma. Another sample is that lncRNA-AC006159.3 in the blood could be used for the diagnostic kit to rapidly speculate the cetuximab-resistant possibility of rectal cancer. Briefly, the lower the expression level of lncRNA-AC006159.3, the higher the possibility of cetuximab resistance. In addition, a patent provided application of miRNA-410 in preparation of a prostatic cancer diagnostic kit.

It is noteworthy that the same RNA may be aberrantly expressed in many types of cancers, which allows the same RNA to be used to diagnose different kinds of cancers. Moreover, the RNA-seq data show that a number of diverse species of ncRNAs are dysregulated in cancer samples compared to normal tissues, suggesting that diagnostic kits can be designed to detect multiple ncRNAs at the same time for more efficient cancer diagnosis.

Conclusion and perspective

The tissue-specific expression patterns, complicated regulatory networks, and emerging roles all suggest that

ncRNAs are not simply debris or side products of transcriptional processes or aberrant splicing; rather, they are important regulatory molecules [102]. New technologies have endlessly emerged with different goals in ncRNA identification in multiple areas of research, including detection of ncRNA expression at the transcriptome scale, identification of novel ncRNA categories, searching for potential functional RNA within specific subcellular compartments, or discovering applicable biomarkers for cancer diagnosis. There are also some ncRNA-associated databases that provide multiple ncRNA information to enable further functional RNA investigations. Moreover, with the increasing number of studies on cancer-associated ncRNAs, translational applications of specific ncRNA identification for clinical diagnosis have been developed, such as diagnostic kits.

It is noteworthy that different RNA categories can be generated from the same regions of DNA, and they can share the same sequences. In addition to small ncRNAs derived from snoRNAs, tRNAs, or rRNAs, which can play a role in the miRNA-like pathway, some long non-polyadenylated transcripts, such as sno-lncRNAs and circRNAs, have also been found to be generated from

Table 3 Developments in diagnostic kits for cancer diagnosis (EPO <https://worldwide.espacenet.com>)

Species	Name	Expression in cancer	Diseases	Application	Patent number
circRNA	hsacirc_0028185	Up	Hepatocellular carcinoma	Cancer auxiliary diagnosis	CN111004850A (2020)
circRNA	hsa_circ_001477	Up	Gastric cancer	Cancer diagnosis	CN110129324A (2019)
circRNA	hsa_circRNA_012515	Up	Non-small cell lung cancer	Cancer diagnosis	CN110592223A (2019)
circRNA	hsa_circRNA_405124 or hsa_circ_0012152	Up	Leukemia	Cancer early diagnosis	CN109593859A (2019)
circRNA	circ_104075	Up	Liver cancer	Cancer diagnosis	CN109161595A (2019)
circRNA	circ3823	Up	Colorectal cancer	Cancer early diagnosis	CN110592220A (2019)
circRNA	hsa_circ_0021977	Up	Breast cancer	Cancer diagnosis	CN109022583A (2018)
circRNA	hsa_circ_0012755	Up	Prostate cancer	Cancer diagnosis	CN108624688A (2018)
circRNA	circ_0047921, circ_0007761 and circ_0056285	Up	Non-small cell lung cancer	Cancer early diagnosis	CN108179190A (2018)
circRNA	hsa-circRPL15-001	Up	Chronic lymphocytic leukemia	Cancer diagnosis	CN109055564A (2018)
circRNA	has_circ_0117909	Up	Acute lymphoblastic leukemia	Cancer diagnosis	CN107937522A (2017)
circRNA	has_circ_0005720	Down			
circRNA	cRNA-ZFR	Up	Bladder cancer	Cancer diagnosis	CN106011139A (2016)
lncRNA	lncRNA-AC006159.3	Down	Colorectal cancer	Cetuximab-resistance diagnosis	CN108949993A (2018)
lncRNA	lncRNAXLOC_004122, linc00467 and lncRNAA1049452	Up	Breast cancer	Cancer bone metastasis diagnosis	CN107699619A (2017)
lncRNA	lncRNA GENE NO.9	Up	Bladder cancer	Cancer diagnosis	CN107267636A (2017)
lncRNA	LINC00516	Up	Lung cancer	Cancer or cancer metastasis diagnosis	CN108998528A (2018)
lncRNA	LSAMP-AS1	Up	Gastric cancer	Cancer diagnosis	CN110628915A (2019)
miRNA	miRNA-4692	Down	Hepatocellular carcinoma	Cancer diagnosis	CN107604065A (2018)
miRNA	miRNA-1266	Up	Endometrial carcinoma	Cancer diagnosis	CN105907883A (2016)
miRNA	miR-320	Down	Cervical cancer	Cancer early diagnosis	CN105506076A (2016)
miRNA	miRNA-2116	Up	Lung adenocarcinoma	Cancer metastasis diagnosis	CN104774966A (2015)
miRNA	miRNA-410	Up	Prostate cancer	Cancer diagnosis	CN104651492A (2015)
miRNA	miRNA-1262	Up	Acute myeloid leukemia	Cancer diagnosis	CN105063052A (2015)

the genetic sequences of well-known ncRNAs [10, 37, 42, 126, 139]. Sno-lncRNAs have the same classical stem-loop as snoRNAs originating from the same genomic regions. However, both types of ncRNAs have been verified to have individual functions and to be important

regulatory molecules in biological processes [42]. Another example is circANRIL, a circRNA formed from the lncRNA ANRIL, which performs functions in apoptosis and proliferation that are the opposite of the functions of ANRIL [227]. Taken together, these results

suggest that transcripts derived from canonical DNA regions have functions in addition to their classical ones by interacting with nonclassical binding molecules or by being located in novel components, which indicates that the transcriptome extends far beyond the genome. A larger RNA world is waiting for us to explore.

High-throughput sequencing with purposeful sample preparation not only has uncovered novel species of ncRNAs but also has mapped the interaction networks and subcellular locations of ncRNAs [221]. In accordance with sequencing data of the RNA-associated interactome and RNA subcellular locations that have been determined at the transcriptome scale, a series of non-coding transcripts, especially lncRNAs and circRNAs, exhibit specific distributions in organelles, protein complexes, or subcellular structures [228]. These results further indicate that ncRNAs are functional molecules playing roles in specific compartments, providing a pool of candidates for us to search for specifically functional ncRNAs. Using these sequencing methods, a huge number of snoRNAs have been found to be enriched on chromatin, robustly suggesting that other potential functions of these well-known small ncRNAs are located in a nonclassical compartment [71]. However, the functions of chromatin-associated snoRNAs remain unanswered. In addition, thousands of potential functional ncRNAs with specific interactions or locations have been discovered and await further investigation.

Due to the features of specific expression patterns in cancers and relatively high stability in serum, plasma, saliva, or urine, ncRNAs especially miRNAs, lncRNAs, and circRNAs are generally considered to have potential as non-invasive diagnostic biomarkers for cancers. A growing number of researchers have provided the suitable ncRNA candidates for diagnosis of different cancers and increasing patents about preparation of ncRNA diagnostic kits for cancer diagnosis have been approved. However, most of these candidate ncRNAs are still in the preclinical stages. In addition, the results of some studies evaluating the potential of ncRNAs as biomarkers are conflicting [20]. Thus, more accurate evaluation of RNA expression pattern in larger cohorts of clinical data are needed to reconcile the controversies.

Abbreviations

ncRNA: Noncoding RNA; lncRNA: Long noncoding RNA; circRNA: Circular RNA; mRNA: Message RNA; H3K4me3: Trimethylation of lysine 4 of histone 3; H3K36me3: Trimethylation of lysine 36 of histone 3; lincRNA: Long intergenic ncRNA; H3K4m1: Monomethylation of lysine 4 of histone 3; H3K27ac: Histone H3 acetylation at lysine 27; eRNA: Enhancer-derived RNA; 3D: 3 Dimensions; pre-rRNA: Precursor RNA; snoRNA: Small nucleolar RNA; rRNA: Ribosomal RNA; tRNA: Transfer RNA; sno-lncRNA: SnoRNA-related lncRNA; rRF: Ribosomal RNA-derived fragment; tsRNA: tRNA-derived small RNA; sdRNA: snoRNA-derived RNA; ORF: Open reading frame; SRA: Steroid receptor RNA activator; Xist: X inactive specific transcript; lnx3: Ligand of numb-protein X 3; Pol II: RNA polymerase II; snoRBP: snoRNA binding protein; RNA-seq: RNA sequencing; XPO1: Exportin 1; CAMSAP1: Calmodulin-regulated spectrin-

associated protein 1; EIF3J: Eukaryotic translation initiation factor 3 subunit J; PAIP2: Poly(A)-binding protein-interacting protein 2; miRNA: MicroRNA; siRNA: Small interfering RNA; piRNA: Piwi-interacting RNA; Ago: Argonaute; RISCs: RNA-induced silencing complexes; RBP: RNA-binding protein; vRNA: Vault RNA; VTRNA: Vault RNA gene; Pol III: RNA polymerase III; Dicer: Ribonuclease III; cDNA: Complementary DNA; HOTAIR: HOX antisense intergenic RNA; NGS: Next-generation sequencing; PRC2: Polycomb repressive complex; scRNA-seq: Single-cell RNA sequencing; FACS: Fluorescence-activated cell sorting; DP-seq: Designed primer-based sequencing; MMLV RT: Moloney murine leukemia virus reverse transcriptase; PCR: Polymerase chain reaction; SUPeR-seq: Single-cell universal poly(A)-independent sequencing; RamDA-seq: Random displacement amplification sequencing; RT-RamDA: RT with random displacement amplification; NSR: Not-so-random primer; Neat1: Nuclear paraspeckle assembly transcript 1; GRO-seq: Global run-on sequencing; RT: Reverse transcription; GRID-seq: Global RNA interactions with DNA by deep sequencing; DSG: Disuccinimidyl glutarate; ssRNA: Single-stranded RNA; dsDNA: A double-stranded DNA; ChAR-seq: Chromatin-associated RNA sequencing; caRNA: Chromatin-associated RNA; TSSs: Transcription start sites; snRNA: Small nuclear RNA; CLASH: Cross-linking ligation and sequencing of hybrids; RIPP LIT: RNA immunoprecipitation and proximity ligation in tandem; MARIO: Mapping the RNA interactome in vivo; PARIS: Psoralen analysis of RNA interactions and structures; LIGR-seq: Ligation of interacting RNA followed by high-throughput sequencing; SPLASH: Sequencing of psoralen cross-linked, ligated, and selected hybrids; RIC: RNA in situ conformation sequencing; ISH: In situ hybridization; smFISH: Single-molecule fluorescence ISH; FISSEQ: Fluorescent in situ RNA sequencing; STARmap: Spatially resolved transcript amplicon readout mapping; CeFra-seq: Biochemical cell fractionation combined with RNA-seq; APEX: Engineered ascorbate peroxidase; RIP: RNA immunoprecipitation; CLIP-seq: Immunoprecipitation followed by deep sequencing; ChIRP-seq: Chromatin isolation by RNA purification followed by high-throughput sequencing

Acknowledgements

Not applicable.

Authors' contributions

Y.Q.C designed the study and drafted the manuscript. Y.M.S prepared the tables and figures and drafted the manuscript. All authors participated in the process of drafting and revising the manuscript. The authors read and approved the final manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (No. 81770174 and 31700719), National Key R&D Program of China (No. 2017YFA0504400), and grants from China postdoctoral foundation (No. 2019M663224) and Sun Yat-sen University (20lqpy115).

Availability of data and materials

The material supporting the conclusion of this study has been included within the article.

Ethics approval and consent to participate

Not applicable for this review.

Consent for publication

Not applicable for this article.

Competing interests

The authors declare that they have no competing interests.

Received: 1 June 2020 Accepted: 27 July 2020

Published online: 10 August 2020

References

- Warner, Soeiro R, Birnboim HC, Girard M, Darnell JE. Rapidly labeled HeLa cell nuclear RNA. I. Identification by zone sedimentation of a heterogeneous fraction separate from ribosomal precursor RNA. *J Mol Biol.* 1966;19:349–61.
- Weinberg RA, Penman S. Small molecular weight monodisperse nuclear RNA. *J Mol Biol.* 1968;38:289–304.

3. Zieve G, Penman S. Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell*. 1976;8:19–31.
4. Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet*. 2001;2:919–29.
5. Pachnis V, Brannan CI, Tilghman SM. The structure and expression of a novel gene activated in early mouse embryogenesis. *EMBO J*. 1988;7:673–81.
6. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 2016;17:47–62.
7. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495:333–8.
8. Li X, Yang L, Chen LL. The biogenesis, functions, and challenges of circular RNAs. *Mol Cell*. 2018;71:428–42.
9. Schimmel P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol*. 2018;19:45–58.
10. Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, et al. A human snoRNA with microRNA-like functions. *Mol Cell*. 2008;32:519–28.
11. Lambert M, Benmoussa A, Provost P. Small non-coding RNAs derived from eukaryotic ribosomal RNA. *Noncoding RNA*. 2019;5:16.
12. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012;482:339–46.
13. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015;21:1253–61.
14. Dong Y, He D, Peng Z, Peng W, Shi W, Wang J, et al. Circular RNAs in cancer: an emerging key player. *J Hematol Oncol*. 2017;10:2.
15. Pekarsky Y, Balatti V, Palamarchuk A, Rizzotto L, Veneziano D, Nigita G, et al. Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. *Proc Natl Acad Sci U S A*. 2016;113:5071–6.
16. Li Y, Zeng C, Hu J, Pan Y, Shan Y, Liu B, et al. Long non-coding RNA-SNHG7 acts as a target of miR-34a to increase GALNT7 level and regulate PI3K/Akt/mTOR pathway in colorectal cancer progression. *J Hematol Oncol*. 2018;11:89.
17. Fort RS, Matho C, Oliveira-Rizzo C, Garat B, Sotelo-Silveira JR, Duhagon MA. An integrated view of the role of miR-130b/301b miRNA cluster in prostate cancer. *Exp Hematol Oncol*. 2018;7:10.
18. Basati G, Khaksarian M, Abbaszadeh S, Lashgarian HE, Marzban A. Cancer stem cells and nanotechnological approaches for eradication. *Stem Cell Investig*. 2019;6:38.
19. Wang WT, Chen TQ, Zeng ZC, Pan Q, Huang W, Han C, et al. The lncRNA LAMP5-AS1 drives leukemia cell stemness by directly modulating DOT1L methyltransferase activity in MLL leukemia. *J Hematol Oncol*. 2020;13:78.
20. Wang WT, Han C, Sun YM, Chen TQ, Chen YQ. Noncoding RNAs in cancer therapy resistance and targeted drug development. *J Hematol Oncol*. 2019;12:55.
21. Zhang M, Xin Y. Circular RNAs: a new frontier for cancer diagnosis and therapy. *J Hematol Oncol*. 2018;11:21.
22. Rastgoo N, Abdi J, Hou J, Chang H. Role of epigenetics-microRNA axis in drug resistance of multiple myeloma. *J Hematol Oncol*. 2017;10:121.
23. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309:1559–63.
24. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature*. 2012;489:101–8.
25. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007;316:1484–8.
26. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128:693–705.
27. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129:823–37.
28. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007;448:553–60.
29. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009;459:108–12.
30. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. 2009;458:223–7.
31. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A*. 2009;106:11667–72.
32. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*. 2010;28:503–10.
33. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol*. 2010;8:e1000384.
34. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating RNAs associate with mediator to enhance chromatin architecture and transcription. *Nature*. 2013;494:497–501.
35. Jiao W, Chen Y, Song H, Li D, Mei H, Yang F, et al. HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNP/p300/EGR1/HPSE axis. *Oncogene*. 2018;37:2728–45.
36. Li S, Xu Z, Sheng J. tRNA-derived small RNA: a novel regulatory small non-coding RNA. *Genes (Basel)*. 2018; 9:246.
37. Chen LL, Yang L. Regulation of circRNA biogenesis. *RNA Biol*. 2015;12:381–8.
38. Danan M, Schwartz S, Edelheit S, Sorek R. Transcriptome-wide discovery of circular RNAs in archaea. *Nucleic Acids Res*. 2012;40:3131–42.
39. Chen LL. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol*. 2016;17:205–11.
40. Lasda E, Parker R. Circular RNAs: diversity of form and function. *RNA*. 2014; 20:1829–42.
41. Dong R, Zhang XO, Zhang Y, Ma XK, Chen LL, Yang L. CircRNA-derived pseudogenes. *Cell Res*. 2016;26:747–50.
42. Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, et al. Long noncoding RNAs with snoRNA ends. *Mol Cell*. 2012;48:219–30.
43. Dinger ME, Pang KC, Mercer TR, Mattick JS. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput Biol*. 2008;4:e1000176.
44. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*. 2015;160:595–606.
45. Huang JZ, Chen M, Chen D, Gao XC, Zhu S, Huang H, et al. A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol Cell*. 2017;68:171–84.
46. Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, et al. Translation of lincRNAs. *Mol Cell*. 2017;66:9–21.
47. Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun*. 2018;9:4475.
48. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. *Mol Cell Biol*. 1990;10:28–36.
49. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, 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.
50. Novikova IV, Hennelly SP, Sanbonmatsu KY. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res*. 2012;40:5034–51.
51. Duret L, Chureau C, Samain S, Weissenbach J, Avner P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science*. 2006;312:1653–5.
52. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;27:i275–82.
53. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41:e74.
54. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427–32.
55. Malone B, Atanassov I, Aeschmann F, Li X, Grosshans H, Dieterich C. Bayesian prediction of RNA translation from ribosome profiling. *Nucleic Acids Res*. 2017;45:2960–72.
56. Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. *Genomics*. 2005;85:1–15.
57. Yan B, Wang ZH, Guo JT. The research strategies for probing the function of long noncoding RNAs. *Genomics*. 2012;99:76–80.
58. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10:57–63.

59. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddloh JA, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat Biotechnol.* 2011;30:99–104.
60. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol.* 2012;30:777–82.
61. Bhargava V, Ko P, Willems E, Mercola M, Subramaniam S. Quantitative transcriptomics using designed primer-based amplification. *Sci Rep.* 2013;3:1740.
62. Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, Imai T, et al. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* 2013;14:R31.
63. Fan X, Zhang X, Wu X, Guo H, Hu Y, Tang F, et al. Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biol.* 2015;16:148.
64. Hayashi T, Ozaki H, Sasagawa Y, Umeda M, Danno H, Nikaido I. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat Commun.* 2018;9:619.
65. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell.* 2007;129:1401–14.
66. Hagemann-Jensen M, Abdullayev I, Sandberg R, Faridani OR. Small-seq for single-cell small-RNA sequencing. *Nat Protoc.* 2018;13:2407–24.
67. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science.* 2008;322:1845–8.
68. Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods.* 2017;14:1198–204.
69. Schofield JA, Duffy EE, Kiefer L, Sullivan MC, Simon MD. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods.* 2018;15:221–5.
70. Chen Y, Wu F, Chen Z, He Z, Wei Q, Zeng W, et al. Acrylonitrile-mediated nascent RNA sequencing for transcriptome-wide profiling of cellular RNA dynamics. *Adv Sci (Weinh).* 2020;7:1900997.
71. Li X, Zhou B, Chen L, Gou LT, Li H, Fu XD. GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol.* 2017;35:940–50.
72. Yan Z, Huang N, Wu W, Chen W, Jiang Y, Chen J, et al. Genome-wide colocalization of RNA-DNA interactions and fusion RNA pairs. *Proc Natl Acad Sci U S A.* 2019;116:3328–37.
73. Bell JC, Jukam D, Teran NA, Risca VI, Smith OK, Johnson WL, et al. Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife.* 2018;7:e27024.
74. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci U S A.* 2011;108:10010–5.
75. Metkar M, Ozadam H, Lajoie BR, Imakaev M, Mirny LA, Dekker J, et al. Higher-order organization principles of pre-translational mRNPs. *Mol Cell.* 2018;72:715–26.
76. Nguyen TC, Cao X, Yu P, Xiao S, Lu J, Biase FH, et al. Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nat Commun.* 2016;7:12023.
77. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell.* 2016;165:1267–79.
78. Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. Global mapping of human RNA-RNA interactions. *Mol Cell.* 2016;62:618–26.
79. Aw JG, Shen Y, Wilm A, Sun M, Lim XN, Boon KL, et al. In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol Cell.* 2016;62:603–17.
80. Cai Z, Cao C, Ji L, Ye R, Wang D, Xia C, et al. RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature.* 2020;582:432–7.
81. Morf J, Wingett SW, Farabella I, Cairns J, Furlan-Magaril M, Jimenez-Garcia LF, et al. RNA proximity sequencing reveals the spatial organization of the transcriptome in the nucleus. *Nat Biotechnol.* 2019;37:793–802.
82. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc.* 2015;10:442–58.
83. Taliaferro JM, Wang ET, Burge CB. Genomic analysis of RNA localization. *RNA Biol.* 2014;11:1040–50.
84. Kaewsapsak P, Shechner DM, Mallard W, Rinn JL, Ting AY. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife.* 2017;6:e29224.
85. Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell.* 2017;66:22–37.
86. Diallo LH, Tatin F, David F, Godet AC, Zamora A, Prats AC, et al. How are circRNAs translated by non-canonical initiation mechanisms? *Biochimie.* 2019;164:45–52.
87. Jiao Y, Meyerowitz EM. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol Syst Biol.* 2010;6:419.
88. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011;147:789–802.
89. Sun YM, Wang WT, Zeng ZC, Chen TQ, Han C, Pan Q, et al. circMYBL2, a circRNA from MYBL2, regulates FLT3 translation by recruiting PTBP1 to promote FLT3-ITD AML progression. *Blood.* 2019;134:1533–46.
90. Struhl K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol.* 2007;14:103–5.
91. Giambruno R, Mihailovich M, Bonaldi T. Mass spectrometry-based proteomics to unveil the non-coding RNA world. *Front Mol Biosci.* 2018;5:90.
92. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet.* 2011;12:861–74.
93. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 2004;116:281–97.
94. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21-and 22-nucleotide RNAs. *Genes Dev.* 2001;15:188–200.
95. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science.* 2007;316:744–7.
96. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, et al. Tiny RNAs associated with transcription start sites in animals. *Nat Genet.* 2009;41:572–8.
97. Ciganda M, Williams N. Eukaryotic 5S rRNA biogenesis. *Wiley Interdiscip Rev RNA.* 2011;2:523–33.
98. Abou Elela S, Nazar RN. Role of the 5.8S rRNA in ribosome translocation. *Nucleic Acids Res.* 1997;25:1788–94.
99. Goodenbour JM, Pan T. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res.* 2006;34:6137–46.
100. Bachelier JP, Cavaille J, Huttenhofer A. The expanding snoRNA world. *Biochimie.* 2002;84:775–90.
101. Hamm J, Darzynkiewicz E, Tahara SM, Mattaj JW. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell.* 1990;62:569–77.
102. Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk? *Front Genet.* 2015;6:2.
103. Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* 2014;15:409.
104. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PLoS Genet.* 2013;9:e1003777.
105. Xing YH, Yao RW, Zhang Y, Guo CJ, Jiang S, Xu G, et al. SLERT regulates DDX21 rings associated with pol I transcription. *Cell.* 2017;169:664–78.
106. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507:455–61.
107. Selenko P, Sprangers R, Stier G, Buhler D, Fischer U, Sattler M. SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol.* 2001;8:27–31.
108. Dupuis-Sandoval F, Poirier M, Scott MS. The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdiscip Rev RNA.* 2015;6:381–97.
109. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatshfte für Chemie/Chemical Monthly.* 1994;125:167–88.
110. Xu L, Sun L, Guan G, Huang Q, Lv J, Yan L, et al. The effects of pH and salts on nucleic acid partitioning during phenol extraction. *Nucleosides Nucleotides Nucleic Acids.* 2019;38:305–20.
111. Nicosia A, Tagliavia M, Costa S. Regeneration of total RNA purification silica-based columns. *Biomed Chromatogr.* 2010;24:1263–4.
112. Fromm B, Harris PD, Bachmann L. MicroRNA preparations from individual monogenean *Gyrodactylus salaris*-a comparison of six commercially available totalRNA extraction kits. *BMC Res Notes.* 2011;4:217.

113. Chen EA, Souzaia T, Herstein JS, Evgrafov OV, Spitsyna VN, Rebolini DF, et al. Effect of RNA integrity on uniquely mapped reads in RNA-Seq. *BMC Res Notes*. 2014;7:753.
114. Veneziano D, Di Bella S, Nigita G, Lagana A, Ferro A, Croce CM. Noncoding RNA: current deep sequencing data analysis approaches and challenges. *Hum Mutat*. 2016;37:1283–98.
115. Modrek B, Lee C. A genomic view of alternative splicing. *Nat Genet*. 2002;30:13–9.
116. Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res*. 2016;26:1277–87.
117. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell*. 2014;159:134–47.
118. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol*. 2015;22:256–64.
119. Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, et al. The landscape of circular RNA in cancer. *Cell*. 2019;176:869–81.
120. Guarniero J, Bezzi M, Jeong JC, Paffenholz SV, Berry K, Naldini MM, et al. Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. *Cell*. 2016;166:1055–6.
121. Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol*. 2005;15:331–41.
122. Seto AG, Kingston RE, Lau NC. The coming of age for Piwi proteins. *Mol Cell*. 2007;26:603–9.
123. Kedersha NL, Rome LH. Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J Cell Biol*. 1986;103:699–709.
124. van Zon A, Mossink MH, Schoester M, Scheffer GL, Scheper RJ, Sonneveld P, et al. Multiple human vault RNAs. Expression and association with the vault complex. *J Biol Chem*. 2001;276:37715–21.
125. Kickhoefer VA, Searles RP, Kedersha NL, Garber ME, Johnson DL, Rome LH. Vault ribonucleoprotein particles from rat and bullfrog contain a related small RNA that is transcribed by RNA polymerase III. *J Biol Chem*. 1993;268:7868–73.
126. Kiss-Laszlo Z, Henry Y, Bachelier JP, Caizergues-Ferrer M, Kiss T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*. 1996;85:1077–88.
127. Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C. Characterization of the RNA content of chromatin. *Genome Res*. 2010;20:899–907.
128. Schubert T, Pusch MC, Diermeier S, Benes V, Kremmer E, Imhof A, et al. Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Mol Cell*. 2012;48:434–44.
129. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007;129:1311–23.
130. Yazaki J, Gregory BD, Ecker JR. Mapping the genome landscape using tiling array technology. *Curr Opin Plant Biol*. 2007;10:534–42.
131. Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, et al. Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 genomes data. *PLoS One*. 2013;8:e58815.
132. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*. 2008;321:956–60.
133. Edgren H, Murumagi A, Kangaspekka S, Nicorici D, Hongisto V, Kleivi K, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol*. 2011;12:R6.
134. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25:1105–11.
135. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915–27.
136. Muller S, Rycak L, Winter P, Kahl G, Koch I, Rotter B. omiRas: a web server for differential expression analysis of miRNAs derived from small RNA-Seq data. *Bioinformatics*. 2013;29:2651–2.
137. Han BW, Wang W, Zamore PD, Weng Z. piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics*. 2015;31:593–5.
138. Hoogstrate Y, Jenster G, Martens-Uzunova ES. FlaiMapper: computational annotation of small ncRNA-derived fragments using RNA-seq high-throughput data. *Bioinformatics*. 2015;31:665–73.
139. Zhang Y, Shi J, Chen Q. tsRNAs: new players in mammalian retrotransposon control. *Cell Res*. 2017;27:1307–8.
140. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13.
141. Baker M. MicroRNA profiling: separating signal from noise. *Nat Methods*. 2010;7:687–92.
142. Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF. Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics*. 2009;25:2298–301.
143. Li Z, Ender C, Meister G, Moore PS, Chang Y, John B. Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res*. 2012;40:6787–99.
144. Giraldez MD, Spengler RM, Etheridge A, Godoy PM, Barczak AJ, Srinivasan S, et al. Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat Biotechnol*. 2018;36:746–57.
145. Wright C, Rajpurohit A, Burke EE, Williams C, Collado-Torres L, Kimos M, et al. Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics*. 2019;20:513.
146. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol*. 2018;14:479–92.
147. Warren L, Bryder D, Weissman IL, Quake SR. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc Natl Acad Sci U S A*. 2006;103:17807–12.
148. Serra L, Chang DZ, Macchietto M, Williams K, Murad R, Lu D, et al. Adapting the smart-seq2 protocol for robust single worm RNA-seq. *Bio Protoc*. 2018;8:e2729.
149. Bhargava V, Head SR, Ordoukhanian P, Mercola M, Subramaniam S. Technical variations in low-input RNA-seq methodologies. *Sci Rep*. 2014;4:3678.
150. Friedel CC, Dolken L. Metabolic tagging and purification of nascent RNA: implications for transcriptomics. *Mol Biosyst*. 2009;5:1271–8.
151. Wissink EM, Vihervaara A, Tippens ND, Lis JT. Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet*. 2019;20:705–23.
152. Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev*. 2011;25:742–54.
153. Baptista MAP, Dolken L. RNA dynamics revealed by metabolic RNA labeling and biochemical nucleoside conversions. *Nat Methods*. 2018;15:171–2.
154. Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol*. 2005;23:1383–90.
155. Fabbri M, Girmila L, Varani G, Calin GA. Decrypting noncoding RNA interactions, structures, and functional networks. *Genome Res*. 2019;29:1377–88.
156. Fu XD. Non-coding RNA: a new frontier in regulatory biology. *Natl Sci Rev*. 2014;1:190–204.
157. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145–66.
158. West JA, Davis CP, Sunwoo H, Simon MD, Sadreyev RI, Wang PI, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell*. 2014;55:791–802.
159. Sridhar B, Rivas-Astroza M, Nguyen TC, Chen W, Yan Z, Cao X, et al. Systematic mapping of RNA-chromatin interactions in vivo. *Curr Biol*. 2017;27:610–2.
160. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, et al. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A*. 2011;108:20497–502.
161. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 2011;44:667–78.
162. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 2013;341:1237973.
163. Kudla G, Wan Y, Helwak A. RNA conformation capture by proximity ligation. *Annu Rev Genomics Hum Genet*. 2020;21(1):14.
164. Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*. 2013;153:654–65.

165. Hearst JE. Psoralen photochemistry and nucleic acid structure. *J Invest Dermatol.* 1981;77:39–44.
166. Huber RG, Lim XN, Ng WC, Sim AYL, Poh HX, Shen Y, et al. Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat Commun.* 2019;10:1408.
167. Li P, Wei Y, Mei M, Tang L, Sun L, Huang W, et al. Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe.* 2018;24:875–86.
168. Sahu A, Singhal U, Chinnaiyan AM. Long noncoding RNAs in cancer: from function to translation. *Trends Cancer.* 2015;1:93–109.
169. Cabili MN, Dunagin MC, McClanahan PD, Bialesch A, Padovan-Merhar O, Regev A, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* 2015;16:20.
170. Hougaard DM, Hansen H, Larsson LI. Non-radioactive in situ hybridization for mRNA with emphasis on the use of oligodeoxynucleotide probes. *Histochem Cell Biol.* 1997;108:335–44.
171. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods.* 2008;5:877–9.
172. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, et al. *Science.* 2018;361:eaat5691.
173. Lee YH, Tan HT, Chung MC. Subcellular fractionation methods and strategies for proteomics. *Proteomics.* 2010;10:3935–56.
174. Sultan M, Amstislavskiy V, Risch T, Schuette M, Dokel S, Ralsler M, et al. Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics.* 2014;15:675.
175. Nathanson L, Xia T, Deutscher MP. Nuclear protein synthesis: a re-evaluation. *RNA.* 2003;9:9–13.
176. Rhee HW, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science.* 2013;339:1328–31.
177. Gilbert C, Kristjuhan A, Winkler GS, Svejstrup JQ. Elongator interactions with nascent mRNA revealed by RNA immunoprecipitation. *Mol Cell.* 2004;14:457–64.
178. Fazzal FM, Han S, Parker KR, Kaewwapsak P, Xu J, Boettiger AN, et al. Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell.* 2019;178:473–90.
179. Li J, Han L, Roebuck P, Diao L, Liu L, Yuan Y, et al. TANRIC: an interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res.* 2015;75:3728–37.
180. Gao Y, Wang P, Wang Y, Ma X, Zhi H, Zhou D, et al. Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.* 2019;47:D1028–33.
181. Liu Y, Zhao M. InCaNet: pan-cancer co-expression network for human lncRNA and cancer genes. *Bioinformatics.* 2016;32:1595–7.
182. Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* 2019;47:D1034–7.
183. Xia S, Feng J, Chen K, Ma Y, Gong J, Cai F, et al. CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res.* 2018;46:D925–9.
184. Ghosal S, Das S, Sen R, Basak P, Chakrabarti J. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front Genet.* 2013;4:283.
185. Fan C, Lei X, Fang Z, Jiang Q, Wu FX. CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database (Oxford).* 2018. <https://doi.org/10.1093/database/bay044>.
186. Xie B, Ding Q, Han H, Wu D. miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics.* 2013;29:638–44.
187. Bhattacharya A, Cui Y. SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. *Nucleic Acids Res.* 2016;44:D1005–10.
188. Wong NW, Chen Y, Chen S, Wang X. OncomiR: an online resource for exploring pan-cancer microRNA dysregulation. *Bioinformatics.* 2018;34:713–5.
189. Ahmed M, Nguyen H, Lai T, Kim DR. miRCancerdb: a database for correlation analysis between microRNA and gene expression in cancer. *BMC Res Notes.* 2018;11:103.
190. Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, et al. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.* 2017;45:D812–8.
191. Chung JF, Chang SJ, Chen CY, Liu SH, Li CY, Chan CH, et al. YM500v3: a database for small RNA sequencing in human cancer research. *Nucleic Acids Res.* 2017;45:D925–31.
192. Zheng LL, Xu WL, Liu S, Sun WJ, Li JH, Wu J, et al. tRF2Cancer: a web server to detect tRNA-derived small RNA fragments (tRFs) and their expression in multiple cancers. *Nucleic Acids Res.* 2016;44:W185–93.
193. Pliatsika V, Loher P, Magee R, Telonis AG, London E, Shigematsu M, 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.* 2018;46:D152–9.
194. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell.* 2015;28:529–40.
195. Guo ZW, Xie C, Li K, Zhai XM, Cai GX, Yang XX, et al. SELER: a database of super-enhancer-associated lncRNA-directed transcriptional regulation in human cancers. *Database (Oxford).* 2019. <https://doi.org/10.1093/database/baz027>.
196. Zhao Z, Wang K, Wu F, Wang W, Zhang K, Hu H, et al. circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. *Cell Death Dis.* 2018;9:475.
197. Ruan H, Xiang Y, Ko J, Li S, Jing Y, Zhu X, et al. Comprehensive characterization of circular RNAs in ~1000 human cancer cell lines. *Genome Med.* 2019;11:55.
198. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* 2009;37:D98–104.
199. Volders PJ, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, et al. LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* 2019;47:D135–9.
200. Gong J, Liu C, Liu W, Xiang Y, Diao L, Guo AY, et al. LNCediting: a database for functional effects of RNA editing in lncRNAs. *Nucleic Acids Res.* 2017;45:D79–84.
201. Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, et al. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 2015;43:D168–73.
202. Ma L, Li A, Zou D, Xu X, Xia L, Yu J, et al. lncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res.* 2015;43:D187–92.
203. Ma L, Cao J, Liu L, Du Q, Li Z, Zou D, et al. LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res.* 2019;47:D128–34.
204. Josset L, Tchitchek N, Gralinski LE, Ferris MT, Eisfeld AJ, Green RR, et al. Annotation of long non-coding RNAs expressed in collaborative cross founder mice in response to respiratory virus infection reveals a new class of interferon-stimulated transcripts. *RNA Biol.* 2014;11:875–90.
205. Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, et al. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.* 2016;44:D203–8.
206. Wu W, Ji P, Zhao F. CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.* 2020;21:101.
207. Glazar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *RNA.* 2014;20:1666–70.
208. Dong R, Ma XK, Li GW, Yang L. CIRCPedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinformatics.* 2018;16:226–33.
209. Xia S, Feng J, Lei L, Hu J, Xia L, Wang J, et al. Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform.* 2017;18:984–92.
210. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014;42:D92–7.
211. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 2018;46:D296–302.
212. Panwar B, Omenn GS, Guan Y. miRmine: a database of human miRNA expression profiles. *Bioinformatics.* 2017;33:1554–60.
213. Liu T, Zhang Q, Zhang J, Li C, Miao YR, Lei Q, et al. EVmiRNA: a database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res.* 2019;47:D89–93.
214. Andres-Leon E, Gonzalez Pena D, Gomez-Lopez G, Pisano DG. miRGate: a curated database of human, mouse and rat miRNA-mRNA targets. *Database (Oxford).* 2015. [doi.org/https://doi.org/10.1093/database/bav035](https://doi.org/10.1093/database/bav035).
215. Kozomara A, Birgaouan M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 2019;47:D155–62.

216. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.* 2018; 46:D239–45.
217. Kuksa PP, Amlie-Wolf A, Katanic Z, Valladares O, Wang LS, Leung YY. DASHR 2.0: integrated database of human small non-coding RNA genes and mature products. *Bioinformatics.* 2019;35:1033–9.
218. Zhang T, Tan P, Wang L, Jin N, Li Y, Zhang L, et al. RNALocate: a resource for RNA subcellular localizations. *Nucleic Acids Res.* 2017;45:D135–8.
219. Mas-Ponte D, Carlevaro-Fita J, Palumbo E, HERNANDEZ Pulido T, Guigo R, Johnson R. LncATLAS database for subcellular localization of long noncoding RNAs. *RNA.* 2017;23:1080–7.
220. Wang Z, Tollervey J, Briese M, Turner D, Ule J. CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. *Methods.* 2009;48:287–93.
221. Hao Y, Wu W, Li H, Yuan J, Luo J, Zhao Y, et al. NPInter v3.0: an upgraded database of noncoding RNA-associated interactions. *Database (Oxford).* 2016; doi.org/<https://doi.org/10.1093/database/baw057>.
222. Wong NK, Huang CL, Islam R, Yip SP. Long non-coding RNAs in hematological malignancies: translating basic techniques into diagnostic and therapeutic strategies. *J Hematol Oncol.* 2018;11:131.
223. Amodio N, Raimondi L, Juli G, Stamato MA, Caracciolo D, Tagliaferri P, et al. MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *J Hematol Oncol.* 2018;11:63.
224. Tie Y, Liu B, Fu H, Zheng X. Circulating miRNA and cancer diagnosis. *Sci China C Life Sci.* 2009;52:1117–22.
225. Rupaimoole R, Calin GA, Lopez-Berestein G, Sood AK. miRNA deregulation in cancer cells and the tumor microenvironment. *Cancer Discov.* 2016;6: 235–46.
226. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn.* 2020;20:453–4.
227. Holdt LM, Stahringer A, Sass K, Pichler G, Kulak NA, Wilfert W, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun.* 2016;7:12429.
228. Carlevaro-Fita J, Johnson R. Global positioning system: understanding long noncoding RNAs through subcellular localization. *Mol Cell.* 2019;73:869–83.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

