Genes & Diseases (2023) 10, 2491–2510

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

REVIEW ARTICLE

Long noncoding RNA study: Genome-wide approaches

Shuang Tao [a], Yarui Hou [a], Liting Diao [a], Yanxia Hu [a], Wanyi Xu [a], Shujuan Xie [a,b,**], Zhendong Xiao [a,*]

[a] *The Biotherapy Center, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510630, China*
[b] *Institute of Vaccine, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong 510630, China*

Received 11 June 2022; received in revised form 9 October 2022; accepted 23 October 2022
Available online 29 November 2022

KEYWORDS
Cancer-related lncRNAs;
Computational pipelines;
Genome-wide study;
Long noncoding RNAs;
Technological approaches

Abstract Long noncoding RNAs (lncRNAs) have been confirmed to play a crucial role in various biological processes across several species. Though many efforts have been devoted to the expansion of the lncRNAs landscape, much about lncRNAs is still unknown due to their great complexity. The development of high-throughput technologies and the constantly improved bioinformatic methods have resulted in a rapid expansion of lncRNA research and relevant databases. In this review, we introduced genome-wide research of lncRNAs in three parts: (i) novel lncRNA identification by high-throughput sequencing and computational pipelines; (ii) functional characterization of lncRNAs by expression atlas profiling, genome-scale screening, and the research of cancer-related lncRNAs; (iii) mechanism research by large-scale experimental technologies and computational analysis. Besides, primary experimental methods and bioinformatic pipelines related to these three parts are summarized. This review aimed to provide a comprehensive and systemic overview of lncRNA genome-wide research strategies and indicate a genome-wide lncRNA research system.

* Corresponding author.
** Corresponding author. Institute of Vaccine, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510630, China
E-mail addresses: xieshj5@mail.sysu.edu.cn (S. Xie), xiaozhd@mail2.sysu.edu.cn (Z. Xiao).
Peer review under responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2022.10.024

Introduction

Rather than protein-coding genes, noncoding transcripts account for the majority of the genome, especially long noncoding RNAs (lncRNAs). Normally, lncRNA refers to the transcript with lengths over 200 nucleotides and does not be translated into proteins.[1] Various subtypes of lncRNAs constitute a highly heterogeneous group including enhancer lncRNAs, sense/antisense lncRNAs, and intronic/intergenic lncRNAs.[2,3] LncRNAs were once thought to be "transcriptional noise", but their purifying selection of promoters, such as substitutions, insertions, deletions, and splicing, indicated functionality.[4] Currently, lncRNAs have been proven to play an important regulatory role in a variety of organismal physiological and pathological conditions by sensory, guiding, scaffolding, or allosteric ability.[5–7] Furthermore, lncRNAs also serve a role to play in different levels of genetic information transmission, such as chromatin reprogramming, transcriptional regulation (e.g., cis-regulation at enhancers), posttranscriptional regulation (e.g., mRNA processing) and translational regulation,[8–13] making lncRNA promising candidate as a biomarker and therapeutic target. We have briefly summarized the origin and function of lncRNAs in Figure 1. Compared to protein-coding messenger RNAs (mRNAs), lncRNAs tend to be confined to chromatin and have great sequence complexity, with reduced stability, shorter transcripts, lower GC content, fewer exons, low polyadenylation efficiency, less alternative splicing efficiency, higher cell-type specificity, lower evolutionary conservation and less expression than mRNAs,[14–16] which resulted in the challenging research status of lncRNAs.

A better understanding of lncRNAs will open up new vistas in the genomic landscape and shed new light on

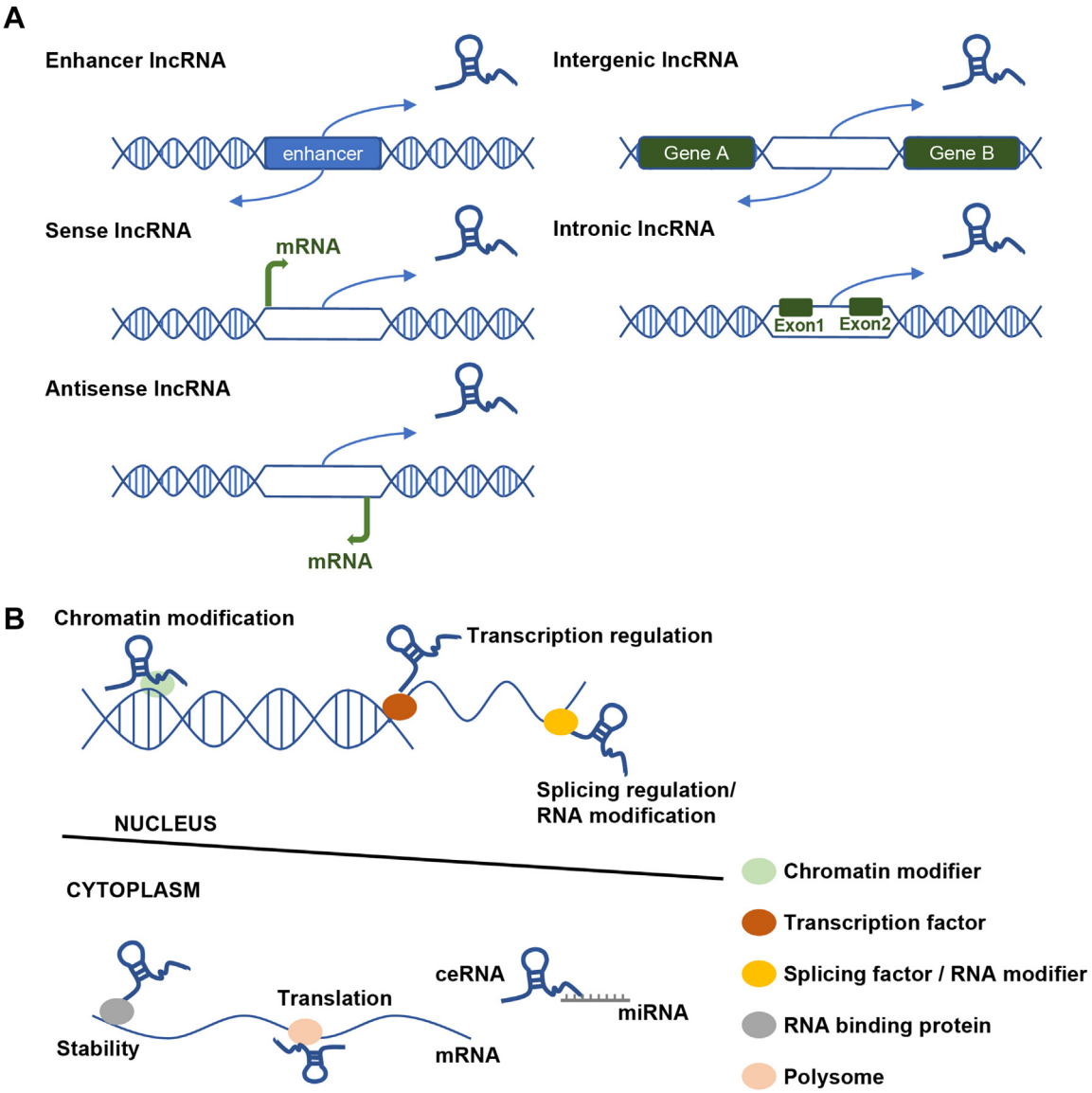

Figure 1 The origin and function of lncRNAs. **(A)** Schematic diagram of the origin of different types of lncRNAs. **(B)** Schematic diagram of the function of lncRNAs.

certain complicated biological mechanisms, providing novel therapeutic targets. However, many efforts have been dwarfed by the complexity of the lncRNA world probably due to their low expression and high natural expression variability.[15] Therefore, genome-scale research will provide indispensable help and boost lncRNA research, given the comprehensiveness of data resources. Here, we will provide a brief overview of genome-wide research strategies for the annotation and characterization of lncRNAs, outlining primary large-scale methods, which can bring new insights into global knowledge of lncRNAs.

Genome-wide identification of novel lncRNAs

Benefiting from the straightforward transmission of information from codons to amino acids of coding genes or messenger RNAs (mRNAs), it is feasible to recognize functional elements such as untranslated regions (UTRs) and open reading frames (ORFs) precisely based only on primary sequencing data.[17] However, the structure-based functional mechanism of lncRNAs, including secondary structures with biological relevance, limits the utilization of *in silico* primary sequencing data analysis.[18]

Identifying novel lncRNAs is the foundation for lncRNAs study. A complete and comprehensive genomic annotation, which refers to full genomic architecture and boundary information, was the cornerstone of the follow-up functional and mechanistic studies. Current approaches to identify novel transcripts at the genome scale are based on numerous raw RNA-sequencing (RNA-seq) data combined with bioinformatic analysis pipelines containing alignments, assembly, quantification, and difference calculations (Fig. 2), and constantly emerging systematic bioinformatic technologies strive for more accurate and comprehensive annotations.

Computational pipelines for RNA-seq-analysis

Efficient and accurate analysis pipelines are essential to decoding the large-scale and great complexity of RNA-seq datasets. The early RNA-seq analysis pipeline published in 2012 contains two main software programs: TopHat and Cufflinks.[19] Raw reads from RNA-seq were firstly aligned by TopHat to the reference genome, and non-annotated reads, which were strong candidates for novel noncoding RNAs, were then assembled into individual transcripts by Cufflinks. TopHat was created on the basis of Bowtie, a popular and efficient read alignment program, to solve the limitations in reads containing introns or large gaps of Bowtie.[20] The Cufflinks package is composed of several utility programs to be used in combination, such as Cuffmerge, which integrates novel transcripts with annotated transcripts for more complete annotations, Cuffcompare, which identifies unannotated transcripts by comparison with the reference genome, and Cuffdiff, which calculates differential transcript expression, splicing and promoter use.[19]

In 2016, a new pipeline was reported with improved speed and accuracy, including HISAT (hierarchical indexing for spliced alignment of transcripts) and StringTie. HISAT, which is much faster with low computer memory, serves the same role as TopHat.[21] Alignments identified by HISAT were assembled into full or partial transcripts by StringTie, which runs faster and generates a more complete transcriptome than other assemblers.[22,23]

Though widely utilized, the pipelines above cannot handle all kinds of RNA-seq analysis because of the need for a reference genome. Therefore, if a well-established reference genome does not exist or is poorly annotated, *de novo* transcriptome assembly without an alignment is required. *De novo* assembler directly assembles RNA-seq reads into transcripts after finding their overlap regions.

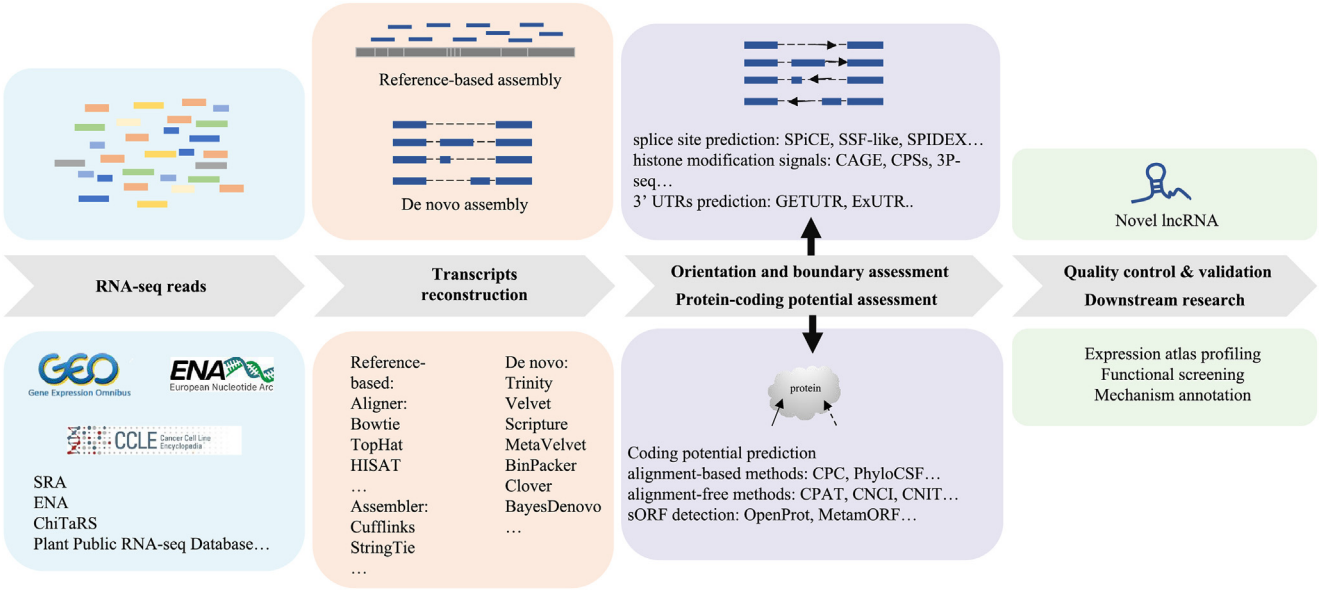

Figure 2 The workflow of genome-wide identification of novel lncRNAs.

The Trinity platform, a *de novo* transcriptome assembler, was first published in 2011 and contains three consecutive working modules, Inchworm for transcript contig generation, Chrysalis for transcript contig clustering and de Bruijn graph, and Butterfly for full-length transcript reporting.[24,25]

Alternative alignment and assembly algorithms

In addition to the main pipelines introduced above, other alignment and assembly packages have been developed to be used in combination with or instead of existing tools. These alternative algorithms were designed with increased accuracy, speed, and some improvements to solve the limitations of existing methods. We briefly summarized these alternative methods with their brief description and presented them in diagrammatic form in Table 1.

Table 1 Alternative algorithms for RNA-seq analysis and transcripts reconstruction.

Name	Brief description	Reference
Alternative algorithms for RNA-seq reads alignment		
SpliceMap	Splice junction detection; 50—100 nt reads	[256]
MapSplice	Splice junction detection; short (<75 bp) and long (≥75 bp) paired reads; higher sensitivity, efficiency and specificity	[257]
GSNAP	Variants and splicing detection; short (as 14 nt) and arbitrarily long single- or paired-end reads	[258]
STAR	>50-fold mapping speed than other aligner	[259]
TopHat2	Alignment across fusion breaks; handle repetitive genomes and pseudogenes	[260]
Minimap2	General purpose; higher speed and accuracy	[261,262]
deSALT	Long RNA-seq read alignment; graph-based	[263]
2passtools	Long-read sequencing reads; filter spurious splice junctions	[264]
Alternative algorithms for transcript assembly		
CEM	Transcripts assembly and abundance estimation; RNA-seq bias capture	[265]
CLASS	Genome-guided assembler; locally reconstruct exons into a splice graph	[266]
Traph	Isoform identification and quantification; based on minimum-cost network flows	[267]
iReckon	Probabilistic approach; isoforms identification and abundance estimation simultaneously	[268]
Bayesembler	Genome-guided assembler; based on a Bayesian model; provide confidence and abundance estimate	[269]
FlipFlop	Isoform identification and quantification; reads cover multiple exons	[270]
TransComb	Genome-guided assembler; based on junction graph	[271]
CLASS2	Genome-guided assembler; extension of CLASS; optimized read patterns and supporting reads	[272]
CIDANE	Genome-based reconstruction; *ab initio* reconstruction; transcripts quantification	[273]
Scallop	Reference-based assembler;	[274]

Filtration of assembled transcripts

Orientation and boundary information

Assembly of RNA-seq reads, especially through *de novo* assembly methods, requires strand information instead of a reference genome to provide orientation and precise boundary information. For the genome-guided assembly approach, the splicing signal could indicate the orientation of some reads spanning exon junctions.[26] Frequently-used computational approaches for splice site prediction are Splicing Prediction in Consensus Elements (SPiCE), Splice-SiteFinder-like (SSF-like), MaxEntScan, NNSplice, HSF, SPANR, dbscSNV tools, Human Splicing Finder (HSF), pre-computed index of splicing variants (SPIDEX), Spliceogen and the APPRIS Database.[27—31] Non-stranded RNA-seq reads are liable to cause incorrect transcript annotation, such as chimeras.[32] Rather, stranded sequencing reads are more beneficial to transcript assembly. It is worth mentioning that non-stranded reads can be converted into stranded reads after direction prediction by k-order Markov chain models (kMC) with maximum likelihood estimation (MLE).[26] For this reason, integration of stranded and non-stranded reads is recommended.

Histone modification signals were once used to identify promoters and active genes.[33] Transcription start sites (TSSs) identified by cap analysis of gene expression (CAGE)[34] as well as cleavage and polyadenylation sites (CPSs) determined using poly(A)-position profiling by sequencing (3P-seq)[35] improve the boundary information of transcription fragments. In addition, GETUTR[36] and ExUTR[37] have been introduced to accurately predict 3′ UTRs. Paired-end ditag (PET) data could provide more efficient boundary information with respect to both end tags.[38]

Protein-coding potential assessment

Classification of coding and noncoding transcripts is a key step of novel lncRNA identification. Due to the species-specificity and functional complexity of lncRNAs, it is hard to discriminate coding and noncoding transcripts merely by some straightforward methods based on intrinsic sequence

Table 1 (*continued*)

Name	Brief description	Reference
	for transcripts with multi-exon or low expression	
scRNAss	Transcripts reconstruction from single-cell RNA-seq data	275
PsiCLASS	Genome-guided assembler; multi-sample approach	276
IntAPT	Phenotype-specific transcripts assembly; <100 samples	277
Velvet	*De novo* short read assembler; de Bruijn graphs	278
Scripture	*Ab initio* transcripts reconstruction	33
MetaVelvet	*De novo* assembler; short reads from multiple species	279
BinPacker	*De novo* assembler; bin-packing strategy; better than Scripture with dog data; short in human data than Trinity	280
Clover	*De novo* assembler; based on de Bruijn graph	281
BayesDenovo	*De novo* assembler; read-guided; Bayesian strategy	282

features and locus information. So constantly emerging methods have been developed to identify noncoding transcripts considering their complex characteristics and functionalities (Table 2). Currently, computational coding potential prediction methods mainly have two categories. On one hand, sequence alignment for homology search or phylogenetic conservation calculation is utilized in several methods like the coding potential calculator (CPC)[39] and PhyloCSF.[40] These alignment-based methods are limited by their dependence on highly conserved protein-coding genes, alignment quality and long work time. It cannot analyze the lncRNAs transcribed from the sense or antisense strand of protein-coding genes. On the other hand, alignment-free methods like the Coding Potential Assessment Tool (CPAT), Coding-Non-Coding Index (CNCI), PLEK (predictor of long non-coding RNAs and messenger RNAs based on an improved k-mer scheme), lncScore, Coding-Non-Coding Identifying Tool (CNIT), CPPred and DeepCPP (deep neural network for coding potential prediction) were developed to solve the limitations of alignment-based approaches. CPAT[41] uses a logistic regression model based on four sequence features to assess coding potential with improved accuracy and speed for the human genome. CNCI[42] employs adjoining nucleotide triplets profiling and outperforms other methods for species with poorly annotated information. PLEK[43] is based on an improved k-mer scheme and support vector machine (SVM) and is good at cross-species prediction. lncScore[44] uses a logistic regression model combined with several ORF- and exon-related features including maximum coding subsequence (MCSS), and it can handle both partial- and full-length data. CNIT[45] is an improved version of CNCI with higher speed and accuracy than CNCI, especially for plants. CPPred[46] is based on sequence features including novel CTD features and the SVM model. The CPPred has better performance in small coding and noncoding transcripts distinguish than other tools. DeepCPP[47] is a deep learning method to estimate the coding potential of RNAs with novel feature representation, ranking, and selecting methods to improve its efficiency.

Table 2 Methods for protein-coding potential assessment and sORF detection.

Name	Description	Reference
Alignment-based methods		
CPC	Based on six sequence features	39
PhyloCSF	Multispecies alignment; statistical comparison of phylogenetic codon models	40
Alignment-free methods		
CPAT	Four sequence features; logistic regression model	41
CNCI	Species with poorly annotated information	42
PLEK	Cross-species prediction	43
lncScore	ORF- and exon-related features including maximum coding subsequence (MCSS); handle both partial- and full-length data	44
CNIT	Improved version of CNCI; especially for plants	45
CPPred	Sequence features including novel CTD features	46
DeepCPP	Deep learning method; novel features representation, ranking and selecting methods	47
sORF detection		
DeepCPP	Deep learning method; novel features representation, ranking and selecting methods	47
sORFs.org	A database containing sORF information across six species	63
OpenProt	ORF prediction resources	64
MetamORF	Database containing sORF information	65

However, it is outdated to regard lncRNAs as "without coding potential" due to the identification of small bioactive peptides encoded by short open reading frames (sORFs, ≤300 bases) in annotated lncRNA transcripts.[48–50] These small peptides (≤100 amino acids), which are called micropeptides or small ORF-encoded peptides (SEPs), have been shown to serve important roles through various biological activities, such as inflammation and oncogenic processes.[51,52] In consideration of the possible translational potential, the identification of lncRNA-encoded micropeptides is indispensable for lncRNA annotation in addition

to coding potential prediction. However, it is still a great challenge to annotate these peptides owing to their low molecular weight, low abundance, and limited annotation methods. Currently, methods for micropeptides detection include high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS) and RNA-Seq.[53,54] With the help of high-throughput sequencing, Ribosome profiling (Ribo-seq) can identify ribosome-related lncRNAs or lncRNAs with translational potential. Ribo-seq is based on ribosome-protected RNA fragments (RPFs) after translation inhibition to map translation status.[55]

Based on the practice of short peptides annotation, several databases have been established for researching of lncRNA coding potential. SmProt[56] curates over 200 thousand small proteins encoded by noncoding regions from eight species. FuncPEP[57] focuses on the functional analysis of ncRNA-encoded peptides. ncEP database[58] is a manually curated database for experimentally validated ncRNA-encoded peptides, which contains 74 peptides across 18 species. cncRNAdb,[59] also a manually curated database, contains translated coding and noncoding RNAs of over 20 species. LncPep[60] is a database, especially for lncRNAs-encoded peptides. It contains more than 10 million lncRNA-encoded peptides across 39 species based on integrated evidence like CPC2, CPAT, m^6A (N^6-methyladenosine modification of RNA), Ribo-seq, and TIS (translation initiation site). TransLnc[61] is another database of lncRNA-encoded peptides. It curates lncRNAs with translation potential from several databases and integrates evidence including m^6A, ribosome association, internal ribosome entry sites (IRESs), ORF prediction, and MS. SPENCER[62] collects ncRNA-coded small peptides based on MS datasets between tumor samples and normal samples. Those emerging databases provide useful tools for lncRNA coding potential study.

For sORF detection, most coding potential prediction methods seem not to have this capacity. DeepCPP[47] is an exception with superior performance on sORF detection. sORFs.org[63] is a database containing sORF information from ribosome profiling across six species. OpenProt[64] is an ORF prediction resource based on Ribo-seq datasets and MS datasets, and it also can predict sORF by annotating non-AUG initiating codons. MetamORF[65] is a database containing sORF information from human and mouse genomes. MetamORF combines both experimental and computational datasets to characterize sORFs information. Other methods for sORF annotation can be found in a previous review.[66]

Quality control and validation

The robustness and accuracy of transcriptome annotation can be estimated by the sensitivity and specificity using parameters such as false negatives (FNs), false positives (FPs), true negatives (TNs), and true positives (TPs) when comparing the assembled transcriptome with reference annotation by Cuffcompare at the base and intron levels.[26] Otherwise, confidence scores can be assigned to transcripts based on annotation status to assess the robustness of annotation.[67] RT-PCR, real-time polymerase chain reaction, and amplification can be employed to assess the quality of transcriptome annotation and RT-PCR-seq can be used to test the splice site loci in annotation.[68] The boundary information of transcriptome annotation can be evaluated by 5′- and 3′- nested- RACEs (rapid amplification of cDNA ends) or transcription factor binding sites (TFBSs) counted upstream of the TSSs and poly(A) signals around the CPSs.[26] To count the number of reads assigned to each genomic feature, which refers to read summarization, featureCounts, and HTSeq-count can be applied.[69,70]

Expanded RNA-seq data resources

The sources of RNA-seq data keep pace with the development of RNA-sequencing technologies. In the earliest period of the 21st century, Sanger sequencing of full-length complementary DNAs from different tissues and developmental stages was the main source of genomic annotation.[71,72] Then, the emergence of massively parallel cDNA sequencing (RNA-seq) boosted the enrichment and improvement of transcription datasets. The combination of rapid amplification of cDNA ends (RACE) and high-density tiling array technologies or long-read RNA-seq, as well as targeted tiling arrays, revealed the deep and complicated transcriptome with known short regions.[73–75] Currently, third-generation long-read sequencing including Pacific Biosciences (PacBio) single-molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies (ONT) nanopore sequencing has been developed.[76,77] Based on PacBio, Iso-seq was developed to yield full-length sequencing reads for the transcriptome.[78] Apart from the above techniques, several emerging sequencing methods were developed to perfect transcriptome annotation, such as RNA Capture Long Seq (CLS), combining RNA capture with third-generation long-read sequencing,[79] highly accurate long high-fidelity (HiFi) reads generated by circular consensus sequencing (CCS)[80] and single-cell and single-nucleus RNA-sequencing.[81] Detailed information about the development timeline and primary methods of RNA-seq can be found in our previous study.[82]

Rapidly developed RNA-sequencing technologies bring a massive resource of RNA-seq reads with constantly improving accuracy and resulting RNA-seq read databases cause great convenience for novel transcripts identification, such as the Gene Expression Omnibus (GEO) database,[83] Sequence Read Archive (SRA), the Cancer Cell Line Encyclopedia (CCLE),[84] The European Nucleotide Archive (ENA),[85] ChiTaRS,[86] Plant Public RNA-seq Database[87] and so on. Based on RNA-seq technologies and raw reads databases, several genomic and transcriptomic research projects were able to carry out, such as RefLnc (Reference catalog of LncRNA),[15] MiTranscriptome,[67] GENCODE,[88] lncRNAdb,[89] NONCODEV6,[90] CHESS,[91] FANTOM (the Functional Annotation of the Mammalian genome),[92] Human BodyMap,[93] and BIGTranscriptome.[26] The accumulation of a large amount of RNA-sequencing data provides a sound basis for the biological annotation of lncRNAs.

The practice of novel lncRNA identification

Based on the above alignment and assembly approaches, genome-wide identification of lncRNAs could be carried out in different species and biological processes. For *Arabidopsis thaliana*, 6,480 long intergenic noncoding RNAs (lincRNAs) were identified by utilizing TopHat, Cufflinks,

and Cuffcompare, and their tiling array-based method was referred to as RepTAS (reproducibility-based tiling array analysis strategy).[94] By RepTAS, also, more than 30 thousand long noncoding natural antisense transcripts (lncNATs) were identified in the Arabidopsis genome and their light reactivity was founded.[95] For long intervening noncoding RNA (lincRNA) identification in humans and mice, a computational protocol was developed. This protocol used Tophat and Cufflinks for transcript alignment and assembly as well as CNCI and ncFANs for coding potential prediction and function prediction.[96] Although much has been accomplished, several remaining problems block our way: unknown total lncRNA number, ignorance of intragenic lncRNAs and monoexonic transcripts, and incomplete annotation of transcripts.[79]

Expression atlas profiling of lncRNAs

It is widely accepted that a strong correlation exists between expression pattern and functionality, so based on increasingly expanded lncRNA databases, expression atlas profiling includes expression level, co-expression analysis, and condition-specific expression characterization provides valuable information for functional lncRNA and their interacting partners in several biological processes as well as disease-related lncRNAs, especially cancer.

Data curation and normalization

Several databases are available for curated lncRNAs, such as EVlncRNAs,[97] LncExpDB,[98] LNCipedia,[99] lncRNAWiki 2.0,[100] LncBook,[101] AnnoLnc,[102] AnnoLnc 2,[103] lncRNAtor,[104] deepBase v2.0,[105] SSTAR (semantic catalog of samples, transcription initiation, and regulators),[106] ncFANs v2.0,[107] and LncSEA.[108] After large-scale data curation, the bias resulting from the variation of different resources, including library size, experimental conditions, and read length, necessitates data normalization. Several data normalization tools can be employed, such as quantile normalization, cyclic loess, and contrast-based methods,[109] reads per kilobase per million mapped reads (RPKM),[110] and DESeq.[111] Although many normalization tools are available, a generally accepted method is absent. Existing algorithms have shortcomings, such as being sensitive to the removal of low-expression genes and lacking alignment across different samples.[112] Therefore, an accurate and powerful normalization method is expected in the future.

Clues from expression

Exploring the temporal and spatial expression patterns of lncRNAs is a powerful method to reveal their function. Algorithms such as ROKU[113] and SpeCond[114] are available. LncRNAs that play fundamental roles during development and differentiation may be spatially and temporally regulated.[115,116] In addition, lncRNAs with high conservation across different species are potential candidates as key regulators of important cellular functions.

Co-expression analysis, such as guilt-by-association analysis, was developed based on the notion that co-expression indicates coregulation and similar functions, thus enabling inference of the potential functions of lncRNAs based on interacting molecules and pathways.[117,118] Genome-wide clustering and network-based approaches have emerged as subtypes of guilt-by-association approaches. Both two methods can be used to identify functional associations of transcript groups by coregulation produced by clustering and networks.[118] Widely used methods of genome-wide clustering are self-organizing maps (SOMs),[119] k-means clustering,[120] and hierarchical clustering.[121] Network construction depends on computational and statistical algorithms such as weighted gene co-expression analysis (WGCNA).[122] After co-expression analysis, enrichment analysis is a necessary step to interpret the biological functions of clustered or grouped transcripts. Such enrichment analysis methods including gene set enrichment analysis (GSEA), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG)[123] and Reactome[124] pathway enrichment analysis as well as Ingenuity Pathway Analysis (IPA),[125] can be used.

Functional screening and cancer association of lncRNAs

Functional screening

Another important approach to characterize functional lncRNAs is the analysis of gain or loss of function to identify crucial players in certain biological states. To achieve the goal of genome-scale functional screening, estimating the effect after activating or depleting lncRNAs on a massive scale in cultured cells is feasible. RNA interference (RNAi)-based approaches have been applied in genome-wide screening for key regulators in several biological processes.[126–128] However, the unintended off-target effects and incomplete depletion have become bottlenecks hindering its extensive application in the research of functional molecules.[129,130] In recent years, the CRISPR/Cas9-based method has made efficient functional screening possible. The CRISPR/Cas9 system comprises CRISPR (clustered regularly interspaced short palindromic repeats) and the Cas9 nuclease (clustered regularly interspaced short palindromic repeats-associated protein 9), which is directed by target-specific guide RNA (gRNA) to create double-strand breaks (DSBs), causing mutations in the desired genome loci.[131] CRISPR-based functional screening utilizes a pooled lentiCRISPR library, in which each lentiviral vector carries Cas9, gRNA, and a selection marker: transduction into cultured cells results in parallel, high-throughput genomic perturbations. After a certain period of culture and phenotypic identification, such as a positive screen or negative screen, hits can be determined based on the depletion or enrichment of gRNAs in selected cells by sequencing.[132–137] The database of sgRNAs, CRISPRlnc, collects validated sgRNAs for lncRNAs of all species and can be employed to effectively design sgRNAs.[138] Various computational algorithms can be used in the data analysis for pooled CRISPR screens, such as MAGeCK,[139] MAGeCK-VISPR,[140] MAGeCKFlute,[141] CRISPRBetaBinomial (CB),[142] HiTSelect,[143] CRISPhieRmix,[144] CERES[145] and JACKS,[146] and

Table 3 Data analysis algorithms for CRISPR screening.

Name	Description	Reference
MAGeCK	Identify both positively and negatively selected genes simultaneously	[139]
MAGeCK-VISPR	Quality control measures, visualization workflow	[140]
MAGeCKFlute	Combines the MAGeCK and MAGeCK-VISPR algorithms; additional downstream analysis functionalities	[141]
CB	Based on the beta-binomial distribution	[142]
HiTSelect	Rigorously selection of screen hits and identification of functionally relevant genes and pathways	[143]
CRISPhieRmix	Using a hierarchical mixture model with a broad-tailed null distribution	[144]
CERES	Estimate gene-dependency levels accounting for the copy number-specific effect	[145]
JACKS	Jointly analysis based on a Bayesian method	[146]
PBNPA	Computing *P* value by permutation twice	[283]

some of their performances have been compared.[147] The algorithms above are listed in Table 3.

Previously, CRISPR-based screening has achieved three types of genomic perturbations: (i) genomic deletion by paired guided RNA associated with Cas9 (CRISPR-del) [148]; (ii) transcriptional activation by catalytically inactive Cas9 (dCas9) fused with activator-like VP64 (CRISPRa) [134]; and (iii) transcriptional repression by dCas9 combined with KRAB domain (CRISPRi).[133] These kinds of screening have been applied to many studies to identify functional lncRNAs from several biological processes, such as normal or cancer cell growth, cell proliferation, drug resistance of tumor cells, the regulation of the transcription factor NF-κB and epidermal homeostasis.[117,148–152] Moreover, unique characteristics of lncRNAs, such as cell type-specificity and cis-/trans-function, could be determined after delving into copy number fold-changes after screening.[117,152] CRISPR/Cas9-based genome-wide screening has already been utilized in the characterization of onco-lncRNAs, and several successful practices in drug resistance and cell proliferation have already been reviewed.[153]

Nevertheless, to conduct high-specificity, sensitive and efficient screening, there are some issues to be noted. First, lncRNAs often tend to overlap with other functional elements, so it is slightly difficult to focus on desired loci without affecting the expression of neighboring genes.[154] Hence, complete, comprehensive, and accurate genomic annotation of lncRNAs is required for the preparation of gRNA libraries. Notably, Liu et al used a genome-wide sgRNA library targeting splicing sites of lncRNAs to cause exon skipping or intron retention, thus disrupting the expression without changing the reading frame.[155] Second, due to the complexity of phenotypic detection techniques, the application of pooled-library screening is limited to cell growth or proliferation, drug similarity, and sortable protein markers. Improving accessibility to even more complex phenotypes and pushing existing boundaries are critical for the characterization of more key players in the noncoding genome. According to recent reports in the literature, imaging-based approaches such as fluorescence in situ hybridization (FISH), live cell microscopy utilizing single-molecule fluorescence time-lapse imaging, single-cell RNA-sequencing, and CyTOF (cytometry by time-of-flight) mass cytometry have been adopted to broaden the application of CRISPR/Cas9-based genome-wide screening to cellular or subcellular biological processes.[156–163] Third, to minimize the false-positive and false-negative rates in high-throughput pooled-based screening, researchers have attempted to embed internal barcodes (iBARs) in gRNAs, sequence labels (RSLs) or unique molecular identifiers (UMIs) to monitor their activity.[164–166] Fourth, knockout induced by CRISPR-Cas9 can be rescued by translation reinitiation, exon skipping, transcriptional adaptation, and other mechanisms.[167,168] Residual expression exerts a great influence on the efficiency of gene editing and the accuracy of functional screening.

In addition to Cas9, CRISPR-Cas13, as a new platform for RNA engineering, has been developed for targeted RNA knockdown with improved specificity, minimal off-target effects, and maintained RNA binding activity.[169,170] Without DNA perturbation, the CRISPR-Cas13 system has great potential to be employed in large-scale functional screening for several lncRNA effector proteins, such as PguCas13b and RfxCas13d.[171] Overall, CRISPR-based functional screening has already been utilized in many aspects of biological research as a crucial large-scale approach. Due to its advantages of high throughput, speed, and specificity, lncRNAs with important functionalities in a certain condition could be identified on a genome-wide scale.

Disease-associated lncRNAs

Disease-associated lncRNAs are usually characterized by specific expression in certain disease conditions and differential expression according to different disease statuses. LncRNAs have been reported to extensively participate in various physiological processes and be closely related to various diseases, like Alzheimer's disease (AD),[172] HIV infection/acquired immunodeficiency syndrome (AIDS),[173] asthma,[174] cardiovascular disease,[175] diabetes,[176] and liver fibrosis.[177] Due to the growing number of disease-associated lncRNAs, multiple databases have been established for their curation and validation, such as TANRIC (The Atlas of Noncoding RNAs in Cancer),[178] Lnc2Cancer (cancer-associated lncRNAs),[179] LncRNADisease 2.0 (lncRNA-disease associations),[180] LincSNP (to identify disease-associated SNPs in human long intergenic noncoding RNAs),[181] and MNDR v3.0.[182]

Cancer-associated lncRNAs

Cancer as a type of disease with a high incidence and fatality rate attracts a lot of attention to its pathogenesis,

progression, and therapeutic strategies. With the increasing number of lncRNAs have been identified and annotated, their role as key regulators during different dimensions of cancer biology was constantly revealed.[183–185] Based on the newest edition of cancer hallmarks like sustaining proliferation signals, activating invasion and metastasis, inducing angiogenesis, resisting cell death, unlocking phenotypic plasticity, reprogramming energy metabolism, genome instability and mutation, avoiding immune destruction and tumor-promoting inflammation,[186] we briefly classified cancer-associated lncRNAs according to different cancer types and present them in diagrammatic form (Table 4 and Fig. 3).

The genome-wide characterization of cancer-related lncRNAs primarily based on databases containing cancer genomic information and computational tools for data analysis. Such cancer genome databases include TCGA (The Cancer Genome Atlas) and ICGC (International Cancer Genome Consortium). According to these databases and high-throughput computational tools, several databases and tools have been developed for the cancer-related lncRNAs and the roles of lncRNAs in cancer have been expanded. The accumulation of somatic mutations in a genome may lead to the tumorigenesis of cancer. ExInAtor[187] is a tool developed for detecting cancer-driver lncRNAs, which are assumed to bear excess somatic mutations. ExInAtor uses a parametric statistical test with the input of short nucleotide variant (SNV) mutations from cancer genomic databases and lncRNA annotation. ncDR[188] and RNAactDrug[189] are databases collecting drug-resistance-related lncRNAs and RNAs, respectively. D-lnc[190] is a database containing the perturbation of drugs on lncRNA expression, which can facilitate the research of lncRNAs as drug resistance biomarkers and drug sensitivity estimation. LncACTdb 3.0[191] is a database containing more than 1000 potential cancer diagnostic and therapeutic lncRNAs biomarkers. lncRNA Lnc2Cancer,[192] which has been updated to version 3.0, is a web resource containing experimentally cancer-related lncRNAs, including their expression, functional mechanism, and clinical relevance. Lnc2Cancer also provides computational tools for RNA-seq and scRNA-seq dataset analysis in the characterization of cancer-related lncRNAs. lncRNAfunc[193] is a database curating lncRNAs from 33 cancer types. lncRNAfunc characterizes these lncRNAs by their interacting partners (such as RNA, DNA, and proteins) and underlying mechanisms (such as alternative splicing and transcription).

Mechanistic annotation

Defining possible interacting partners, functional mechanisms and putative functional elements within the

Table 4 A brief summary of cancer-associated lncRNAs.

lncRNA	Cancer type	Related cancer hallmark	Role	Reference
LINC01123	Lung cancer	Reprogramming energy metabolism	Positive	284
AFAP1-AS1	Lung cancer	Activating invasion and metastasis	Positive	285
LCAT3	Lung cancer	Activating invasion and metastasis; sustaining proliferative signaling	Positive	286
LINC00312	Lung cancer	Inducing angiogenesis	Positive	287
MALAT1	Breast cancer	Activating invasion and metastasis	Negative	288
LINK-A	Breast cancer	Avoiding immune destruction	Positive	289
NKILA	Breast cancer	Avoiding immune destruction	Positive	290
H19	Breast cancer	Resisting cell death	Positive	291
HOTAIR	Gastric cancer	Tumor-promoting inflammation	Positive	292
HOXC-AS3	Gastric cancer	Sustaining proliferative signaling; activating invasion and metastasis	Positive	293
GMAN	Gastric cancer	Activating invasion and metastasis	Positive	11
MALAT1	Gastric cancer	Resisting cell death	Negative	294
LncTCF7	Liver cancer	Unlocking phenotypic plasticity	Positive	295
lncRNA-ATB	Liver cancer	Activating invasion and metastasis	Positive	296
lnc-EGFR	Liver cancer	Avoiding immune destruction	Positive	297
LINC00662	Liver cancer	Genome instability and mutation	Positive	298
CCAT2	Colorectal cancer	Genome instability and mutation	Positive	299
FLANC	Colorectal cancer	Inducing angiogenesis; activating invasion and metastasis; sustaining proliferative signaling	Positive	300
NEAT1-1	Prostate cancer	Activating invasion and metastasis	Positive	301
SChLAP1	Prostate cancer	Activating invasion and metastasis	Positive	302
LNMAT2	Bladder cancer	Activating invasion and metastasis	Positive	303
SOX2OT	Bladder cancer	Unlocking phenotypic plasticity	Positive	304
HOXB-AS3	Leukemia	Sustaining proliferative signaling	Positive	305
IUR1	Leukemia	Resisting cell death	Negative	306
NORAD	Pancreatic cancer	Activating invasion and metastasis	Positive	307
XLOC_000647	Pancreatic cancer	Activating invasion and metastasis	Negative	308

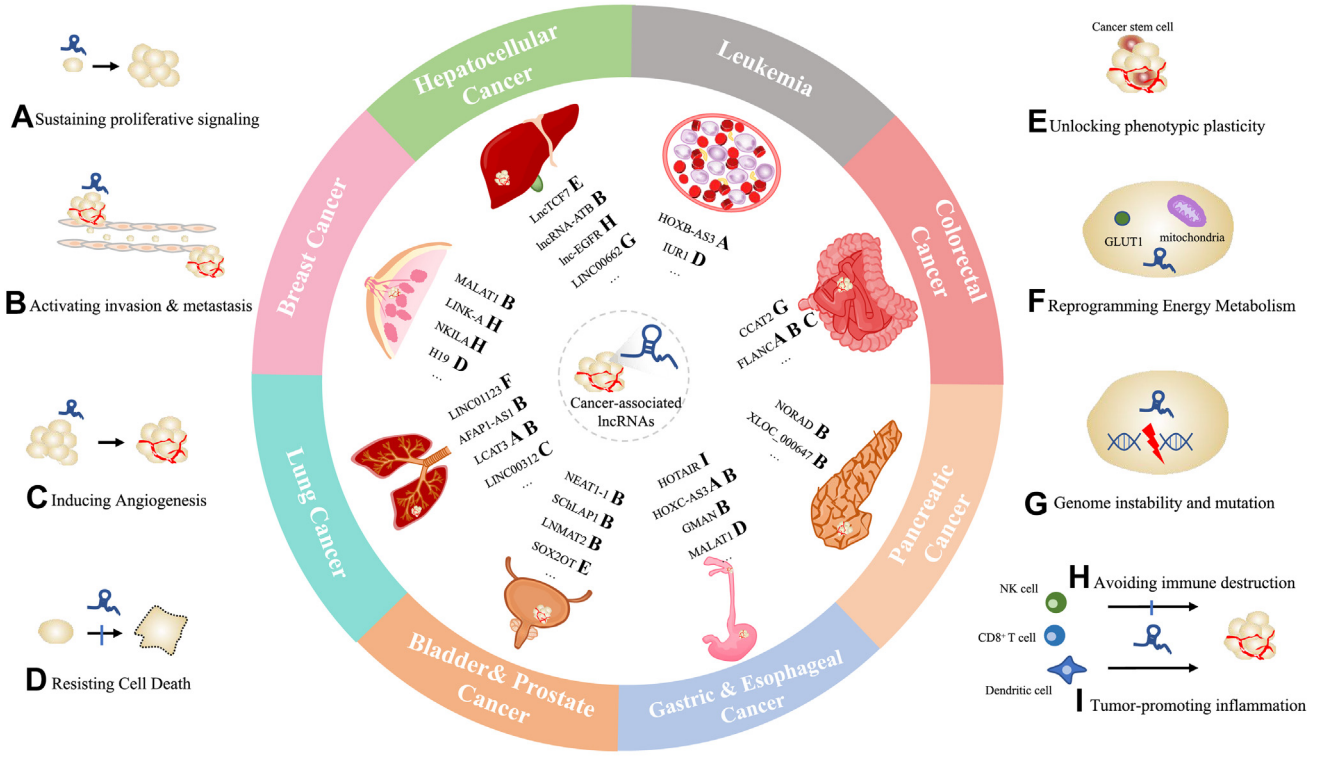

Figure 3 Summary of cancer-associated lncRNAs based on different cancer types and cancer hallmarks.

transcripts is another difficult problem after functional characterization. Systemic identification of lncRNA action targets will greatly promote a better understanding of lncRNA function and the effective design of follow-up experiments. Molecules that bind to lncRNAs and come into play can be identified according to three types: chromatin, RNA, and protein. RNA structure and interactions between lncRNAs and various cellular molecules are fundamental to lncRNA functions. Considering a large number of lncRNAs and their targets, it is a challenging and promising field to map lncRNA interactions on a large scale to obtain a comprehensive view of lncRNA mechanisms. Several methods have been developed to map lncRNA-chromatin, RNA, and protein interactions in a time- or genome-scale manner. Ligation, purification, and sequencing are the common key steps of these methods. Emerging computational algorithms take advantage of large-scale data sources from different experiments and datasets and play an increasingly important role. For example, the NPInter v4.0 database curated experimentally verified interactions between lncRNAs and genomic DNA, RNA, or proteins.[194] Available methods for the prediction of lncRNA interaction partners are summarized in Table 5.

RNA structure

The functional and regulatory plasticity of lncRNAs mostly due to their complex secondary and tertiary structure. Compared to the primary sequence, other factors like thermodynamic features, RNA chaperones, and interacting molecules have a stronger influence on lncRNA structure. During multiple biological processes, the presence of dynamic conformation rearrangement makes it complicated to profile lncRNA structure.[195,196] Biological functionality and interacting partners of lncRNAs can partly be inferred according to their specific structural domains. By utilizing the genome-wide and large-scale methods, lncRNAs with similar structural modes as well as the dynamic conformation change of one certain lncRNA can be well characterized, which will make a significant contribution to understanding the action modes of lncRNAs.

Previous high-throughput methods based on different ribonucleases (RNases) specific for single- or double-strand regions in RNA include PARS (Parallel Analysis of RNA Structure)[197] and FragSeq (fragmentation sequencing).[198] The accuracy of these methods relies on various ribonucleases with different specificities. PARS outputs comprehensive information about single- and double-stranded RNA, but FragSeq has a preference for RNAs that are shorter than 200 nt. SHAPE (Selective 2′-Hydroxyl Acylation analyzed by Primer Extension) and SHAPE-Seq (SHAPE combined with multiplexed paired-end deep sequencing)[199] are chemical methods based on nucleotide acylation, which can act as a termination signal in reverse transcription and be mapped by following sequencing of consequent cDNA fragments. Several chemical reagents can be used to produce RNA modification and transfer RNA information into cDNA by reverse transcription, providing clues for structural modeling.[199] The above high-throughput methods bring

Table 5 Approaches for identification of lncRNA interaction partners.

Name	Description	Reference
LncRNA-Chromatin		
RAP	Specifically selective purification of endogenous RNA complexes	209
CHART	Hybridization-based technique that specifically enriches endogenous RNAs along with their targets	210
ChIRP	Tiling oligonucleotides retrieve specific lncRNAs with bound protein and DNA sequences	211
RAT-seq	*In situ* reverse transcription with specific complementary primers; isolation of biotin-chromatin complex by streptavidin beads; genome-wide measurement of RNA interactions	309
GRID-seq	Link RNA to chromatin in situ by a bivalent linker on fixed nuclei; comprehensive identification	213
MARGI	Forms RNA-DNA chimeric sequences by proximity ligation containing RNA-ligation and DNA-ligation; comprehensive identification	214
TDF	Detect triplexes and characterize DNA-binding domains and DNA targets statistically	216
Triplex-Inspector	A computational aid for the sophisticated design of sequence-specific ligands and selection of appropriate targets	217
Triplexator	Integrates all aspects of triplex formation and assess the potential of triplex formation	218
R-loopDB	A database that collects all R-loop forming sequences	219
QmRLFS-finder	Predicts R-loop forming sequences in nucleic acid sequences	220
LncRNA-RNA		
RAP-RNA	Systematically map RNA—RNA interactions based on RAP	222
SPLASH	Maps pairwise RNA interactions *in vivo* genome-wide	223
CLASH	High-throughput method to identify intramolecular and intermolecular RNA—RNA interactions	224
PARIS	Employs psoralen-derivative 4′-aminomethyltrioxsalen (AMT) as crosslinker followed by purification and sequencing	225
MARIO	Depends on RNA-binding proteins and a biotinylated RNA linker to form a chimeric RNA and subjected to sequencing after isolation	226
LIGR-seq	Global-scale mapping of RNA—RNA duplexes crosslinked *in vivo*	227
starBase v2.0	Systematically identify the RNA—RNA and RNA-DNA interaction	228
LnCeCell	A database of lncRNA-related ceRNA	231
LncACTdb 3.0	A database of lncRNA-related ceRNA	191
LncRNA-Protein		
lncRNA-pulldown	Identify interacting proteins of specific lncRNA	237
CLIP-seq	Study protein-lncRNA interactions of RNA-binding protein (RBP) binding sites in transcriptome-wide	238
ChIRP-MS	Utilizing mass spectrometry combined with ChIRP	239
RAP-MS	Identification of direct and specific protein interaction partners of a specific RNA molecule	240
CLIPdb	A database to describe RBP-RNA interactions based on CLIP-Seq data	242
starBase v2.0	Systematically identify the RNA—RNA and RNA-DNA interaction	228
Global score	Computational algorithm for quantitatively prediction of overall binding ability and contact between lncRNA and protein	243
LPBNI	Computational algorithm to identify potential lncRNA-interacting proteins	244
HLPI-Ensemble	Computational algorithm for lncRNA-protein interaction based on ensemble strategy	245
catRAPID	Investigate Xist associations with a number of proteins	246

valuable data resources for subsequent computational RNA structural profiling, such as SeqFold[200] and CROSS (Computational Recognition of Secondary Structure).[201]

Subcellular localization

The spatial distribution of lncRNA determines where it plays a role, which may provide essential indicative information of their functional mechanism. Some existing databases collected information on the subcellular localization of lncRNAs from literature mining, RNA-seq data, and other databases, or combined them all. These databases include lncSLdb,[202] LncATLAS,[203] and RNALocate v2.0.[204] In addition to information-collecting databases, lncRNA subcellular localization is also predicted by computational pipelines and methods. Previous computational methods based on k-

mer features have a major drawback of sequence order information lost, so a new method named DeepLncLoc solved this problem by utilizing sequence embedding.[205] The complexity of lncRNA function caused by their multiple subcellular localization will be a promising and challenging trend for lncRNA study.

LncRNA-chromatin

Several lncRNAs have been proven to directly act with chromatin in a cis- or trans-manner, binding to chromatin for complex structure formation and recruiting histone- and DNA-modifying enzymes to epigenetically regulate target genes.[206,207] To identify the interaction between a known lncRNA and all possible chromatin targets, methods using complementary oligonucleotides containing biotin moieties followed by sequencing to determine the targets, such as RNA antisense purification (RAP),[208,209] capture hybridization analysis of RNA targets (CHART),[210] chromatin isolation by RNA purification (ChIRP)[211] and reverse transcription-associated trap sequencing (RAT-Seq),[212] are available. When a genome-wide view of all potential RNA-chromatin interaction and binding sites is sought, the goal can be achieved by Global RNA Interactions with DNA by deep sequencing (GRID-seq) [213] and mapping RNA-genome interactions (MARGI) [214]; a GENCODE-based pipeline for lncRNA-DNA interaction analysis was reported to identify cross-species and genome-wide potential genome and epigenome editing targeted sites.[215] Triple-helix (triplex) and R-loop (three-stranded RNA and DNA hybrid structures) forming potentials can also provide evidence for lncRNA-DNA interactions. Triplex domain finder (TDF), Triplex-Inspector, and Triplexator, another computational method, were developed to detect triplex formation and predict potential binding sites.[216–218] Prediction of the R-loop forming sequence (RLFS) by R-loopDB and quantitative model of RLFS finder (QmRLFS-finder) provides a novel perspective on lncRNA-DNA interactions and reveals a wide range of functions, such as transcriptional regulation and epigenetic modification.[219,220]

LncRNA-RNA

LncRNA-RNA interactions reveal another dimension of lncRNA function in living cells. Base pairing and the formation of joint secondary structures are needed in the interaction between RNA and RNA.[221] Previous methods to identify RNA–RNA interactions are limited to specific known objects and thus preclude a general view of lncRNA interactions. Global mapping approaches such as RAP-RNA,[222] sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH),[223] UV cross-linking, ligation, and sequencing of hybrids (CLASH),[224] psoralen analysis of RNA interactions and structures (PARIS),[225] mapping RNA interactome *in vivo* (MARIO),[226] and ligation of interacting RNA followed by high-throughput sequencing (LIGR-seq)[227] elucidate the global landscape of RNA–RNA interactions and enable discovery of alternative and long-range structures of lncRNAs. An integrated pipeline utilized the combination of Raccess, TanTan, LAST, IntaRNA, and RactIP programs to conduct a comprehensive prediction of lncRNA-RNA interactions in a whole human transcriptome, and the resulting database containing predicted interactions was developed.[221] StarBase v2.0 provides comprehensive lncRNA-miRNA interactions that were validated by experiments.[228] Even so, a comparison between computational tools revealed that large-scale prediction of RNA–RNA predictions is still challenging at this stage and suggested a normalization strategy to improve accuracy.[229] A subset of lncRNAs may compete with microRNAs to regulate mRNAs, playing a role as competing endogenous RNAs (ceRNAs).[230] LnCeCell[231] and LncACTdb 3.0 [191] are databases including lncRNA-related ceRNA in tumors and other diseases, indicating their subcellular location and functionality in tumorigenesis and cancer development.

LncRNA–protein

LncRNA-protein interactions enable lncRNAs to extend their scope of functionality. Epigenetic regulation can be mediated by the recruitment of proteins, such as lncRNA Gm15055, PARTICLE, and TUG1, which recruit the polycomb repressive complex, PRC2, to epigenetically silence the expression of target genes.[232–234] LncRNAs can also interact with proteins to conduct posttranscriptional regulation, such as stability regulation and RNA editing.[235,236] Biotinylated RNA pulldown assays can be used to identify the interaction with proteins of specific lncRNAs.[237] Methods such as cross-linking and immunoprecipitation followed by sequencing (CLIP-seq),[238] ChIRP-MS,[239] and RAP-MS,[240] and iDRiP[241] were developed to reveal specific noncoding RNA-protein complexes at the genome scale. These methods have already been summarized and comprised. The resulting databases StarBase v2.0 and CLIPdb contain several datasets of RBP sites and represent a valuable resource for posttranscriptional regulatory research.[228,242]

Computational algorithms such as Global Score, lncRNA-protein bipartite network inference (LPBNI), HLPI-Ensemble, and catRAPID are available. Global Score has the ability to predict overall binding ability and contact between lncRNAs and proteins quantitatively.[243] LPBNI makes use of well-characterized lncRNA-protein interactions to predict potential interacting proteins.[244] HLPI-Ensemble models utilize ensemble strategies on the basis of machine learning algorithms.[245] CatRAPID was developed to identify associations of long noncoding RNA X-inactive-specific transcripts (Xist) and proteins in various processes.[246]

Perspectives

Though the landscape of lncRNAs has been extended to massive aspects of biological processes, researchers never stop expanding the functional roles of this massive group. Based on large-scale research strategies of lncRNAs, growing interest has been devoted to the utilization of lncRNAs as powerful tools of genetic regulation. Genetic regulation by lncRNAs is reflected in epigenetic modification, transcriptional regulation, and posttranscriptional regulation. The identification of these regulatory lncRNAs globally is crucial for further understanding and the

transformation of regulatory lncRNAs to molecular tools is significant and promising. For example, RNA enhancement (RNAe) has been developed to regulate protein expression at the posttranscriptional and pretranslational levels without affecting mRNA levels. This technology is based on the modification of antisense RNA with the repeat element SINEB2 and is universally applicable.[247]

A class of ncRNA, which were transcribed from enhancers and called enhancer RNA (eRNA), have been proven to gain increased attention by their role in transcriptional regulation. Enhancers used to be considered as a regulatory DNA element to active transcription of a target gene by binding with transcription factors. Increasing evidence shows that enhancers can be transcribed into noncoding RNA through recruiting polymerase II across multiple cell types and species, and this transcriptional activity can promote the expression of nearby target genes.[248] The imbalance between the growing number and insufficient understanding of eRNA highlights that it is important to decode the systemic landscape of eRNA. Early detection of eRNA based on RNA-seq (alignment between transcripts and enhancer region) and CHIP-seq (histone modification of active enhancers).[249] Existing databases like the FANTOM project and ENCODE project provide a large number of enhancer annotations for potential eRNA prediction.[250] HeRA (Human enhancer RNA Atlas)[251] is a web resource containing eRNAs characterized in different human tissues as well as their interaction networks, such as regulators and targets.

Nascent RNA profiling is another high-throughput method for eRNA detection. Based on RNA polymerase detection, GRO (global run-on sequencing),[252] PRO-seq (precision nuclear run-on sequencing),[253] and ChRO-seq (chromatin run-on and sequencing)[254] can map nascent transcription regions by the location, density, and orientation of RNA polymerases. For the data processing of nascent RNA analysis, a computational pipeline called PEPPRO can be used.[255]

In conclusion, genome-wide approaches have become a crucial method for lncRNA research and enable a comprehensive view, including their expression, function, and interaction, of this massive group of transcripts. With the help of these approaches, lncRNAs with specific expressions and crucial functions can be annotated and subsequently serve as powerful tools in biological research and medical practice. Our review can provide information for a comprehensive research system of lncRNA genome-wide study.

Author contributions

S.T. wrote and edited this manuscript and created figures and tables. Y.R.H., L.T.D., Y.X.H., and W.Y.X. revised and perfected the manuscript. S.J.X. and Z.D.X. conceived the idea and guided the preparation of the manuscript. All authors read and approved the final version of the manuscript.

Conflict of interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No. 82273158, 81974436, 81900559, and 31701116), the Natural Science Foundation of Guangdong Province, China (No. 2022A1515010704, 2022A1515011413, and 2021A1515010477), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (China) (No. 2019ZT08Y485), Guangzhou Science and Technology Plan Projects (China) (No. 202102020104), and the Guangdong Province Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation (China) (No. 2020B1212060018).

References

1. Maher B. ENCODE: the human encyclopaedia. *Nature*. 2012; 489(7414):46—48.
2. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012;81:145—166.
3. Boon RA, Jaé N, Holdt L, et al. Long noncoding RNAs: from clinical genetics to therapeutic targets? *J Am Coll Cardiol*. 2016;67(10):1214—1226.
4. Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? evidence for selection within long noncoding RNAs. *Genome Res*. 2007;17(5):556—565.
5. Kapranov P, St Laurent G, Raz T, et al. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biol*. 2010;8:149.
6. Clark MB, Choudhary A, Smith MA, et al. The dark matter rises: the expanding world of regulatory RNAs. *Essays Biochem*. 2013;54:1—16.
7. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell*. 2013;152(6): 1298—1307.
8. Stojic L, Niemczyk M, Orjalo A, et al. Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions. *Nat Commun*. 2016;7: 10406.
9. Martianov I, Ramadass A, Serra Barros A, et al. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*. 2007;445(7128):666—670.
10. Tripathi V, Ellis JD, Shen Z, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010; 39(6):925—938.
11. Zhuo W, Liu Y, Li S, et al. Long noncoding RNA *GMAN*, upregulated in gastric cancer tissues, is associated with metastasis in patients and promotes translation of ephrin A1 by competitively binding GMAN-AS. *Gastroenterology*. 2019; 156(3):676—691.
12. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154(1):26—46.
13. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1(5):391—407.
14. Melé M, Mattioli K, Mallard W, et al. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res*. 2017;27(1):27—37.
15. Jiang S, Cheng SJ, Ren LC, et al. An expanded landscape of human long noncoding RNA. *Nucleic Acids Res*. 2019;47(15): 7842—7856.
16. Schlackow M, Nojima T, Gomes T, et al. Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol Cell*. 2017;65(1):25—38.
17. Mattick JS. A new paradigm for developmental biology. *J Exp Biol*. 2007;210(Pt 9):1526—1547.

18. St Laurent III G, Wahlestedt C. Noncoding RNAs: couplers of analog and digital information in nervous system function? *Trends Neurosci*. 2007;30(12):612—621.
19. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562—578.
20. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
21. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357—360.
22. Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650—1667.
23. Pertea M, Pertea GM, Antonescu CM, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290—295.
24. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644—652.
25. Haas BJ, Papanicolaou A, Yassour M, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494—1512.
26. You BH, Yoon SH, Nam JW. High-confidence coding and noncoding transcriptome maps. *Genome Res*. 2017;27(6):1050—1062.
27. Leman R, Gaildrat P, Le Gac G, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Res*. 2018;46(15):7913—7923.
28. Moles-Fernández A, Duran-Lozano L, Montalban G, et al. Computational tools for splicing defect prediction in breast/ovarian cancer genes: how efficient are they at predicting RNA alterations? *Front Genet*. 2018;9:366.
29. Tiulpakov A, Zubkova N, Makretskaya N, et al. Minigene splicing assessment of 20 novel synonymous and intronic glucokinase gene variants identified in patients with maturity-onset diabetes of the young. *Hum Mutat*. 2020;41(1):129—132.
30. Monger S, Troup M, Ip E, et al. Spliceogen: an integrative, scalable tool for the discovery of splice-altering variants. *Bioinformatics*. 2019;35(21):4405—4407.
31. Rodriguez JM, Maietta P, Ezkurdia I, et al. APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res*. 2013;41(Database issue):D110—D117.
32. Garber M, Grabherr MG, Guttman M, et al. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods*. 2011;8(6):469—477.
33. Guttman M, Garber M, Levin JZ, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*. 2010;28(5):503—510.
34. Kawaji H, Lizio M, Itoh M, et al. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. *Genome Res*. 2014;24(4):708—717.
35. Jan CH, Friedman RC, Ruby JG, et al. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature*. 2011;469(7328):97—101.
36. Kim M, You BH, Nam JW. Global estimation of the 3' untranslated region landscape using RNA sequencing. *Methods*. 2015;83:111—117.
37. Huang Z, Teeling EC. ExUTR: a novel pipeline for large-scale prediction of 3'-UTR sequences from NGS data. *BMC Genom*. 2017;18(1):847.
38. Kang YN, Lai DP, Ooi HS, et al. Genome-wide profiling of untranslated regions by paired-end ditag sequencing reveals unexpected transcriptome complexity in yeast. *Mol Genet Genom*. 2015;290(1):217—224.
39. Kong L, Zhang Y, Ye ZQ, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(Web Server issue):W345—W349.
40. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;27(13):i275—i282.
41. Wang L, Park HJ, Dasari S, et al. CPAT: coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41(6):e74.
42. Sun L, Luo H, Bu D, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41(17):e166.
43. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinf*. 2014;15(1):311.
44. Zhao J, Song X, Wang K. lncScore: alignment-free identification of long noncoding RNA from assembled novel transcripts. *Sci Rep*. 2016;6:34838.
45. Guo JC, Fang SS, Wu Y, et al. CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition. *Nucleic Acids Res*. 2019;47(W1):W516—W522.
46. Tong X, Liu S. CPPred: coding potential prediction based on the global description of RNA sequence. *Nucleic Acids Res*. 2019;47(8):e43.
47. Zhang Y, Jia C, Fullwood MJ, et al. DeepCPP: a deep neural network based on nucleotide bias information and minimum distribution similarity feature selection for RNA coding potential prediction. *Briefings Bioinf*. 2021;22(2):2073—2084.
48. Kondo T, Plaza S, Zanet J, et al. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila embryogenesis*. *Science*. 2010;329(5989):336—339.
49. Nelson BR, Makarewich CA, Anderson DM, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*. 2016;351(6270):271—275.
50. Rion N, Rüegg MA. LncRNA-encoded peptides: more than translational noise? *Cell Res*. 2017;27(5):604—605.
51. Wu P, Mo Y, Peng M, et al. Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA. *Mol Cancer*. 2020;19(1):22.
52. Chen Y, Ho L, Tergaonkar V. sORF-Encoded MicroPeptides: new players in inflammation, metabolism, and precision medicine. *Cancer Lett*. 2021;500:263—270.
53. Flower CT, Chen L, Jung HJ, et al. An integrative proteogenomics approach reveals peptides encoded by annotated lincRNA in the mouse kidney inner medulla. *Physiol Genom*. 2020;52(10):485—491.
54. Wang B, Hao J, Pan N, et al. Identification and analysis of small proteins and short open reading frame encoded peptides in Hep3B cell. *J Proteonomics*. 2021;230:103965.
55. Bonilauri B, Holetz FB, Dallagiovanna B. Long non-coding RNAs associated with ribosomes in human adipose-derived stem cells: from RNAs to microproteins. *Biomolecules*. 2021;11(11):1673.
56. Hao Y, Zhang L, Niu Y, et al. SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Briefings Bioinf*. 2018;19(4):636—643.
57. Dragomir MP, Manyam GC, Ott LF, et al. FuncPEP: a database of functional peptides encoded by non-coding RNAs. *Non-coding RNA*. 2020;6(4):41.

58. Liu H, Zhou X, Yuan M, et al. ncEP: a manually curated database for experimentally validated ncRNA-encoded proteins or peptides. *J Mol Biol*. 2020;432(11):3364—3368.
59. Huang Y, Wang J, Zhao Y, et al. cncRNAdb: a manually curated resource of experimentally supported RNAs with both protein-coding and noncoding function. *Nucleic Acids Res*. 2021; 49(D1):D65—D70.
60. Liu T, Wu J, Wu Y, et al. LncPep: a resource of translational evidences for lncRNAs. *Front Cell Dev Biol*. 2022;10:795084.
61. Lv D, Chang Z, Cai Y, et al. TransLnc: a comprehensive resource for translatable lncRNAs extends immunopeptidome. *Nucleic Acids Res*. 2022;50(D1):D413—D420.
62. Luo X, Huang Y, Li H, et al. SPENCER: a comprehensive database for small peptides encoded by noncoding RNAs in cancer patients. *Nucleic Acids Res*. 2022;50(D1): D1373—D1381.
63. Olexiouk V, Menschaert G. Using the sORFs.Org database. *Curr Protoc Bioinformatics*. 2019;65(1):e68.
64. Brunet MA, Lucier JF, Levesque M, et al. OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Res*. 2021;49(D1): D380—D388.
65. Choteau SA, Wagner A, Pierre P, et al. MetamORF: A repository of unique short open reading frames identified by both experimental and computational approaches for gene and metagene analyses. *Database (Oxford)*. 2021;2021: baab032.
66. Leong AZX, Lee PY, Mohtar MA, et al. Short open reading frames (sORFs) and microproteins: an update on their identification and validation measures. *J Biomed Sci*. 2022; 29(1):19.
67. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015;47(3):199—208.
68. Howald C, Tanzer A, Chrast J, et al. Combining RT-PCR-seq and RNA-seq to catalog all genic elements encoded in the human genome. *Genome Res*. 2012;22(9):1698—1710.
69. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923—930.
70. Anders S, Pyl PT, Huber W. HTSeq: a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166—169.
71. Kawai J, Shinagawa A, Shibata K, et al. Functional annotation of a full-length mouse cDNA collection. *Nature*. 2001; 409(6821):685—690.
72. Seki M, Narusaka M, Kamiya A, et al. Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science*. 2002; 296(5565):141—145.
73. Kapranov P, Drenkow J, Cheng J, et al. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res*. 2005; 15(7):987—997.
74. Lagarde J, Uszczynska-Ratajczak B, Santoyo-Lopez J, et al. Extension of human lncRNA transcripts by RACE coupled with long-read high-throughput sequencing (RACE-Seq). *Nat Commun*. 2016;7:12339.
75. Mercer TR, Gerhardt DJ, Dinger ME, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat Biotechnol*. 2011;30(1):99—104.
76. Zhu L, Zhong J, Jia X, et al. Precision methylome characterization of *Mycobacterium* tuberculosis complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. *Nucleic Acids Res*. 2016;44(2):730—743.
77. Giesselmann P, Brändl B, Raimondeau E, et al. Analysis of short tandem repeat expansions and their methylation state with nanopore sequencing. *Nat Biotechnol*. 2019;37(12): 1478—1481.
78. Sharon D, Tilgner H, Grubert F, et al. A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol*. 2013;31(11):1009—1014.
79. Lagarde J, Uszczynska-Ratajczak B, Carbonell S, et al. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet*. 2017;49(12): 1731—1740.
80. Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol*. 2019; 37(10):1155—1162.
81. Ding J, Adiconis X, Simmons SK, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol*. 2020;38(6):737—746.
82. Hong M, Tao S, Zhang L, et al. RNA sequencing: new technologies and applications in cancer research. *J Hematol Oncol*. 2020;13(1):166.
83. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: archive for functional genomics data sets: 10 years on. *Nucleic Acids Res*. 2011;39:D1005—D1010. Database issue.
84. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603—607.
85. Cummins C, Ahamed A, Aslam R, et al. The European nucleotide archive in 2021. *Nucleic Acids Res*. 2022;50(D1): D106—D110.
86. Frenkel-Morgenstern M, Gorohovski A, Lacroix V, et al. ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data. *Nucleic Acids Res*. 2013; 41:D142—D151. Database issue.
87. Yu Y, Zhang H, Long Y, et al. Plant Public RNA-seq Database: a comprehensive online database for expression analysis of ~45 000 plant public RNA-Seq libraries. *Plant Biotechnol J*. 2022;20(5):806—808.
88. Frankish A, Diekhans M, Jungreis I, et al. Gencode 2021. *Nucleic Acids Res*. 2021;49(D1):D916—D923.
89. Quek XC, Thomson DW, Maag JLV, et al. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res*. 2015;43:D168—D173. Database issue.
90. Zhao L, Wang J, Li Y, et al. NONCODEV6:an updated database dedicated to long non-coding RNA annotation in both animals and plants. *Nucleic Acids Res*. 2021;49(D1):D165—D171.
91. Pertea M, Shumate A, Pertea G, et al. CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol*. 2018;19(1):208.
92. Abugessaisa I, Ramilowski JA, Lizio M, et al. FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. *Nucleic Acids Res*. 2021; 49(D1):D892—D898.
93. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25(18): 1915—1927.
94. Liu J, Jung C, Xu J, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell*. 2012;24(11):4333—4345.
95. Wang H, Chung PJ, Liu J, et al. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res*. 2014;24(3):444—453.
96. Luo H, Bu D, Sun L, et al. Identification and function annotation of long intervening noncoding RNAs. *Briefings Bioinf*. 2017;18(5):789—797.
97. Zhou B, Ji B, Liu K, et al. EVLncRNAs 2.0:an updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res*. 2021;49(D1):D86—D91.

98. Li Z, Liu L, Jiang S, et al. LncExpDB: an expression database of human long non-coding RNAs. *Nucleic Acids Res*. 2021;49(D1):D962—D968.
99. Volders PJ, Anckaert J, Verheggen K, et al. LNCipedia 5:towards a reference set of human long non-coding RNAs. *Nucleic Acids Res*. 2019;47(D1):D135—D139.
100. Liu L, Li Z, Liu C, et al. LncRNAWiki 2.0:a knowledgebase of human long non-coding RNAs with enhanced curation model and database system. *Nucleic Acids Res*. 2022;50(D1):D190—D195.
101. Ma L, Cao J, Liu L, et al. LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res*. 2019;47(D1):D128—D134.
102. Hou M, Tang X, Tian F, et al. AnnoLnc: a web server for systematically annotating novel human lncRNAs. *BMC Genom*. 2016;17(1):931.
103. Ke L, Yang DC, Wang Y, et al. AnnoLnc2:the one-stop portal to systematically annotate novel lncRNAs for human and mouse. *Nucleic Acids Res*. 2020;48(W1):W230—W238.
104. Park C, Yu N, Choi I, et al. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics*. 2014;30(17):2480—2485.
105. Zheng LL, Li JH, Wu J, et al. deepBase v2.0:identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res*. 2016;44(D1):D196—D202.
106. Abugessaisa I, Shimoji H, Sahin S, et al. FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki. *Database (Oxford)*. 2016;2016:baw105.
107. Zhang Y, Bu D, Huo P, et al. ncFANs v2.0:an integrative platform for functional annotation of non-coding RNAs. *Nucleic Acids Res*. 2021;49(W1):W459—W468.
108. Chen J, Zhang J, Gao Y, et al. LncSEA: a platform for long non-coding RNA related sets and enrichment analysis. *Nucleic Acids Res*. 2021;49(D1):D969—D980.
109. Bolstad BM, Irizarry RA, Åstrand M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185—193.
110. Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods*. 2008;5(7):621—628.
111. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
112. Lin Y, Golovnina K, Chen ZX, et al. Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. *BMC Genom*. 2016;17:28.
113. Kadota K, Ye J, Nakai Y, et al. ROKU: a novel method for identification of tissue-specific genes. *BMC Bioinf*. 2006;7:294.
114. Cavalli FM, Bourgon R, Huber W, et al. SpeCond: a method to detect condition-specific gene expression. *Genome Biol*. 2011;12(10):R101.
115. Forouzmand E, Owens NDL, Blitz IL, et al. Developmentally regulated long non-coding RNAs in Xenopus tropicalis. *Dev Biol*. 2017;426(2):401—408.
116. Tang Z, Wu Y, Yang Y, et al. Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutional conservation in *Sus scrofa*. *Sci Rep*. 2017;7:43166.
117. Bester AC, Lee JD, Chavez A, et al. An integrated genome-wide CRISPRa approach to functionalize lncRNAs in drug resistance. *Cell*. 2018;173(3):649—664.e20.
118. Stuart JM, Segal E, Koller D, et al. A gene-coexpression network for global discovery of conserved genetic modules. *Science*. 2003;302(5643):249—255.
119. Weber M, Teeling H, Huang S, et al. Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics. *ISME J*. 2011;5(5):918—928.
120. Dai L, Li J, Dong Z, et al. Temporal expression and functional analysis of long non-coding RNAs in colorectal cancer initiation. *J Cell Mol Med*. 2019;23(6):4127—4138.
121. Yu CK, Xu T, Assoian RK, et al. Mining the stiffness-sensitive transcriptome in human vascular smooth muscle cells identifies long noncoding RNA stiffness regulators. *Arterioscler Thromb Vasc Biol*. 2018;38(1):164—173.
122. Liu H, Sun Y, Tian H, et al. Characterization of long non-coding RNA and messenger RNA profiles in laryngeal cancer by weighted gene co-expression network analysis. *Aging*. 2019;11(22):10074—10099.
123. Xiong HG, Li H, Xiao Y, et al. Long noncoding RNA MYOSLID promotes invasion and metastasis by modulating the partial epithelial-mesenchymal transition program in head and neck squamous cell carcinoma. *J Exp Clin Cancer Res*. 2019;38(1):278.
124. Liu P, Xie X, Yang A, et al. Melatonin regulates breast cancer progression by the lnc010561/miR-30/$FKBP_3$ axis. *Mol Ther Nucleic Acids*. 2020;19:765—774.
125. Yang WS, Lin TY, Chang L, et al. HIV-1 tat interacts with a kaposi's sarcoma-associated herpesvirus reactivation-upregulated antiangiogenic long noncoding RNA, LINC00313, and antagonizes its function. *J Virol*. 2020;94(3):e01280-19.
126. Berns K, Hijmans EM, Mullenders J, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*. 2004;428(6981):431—437.
127. Boutros M, Kiger AA, Armknecht S, et al. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science*. 2004;303(5659):832—835.
128. Seiler J, Breinig M, Caudron-Herger M, et al. The lncRNA VELUCT strongly regulates viability of lung cancer cells despite its extremely low abundance. *Nucleic Acids Res*. 2017;45(9):5458—5469.
129. Jackson AL, Burchard J, Schelter J, et al. Widespread siRNA off-target transcript silencing mediated by seed region sequence complementarity. *RNA*. 2006;12(7):1179—1187.
130. Echeverri CJ, Beachy PA, Baum B, et al. Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat Methods*. 2006;3(10):777—779.
131. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819—823.
132. Wang T, Wei JJ, Sabatini DM, et al. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*. 2014;343(6166):80—84.
133. Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014;343(6166):84—87.
134. Konermann S, Brigham MD, Trevino AE, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015;517(7536):583—588.
135. Zhou Y, Zhu S, Cai C, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*. 2014;509(7501):487—491.
136. Chen S, Sanjana NE, Zheng K, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*. 2015;160(6):1246—1260.
137. Koike-Yusa H, Li Y, Tan EP, et al. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol*. 2014;32(3):267—273.
138. Chen W, Zhang G, Li J, et al. CRISPRlnc: a manually curated database of validated sgRNAs for lncRNAs. *Nucleic Acids Res*. 2019;47(D1):D63—D68.
139. Li W, Xu H, Xiao T, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*. 2014;15(12):554.
140. Li W, Köster J, Xu H, et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol*. 2015;16:281.

141. Wang B, Wang M, Zhang W, et al. Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc*. 2019;14(3):756—780.
142. Jeong HH, Kim SY, Rousseaux MWC, et al. Beta-binomial modeling of CRISPR pooled screen data identifies target genes with greater sensitivity and fewer false negatives. *Genome Res*. 2019;29(6):999—1008.
143. Diaz AA, Qin H, Ramalho-Santos M, et al. HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. *Nucleic Acids Res*. 2015;43(3):e16.
144. Daley TP, Lin Z, Lin X, et al. CRISPhieRmix: a hierarchical mixture model for CRISPR pooled screens. *Genome Biol*. 2018;19(1):159.
145. Meyers RM, Bryan JG, McFarland JM, et al. Computational correction of copy number effect improves specificity of CRISPR—Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49(12):1779—1784.
146. Allen F, Behan F, Khodak A, et al. JACKS: joint analysis of CRISPR/Cas9 knockout screens. *Genome Res*. 2019;29(3):464—471.
147. Bodapati S, Daley TP, Lin X, et al. A benchmark of algorithms for the analysis of pooled CRISPR screens. *Genome Biol*. 2020;21(1):62.
148. Zhu S, Li W, Liu J, et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat Biotechnol*. 2016;34(12):1279—1286.
149. Joung J, Engreitz JM, Konermann S, et al. Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature*. 2017;548(7667):343—346.
150. Covarrubias S, Robinson EK, Shapleigh B, et al. CRISPR/Cas-based screening of long non-coding RNAs (lncRNAs) in macrophages with an NF-κB reporter. *J Biol Chem*. 2017;292(51):20911—20920.
151. Cai P, Otten ABC, Cheng B, et al. A genome-wide long noncoding RNA CRISPRi screen identifies *PRANCR* as a novel regulator of epidermal homeostasis. *Genome Res*. 2020;30(1):22—34.
152. Liu SJ, Horlbeck MA, Cho SW, et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*. 2017;355(6320):aah7111.
153. Esposito R, Bosch N, Lanzós A, et al. Hacking the cancer genome: profiling therapeutically actionable long non-coding RNAs using CRISPR-Cas9 screening. *Cancer Cell*. 2019;35(4):545—557.
154. Goyal A, Myacheva K, Groß M, et al. Challenges of CRISPR/-Cas9 applications for long non-coding RNA genes. *Nucleic Acids Res*. 2017;45(3):e12.
155. Liu Y, Cao Z, Wang Y, et al. Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. *Nat Biotechnol*. 2018;36:1203—1210.
156. Wang C, Lu T, Emanuel G, et al. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc Natl Acad Sci USA*. 2019;116(22):10842—10851.
157. Dixit A, Parnas O, Li B, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*. 2016;167(7):1853—1866.
158. Adamson B, Norman TM, Jost M, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*. 2016;167(7):1867—1882.
159. Jaitin DA, Weiner A, Yofe I, et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell*. 2016;167(7):1883—1896.
160. Datlinger P, Rendeiro AF, Schmidl C, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*. 2017;14(3):297—301.
161. Lawson MJ, Camsund D, Larsson J, et al. *In situ* genotyping of a pooled strain library after characterizing complex phenotypes. *Mol Syst Biol*. 2017;13(10):947.
162. Emanuel G, Moffitt JR, Zhuang X. High-throughput, image-based screening of pooled genetic-variant libraries. *Nat Methods*. 2017;14(12):1159—1162.
163. Wroblewska A, Dhainaut M, Ben-Zvi B, et al. Protein barcodes enable high-dimensional single-cell CRISPR screens. *Cell*. 2018;175(4):1141—1155.
164. Zhu S, Cao Z, Liu Z, et al. Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Genome Biol*. 2019;20(1):20.
165. Michlits G, Hubmann M, Wu SH, et al. CRISPR-UMI: single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat Methods*. 2017;14(12):1191—1197.
166. Schmierer B, Botla SK, Zhang J, et al. CRISPR/Cas9 screening using unique molecular identifiers. *Mol Syst Biol*. 2017;13(10):945.
167. Smits AH, Ziebell F, Joberty G, et al. Biological plasticity rescues target activity in CRISPR knock outs. *Nat Methods*. 2019;16(11):1087—1093.
168. El-Brolosy MA, Kontarakis Z, Rossi A, et al. Genetic compensation triggered by mutant mRNA degradation. *Nature*. 2019;568(7751):193—197.
169. Abudayyeh OO, Gootenberg JS, Essletzbichler P, et al. RNA targeting with CRISPR-Cas13. *Nature*. 2017;550(7675):280—284.
170. Zhang C, Konermann S, Brideau NJ, et al. Structural basis for the RNA-guided ribonuclease activity of CRISPR-Cas13d. *Cell*. 2018;175(1):212—223.
171. Wessels HH, Méndez-Mancilla A, Guo X, et al. Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat Biotechnol*. 2020;38(6):722—727.
172. Lauretti E, Dabrowski K, Praticò D. The neurobiology of non-coding RNAs and Alzheimer's disease pathogenesis: pathways, mechanisms and translational opportunities. *Ageing Res Rev*. 2021;71:101425.
173. Hartana CA, Rassadkina Y, Gao C, et al. Long noncoding RNA MIR4435-2HG enhances metabolic function of myeloid dendritic cells from HIV-1 elite controllers. *J Clin Invest*. 2021;131(9):146136.
174. Xia L, Wang X, Liu L, et al. *lnc-BAZ2B* promotes M2 macrophage activation and inflammation in children with asthma through stabilizing *BAZ2B* pre-mRNA. *J Allergy Clin Immunol*. 2021;147(3):921—932.
175. Liu N, Kataoka M, Wang Y, et al. LncRNA LncHrt preserves cardiac metabolic homeostasis and heart function by modulating the LKB1-AMPK signaling pathway. *Basic Res Cardiol*. 2021;116(1):48.
176. Yu P, Guo J, Li J, et al. lncRNA-H19 in fibroblasts promotes wound healing in diabetes. *Diabetes*. 2022;71(7):1562—1578.
177. Liu R, Li X, Zhu W, et al. Cholangiocyte-derived exosomal long noncoding RNA H19 promotes hepatic stellate cell activation and cholestatic liver fibrosis. *Hepatology*. 2019;70(4):1317—1335.
178. Li J, Han L, Roebuck P, et al. TANRIC: an interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res*. 2015;75(18):3728—3737.
179. Ning S, Zhang J, Wang P, et al. Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res*. 2016;44(D1):D980—D985.
180. Bao Z, Yang Z, Huang Z, et al. LncRNADisease 2.0:an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res*. 2019;47(D1):D1034—D1037.
181. Gao Y, Li X, Shang S, et al. LincSNP 3.0:an updated database for linking functional variants to human long non-coding RNAs, circular RNAs and their regulatory elements. *Nucleic Acids Res*. 2021;49(D1):D1244—D1250.
182. Ning L, Cui T, Zheng B, et al. MNDR v3.0:mammal ncRNA-disease repository with increased coverage and annotation. *Nucleic Acids Res*. 2021;49(D1):D160—D164.

183. Liu SJ, Dang HX, Lim DA, et al. Long noncoding RNAs in cancer metastasis. *Nat Rev Cancer*. 2021;21(7):446—460.
184. Winkle M, El-Daly SM, Fabbri M, et al. Noncoding RNA therapeutics - challenges and potential solutions. *Nat Rev Drug Discov*. 2021;20(8):629—651.
185. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell*. 2016;29(4):452—463.
186. Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov*. 2022;12(1):31—46.
187. Lanzós A, Carlevaro-Fita J, Mularoni L, et al. Discovery of cancer driver long noncoding RNAs across 1112 tumour genomes: new candidates and distinguishing features. *Sci Rep*. 2017;7:41544.
188. Dai E, Yang F, Wang J, et al. ncDR: a comprehensive resource of non-coding RNAs involved in drug resistance. *Bioinformatics*. 2017;33(24):4010—4011.
189. Dong Q, Li F, Xu Y, et al. RNAactDrug: a comprehensive database of RNAs associated with drug sensitivity from multi-omics data. *Briefings Bioinf*. 2020;21(6):2167—2174.
190. Jiang W, Qu Y, Yang Q, et al. D-lnc: a comprehensive database and analytical platform to dissect the modification of drugs on lncRNA expression. *RNA Biol*. 2019;16(11):1586—1591.
191. Wang P, Guo Q, Qi Y, et al. LncACTdb 3.0:an updated database of experimentally supported ceRNA interactions and personalized networks contributing to precision medicine. *Nucleic Acids Res*. 2022;50(D1):D183—D189.
192. Gao Y, Shang S, Guo S, et al. Lnc2Cancer 3.0:an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. *Nucleic Acids Res*. 2021;49(D1):D1251—D1258.
193. Yang M, Lu H, Liu J, et al. lncRNAfunc: a knowledgebase of lncRNA function in human cancer. *Nucleic Acids Res*. 2022; 50(D1):D1295—D1306.
194. Teng X, Chen X, Xue H, et al. NPInter v4.0:an integrated database of ncRNA interactions. *Nucleic Acids Res*. 2020; 48(D1):D160—D165.
195. Graf J, Kretz M. From structure to function: route to understanding lncRNA mechanism. *Bioessays*. 2020;42(12): e2000027.
196. Kotar A, Foley HN, Baughman KM, et al. Advanced approaches for elucidating structures of large RNAs using NMR spectroscopy and complementary methods. *Methods*. 2020;183: 93—107.
197. Solomon O, Di Segni A, Cesarkas K, et al. RNA editing by ADAR1 leads to context-dependent transcriptome-wide changes in RNA secondary structure. *Nat Commun*. 2017;8(1):1440.
198. Underwood JG, Uzilov AV, Katzman S, et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods*. 2010;7(12):995—1001.
199. Lucks JB, Mortimer SA, Trapnell C, et al. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc Natl Acad Sci U S A*. 2011;108(27):11063—11068.
200. Ouyang Z, Snyder MP, Chang HY. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res*. 2013;23(2): 377—387.
201. Delli Ponti R, Marti S, Armaos A, et al. A high-throughput approach to profile RNA structure. *Nucleic Acids Res*. 2017; 45(5):e35.
202. Wen X, Gao L, Guo X, et al. lncSLdb: a resource for long non-coding RNA subcellular localization. *Database*. 2018;2018: 1—6.
203. Mas-Ponte D, Carlevaro-Fita J, Palumbo E, et al. LncATLAS database for subcellular localization of long noncoding RNAs. *RNA*. 2017;23(7):1080—1087.
204. Cui T, Dou Y, Tan P, et al. RNALocate v2.0:an updated resource for RNA subcellular localization with increased coverage and annotation. *Nucleic Acids Res*. 2022;50(D1): D333—D339.
205. Zeng M, Wu Y, Lu C, et al. DeepLncLoc: a deep learning framework for long non-coding RNA subcellular localization prediction based on subsequence embedding. *Briefings Bioinf*. 2022;23(1):bbab360.
206. Fu XD. Non-coding RNA: a new frontier in regulatory biology. *Natl Sci Rev*. 2014;1(2):190—204.
207. West JA, Davis CP, Sunwoo H, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell*. 2014;55(5):791—802.
208. Engreitz JM, Pandya-Jones A, McDonel P, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 2013;341(6147): 1237973.
209. Engreitz J, Lander ES, Guttman M. RNA antisense purification (RAP) for mapping RNA interactions with chromatin. *Methods Mol Biol*. 2015;1262:183—197.
210. Simon MD, Wang CI, Kharchenko PV, et al. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci USA*. 2011;108(51):20497—20502.
211. Chu C, Qu K, Zhong FL, et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 2011;44(4):667—678.
212. Sun J, Li W, Sun Y, et al. A novel antisense long noncoding RNA within the *IGF1R* gene locus is imprinted in hematopoietic malignancies. *Nucleic Acids Res*. 2014;42(15):9588—9601.
213. Li X, Zhou B, Chen L, et al. GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol*. 2017;35(10):940—950.
214. Sridhar B, Rivas-Astroza M, Nguyen TC, et al. Systematic mapping of RNA-chromatin interactions *in vivo*. *Curr Biol*. 2017;27(4):602—609.
215. Lin J, Wen Y, He S, et al. Pipelines for cross-species and genome-wide prediction of long noncoding RNA binding. *Nat Protoc*. 2019;14(3):795—818.
216. Kuo CC, Hänzelmann S, Sentürk Cetin N, et al. Detection of RNA-DNA binding sites in long noncoding RNAs. *Nucleic Acids Res*. 2019;47(6):e32.
217. Buske FA, Bauer DC, Mattick JS, et al. Triplex-Inspector: an analysis tool for triplex-mediated targeting of genomic loci. *Bioinformatics*. 2013;29(15):1895—1897.
218. Buske FA, Bauer DC, Mattick JS, et al. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res*. 2012;22(7):1372—1381.
219. Wongsurawat T, Jenjaroenpun P, Kwoh CK, et al. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res*. 2012; 40(2):e16.
220. Jenjaroenpun P, Wongsurawat T, Yenamandra SP, et al. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res*. 2015;43(W1):W527—W534.
221. Terai G, Iwakiri J, Kameda T, et al. Comprehensive prediction of lncRNA-RNA interactions in human transcriptome. *BMC Genom*. 2016;17(Suppl 1):12.
222. Engreitz JM, Sirokman K, McDonel P, et al. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell*. 2014;159(1): 188—199.
223. Aw JGA, Shen Y, Wilm A, et al. *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol Cell*. 2016;62(4):603—617.
224. Kudla G, Granneman S, Hahn D, et al. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci U S A*. 2011;108(24):10010—10015.
225. Lu Z, Zhang QC, Lee B, et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell*. 2016; 165(5):1267—1279.

226. Nguyen TC, Cao X, Yu P, et al. Mapping RNA-RNA interactome and RNA structure *in vivo* by MARIO. *Nat Commun*. 2016;7:12023.
227. Sharma E, Sterne-Weiler T, O'Hanlon D, et al. Global mapping of human RNA-RNA interactions. *Mol Cell*. 2016;62(4):618—626.
228. Li JH, Liu S, Zhou H, et al. starBase v2.0:decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92—D97.
229. Antonov IV, Mazurov E, Borodovsky M, et al. Prediction of lncRNAs and their interactions with nucleic acids: benchmarking bioinformatics tools. *Briefings Bioinf*. 2019;20(2):551—564.
230. Li B, Huang N, Wei S, et al. lncRNA TUG1 as a ceRNA promotes PM exposure-induced airway hyper-reactivity. *J Hazard Mater*. 2021;416:125878.
231. Wang P, Guo Q, Hao Y, et al. LnCeCell: a comprehensive database of predicted lncRNA-associated ceRNA networks at single-cell resolution. *Nucleic Acids Res*. 2021;49(D1):D125—D133.
232. Liu GY, Zhao GN, Chen XF, et al. The long noncoding RNA Gm15055 represses Hoxa gene expression by recruiting PRC2 to the gene cluster. *Nucleic Acids Res*. 2016;44(6):2613—2627.
233. O'Leary VB, Ovsepian SV, Carrascosa LG, et al. *PARTICLE*, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep*. 2015;11(3):474—485.
234. Huang MD, Chen WM, Qi FZ, et al. Long non-coding RNA TUG1 is up-regulated in hepatocellular carcinoma and promotes cell growth and apoptosis by epigenetically silencing of KLF_2. *Mol Cancer*. 2015;14:165.
235. Kretz M, Siprashvili Z, Chu C, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature*. 2013;493(7431):231—235.
236. Salameh A, Lee AK, Cardó-Vila M, et al. PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. *Proc Natl Acad Sci U S A*. 2015;112(27):8403—8408.
237. Chen J, Ke S, Zhong L, et al. Long noncoding RNA MALAT1 regulates generation of reactive oxygen species and the insulin responses in male mice. *Biochem Pharmacol*. 2018;152:94—103.
238. Li JH, Liu S, Zheng LL, et al. Discovery of protein-lncRNA interactions by integrating large-scale CLIP-seq and RNA-seq datasets. *Front Bioeng Biotechnol*. 2014;2:88.
239. Chu C, Zhang QC, da Rocha ST, et al. Systematic discovery of xist RNA binding proteins. *Cell*. 2015;161(2):404—416.
240. McHugh CA, Guttman M. RAP-MS: a method to identify proteins that interact directly with a specific RNA molecule in cells. *Methods Mol Biol*. 2018;1649:473—488.
241. Chu HP, Minajigi A, Chen Y, et al. iDRiP for the systematic discovery of proteins bound directly to noncoding RNA. *Nat Protoc*. 2021;16(7):3672—3694.
242. Yang YCT, Di C, Hu B, et al. CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genom*. 2015;16(1):51.
243. Cirillo D, Blanco M, Armaos A, et al. Quantitative predictions of protein interactions with long noncoding RNAs. *Nat Methods*. 2016;14(1):5—6.
244. Ge M, Li A, Wang M. A bipartite network-based method for prediction of long non-coding RNA-protein interactions. *Dev Reprod Biol*. 2016;14(1):62—71.
245. Hu H, Zhang L, Ai H, et al. HLPI-Ensemble: prediction of human lncRNA-protein interactions based on ensemble strategy. *RNA Biol*. 2018;15(6):797—806.
246. Agostini F, Cirillo D, Bolognesi B, et al. X-inactivation: quantitative predictions of protein interactions in the Xist network. *Nucleic Acids Res*. 2013;41(1):e31.
247. Yao Y, Jin S, Long H, et al. RNAe: an effective method for targeted protein translation enhancement by artificial non-coding RNA with SINEB2 repeat. *Nucleic Acids Res*. 2015;43(9):e58.
248. Kim TK, Hemberg M, Gray JM. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harbor Perspect Biol*. 2015;7(1):a018622.
249. Kim TK, Hemberg M, Gray JM, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010;465(7295):182—187.
250. Zhang Z, Lee JH, Ruan H, et al. Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat Commun*. 2019;10(1):4562.
251. Zhang Z, Hong W, Ruan H, et al. *HeRA*: an atlas of enhancer RNAs across human tissues. *Nucleic Acids Res*. 2021;49(D1):D932—D938.
252. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008;322(5909):1845—1848.
253. Mahat DB, Kwak H, Booth GT, et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc*. 2016;11(8):1455—1476.
254. Chu T, Rice EJ, Booth GT, et al. Chromatin Run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat Genet*. 2018;50(11):1553—1564.
255. Smith JP, Dutta AB, Sathyan KM, et al. PEPPRO: quality control and processing of nascent RNA profiling data. *Genome Biol*. 2021;22(1):155.
256. Au KF, Jiang H, Lin L, et al. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res*. 2010;38(14):4570—4578.
257. Wang K, Singh D, Zeng Z, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*. 2010;38(18):e178.
258. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*. 2010;26(7):873—881.
259. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15—21.
260. Kim D, Pertea G, Trapnell C, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36.
261. Li H. Minimap2:pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094—3100.
262. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37(23):4572—4574.
263. Liu B, Liu Y, Li J, et al. deSALT: fast and accurate long transcriptomic read alignment with de Bruijn graph-based index. *Genome Biol*. 2019;20(1):274.
264. Parker MT, Knop K, Barton GJ, et al. 2passtools: two-pass alignment using machine-learning-filtered splice junctions increases the accuracy of intron detection in long-read RNA sequencing. *Genome Biol*. 2021;22(1):72.
265. Li W, Jiang T. Transcriptome assembly and isoform expression level estimation from biased RNA-Seq reads. *Bioinformatics*. 2012;28(22):2914—2921.
266. Song L, Florea L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinf*. 2013;14(Suppl 5):S14.
267. Tomescu AI, Kuosmanen A, Rizzi R, et al. A novel Min-cost flow method for estimating transcript expression with RNA-Seq. *BMC Bioinf*. 2013;14(Suppl 5):S15.
268. Mezlini AM, Smith EJM, Fiume M, et al. iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data. *Genome Res*. 2013;23(3):519—529.
269. Maretty L, Sibbesen JA, Krogh A. Bayesian transcriptome assembly. *Genome Biol*. 2014;15(10):501.
270. Bernard E, Jacob L, Mairal J, et al. Efficient RNA isoform identification and quantification from RNA-Seq data with network flows. *Bioinformatics*. 2014;30(17):2447—2455.

271. Liu J, Yu T, Jiang T, et al. TransComb: genome-guided transcriptome assembly via combing junctions in splicing graphs. *Genome Biol*. 2016;17(1):213.
272. Song L, Sabunciyan S, Florea L. CLASS2:accurate and efficient splice variant annotation from RNA-seq reads. *Nucleic Acids Res*. 2016;44(10):e98.
273. Canzar S, Andreotti S, Weese D, et al. CIDANE: comprehensive isoform discovery and abundance estimation. *Genome Biol*. 2016;17:16.
274. Shao M, Kingsford C. Accurate assembly of transcripts through phase-preserving graph decomposition. *Nat Biotechnol*. 2017; 35(12):1167—1169.
275. Liu J, Liu X, Ren X, et al. scRNAss: a single-cell RNA-seq assembler via imputing dropouts and combing junctions. *Bioinformatics*. 2019;35(21):4264—4271.
276. Song L, Sabunciyan S, Yang G, et al. A multi-sample approach increases the accuracy of transcript assembly. *Nat Commun*. 2019;10(1):5000.
277. Shi X, Neuwald AF, Wang X, et al. IntAPT: integrated assembly of phenotype-specific transcripts from multiple RNA-seq profiles. *Bioinformatics*. 2021;37(5):650—658.
278. Zerbino DR, Birney E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18(5):821—829.
279. Namiki T, Hachiya T, Tanaka H, et al. MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Res*. 2012;40(20): e155.
280. Liu J, Li G, Chang Z, et al. BinPacker: packing-based *de novo* transcriptome assembly from RNA-seq data. *PLoS Comput Biol*. 2016;12(2):e1004772.
281. Hsieh MF, Lu CL, Tang CY. Clover: a clustering-oriented *de novo* assembler for Illumina sequences. *BMC Bioinf*. 2020; 21(1):528.
282. Shi X, Wang X, Neuwald AF, et al. A Bayesian approach for accurate *de novo* transcriptome assembly. *Sci Rep*. 2021; 11(1):17663.
283. Jia G, Wang X, Xiao G. A permutation-based non-parametric analysis of CRISPR screen data. *BMC Genom*. 2017;18(1):545.
284. Hua Q, Jin M, Mi B, et al. LINC01123, a c-Myc-activated long non-coding RNA, promotes proliferation and aerobic glycolysis of non-small cell lung cancer through miR-199a-5p/c-Myc axis. *J Hematol Oncol*. 2019;12(1):91.
285. Zhong Y, Yang L, Xiong F, et al. Long non-coding RNA AFAP1-AS1 accelerates lung cancer cells migration and invasion by interacting with $SNIP_1$ to upregulate c-Myc. *Signal Transduct Targeted Ther*. 2021;6(1):240.
286. Qian X, Yang J, Qiu Q, et al. LCAT3, a novel m6A-regulated long non-coding RNA, plays an oncogenic role in lung cancer via binding with $FUBP_1$ to activate c-MYC. *J Hematol Oncol*. 2021;14(1):112.
287. Peng Z, Wang J, Shan B, et al. The long noncoding RNA LINC00312 induces lung adenocarcinoma migration and vasculogenic mimicry through directly binding YBX1. *Mol Cancer*. 2018;17(1):167.
288. Kim J, Piao HL, Kim BJ, et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat Genet*. 2018;50(12): 1705—1715.
289. Hu Q, Ye Y, Chan LC, et al. Oncogenic lncRNA downregulates cancer cell antigen presentation and intrinsic tumor suppression. *Nat Immunol*. 2019;20(7):835—851.
290. Huang D, Chen J, Yang L, et al. NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. *Nat Immunol*. 2018;19(10):1112—1125.
291. Wang J, Xie S, Yang J, et al. The long noncoding RNA H19 promotes tamoxifen resistance in breast cancer via autophagy. *J Hematol Oncol*. 2019;12(1):81.
292. Wei Z, Chen L, Meng L, et al. LncRNA HOTAIR promotes the growth and metastasis of gastric cancer by sponging miR-1277-5p and upregulating COL5A1. *Gastric Cancer*. 2020; 23(6):1018—1032.
293. Zhang E, He X, Zhang C, et al. A novel long noncoding RNA HOXC-AS3 mediates tumorigenesis of gastric cancer by binding to YBX1. *Genome Biol*. 2018;19(1):154.
294. Hu Y, Yu Y, You S, et al. Long noncoding RNA MALAT1 regulates autophagy associated chemoresistance via miR-23b-3p sequestration in gastric cancer. *Mol Cancer*. 2017;16(1):174.
295. Wang Y, He L, Du Y, et al. The long noncoding RNA *lncTCF7* promotes self-renewal of human liver cancer stem cells through activation of Wnt signaling. *Cell Stem Cell*. 2015; 16(4):413—425.
296. Yuan JH, Yang F, Wang F, et al. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell*. 2014;25(5):666—681.
297. Jiang R, Tang J, Chen Y, et al. The long noncoding RNA lnc-EGFR stimulates T-regulatory cells differentiation thus promoting hepatocellular carcinoma immune evasion. *Nat Commun*. 2017;8:15129.
298. Guo T, Gong C, Wu P, et al. LINC00662 promotes hepatocellular carcinoma progression via altering genomic methylation profiles. *Cell Death Differ*. 2020;27(7):2191—2205.
299. Chen B, Dragomir MP, Fabris L, et al. The long noncoding RNA *CCAT2* induces chromosomal instability through BOP_1-AURKB signalin. *Gastroenterology*. 2020;159(6):2146—2162.
300. Pichler M, Rodriguez-Aguayo C, Nam SY, et al. Therapeutic potential of FLANC, a novel primate-specific long non-coding RNA in colorectal cancer. *Gut*. 2020;69(10):1818—1831.
301. Wen S, Wei Y, Zen C, et al. Long non-coding RNA NEAT1 promotes bone metastasis of prostate cancer through N6-methyladenosine. *Mol Cancer*. 2020;19(1):171.
302. Prensner JR, Iyer MK, Sahu A, et al. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet*. 2013;45(11): 1392—1398.
303. Chen C, Luo Y, He W, et al. Exosomal long noncoding RNA LNMAT2 promotes lymphatic metastasis in bladder cancer. *J Clin Invest*. 2020;130(1):404—421.
304. Zhan Y, Chen Z, He S, et al. Long non-coding RNA SOX2OT promotes the stemness phenotype of bladder cancer cells by modulating SOX2. *Mol Cancer*. 2020;19(1):25.
305. Papaioannou D, Petri A, Dovey OM, et al. The long non-coding RNA HOXB-AS3 regulates ribosomal RNA transcription in NPM_1-mutated acute myeloid leukemia. *Nat Commun*. 2019;10(1): 5351.
306. Ma Y, Guo G, Li T, et al. A novel imatinib-upregulated long noncoding RNA plays a critical role in inhibition of tumor growth induced by Abl oncogenes. *Mol Cancer*. 2022;21(1):5.
307. Li H, Wang X, Wen C, et al. Long noncoding RNA NORAD, a novel competing endogenous RNA, enhances the hypoxia-induced epithelial-mesenchymal transition to promote metastasis in pancreatic cancer. *Mol Cancer*. 2017;16(1):169.
308. Hu H, Wang Y, Ding X, et al. Long non-coding RNA $XLOC_{000647}$ suppresses progression of pancreatic cancer and decreases epithelial-mesenchymal transition-induced cell invasion by down-regulating NLRP3. *Mol Cancer*. 2018;17(1):18.
309. Du Z, Wen X, Wang Y, et al. Chromatin lncRNA Platr10 controls stem cell pluripotency by coordinating an intrachromosomal regulatory network. *Genome Biol*. 2021;22(1):233.